1-1-2015

Compressive Sensing with Redundant Dictionaries and Structured Measurements

Felix Krahmer
Technische Universitat Munchen

Deanna Needell
Claremont McKenna College

Rachel Ward
University of Texas at Austin

Recommended Citation

COMPRESSIVE SENSING WITH REDUNDANT DICTIONARIES AND STRUCTURED MEASUREMENTS

FELIX KRAHMER, DEANNA NEEDELL, AND RACHEL WARD

ABSTRACT. Consider the problem of recovering an unknown signal from undersampled measurements, given the knowledge that the signal has a sparse representation in a specified dictionary D. This problem is now understood to be well-posed and efficiently solvable under suitable assumptions on the measurements and dictionary, if the number of measurements scales roughly with the sparsity level. One sufficient condition for such is the D-restricted isometry property (D-RIP), which asks that the sampling matrix approximately preserve the norm of all signals which are sufficiently sparse in D. While many classes of random matrices are known to satisfy such conditions, such matrices are not representative of the structural constraints imposed by practical sensing systems. We close this gap in the theory by demonstrating that one can subsample a fixed orthogonal matrix in such a way that the D-RIP will hold, provided this basis is sufficiently incoherent with the sparsifying dictionary D. We also extend this analysis to allow for weighted sparse expansions. Consequently, we arrive at compressive sensing recovery guarantees for structured measurements and redundant dictionaries, opening the door to a wide array of practical applications.

1. INTRODUCTION

1.1. **Compressive Sensing.** The compressive sensing paradigm, as first introduced by Candès and Tao [CT05b] and Donoho [Don06], is based on using available degrees of freedom in a sensing mechanism to tune the measurement system so as to allow for efficient recovery of a particular type of signal or image of interest from as few measurements as possible. A model assumption that allows for such signal recovery is *sparsity*: the signal can be well-approximated by just a few elements of a given representation system.

Often, a near-optimal strategy is to choose the measurements completely at random, for example following a Gaussian distribution [RV08]. Typically, however, some additional structure is imposed by the application at hand, and randomness can only be infused in the remaining degrees of freedom. For example, magnetic resonance imaging (MRI) is known to be well modeled by inner products with Fourier basis vectors. This structure cannot be changed, and the only aspect that can be decided at random is which Fourier basis vectors to select.

An important difference between completely random measurement systems and structured random measurement systems is in the aspect of universality. Gaussian measurement systems, among many other systems with minimal imposed structure, are oblivious to the basis in which the underlying signal is sparse, and achieve equal reconstruction quality for all different orthonormal basis representations. For structured measurement systems, this is, in general, no longer the case. When the measurements are uniformly subsampled from an orthonormal basis such as the Fourier basis, one requires, for example, that the measurement basis and the sparsity basis are *incoherent*; the inner products between vectors from the two bases are small. If the two bases are not incoherent, more refined concepts of incoherence are required [KW14, AHPR13]. An important example is that of the Fourier measurement basis and a wavelet sparsity basis. Since both contain the constant vector, they are maximally coherent.

1.2. **Motivation.** All of the above mentioned works, however, exclusively cover the case of sparsity in *orthonormal* basis representations. On the other hand, there are a vast number of applications in which sparsity is expressed not in terms of a basis but in terms of a redundant, often highly overcomplete, dictionary. Specifically, if $f \in \mathbb{C}^n$ is the signal of interest to be recovered, then one expresses $f = Dx$,

where $\boldsymbol{D} \in \mathbb{C}^{n \times N}$ is an overcomplete dictionary and $\boldsymbol{x} \in \mathbb{C}^N$ is a sparse (or nearly sparse) coefficient vector. Redundancy is widespread in practice, either because no sparsifying orthonormal basis exists for the signal class of interest, or because the redundancy itself is useful and allows for a significantly larger, richer class of signals which can be sparsely represented in the resulting dictionary. For example, it has been well documented that overcompleteness is the key to a drastic reduction in artifacts and recovery error in the denoising framework [SED04, SFM07].

In the compressive sensing framework, results using overcomplete dictionaries are motivated by the broad array of tight (and often Parseval) frames appearing in practical applications. For example, if one assumes sparsity with respect to the Discrete Fourier Transform (DFT), this is implicitly assuming that the signal is well-represented by frequencies lying along the lattice of the DFT. To allow for more flexibility in this rigid assumption, one instead may employ the oversampled DFT frame, containing frequencies on a much finer grid, or even over intervals of varying widths. Gabor frames are used in imaging as well as radar and sonar applications, which are often highly redundant [Mal99]. Many of the most widely-used frames in imaging applications such as undecimated wavelet frames [Dut89, SED04], curvelets [CD04], shearlets [LLKW05, ELL08], framelets [CCS08], and many others, are overcomplete with highly correlated columns.

1.3. **Related Work.** Due to the abundance of relevant applications, a number of works have studied compressive sensing for overcomplete frames. The first work on this topic aimed to recover the coefficient vector $\boldsymbol{x}$ directly, and thus required strong incoherence assumptions on the dictionary $\boldsymbol{D}$ [RSV08]. More recently, it was noted that if one instead aims to recover $\boldsymbol{f}$ rather than $\boldsymbol{x}$, recovery guarantees can be obtained under weaker assumptions. Namely, one only needs that the measurement matrix $\boldsymbol{A}$ respects the norms of signals which are sparse in the dictionary $\boldsymbol{D}$. To quantify this, Candès et al. [CENR10] define the $\boldsymbol{D}$-restricted isometry property ($\boldsymbol{D}$-RIP in short, see Definition 2.1 below). For measurement matrices that have this property, a number of algorithms have been shown to guarantee recovery under certain assumptions. Optimization approaches such as ℓ_1-analysis [EMR07, CENR10, GNE$^+$14, NDEG13, LML12, RK13, Gir14] and greedy approaches [DNW12, GNE$^+$14, GN13, PE13, GN14] have been studied.

