

On the Absence of Uniform Recovery in Many Real-World Applications of Compressed Sensing and the Restricted Isometry Property and Nullspace Property in Levels*

Alexander Bastounis[†] and Anders C. Hansen[‡]

Abstract. The purpose of this paper is twofold. The first is to point out that the property of uniform recovery, meaning that all sparse vectors are recovered, does not hold in many applications where compressed sensing is successfully used. This includes fields like magnetic resonance imaging (MRI), nuclear magnetic resonance computerized tomography, electron tomography, radio interferometry, helium atom scattering, and fluorescence microscopy. We demonstrate that for natural compressed sensing matrices involving a level based reconstruction basis (e.g., wavelets), the number of measurements required to recover all s -sparse signals for reasonable s is excessive. In particular, uniform recovery of all s -sparse signals is quite unrealistic. This realization explains why the restricted isometry property (RIP) is insufficient for explaining the success of compressed sensing in various practical applications. The second purpose of the paper is to introduce a new framework based on a generalized RIP and a generalized nullspace property that fit the applications where compressed sensing is used. We demonstrate that the shortcomings previously used to prove that uniform recovery is unreasonable no longer apply if we instead ask for structured recovery that is uniform only within each of the levels. To examine this phenomenon, a new tool, termed the “restricted isometry property in levels” (RIP_L) is described and analyzed. Furthermore, we show that with certain conditions on the RIP_L , a form of uniform recovery within each level is possible. Fortunately, recent theoretical advances made by Li and Adcock demonstrate the existence of large classes of matrices that satisfy the RIP_L . Moreover, such matrices are used extensively in applications such as MRI. Finally, we conclude the paper by providing examples that demonstrate the optimality of the results obtained.

Key words. structured compressed sensing, restricted isometry property in levels, structured sampling, flip test

AMS subject classifications. 94A08, 94A20, 42C40, 65R32, 92C55

DOI. 10.1137/15M1043972

1. Introduction. Compressed sensing, introduced by Candès, Romberg, and Tao [18] and Donoho [30], has been one of the important new developments in applied mathematics in the last decade [9, 15, 24, 27, 31, 32, 33, 35, 42, 60]. By introducing a nonlinear reconstruction method via convex optimization and randomization in the sampling procedure, one can circumvent traditional sampling barriers when reconstructing vectors that are sparse or

*Received by the editors October 15, 2015; accepted for publication (in revised form) November 3, 2016; published electronically March 15, 2017.

<http://www.siam.org/journals/siims/10-1/M104397.html>

Funding: The work of the first author was supported by RCUK/Engineering and Physical Science Research Council (EPSRC) grant EP/H023348/1. The work of the second author was supported by EPSRC grant EP/L003457/1 and a Royal Society University research fellowship.

[†]CCA, Centre for Mathematical Sciences, University of Cambridge, Cambridge CB3 0WA, UK (A.Bastounis@maths.cam.ac.uk).

[‡]DAMTP, Centre for Mathematical Sciences, University of Cambridge, Cambridge CB3 0WA, UK, and Department of Mathematics, University of Oslo, 0316 Oslo, Norway (A.Hansen@damtp.cam.ac.uk).

compressible, meaning that they have few nonzero coefficients or can be approximated well by vectors with few nonzero coefficients.

Under certain conditions on the matrix U , every “ s -sparse” vector x (i.e., any vector x with at most s nonzero entries) can be recovered by observing the values of Ux and employing ℓ^1 minimization. Remarkably this can be achieved even if U is singular. If any s -sparse vector can be perfectly recovered in this way, we say that *uniform recovery of order s is possible*.

Given the recent substantial interest in uniform recovery it is natural to ask whether this intriguing mathematical concept is actually observed in many of the applications where compressed sensing is applied. Certain conditions on U , like the restricted isometry property (RIP) (see Definition 1.2) and the nullspace property of order s (see Definition 1.3) imply uniform recovery of order s . However, for general matrices U it is difficult to check that these properties hold. Indeed, it is shown in [62] that verifying that the RIP holds (and thus order s uniform recovery is possible) for general U is an NP hard problem. However, there is a simple test that can be used to show that certain matrices cannot achieve uniform recovery of order s for reasonable values of s . This is called the *flip test*. As this test reveals, there are a significant number of practical applications where uniform recovery is not the correct model for compressed sensing. This list of applications includes magnetic resonance imaging (MRI) [37, 52], other areas of medical imaging such as computerized tomography (CT) [23, 39], nuclear magnetic resonance (NMR) [43], electron tomography [36, 49], as well as other fields such as fluorescence microscopy [59, 61], surface scattering such as helium atom scattering [44], and radio interferometry [53].

We will thoroughly document the lack of uniform recovery of order s in this paper and explain why it does not hold for reasonable s in many applications. It is then natural to ask whether there might be an alternative to uniform recovery of order s that may be more suitable for the actual real-world compressed sensing applications. With this in mind, we shall generalize uniform recovery of order s to a level based uniform recovery, which we term *uniform recovery of order (\mathbf{s}, \mathbf{M})* . Numerical experiments will suggest that uniform recovery of order (\mathbf{s}, \mathbf{M}) is better suited to many of the applications where compressed sensing is used than uniform recovery of order s . We will extend the concepts of the RIP and the nullspace property to this setting with the introduction of the *RIP in levels* and the ℓ^2 robust nullspace property of order (\mathbf{s}, \mathbf{M}) .

1.1. Compressed sensing. We shall begin by discussing the general ideas of compressed sensing as it is used in linear inverse problems. Consider the problem of recovering information $x \in \mathbb{C}^n$ from a scanning device, represented by an invertible matrix $M \in \mathbb{C}^{n \times n}$, given observed measurements $y := Mx$. In general, we require knowledge of every element of y to be able to accurately recover x without additional structure. Indeed, let $\Omega = \{\alpha_1, \alpha_2, \dots, \alpha_m\}$ with $1 \leq \alpha_1 < \alpha_2 < \alpha_3 < \dots < \alpha_m \leq n$ and define the projection map $P_\Omega : \mathbb{C}^n \rightarrow \mathbb{C}^m$ so that $P_\Omega(x_1, x_2, \dots, x_n) := (x_{\alpha_1}, x_{\alpha_2}, \dots, x_{\alpha_m})$. If m is strictly less than n then for a given y there are at least two distinct vectors $x_1 \in \mathbb{C}^n$ and $x_2 \in \mathbb{C}^n$ with $P_\Omega y = P_\Omega Mx_1 = P_\Omega Mx_2$, so that knowledge of $P_\Omega y$ will not allow us to distinguish between multiple candidates for x . Ideally though we would like to be able to take $m \ll n$ to reduce either the computational or financial costs associated with using the scanning device M .

So far, we have not assumed any additional structure on x . However, let us consider the case where the vector x consists mostly of zeros. More precisely, we make the following definition.

Definition 1.1 (sparsity). *A vector $x \in \mathbb{C}^n$ is said to be s -sparse for some natural number s if $|\text{supp}(x)| \leq s$, where $\text{supp}(x)$ denotes the support of x .*

The key to compressed sensing is the fact that, under certain conditions, any minimizer

$$(P1) \quad z \in \arg \min \|\hat{x}\|_1 \text{ such that } Ux = U\hat{x}$$

of the ℓ^1 basis pursuit (BP) problem (where $U := P_\Omega M$ is a matrix in $\mathbb{C}^{m \times n}$) gives a good approximation to x . Indeed, one would like to be able to use minimizers of (P1) to recover x whenever x is s -sparse, and more generally if x is close to an s -sparse vector then we might expect solutions to (P1) to be close to x . We can encapsulate this statement mathematically by making the following definition.

Definition 1.2 (uniform recovery of order s). *Let s be a positive integer. We say that uniform recovery of order s is possible for the matrix U if solutions \tilde{x} to (P1) satisfy*

$$(1.1) \quad \|x - \tilde{x}\|_1 \leq C\sigma_s(x)_1$$

for some constant C independent of x , where $\sigma_s(x)_1 := \min\{\|x - \hat{x}_2\|_1 \text{ such that } \hat{x}_2 \text{ is } s\text{-sparse}\}$. Note that (1.1) implies that all s -sparse x are recovered exactly by solving (P1), since if x is s -sparse then $\sigma_s(x)_1 = 0$.

Proving that uniform recovery of order s is possible for the matrix U is an inherently complicated task. To simplify this task, the nullspace property and RIP have been introduced (see [19] and [35] for more information). More specifically, the nullspace property is defined as follows.

Definition 1.3 (ℓ^2 RNSP of order s). *A matrix $U \in \mathbb{C}^{m \times n}$ is said to satisfy the ℓ^2 robust nullspace property (ℓ^2 RNSP) of order s if there is a $\rho \in (0, 1)$ and a $\tau > 0$ such that for all vectors $v \in \mathbb{C}^n$ and all S which are subsets of $\{1, 2, 3, \dots, n\}$ with $|S| \leq s$, we have $\|v_S\|_2 \leq \rho\|v_{S^c}\|_1/\sqrt{s} + \tau\|Uv\|_2$.*

The RIP is defined in terms of the restricted isometry constant (RIC) δ_s . A matrix is said to have the RIP if $\delta_s < 1$.

Definition 1.4 (restricted isometry property). *The RIC of order s for a matrix $U \in \mathbb{C}^{m \times n}$, denoted by δ_s , is the minimal $\delta \geq 0$ such that*

$$(1.2) \quad (1 - \delta)\|x\|_2^2 \leq \|Ux\|_2^2 \leq (1 + \delta)\|x\|_2^2$$

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

It is well known (e.g., [6] and [35]) that if U satisfies the ℓ^2 RNSP or the RIC of order s is sufficiently small, then (1.1) is satisfied when finding minimizers \tilde{x} of the BP problem (P1).

Remark 1.5. In fact, the RIP and RNSP both imply a stronger result. Suppose that instead of seeing Ux , we see a noisy version $v := Ux + \nu$ for some noise vector ν with

$\|\nu\|_2 \leq \epsilon$. Instead of finding minimizers of (P1), we can try to recover x by finding minimizers to the modified ℓ^1 minimization problem (so that x itself is a feasible solution)

$$(P2) \quad z \in \arg \min \|\hat{x}\|_1 \text{ such that } \|v - U\hat{x}\|_2 \leq \epsilon.$$

Then any minimizer \tilde{x} to (P2) will satisfy both

$$(1.3) \quad \|x - \tilde{x}\|_1 \leq C\sigma_s(x)_1 + D\epsilon\sqrt{s},$$

$$(1.4) \quad \|x - \tilde{x}\|_2 \leq \frac{C\sigma_s(x)_1}{\sqrt{s}} + D\epsilon$$

provided that U satisfies the ℓ^2 RNSP (e.g., [35, Theorem 4.22]) or δ_s is sufficiently small (e.g., [12, 13, 16, 34] and [14, 28, 64] for optimal conditions). In later chapters we shall introduce the ℓ^2 RNSP in levels and the RIP in levels which will also have a similar resilience to noise.

2. The absence of the uniform recovery and the flip test.

2.1. The flip test. Although uniform recovery seems convenient, it is in general very difficult to verify that uniform recovery of order s is possible for a matrix U . In fact, showing that the RIC of an arbitrary matrix is below a certain value is an NP hard problem [62]. However, some special cases for U do exhibit uniform recovery of order s (e.g., with high probability, Gaussian and Bernoulli matrices can achieve uniform recovery [20]). Even though it is hard to show that uniform recovery is possible for a general matrix U , there is a simple test (the “flip test,” introduced in [2]) that shows that there are a variety of matrices used in practical applications for which uniform recovery grossly underestimates the effectiveness of compressed sensing.

Flip test (sparse vectors). Suppose we are given $U \in \mathbb{C}^{m \times n}$, $s_1 \in \mathbb{N}$, and an s_1 -sparse vector x^1 that is perfectly recovered by finding minimizers of the BP problem (P1) using U and $x = x^1$. We now want to test if this recovery is uniform.

1. Let Q be an operator that permutes the entries of x^1 and let $x^2 = Qx^1$. Run the BP problem (P1) with $x = x^2$ to try to recover x^2 from Ux^2 and obtain a minimizer \tilde{x}^2 . Compare x^1 and $Q^{-1}\tilde{x}^2$. If $x^1 \neq Q^{-1}\tilde{x}^2$ then we do not have uniform recovery of order s_1 .
2. If the test in the first step failed, we want to test how far we were from uniform recovery of order s_1 . We want to see how many coefficients s_2 of x^2 one could hope to recover uniformly. Select a nonzero coefficient of x^2 and set it to zero and call this new vector h^1 . If we recover h^1 by using BP with $x = h^1$ in (P1), then set $s_2 = s_1 - 1$. If not set n nonzero coefficients of x^2 to zero to obtain h^n and repeat until h^n is recovered exactly by using BP with $x = h^n$ in (P1). Let

$$s_2 = s_1 - n.$$

3. If the first step succeeds, retry it with many different permutation matrices Q . If this succeeds for a large variety of such Q then this is an indicator (but not a mathematical proof) that we may have uniform recovery.

The particular choice of Q that was given in [2] was the permutation Q_{reverse} that reverses order—namely, if $x \in \mathbb{C}^n$ then

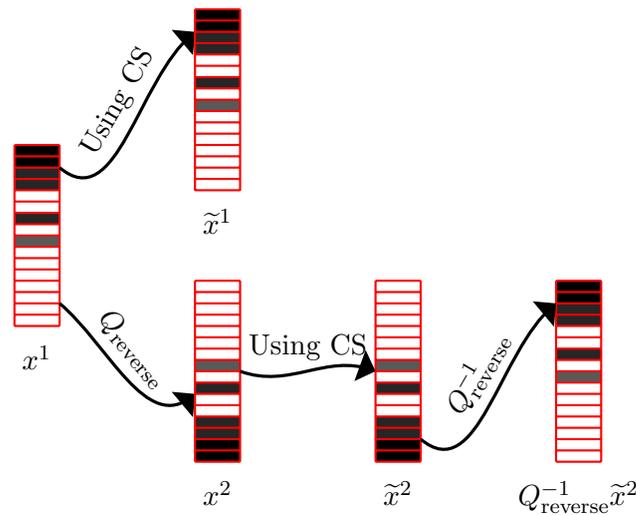


Figure 1. A graphical demonstration of the flip test for matrices which exhibit the uniform recovery where x^1 is a vector exactly recovered by minimizers of (P1). Darker colors denote larger values. If uniform recovery of a sufficiently high order holds, then $Q_{\text{reverse}}^{-1} \tilde{x}^2 = \tilde{x}^1$.

$$Q_{\text{reverse}}(x_1, x_2, \dots, x_{n-1}, x_n) = (x_n, x_{n-1}, \dots, x_2, x_1).$$

A graphical demonstration and summary of the expected results of the flip test with a matrix U that exhibits uniform recovery is given in Figure 1.

We have performed the flip test on Fourier and Hadamard matrices in combinations with wavelet transforms. In particular, the U used in the test is of the form

$$U = P_{\Omega} \text{DFT} \cdot \text{DWT}_N^{-1} \quad \text{or} \quad U = P_{\Omega} \text{HAD} \cdot \text{DWT}_N^{-1}$$

for different successful sampling patterns Ω . The notation DFT, HAD, and DWT_N is used throughout this article to represent the discrete Fourier transform, the Hadamard transform, and the discrete wavelet transform (with Daubechies wavelets with N vanishing moments), respectively. These different types of matrices are represented in a variety of applications including (but not limited to) MRI [37, 52], radio interferometry [53], helium atom scattering [44], electron tomography [36, 49], CT [23, 39], fluorescence microscopy [59, 61], and NMR [43].

In Figure 2 and Table 1 we have displayed the results of the flip test. Note the failure of uniform recovery displayed visually in Figure 2. More quantitatively, observe the substantial differences between s_1 and s_2 in Table 1. It is worth noting that even with 97% sampling as in the second row of Figure 2, there is still a vast difference between \tilde{x}^1 and $Q^{-1}x^2$. Although this may seem surprising at first, this is a consequence of the near block diagonal structure of the matrix $\text{DFT} \cdot \text{DWT}_N^{-1}$ (see Figure 3 and Remark 2.2). The high Fourier frequencies (which, due to the block diagonality, correspond to the finer detail wavelet coefficients) are heavily subsampled since the finer detail coefficients are highly sparse. However, when the wavelet coefficients are flipped, we are now subsampling the Fourier frequencies corresponding to the nonsparse coarse wavelet coefficients. Thus the recovery is poor and we get the results of Figure 2. Note that the flip test will fail in a similar manner if we replace wavelets with

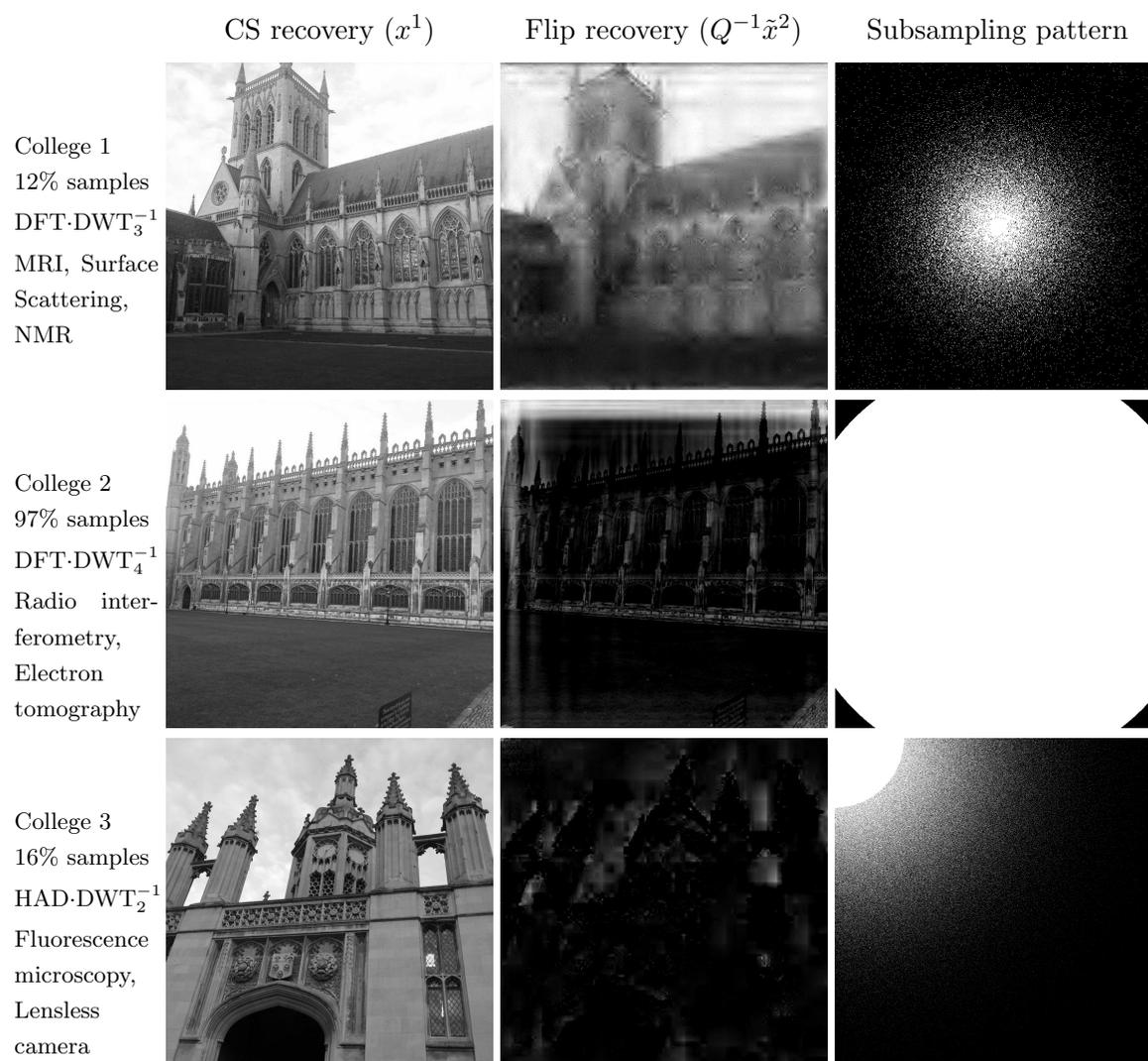


Figure 2. Results of the flip test for different compressed sensing matrices frequently used in applications.

