On fundamentals of models and sampling in compressed sensing

Bogdan Roman
Univ. of Cambridge

Alex Bastounis
Univ. of Cambridge

Ben Adcock
Simon Fraser Univ.

Anders C. Hansen
Univ. of Cambridge

Abstract

Sparsity is the traditional mainstay of the majority of compressed sensing. However, recent evidence suggests that in many applications where the sampling mechanism is fixed one does not recover all sparse vectors, but rather a much smaller subclass possessing greater structure. We therefore pose the following question: *What type of structured signals does compressed sensing actually recover?* To answer this question we provide an array of tests known as *flip tests* that aid one in determining the right structured sparsity model for a given problem. Only once the correct model has been determined can one provide mathematics that accurately describe the recovery properties of compressed sensing. In the case of wavelets (or, more generally, X-lets) we suggest a structured sparsity model that aligns with the corresponding flip tests, and demonstrate improved recovery guarantees.

The fact that such sampling mechanisms recover the sparsity structure common to many signals of interest, e.g. images, leads us to question the conceived compressed sensing wisdom of complete random, or generally, incoherent sampling (e.g. random sub-Gaussian, permuted Fourier, expanders etc.). In particular, we pose the following question: *In applications where the sampling mechanism can be chosen, does structured sampling outperform incoherent sampling?* We answer this question affirmatively for certain cases of interest.

In investigating these two themes, we find that there are many questions at the heart of compressed sensing that require us to revisit its fundamentals. Aiming to spur new research in this direction, we conclude this note with a list of open problems and topics for future research.

1 Introduction

Compressed sensing (CS) [8, 10, 11, 12] has led to an important shift in our perspective on the task of sampling and reconstruction of signals. Given a signal $x \in \mathbb{C}^N$ possessing a certain structure, often referred to as sparsity, and an appropriate sensing mechanism $A \in \mathbb{C}^{m \times N}$, CS states that one recover x from noisy measurements $y = Ax + e$ using far fewer samples than was traditionally thought necessary. Practically, this can be achieved by a number of recovery algorithms, including, for example the ℓ^1 minimization

$$\min_{z \in \mathbb{C}^n} \|z\|_1 \quad \text{s.t.} \quad \|Ax - y\|_2 \leq \delta, \tag{1.1}$$

where $\|e\|_2 \leq \delta$. The key to this is the assumed structure of x. We say that x is s-sparse if it has at most s nonzero entries, regardless of their locations. There are now a myriad of theoretical results describing sufficient (and sometimes necessary) conditions for recovering an s-sparse vector from measurements y by solving (1.1) or other appropriate algorithms.

There is however one mismatch: sparsity turns out to be an inaccurate model in many applications of CS including Magnetic Resonance Imaging (MRI), Computerised Tomography (CT), Electron Microscopy (EM), Fluorescence Microscopy (FM), Helium Atom Scattering (HAS), Radio Interferometry (RI), and Hadamard Spectroscopy (HS). More precisely, one does not actually recover all sparse vectors, but only a small subset of sparse vectors possessing substantial additional structure. Given the large body of work on the sparsity model this may come as a surprise. Yet, it is possible to observe this phenomenon through the so-called *flip test* [2, 23]. We recall this test in Section 2.1.

If sparsity is not the correct model for CS in many applications, one key question that arises is:

(1) *What kind of structured signals does one actually recover with CS?*

To address this general problem we suggest subdividing into two more basic questions:

(1.a) *Given a sampling mechanism, does one recover an arbitrary sparse signal?*

(1.b) *If not, what kind of sparsity structure does one recover?*

Question (1.b) is hard to answer. A typical trial and error approach would be to conjecture a structured sparsity model and then ask the following:

(1.c) *Given a sampling mechanism and a structured sparsity model, does one recover an arbitrary signal in this class?*

Striving towards an answer for (1) we introduce the *extended flip test* below. In particular, this is a numerical test that allows one to investigate questions (1.a) and (1.c) above.

In addition, in many applications of CS such as those listed above, the sampling mechanism is itself not just random, but also highly structured [2, 16, 17, 19, 21, 23, 25, 29], and so it is perhaps not too surprising that one recovers only signals with a particular structure. This raises a second fundamental question: if we were able to design the sampling mechanism (this is impossible in many applications, e.g. MRI, but we ask this as a basic question), should we follow standard practice and use incoherent sampling, e.g. sub-Gaussian random matrices, which can recover any s-sparse signal, or should we choose a sampling mechanism and strategy that recovers only structured sparse signals, such as natural images? In other words,

(2) *Does structured sampling outperform incoherent sampling?*

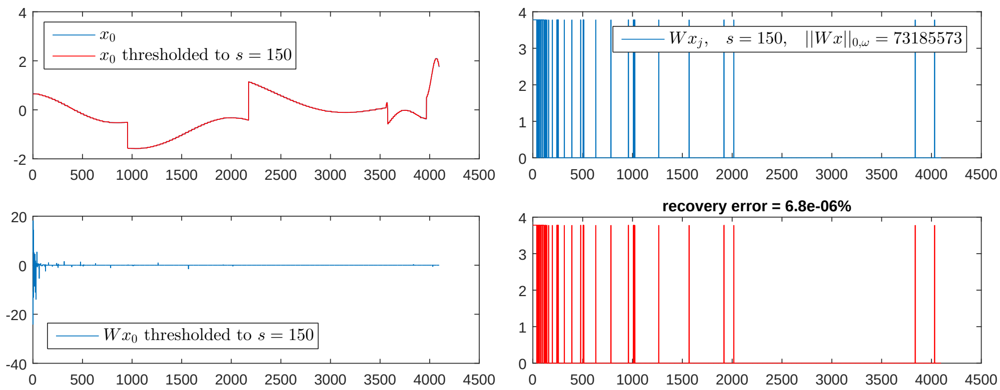
Section 3 endeavours to answer this question both numerically and mathematically, and the answer is: yes, provided the signal is structured. We thus conclude that incoherent sampling is indeed substantially less than optimal for recovering natural signals such as images.

2 Towards an accurate model: The generalized flip test

The original *flip test* [2, 23] was designed to answer questions (1.a) and (1.c) above. It showed that the standard Restricted Isometry Property (RIP) and sparsity model do not explain the recovery seen in the applications dealing with structured sampling operators and structured signals (virtually all applications listed above). This was done by flipping the sparsity basis coefficients, which completely preserves the signal sparsity, and then recovering this new signal using the same sampling operator and algorithm. We now generalize the flip test to allow for any sparsity model, any signal class, any sampling operator, and any recovery algorithm. The original flip test is a special case of this extended flip test. To show its applicability, we apply it to three existing sparsity models when considering Fourier and Hadamard measurements and X-lets as the sparsity representation: the *classic sparsity* model [8, 10], the *weighted sparsity* model [22], and the *sparsity in levels* model [2, 23]. As we shall see, in these cases, the third model is the most realistic.