This paper establishes the $\boldsymbol{D}$-RIP for structured random measurements formed by subsampling orthonormal bases, allowing for these types of recovery results to be utilized in more realistic settings. To date, most random matrix constructions known to yield $\boldsymbol{D}$-RIP matrices with high probability are random matrices with a certain concentration property. As shown in [CENR10], such a property implies that for arbitrary dictionary $\boldsymbol{D}$, one obtains the $\boldsymbol{D}$-RIP with high probability. This can be interpreted as a dictionary version of the universality property discussed above. It has now been shown that several classes of random matrices satisfy this property as well as subsampled structured matrices, after applying random column signs [DG03, KW11]. The matrices in both of these cases are motivated by application scenarios, but typically in applications they appear without the randomized column signs. In many cases one is not able to apply column signs in practice; for example in cases where the measurements are fixed such as in MRI, one has no choice but to use unsigned Fourier samples and cannot pre-process the data to incorporate column signs. Without these signs however, such measurement ensembles will not work for arbitrary dictionaries in general. This is closely related to the underlying RIP matrix constructions not being universal. For example, it is clear that randomly subsampled Fourier measurements will fail for the oversampled Fourier dictionary (for reasons similar to the basis case). In this work, we address this issue, deriving recovery guarantees that take into account the dictionary. Similar to the basis case, our analysis will be coherence based. A similar approach has also been taken by Poon in [Poo15b] (completed after the first version of our paper), for an infinite dimensional version of the problem.

1.4. **Contribution.** Our main result shows that a wide class of orthogonal matrices having uniformly bounded entries can be subsampled to obtain a matrix that has the $\boldsymbol{D}$-RIP and hence yields recovery guarantees for sparse recovery in the setting of redundant dictionaries. As indicated above, our technical estimates below will imply such guarantees for various algorithms. As an example we focus on the method of ℓ_1-analysis, for which the first $\boldsymbol{D}$-RIP based guarantees were available [CENR10]. Our

technical estimates will also provide more general guarantees for weighted ℓ_1-analysis minimization (a weighted version of (P_1), see $(P_{1,w})$ in Section 3 for details) in case one has prior knowledge of the underlying sparsity support.

Recall that the method of ℓ_1-analysis consists of estimating a signal $\boldsymbol{f}$ from noisy measurements $\boldsymbol{y} = \boldsymbol{Af} + \boldsymbol{e}$ by solving the convex minimization problem

$$\boldsymbol{f}^\sharp = \operatorname*{argmin}_{\tilde{\boldsymbol{f}} \in \mathbb{C}^n} \|\boldsymbol{D}^* \tilde{\boldsymbol{f}}\|_1 \quad \text{such that} \quad \|\boldsymbol{A}\tilde{\boldsymbol{f}} - \boldsymbol{y}\|_2 \le \varepsilon, \tag{P_1}$$

where ε is the noise level, that is, $\|\boldsymbol{e}\|_2 \le \varepsilon$. The ℓ_1-analysis method (like alternative approaches) assumes that for the signal $\boldsymbol{f} = \boldsymbol{Dx}$, not only is the underlying (synthesis) coefficient sequence $\boldsymbol{x}$ sparse (typically unknown and hard to obtain), but also the analysis coefficients $\boldsymbol{D}^* \boldsymbol{f}$ are nearly sparse, i.e., dominated by a few large entries. We refer the reader to Theorem 2.2 below for the precise formulation of the resulting recovery guarantees (as derived in [CENR10]). The assumption has been observed empirically for many dictionaries used in practice such as the Gabor frame, undecimated wavelets, curvelets, etc. (see, e.g., [CENR10] for a detailed description of such frames) and is also key for a number of thresholding approaches to signal denoising [CD04, KL07, ELL08].

A related and slightly stronger signal model, in which the analysis vector $\boldsymbol{D}^* \boldsymbol{f}$ is sparse or nearly sparse, has been considered independently from coefficient sparsity (e.g., [NDEG13]), and is commonly called the co-sparsity model.

The results in this paper need a similar, but somewhat weaker assumption to hold for all signals corresponding to sparse synthesis coefficients $\boldsymbol{x}$. Namely, one needs to control the localization factor as we now introduce.

Definition 1.1. For a dictionary $\boldsymbol{D} \in \mathbb{C}^{n \times N}$ and a sparsity level s, we define the *localization factor* as

$$\eta_{s,\boldsymbol{D}} = \eta \stackrel{\text{def}}{=} \sup_{\|\boldsymbol{Dz}\|_2 = 1, \|\boldsymbol{z}\|_0 \le s} \frac{\|\boldsymbol{D}^* \boldsymbol{Dz}\|_1}{\sqrt{s}}. \tag{1.1}$$

In the following we will mainly consider dictionaries, which form Parseval frames, that is, $\boldsymbol{D}^*$ is an isometry. Then the term localization factor is appropriate because this quantity can be viewed as the factor by which sparsity is preserved under the gram matrix map $\boldsymbol{D}^* \boldsymbol{D}$, compared to the case where $\boldsymbol{D}$ is an orthonormal basis and $\boldsymbol{D}^* \boldsymbol{D} = \mathbf{I}_n$ and $\eta \equiv 1$ by Cauchy-Schwarz. For a general family of such frames parameterized by the redundancy N/n, η will increase with the redundancy; families of dictionaries where such growth is relatively slow will be of interest. For example, new results on the construction of unit-norm tight frames give a constructive method to generate *spectral tetris* frames [CFM$^+$11], whose gram matrices have at most $2\lceil N/n \rceil + 6$ non-zeros per row, guaranteeing that η is proportional only to the redundancy factor N/n. In fact, one can show that these are sparsest frames possible [CHKK11], suggesting that families of tight frames with localization factor depending linearly on the redundancy factor should be essentially optimal for ℓ_1-analysis reconstruction. We pose as a problem for subsequent work to obtain such estimates for dictionaries of practical interest, discussing a few examples in Section 3. First, to illustrate that our results go strictly beyond existing theory, we show that harmonic frames (frames constructed by removing the high frequency rows of the DFT – see (3.3)) with small redundancy indeed have a bounded localization factor. To our knowledge, this important case is not covered by existing theories. In addition, we also bound the localization factor for redundant Haar wavelet frames. We expect, however, that it will be difficult to efficiently compute the localization factor of an arbitrary dictionary. For bounded localization factors, we prove the following theorem.