Table 1

A table displaying the number of nonzeros that are recovered exactly by various operators. s_1 represents the number of nonzeros that can be recovered from a standard image, whereas s_2 represents the nonzeros recoverable after flipping.

Image	Operator	s_1	s_2
College 1	DFT·DWT ₃ ⁻¹	121,923	329
College 2	DFT·DWT ₄ ⁻¹	1,850,917	143
College 3	HAD·DWT ₂ ⁻¹	167,772	4

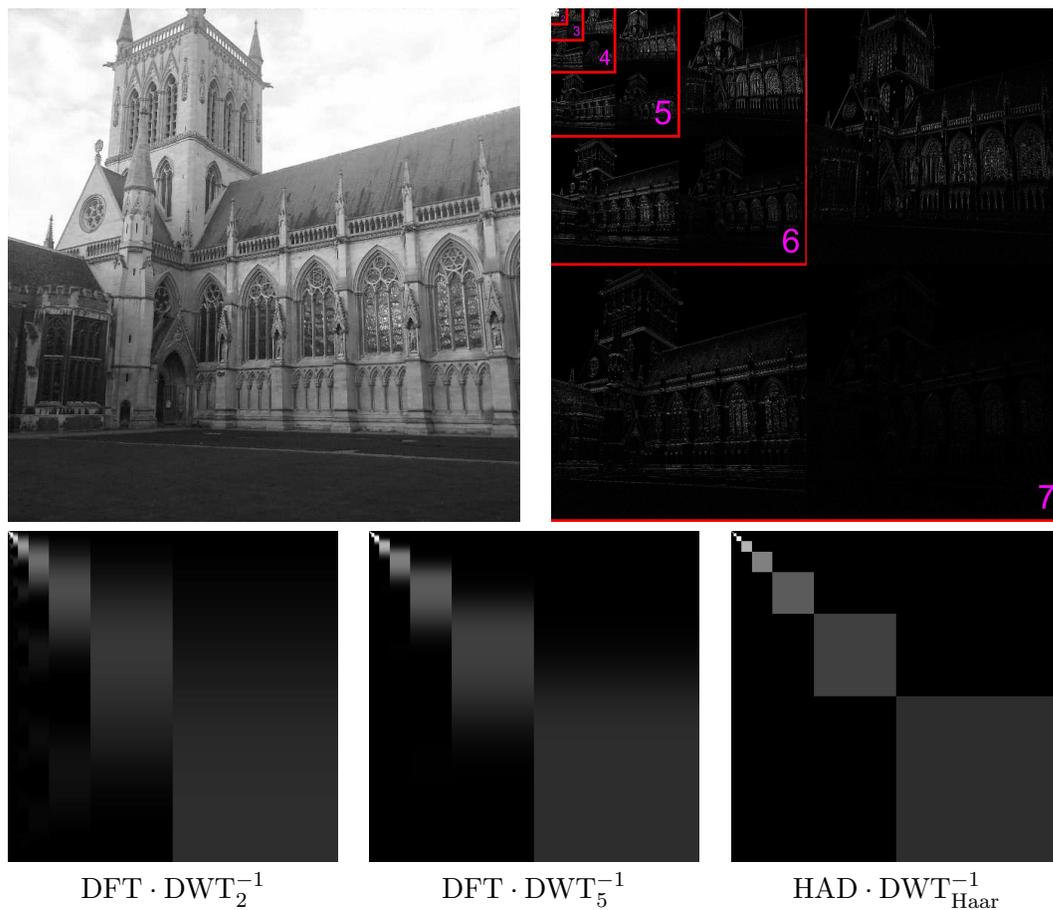


Figure 3. First row: an image and its wavelet coefficients, where a brighter color corresponds to a larger value. Second row: absolute values of a variety of compressive sensing matrices. The block diagonal structure allows us to fully sample rows that correspond to the coarser wavelet levels and subsample the rows that correspond to the finer wavelet levels.

other popular frames such as curvelets, contourlets, or shearlets [17, 29, 48]. We thus need to consider more structure than just sparsity to explain the success of compressed sensing in these applications.

Remark 2.1 (sparsity cannot be the right model). The flip test reveals that sparsity cannot be the correct model for these examples in compressed sensing. When the values of s_1 and s_2 are (for example)

$$s_1 = 121,923, \quad s_2 = 329,$$

it is hard to argue that one recovers s -sparse vectors for a representative s when the location of the nonzero coefficients is arbitrary. On the contrary, as the flip test reveals, the location of the nonzero coefficients is highly important. If sparsity is not the correct model one needs to revise the model in order to find a more realistic description. Moreover, the concepts of the nullspace property of order s and the RIP no longer apply if sparsity is not the correct model for compressed sensing. Of course, it could be the case that the RIC of order 329 is

sufficiently small to allow uniform recovery of order 329; however, that has nothing to do with the successful recovery of the image with $s = 121,923$ nonzero wavelet coefficients.

Remark 2.2 (large coherence and almost block diagonality). One can understand the lack of uniform recovery from simply looking at Figure 3. The blocks in the matrices correspond to the different scales in the wavelet expansion that give a level structure. The wavelet coefficients get relatively sparser in the finer levels and this corresponds to the blocks in the matrices where the absolute values (coherence; see Definition 3.8) decrease. The coherence is not uniformly small. In fact, it is very big in the upper left corner and then decreases with the levels. As is well known [35], it is the uniform small coherence that is the key property to prove uniform recovery. It should be noted that there have been attempts in applications to change the measurements in, for example, MRI and radio interferometry, in order to make the coherence smaller in the first levels. This is called the spread spectrum technique [55, 56].

2.2. Weighted sparsity. Consideration of a different explanation for the success of compressed sensing that includes more structure than just plain sparsity is not a novel idea. Indeed, *weighted sparsity* and the *weighted RIP* were described in [58] as a structured alternative to sparsity and the RIP. To describe this approach, we shall begin by defining *weighted sparsity*. More specifically, given a collection of weights $\omega := (\omega_1, \omega_2, \dots, \omega_n) \in \mathbb{R}^n$ with $\omega_j \geq 1$ for each j , a vector $x \in \mathbb{C}^n$ is said to be (ω, s) -weighted sparse if the weighted ℓ^0 norm, $\|x\|_{\omega,0} := \sum_{j \in \text{supp}(x)} \omega_j^2$, satisfies $\|x\|_{\omega,0} < s$. We can similarly extend the ℓ^1 norm to a weighted ℓ^1 norm by defining $\|x\|_{\omega,1} := \sum_{j=1}^n \omega_j x_j$ and then examine *weighted ℓ^1 minimization* in the same way that we can discuss ℓ^1 minimization. A preliminary idea to deal with the difficulties raised in section 2.1 is to argue that instead of expecting uniform recovery of order s as in equation (1.1) to hold whenever \tilde{x} is a minimizer of (P1), we should hope for uniform recovery of order (ω, s) to hold. More specifically,

$$(2.1) \quad \|\tilde{x} - x\|_1 \leq C \sigma_{\omega,s}(x)_1,$$

where $\sigma_{\omega,s}(x)_1 := \min\{\|x - \hat{x}_2\|_1 \text{ such that } \hat{x}_2 \text{ is } (\omega, s)\text{-weighted sparse}\}$ and C is a fixed constant. This is further motivated by the success of such an approach to the recovery of smooth functions from undersampled measurements [58] and the improvements seen by applying weighted ℓ^1 techniques to random Gaussian matrices [46].

2.2.1. The insufficiency of uniform recovery of weighted sparse vectors through ℓ^1 minimization with wavelets. Unfortunately, there are issues with this method when applied to problems involving a level based construction basis such as wavelets like in section 2.1. These are more thoroughly documented in [1], but we shall provide a brief outline here. Just as the flip test demonstrates that in many examples relevant to practical applications the class of s -sparse vectors is too big and contains objects that cannot be recovered by ℓ^1 minimization, we have the same phenomenon for weighted sparsity. We find that for problems involving a level based reconstruction basis, and for any choice of weights ω , the class of (ω, s) -sparse vectors is too large and contains vectors that cannot be recovered by either weighted- ℓ^1 or ℓ^1 minimization.

In our specific setting above, this means that we have a “natural” image with wavelet coefficients w that is recovered exactly and a vector w' with $\|w'\|_{(\omega,0)} \leq \|w\|_{(\omega,0)}$ which is

not recovered. Therefore, either $\|w\|_{(\omega,0)} > s$ and w is not (ω, s) -sparse (so a theory based on weighted sparsity does not explain why w is recovered) or w' is (ω, s) -sparse (but not recovered, so that the class of (ω, s) -sparse vectors is too large and inequality (2.1) does not hold).

We can show these results by expanding the “flip test” from section 2.1. The result is the *flip test for weighted sparse vectors*.

Flip test (weighted sparse vectors). Suppose we are given $U \in \mathbb{C}^{m \times n}$, a collection of weights $\omega \in \mathbb{C}^n$, and a vector $x^1 \in \mathbb{C}^n$ that is perfectly recovered by finding a minimizer of the BP problem (P1) using U and $x = x^1$. Set s to be the minimal value so that x^1 is (ω, s) -sparse. We now want to test if this recovery is uniform across all (ω, s) -weighted sparse vectors.

1. Let Q be an operator that permutes the entries of x^1 and let $v = Qx^1$. Repeatedly set individual coefficients of v to be 0 until v is also (ω, s) -sparse. Call this new vector x^2 .
2. Run the BP problem (P1) with $x = x^2$ to try to recover x^2 from Ux^2 and obtain a minimizer \tilde{x}^2 .
3. If \tilde{x}^2 is not recovered exactly with this method, we do not have uniform recovery of (ω, s) -weighted sparse vectors.
4. Retry steps 1 to 3 with many different permutation matrices Q . If this succeeds over a large variety of such Q then this is an indicator (but not a mathematical proof) that we may have uniform recovery. A single failure, however, demonstrates that we do not have uniform recovery of (ω, s) -weighted sparse vectors.

Figure 4 displays some examples where the flip test implies a lack of uniform recovery of weighted sparse vectors (either because s is too small to explain the perfect recovery observed or s is too large and there are too many vectors that are (ω, s) -weighted sparse as in the previous discussion). Thus, weighted sparsity is insufficient to explain the success of compressed sensing when using wavelets and other X-lets. In Figure 4 we have displayed only the result of using ℓ^1 recovery; however, the results are the same when using weighted ℓ^1 . See [1] for a thorough discussion of this phenomenon. We shall provide additional insight as to why weighted sparsity is insufficient in section 3.1.4.

Remark 2.3. It must be emphasized that weighted sparsity and the weighted RIP were developed in [58] for the purpose of recovering smooth functions with polynomials. Thus, one should not expect the weighted RIP to hold for wavelets. Conversely, the RIP in levels may not work for polynomials, as unlike wavelets there is no level structure present. Moreover, in [46] the weighted approach is used in combination with random Gaussian measurements, which is very different from the setup in this paper. These facts demonstrate the subtleties of compressed sensing theory and that we are in need of a collection of much more specific theorems using different sparsity models that depend on the problem.

Remark 2.4. The matrices discussed in sections 2.1 and 2.2.1 focused on matrices that can be used to solve finite dimensional models of the real-world compressive sensing applications. In some circumstances, it has been shown that this does not match the original infinite dimensional problem and a different finite dimensional approximation is needed [3, 4, 10, 37, 38, 54]. It should be noted that the preceding flip tests could easily be adapted to this infinite dimensional setting, and thus uniform recovery of either s -sparse vectors or (ω, s) -weighted sparse vectors will still be unattainable for descriptive values of s .

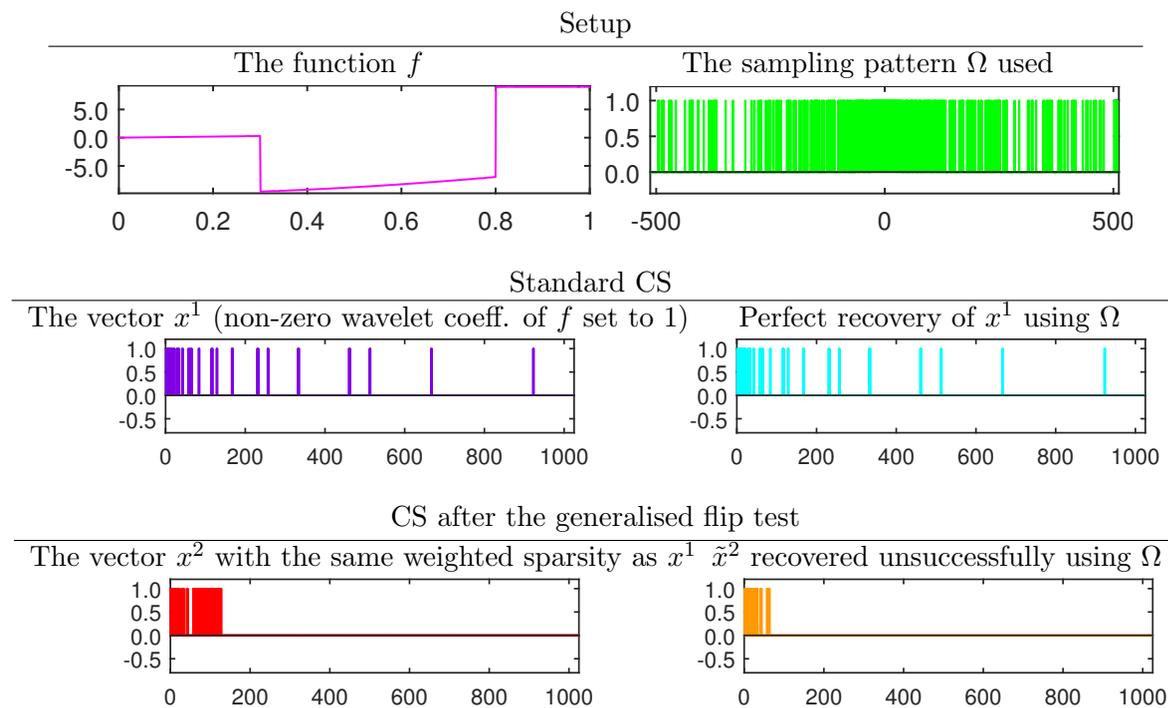


Figure 4. The figure displays the flip test for weighted sparse vectors with the function $f(x) = \sin(x)\mathbb{1}_{[0,0.3]} - 10\cos(x)\mathbb{1}_{(0.3,0.8]} + 9\mathbb{1}_{(0.8,1]}$ after its wavelet coefficients are thresholded so that perfect recovery is possible. Recovery was done using a subsampled one-dimensional (1D) Fourier to wavelet matrix, with Daubechies 3 wavelets and ℓ^1 minimization. The weights on the coefficients in level i were given by 2^i . Similar results follow for other weights and also for recovery with weighted ℓ^1 minimization.

3. An extended theory for compressed sensing. The current mathematical theory for compressive sensing revolves around a few key ideas. These are the concepts of sparsity, incoherence, uniform subsampling, and uniform recovery of order s . In [4] and [59], it was shown that these concepts are absent for a large class of compressed sensing problems. To solve this problem, the extended concepts of *asymptotic sparsity*, *asymptotic incoherence*, and *multilevel sampling* were introduced. We now introduce the fourth extended concept in the new theory of compressive sensing: uniform recovery of order (\mathbf{s}, \mathbf{M}) . To accomplish this, we shall extend the definitions of nullspace property and RIP of order s to a pair of new concepts—the RIP in levels and the nullspace property of order (\mathbf{s}, \mathbf{M}) .

3.1. A level based alternative to sparsity: (\mathbf{s}, \mathbf{M}) -sparsity. The examples given in Figure 2 all involve reconstructing in a basis that is divided into various levels. It is this level based structure that prevents us from observing uniform recovery of order s and necessitates a new theory based on a different kind of sparsity. We shall demonstrate this new theory with wavelets, which we give a brief description of in the following section. Despite our focus on wavelets in the next few pages, it should be noted that our work applies equally to all level based reconstruction bases.

3.1.1. Wavelets. A multiresolution analysis (as defined in [25, 26, 51]) for $L^2(X)$ (where X is an interval or a square) is formed by constructing increasing scaling spaces $(V_j)_{j=0}^\infty$ and wavelet spaces $(W_j)_{j=0}^\infty$ with $V_j, W_j \subset L^2(X)$ so that

1. if $f(\cdot) \in V_j$, then $f(2\cdot) \in V_{j+1}$, and vice versa;
2. $\overline{\bigcup_{j=0}^\infty V_j} = L^2(X)$ and $\bigcap_{j=0}^\infty V_j = \{0\}$;
3. W_j is the orthogonal complement of V_j in V_{j+1} .

The wavelet expansion of a function f is an expansion in terms of basis elements of V_0 (the *scaling level*) and W_j for $j \geq 0$ (the *wavelet levels*, said to be increasingly fine as j increases). For natural images f , the largest coefficients in the wavelet expansion of f appear in the levels corresponding to smaller j (the *coarser levels*). Closer examination of the relative sparsity in each level also reveals a pattern: let w be the collection of wavelet coefficients of f and for a given level k let S^k be the indices of all wavelet coefficients of f in the k th level. Additionally, let \mathcal{M}_n be the largest (in absolute value) n wavelet coefficients of f . Given $\epsilon \in [0, 1]$, we define the functions $s(\epsilon)$ and $s_k(\epsilon)$ (as in [4]) by

$$s(\epsilon) := \min \left\{ n : \|w_{\mathcal{M}_n}\|_2 = \sqrt{\sum_{i \in \mathcal{M}_n} |w_i|^2} \geq \epsilon \sqrt{\sum_{k \in \mathbb{N}} \sum_{i \in S^k} |w_i|^2} = \epsilon \|w\|_2 \right\},$$

$$s_k(\epsilon) := |\mathcal{M}_{s(\epsilon)} \cap S^k|.$$

More succinctly, $s_k(\epsilon)$ represents the relative sparsity of the wavelet coefficients of f at the k th scale. If an image is very well represented by wavelets, we would like $s_k(\epsilon)$ to be as small as possible for ϵ close to 1. However, one can make the following observation: then the ratios $s_k(\epsilon)/|S^k|$ decay very rapidly for a fixed ϵ . Numerical examples showing this phenomenon with Haar wavelets are displayed in Figure 5. Summarizing, we observe that images taken from the real world are sparse with a structure which the traditional RIP ignores.