The generalized flip test is as follows:
(i) Signal model. Decide on the type of signals of interest in, e.g. 1D piecewise smooth functions, 2D images, smooth functions, etc. Create a discrete vector x coming from the discretization of this desired signal model.
(ii) Sparse transform. Choose a sparsifying transform W such that Wx is sufficiently sparse. See more details below in step (vi).
(iii) Measurements and sampling strategy. Choose a measurement operator $A \in \mathbb{C}^{m\times N}$ e.g. Fourier, Hadamard, Gaussian, etc. For orthogonal operators, such as Fourier, Hadamard, DCT and others, also specify the subsampling strategy for selecting the m rows, e.g. uniform random, power law [17], half-half [10, 27], multilevel [2, 23] etc.
(iv) Recovery algorithm. Choose a recovery algorithm $\Delta : \mathbb{C}^m \to \mathbb{C}^N$. For example ℓ^1 minimization, i.e. solving $\min_{z\in\mathbb{C}^N} ||z||_1$ s.t. $AW^{-1}z = Ax$.
(v) Sparsity structure model. Choose the model of the sparsity structure to be tested. For example, the classic sparsity model [8, 10, 11, 12], weighted sparsity [22], sparsity in levels [2, 23], or another structured model.
(vi) The test. Perform the following experiment:

- Generate a signal x_0 from the model in (i) and obtain Wx_0 using the sparsifying transform from (ii), then threshold Wx_0 so that it is perfectly sparse (see next step for how to determine the threshold level). If the sparsity model from (v) does not depend on the magnitudes of the entries in Wx_0 (e.g. models such as sparsity, weighted sparsity and sparsity in levels) then set all the non-zero entries in Wx_0 to a positive constant α, i.e. obtain an x so that $(Wx)_i = \alpha$ for all $i \in \mathrm{supp}(Wx_0)$. This is to make all non-zeros equally important in order to avoid small coefficients giving false positives in the next step.
- Perform a reconstruction with measurements Ax using the sampling operator from (iii) and the recovery algorithm from (iv). If x is not recovered exactly (within a low tolerance) then decrease the thresholding level for Wx_0 in the above step to obtain a sparser x and repeat until x is recovered exactly.
- Create several new signals $x_1, x_2, ...$ such that Wx_j are in the same structured sparsity class from (iv) as Wx is. Specifically, ensure that all vectors Wx and Wx_j give the same value under the sparsity measure defined in the model from (iv), with non-zero entries set to some positive α if the sparsity model does not involve magnitudes. For example, for the classic sparsity model, all Wx_j must have the same number of non-zero entries as Wx and magnitudes equal to some positive α; for the weighted sparsity model, all Wx_j must have the same weighted ℓ^0 norm and non-zero entries set to some positive α; etc.
- Obtain a recovery for each x_j using the same operator A and algorithm Δ. Ensure the recovery is consistent by averaging over several trials if A entails any kind of randomization.



x_0 (top) and thresholded Wx_0 (bottom) $\quad$ Wx (top) and its ℓ^1 recovery (bottom)

Figure 1: Piecewise smooth 1D signal x_0 and recovery into Haar wavelets from $m = 1000$ Hadamard samples taken using a half-half scheme (first $m/2$ samples taken fully from the lower ordered rows, and the other $m/2$ uniformly at random from the remaining rows), which is known to be a good all round strategy [10, 27]. Here Wx was thresholded to $s = 150$ Haar coefficients.

(vii) The conclusions. First, if any of the recovery tests failed then the structured sparsity model chosen in (v) is not appropriate for the signal model, sparse transform, measurements and recovery algorithm chosen in (i)–(iv) respectively, and the conjectured model can be ruled out. Second, if the recovery of sufficiently many and different x_j is successful, this suggests that the structured sparsity model could be correct - though this is never a complete validation of the model.

We shall now test three existing sparsity models for structured 1D and 2D signals using structured operators. Figure 1 and Figure 2 show the signals x_0, Wx_0, Wx (thresholded) and their reconstructions via ℓ^1 minimization, which is the recovery algorithm we shall use.

2.1 The *classic sparsity* model and X-lets

The classic sparsity model is the first CS model [8, 10, 11, 12] and states that the location of the non-zero entries of Wx is not important, only the sparsity measure $s = |\operatorname{supp}(Wx)|$ is important. In this model, operators that satisfy the RIP with appropriate constant can recover all s-sparse vectors when using ℓ^1 minimization.

Definition 2.1 (Classic RIP). *A matrix $U \in \mathbb{C}^{m\times N}$ is said to satisfy the RIP of order s with constant δ_s if the following holds for all s-sparse vectors $y \in \mathbb{C}^N$*

$$(1-\delta_s)\|y\|_2^2 \le \|Uy\|_2^2 \le (1+\delta_s)\|y\|_2^2.$$

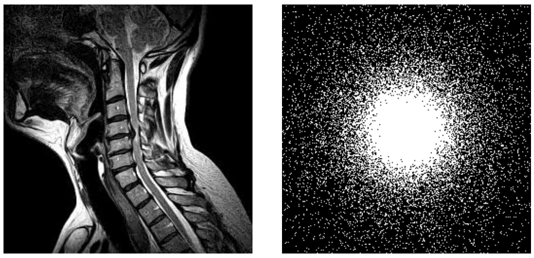
The flip test for this model is the original flip test introduced in [2]. Specifically, we take A to be the Fourier operator, W to be any discrete wavelet transform, x to be a natural image, and the new signal x_j is generated so that Wx_j is the flipped version of Wx, i.e. $(Wx_j)_k = (Wx)_{N-k+1}$, thus being in the same sparsity model as the sparsity measure is preserved since $|\operatorname{supp}(Wx)| = |\operatorname{supp}(Wx_j)| = s$.

Figures 3 and 4 show a generalized flip test for this model. The marked differences between the two recoveries demonstrates that the the classic sparsity model and RIP are not appropriate for this class of operators and signals. Simply put, one does not recover all s-sparse vectors, the location of the non-zero entries in Wx is actually important when dealing with structured operators and structured signals. The same conclusion was reached when we repeated this experiment for a large number of various natural images and combinations of Fourier, Hadamard and DCT measurements, see [2, 23].

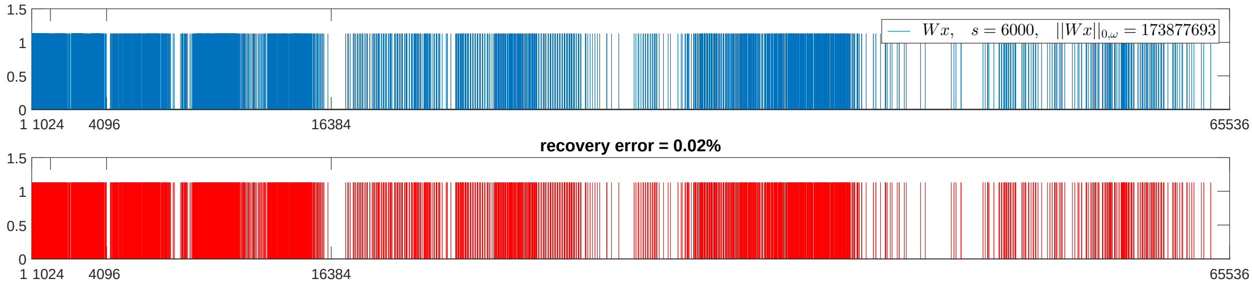
2.2 The *weighted sparsity* model and X-lets

The weighted sparsity model [22] extends the classic sparsity model by making some non-zero entries in Wx be more important than others by using weights. It is in this sense a more structured model than the classic sparsity model, and it similarly defines a weighted RIP. Just as with the classic RIP, matrices that satisfy the weighted RIP can recover all weighted s-sparse signals when using weighted l^1 minimization with matching weights [22].