Theorem 1.2. *Fix a sparsity level $s < N$. Let $\boldsymbol{D} \in \mathbb{C}^{n \times N}$ be a Parseval frame – i. e., $\boldsymbol{DD}^*$ is the identity – with columns $\{\boldsymbol{d}_1, \ldots, \boldsymbol{d}_N\}$, and let $\boldsymbol{B} = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\}$ be an orthonormal basis of $\mathbb{C}^n$ which is incoherent to $\boldsymbol{D}$ in the sense that*

$$\sup_{i \in [n]} \sup_{j \in [N]} |\langle \boldsymbol{b}_i, \boldsymbol{d}_j \rangle| \le K n^{-1/2} \tag{1.2}$$

for some constant $K \ge 1$.

Consider the (unweighted) localization factor $\eta = \eta_{s,D}$ as in Definition 1.1. Construct $\widetilde{B} \in \mathbb{C}^{m \times n}$ by sampling row vectors from B i.i.d. uniformly at random. Provided

$$m \geq CsK^2\eta^2 \log^3(s\eta^2) \log(N), \tag{1.3}$$

then with probability $1 - N^{-\log^3(2s)}$, $\sqrt{\frac{n}{m}}\widetilde{B}$ exhibits uniform recovery guarantees for ℓ_1-analysis. That is, for every signal f, the solution $f^\sharp$ of the minimization problem (P_1) with $y = \sqrt{\frac{n}{m}}\widetilde{B}f + e$ for noise e with $\|e\|_2 \leq \varepsilon$ satisfies

$$\|f^\sharp - f\|_2 \leq C_1\varepsilon + C_2 \frac{\|D^*f - (D^*f)_s\|_1}{\sqrt{s}}. \tag{1.4}$$

Here, $C, C_1,$ and C_2 are absolute constants independent of the dimensions and the signal f.

Above and throughout, $(u)_s$ denotes the best s-sparse approximation to a signal u, that is, $(u)_s \stackrel{\text{def}}{=} \arg\min_{\|z\|_0 \leq s}\|u - z\|_2$.

Remarks.

1. Our stability result (1.4) for ℓ_1 analysis guarantees recovery for the particular $f = Dz$ only up to the scaled ℓ_1 norm of the tail of D^*Dz. In fact, the quantity $\mathscr{E}_f^* = \frac{\|D^*f - (D^*f)_s\|_1}{\sqrt{s}}$, referred to as the *unrecoverable energy* of the signal f in D [CENR10], is closely related to the localization factor η_s:

$$\eta_s = \sup_{f=Dz: \|f\|_2=1, \|z\|_0 \leq s} \frac{\|D^*f\|_1}{\sqrt{s}} \leq \sup_{f=Dz: \|f\|_2=1, \|z\|_0 \leq s} \mathscr{E}_f^* + 1$$

2. As mentioned, the proof of this theorem will proceed via the D-RIP, so our analysis yields similar guarantees for other recovery methods as well.

3. It is interesting to remark on the role of incoherence in this setting. While prior work in compressed sensing has required the measurement matrix $\widetilde{B}$ itself be incoherent, we now ask instead for incoherence between the basis from which measurements are selected and the dictionary D, rather than incoherence within D. Of course, note that the coherence of the dictionary D itself impacts the localization factor $\eta_{s,D}$. We will see later in Theorem 3.1 that the incoherence requirement between the basis and dictionary can be weakened even further by using a weighted approach.

4. The assumption that D is a Parseval frame is not necessary but made for convenience of presentation. Several results have been shown for frames which are not tight, e.g., [LML12, RK13, Gir14]. Indeed, any frame $D \in \mathbb{C}^{n \times N}$ with linearly independent rows is such that

$$\tilde{D} := (DD^*)^{-1/2}D$$

is a tight frame. As observed in the recent paper [Fou15], in this case one can use for sampling the adapted matrix $\tilde{A} = (\tilde{B}(DD^*)^{-1/2})$, as measuring the signal $f = Dz$ through $y = \tilde{A}f + e$ is equivalent to measuring the signal $\tilde{f} = \tilde{D}z$ through $y = \tilde{B}\tilde{f} + e$. Working through this extension poses an interesting direction for future work.

1.5. **Organization.** The rest of the paper is organized as follows. We introduce some notation and review some technical background in Section 2 before presenting a technical version of our main result in Section 3, which demonstrates that bases which are incoherent with a dictionary D can be randomly subsampled to obtain D-RIP matrices. In that section, we also discuss some relevant applications and the implications of our results, including the regime where the sampling basis is coherent to the dictionary. The proof of our main theorem is presented in the final section.

2. NOTATION AND TECHNICAL BACKGROUND

Throughout the paper, we write $C, C', C_1, C_2, \ldots$ to denote absolute constants; their values can change between different occurrences. We write $\mathbf{I}_n$ to denote the $n \times n$ identity matrix and we denote by $[n]$ the set $\{1, 2, \ldots, n\}$. For a vector u, we denote the jth index by $u[j]$ or $u(j)$ or u_j, depending on which notation is the most clear in any given context; its restriction to only those entries indexed by a subset

Λ is donated by u_Λ. These notations should not be confused with the notation $(\boldsymbol{u})_s$, which we use to denote the best s-sparse approximation to $\boldsymbol{u}$, that is, $(\boldsymbol{u})_s \overset{\text{def}}{=} \arg\inf_{\|z\|_0 \le s} \|\boldsymbol{u} - \boldsymbol{z}\|_2$. Similarly for a matrix $\boldsymbol{A}$, $\boldsymbol{A}_j$ is the jth column and $\boldsymbol{A}_\Lambda$ is the submatrices consisting of the columns with indices in Λ.

2.1. **The unweighted case with incoherence.** Recall that a dictionary $\boldsymbol{D} \in \mathbb{C}^{n \times N}$ is a *Parseval frame* if $\boldsymbol{D}\boldsymbol{D}^* = \boldsymbol{I}_n$ and that a vector $\boldsymbol{x}$ is *s-sparse* if $\|\boldsymbol{x}\|_0 := |\operatorname{supp}(\boldsymbol{x})| \le s$. Then the restricted isometry property with respect to the dictionary $\boldsymbol{D}$ ($\boldsymbol{D}$-RIP) is defined as follows.