3.1.2. (s, M)-sparsity and uniform recovery of order (s, M). Uniform recovery of order s suggests that we are able to recover all s -sparse vectors exactly, independent of which levels the s -sparse vectors are primarily supported on. Instead of such a stringent requirement, we can take advantage of the structure of our problem, a concept that is already popular from the recovery point of view [7, 40, 41, 63]. We have observed that, for wavelets, $s_k(\epsilon)/|S^k|$ decays rapidly as $k \rightarrow \infty$ (see Figure 5). To further understand this phenomenon, in [4] the concept of (s, M)-sparsity was introduced.

Definition 3.1 ((s, M)-sparsity). Let $\mathbf{M} = (M_0, M_1, \dots, M_l) \in \mathbb{N}^{l+1}$ with $1 \leq M_1 < M_2 < \dots < M_l$ and $M_0 = 0$, where the natural number l is called the number of levels. Additionally, let $\mathbf{s} = (s_1, s_2, \dots, s_l) \in \mathbb{N}^l$ with $s_i \leq M_i - M_{i-1}$. We call (\mathbf{s}, \mathbf{M}) a sparsity pattern.

A set Λ of integers is said to be (s, M)-sparse if $\Lambda \subset \{M_0 + 1, M_0 + 2, \dots, M_l\}$ and for each $i \in \{1, 2, \dots, l\}$, we have $|\Lambda \cap \{M_{i-1} + 1, M_{i-1} + 2, \dots, M_i\}| \leq s_i$. A vector is said to be (s, M)-sparse if its support is an (s, M)-sparse set. The collection of (s, M)-sparse vectors is denoted by $\Sigma_{\mathbf{s}, \mathbf{M}}$. We can also define $\sigma_{\mathbf{s}, \mathbf{M}}(x)_1$ as a natural extension of $\sigma_s(x)_1$. Namely, $\sigma_{\mathbf{s}, \mathbf{M}}(x)_1 := \min_{\hat{x} \in \Sigma_{\mathbf{s}, \mathbf{M}}} \|x - \hat{x}\|_1$.

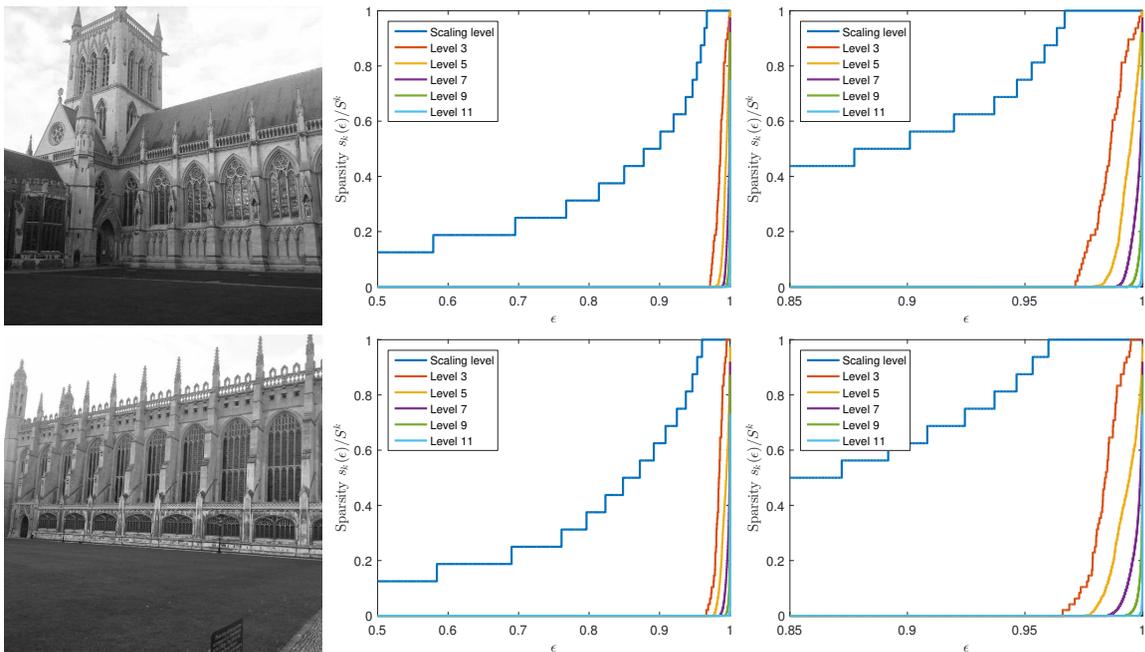


Figure 5. The relative sparsity of Haar wavelet coefficients of two images. The leftmost column displays the image in question. The middle and final columns display the values of $s_k(\epsilon)$ for $\epsilon \in [0.5, 1]$ and $\epsilon \in [0.85, 1]$, respectively, where k represents a wavelet level. Of particular importance is the rapid decay of $s_k(\epsilon)$ as k grows larger. “Scaling level” denotes the case where k corresponds to the scaling level.

Remark 3.2. If (\mathbf{s}, \mathbf{M}) is a sparsity pattern, we will sometimes refer to $(a\mathbf{s}, \mathbf{M})$ -sparse sets for some natural number a even though as_i may be larger than $M_i - M_{i-1}$. To make sense of such a statement, we define (in this context)

$$as := (\min(as_1, M_1 - M_0), \min(as_2, M_2 - M_1), \dots, \min(as_l, M_l - M_{l-1})).$$

Let us now look at a specific case where (\mathbf{s}, \mathbf{M}) represent wavelet levels (again, we emphasize that wavelets are simply one example of a level based system and that our work is more general). Roughly speaking, we can choose \mathbf{s} and \mathbf{M} (we set \mathbf{M} so that $M_{i-1} + 1$ is the first index for the i th wavelet level) such that x is (\mathbf{s}, \mathbf{M}) -sparse if it has fewer nonzero coefficients in the finer wavelet levels. As with uniform recovery of order s , we ask for minimizers \tilde{x} to (P1). Instead of asking for (1.1), we might expect

$$(3.1) \quad \|x - \tilde{x}\|_1 \leq C\sigma_{\mathbf{s}, \mathbf{M}}(x)_1$$

for some C independent of x . If these conditions are satisfied then we say that *uniform recovery of order (\mathbf{s}, \mathbf{M}) is possible for the matrix U* .

3.1.3. The flip test in levels. In section 2.1, we saw that there was a simple test that uniform recovery of order s is not an accurate explanation for why compressed sensing is effective with some matrices U . However, the argument in section 2.1 does not apply if we

expect uniform recovery of order (\mathbf{s}, \mathbf{M}) instead of uniform recovery of order s (since (1.1) will no longer hold for minimizers \tilde{x} of (P1)).

Flip test in levels ((\mathbf{s}, \mathbf{M})-sparse vectors). Suppose we are given $U \in \mathbb{C}^{m \times n}$, a sparsity pattern (\mathbf{s}, \mathbf{M}) , and an (\mathbf{s}, \mathbf{M}) -sparse vector x^1 that is perfectly recovered by finding a minimizer of the BP problem (P1) using U and $x = x^1$. We now want to test if this recovery is uniform.

1. Let Q be a randomly chosen permutation with $Q(\Sigma_{\mathbf{s}, \mathbf{M}}) = \Sigma_{\mathbf{s}, \mathbf{M}}$ and let $x^2 = Qx^1$. Run the BP problem (P1) with $x = x^2$ to try to recover x^2 from Ux^2 and obtain a minimizer \tilde{x}^2 . Compare x^1 and $Q^{-1}\tilde{x}^2$. If $x^1 \neq Q^{-1}\tilde{x}^2$ then we do not have uniform recovery of order (\mathbf{s}, \mathbf{M}) (since x^2 is (\mathbf{s}, \mathbf{M}) -sparse).
2. If the first step succeeds, retry it with many different permutation matrices Q satisfying $Q(\Sigma_{\mathbf{s}, \mathbf{M}}) = \Sigma_{\mathbf{s}, \mathbf{M}}$. If this succeeds for a large variety of such Q then this is an indicator (but not a mathematical proof) that we may have uniform recovery.

The requirement that $Q(\Sigma_{\mathbf{s}, \mathbf{M}}) = \Sigma_{\mathbf{s}, \mathbf{M}}$ now requires us to consider different permutations than a simple reverse permutation as in section 2.1. A natural adaptation of Q_{reverse} to this new “flip test in levels” is a permutation that just reverses coefficients within each wavelet level. Figure 6 displays what happens when we attempt to do the flip test with this permutation. In this case, we see that the performance of compressed sensing reconstruction under flipping and the performance of standard compressed sensing reconstruction are very similar. This suggests that uniform recovery within the class of (\mathbf{s}, \mathbf{M}) -sparse vectors (as in (3.1)) is possible with a variety of practical compressive sensing matrices. Indeed, in Table 2 we also consider a collection of randomly generated Q with $Q(\Sigma_{\mathbf{s}, \mathbf{M}}) = \Sigma_{\mathbf{s}, \mathbf{M}}$. We see that perfect recovery of Qx^1 is possible for a wide variety of permutation matrices Q .

Remark 3.3. Throughout this article we have used a variety of numerical tests to demonstrate that in a collection of compressive imaging applications the concepts of uniform recovery of sparse vectors or uniform recovery of weighted sparse vectors are not appropriate to explain the success of compressed sensing. By contrast, there is evidence (like the flip test in levels) to suggest that uniform recovery of (\mathbf{s}, \mathbf{M}) -sparse vectors is the right model to explain why compressive imaging works with applications using matrices such as $\text{DFT} \cdot \text{IWT}$ or $\text{HAD} \cdot \text{IWT}$. Further detail on this claim is provided in [1].

Table 2

Flip test in levels with randomly generated permutations. The table displays the flip test in levels for various images x^1 as in Figure 6 permuted using Q to form $x^2 = Qx^1$. Each image was processed with a fixed subsampling pattern and 1000 randomly generated permutations as described in section 3.1.3.

Image	Subsampling percentage	Matrix	Permutations where x^2 was perfectly recovered	Permutations where x^2 was not perfectly recovered
College 1	12.48%	$\text{DFT} \cdot \text{DWT}_3^{-1}$	1000	0
College 2	97.17%	$\text{DFT} \cdot \text{DWT}_4^{-1}$	1000	0
College 3	15.54%	$\text{HAD} \cdot \text{DWT}_2^{-1}$	1000	0

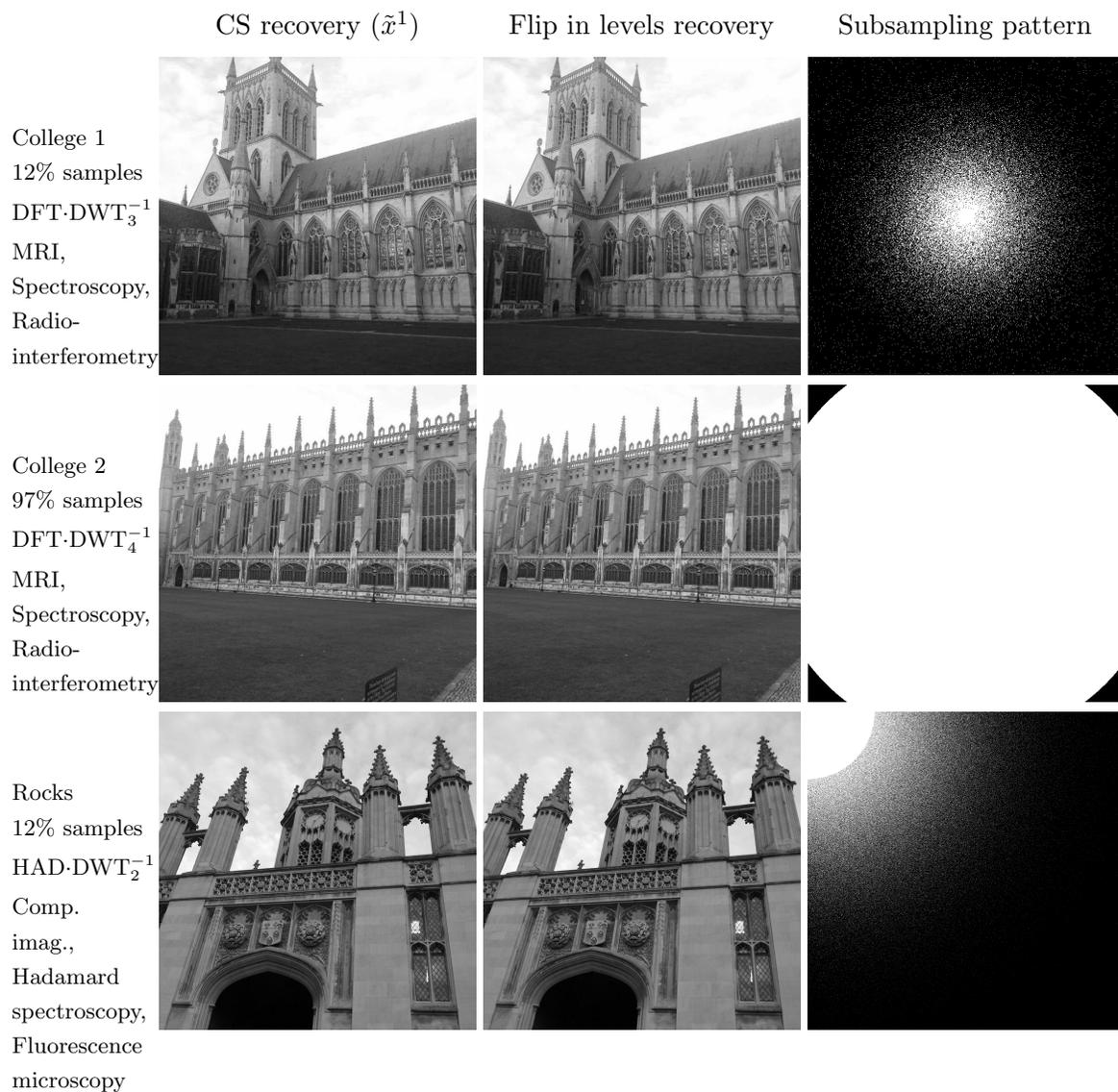


Figure 6. Results of the flip test when the flipping preserves the sparsity within the levels.

3.1.4. Relating (s, M) -sparsity and weighted sparsity. The “flip test in levels” suggests that for many compressed sensing problems, there are \mathbf{s} and \mathbf{M} such that all (\mathbf{s}, \mathbf{M}) -sparse vectors are recovered equally well by ℓ^1 minimization. With this in mind, we are now in a position to provide additional details on why the same is not the case for weighted sparsity. Indeed, one can easily state and prove the following theorem (see [1] for details).

Theorem 3.4 (the relationship between weighted sparsity and (\mathbf{s}, \mathbf{M}) -sparsity). *Let (\mathbf{s}, \mathbf{M}) have l levels (with $l > 2$) and fix $r < l$. Suppose that the collection of (\mathbf{s}, \mathbf{M}) -sparse vectors are all (ω, X) -weighted sparse for some X . Then there is an l_0 with $r < l_0 < l$ such that the collection of $(\tilde{\mathbf{s}}, \mathbf{M})$ -sparse are also (ω, X) -weighted sparse, where*

$$\tilde{\mathbf{s}} = (\underbrace{s_1, s_2, \dots, s_r}_r, \underbrace{0, 0, \dots, 0}_{l_0-1-r}, (l-r)s_{l_0}, 0, \dots, 0).$$

In particular, the set of (\mathbf{s}, \mathbf{M}) -sparse and (ω, X) -weighted sparse vectors are not the same.

The use of this theorem becomes apparent if we consider Figure 4. As in the second row of Figure 3, the Fourier to wavelet matrix in Figure 4 is well approximated by block diagonal matrices. This block diagonality structure means that we can design our sampling pattern so that information corresponding to coarser wavelet levels is more readily captured than the information corresponding to the finer wavelet levels. Since the finer wavelet levels are relatively more sparse (see the first row of Figure 3 and Figure 5) we can design a sampling pattern to effectively capture images (note, however, that these variable density/multilevel sampling schemes have been discussed in [4, 8, 22, 47, 52, 57, 68] and structured sampling in [21, 67]). To utilize these ideas we choose a sampling pattern so that the first r levels will be fully sampled, but after that subsampling occurs and this is where we run into difficulties with weighted sparsity. If we suppose that recovering all vectors with s_k nonzero coefficients in the indices corresponding to the k th wavelet level takes Ω_k measurements in that level, then recovering all weighted sparse vectors requires $(l-r)\Omega_k$ measurements for some k . Unfortunately, this leads to weighted sparsity overestimating the number of measurements required to recover all vectors of interest. Unless we substantially oversample the finer wavelet levels then we are unable to see uniform recovery of weighted sparse vectors.

3.1.5. The ℓ^2 robust nullspace property of order (\mathbf{s}, \mathbf{M}) and the RIP in levels. Given the success of the “flip test in levels,” let us now try to find a sufficient condition on a matrix $U \in \mathbb{C}^{m \times n}$ that allows us to conclude that uniform recovery of order (\mathbf{s}, \mathbf{M}) is possible for U . If the RIP implies uniform recovery of order s then the obvious idea is to extend the RIP to a so-called RIP in levels, defined as follows.

Definition 3.5 (RIP in levels). For a given sparsity pattern (\mathbf{s}, \mathbf{M}) and matrix $U \in \mathbb{C}^{m \times n}$, the RIP in levels (RIP_L) constant of order (\mathbf{s}, \mathbf{M}) (RIC_L), denoted by $\delta_{\mathbf{s}, \mathbf{M}}$, is the smallest $\delta > 0$ such that

$$(1 - \delta)\|x\|_2^2 \leq \|Ux\|_2^2 \leq (1 + \delta)\|x\|_2^2$$

for all $x \in \Sigma_{\mathbf{s}, \mathbf{M}}$.

We will see that the RIP in levels allows us to obtain error estimates on $\|x - \tilde{x}\|_1$ and $\|x - \tilde{x}\|_2$ with \tilde{x} set to be a minimizer of (P1). Similar error estimates can be obtained if U satisfies the ℓ^2 robust nullspace property of order (\mathbf{s}, \mathbf{M}) , a natural generalization of the nullspace property of order s .

Definition 3.6 (the order (\mathbf{s}, \mathbf{M}) ℓ^2 RNSP). A matrix $U \in \mathbb{C}^{m \times n}$ satisfies the ℓ^2 robust nullspace property of order (\mathbf{s}, \mathbf{M}) if there is a $\rho \in (0, 1)$ and a $\tau > 0$ such that

$$(3.2) \quad \|v_S\|_2 \leq \frac{\rho}{\sqrt{s}} \|v_{S^c}\|_1 + \tau \|Uv\|_2$$

for all (\mathbf{s}, \mathbf{M}) -sparse sets S and vectors $v \in \mathbb{C}^n$.