Definition 2.2 (Weighted sparsity). *Given a vector $y \in \mathbb{C}^N$ and a vector of weights $\omega := (\omega_1, \omega_2, \ldots, \omega_N) \in \mathbb{R}^N$ with $\omega_j \ge 1$ for each j, we define the weighted ℓ_0 norm as $\|y\|_{\omega,0} := \sum_{j\in\operatorname{supp}(w)} \omega_j^2$, and y is said to be (ω, s)-sparse for some $s > 0$ if $\|y\|_{\omega,0} \le s$. We also define $\Sigma_{\omega,s}$ to be the set of all $y \in \mathbb{C}^N$ such that $\|y\|_{\omega,0} \le s$.*



2D signal x_0 and Fourier sampling pattern



Wx (top) and its ℓ^1 recovery (bottom)

Figure 2: Natural 2D signal at 256×256 resolution and recovery into Haar wavelets from $m = 18000$ Fourier samples taken using a power law $\sim (k_1 + k_2)^{-3/2}$ as in [17]. Here Wx was thresholded to $s = 6000$ Haar coefficients.

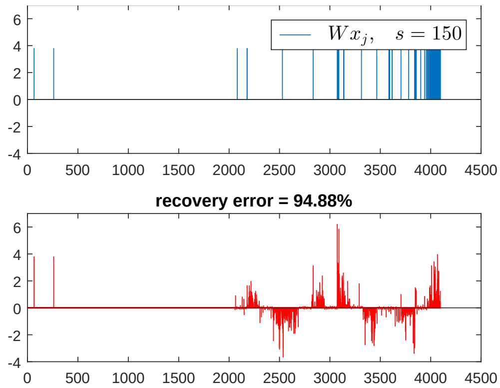


Figure 3: Flip test for the classic sparsity model with piecewise smooth 1D signals. Wx_j (top) and its ℓ^1 recovery (bottom). All test elements, except x_j, are identical to those used in Figure 1.

Definition 2.3 (Weighted RIP). *A matrix $U \in \mathbb{C}^{m\times N}$ is said to satisfy the weighted RIP of order (ω, s) with constant $\delta_{\omega,s}$ if the following holds for all (ω, s)-sparse vectors $y \in \mathbb{C}^N$:*

$$(1 - \delta_{\omega,s})\|y\|_2^2 \leq \|Uy\|_2^2 \leq (1 + \delta_{\omega,s})\|y\|_2^2$$

We shall now test whether the weighted sparsity model and the weighted RIP are appropriate for the structured signals and structured operators typically seen in practice. Let W be the discrete Haar wavelet transform. To be in the same model, the *flipped* signals Wx_j must preserve the weighted ℓ_0 norm of the original signal Wx for the given weights ω i.e. $||Wx||_{\omega,0} = ||Wx_j||_{\omega,0}$. The weights ω_i were chosen to be equal within each wavelet scale, but different across scales, which follows the wisdom of weighted ℓ^1 recovery [22]. Here we chose $\omega_i = 2^{\lceil \log_2 i \rceil}, i = 1, 2, \ldots$, however we get similar results if $\omega_i = 2^{\alpha \log_2 i}$ with $0 < \alpha < 1$.

Figure 5 and Figure 6 show generalized flip tests for this model. It is clear that a similar conclusion as with the classic sparsity model can be drawn: the weighted sparsity model and the weighted RIP do not hold for structured operators and structured signals of this kind. The weighted sparse class is still too big, i.e. it contains too many vectors that are unrealistic to recover. The test results can also be explained mathematically, see Theorem (2.7) which shows how the weighted sparsity class is indeed too big. We note that we also performed the recovery using weighted ℓ^1 minimization and the result is the same.

Remark 2.4 (Weighted sparsity for other models) The fact that weighted sparsity may not be the right model for wavelets or other X-lets does not mean that it is not accurate for other models. In [22] it was shown that weighted

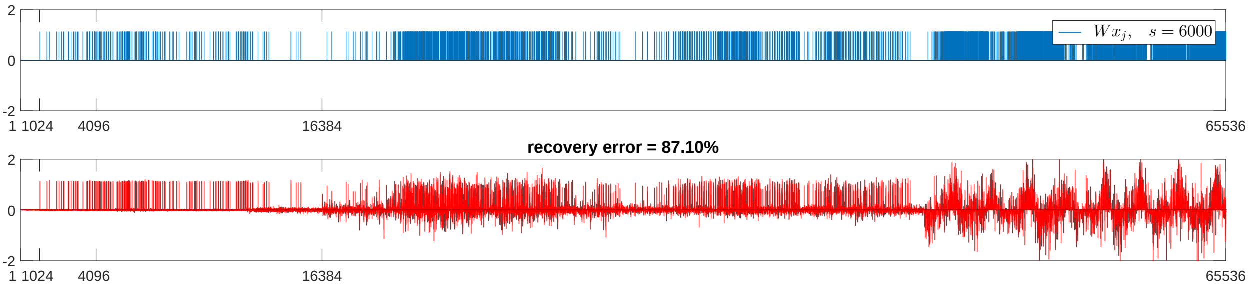


Figure 4: Flip test for the classic sparsity model with natural 2D signals. Wx_j (top) and its ℓ^1 recovery (bottom). All test elements, except x_j, are identical to those used in Figure 2.

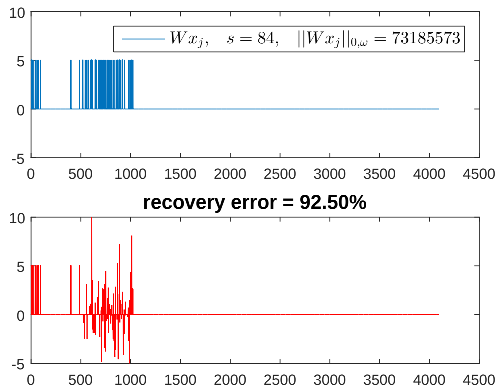


Figure 5: Flip test for the weighted sparsity model with piecewise smooth 1D signals. Wx_j (top) and its ℓ^1 recovery (bottom). All test components, except x_j, are identical to those used in Figure 1.

sparsity is an accurate model for pointwise sampling of smooth functions using polynomials as the sparse representation. This emphasises one of the main messages of the extended flip test: different sampling and recovery bases yield different structured sparsity.

2.3 The *sparsity in levels* model and X-lets

Another structured sparsity model is sparsity in levels [2, 23]. It is motivated by the fact that any X-lets have a particular level structure according to their scales. As seen below, this model defines a vector $\{s_k\}$ of local sparsities and level boundaries $\{M_k\}$ in order to capture sparsity in a more refined manner. As such, it is expected to contain a smaller class of signals. This model also has its own RIP variant.