Definition 2.1 ([CENR10])**.** Fix a dictionary $\boldsymbol{D} \in \mathbb{C}^{n \times N}$ and matrix $\boldsymbol{A} \in \mathbb{C}^{m \times n}$. The matrix $\boldsymbol{A}$ satisfies the $\boldsymbol{D}$-RIP with parameters δ and s if

$$(1 - \delta)\|\boldsymbol{D}\boldsymbol{x}\|_2^2 \le \|\boldsymbol{A}\boldsymbol{D}\boldsymbol{x}\|_2^2 \le (1 + \delta)\|\boldsymbol{D}\boldsymbol{x}\|_2^2 \tag{2.1}$$

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

Note that when $\boldsymbol{D}$ is the identity, this definition reduces to the standard definition of the restricted isometry property [CT05a]. Under such an assumption on the measurement matrix, the following results bound the reconstruction error for the ℓ_1-analysis method.

Theorem 2.2 ([CENR10])**.** *Let $\boldsymbol{D}$ be a tight frame, $\varepsilon > 0$, and consider a matrix $\boldsymbol{A}$ that has the $\boldsymbol{D}$-RIP with parameters $2s$ and $\delta < 0.08$. Then for every signal $\boldsymbol{f} \in \mathbb{C}^n$, the reconstruction $\boldsymbol{f}^\sharp$ obtained from noisy measurements $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{f} + \boldsymbol{e}$, $\|\boldsymbol{e}\|_2 \le \varepsilon$, via the ℓ_1-analysis problem (P_1) above satisfies*

$$\|\boldsymbol{f} - \boldsymbol{f}^\sharp\|_2 \le \varepsilon + \frac{\|\boldsymbol{D}^*\boldsymbol{f} - (\boldsymbol{D}^*\boldsymbol{f})_s\|_1}{\sqrt{s}}. \tag{2.2}$$

Thus as indicated in the introduction, one obtains good recovery for signals $\boldsymbol{f}$ whose analysis coefficients $\boldsymbol{D}^*\boldsymbol{f}$ have a suitable decay.

Both the RIP and the $\boldsymbol{D}$-RIP are closely related to the Johnson-Lindenstrauss lemma [HV11, AC09, BDDW08]. Recall the classical variant of the lemma states that for any $\varepsilon \in (0, 0.5)$ and points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_d \in \mathbb{R}^n$, that there exists a Lipschitz function $f : \mathbb{R}^n \to \mathbb{R}^m$ for some $m = O(\varepsilon^{-2} \log d)$ such that

$$(1 - \varepsilon)\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 \le \|f(\boldsymbol{x}_i) - f(\boldsymbol{x}_j)\|_2^2 \le (1 + \varepsilon)\|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2^2 \tag{2.3}$$

for all $i, j \in \{1, \dots, d\}$. Recent improvements have been made to this statement which in particular show that the map f can be taken as a random linear mapping which satisfies (2.3) with high probability (see, e.g., [Ach03] for an account of such improvements). Indeed, any matrix $\boldsymbol{A} \in \mathbb{C}^{m \times n}$ which for a fixed[1] vector $\boldsymbol{z} \in \mathbb{C}^n$ satisfies

$$\mathbb{P}\left((1 - \delta)\|\boldsymbol{z}\|_2^2 \le \|\boldsymbol{A}\boldsymbol{z}\|_2^2 \le (1 + \delta)\|\boldsymbol{z}\|_2^2\right) \le C e^{-cm}$$

will satisfy the $\boldsymbol{D}$-RIP with high probability as long as m is at least on the order of $s \log(n/s)$ [CENR10]. From this, any matrix satisfying the Johnson-Lindenstrauss lemma will also satisfy the $\boldsymbol{D}$-RIP (see [BDDW08] for the proof of the RIP for such matrices, which directly carries over to the $\boldsymbol{D}$-RIP). Random matrices known to have this property include matrices with independent subgaussian entries (such as Gaussian or Bernoulli matrices), see for example [DG03]. Moreover, it is shown in [KW11] that any matrix that satisfies the classical RIP will satisfy the Johnson-Lindenstrauss lemma and thus the $\boldsymbol{D}$-RIP with high probability after randomizing the signs of the columns. The latter construction allows for structured random matrices with fast multiplication properties such as randomly subsampled Fourier matrices (in combination with the results from [RV08]) and matrices representing subsampled random convolutions (in combination with the results from [RRT12, KMR14]); in both cases, however, again with randomized column signs. While this gives an abundance of such matrices, as mentioned above, it is not always practical or possible to apply random column signs in the sampling procedure.

---

[1] By "fixed" we mean to emphasize that this probability bound must occur for a *single* vector $\boldsymbol{z}$ rather than for all vectors $\boldsymbol{z}$ like one typically sees in the restricted isometry property.

An important general set-up for structured random sensing matrices known to satisfy the regular RIP is the framework of *bounded orthonormal systems*, which includes as a special case the subsampled discrete Fourier transform measurements (without column signs randomized). Such measurements are the only natural measurements possible in many physical systems where compressive sensing is of interest, such as in MRI, radar, and astronomy [LDP07, BS07, HS09, BSO08], as well as in applications to polynomial interpolation [RW12, BDWZ12, RW11] and uncertainty quantification [HD15]. In the following, we recall this set-up (in the discrete setting), see [Rau10, Sec.4] for a detailed account of bounded orthonormal systems and examples.

Definition 2.3 (Bounded orthonormal system). Consider a probability measure ν on the discrete set $[n]$ and a system $\{\boldsymbol{r}_j \in \mathbb{C}^n,\ j \in [n]\}$ that is orthonormal with respect to ν in the sense that

$$\sum_i \boldsymbol{r}_k(i)\boldsymbol{r}_j(i)\nu_i = \delta_{j,k}$$

where $\delta_{j,k} = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{else.} \end{cases}$ is the Kronecker delta function. Suppose further that the system is uniformly bounded: there exists a constant $K \geq 1$ such that

$$\sup_{i\in[n]} \sup_{j\in[n]} |\boldsymbol{r}_j(i)| \leq K. \tag{2.4}$$

Then the matrix $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ whose rows are $\boldsymbol{r}_j$ is called a *bounded orthonormal system* matrix.