3.2. Matrices with a small RIP_L constant. To see how matrices similar to the ones in Figure 6 have a small RIC_L , we will first explain how the sampling patterns in Figure 6 were obtained. Earlier work on compressive sensing suggested that sampling should be done uniformly at random (see [35]). Because of the near block diagonality of the matrices $\text{DFT} \cdot \text{IWT}$ and $\text{HAD} \cdot \text{IWT}$ for a variety of wavelets, instead of sampling uniformly at random, it is better to sample using a *multilevel* structure, where the percentage subsampling of each block depends on the relative importance of the corresponding wavelet coefficients (as done in [4], among others). More precisely, we can make the following definition.

Definition 3.7 (multilevel sampling). Let N be the dimension of the ambient measurement space. We set $\mathbf{N} := (N_0, N_1, \dots, N_l)$ and $\mathbf{m} := (m_1, m_2, \dots, m_l)$ so that $0 = N_0 < N_1 < N_2 \cdots < N_l = N$ and $m_k \leq N_k - N_{k-1}$ for every $k = 1, \dots, l$. For each such k , set $\Omega_k = \{t_{k,1}, \dots, t_{k,m_k}\}$, where the $t_{k,j}$ for $j = 1, \dots, m_k$ are selected independently and uniformly at random from $\{N_{k-1} + 1, \dots, N_k\}$. We call $\Omega = \Omega_{\mathbf{N}, \mathbf{m}} = \Omega_1 \cup \Omega_2 \cdots \cup \Omega_l$ a (\mathbf{N}, \mathbf{m}) -multilevel subsampling scheme

We define the *coherence* of U to determine how effective recovery is with equation (P1) for matrices $U = P_\Omega M$, where M is an isometry and where Ω is taken uniformly at random.

Definition 3.8 (coherence). The coherence of a matrix $M \in \mathbb{C}^{N \times N}$, denoted by $\mu(M)$, is the quantity

$$\mu(M) := \max_{i,j=1,\dots,N} |M_{i,j}|^2.$$

If $\mu(M)$ is sufficiently small and Ω is a sufficiently large set taken uniformly at random from $1, 2, \dots, N$ then with high probability $P_\Omega M$ will satisfy the RIP (and therefore exhibit uniform recovery). The related quantity with uniform multilevel subsampling is *coherence in levels*.

Definition 3.9 (local coherence). Given a sparsity pattern (\mathbf{s}, \mathbf{M}) and a uniform (\mathbf{m}, \mathbf{N}) -multilevel subsampling scheme, we define the (j, k) coherence in levels of the matrix M with respect to (\mathbf{s}, \mathbf{M}) and (\mathbf{m}, \mathbf{N}) to be the value $\mu_{j,k}(\mathbf{N}, \mathbf{M})$ where

$$(3.3) \quad \mu_{j,k}(M) = \max\{|M_{s,t}|^2 \text{ such that } s \in \{N_{j-1} + 1, \dots, N_j\}, t \in \{M_{k-1} + 1, \dots, M_k\}\}.$$

If the coherence in levels of a matrix is sufficiently small and Ω is a sufficiently dense uniform multilevel subsampling scheme then we can show that $P_\Omega M$ satisfies the RIP_L . Indeed in [50] the following result was shown.

Theorem 3.10 (the existence of RIP_L matrices). There exists a constant $C > 0$ with the following property: let $M \in \mathbb{C}^{N \times N}$ be an isometry, $l \in \mathbb{N}$, and ϵ, δ real numbers such that $0 < \epsilon, \delta < 1$. Let l_0 be a natural number with $0 \leq l_0 \leq l$ and $\Omega_{\mathbf{N}, \mathbf{m}}$ be an (\mathbf{N}, \mathbf{m}) -multilevel subsampling scheme and (\mathbf{s}, \mathbf{M}) be a sparsity pattern. Suppose that $m_k = N_k - N_{k-1}$ for $k = 1, 2, \dots, l_0$ and that

$$(3.4) \quad m_k \geq C\delta^{-2}(N_k - N_{k-1}) \left(\sum_{r=1}^l \mu_{k,r} s_r \right) (l \log(2\tilde{m}) \log(2N) \log^2(2s) + \log(\epsilon^{-1}))$$

$k > l_0$, where $\tilde{m} = m_{l_0+1} + \dots + m_l$. Then with probability at least $1 - \epsilon$, the matrix $U = P_\Omega M$ satisfies the RIP_L of order (\mathbf{s}, \mathbf{M}) with constant $\delta_{\mathbf{s}, \mathbf{M}} \leq \delta$.

In particular, it is possible to use results from [5] to obtain a condition on the number of measurements taken that guarantees that matrices of the form $P_\Omega \text{DFT} \cdot \text{Haar}^{-1}$ satisfy the RIP_L where Ω is a uniform multilevel subsampling scheme and DFT and Haar^{-1} are the 1D discrete Fourier transform and 1D inverse Haar wavelet transform, respectively. More precisely, Corollary 3.3 from [50] says the following.

Theorem 3.11 (the RIP_L of the $\text{DFT} \cdot \text{Haar}^{-1}$ matrix). *There exists a constant $C > 0$ with the following property: let $N = 2^l$ for some $l \in \mathbb{N}$, where l is the number of wavelet levels and ϵ, δ real numbers such that $0 < \epsilon, \delta < 1$. Set $S^k = 2^{\max(k-1, 1)}$ to be the number of wavelet coefficients in the k th level. Suppose that m_k satisfies*

$$(3.5) \quad m_k \geq C\delta^{-2} \left(s_k + \sum_{r=l_0+1}^l 2^{-|k-r|} s_r \right) (\log(2m) \log^2(2N) \log^2(2s) + \log(\epsilon^{-1}))$$

for $k = 1, 2, \dots, l$. Then if Ω is a (\mathbf{N}, \mathbf{m}) -multilevel sampling scheme with $\mathbf{N} = (0, S^1, S^1 + S^2, \dots, \sum_{r=1}^l S^k)$ and $\mathbf{m} = (m_1, m_2, \dots, m_l)$ then the matrix $P_\Omega \text{DFT} \cdot \text{Haar}^{-1}$ satisfies $\delta_{\mathbf{s}, \mathbf{M}} < \delta$ for $\mathbf{M} := \mathbf{N}$ and $\mathbf{s} := (s_1, s_2, \dots, s_l)$ with probability exceeding $1 - \epsilon$.

It is possible to give further examples of matrices with a small RIP_L constant by examining their coherence in levels and employing Theorem 3.10. Bounds on the coherence in levels for a variety of matrices can be examined by using tools developed in papers such as [45]. Moreover, one can obtain a version of Theorem 3.11 for Hadamard matrices combined with Haar wavelets as the resulting matrix $U = \text{HAD} \cdot \text{DWT}_{\text{Haar}}^{-1}$ is completely block diagonal (see Figure 3); however, this is beyond the scope of this paper.

4. Main results. If a matrix $U \in \mathbb{C}^{m \times n}$ satisfies the RIP then $(1 - \delta_s) < \|u_i\|_2^2 < (1 + \delta_s)$ for each column u_i of U . To ensure that we have similar control over $\|u_i\|_2$ with the RIP_L we make the following two definitions.

Definition 4.1 (ratio constant). *The ratio constant of a sparsity pattern (\mathbf{s}, \mathbf{M}) , which we denote by $\eta_{\mathbf{s}, \mathbf{M}}$, is given by $\eta_{\mathbf{s}, \mathbf{M}} := \max_{i,j} s_i/s_j$.*

If the sparsity pattern (\mathbf{s}, \mathbf{M}) has l levels and there is a $j \in \{1, 2, \dots, l\}$ for which $s_j = 0$ then we write $\eta_{\mathbf{s}, \mathbf{M}} = \infty$.

Definition 4.2 (covering a matrix). *A sparsity pattern (\mathbf{s}, \mathbf{M}) is said to cover a matrix $U \in \mathbb{C}^{m \times n}$ if*

1. $\eta_{\mathbf{s}, \mathbf{M}} < \infty$,
2. $M_l \geq n$, where l is the number of levels for (\mathbf{s}, \mathbf{M}) .

If a sparsity pattern does not cover U because it fails to satisfy either 1 or 2 from the definition of a sparsity pattern covering a matrix U then we cannot guarantee recovery of (\mathbf{s}, \mathbf{M}) -sparse vectors, even in the case that $\delta_{\mathbf{s}, \mathbf{M}} = 0$. We shall justify the necessity of both conditions using two counterexamples. First, we shall provide a matrix U , a sparsity pattern (\mathbf{s}, \mathbf{M}) , and an (\mathbf{s}, \mathbf{M}) -sparse vector $x_1 \in \mathbb{C}^n$ such that $\eta_{\mathbf{s}, \mathbf{M}} = \infty$, $\delta_{\mathbf{s}, \mathbf{M}} = 0$ and x_1 is not recovered by standard ℓ^1 minimization. Indeed, consider the following:

$$U = \begin{pmatrix} 1 & 2 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{s} = (1, 0), \quad \mathbf{M} = (0, 1, 2), \quad x_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

By the definition of $\eta_{\mathbf{s},\mathbf{M}}$, we have $\eta_{\mathbf{s},\mathbf{M}} = \infty$ and it is obvious that $\delta_{\mathbf{s},\mathbf{M}} = 0$. Furthermore, even without noise, x_1 does not solve the minimization problem $\min \|\tilde{x}\|_1$ such that $Ux_1 = U\tilde{x}$. This can easily be seen by observing that $Ux_1 = Ux_2$ with $\|x_2\|_1 = \frac{1}{2}$, where $x_2 := (0, 1/2)^T$. It is therefore clear that Assumption 1 is necessary. We shall now provide an explanation for why Assumption 2 is also a requirement if we wish for the RIP_L to be a sufficient condition for the recovery of (\mathbf{s}, \mathbf{M}) -sparse vectors. This time, consider the following combination of U , (\mathbf{s}, \mathbf{M}) , and x_1 :

$$U = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \end{pmatrix}, \quad \mathbf{s} = (1), \quad \mathbf{M} = (0, 1), \quad x_1 = (1, 0, 0)^T;$$

and again, even though $\delta_{\mathbf{s},\mathbf{M}} = 0$, recovery is not possible because $Ux_1 = Ux_2$ with $\|x_2\|_1 = 1/2$ where $x_2 := (0, 0, 1/2)^T$.

We shall therefore try to prove that uniform recovery of order (\mathbf{s}, \mathbf{M}) is possible with the RIP_L under the assumption that (\mathbf{s}, \mathbf{M}) covers U . To do this, we need one further definition. In (1.4) the bound on $\|x - \tilde{x}\|_2$ involves \sqrt{s} . This arises because s is the maximum number of nonzero values that could be in an s -sparse vector. The equivalent for (\mathbf{s}, \mathbf{M}) -sparse vectors is the following.

Definition 4.3 (number of elements of a sparsity pattern). *The number of elements of a sparsity pattern (\mathbf{s}, \mathbf{M}) , which we denote by \tilde{s} , is given by $\tilde{s} := s_1 + s_2 + \dots + s_l$.*

To prove that a sufficiently small RIP in levels constant implies uniform recovery of order (\mathbf{s}, \mathbf{M}) , it is natural to adapt the steps used in [6] to prove that the RIP implies uniform recovery of order s . This adaptation yields a sufficient condition for recovery even in the noisy case.

Theorem 4.4 (RIP_L recovery theorem). *Let (\mathbf{s}, \mathbf{M}) be a sparsity pattern with l levels and ratio constant $\eta_{\mathbf{s},\mathbf{M}}$. Suppose that the matrix $U \in \mathbb{C}^{m \times n}$ is covered by (\mathbf{s}, \mathbf{M}) and has a RIP_L constant $\delta_{2\mathbf{s},\mathbf{M}}$ satisfying*

$$(4.1) \quad \delta_{2\mathbf{s},\mathbf{M}} < \frac{1}{\sqrt{l(\sqrt{\eta_{\mathbf{s},\mathbf{M}}} + \frac{1}{4})^2 + 1}}.$$

Furthermore, suppose that $x \in \mathbb{C}^n$ and $v = Ux + \nu$ where $\|\nu\|_2 \leq \epsilon$. Then any $\tilde{x} \in \mathbb{C}^n$ which are minimizers of the noisy ℓ^1 minimization problem (P2) also satisfy

$$(4.2) \quad \|x - \tilde{x}\|_1 \leq C_1 \sigma_{\mathbf{s},\mathbf{M}}(x)_1 + D_1 \sqrt{\tilde{s}} \epsilon \quad \text{and}$$

$$(4.3) \quad \|x - \tilde{x}\|_2 \leq \frac{\sigma_{\mathbf{s},\mathbf{M}}(x)_1}{\sqrt{\tilde{s}}} \left(C_2 + C'_2 \sqrt[4]{l\eta_{\mathbf{s},\mathbf{M}}} \right) + \epsilon \left(D_2 + D'_2 \sqrt[4]{l\eta_{\mathbf{s},\mathbf{M}}} \right),$$

where C_1, C_2, C'_2, D_1, D_2 , and D'_2 depend only on $\delta_{2\mathbf{s},\mathbf{M}}$. Note that for the noiseless case $\epsilon = 0$, (4.2) reduces to uniform recovery of order (\mathbf{s}, \mathbf{M}) and (P2) reduces to (P1).

This result allows uniform recovery within the class of (\mathbf{s}, \mathbf{M}) -sparse vectors but the requirement on $\delta_{2\mathbf{s},\mathbf{M}}$ depends on l and $\eta_{\mathbf{s},\mathbf{M}}$. We make the following observations:

1. If we pick a sparsity pattern that uses lots of levels then we will require a smaller RIP_L constant (note, however, that if we work with wavelets and \mathbf{M} corresponds to wavelet levels then the number of levels l is approximately $\log_2(N)$, so that the RIP_L constant only has to shrink like $1/\sqrt{\log_2(N)}$).
2. If we pick a sparsity pattern with fewer levels then to explain the excellent observed recovery in applications we shall have choose s_i so that s_i/s_j is correspondingly larger for distinct i and j .
3. If the RIP_L constant $\delta_{2s,\mathbf{M}}$ is sufficiently small so that the conclusion of Theorem 4.4 holds, the bound on $\|x - \tilde{x}\|_2$ is weaker than the bound (1.4) obtained using the RIP.

As a consequence of these observations, at first glance it may appear that the results we have obtained with the RIP_L are weaker than those given using the standard RIP. However, Theorem 4.4 is stronger than any theorem based around uniform recovery in two senses. First, if one considers a sparsity pattern with one level then the bounds (4.2) and (4.3) reduce to (1.3) and (1.4), respectively. Second, neither (1.1) nor (1.3) applies at all if we do not have uniform recovery. Therefore, for the examples given in Figure 2, (1.1) and (1.3) are not applicable.

Ideally, it would be possible to find a constant C such that if the RIP_L constant is smaller than C then recovery of all (\mathbf{s}, \mathbf{M}) -sparse vectors would be possible. Unfortunately, we shall demonstrate that this is impossible in Theorems 4.5 and 4.6. Indeed, in some sense Theorem 4.4 is optimal in l and $\eta_{s,\mathbf{M}}$, as the following results confirm.

Theorem 4.5 (RIP_L dependence on the ratio constant). *Fix $a \in \mathbb{N}$ and $f : \mathbb{R} \rightarrow \mathbb{R}$ such that $f(\eta_{s,\mathbf{M}}) = o(\eta_{s,\mathbf{M}}^{\frac{1}{2}})$. Then there are $m, n \in \mathbb{N}$, a matrix $U' \in \mathbb{C}^{m \times n}$ and a sparsity pattern (\mathbf{s}, \mathbf{M}) with two levels that covers U' such that the RIP_L constant $\delta_{as,\mathbf{M}}$ and ratio constant $\eta_{s,\mathbf{M}}$ satisfy*

$$(4.4) \quad \delta_{as,\mathbf{M}} \leq \frac{1}{|f(\eta_{s,\mathbf{M}})|}$$

but there is an (\mathbf{s}, \mathbf{M}) -sparse z^1 such that

$$z^1 \notin \arg \min \|z\|_1 \text{ such that } U'z = U'z^1.$$

Roughly speaking, Theorem 4.5 says that if we fix the number of levels and try to replace the condition

$$\delta_{2s,\mathbf{M}} < \frac{1}{\sqrt{l(\sqrt{\eta_{s,\mathbf{M}}} + \frac{1}{4})^2 + 1}}$$

with a condition of the form $\delta_{2s,\mathbf{M}} < (\eta_{s,\mathbf{M}})^{-\alpha/2} / (C\sqrt{l})$ for some constant C and some $\alpha < 1$ then the conclusion of Theorem 4.4 ceases to hold. In particular, the requirement on $\delta_{2s,\mathbf{M}}$ cannot be independent of $\eta_{s,\mathbf{M}}$. The parameter a in the statement of Theorem 4.5 says that we cannot simply fix the issue by changing $\delta_{2s,\mathbf{M}}$ to $\delta_{3s,\mathbf{M}}$ or any further multiple of \mathbf{s} .

Similarly, we can state and prove a similar theorem that shows that the dependence on the number of levels, l , cannot be ignored.

Theorem 4.6 (RIP_L dependence on the number of levels). Fix $a \in \mathbb{N}$ and $f : \mathbb{R} \rightarrow \mathbb{R}$ such that $f(l) = o(l^{\frac{1}{2}})$. Then there are $m, n \in \mathbb{N}$, a matrix $U' \in \mathbb{C}^{m \times n}$, and a sparsity pattern (\mathbf{s}, \mathbf{M}) that covers U' with ratio constant $\eta_{\mathbf{s}, \mathbf{M}} = 1$ and l levels such that the RIP_L constant $\delta_{a\mathbf{s}, \mathbf{M}}$ corresponding to U' satisfies $\delta_{a\mathbf{s}, \mathbf{M}} \leq 1/|f(l)|$ but there is an (\mathbf{s}, \mathbf{M}) -sparse z^1 such that

$$z^1 \notin \arg \min \|z\|_1 \text{ such that } U'z = U'z^1.$$

Furthermore, Theorem 4.7 shows that the ℓ^2 error estimate on $\|x - \tilde{x}\|_2$ is optimal up to constant terms.