Definition 2.5 (Sparsity in levels)**.** *Let $y \in \mathbb{C}^N$. For $r \in \mathbb{N}$ let $\mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r$ and $\mathbf{s} = (s_1, \ldots, s_r) \in \mathbb{N}^r$, with $s_k \leq M_k - M_{k-1}$, $k = 1, \ldots, r$, where $M_0 = 0$. We say that y is $(\mathbf{s}, \mathbf{M})$-sparse if, for each $k = 1, \ldots, r$ we have $|\Delta_k| \leq s_k$, where*

$$\Delta_k := \operatorname{supp}(y) \cap \{M_{k-1} + 1, \ldots, M_k\}.$$

We write $\Sigma_{\mathbf{s},\mathbf{M}}$ for the set of $(\mathbf{s}, \mathbf{M})$-sparse vectors and define the best $(\mathbf{s}, \mathbf{M})$-term approximation as

$$\sigma_{\mathbf{s},\mathbf{M}}(y)_1 = \min_{z \in \Sigma_{\mathbf{s},\mathbf{M}}} \|y - z\|_1.$$

Definition 2.6 (RIP in Levels)**.** *Given an r-level sparsity pattern $(\mathbf{s}, \mathbf{M})$, where $M_r = N$, we say that the matrix $U \in \mathbb{C}^{m \times N}$ satisfies the* RIP in levels *(RIP$_L$) with RIP$_L$ constant $\delta_{\mathbf{s},\mathbf{M}} \geq 0$ if for all $y \in \Sigma_{\mathbf{s},\mathbf{M}}$ we have*

$$(1 - \delta_{\mathbf{s},\mathbf{M}})\|y\|_2^2 \leq \|Uy\|_2^2 \leq (1 + \delta_{\mathbf{s},\mathbf{M}})\|y\|_2^2.$$

Similarly to the other RIP concepts the RIP in levels implies recovery of all $(\mathbf{s}, \mathbf{M})$ sparse vectors [4, 6]. We shall test this model in the same manner, with an interest towards structured operators and structured signals. Let the sparsity transform W be a wavelet transform and let the level boundaries $\mathbf{M}$ correspond to the wavelet scale boundaries. For this model, the *flipped* signals Wx_j must have the same $(\mathbf{s}, \mathbf{M})$ sparsity as Wx, i.e. the local sparsities and the level boundaries must be preserved. In other words, we can move coefficients within wavelet levels, but not across levels.

Figure 7 and Figure 8 show results of the flip test. We ran the same test for various other structured signals and structured operators and the results were consistent. As suggested by the results, sparsity in levels seems to be a

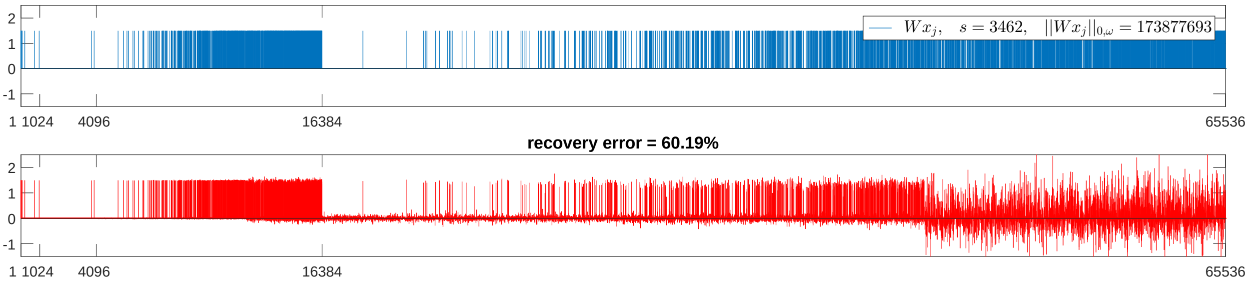


Figure 6: Flip test for the weighted sparsity model with natural 2D signals. Wx_j (top) and its ℓ^1 recovery (bottom). All test elements, except x_j, are identical to those used in Figure 2.

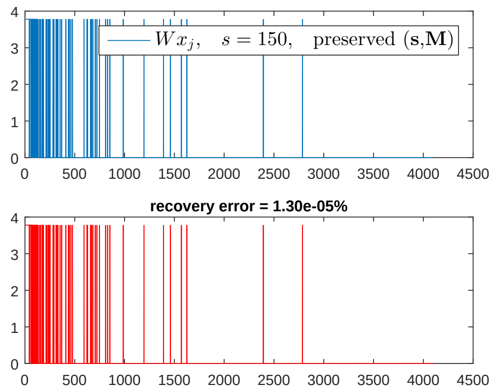


Figure 7: Flip test for the sparsity in levels model with piecewise smooth 1D signals. Wx_j (top) and its ℓ^1 recovery (bottom). All test components, except x_j, are identical to those used in Figure 1.

class that is actually recovered. As previously stated though, the flip test cannot guarantee that this is true for the entire class, since the flip test cannot entirely prove a model correct (unless it tests all signals in the class, which is infeasible). However, the variety of structured signals and structured operators we tested with success does provide reassurance that this model is more realistic than the classic sparsity and weighted sparsity models. Theoretically, the following theorem shows why the weighted sparse model yields a class of signals that is too big in comparison with sparsity in levels:

Theorem 2.7. *Let* $(\mathbf{s}, \mathbf{M})$ *have* r *levels and fix a nonempty* $L \subset \{1, 2, \ldots, r\}$*. Suppose that the collection of* $(\mathbf{s}, \mathbf{M})$*-sparse vectors, denoted by* $\Sigma_{\mathbf{s},\mathbf{M}}$*, is a subset of* $\Sigma_{\omega,X}$ *for some* ω, X*. Then there is an* $l_0 \in L$ *such that* $\Sigma_{\tilde{\mathbf{s}},\mathbf{M}} \subset \Sigma_{\omega,X}$*, where* $\tilde{\mathbf{s}} = (\tilde{s}_1, \tilde{s}_2, \ldots, \tilde{s}_r)$ *and*

$$\tilde{s}_i = \begin{cases} s_i & \text{if } i \in L^c \\ K & \text{if } i = l_0 \\ 0 & \text{otherwise}, \end{cases} \qquad K = \min\{|L|s_{l_0}, M_{l_0} - M_{l_0-1}\}$$

A direct consequence of this theorem is the following corollary that describes why the weighted sparsity class may be too big for wavelets. See Remark 2.9 for a thorough discussion. Note that this result is independent of the weights chosen, thus, changing the weights will not solve the problem.

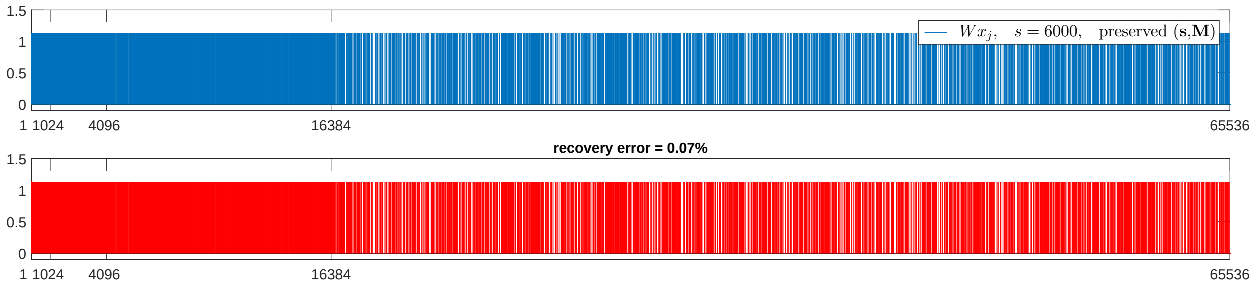


Figure 8: Flip test for the sparsity in levels model with natural 2D signals. Wx_j (top) and its ℓ^1 recovery (bottom). All test elements, except x_j, are identical to those used in Figure 2.