Drawing m indices $i_1, i_2, \ldots, i_m$ independently from the orthogonalization measure ν, the sampling matrix $\widetilde{\boldsymbol{A}} \in \mathbb{C}^{m\times n}$ whose rows are indexed by the (re-normalized) sampled vectors $\sqrt{\frac{1}{m}}\boldsymbol{r}_j(i_k) \in \mathbb{C}^n$ will have the restricted isometry property with high probability (precisely, with probability exceeding $1 - n^{-C\log^3(s)}$) provided the number of measurements satisfies

$$m \geq CK^2 s \log^2(s) \log(n). \tag{2.5}$$

This result was first shown in the case where ν is the uniform measure by Rudelson and Vershynin [RV08], for a slightly worse dependence of m on the order of $s\log^2 s \log(s\log n)\log(n)$. These results were subsequently extended to the general bounded orthonormal system set-up by Rauhut [Rau10], and the dependence of m was slightly improved to $s\log^3 s \log n$ in [RW13].

An important special case where these results can be applied is that of *incoherence* between the measurement and sampling bases. Here the *coherence* between two sets of vectors $\Phi = \{\phi_i\}$ and $\Psi = \{\psi_j\}$ is given by $\mu = \sup_{i,j}|\langle\phi_i,\psi_j\rangle|$. Two orthonormal bases Φ and Ψ of $\mathbb{C}^n$ are called incoherent if $\mu \leq Kn^{-1/2}$. In this case, the renormalized system $\tilde{\Phi} = \{\sqrt{n}\phi_i\Psi^*\}$ is an orthonormal system with respect to the uniform measure, which is bounded by K. Then the above results imply that signals which are sparse in basis Ψ can be reconstructed from inner products with a uniformly subsampled subset of basis Φ. These incoherence-based guarantees are a standard criterion to ensure signal recovery, such results had first been observed in [CR07].

2.2. **Generalization to a weighted setup and to local coherence.** Recently, the criterion of incoherence has been generalized to the case where only *most* of the inner products between sparsity basis vectors and measurement basis vectors are bounded. If some vectors in the measurement basis yield larger inner products, one can adjust the sampling measure and work with a preconditioned measurement matrix [KW14], and if some vectors in the sparsity basis yield larger inner products, one can incorporate weights into the recovery schemes and work with a weighted sparsity model [RW13]. Our results can accommodate both these modifications. In the remainder of this section, we will recall some background on these frameworks from [KW14, RW13] and formulate our definitions for the more general setup.

Consider a set of positive weights $\omega = (\omega_j)_{j \in [N]}$. Associate to these weights the norms

$$\|\boldsymbol{x}\|_{\omega,p} := \left(\sum_j |x_j|^p \omega_j^{2-p} \right)^{1/p}, \quad 0 < p \leq 2, \tag{2.6}$$

along with the weighted "ℓ_0-norm", or weighted sparsity: $\|\boldsymbol{x}\|_{\omega,0} := \sum_{j:|x_j|>0} \omega_j^2$, which equivalently is defined as the limit $\|\boldsymbol{x}\|_{\omega,0} = \lim_{p \to 0} \|\boldsymbol{x}\|_{\omega,p}^p$. We say that a vector $\boldsymbol{x}$ is **weighted s-sparse** if $\|\boldsymbol{x}\|_{\omega,0} := \sum_{j:|x_j|>0} \omega_j^2 \leq s$. In line with this definition, the weighted size of a finite set $\Lambda \subset \mathbb{N}$, is given by $\omega(\Lambda) := \sum_{j \in \Lambda} \omega_j^2$; thus a vector is weighted s-sparse if its support has weighted size at most s. When $\omega_j \geq 1$ for all j, the weighted sparsity of a vector is at least as large as its unweighted sparsity, so that the class of weighted s-sparse vectors is a strict subset of the s-sparse vectors. We make this assumption in the remainder. Note in particular the special cases $\|\boldsymbol{x}\|_{\omega,1} = \sum_j |x_j| \omega_j$ and $\|\boldsymbol{x}\|_{\omega,2} = \|\boldsymbol{x}\|_2 = \sqrt{\sum_j |x_j|^2}$; by Cauchy-Schwarz, it follows that $\|\boldsymbol{x}\|_{\omega,1} \leq \sqrt{s}\|\boldsymbol{x}\|_2$ if $\boldsymbol{x}$ is weighted s-sparse. Indeed, we can extend the notions of localization factor (1.1) and $\boldsymbol{D}$-RIP (2.1) to the weighted sparsity setting. It should be clear from context which definition we refer to in the remainder.

Definition 2.4. For a dictionary $\boldsymbol{D} \in \mathbb{C}^{n \times N}$, weights $\omega_1, \omega_2, \ldots, \omega_N \geq 1$, and a sparsity level s, we define the (weighted) *localization factor* as

$$\eta_{\omega,s,\boldsymbol{D}} = \eta \stackrel{\text{def}}{=} \sup_{\|\boldsymbol{Dz}\|_2=1, \|\boldsymbol{z}\|_{\omega,0} \leq s} \frac{\|\boldsymbol{D}^* \boldsymbol{Dz}\|_{\omega,1}}{\sqrt{s}}, \tag{2.7}$$

Definition 2.5. Fix a dictionary $\boldsymbol{D} \in \mathbb{C}^{n \times N}$, weights $\omega_1, \omega_2, \ldots, \omega_N \geq 1$, and matrix $\boldsymbol{A} \in \mathbb{C}^{m \times n}$. The matrix $\boldsymbol{A}$ satisfies the $\boldsymbol{D}$-ωRIP with parameters δ and s if

$$(1-\delta)\|\boldsymbol{Dx}\|_2^2 \leq \|\boldsymbol{ADx}\|_2^2 \leq (1+\delta)\|\boldsymbol{Dx}\|_2^2 \tag{2.8}$$

for all **weighted s-sparse** vectors $\boldsymbol{x} \in \mathbb{C}^N$.