Theorem 4.7 (RIP_L ℓ^2 error optimality). The ℓ^2 result (4.3) in Theorem 4.4 is sharp in the following sense:

1. For a fixed $a \in \mathbb{N}$ and any functions $f, g : \mathbb{R} \rightarrow \mathbb{R}$ such that $f(\eta) = o(\eta^{\frac{1}{4}})$ and $g(\eta) = O(\sqrt{\eta})$, there are natural numbers m and n , a matrix $U' \in \mathbb{C}^{m \times n}$, and a sparsity pattern (\mathbf{s}, \mathbf{M}) with two levels such that
 - (\mathbf{s}, \mathbf{M}) covers U' ,
 - the RIP_L constant corresponding to the sparsity pattern $(a\mathbf{s}, \mathbf{M})$, denoted by $\delta_{a\mathbf{s}, \mathbf{M}}$, satisfies $\delta_{a\mathbf{s}, \mathbf{M}} \leq 1/|g(\eta_{\mathbf{s}, \mathbf{M}})|$,
 - there exist vectors z and z^1 such that $U'(z - z^1) = 0$ and $\|z\|_1 \leq \|z^1\|_1$ but

$$\|z - z^1\|_2 > \frac{f(\eta_{\mathbf{s}, \mathbf{M}})}{\sqrt{\tilde{s}}} \sigma_{\mathbf{s}, \mathbf{M}}(z^1)_1.$$

2. For a fixed $a \in \mathbb{N}$ and any functions $f, g : \mathbb{R} \rightarrow \mathbb{R}$ such that $f(l) = o(l^{\frac{1}{4}})$ and $g(l) = O(\sqrt{l})$, there are natural numbers m and n , a matrix $U' \in \mathbb{C}^{m \times n}$, and a sparsity pattern (\mathbf{s}, \mathbf{M}) with $\eta_{\mathbf{s}, \mathbf{M}} = 1$ such that
 - (\mathbf{s}, \mathbf{M}) covers U' ,
 - the RIP_L constant corresponding to the sparsity pattern $(a\mathbf{s}, \mathbf{M})$, denoted by $\delta_{a\mathbf{s}, \mathbf{M}}$, satisfies $\delta_{a\mathbf{s}, \mathbf{M}} \leq 1/|g(l)|$,
 - there exist vectors z and z^1 such that $U'(z - z^1) = 0$ and $\|z\|_1 \leq \|z^1\|_1$ but

$$\|z - z^1\|_2 > \frac{f(l)}{\sqrt{\tilde{s}}} \sigma_{\mathbf{s}, \mathbf{M}}(z^1)_1.$$

As with the RIP in levels, we can obtain results on recovery using the ℓ^2 robust nullspace property of order (\mathbf{s}, \mathbf{M})

Theorem 4.8 (ℓ^2 RNSP of order (\mathbf{s}, \mathbf{M}) recovery theorem). Suppose that a matrix $U \in \mathbb{C}^{m \times n}$ satisfies the ℓ^2 robust nullspace property of order (\mathbf{s}, \mathbf{M}) with constants $\rho \in (0, 1)$ and $\tau > 0$. Let $x \in \mathbb{C}^n$ and $y \in \mathbb{C}^m$ satisfy $\|Ux - y\|_2 < \epsilon$. Then any solutions \tilde{x} of the ℓ^1 minimization problem

$$\min_{\hat{x} \in \mathbb{C}^n} \|\hat{x}\|_1 \text{ subject to } \|U\hat{x} - y\|_2 \leq \epsilon$$

satisfy

$$(4.5) \quad \|\tilde{x} - x\|_1 \leq A_1 \sigma_{\mathbf{s}, \mathbf{M}}(x)_1 + C_1 \epsilon \sqrt{\tilde{s}},$$

$$(4.6) \quad \|\tilde{x} - x\|_2 \leq \frac{\sigma_{\mathbf{s}, \mathbf{M}}(x)_1}{\sqrt{\tilde{s}}} \left(A_2 + B_2 \sqrt[4]{l \eta_{\mathbf{s}, \mathbf{M}}} \right) + 2\epsilon \left(C_2 + D_2 \sqrt[4]{l \eta_{\mathbf{s}, \mathbf{M}}} \right),$$

where

$$A_1 := \frac{2+2\rho}{1-\rho}, \quad C_1 := \frac{4\tau}{1-\rho}, \quad A_2 := \frac{2\rho+2\rho^2}{1-\rho},$$

$$B_2 := \frac{(2\sqrt{\rho}+1)(1+\rho)}{1-\rho}, \quad C_2 := \frac{\rho\tau+\tau}{1-\rho}, \quad \text{and} \quad D_2 := \frac{4\sqrt{\rho}\tau+3\tau-\rho\tau}{2-2\rho}.$$

This theorem explains where the dependence on $\eta_{\mathbf{s},\mathbf{M}}$ and l in (4.3) emerges from. One technique for showing that the RIP implies uniform recovery of order s is to prove that a sufficiently small RIC implies the nullspace property (for example, this method is used in [35]). In a similar way, we prove the ℓ^2 error estimate in Theorem 4.4 by showing that a sufficiently small RIP_L constant implies the robust ℓ^2 nullspace property of order (\mathbf{s}, \mathbf{M}) . The ℓ^2 error estimate (4.6) follows and we are left with a dependence on $\sqrt[4]{l\eta_{\mathbf{s},\mathbf{M}}}$ in the right-hand side of (4.3). As before, we can show that this is optimal. We do this in Theorem 4.9.

Theorem 4.9 (ℓ^2 RNSP of order (\mathbf{s}, \mathbf{M}) optimality). *The result in Theorem 4.8 is sharp, in the following sense:*

1. For any $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ satisfying $f(\rho, \tau, \eta) = o(\eta^{\frac{1}{4}})$ for fixed $\rho \in (0, 1)$ and $\tau > 0$, there are natural numbers m and n , a matrix $U' \in \mathbb{C}^{m \times n}$, and a sparsity pattern (\mathbf{s}, \mathbf{M}) with ratio constant $\eta_{\mathbf{s},\mathbf{M}}$ and two levels such that

- (\mathbf{s}, \mathbf{M}) covers U' ,
- U' satisfies the ℓ^2 robust nullspace property of order (\mathbf{s}, \mathbf{M}) with constants $\rho \in (0, 1)$ and $\tau > 0$,
- there exist vectors z and z^1 such that $U'(z - z^1) = 0$ and $\|z\|_1 \leq \|z^1\|_1$ but

$$\|z - z^1\|_2 > \frac{f(\rho, \tau, \eta_{\mathbf{s},\mathbf{M}})}{\sqrt{s}} \sigma_{\mathbf{s},\mathbf{M}}(z^1)_1.$$

2. For any $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ satisfying $f(\rho, \tau, l) = o(l^{\frac{1}{4}})$ for fixed $\rho \in (0, 1)$ and $\tau > 0$, there are natural numbers m and n , a matrix $U' \in \mathbb{C}^{m \times n}$, and a sparsity pattern (\mathbf{s}, \mathbf{M}) with ratio constant $\eta_{\mathbf{s},\mathbf{M}} = 1$ and l levels such that

- (\mathbf{s}, \mathbf{M}) covers U' ,
- U' satisfies the ℓ^2 robust nullspace property of order (\mathbf{s}, \mathbf{M}) with constants $\rho \in (0, 1)$ and $\tau > 0$,
- there exist vectors z and z^1 such that $U'(z - z^1) = 0$ and $\|z\|_1 \leq \|z^1\|_1$ but

$$\|z - z^1\|_2 > \frac{f(\rho, \tau, l)}{\sqrt{s}} \sigma_{\mathbf{s},\mathbf{M}}(z^1)_1.$$

The conclusions that we can draw from the above theorems are the following:

1. The RIP_L will guarantee uniform recovery of order (\mathbf{s}, \mathbf{M}) , provided that the RIP_L constant is sufficiently small (Theorem 4.4).
2. The requirement on the RIP_L constant to achieve uniform recovery of order (\mathbf{s}, \mathbf{M}) is dependent on $\sqrt{\eta_{\mathbf{s},\mathbf{M}}}$ and \sqrt{l} . This is optimal up to constants (Theorem 4.5 and Theorem 4.6).
3. When compared to the error estimates obtained using the RIP, the ℓ^2 error when using the RIP_L has additional factors of the form $\sqrt[4]{l}$ and $\sqrt[4]{\eta_{\mathbf{s},\mathbf{M}}}$. Again, these are optimal up to constants (Theorem 4.7).

4. The same additional factors of the form $\sqrt[4]{l}$ and $\sqrt[4]{\eta_{\mathbf{s}, \mathbf{M}}}$ on the ℓ^2 error estimate (Theorem 4.8) are also present with the robust ℓ^2 nullspace property of order (\mathbf{s}, \mathbf{M}) .
5. These factors are optimal up to constants, so that even if we ignore the RIP_L and still try to prove results using the ℓ^2 robust nullspace property of order (\mathbf{s}, \mathbf{M}) then we would be unable to improve the ℓ^2 error (Theorem 4.9).

With these results, we have demonstrated that the RIP in levels may be able to explain why permutations within levels are possible and why more general permutations are impossible with compressed sensing for the matrices in 2 (similar numerical arguments can be used whenever the matrix is nearly block diagonal). The results that we have obtained give a sufficient condition on the RIP in levels constant that guarantees (\mathbf{s}, \mathbf{M}) -sparse recovery. Furthermore, we have managed to demonstrate that this condition and the conclusions that follow from it are optimal up to constants.

5. Conclusions and open problems. The flip test demonstrates that in many practical applications the ability to recover sparse signals depends on the structure of the sparsity, so that a tool that guarantees uniform recovery of all s -sparse signals does not apply. In particular, this is the case when considering the broadly applicable problem of recovery from samples of Fourier or Hadamard measurements. The flip test in levels suggests that reasonable sampling schemes provide a different form of uniform recovery, namely, the recovery of (\mathbf{s}, \mathbf{M}) -sparse signals. It is therefore natural to try to find theoretical tools that are able to analyze and describe this phenomenon. To generalize the RIP, we proposed the RIP_L . Using the results from [50], in section 3.2 we were able to produce a collection of matrices which exhibit the RIP_L . However, we are now left with the following fundamental problems:

- Is it possible to expand the analysis done in [50] and section 3.2 to show that a wider variety of randomly sampled matrices exhibit the RIP_L with a high probability? As in section 3.2, we believe this can be achieved using results from, for example, [45].
- Can one improve on the number of measurements required in Theorems 3.10 and 3.11 (perhaps by reducing log factors) or else show that the bounds in (3.4) and (3.5) are optimal?
- In section 3.2 we considered the size of the RIC_L constant $\delta_{\mathbf{s}, \mathbf{M}}$ when the measurement locations Ω were selected at random using a multilevel subsampling scheme. Instead, suppose that the measurements Ω and the matrix M are fixed. We can then ask the important open question: for which (\mathbf{s}, \mathbf{M}) does the matrix $P_\Omega M$ satisfy the RIP_L with sufficiently small constant $\delta_{\mathbf{s}, \mathbf{M}}$ for Theorem 4.4 to apply?

Note that these problems are vast as the sampling patterns will depend not only on the sparsity patterns but of course also on the sampling basis and recovery basis (or frame). Thus, covering all interesting cases relevant to practical applications will yield an incredibly rich mathematical theory.

6. Proofs. We shall present the proofs in an arrangement different from the order in which their statements were presented. The first proof that we shall present is that of Theorem 4.8.

6.1. Proof of Theorem 4.8. We begin with the following lemma.

Lemma 6.1. *Suppose that $U \in \mathbb{C}^{m \times n}$ satisfies the ℓ^2 robust nullspace property of order (\mathbf{s}, \mathbf{M}) with constants $\rho \in (0, 1)$ and $\tau > 0$. Fix $v \in \mathbb{C}^n$, and let S be an (\mathbf{s}, \mathbf{M}) -sparse set such*

that $|S| = \tilde{s}$ and the property that if T is an (\mathbf{s}, \mathbf{M}) -sparse set, we have $\|v_S\|_1 \geq \|v_T\|_1$. Then

$$\|v\|_2 \leq \frac{\|v_{S^c}\|_1}{\sqrt{\tilde{s}}} \left[\rho + \left(\sqrt{\rho} + \frac{1}{2} \right) \sqrt[4]{l\eta_{\mathbf{s}, \mathbf{M}}} \right] + \tau \|Uv\|_2 \left[\frac{\sqrt[4]{l\eta_{\mathbf{s}, \mathbf{M}}}}{2} + 1 \right].$$

Proof. For $i = 1, 2, \dots, l$, we define S_0^i to be $S_0^i = S \cap \{M_{i-1} + 1, M_{i-1} + 2, \dots, M_i\}$ (i.e., S_0^i is the elements of S that are in the i th level). Let $m = \max_{i=1,2,\dots,l} \min_{j \in S_0^i} |v_j|$. Since $|S_0^i| = s_i$ (otherwise $|S| < \tilde{s}$), we can see that given any $i = 1, 2, \dots, l$

$$\|v_S\|_2 = \sqrt{\sum_{n \in S} |v_n|^2} \geq \sqrt{\sum_{j \in S_0^i} |v_j|^2} \geq \sqrt{s_i} \min_{j \in S_0^i} |v_j| \geq \min_{k=1,2,\dots,l} \sqrt{s_k} \min_{j \in S_0^i} |v_j|$$

so that $\|v_S\|_2 \geq m \min_{k=1,2,\dots,l} \sqrt{s_k}$. Furthermore, $|v_j| \leq m$ for each $j \in S^c$; otherwise there is an (\mathbf{s}, \mathbf{M}) -sparse T with $\|v_T\|_1 > \|v_S\|_1$. Therefore $\|v_{S^c}\|_2^2 = \sum_{j \in S^c} |v_j|^2 \leq \sum_{j \in S^c} m |v_j| \leq \frac{\|v_{S^c}\|_1 \|v_S\|_2}{\min_{k=1,2,\dots,l} \sqrt{s_k}}$. By the ℓ^2 robust nullspace property of order (\mathbf{s}, \mathbf{M}) , $\|v_{S^c}\|_1 \|v_S\|_2 \leq \frac{\rho}{\sqrt{\tilde{s}}} \|v_{S^c}\|_1^2 + \tau \|Uv\|_2 \|v_{S^c}\|_1$. Since $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ whenever $a, b > 0$,

$$(6.1) \quad \|v_{S^c}\|_2 \leq \frac{1}{\min \sqrt[4]{s_i}} \left(\frac{\sqrt{\rho}}{\sqrt[4]{\tilde{s}}} \|v_{S^c}\|_1 + \sqrt{\tau \|Uv\|_2 \|v_{S^c}\|_1} \right).$$

Using the arithmetic-geometric mean inequality,

$$\sqrt{\tau \|Uv\|_2 \|v_{S^c}\|_1} = \sqrt{\tau \|Uv\|_2 \sqrt[4]{\tilde{s}} \frac{\|v_{S^c}\|_1}{\sqrt[4]{\tilde{s}}}} \leq \frac{\tau \|Uv\|_2 \sqrt[4]{\tilde{s}}}{2} + \frac{\|v_{S^c}\|_1}{2 \sqrt[4]{\tilde{s}}}.$$

Therefore, (6.1) yields

$$\begin{aligned} \|v_{S^c}\|_2 &\leq \frac{1}{\min \sqrt[4]{s_i}} \left(\frac{\sqrt{\rho}}{\sqrt[4]{\tilde{s}}} \|v_{S^c}\|_1 + \frac{\|v_{S^c}\|_1}{2 \sqrt[4]{\tilde{s}}} + \frac{\tau \|Uv\|_2 \sqrt[4]{\tilde{s}}}{2} \right) \\ &\leq \frac{\|v_{S^c}\|_1}{\sqrt[4]{\tilde{s}} \min \sqrt[4]{s_i}} \left(\sqrt{\rho} + \frac{1}{2} \right) + \frac{\tau \|Uv\|_2 \sqrt[4]{l\eta_{\mathbf{s}, \mathbf{M}}}}{2} \end{aligned}$$

because $\frac{\tilde{s}}{\min s_i} \leq l\eta_{\mathbf{s}, \mathbf{M}}$. Once again, employing the ℓ^2 nullspace property gives

$$\begin{aligned} \|v\|_2 &\leq \|v_S\|_2 + \|v_{S^c}\|_2 \leq \frac{\rho}{\sqrt{\tilde{s}}} \|v_{S^c}\|_1 + \tau \|Uv\|_2 + \frac{\|v_{S^c}\|_1}{\sqrt[4]{\tilde{s}} \min \sqrt[4]{s_i}} \left(\sqrt{\rho} + \frac{1}{2} \right) + \frac{\tau \|Uv\|_2 \sqrt[4]{l\eta_{\mathbf{s}, \mathbf{M}}}}{2} \\ &\leq \frac{\|v_{S^c}\|_1}{\sqrt{\tilde{s}}} \left[\rho + \left(\sqrt{\rho} + \frac{1}{2} \right) \frac{\sqrt[4]{\tilde{s}}}{\min \sqrt[4]{s_i}} \right] + \tau \|Uv\|_2 \left[\frac{\sqrt[4]{l\eta_{\mathbf{s}, \mathbf{M}}}}{2} + 1 \right] \\ &\leq \frac{\|v_{S^c}\|_1}{\sqrt{\tilde{s}}} \left[\rho + \left(\sqrt{\rho} + \frac{1}{2} \right) \sqrt[4]{l\eta_{\mathbf{s}, \mathbf{M}}} \right] + \tau \|Uv\|_2 \left[\frac{\sqrt[4]{l\eta_{\mathbf{s}, \mathbf{M}}}}{2} + 1 \right]. \quad \blacksquare \end{aligned}$$

We shall see that the remaining error estimates will follow from the ℓ^1 robust nullspace property (see [35, Definition 4.17]). This definition and its generalization to (\mathbf{s}, \mathbf{M}) -sparse sets are as follows:

Definition 6.2. A matrix $U \in \mathbb{C}^{m \times n}$ satisfies the ℓ^1 robust nullspace property relative to S with constants $\rho \in (0, 1)$ and $\tau' > 0$ if

$$(6.2) \quad \|v_S\|_1 \leq \rho \|v_{S^c}\|_1 + \tau' \|Uv\|_2$$

for any $v \in \mathbb{C}^n$. We say that U satisfies the ℓ^1 robust nullspace property of order (\mathbf{s}, \mathbf{M}) if (6.2) holds for any (\mathbf{s}, \mathbf{M}) -sparse sets S .

It is easy to see that if U satisfies the ℓ^2 robust nullspace property of order (\mathbf{s}, \mathbf{M}) with constants ρ and τ then, for any (\mathbf{s}, \mathbf{M}) -sparse set S , U also satisfies the ℓ^1 robust nullspace property relative to S with constants ρ and $\tau\sqrt{\tilde{s}}$. Indeed, assume that U satisfies the ℓ^2 robust nullspace property of order (\mathbf{s}, \mathbf{M}) with constants ρ and τ . Then (by the Cauchy–Schwarz inequality) $\|v_S\|_1 \leq \sqrt{\tilde{s}} \|v_S\|_2 \leq \rho \|v_{S^c}\|_1 + \tau\sqrt{\tilde{s}} \|Uv\|_2$.

An immediate conclusion of the robust nullspace property is the following, proven in [35] as Theorem 4.20.

Lemma 6.3. Suppose that $U \in \mathbb{C}^{m \times n}$ satisfies the ℓ^1 robust nullspace property with constants $\rho \in (0, 1)$ and τ' relative to a set S . Then for any complex vectors $x, z \in \mathbb{C}^n$, we have

$$\|z - x\|_1 \leq \frac{1 + \rho}{1 - \rho} (\|z\|_1 - \|x\|_1 + 2\|x_{S^c}\|_1) + \frac{2\tau'}{1 - \rho} \|U(z - x)\|_2.$$

We can use this lemma to show the following important result, which is similar both in proof and statement to Theorem 4.19 in [35].