Figure 9: A demonstration of the near block diagonality of typical CS matrices. From left to right, the images represent the absolute values of Fourier to DB2, Fourier to DB10 and Hadmard to Haar matrices. Black corresponds to the value 1 and white to the value 0.

Corollary 2.8. *Let* $(\mathbf{s}, \mathbf{M})$ *have* r *levels and fix* $l \in \{0, \ldots, r-1\}$. *Suppose that* $\Sigma_{\mathbf{s},\mathbf{M}} \subset \Sigma_{\omega,X}$ *for some* ω, X. *Then there is an* l_0 *with* $l < l_0 \leq r$ *such that* $\Sigma_{\tilde{\mathbf{s}},\mathbf{M}} \subset \Sigma_{\omega,X}$, *where*

$$\tilde{\mathbf{s}} = (\underbrace{s_1, s_2, \ldots, s_l}_{l}, \underbrace{0, 0, \ldots, 0}_{l_0-1-l}, K, 0, \ldots, 0), \qquad K = \min\{(r-l)s_{l_0}, M_{l_0} - M_{l_0-1}\}.$$

Remark 2.9 (The sparse and weighted sparse classes are too big for wavelets) Note that the Fourier to wavelet matrix $U = U_{\text{dft}} V_{\text{dwt}}^{-1} \in \mathbb{C}^{N\times N}$, where U_{dft} denotes the discrete Fourier transform and V_{dwt} denotes the discrete wavelet, transform is almost block diagonal, see Figure 9 as well as [4]. This phenomenon also happens with the Hadamard to wavelet matrix. Hence, to simplify our explanation we will assume that U is block diagonal with levels dictated by the wavelet scales.

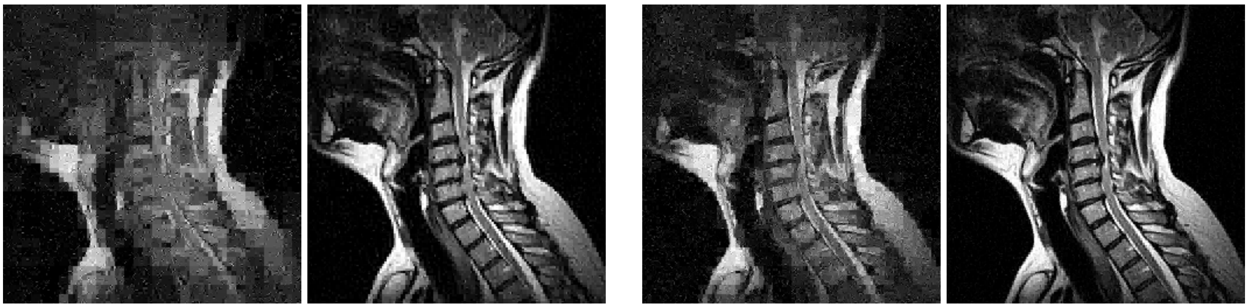
Assuming this block diagonality, any successful sampling scheme that can recover the wavelet coefficients must have in the kth level a set of samples Ω_k that is proportional in size to the number of important coefficients s_k in the kth wavelet level. What Corollary 2.8 states is that there will always exist an l_0 such that the set of weighted sparse vectors includes a vector that has $(r-l)s_{l_0}$ coefficients in the l_0th level. Hence, if Ω_{l_0} was a set of sampling points that were sufficient to recover a set of s_{l_0} coefficients in the l_0th level and no more, then if one instead assumes a weighted sparse model one would end up needing on the order of $(r-l)|\Omega|_{l_0}$ samples to recover the $(r-l)s_{l_0}$ possible coefficients in the l_0th level in this model. In particular, $|\Omega_{l_0}|$ is not large enough and this is why the flip test fails for the weighted sparse model. This 'cramming of coefficients' phenomenon described by Corollary 2.8 is illustrated in Figure 5. Note that such cramming can be completely controlled by the parameter l and can essentially occur in any levels of the wavelet coefficients. Moreover, due to the factor $r-l$, the amount of cramming allowed for weighted sparse vectors becomes worse with the number of levels. In other words, Corollary 2.8 reveals precisely why the weighted sparsity class is too large, and therefore fails as a model to describe this a setup.

Remark 2.10 (Sparsity in levels may not be the right model for polynomials) Just as weighted sparsity does not capture the essence of the level structure with wavelets given natural signals, sparsity in levels may not be the correct structured sparsity model for, say, polynomials. This underlines one of the important messages of the extended flip test, namely, different setups will yield different structured sparsity models. The key is to find the appropriate models.

3 Does structured sampling outperform incoherent sampling?

Being interested in structured signals, and having discussed sparsity models, we turn our attention to problems where one has the freedom to design the sampling mechanism, such as the single pixel camera [28], lenseless camera [15] or fluorescence microscopy [27, 23], which can implement either incoherent matrices (e.g. random sub-Gaussian, expanders etc.) or structured matrices (e.g. Hadamard or DCT).

Can we outperform incoherent sampling? At first that appears to be difficult as one typically needs $m \gtrsim s\log(N)$ samples to recover all s-sparse vectors from incoherent measurements. This bound is optimal for recovering sparse vectors so it seems hard to believe that one can do better. However, the context changes when we restrict the class of s-sparse signals to signals that have substantially more structure, such as natural images. As shown empirically and discussed in [23], when dealing with such signals, using a variable density sampling procedure termed multilevel (defined below) and either Fourier or Hadamard measurements one can substantially outperform sampling with incoherent matrices whenever the sparsifying transform consists of wavelets, X-lets or total variation. In fact, [23] showed that one can substantially outperform incoherent sampling even when state of the art structured recovery algorithms are used (e.g. Model-based [5], TurboAMP [24], Bayesian CS [14] etc.), simply by using structured sampling and



12.5% Gaussian | 12.5% Multilevel DFT | 25% Gaussian | 25% Multilevel DFT

Figure 10: Comparison between incoherent sampling and structured sampling, recovered into Haar at 256×256.

standard ℓ^1 recovery. See [23] for examples and in-depth discussion and comparison of structured versus incoherent sampling.

While [23] provides an in-depth investigation of this phenomenon, here we provide the first theoretical results which seek to it mathematically. We commence by defining the sampling scheme.