When $\boldsymbol{D}$ is an orthonormal matrix, this definition reduces to the definition ωRIP from [RW13]. More generally, weights allow the flexibility to incorporate prior information about the support set and allow for weaker assumptions on the dictionary $\boldsymbol{D}$. In particular, a larger weight assigned to a dictionary element will allow for this element to have larger inner products with the measurement basis vectors. In this regard, a basis version of our result is the following variant of a theorem from [RW13]:

Proposition 2.6 (from [RW13]). *Fix a probability measure ν on $[n]$, sparsity level $s < n$, and constant $0 < \delta < 1$. Let $\boldsymbol{D} \in \mathbb{C}^{n \times n}$ be an orthonormal matrix. Let $\boldsymbol{A}$ be an orthonormal system with respect to ν as in Definition 2.3 with rows denoted by $\boldsymbol{r}_i$, and consider weights $\omega_1, \omega_2, \ldots, \omega_n \geq 1$ such that $\max_i |\langle \boldsymbol{r}_i, \boldsymbol{d}_j \rangle| \leq \omega_j$. Construct an $m \times n$ submatrix $\widetilde{\boldsymbol{A}}$ of $\boldsymbol{A}$ by sampling rows of $\boldsymbol{A}$ according to the measure ν and renormalizing each row by $\sqrt{\frac{1}{m}}$. Provided*

$$m \geq C\delta^{-2}s\log^3(s)\log(n), \tag{2.9}$$

then with probability $1 - n^{c\log^3 s}$, the submatrix $\widetilde{\boldsymbol{A}}$ satisfies the ωRIP with parameters s and δ.

An adjusted sampling density for the case when certain vectors in the measurement basis yield larger inner products is obtained via the **local coherence** as defined in the following.

Definition 2.7 (Local coherence, [KW14, RW12]). The *local coherence* of a set $\Phi = \{\varphi_i\}_{i=1}^k \subset \mathbb{C}^n$ with respect to another set $\Psi = \{\psi_j\}_{j=1}^\ell \subseteq \mathbb{C}^n$ is the function $\mu^{\text{loc}}(\Phi, \Psi) \in \mathbb{R}^k$ defined coordinate-wise by

$$\mu_i^{\text{loc}}(\Phi, \Psi) = \mu_i^{loc} = \sup_{1 \leq j \leq \ell} |\langle \varphi_i, \psi_j \rangle|, \qquad i = 1, 2, \ldots, k.$$

When Φ and Ψ are orthonormal bases, and a subset of Φ is used for sampling while Ψ is the basis in which sparsity is assumed, the main point is that renormalizing the different vectors φ_i in the sampling basis by respective factors $\frac{1}{\mu_i^{loc}}$ does not affect orthogonality of the rows of $\Phi^* \Psi$. In this way, the larger inner products can be reduced in size. To compensate for this renormalization and retain the orthonormal system property in the sense of Definition 2.3, one can then adjust the sampling measure. The resulting preconditioned measurement system with variable sampling density will then yield better bounds K in Equation (2.5). This yields the following estimate for the restricted isometry property.

Proposition 2.8 ([KW14, RW12])**.** *Let* $\Phi = \{\varphi_j\}_{j=1}^n$ *and* $\Psi = \{\psi_k\}_{k=1}^n$ *be orthonormal bases of* $\mathbb{C}^n$. *Assume the local coherence of* Φ *with respect to* Ψ *is pointwise bounded by the function* $\kappa/\sqrt{n} \in \mathbb{C}^n$, *that is* $\sup_{1 \le k \le n} |\langle \varphi_j, \psi_k \rangle| \le \kappa_j$. *Suppose*

$$m \ge C\delta^{-2} \|\kappa\|_2^2 s \log^3(s) \log(n), \tag{2.10}$$

and choose m (possibly not distinct) indices $j \in [n]$ *i.i.d. from the probability measure* ν *on* $[n]$ *given by*

$$\nu(j) = \frac{\kappa_j^2}{\|\kappa\|_2^2}.$$

Call this collection of selected indices Ω *(which may possibly be a multiset). Consider the matrix* $A \in \mathbb{C}^{m \times n}$ *with entries*

$$A_{j,k} = \langle \varphi_j, \psi_k \rangle, \quad j \in \Omega, k \in [n], \tag{2.11}$$

and consider the diagonal matrix $W = \mathrm{diag}(w) \in \mathbb{C}^{n \times n}$ *with* $w_j = \|\kappa\|_2/\kappa_j$. *Then with probability at least* $1 - n^{-c \log^3(s)}$, *the preconditioned matrix* $\frac{1}{\sqrt{m}} W A$ *has the restricted isometry property with parameters* δ *and* s.

Remark. In case Φ and Ψ are incoherent, or if $\kappa_j = K n^{-1/2}$ uniformly for all j, then local coherence sampling as above reduces to the previous results for incoherent systems: in this case, the associated probability measure ν is uniform, and the preconditioned matrix reduces to $\frac{1}{\sqrt{m}} W A = \sqrt{\frac{n}{m}} A$.

Our main result on D-RIP for redundant systems and structured measurements (Theorem 3.1) implies a strategy for extending local coherence sampling theory to dictionaries, of which Theorem 1.2 is a special case. Indeed, if $\Psi = \{\psi_j\}$, more generally, is a Parseval frame and $\Phi = \{\varphi_i\}$ is an orthonormal matrix, then renormalizing the different vectors φ_i in the sampling basis by respective factors $\frac{1}{\mu_i^{loc}}$ still does not affect orthogonality of the rows of $\Phi^* \Psi$. We have the following corollary of our main result; we will establish a more general result, Theorem 3.1 in the next section, from which both Theorem 1.2 and Corollary 2.9 follow.

Corollary 2.9. *Fix a sparsity level* $s < N$, *and constant* $0 < \delta < 1$. *Let* $D \in \mathbb{C}^{n \times N}$ *be a Parseval frame with columns* $\{d_1, \ldots, d_N\}$, *let* B *with rows* $\{b_1, \ldots, b_n\}$ *be an orthonormal basis of* $\mathbb{C}^n$, *and assume the local coherence of* D *with respect to* B *is pointwise bounded by the function* $\kappa \in \mathbb{C}^n$, *that is* $\sup_{1 \le j \le N} |\langle d_j, b_k \rangle| \le \kappa_k$.