Lemma 6.4. Suppose that a matrix $U \in \mathbb{C}^{m \times n}$ satisfies the ℓ^1 robust nullspace property of order (\mathbf{s}, \mathbf{M}) with constants $\rho \in (0, 1)$ and $\tau' > 0$. Furthermore, suppose that $\|Ux - y\|_2 \leq \epsilon$. Then any solutions \tilde{x} to the ℓ^1 minimization problem

$$\min_{\hat{x} \in \mathbb{C}^n} \|\hat{x}\|_1 \text{ subject to } \|U\hat{x} - y\|_2 \leq \epsilon$$

satisfy

$$\|x - \tilde{x}\|_1 \leq \frac{2 + 2\rho}{1 - \rho} \sigma_{\mathbf{s}, \mathbf{M}}(x)_1 + \frac{4\tau'\epsilon}{1 - \rho}.$$

Proof. By Lemma 6.3, for any (\mathbf{s}, \mathbf{M}) -sparse set S

$$\|\tilde{x} - x\|_1 \leq \frac{1 + \rho}{1 - \rho} (\|\tilde{x}\|_1 - \|x\|_1 + 2\|x_{S^c}\|_1) + \frac{2\tau'}{1 - \rho} \|U(\tilde{x} - x)\|_2.$$

Because both $\|Ux - y\|_2$ and $\|U\tilde{x} - y\|_2$ are smaller than or equal to ϵ , $\|Ux - U\tilde{x}\|_2 \leq 2\epsilon$. Furthermore, because \tilde{x} has minimal ℓ^1 norm, $\|\tilde{x}\|_1 - \|x\|_1 \leq 0$.

Thus $\|x - \tilde{x}\|_1 \leq \frac{2+2\rho}{1-\rho} \|x_{S^c}\|_1 + \frac{4\tau'\epsilon}{1-\rho}$. If we take S to be the (\mathbf{s}, \mathbf{M}) -sparse set which maximizes $\|x_S\|_1$, then

$$\|x - \tilde{x}\|_1 \leq \frac{2 + 2\rho}{1 - \rho} \sigma_{\mathbf{s}, \mathbf{M}}(x)_1 + \frac{4\tau'\epsilon}{1 - \rho}. \quad \blacksquare$$

We can combine these results to complete the proof of Theorem 4.8. Indeed, (4.5) follows immediately from Lemma 6.4 and the fact that U satisfies the ℓ^1 robust nullspace property with constants ρ and $\tau\sqrt{s}$. To prove (4.6), we can simply set $v = x - \tilde{x}$ in Lemma 6.1 to see that

$$\begin{aligned} \|x - \tilde{x}\|_2 &\leq \frac{\|(x - \tilde{x})_{S^c}\|_1}{\sqrt{s}} \left[\rho + \left(\sqrt{\rho} + \frac{1}{2} \right) \sqrt[4]{l\eta_{s,\mathbf{M}}} \right] + \tau \|U(x - \tilde{x})\|_2 \left[\frac{\sqrt[4]{l\eta_{s,\mathbf{M}}}}{2} + 1 \right] \\ &\leq \frac{\|x - \tilde{x}\|_1}{\sqrt{s}} \left[\rho + \left(\sqrt{\rho} + \frac{1}{2} \right) \sqrt[4]{l\eta_{s,\mathbf{M}}} \right] + 2\tau\epsilon \left[\frac{\sqrt[4]{l\eta_{s,\mathbf{M}}}}{2} + 1 \right] \end{aligned}$$

and the result follows from (4.5).

6.2. Proof of Theorem 4.4. It will suffice to prove that the conditions on $\delta_{s,\mathbf{M}}$ and (s, \mathbf{M}) in Theorem 4.4 imply the ℓ^2 robust nullspace property. To show this, we begin by stating the following inequality, proven in [11].

Lemma 6.5 (the norm inequality for ℓ^1 and ℓ^2). *Let $v = (v_1, v_2, \dots, v_s)$, where $v_1 \geq v_2 \geq v_3 \geq \dots \geq v_s$. Then*

$$\|v\|_2 \leq \frac{1}{\sqrt{s}} \|v\|_1 + \frac{\sqrt{s}}{4} (v_1 - v_s).$$

We will now prove the following additional lemma, which is almost identical in statement and proof to that of Lemma 6.1 in [6].

Lemma 6.6. *Suppose that $x, y \in \Sigma_{s,\mathbf{M}}$ and that*

$$(6.3) \quad \|Ux\|_2^2 - \|x\|_2^2 = t\|x\|_2^2.$$

Additionally, suppose that x and y are orthogonal. Then $|\langle Ux, Uy \rangle| \leq \sqrt{\delta_{2s,\mathbf{M}}^2 - t^2} \|x\|_2 \|y\|_2$, where $\delta_{2s,\mathbf{M}}$ is the restricted isometry constant corresponding to the sparsity pattern $(2s, \mathbf{M})$ and the matrix U .

Proof. Without loss of generality, we can assume that $\|x\|_2 = \|y\|_2 = 1$. Note that for $\alpha, \beta \in \mathbb{R}$ and $\gamma \in \mathbb{C}$, the vectors $\alpha x + \gamma y$ and $\beta x - \gamma y$ are contained in $\Sigma_{2s,\mathbf{M}}$. Therefore,

$$(6.4) \quad \|U(\alpha x + \gamma y)\|_2^2 \leq (1 + \delta_{2s,\mathbf{M}}) \|\alpha x + \gamma y\|_2^2 = (1 + \delta_{2s,\mathbf{M}})(\alpha^2 + |\gamma|^2),$$

where the last line follows because $\langle x, y \rangle = 0$ (from the orthogonality of x and y). Similarly,

$$(6.5) \quad -\|U(\beta x - \gamma y)\|_2^2 \leq -(1 - \delta_{2s,\mathbf{M}})(\beta^2 + |\gamma|^2).$$

We will now add these two inequalities. On the one hand (by using the assumption in (6.3) and the fact that α, β are real), we have

$$\begin{aligned} \|U(\alpha x + \gamma y)\|_2^2 - \|U(\beta x - \gamma y)\|_2^2 &= \alpha^2 \|Ux\|_2^2 + 2\text{Re}(\alpha\bar{\gamma} \langle Ux, Uy \rangle) + |\gamma|^2 \|Uy\|_2^2 \\ &\quad - (\beta^2 \|Ux\|_2^2 - 2\text{Re}(\beta\bar{\gamma} \langle Ux, Uy \rangle) + |\gamma|^2 \|Uy\|_2^2) \\ &= (1 + t)(\alpha^2 - \beta^2) + 2(\alpha + \beta)\text{Re}(\bar{\gamma} \langle Ux, Uy \rangle), \end{aligned}$$

and on the other hand (from (6.4) and (6.5))

$$\|U(\alpha x + \gamma y)\|_2^2 - \|U(\beta x - \gamma y)\|_2^2 \leq \delta_{2s, \mathbf{M}} (\alpha^2 + \beta^2 + 2|\gamma|^2) + \alpha^2 - \beta^2.$$

Therefore,

$$(1+t)(\alpha^2 - \beta^2) + 2(\alpha + \beta)\operatorname{Re}(\bar{\gamma}\langle Ux, Uy \rangle) \leq \delta_{2s, \mathbf{M}} (\alpha^2 + \beta^2 + 2|\gamma|^2) + \alpha^2 - \beta^2.$$

After choosing γ so that $\operatorname{Re}(\bar{\gamma}\langle Ux, Uy \rangle) = |\langle Ux, Uy \rangle|$ we obtain

$$(6.6) \quad |\langle Ux, Uy \rangle| \leq \frac{1}{2\alpha + 2\beta} [(\delta_{2s, \mathbf{M}} - t)\alpha^2 + (\delta_{2s, \mathbf{M}} + t)\beta^2 + 2\delta_{2s, \mathbf{M}}]$$

because $|\gamma| = 1$. By the definition of the RIP in levels constant, $\delta_{2s, \mathbf{M}} \geq \delta_{s, \mathbf{M}}$ and so

$$(6.7) \quad |t| = \left| \|Ux\|_2^2 - \|x\|_2^2 \right| \leq \delta_{s, \mathbf{M}} \leq \delta_{2s, \mathbf{M}}.$$

If equality holds in (6.7), then we can set $\beta = 0$ and send $\alpha \rightarrow \infty$ in (6.6) to obtain the required result. Otherwise, (6.7) implies that $\sqrt{\frac{\delta_{2s, \mathbf{M}} + t}{\delta_{2s, \mathbf{M}} - t}} \in \mathbb{R}$ and so we can set $\alpha = \sqrt{\frac{\delta_{2s, \mathbf{M}} + t}{\delta_{2s, \mathbf{M}} - t}}$ and $\beta = \frac{1}{\alpha}$ in (6.6). With these values, we obtain

$$\begin{aligned} |\langle Ux, Uy \rangle| &\leq \frac{\alpha}{2\alpha^2 + 2} (\delta_{2s, \mathbf{M}} + t + \delta_{2s, \mathbf{M}} - t + 2\delta_{2s, \mathbf{M}}) \leq \frac{4\delta_{2s, \mathbf{M}}\alpha(\delta_{2s, \mathbf{M}} - t)}{4\delta_{2s, \mathbf{M}}} \\ &\leq \sqrt{\delta_{2s, \mathbf{M}}^2 - t^2}. \end{aligned} \quad \blacksquare$$

Proof of Theorem 4.4. Let $x \in \mathbb{C}^m$ be an arbitrary m dimensional complex vector, and let

$$x^i := x_{\{M_{i-1}+1, M_{i-1}+2, \dots, M_i\}}$$

denote the i th level of x . For an arbitrary vector $v = (v_1, v_2, \dots, v_n)$, we define $|v|$ to be the vector $(|v_1|, |v_2|, \dots, |v_n|)$. Let S_0^i denote the indexes of the s_i th largest elements of $|x^i|$, and $S_0 := \bigcup_{i=1}^l S_0^i$. We then define S_1^i to be the indexes of the s_i th largest elements of $|x^i|$ that are not contained in S_0^i (if there are fewer than s_i elements remaining, we simply take the indexes of any remaining elements of $|x^i|$ and define $S_1 := \bigcup_{i=1}^l S_1^i$). In general, we can make a similar definition to form a collection of index sets labeled $(S_j^i)_{i=1,2,\dots,l, j=1,2,\dots}$ and corresponding (s, \mathbf{M}) -sparse S_j .

These definitions and the fact that (s, \mathbf{M}) covers U implies that if $\Omega = \bigcup_{j \geq 0} S_j$ then $x_\Omega = x$. By the definition of S_0 , $\|x_\Lambda\|_2 \leq \|x_{S_0}\|_2$ and $\|x_{S_0^c}\|_1 \leq \|x_{\Lambda^c}\|_1$ whenever Λ is (s, \mathbf{M}) -sparse. It will suffice to verify that

$$(6.8) \quad \sqrt{\tilde{s}}\|x_{S_0}\|_2 \leq \rho\|x_{S_0^c}\|_1 + \tau\sqrt{\tilde{s}}\|Ux\|_2$$

holds for some $\rho \in (0, 1)$ and $\tau > 0$. Indeed, if (6.8) holds then for (s, \mathbf{M}) -sparse sets Λ ,

$$\begin{aligned} \sqrt{\tilde{s}}\|x_\Lambda\|_2 &\leq \sqrt{\tilde{s}}\|x_{S_0}\|_2 \leq \rho\|x_{S_0^c}\|_1 + \tau\sqrt{\tilde{s}}\|Ux\|_2 \\ &\leq \rho\|x_{\Lambda^c}\|_1 + \tau\sqrt{\tilde{s}}\|Ux\|_2 \end{aligned}$$

as required in Theorem 4.8. Set

$$(6.9) \quad \|Ux_{S_0}\|_2^2 = (1+t)\|x_{S_0}\|_2^2.$$

Clearly, $|t| \leq \delta_{s,M}$. Then

$$(6.10) \quad \|Ux_{S_0}\|_2^2 = \langle Ux_{S_0}, Ux_{S_0} \rangle = \langle Ux_{S_0}, Ux \rangle - \sum_{j \geq 1} \langle Ux_{S_0}, Ux_{S_j} \rangle,$$

where we have used $x_\Omega = x$. Using the Cauchy–Schwarz inequality and (6.9) yields

$$(6.11) \quad |\langle Ux_{S_0}, Ux \rangle| \leq \|Ux_{S_0}\|_2 \|Ux\|_2 \leq \sqrt{1+t} \|x_{S_0}\|_2 \|Ux\|_2.$$

Furthermore, we can use Lemma 6.6 to see that

$$(6.12) \quad \left| \sum_{j \geq 1} \langle Ux_{S_0}, Ux_{S_j} \rangle \right| \leq \sqrt{\delta_{2s,M}^2 - t^2} \sum_{j \geq 1} \|x_{S_0}\|_2 \|x_{S_j}\|_2 \leq \|x_{S_0}\|_2 \sqrt{\delta_{2s,M}^2 - t^2} \sum_{i=1}^l \sum_{j \geq 1} \|x_{S_j^i}\|_2.$$

Combining (6.9), (6.10), (6.11), and (6.12) yields

$$(6.13) \quad (1+t)\|x_{S_0}\|_2^2 \leq \sqrt{1+t} \|x_{S_0}\|_2 \|Ux\|_2 + \|x_{S_0}\|_2 \sqrt{\delta_{2s,M}^2 - t^2} \sum_{i=1}^l \sum_{j \geq 1} \|x_{S_j^i}\|_2.$$

If $|S_j^i| = s_i$ then let $x_{i,j}^+$ (correspondingly $x_{i,j}^-$) be the largest element of $|x_{S_j^i}|$ (correspondingly the smallest element of $|x_{S_j^i}|$). If S_j^i is nonempty with fewer than s_i elements then we set $x_{i,j}^+$ to be the largest element of $|x_{S_j^i}|$ and $x_{i,j}^- = 0$. Finally, when $S_j^i = \emptyset$, we let $x_{i,j}^+ = x_{i,j}^- = 0$. It is clear then that $x_{i,j+1}^+ \leq x_{i,j}^-$.

Since $x_{S_j^i}$ contains at most s_i nonzero elements, we can apply the norm inequality for ℓ^1 and ℓ^2 (Lemma 6.5) to obtain

$$\|x_{S_j^i}\|_2 \leq \frac{1}{\sqrt{s_i}} \|x_{S_j^i}\|_1 + \frac{\sqrt{s_i}}{4} (x_{i,j}^+ - x_{i,j}^-)$$

for any $i = 1, 2, \dots, l$ and $j \in \mathbb{N}$. Therefore

$$\begin{aligned} \sum_{j \geq 1} \|x_{S_j^i}\|_2 &\leq \sum_{j \geq 1} \left(\frac{1}{\sqrt{s_i}} \|x_{S_j^i}\|_1 \right) + \frac{\sqrt{s_i}}{4} \sum_{j \geq 1} (x_{i,j}^+ - x_{i,j}^-) \\ &= \sum_{j \geq 1} \left(\frac{1}{\sqrt{s_i}} \|x_{S_j^i}\|_1 \right) + \frac{\sqrt{s_i}}{4} \left(x_{i,1}^+ + \sum_{j \geq 2} x_{i,j}^+ - \sum_{j \geq 1} x_{i,j}^+ \right) \\ &= \sum_{j \geq 1} \left(\frac{1}{\sqrt{s_i}} \|x_{S_j^i}\|_1 \right) + \frac{\sqrt{s_i}}{4} \left(x_{i,1}^+ + \sum_{j \geq 1} (x_{i,j+1}^+ - x_{i,j}^-) \right) \\ &\leq \sum_{j \geq 1} \left(\frac{1}{\sqrt{s_i}} \|x_{S_j^i}\|_1 \right) + \frac{\sqrt{s_i}}{4} x_{i,1}^+, \end{aligned}$$

where the last inequality follows because $x_{i,j+1}^+ - x_{i,j}^- \leq 0$. Additionally, $\sqrt{s_i}x_{i,1}^+ = \sqrt{s_i}\|x_{S_1^i}\|_\infty \leq \|x_{S_0^i}\|_2$ because each element of $|x_{S_0^i}|$ is larger than $x_{i,1}^+$. We conclude that

$$\begin{aligned} \sum_{i=1}^l \sum_{j \geq 1} \|x_{S_j^i}\|_2 &\leq \sum_{j \geq 1} \sum_{i=1}^l \frac{1}{\sqrt{s_i}} \|x_{S_j^i}\|_1 + \sum_{i=1}^l \frac{1}{4} \|x_{S_0^i}\|_2 \leq \frac{1}{\min \sqrt{s_i}} \sum_{j \geq 1} \sum_{i=1}^l \|x_{S_j^i}\|_1 + \frac{1}{4} \sqrt{l} \|x_{S_0}\|_2 \\ &\leq \frac{1}{\min \sqrt{s_i}} \sum_{j \geq 1} \|x_{S_j}\|_1 + \frac{1}{4} \sqrt{l} \|x_{S_0}\|_2 \leq \frac{1}{\min \sqrt{s_i}} \left\| x \bigcup_{j \geq 1} S_j \right\|_1 + \frac{1}{4} \sqrt{l} \|x_{S_0}\|_2, \end{aligned}$$

where the second inequality follows from the Cauchy–Schwarz inequality applied to

$$\underbrace{(1, 1, \dots, 1)}_l$$

and $(\|x_{S_0^1}\|_2, \|x_{S_0^2}\|_2, \dots, \|x_{S_0^l}\|_2)$ and the third and fourth inequalities follow from the disjoint supports of the vectors $x_{S_j^i}$ and $x_{S_{j'}^{i'}}$ whenever $i \neq i'$ or $j \neq j'$. By $x_\Omega = x$ and the disjointedness of S_i, S_j for $i \neq j$, $\bigcup_{j \geq 1} S_j = S_0^c$ so

$$(6.14) \quad \sum_{i=1}^l \sum_{j \geq 1} \|x_{S_j^i}\|_2 \leq \frac{1}{\min \sqrt{s_i}} \|x_{S_0^c}\|_1 + \frac{1}{4} \sqrt{l} \|x_{S_0}\|_2.$$

Dividing (6.13) by $\|x_{S_0}\|_2$ and employing (6.14) yields

$$(6.15) \quad (1+t)\|x_{S_0}\|_2 \leq \sqrt{1+t} \|Ux\|_2 + \sqrt{\delta_{2s,M}^2 - t^2} \left(\frac{1}{\min \sqrt{s_i}} \|x_{S_0^c}\|_1 + \frac{1}{4} \sqrt{l} \|x_{S_0}\|_2 \right).$$

Let $g(t) := \frac{\delta_{2s,M}^2 - t^2}{(1+t)^2}$ for $|t| \leq \delta_{2s,M}$. It is clear that $g(\delta_{2s,M}) = g(-\delta_{2s,M}) = 0$. Furthermore, g is differentiable. Therefore g attains its maximum at t_{\max} , where $g'(t_{\max}) = 0$. A simple calculation shows us that $t_{\max} = -\delta_{2s,M}^2$ (note that by the assumption (4.1), $\delta_{2s,M}^2 \leq \delta_{2s,M}$).