Definition 3.1 (Multilevel random sampling)**.** *Let $r \in \mathbb{N}$, $\mathbf{N} = (N_1, \ldots, N_r) \in \mathbb{N}^r$ with $1 \le N_1 < \ldots < N_r$, $\mathbf{m} = (m_1, \ldots, m_r) \in \mathbb{N}^r$, with $m_k \le N_k - N_{k-1}$, $k = 1, \ldots, r$, and suppose that $\Omega_k \subseteq \{N_{k-1}+1, \ldots, N_k\}$, $\quad |\Omega_k| = m_k$, $\quad k = 1, \ldots, r$, are chosen uniformly at random, where $N_0 = 0$. We refer to the set $\Omega = \Omega_{\mathbf{N},\mathbf{m}} = \Omega_1 \cup \ldots \cup \Omega_r$ as an $(\mathbf{N}, \mathbf{m})$-multilevel sampling scheme.*

First we define the discrete Fourier transform U_{dft}. Let $x = \{x(t)\}_{t=0}^{N-1} \in \mathbb{C}^N$ be a signal and the Fourier transform of x be $\mathcal{F}x(\omega) = N^{-1/2} \sum_{t=1}^{N} x(t) \mathrm{e}^{2\pi \mathrm{i} \omega t/N}$, with $\omega \in \mathbb{R}$, then write $F \in \mathbb{C}^{N \times N}$ for the corresponding matrix, so that $Fx = \{\mathcal{F}x(\omega)\}_{\omega=-N/2+1}^{N/2}$. We then let U_{dft} be the row permuted version of F where frequencies are reordered according to the bijection $\theta : \mathbb{Z} \to \mathbb{N}$ defined by $\theta(0) = 1$, $\theta(1) = 2$, $\theta(-1) = 3$ etc.

Theorem 3.2 (Fourier to Haar)**.** *Let $\epsilon \in (0, \mathrm{e}^{-1}]$ and $U = U_{\text{dft}} V_{\text{dwt}}^{-1} \in \mathbb{C}^{N \times N}$, where V_{dwt} denotes the discrete Haar transform. Let $x \in \mathbb{C}^N$. Suppose that $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ is a multilevel sampling scheme and $(\mathbf{s}, \mathbf{M})$ is a multilevel sparsity structure as described above where $\mathbf{M} = \mathbf{N}$ correspond to levels defined by the wavelet scales (where potentially several scales could be combined into one level). Moreover, suppose that $s_1 = M_1$ and $s_1 \le s_2$. If*

$$m_1 = M_1, \qquad m_j \gtrsim \left(s_j + \sum_{l=2, l \neq j}^{r} 2^{-\frac{|j-l|}{2}} s_l \right) \log(\epsilon^{-1}) \log(N), \quad j = 2, \ldots, r, \tag{3.1}$$

then any minimiser z of (1.1) with $A = P_\Omega U$ satisfies

$$\|z - x\| \le C \left(\delta \sqrt{D} (1 + E\sqrt{s}) + \sigma_{\mathbf{s},\mathbf{M}}(x) \right), \tag{3.2}$$

with probability exceeding $1 - s\epsilon$, where $s = s_1 + \ldots + s_r$, C is a universal constant, $D = 1 + \frac{\sqrt{\log_2(6\epsilon^{-1})}}{\log_2(4EN\sqrt{s})}$ and $E = \max_{j=1,\ldots,r}\{(N_j - N_{j-1})/m_j\}$. In particular, the total number of measurements $m = M_1 + m_2 + \ldots + m_r$ satisfies

$$m \ge M_1 + C\,(s_2 + \ldots + s_r) \log(\epsilon^{-1}) \log(N). \tag{3.3}$$

Remark 3.3 (Losing the log factor) The key point of this theorem is that we lose the log factors in the M_1 term. In particular, the number of measurements needed to recover the first M_1 coefficients is exactly equal to M_1, and is independent of the resolution N. Conversely, if one were to use an incoherent sampling strategy then the total number of measurements would be

$$m \ge C(M_1 + s_2 + \ldots + s_r) \log(\epsilon^{-1}) \log(N).$$

Note that the coarse scale wavelet coefficients typically carry much of the signal's energy. Hence accurate recovery of these coefficients using as few measurements as possible is important for a good reconstruction. The improved recovery properties of structured sampling, based on recovering such coefficients more efficiently than in the incoherent setting, is demonstrated theoretically by Theorem 3.2. An experiment verifying this fact is shown in Figure 10.

4 Open problems and proofs

The purpose of this note was to stimulate interest in revisiting the fundamentals of CS by demonstrating firstly that structured sparsity is what one actually recovers in many applications of CS, and secondly that structured sampling can bring significant performance gains in practice over incoherent sampling, even in scenarios where the sampling operator can be designed. We conclude with a list of open problems and challenges.

(i) **Structured sampling and structured sparsity.** Characterise the right structured sparsity models for different types of problems and different types of sparsity bases or frames (polynomials, wavelets [9], shearlets [18], curvelets [7], total variation [17, 20] etc.)

(ii) Provide a mathematical theory for the sampling strategies that recover structured sparse signals from (i). Note that different sampling strategies and different sparsity bases/frames may correspond to different structured sparsity models.

(iii) Provide sharp recovery guarantees for nonuniform recovery of *a* signal in the structured sparsity class.

(iv) Generalize standard compressed sensing tools such as the null space property and the restricted isometry property to the structured setting, and prove that the appropriate measurement matrices A satisfy both.

(v) Provide a mathematical theory to explain why structured sampling outperforms incoherent sampling even when structured recovery algorithms are used, as shown empirically in [23].

(vi) Improve structured sparsity models to take into account coefficient magnitudes.

(vii) **Deterministic sampling.** Some key CS applications do not allow for much randomness when sampling, yet deterministic sampling in those cases has been shown to work well in practice. Here one expects to recover only a small structured subset of the set of sparse vectors. However, most problems in these instances are highly structured as well, which gives an intuitive reason as to why this should work. Can we prove prove this mathematically?

(viii) Characterise the right structured sparsity models for different types of problems where the sampling procedure is deterministic.

(ix) **Outperforming ℓ^1 minimization.** The extended flip test shows that in many cases the signal class of interest is smaller than the structured sparsity class that is actually being recovered by ℓ^1 minimization. This therefore raises the question: can we design structured recovery algorithms that improve on the results obtained by structured sampling and standard ℓ^1 minimization? Do such algorithms exist when the sampling procedure is optimal for a particular problem? Essentially, we would like an algorithm capable of recovering smaller structured sparsity classes than what ℓ^1 recovers. For example, for natural images, ℓ^1 recovers a larger class, which includes many non-natural images, even with structured sampling. In other words, ℓ^1 seems suboptimal.

Note that some of these problems are inspired by some of the discussions in [26]. Also, the questions above are valid regardless if the model is finite or infinite-dimensional [1, 13].

4.1 Proofs of Theorems 2.7 and 3.2

Proof of Theorem 2.7. Let $i \in \{1, \dots, l\}$ and $n \leq M_i - M_{i-1}$. Define $\Omega_i(n)$ to be the set of indices of the n largest elements of the weights $\{\omega_{M_{i-1}+1}, \omega_{M_{i-1}+2}, \dots, \omega_{M_i}\}$. Furthermore, define

$$H_i(n) := \sum_{j \in \Omega_i(n)} w_j^2, \quad X' := X - \sum_{i \in L^c} H_i(s_i)$$

and

$$R_i(n) := \frac{X'}{H_i(n)},$$

where we recall that X stems from $\Sigma_{\omega,X}$ in particular $\|x\|_{\omega,0} \leq X$ for all $x \in \Sigma_{\omega,X}$. From these definitions, it is easy to see that $H_i(m)\frac{n}{m} \geq H_i(n)$ for $n \geq m$. Indeed,

$$\begin{aligned} H_i(n) &\leq H_i(m) + (n-m) \min_{j \in \Omega_i(m)} w_j^2 \\ &\leq H_i(m) + \frac{(n-m)H_i(m)}{m} = \frac{n}{m} H_i(m) \end{aligned} \tag{4.1}$$