Consider the probability measure ν *on* $[n]$ *given by* $\nu(k) = \frac{\kappa_k^2}{\|\kappa\|_2^2}$ *along with the diagonal matrix* $W = \mathrm{diag}(w) \in \mathbb{C}^{n \times n}$ *with* $w_k = \|\kappa\|_2/\kappa_k$. *Note that the renormalized system* $\Phi = \frac{1}{\sqrt{m}} W B$ *is an orthonormal system with respect to* ν, *bounded by* $\|\kappa\|^2$. *Construct* $\widetilde{B} \in \mathbb{C}^{m \times n}$ *by sampling vectors from* B *i.i.d. from the measure* ν, *and consider the localization factor* $\eta = \eta_{D,s}$ *of the frame* D *as in Definition 1.1,*

As long as

$$m \ge C\delta^{-2} \eta^2 \|\kappa\|_2^2 s \log^3(s\eta^2) \log(N)$$

then with probability $1 - N^{-\log^3(2s)}$, *the following holds for every signal* f: *the solution* $f^\sharp$ *of the weighted* ℓ_1-*analysis problem*

$$f^\sharp = \operatorname*{argmin}_{\tilde{f} \in \mathbb{C}^n} \|D^* \tilde{f}\|_1 \quad \text{such that} \quad \|\frac{1}{\sqrt{m}} W(\widetilde{B} \tilde{f} - y)\|_2 \le \varepsilon, \tag{$P_{1,w}$}$$

with $y = Bf + e$ for noise e with weighted error $\|\frac{1}{\sqrt{m}}We\|_2 \leq \varepsilon$ satisfies

$$\|f^\sharp - f\|_2 \leq C_1\varepsilon + C_2\frac{\|D^*f - (D^*f)_s\|_1}{\sqrt{s}}. \tag{2.12}$$

Here C, C_1, and C_2 are absolute constants independent of the dimensions and the signal f.

The resulting bound involves a weighted noise model, which can, in the worst case, introduce an additional factor of $\sqrt{\frac{n}{m}}\max_i\sqrt{\nu(i)}$. We believe however that this noise model is just an artifact of the proof technique, and that the stability results in Corollary 2.9 should hold for the standard noise model using the standard ℓ_1-analysis problem (P_1). In the important case where D is a wavelet frame and B is the orthonormal discrete Fourier matrix, we believe that *total variation minimization*, like ℓ_1-analysis, should give stable and robust error guarantees with the standard measurement noise model. Indeed, such results were recently obtained for variable density sampling in the case of orthonormal wavelet sparsity basis [Poo15a], improving on previous bounds for total variation minimization [KW14, NW13b, NW13a, AHPR13]. Generalizing such results to the dictionary setting is indeed an interesting direction for future research.

3. OUR MAIN RESULT AND ITS APPLICATIONS

As mentioned in the previous section, the case of incoherence between the sampling basis and the sparsity dictionary, as in Theorem 1.2, is a special case of a bounded orthonormal system. The following more technical formulation of our main result in the framework of such systems covers both weighted sparsity models and local coherence and as we will see, implies Theorem 1.2. The proof of Theorem 3.1 is presented in Section 4.

Theorem 3.1. *Fix a probability measure ν on $[N]$, sparsity level $s < N$, and constant $0 < \delta < 1$. Let $D \in \mathbb{C}^{n \times N}$ be a Parseval frame. Let A be an orthonormal systems matrix with respect to ν with rows r_i as in Definition 2.3, and consider weights $\omega_1, \omega_2, \ldots, \omega_N \geq 1$ such that*

$$\max_i |\langle r_i, d_j \rangle| \leq \omega_j. \tag{3.1}$$

Define the unrecoverable energy $\eta = \eta_{D,s}$ as in Definition 2.7. Construct an $m \times n$ submatrix $\widetilde{A}$ of A by sampling rows of A according to the measure ν. Then as long as

$$\begin{aligned}
m &\geq C\delta^{-2}s\eta^2\log^3(s\eta^2)\log(N), \quad and \\
m &\geq C\delta^{-2}s\eta^2\log(1/\gamma)
\end{aligned} \tag{3.2}$$

then with probability $1 - \gamma$, the normalized submatrix $\sqrt{\frac{1}{m}}\widetilde{A}$ satisfies the D-ωRIP with parameters s and δ.

Proof of Theorem 1.2 assuming Theorem 3.1: Let B, D and s be given as in Theorem 1.2. We will apply Theorem 3.1 with $\omega_j = 1$ for all j, with ν the uniform measure on $[n]$, sparsity level $2s$, $\gamma = N^{-\log^3(2s)}$, $\delta = 0.08$, and matrix $A = \sqrt{n}B$. We first note that since B is an orthonormal basis, that $\sqrt{n}B$ is an orthonormal systems matrix with respect to the uniform measure ν as in Definition 2.3. In addition, (1.2) implies (3.1) for matrix $A = \sqrt{n}B$ with $K = \omega_j = 1$. Furthermore, setting $\gamma = N^{-\log^3(2s)}$, (1.3) implies both inequalities of (3.2) (adjusting constants appropriately). Thus the assumptions of Theorem 3.1 are in force. Theorem 3.1 then guarantees that with probability $1 - \gamma = 1 - N^{-\log^3(2s)}$, the uniformly subsampled matrix $\sqrt{\frac{1}{m}}(\sqrt{n}\widetilde{B})$ satisfies the D-ωRIP with parameters $2s$ and δ and weights $\omega_j = 1$. A simple calculation and the definition of the D-ωRIP (Definition 2.5) shows that this implies the D-RIP with parameters $2s$ and δ. By Theorem 2.2, (1.4) holds and this completes the proof. $\square$

Proof of Corollary 2.9 assuming Theorem 3.1. Corollary 2.9 is implied by Theorem 3.1 because the pre-conditioned matrix $\Phi = \frac{1}{\sqrt{m}} W B$ formed by normalizing the ith row of B by $\|\kappa\|_2/\kappa_j$ constitutes an orthonormal system with respect to the probability measure $\nu(i)$ and uniform weights $\omega_j = \|\kappa\|_2$. Then the preconditioned sampled matrix $\widetilde{\Phi}$ satisfies the D-RIP for sparsity $2s$ and parameter δ according to Theorem 3.1 with probability $1 - \gamma = 1 - N^{-\log^3(2s)}$. Applying Theorem 2.2 to the preconditioned sampling matrix $\widetilde{\Phi}$ and Parseval frame D produces the results in Corollary 2.9.   □

3.1. **Example: Harmonic frame with L more vectors than dimensions.** It remains to find examples of a dictionary with bounded localization factor and an associated measurement system for which incoherence condition (1.2) holds. Our main example is that of sampling a signal that is sparse in an oversampled Fourier system, a so-called *harmonic frame* [VW04]; the measurement system is the standard basis. Indeed, one can see by direct calculation that the standard basis is incoherent in the sense of (1.2) to any set of Fourier vectors, even of non-integer frequencies. We will now show that if the number of frame vectors exceeds the dimension only by a constant, such a dictionary will also have bounded localization factor. This setup is a simple example, but our results apply, and it is not covered by previous theory.