Thus $g(t) \leq g(-\delta_{2s,M}^2) = \frac{\delta_{2s,M}^2}{1 - \delta_{2s,M}^2}$. Additionally, $\frac{1}{\sqrt{1+t}} \leq \frac{1}{\sqrt{1 - \delta_{2s,M}^2}}$. Combining this with (6.15) yields

$$\begin{aligned} \|x_{S_0}\|_2 &\leq \frac{1}{\sqrt{1+t}} \|Ux\|_2 + \sqrt{g(t)} \left(\frac{1}{\min \sqrt{s_i}} \|x_{S_0^c}\|_1 + \frac{1}{4} \sqrt{l} \|x_{S_0}\|_2 \right) \\ &\leq \frac{1}{\sqrt{1 - \delta_{2s,M}^2}} \|Ux\|_2 + \frac{\delta_{2s,M}}{\sqrt{1 - \delta_{2s,M}^2}} \left(\frac{1}{\min \sqrt{s_i}} \|x_{S_0^c}\|_1 + \frac{1}{4} \sqrt{l} \|x_{S_0}\|_2 \right). \end{aligned}$$

A simple rearrangement gives

$$(6.16) \quad \|x_{S_0}\|_2 \leq \frac{\sqrt{1 + \delta_{2s,M}}}{\sqrt{1 - \delta_{2s,M}^2} - \delta_{2s,M} \sqrt{l}/4} \|Ux\|_2 + \frac{\delta_{2s,M}}{\min \sqrt{s_i} \left(\sqrt{1 - \delta_{2s,M}^2} - \delta_{2s,M} \sqrt{l}/4 \right)} \|x_{S_0^c}\|_1$$

provided

$$(6.17) \quad \sqrt{1 - \delta_{2s,M}^2} - \delta_{2s,M} \sqrt{l}/4 > 0.$$

Multiplying (6.16) by $\sqrt{\tilde{s}}$ yields

$$\begin{aligned} \sqrt{\tilde{s}}\|x_{S_0}\| &\leq \sqrt{\tilde{s}} \frac{\sqrt{1 + \delta_{2s,M}}}{\sqrt{1 - \delta_{2s,M}^2 - \delta_{2s,M}\sqrt{l}/4}} \|Ux\|_2 + \frac{\delta_{2s,M}\sqrt{\tilde{s}}}{\min \sqrt{s_i} \left(\sqrt{1 - \delta_{2s,M}^2 - \delta_{2s,M}\sqrt{l}/4}\right)} \|x_{S_0^c}\|_1 \\ &\leq \tau\sqrt{\tilde{s}} \|Ux\|_2 + \frac{\delta_{2s,M}}{\sqrt{1 - \delta_{2s,M}^2 - \delta_{2s,M}\sqrt{l}/4}} \sqrt{\sum_{k=1}^l \frac{s_k}{\min s_i}} \|x_{S_0^c}\|_1 \\ &\leq \tau\sqrt{\tilde{s}} \|Ux\|_2 + \frac{\delta_{2s,M}\sqrt{l\eta_{s,M}}}{\sqrt{1 - \delta_{2s,M}^2 - \delta_{2s,M}\sqrt{l}/4}} \|x_{S_0^c}\|_1, \end{aligned}$$

where $\tau = \frac{\sqrt{1+\delta_{2s,M}}}{\sqrt{1-\delta_{2s,M}^2-\delta_{2s,M}\sqrt{l}/4}}$. It is clear that (6.8) is satisfied if condition (6.17) holds and

$$(6.18) \quad \frac{\delta_{2s,M}\sqrt{l\eta_{s,M}}}{\sqrt{1 - \delta_{2s,M}^2 - \delta_{2s,M}\sqrt{l}/4}} < 1 \quad \text{or equivalently} \quad \delta_{2s,M} < \frac{1}{\sqrt{l(\sqrt{\eta_{s,M}} + \frac{1}{4})^2 + 1}}$$

while (6.17) is equivalent to $\delta_{2s,M} < \frac{1}{\sqrt{\frac{l}{16} + 1}}$. Since

$$\frac{1}{\sqrt{l(\sqrt{\eta_{s,M}} + \frac{1}{4})^2 + 1}} \leq \frac{1}{\sqrt{\frac{l}{16} + 1}}$$

it will suffice for (6.18) to hold, completing the proof. ■

6.3. Proofs of Theorems 4.5 and 4.6.

Proof of Theorem 4.5. The ideas behind the counterexample in this proof are similar to those in Theorem 3.2 in [13]. We prove this theorem in three stages. First we shall construct the matrix U . Next we shall show that our construction does indeed have a RIP in levels constant satisfying (4.4). Finally, we shall explain why z^1 exists.

Step I. Set $n = C + C^2$, where the nonnegative integer C is much greater than a (we shall give a precise choice of C later). Let $x^1 \in \mathbb{C}^n$ be the vector

$$x^1 := \lambda(\underbrace{C, C, \dots, C}_C, \underbrace{1, 1, \dots, 1}_{C^2}).$$

With this definition, the first C elements of x^1 have value $C\lambda$ and the next C^2 elements have value λ . Our (s, M) sparsity pattern is given by $s = (1, C^2)$ and $M = (0, C, C + C^2)$. Clearly, by the definition of the ratio constant, $\eta_{s,M} = C^2$ (in particular, $\eta_{s,M}$ is finite). Choose $\lambda = \frac{1}{\sqrt{C^3 + C^2}}$ so that $\|x^1\|_2 = 1$. By using this fact, we can form an orthonormal basis of \mathbb{C}^{C+C^2} that includes x^1 . We can write this basis as $(x^i)_{i=1}^{C+C^2}$. Finally, for a vector $v \in \mathbb{C}^{C+C^2}$, we define the linear map U by

$$U'v := \sum_{i=2}^{C+C^2} v^i x^i, \quad \text{where} \quad v = \sum_{i=1}^{C+C^2} v^i x^i.$$

In particular, notice that the nullspace of U is precisely the space spanned by x^1 and that $v^i = \langle v, x^i \rangle$.

Step II. Let γ be an $(\mathbf{as}, \mathbf{M})$ sparse vector. Our aim will be to estimate $|\|U\gamma\|_2^2 - \|\gamma\|_2^2|$. Clearly, $\|U\gamma\|_2^2 - \|\gamma\|_2^2 = -|\gamma^1|^2$, where γ^1 is the coefficient of x^1 in the expansion of γ in the basis (x^i) . Therefore, to show that U satisfies the RIP_L we will only need to bound $|\gamma^1| = |\langle \gamma, x^1 \rangle|$. Let S be the support of γ . Then

$$|\langle \gamma, x^1 \rangle| = |\langle \gamma_S, x^1 \rangle| = |\langle \gamma, x_S^1 \rangle| \leq \|\gamma\|_2 \|x_S^1\|_2 \leq \lambda \|\gamma\|_2 \sqrt{aC^2 + C^2},$$

where we have used Cauchy–Schwarz in the first inequality and in the second inequality we have used the fact that x_S^1 has at most a elements of size λC and at most C^2 elements of size λ . From the definition of λ we get $|\langle \gamma, x^1 \rangle| \leq \sqrt{\frac{a+1}{C+1}} \|\gamma\|_2$. Therefore,

$$|\|U\gamma\|_2^2 - \|\gamma\|_2^2| = |\langle \gamma, x^1 \rangle|^2 \leq \frac{a+1}{C+1} \|\gamma\|_2^2.$$

By the assumption that $f(x) = o(x^{\frac{1}{2}})$, we can find a $C \in \mathbb{N}$ sufficiently large so that $\frac{a+1}{C+1} \leq \frac{1}{|f(C^2)|}$. Then $\delta_{\mathbf{as}, \mathbf{M}} \leq \frac{1}{|f(\eta_{\mathbf{s}, \mathbf{M}})|}$ as claimed.

Step III. Let

$$z^1 := (C, \underbrace{0, 0, \dots, 0}_{C-1}, \underbrace{1, 1, \dots, 1}_{C^2}), \quad z^2 := (0, \underbrace{C, C, \dots, C}_{C-1}, \underbrace{C, C, 0, 0, \dots, 0}_{C^2}).$$

It is clear that z^1 is (\mathbf{s}, \mathbf{M}) -sparse. Additionally, $\|z^1\|_1 = C^2 + C$ and $\| -z^2 \|_1 = (C-1)C = C^2 - C$. Because $U(z^1 + z^2) = U(x^1)/\lambda = 0$, we have $U(-z^2) = U(z^1)$. Since the kernel of U is of dimension 1, the only vectors z which satisfy $U(z) = U(z^1)$ are $z = z^1$ and $z = -z^2$. Moreover, $\|z^1\|_1 > \| -z^2 \|_1$. Consequently

$$z^1 \notin \arg \min \|z\|_1 \text{ such that } Uz = Uz^1. \quad \blacksquare$$

Proof of Theorem 4.6. The proof of this theorem is almost identical to that of Theorem 4.5, so we shall omit details here. Again, we set x^1 so that

$$x^1 := \lambda (\underbrace{C, C, \dots, C}_C, \underbrace{1, 1, \dots, 1}_{C^2}),$$

where $C \gg a$. We choose λ so that $\|x^1\|_2 = 1$. In contrast to the proof of Theorem 4.5, we take

$$\mathbf{s} = (\underbrace{1, 1, 1, \dots, 1}_{C^2+1}), \quad \mathbf{M} = (0, C, C+1, \dots, C+C^2-1, C+C^2).$$

This time, there are $C^2 + 1$ levels and the ratio constant $\eta_{\mathbf{s}, \mathbf{M}}$ is equal to 1. Once again, we produce an orthonormal basis of \mathbb{C}^{C+C^2} that includes x^1 , which we label $(x^i)_{i=1}^{C+C^2}$, and we define the linear map U by

$$U^i v := \sum_{i=2}^{C+C^2} v^i x^i, \quad \text{where } v = \sum_{i=1}^{C+C^2} v^i x^i.$$

The same argument as before proves that for any (as, \mathbf{M}) -sparse γ ,

$$| \|U'\gamma\|_2^2 - \|\gamma\|_2^2 | \leq \frac{a+1}{C+1} \|\gamma\|_2^2,$$

and again, taking C sufficiently large so that $\frac{a+1}{C+1} \leq \frac{1}{|f(C^2+1)|}$ yields $\delta_{as, \mathbf{M}} \leq \frac{1}{|f(l)|}$. The proof of the existence of z^1 is identical to Step III in the proof of Theorem 4.5. ■

6.4. Proof of Theorem 4.7.

Proof. Once again, we prove this theorem in three stages. First we shall construct the matrix U' . Next, we shall show that the matrix U' has a sufficiently small RIP_L constant. Finally, we shall explain why both z^1 and z exist.

Step I. Let x^1 be the vector

$$x^1 := \lambda(\underbrace{0, 0, \dots, 0}_{C^2}, \underbrace{1, 1, \dots, 1}_{\omega(\rho, C)+1}),$$

where $\omega(\rho, C) = \text{ceil}(\frac{2C}{\rho})$ for a fixed $\rho \in (0, 1)$ which we will specify later, $\text{ceil}(a)$ denotes the smallest integer greater than or equal to a , and C is an integer greater than 1. In other words, the first C^2 elements of x^1 have value 0 and the next $\omega(\rho, C) + 1$ elements have value λ . We choose λ so that $\|x^1\|_2 = 1$ and C so that $C^2 > \omega(\rho, C)$ and choose our (\mathbf{s}, \mathbf{M}) sparsity pattern so that $\mathbf{s} = (C^2, 1)$ and $\mathbf{M} = (0, C^2, C^2 + \omega(\rho, C) + 1)$. By the definition of the ratio constant, $\eta_{\mathbf{s}, \mathbf{M}} = C^2$ (in particular, $\eta_{\mathbf{s}, \mathbf{M}}$ is finite). Because $\|x^1\|_2 = 1$, we can form an orthonormal basis of $\mathbb{C}^{C^2 + \omega(\rho, C) + 1}$ that includes x^1 , which we can write as $(x^i)_{i=1}^{C^2 + \omega(\rho, C) + 1}$. Finally, for a vector $v \in \mathbb{C}^{C^2 + \omega(\rho, C) + 1}$, we define the linear map U' by

$$U'v := \frac{\sqrt{2}w}{\tau}, \text{ where } w = \sum_{i=2}^{C^2 + \omega(\rho, C) + 1} v^i x^i \text{ and } v = v^1 x^1 + w.$$

In particular, notice that the nullspace of U' is precisely the space spanned by x^1 and that $v^i = \langle v, x^i \rangle$.

Step II. Let γ be an (as, \mathbf{M}) sparse vector. For the purposes of proving Theorem 4.7, it will suffice to take $\tau = \sqrt{2}$. Then

$$\|U'\gamma\|_2^2 - \|\gamma\|_2^2 = -|\gamma^1|^2,$$

where γ^1 is the coefficient corresponding to x^1 in the expansion of γ in the basis (x^i) . As in the proof of Theorem 4.5, $|\gamma^1| = |\langle \gamma, x^1 \rangle|$. Let S be the support of γ . Then

$$|\langle \gamma_S, x^1 \rangle| = |\langle \gamma, x_S^1 \rangle| \leq \|\gamma\|_2 \|x_S^1\|_2 \leq \lambda \|\gamma\|_2 \sqrt{a},$$

where we have used Cauchy–Scharwz in the first inequality and in the second inequality we have used the fact that x_S^1 has at most a elements of size λ . It is easy to see that $\lambda = \frac{1}{\sqrt{\omega(\rho, C)+1}}$.

Therefore,

$$| \|U'\gamma\|_2^2 - \|\gamma\|_2^2 | = |\langle \gamma, x^1 \rangle|^2 \leq \frac{a}{\omega(\rho, C) + 1} \|\gamma\|_2^2 \leq \frac{\rho a}{2C}$$

because $\omega(\rho, C) \geq \frac{2C}{\rho}$. By the assumption that $g(\eta_{\mathbf{s}, \mathbf{M}}) \leq \frac{1}{A} \sqrt{\eta_{\mathbf{s}, \mathbf{M}}}$ for some $A > 0$ and $\eta_{\mathbf{s}, \mathbf{M}}$ sufficiently large, and the fact that $\eta_{\mathbf{s}, \mathbf{M}} = C^2$, we must have $\frac{A}{C} \leq \frac{1}{g(\eta_{\mathbf{s}, \mathbf{M}})}$. If we take ρ sufficiently small and C sufficiently large, then

$$\delta_{as, \mathbf{M}} < \frac{\rho a}{2C} \leq \frac{A}{C} \leq \frac{1}{g(\eta_{\mathbf{s}, \mathbf{M}})},$$

as claimed.

Step III. Let $z^1 := x^1$ and set z to be the 0 vector in $\mathbb{C}^{C^2 + \omega(\rho, C) + 1}$. Because x^1 is in the kernel of U' , $U'(z - z^1) = 0$. Furthermore, it is obvious that $\|z\|_1 \leq \|z^1\|_1$. Additionally, $\|z^1\|_2 = 1$ and

$$\frac{\sigma_{\mathbf{s}, \mathbf{M}}(z^1)_1}{\sqrt{\tilde{s}}} = \lambda \frac{\omega(\rho, C)}{\sqrt{C^2 + 1}} \leq \frac{\omega(\rho, C)}{\sqrt{\omega(\rho, C)(C^2 + 1)}} \leq \sqrt{\frac{2C + 1}{\rho(C^2 + 1)}} \leq \sqrt{\frac{3}{\rho \sqrt{\eta_{\mathbf{s}, \mathbf{M}}}}}$$

since $\eta_{\mathbf{s}, \mathbf{M}} = C^2$ and $\omega(\rho, C) \leq 2C/\rho + 1 \leq (2C + 1)/\rho$. Because $f(\eta_{\mathbf{s}, \mathbf{M}}) = o(\eta_{\mathbf{s}, \mathbf{M}}^{\frac{1}{4}})$,

$$\frac{\sigma_{\mathbf{s}, \mathbf{M}}(z^1)_1}{\sqrt{\tilde{s}}} f(\eta_{\mathbf{s}, \mathbf{M}}) \rightarrow 0, \quad \eta_{\mathbf{s}, \mathbf{M}} \rightarrow \infty.$$

The desired result follows by taking $\eta_{\mathbf{s}, \mathbf{M}}$ sufficiently large so that

$$\|z - z^1\|_2 = 1 > \frac{\sigma_{\mathbf{s}, \mathbf{M}}(z^1)_1}{\sqrt{\tilde{s}}} f(\eta_{\mathbf{s}, \mathbf{M}}). \quad \blacksquare$$

Proof of part 2. The proof of part 2 follows with a few alterations to the previous case. We now use the sparsity pattern

$$\mathbf{s} = (\underbrace{1, 1, 1, \dots, 1}_{C^2}, 1) \quad \text{and} \quad \mathbf{M} = (0, 1, 2, \dots, C^2, C^2 + \omega(\rho, C) + 1).$$

In this case, $\eta_{\mathbf{s}, \mathbf{M}} = 1$ and $l = C^2 + 1$. The result follows by simply employing the same matrix U' with this new sparsity pattern. \blacksquare

6.5. Proofs of Theorem 4.9. The counterexample for Theorem 4.9 is the same as the one used in the proof of Theorem 4.7. In that case, the matrix depended on three parameters: C, τ , and ρ . We show that U' satisfies the ℓ^2 robust nullspace property of order (\mathbf{s}, \mathbf{M}) with parameters ρ and τ . The existence of z^1 and z is identical to Step III in the proof of Theorem 4.7.

Proof of part 1. First, if $T \subset S$ then for any $v \in \mathbb{C}^{C^2 + \omega(\rho, C) + 1}$, we have

$$\|v_T\|_2 \leq \|v_S\|_2 \quad \text{and} \quad \frac{\rho}{\sqrt{\tilde{s}}} \|v_{T^c}\|_1 + \tau \|U'v\|_2 \geq \frac{\rho}{\sqrt{\tilde{s}}} \|v_{S^c}\|_1 + \tau \|U'v\|_2$$

so it will suffice to prove that U' satisfies (3.2) for (\mathbf{s}, \mathbf{M}) -sparse sets S with $|S| = \tilde{s}$.