By the conditions of the theorem, we have

$$\sum_{i \in L} H_i(s_i) \leq X' \tag{4.2}$$

since $\Sigma_{\mathbf{s},\mathbf{M}} \subset \Sigma_{\mathbf{w},X}$ and so $\sum_{i=1}^{l} H_i(s_i) \leq X$, yielding (4.2). Define, for $m \leq M_j - M_{j-1}$, the vector $\tilde{s}^j(m) := (\tilde{s}^j_1, \tilde{s}^j_2, \dots, \tilde{s}^j_l)$, where

$$\tilde{s}^j_i = \begin{cases} s_i & \text{if } i \in L^c \\ m & \text{if } i = j \\ 0 & \text{otherwise} \end{cases}$$

it is clear that $\Sigma_{\tilde{\mathbf{s}}^{\mathbf{j}}(m),\mathbf{M}} \subset \Sigma_{\omega,X}$ whenever $R_j(m) \geq 1$, because if x is $(\tilde{\mathbf{s}}^{\mathbf{j}}(m), \mathbf{M})$-sparse, then $\|x\|_{\omega,0} \leq H_j(m) + (X - X') \leq X$ by the definition of $R_j(m)$. Finally, by using (4.1) and (4.2) (in the first inequality below) then

$$\max_{j \in L} R_j(ls_j) \geq \max_{j \in L} \frac{\sum\limits_{i \in L} H_i(s_i)}{|L| H_j(s_j)} \geq \frac{\min\limits_{i \in L} |L| H_i(s_i)}{\min\limits_{j \in L} |L| H_j(s_j)} = 1.$$

Since the maximum on the left hand side is taken over a finite set, it is attained at some l_0, and the theorem follows. □

Proof of Theorem 3.2. The proof is based on an abstract theorem from [2] as well as some estimates from [3]. Before we can apply these we need to recall some basics from [2].

Definition 4.1 (Relative sparsity)**.** *Let U be an isometry of either $\mathbb{C}^N$ or $l^2(\mathbb{N})$. For $\mathbf{N} = (N_1, \dots, N_r) \in \mathbb{N}^r$, $\mathbf{M} = (M_1, \dots, M_r) \in \mathbb{N}^r$ with $1 \leq N_1 < \dots < N_r$ and $1 \leq M_1 < \dots < M_r$, $\mathbf{s} = (s_1, \dots, s_r) \in \mathbb{N}^r$ and $1 \leq k \leq r$, the k^{th} relative sparsity is given by $S_k = S_k(\mathbf{N}, \mathbf{M}, \mathbf{s}) = \max_{\eta \in \Theta} \|P_{N_k}^{N_{k-1}} U \eta\|^2$, where $N_0 = M_0 = 0$ and Θ is the set*

$$\Theta = \{\eta : \|\eta\|_{l^\infty} \leq 1, |\mathrm{supp}(P_{M_l}^{M_{l-1}} \eta)| = s_l,\ l = 1, \dots, r\}.$$

Relative sparsity is crucial for controlling so-called interference between different levels. A key in the proof is to be able to estimate the $S_k(\mathbf{N}, \mathbf{M}, \mathbf{s})$.

Definition 4.2 (Local coherence)**.** *Let U be an isometry of either $\mathbb{C}^N$ or $l^2(\mathbb{N})$. If $\mathbf{N} = (N_1, \dots, N_r) \in \mathbb{N}^r$ and $\mathbf{M} = (M_1, \dots, M_r) \in \mathbb{N}^r$ with $1 \leq N_1 < \dots N_r$ and $1 \leq M_1 < \dots < M_r$ the $(k,l)^{\text{th}}$ local coherence of U with respect to $\mathbf{N}$ and $\mathbf{M}$ is given by*

$$\mu_{\mathbf{N},\mathbf{M}}(k,l) = \sqrt{\mu(P_{N_k}^{N_{k-1}} U P_{M_l}^{M_{l-1}}) \cdot \mu(P_{N_k}^{N_{k-1}} U)}, \quad k, l = 1, \dots, r,$$

where $N_0 = M_0 = 0$ and P_b^a denotes the projection matrix corresponding to indices $\{a+1, \dots, b\}$.

To estimate the relative sparsity and local coherence in the Fourier to Haar case we will rely on some previous calculations done in [3], and thus we need to make a small adjustment to be able to use the results directly. Note that, due to the definition of U_{dft} and the multilevel sampling corresponding to the levels our sampling is equivalent to sampling the classical discrete Fourier transform F in the following way. Recall that the rows of F are indexed over $\{-n/2+1, \dots, n/2\}$. We divide this set into r frequency bands. Let $W_0 = \{0, 1\}$, and

$$W_j = \{-2^j + 1, \dots, -2^{j-1}\} \cup \{2^{j-1} + 1, \dots, 2^j\}, \quad j = 1, \dots, r-1, \tag{4.3}$$

and note that $W_0, \dots, W_{r-1}$ form a disjoint partition of $\{-n/2+1, \dots, n/2\}$. Observe that

$$|W_0| = 2, \qquad |W_j| = 2^j, \quad j = 1, \dots, r-1.$$

For $j = 0, \dots, r-1$, we now choose the index set $\Omega_j \subseteq W_j$ uniformly at random of size $|\Omega_j| = m_j$, and finally we have

$$\Omega = \Omega_0 \cup \dots \cup \Omega_{r-1}, \qquad |\Omega| = m = m_0 + \dots + m_{r-1}. \tag{4.4}$$

Since this form of sampling the matrix F is equivalent to the way of sampling U_{dft} described in the theorem we can continue with estimates of the relative sparsity and local coherence with respect to FV_{dwt}^{-1} that were proved in [3]. More precisely we have

$$S_j \lesssim \sum_{l=0}^{r-1} 2^{-|j-l|/2} s_{l+1}, \qquad j = 1, \dots, r, \tag{4.5}$$

and

$$\mu_{\mathbf{N},\mathbf{M}}(j,l) \lesssim 2^{-(j-1)} 2^{-|j-l|/2}, \quad j, l = 1, \dots, r. \tag{4.6}$$

We can now insert these estimates into one of the main theorems from [2] to prove our theorems. We use the following result from [2]. In particular, let $U \in \mathbb{C}^{N \times N}$ is an isometry, $x \in \mathbb{C}^N$, $\Omega = \Omega_{\mathbf{N},\mathbf{m}}$ is a multilevel sampling scheme,

$(\mathbf{s}, \mathbf{M})$ a multilevel sparsity structure, where $\mathbf{M} = (M_1, \ldots, M_r) \in \mathbb{N}^r$, $M_r = N$, and $\mathbf{s} = (s_1, \ldots, s_r) \in \mathbb{N}^r$, such that the following holds: for $\epsilon \in (0, e^{-1}]$ and $1 \leq k \leq r$,

$$\begin{aligned} m_k &= \min\{N_k - N_{k-1}, \tilde{m}_k\} \\ \tilde{m}_k &\gtrsim (N_k - N_{k-1}) \cdot \log(\epsilon^{-1}) \cdot \left(\sum_{l=1}^{r} \mu_{\mathbf{N},\mathbf{M}}(k, l) \cdot s_l \right) \cdot \log(N), \end{aligned} \tag{4.7}$$