As before, we set $U'v := \sqrt{2}w/\tau$, where w is defined as in the proof of Theorem 4.7. Let us consider a set S such that $|S| = \tilde{s}$. Because w_S and w_{S^c} have disjoint support,

by the Cauchy–Schwarz inequality applied to the vectors $(1, 1)$ and $(\|w_S\|_2, \|w_{S^c}\|_2)$ we get $\sqrt{2}\|w\|_2 \geq \|w_S\|_2 + \|w_{S^c}\|_2$. Therefore,

$$(6.19) \quad \tau\|U'v\|_2 \geq \sqrt{2}\|w\|_2 \geq \|w_S\|_2 + \|w_{S^c}\|_2.$$

Furthermore, because $|S| = \tilde{s} \geq |S^c|$ (recall that $|S| = C^2 + 1$ and that C was chosen so that $C^2 > \omega(\rho, C) = |S^c|$) and $\rho \in (0, 1)$

$$(6.20) \quad \|w_{S^c}\|_2 \geq \frac{1}{\sqrt{|S^c|}}\|w_{S^c}\|_1 \geq \frac{1}{\sqrt{\tilde{s}}}\|w_{S^c}\|_1 \geq \frac{\rho}{\sqrt{\tilde{s}}}\|w_{S^c}\|_1.$$

Combining (6.19) and (6.20) gives

$$(6.21) \quad \begin{aligned} \tau\|U'v\|_2 + \frac{\rho}{\sqrt{\tilde{s}}}\|v_{S^c}\|_1 &\geq \|w_S\|_2 + \frac{\rho}{\sqrt{\tilde{s}}}\|w_{S^c}\|_1 + \frac{\rho}{\sqrt{\tilde{s}}}\|v_{S^c}\|_1 \geq \|w_S\|_2 + \frac{\rho}{\sqrt{\tilde{s}}}\|v_{S^c} - w_{S^c}\|_1 \\ &\geq \|w_S\|_2 + \frac{\rho}{\sqrt{\tilde{s}}}\|v^1 x_{S^c}^1\|_1. \end{aligned}$$

We shall now aim to bound $\|v^1 x_S^1\|_2$ in terms of $\|v^1 x_{S^c}^1\|_1$. We have

$$(6.22) \quad \|v^1 x_S^1\|_2 \leq \lambda|v^1|$$

since at most one element of x_S^1 is nonzero and its value will be at most λ . Additionally, since each element of $x_{S^c}^1$ has value λ and there are at least $\frac{2C}{\rho}$ of them

$$\rho\|v^1 x_{S^c}^1\|_1 = \rho|v^1|\|x_{S^c}^1\|_1 \geq \frac{2\lambda C}{\rho}\rho|v^1| \geq 2\lambda C|v^1|.$$

Therefore,

$$(6.23) \quad \frac{\rho}{\sqrt{\tilde{s}}}\|v^1 x_{S^c}^1\|_1 \geq \frac{2\lambda C}{\sqrt{C^2 + 1}}|v^1| \geq \lambda|v^1|.$$

Using (6.22) and (6.23), we have $\|v^1 x_S^1\|_2 \leq \frac{\rho}{\sqrt{\tilde{s}}}\|v^1 x_{S^c}^1\|_1$. We can conclude the proof that U' satisfies the ℓ^2 robust nullspace property by combining this result with (6.21) as follows:

$$\|v_S\|_2 \leq \|v^1 x_S^1\|_2 + \|w_S\|_2 \leq \frac{\rho}{\sqrt{\tilde{s}}}\|v^1 x_{S^c}^1\|_1 + \|w_S\|_2 \leq \tau\|U'v\|_2 + \frac{\rho}{\sqrt{\tilde{s}}}\|v_{S^c}\|_1. \quad \blacksquare$$

Proof of part 2. The proof of part 2 is identical. We simply adapt the sparsity pattern so that

$$\mathbf{s} = (\underbrace{1, 1, 1, \dots, 1}_{C^2}, 1) \text{ and } \mathbf{M} = (0, 1, 2, \dots, C^2, C^2 + \omega(\rho, C) + 1).$$

We can apply the proceeding argument with this new sparsity pattern to obtain the required result. \blacksquare

Acknowledgments. The authors would like to thank Arash Amini for asking an excellent question, during a talk given at EPFL, about the possibility of having a theory for the RIP in levels. It was this question that sparked the research leading to this paper. The authors would also like to thank Ben Adcock for many great suggestions and input. Finally, the authors would like to thank Bogdan Roman for his valuable discussions concerning the implementation of the examples using the hadamard transformation. All the numerical computations were done with the SPGL1 package [65, 66].

REFERENCES

- [1] B. ADCKOCK, A. BASTOUNIS, A. C. HANSEN, AND B. ROMAN, *On Fundamentals of Models and Sampling in Compressed Sensing*, preprint, 2015.
- [2] B. ADCKOCK, A. HANSEN, B. ROMAN, AND G. TESCHKE, *Chapter four—generalized sampling: Stable reconstructions, inverse problems and compressed sensing over the continuum*, in *Adv. Imaging Electron Phys.*, Elsevier, Amsterdam, 2014, pp. 187–279, <https://dx.doi.org/10.1016/B978-0-12-800146-2.00004-7>.
- [3] B. ADCKOCK AND A. C. HANSEN, *Generalized sampling and infinite-dimensional compressed sensing*, *Found. Comput. Math.*, 16 (2016), pp. 1263–1323.
- [4] B. ADCKOCK, A. C. HANSEN, C. POON, AND B. ROMAN, *Breaking the Coherence Barrier: A New Theory for Compressed Sensing*, [arXiv:1302.0561](https://arxiv.org/abs/1302.0561), 2014.
- [5] B. ADCKOCK, A. C. HANSEN, AND B. ROMAN, *A note on compressed sensing of structured sparse wavelet coefficients from subsampled fourier measurements*, *IEEE Signal Process. Lett.*, 23 (2016), pp. 732–736, <https://doi.org/10.1109/LSP.2016.2550101>.
- [6] J. ANDERSSON AND J. STRÖMBERG, *On the theorem of uniform recovery of random sampling matrices*, *IEEE Trans. Inform. Theory*, 60 (2014), pp. 1700–1710, <https://doi.org/10.1109/TIT.2014.2300092>.
- [7] R. G. BARANIUK, V. CEVHER, M. F. DUARTE, AND C. HEDGE, *Model-based compressive sensing*, *IEEE Trans. Inform. Theory*, 56 (2010), pp. 1982–2001.
- [8] J. BIGOT, C. BOYER, AND P. WEISS, *An Analysis of Block Sampling Strategies in Compressed Sensing*, [arXiv:1305.4446](https://arxiv.org/abs/1305.4446), 2013.
- [9] H. BOCHE, R. CALDERBANK, G. KUTYNIOK, AND J. VYBÍRAL, *Compressed Sensing and Its Applications*, Springer, New York, 2015.
- [10] A. BOURRIER, M. DAVIES, T. PELEG, P. PEREZ, AND R. GRIBONVAL, *Fundamental performance limits for ideal decoders in high-dimensional linear inverse problems*, *IEEE Trans. Inform. Theory*, 60 (2014), pp. 7928–7946, <https://doi.org/10.1109/TIT.2014.2364403>.
- [11] T. T. CAI, L. WANG, AND G. XU, *New bounds for restricted isometry constants*, *IEEE Trans. Inform. Theory*, 56 (2010), pp. 4388–4394, <https://doi.org/10.1109/TIT.2010.2054730>.
- [12] T. T. CAI, L. WANG, AND G. XU, *Shifting inequality and recovery of sparse signals*, *IEEE Trans. Signal Process.*, 58 (2010), pp. 1300–1308, <https://doi.org/10.1109/TSP.2009.2034936>.
- [13] T. T. CAI AND A. ZHANG, *Sharp RIP bound for sparse signal and low-rank matrix recovery*, *Appl. Comput. Harmon. Anal.*, 35 (2013), pp. 74–93, <https://doi.org/10.1016/j.acha.2012.07.010>.
- [14] T. T. CAI AND A. ZHANG, *Sparse representation of a polytope and recovery of sparse signals and low-rank matrices*, *IEEE Trans. Inform. Theory*, 60 (2014), pp. 122–132.
- [15] E. J. CANDÈS, *An introduction to compressive sensing*, *IEEE Signal Process. Mag.*, 25 (2008), pp. 21–30.
- [16] E. J. CANDÈS, *The restricted isometry property and its implications for compressed sensing*, *C. R. Math. Acad. Sci. Paris*, 346 (2008), pp. 589–592, <https://doi.org/10.1016/j.crma.2008.03.014>.
- [17] 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 (2004), pp. 219–266.
- [18] E. J. CANDÈS, J. ROMBERG, AND T. TAO, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, *IEEE Trans. Inform. Theory*, 52 (2006), pp. 489–509.
- [19] E. J. CANDÈS AND T. TAO, *Decoding by linear programming*, *IEEE Trans. Inform. Theory*, 51 (2005), pp. 4203–4215, <https://doi.org/10.1109/TIT.2005.858979>.

- [20] E. J. CANDÈS AND T. TAO, *Near-optimal signal recovery from random projections: Universal encoding strategies?*, IEEE Trans. Inform. Theory, 52 (2006), pp. 5406–5425, <https://doi.org/10.1109/TIT.2006.885507>.
- [21] W. R. CARSON, M. CHEN, M. R. D. RODRIGUES, R. CALDERBANK, AND L. CARIN, *Communications-inspired projection design with application to compressive sensing*, SIAM J. Imaging Sci., 5 (2012), pp. 1185–1212.
- [22] N. CHAUFFERT, P. CIUCIU, J. KAHN, AND P. WEISS, *Variable density sampling with continuous sampling trajectories*, SIAM J. Imaging Sci., 7 (2014), pp. 1962–1992.
- [23] K. CHOI, J. WANG, L. ZHU, T.-S. SUH, S. BOYD, AND L. XING, *Compressed sensing based cone-beam computed tomography reconstruction with a first-order method*, Medical Phys., 37 (2010), pp. 5113–5125.
- [24] A. COHEN, W. DAHMEN, AND R. DEVORE, *Compressed sensing and best k -term approximation*, J. Amer. Math. Soc., 22 (2009), pp. 211–231.
- [25] I. DAUBECHIES, *Orthonormal bases of compactly supported wavelets*, Comm. Pure Appl. Math., 41 (1988), pp. 909–996, <https://doi.org/10.1002/cpa.3160410705>.
- [26] I. DAUBECHIES, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conf. Ser. in Appl. Math., SIAM, Philadelphia, 1992, <https://doi.org/10.1137/1.9781611970104>.
- [27] M. A. DAVENPORT, M. F. DUARTE, Y. C. ELДАР, AND G. KUTYNIOK, *Introduction to compressed sensing*, in Compressed Sensing: Theory and Applications, Cambridge University Press, Cambridge, UK, 2011.
- [28] M. E. DAVIES AND R. GRIBONVAL, *Restricted isometry constants where ℓ^p sparse recovery can fail for $0 \ll p \leq 1$* , IEEE Trans. Inform. Theory, 55 (2009), pp. 2203–2214, <https://doi.org/10.1109/TIT.2009.2016030>.
- [29] M. N. DO AND M. VETTERLI, *The contourlet transform: An efficient directional multiresolution image representation*, IEEE Trans. Image Process., 14 (2005), pp. 2091–2106, <https://doi.org/10.1109/TIP.2005.859376>.
- [30] D. L. DONOHO, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.
- [31] D. L. DONOHO AND J. TANNER, *Counting faces of randomly-projected polytopes when the projection radically lowers dimension*, J. Amer. Math. Soc., 22 (2009), pp. 1–53.
- [32] Y. C. ELДАР AND G. KUTYNIOK, EDS., *Compressed Sensing: Theory and Applications*, Cambridge University Press, Cambridge, UK, 2012.
- [33] M. FORNASIER AND H. RAUHUT, *Compressive sensing*, in Handbook of Mathematical Methods in Imaging, Springer, New York, 2011, pp. 187–228.
- [34] S. FOUCAIT, *A note on guaranteed sparse recovery via ℓ_1 -minimization*, Appl. Comput. Harmon. Anal., 29 (2010), pp. 97–103, <https://doi.org/10.1016/j.acha.2009.10.004>.
- [35] S. FOUCAIT AND H. RAUHUT, *A Mathematical Introduction to Compressive Sensing*, Applied and Numerical Harmonic Analysis, Springer, New York, 2013, <https://doi.org/10.1007/978-0-8176-4948-7>.
- [36] M. D. GUAY, W. CZAJA, M. A. ARONOVA, AND R. D. LEAPMAN, *Compressed sensing electron tomography for determining biological structure*, Scientific Reports, 6 (2016), 27614, <http://dx.doi.org/10.1038/srep27614>.
- [37] 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 (2011), pp. 1649–1660.
- [38] M. GUERQUIN-KERN, L. LEJEUNE, K. P. PRUESSMANN, AND M. UNSER, *Realistic analytical phantoms for parallel Magnetic Resonance Imaging*, IEEE Trans. Med. Imaging, 31 (2012), pp. 626–636.
- [39] S. HASHEMI, S. BEHESHTI, P. R. GILL, N. S. PAUL, AND R. S. C. COBBOLD, *Accelerated compressed sensing based CT image reconstruction*, Comput. Math. Methods Medicine, 2015 (2015), 161797, <https://doi.org/10.1155/2015/161797>.
- [40] L. HE AND L. CARIN, *Exploiting structure in wavelet-based Bayesian compressive sensing*, IEEE Trans. Signal Process., 57 (2009), pp. 3488–3497.
- [41] L. HE, H. CHEN, AND L. CARIN, *Tree-structured compressive sensing with variational Bayesian analysis*, IEEE Signal Process. Letters, 17 (2010), pp. 233–236.
- [42] M. A. HERMAN AND T. STROHMER, *High-resolution radar via compressed sensing*, IEEE Trans. Signal Process., 57 (2009), pp. 2275–2284.

- [43] D. J. HOLLAND, M. J. BOSTOCK, L. F. GLADDEN, AND D. NIETLISPACH, *Fast multidimensional NMR spectroscopy using compressed sensing*, *Angew. Chemie*, 50 (2011), pp. 6548–6551, <https://doi.org/10.1002/anie.201100440>.
- [44] A. JONES, A. TAMTÖGL, I. CALVO-ALMAZÁN, AND A. HANSEN, *Continuous compressed sensing for surface dynamical processes with helium atom scattering*, *Scientific Reports*, 6 (2016), 27776, <http://dx.doi.org/10.1038/srep27776>.
- [45] A. D. JONES, B. ADCOCK, AND A. C. HANSEN, *On asymptotic incoherence and its implications for compressed sensing of inverse problems*, *IEEE Trans. Inform. Theory*, 62 (2016), pp. 1020–1037, <https://doi.org/10.1109/TIT.2015.2508562>.
- [46] M. A. KHAJEHNEJAD, W. XU, A. S. AVESTIMEHR, AND B. HASSIBI, *Weighted l_1 minimization for sparse recovery with prior information*, in *Proceedings of the IEEE International Conference on Symposium on Information Theory*, Vol. 1, Piscataway, NJ, 2009, pp. 483–487, <http://dl.acm.org/citation.cfm?id=1701495.1701594>.
- [47] F. KRAHMER AND R. WARD, *Stable and robust sampling strategies for compressive imaging*, *IEEE Trans. Image Process.*, 23 (2014), pp. 612–622, <http://num.math.uni-goettingen.de/~f.krahmer/KW13.pdf>.
- [48] G. KUTYNIOK, J. LEMVIG, AND W.-Q. LIM, *Compactly supported shearlets*, in *Approximation Theory XIII: San Antonio 2010*, M. Neamtu and L. Schumaker, eds., *Springer Proc. Math.* 13, Springer, New York, 2012, pp. 163–186, https://doi.org/10.1007/978-1-4614-0772-0_10.
- [49] R. LEARY, Z. SAGHI, P. A. MIDGLEY, AND D. J. HOLLAND, *Compressed sensing electron tomography*, *Ultramicroscopy*, 131 (2013), pp. 70–91, <https://doi.org/10.1016/j.ultramic.2013.03.019>.
- [50] C. LI AND B. ADCOCK, *Compressed Sensing with Local Structure: Uniform Recovery Guarantees for the Sparsity in Levels Class*, [arXiv:1601.01988](https://arxiv.org/abs/1601.01988), 2016.
- [51] A. K. LOUIS, P. MAASS, AND A. RIEDER, *Wavelets: Theorie und Anwendungen*, Teubner Stud. Math., B. G. Teubner, Stuttgart, 1998, <https://doi.org/10.1007/978-3-322-80136-4>.
- [52] M. LUSTIG, D. L. DONOHO, AND J. M. PAULY, *Sparse MRI: The application of compressed sensing for rapid MRI imaging*, *Magn. Reson. Imaging*, 58 (2007), pp. 1182–1195.
- [53] J. D. MCEWEN AND Y. WIAUX, *Compressed sensing for radio interferometric imaging: Review and future direction*, in *Proceedings of the 18th IEEE International Conference on Image Processing*, 2011, pp. 1313–1316, <https://doi.org/10.1109/ICIP.2011.6115677>.
- [54] G. PUY, M. E. DAVIES, AND R. GRIBONVAL, *Linear embeddings of low-dimensional subsets of a Hilbert space to R^m* , in *Proceedings of EUSIPCO—23rd European Signal Processing Conference*, 2015.
- [55] G. PUY, J. P. MARQUES, R. GRUETTER, J. P. THIRAN, D. V. D. VILLE, P. VANDERGHEYNST, AND Y. WIAUX, *Spread spectrum magnetic resonance imaging*, *IEEE Trans. Medical Imaging*, 31 (2012), pp. 586–598, <https://doi.org/10.1109/TMI.2011.2173698>.
- [56] G. PUY, P. VANDERGHEYNST, R. GRIBONVAL, AND Y. WIAUX, *Universal and efficient compressed sensing by spread spectrum and application to realistic fourier imaging techniques*, *EURASIP J. Adv. Signal Process.*, 2012 (2012), pp. 1–13.
- [57] G. PUY, P. VANDERGHEYNST, AND Y. WIAUX, *On variable density compressive sampling*, *IEEE Signal Process. Lett.*, 18 (2011), pp. 595–598.
- [58] H. RAUHUT AND R. WARD, *Interpolation via weighted minimization*, *Appl. Comput. Harmon. Anal.*, 40 (2016), pp. 321–351, <https://doi.org/10.1016/j.acha.2015.02.003>.
- [59] B. ROMAN, B. ADCOCK, AND A. C. HANSEN, *On Asymptotic Structure in Compressed Sensing*, [arXiv:1406.4178](https://arxiv.org/abs/1406.4178), 2015.
- [60] J. ROMBERG, *Imaging via compressive sampling*, *IEEE Signal Process. Mag.*, 25 (2008), pp. 14–20.
- [61] V. STUDER, J. BOBIN, M. CHAHID, H. MOUSSAVI, E. CANDÈS, AND M. DAHAN, *Compressive fluorescence microscopy for biological and hyperspectral imaging*, *Proc. Natl. Acad. Sci. USA*, 109 (2011), pp. 1679–1687.
- [62] A. M. TILLMANN AND M. E. PFETSCH, *The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing*, *IEEE Trans. Inform. Theory*, 60 (2014), pp. 1248–1259, <https://doi.org/10.1109/TIT.2013.2290112>.
- [63] Q. TRAN-DINH AND V. CEVHER, *A Primal-Dual Algorithmic Framework for Constrained Convex minimization*, [arXiv:1406.5403v2](https://arxiv.org/abs/1406.5403v2), 2014.
- [64] Y. TRAONMILIN AND R. GRIBONVAL, *Stable Recovery of Low-Dimensional Cones in Hilbert Spaces: One RIP to Rule Them All*, [arXiv:1510.00504](https://arxiv.org/abs/1510.00504), 2015.

- [65] E. VAN DEN BERG AND M. P. FRIEDLANDER, *SPGL1: A Solver for Large-Scale Sparse Reconstruction*, <http://www.cs.ubc.ca/labs/scl/spgl1> (2007).
- [66] E. VAN DEN BERG AND M. P. FRIEDLANDER, *Probing the Pareto frontier for basis pursuit solutions*, *SIAM J. Sci. Comput.*, 31 (2008), pp. 890–912, <https://doi.org/10.1137/080714488>.
- [67] L. WANG, D. CARLSON, M. R. D. RODRIGUES, D. WILCOX, R. CALDERBANK, AND L. CARIN, *Designed measurements for vector count data*, in *Advances in Neural Information Processing Systems*, MIT Press, Cambridge, MA, 2013, pp. 1142–1150.
- [68] Q. WANG, M. ZENGE, H. E. CETINGUL, E. MUELLER, AND M. S. NADAR, *Novel sampling strategies for sparse MR image reconstruction*, presented at the Annual Meeting of the International Society for Magnetic Resonance in Medicine, Milan, 2014.