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

$$1 \gtrsim \sum_{k=1}^{r} \left(\frac{N_k - N_{k-1}}{\hat{m}_k} - 1 \right) \cdot \mu_{\mathbf{N},\mathbf{M}}(k, l) \cdot S_k, \quad \forall\, l = 1, \ldots, r, \tag{4.8}$$

and $N_0 = 0$. Then any minimiser z of (1.1) with $A = P_\Omega U$ satisfies

$$\|z - x\| \leq C \left(\delta \sqrt{D} (1 + E\sqrt{s}) + \sigma_{\mathbf{s},\mathbf{M}}(x) \right).$$

Hence, to prove our theorems we simply need to insert our previous estimates into (4.7) and (4.8). More precisely, to prove Theorem 3.2 we get that (4.7) and (4.6) yield the following conditions on m_j:

$$\begin{aligned} m_j &= \min\{N_j - N_{j-1}, \tilde{m}_j\} \\ \tilde{m}_j &\gtrsim \left(s_j + \sum_{\substack{l=1 \\ l \neq j}}^{r} 2^{-\frac{|j-l|}{2}} s_l \right) \log(\epsilon^{-1}) \log(N), \quad j = 1, \ldots, r. \end{aligned} \tag{4.9}$$

Also, by using (4.8), (4.5) and (4.6) give the conditions

$$m_j \gtrsim \hat{m}_j \log(\epsilon^{-1}) \log(N), \quad 1 \gtrsim \sum_{j=2}^{r} \left(\frac{N_j - N_{j-1}}{\hat{m}_j} - 1 \right) 2^{-(j-1)} 2^{-|j-l|/2} \sum_{k=1}^{r-1} 2^{-|j-k|/2} s_{k+1}, \tag{4.10}$$

for $j = 2, \ldots, r$. Using the assumption of the theorem and the facts that

$$N_j - N_{j-1} = 2^{-(j-1)}, \quad ((N_j - N_{j-1})/\hat{m}_j - 1) \leq (N_j - N_{j-1})/\hat{m}_j,$$

and that

$$\sum_{j=2}^{r} 2^{-|j-l|/2} \lesssim 1, \qquad l = 1, \ldots, r,$$

then (4.10) is satisfied if m_j satisfies (4.9). As (3.1) obviously implies (4.9) the theorem is proved. □

References

[1] B. Adcock and A. C. Hansen. Generalized sampling and infinite-dimensional compressed sensing. *Found. Comp. Math.*, (to appear).

[2] B. Adcock, A. C. Hansen, C. Poon, and B. Roman. Breaking the coherence barrier: A new theory for compressed sensing. *arXiv:1302.0561*, 2014.

[3] B. Adcock, A. C. Hansen, and B. Roman. A note on compressed sensing of structured sparse wavelet coefficients from subsampled Fourier measurements. *arXiv:1403.6541*, 2014.

[4] B. Adcock, A. C. Hansen, and B. Roman. The quest for optimal sampling: computationally efficient, structure-exploiting measurements for compressed sensing. In *Compressed Sensing and Its Applications (to appear)*. Springer, 2014.

[5] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hedge. Model-based compressive sensing. *IEEE T Inf Theory*, 56(4):1982–2001, 2010.

[6] A. Bastounis and A. C. Hansen. On the absence of the rip in real-world applications of compressed sensing and the rip in levels. *CoRR*, abs/1411.4449, 2014.

[7] E. J. Candès and D. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. *Comm. Pure Appl. Math.*, 57(2):219–266, 2004.

[8] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE T Inf Theory*, 52(2):489–509, 2006.

[9] I. Daubechies. Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 41(7):909—996, 1988.

[10] D. L. Donoho. Compressed sensing. *IEEE T Inf Theory*, 52(4):1289–1306, 2006.

[11] Y. C. Eldar and G. Kutyniok, editors. *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[12] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Birkhäuser, 2013.

[13] M. Guerquin-Kern, M. Häberlin, K. P. Pruessmann, and M. Unser. A fast wavelet-based reconstruction method for Magnetic Resonance Imaging. *IEEE Trans. Med. Imaging*, 30(9):1649–1660, 2011.

[14] L. He and L. Carin. Exploiting structure in wavelet-based Bayesian compressive sensing. *IEEE T Signal Proc*, 57(9):3488–3497, 2009.

[15] G. Huang, H. Jiang, K. Matthews, and P. A. Wilford. Lensless imaging by compressive sensing. In *IEEE Intl Conf Image Proc,*, pages 2101–2105, 2013.

[16] F. Krahmer and H. Rauhut. Structured random measurements in signal processing. *arXiv:1401.1106v2*, 2014.

[17] F. Krahmer and R. Ward. Stable and robust recovery from variable density frequency samples. *IEEE Trans. Image Proc. (to appear)*, 2014.

[18] G. Kutyniok, J. Lemvig, and W.-Q. Lim. Compactly supported shearlets. In M. Neamtu and L. Schumaker, editors, *Approximation Theory XIII: San Antonio 2010*, volume 13 of *Springer Proceedings in Mathematics*, pages 163–186. Springer New York, 2012.

[19] M. Lustig, D. L. Donoho, and J. M. Pauly. Sparse MRI: the application of compressed sensing for rapid MRI imaging. *Magn. Reson. Imaging*, 58(6):1182–1195, 2007.

[20] C. Poon. On the role of total variation in compressed sensing. *SIAM J Imaging Sci*, (to appear).

[21] G. Puy, P. Vandergheynst, and Y. Wiaux. On variable density compressive sampling. *IEEE Sig Proc Lett*, 18:595–598, 2011.

[22] H. Rauhut and R. Ward. Interpolation via weighted l^1 minimization. *arXiv:1401.1106v2*, 2014.

[23] B. Roman, B. Adcock, and A. Hansen. On asymptotic structure in compressed sensing. *arXiv:1406.4178*, 2014.

[24] S. Som and P. Schniter. Compressive imaging using approximate message passing and a markov-tree prior. *IEEE T Signal Proc*, 60(7):3439–3448, 2012.

[25] A. F. Stalder, M. Schmidt, H. H. Quick, M. Schlamann, S. Maderwald, P. Schmitt, Q. Wang, M. S. Nadar, and M. O. Zenge. Highly undersampled contrast-enhanced mra with iterative reconstruction: Integration in a clinical setting. *Magnetic Resonance in Medicine*, pages n/a–n/a, 2014.

[26] T. Strohmer. Measure what should be measured: progress and challenges in compressive sensing. *IEEE Signal Process. Letters*, 19(12):887–893, 2012.

[27] V. Studer, J. Bobin, M. Chahid, H. Moussavi, E. Candès, and M. Dahan. Compressive fluorescence microscopy for biological and hyperspectral imaging. *Natl Acad Sci USA*, 109(26):1679–1687, 2011.

[28] D. Takhar, J. N. Laska, M. B. Wakin, M. F. Duarte, D. Baron, S. Sarvotham, K. F. Kelly, and R. G. Baraniuk. A new compressive imaging camera architecture using optical-domain compression. In *Comput Imaging IV at SPIE Electronic Imaging*, pages 43–52, 2006.

[29] Q. Wang, M. Zenge, H. E. Cetingul, E. Mueller, and M. S. Nadar. Novel sampling strategies for sparse mr image reconstruction. In *Proc. Intl. Soc. Mag. Reson. Med., ISMRM'14*, volume 22, page 1549, 2014.