More precisely, we fix $L \in \mathbb{N}$ and consider $N = n + L$ vectors in dimension n. We assume that L is such that $Ls \le \frac{N}{4}$. Then the harmonic frame is defined via its frame matrix $D = (d_{jk})$, which results from the $N \times N$ discrete Fourier transform matrix $F = (f_{jk})$ by deleting the last L rows. That is, we have

$$d_{jk} = \tfrac{1}{\sqrt{n+L}} \exp(\tfrac{2\pi i j k}{n+L}) \tag{3.3}$$

for $j = 1 \dots n$ and $k = 1 \dots N$. The corresponding Gram matrix satisfies $(D^* D)_{ii} = \frac{n}{n+L}$ and, for $j \ne k$, by orthogonality of F,

$$|(D^* D)_{jk}| = \left| (F^* F)_{jk} - \sum_{\ell=n+1}^{N} \bar{f}_{\ell j} f_{\ell k} \right| = \left| \sum_{\ell=n+1}^{N} \tfrac{1}{n+L} \exp(\tfrac{-2\pi i \ell j}{n+L}) \exp(\tfrac{2\pi i \ell k}{n+L}) \right| \le \tfrac{L}{n+L}$$

As a consequence, we have for z s-sparse and $j \notin \operatorname{supp} z$

$$|(D^* D z)[j]| = \left| \sum_{k \in \operatorname{supp} z} (D^* D)_{jk} z[k] \right| \le \tfrac{L}{n+L} \|z\|_1$$

where we write $z[k]$ to denote the kth index of the vector z. Similarly, for $j \in \operatorname{supp} z$, we obtain

$$|(D^* D z)[j] - z[j]| = \left| \sum_{k \in \operatorname{supp} z, k \ne j} (D^* D)_{jk} z[k] \right| \le \tfrac{L}{n+L} \|z\|_1.$$

So we obtain, using that D^* is an isometry,

$$\|Dz\|_2^2 = \|D^* D z\|_2^2 \ge \sum_{k \in \operatorname{supp} z} \big((D^* D z)[k]\big)^2 \ge \sum_{k \in \operatorname{supp} z} \big(|z[k]| - |(D^* D z)[k] - z[k]|\big)^2$$

$$\ge \sum_{k \in \operatorname{supp} z} |z[k]|^2 - 2 \tfrac{L}{n+L} \|z\|_1 |z[k]|$$

$$= \|z\|_2^2 - \tfrac{2L}{n+L} \|z\|_1^2 \ge (1 - \tfrac{2Ls}{N}) \|z\|_2^2 \ge \tfrac{1}{2} \|z\|_2^2.$$

That is, for z with $\|Dz\|_2 = 1$, one has $\|z\|_2 \le \sqrt{2}$. Consequently,

$$\eta = \sup_{\|Dz\|_2=1, \|z\|_0 \le s} \frac{\|D^* D z\|_1}{\sqrt{s}} = \sup_{\|Dz\|_2=1, \|z\|_0 \le s} \frac{\|(D^* D z)|_{(\operatorname{supp} z)}\|_1 + \|(D^* D z)|_{(\operatorname{supp} z)^c}\|_1}{\sqrt{s}}$$

$$\le \sup_{\|Dz\|_2=1, \|z\|_0 \le s} \|D^* D z\|_2 + \frac{\|(D^* D z)|_{(\operatorname{supp} z)^c}\|_1}{\sqrt{s}} = 1 + \sup_{\|Dz\|_2=1, \|z\|_0 \le s} \frac{1}{\sqrt{s}} \sum_{j \in (\operatorname{supp} z)^c} |(D^* D z)[j]|$$

$$\le 1 + \sup_{\|Dz\|_2=1, \|z\|_0 \le s} \tfrac{1}{\sqrt{s}} (N - s) \tfrac{L\|z\|_1}{N} \le 1 + \tfrac{(N-s)L\|z\|_2}{N} \le 1 + L\sqrt{2}.$$

                                                                                    □

Remark. We note here that the dependence on L seems quite pessimistic. Indeed, one often instead considers the redundancy of the dictionary, namely $r = N/n$. In this case, even for redundancy $r = 2$ we would have $L = n$ and require a full set of measurements according to this argument. However, we conjecture that rather than a dependence on L one can reduce the dependence to one on the redundancy r. Intuitively, due to the restriction of Parseval normalization on the frame $\boldsymbol{D}$, there is a tradeoff between redundancy and coherence; an increase in redundancy should be in some way balanced by a decrease in coherence. This conjecture is supported by numerical experiments and by theoretical lower bounds for robust recovery in the co-sparse model [GPV15], but we leave a further investigation to future work.

This example also plays an important role in the so-called *off the grid* compressed sensing setting [TBSR12, SB12]. In this framework, the signal frequencies are not assumed to lie on a lattice but instead can assume any values in a continuous interval. When the signal parameters do not lie on a lattice, the signal may not be truly sparse in the discrete dictionary, and refining the grid to incorporate finer frequencies may lead to numerical instability. In addition, typical compressed results are difficult to apply under discretization, making off the grid approaches advantageous. The example we discuss above gives a possible compromise to nearly on the grid recovery from linear measurements.

Another line of research that relates to this example is that of superresolution (see, e.g., [CFG14, CFG13] and many followup works). These works study the recovery of frequency sparse signals from equispaced samples. No assumptions regarding an underlying grid are made, but rather one assumes a separation distance between the active frequencies. In this sense, the nearly-on-the-grid example just discussed satisfies their assumption as the corresponding separation distance is close to the grid spacing. However, the nature of the resulting guarantees is somewhat different. For example, in these works, every deviation from an exactly sparse signal must be treated as noise, whereas in our result above we have an additional term capturing the compressibility of a signal. We think that an in-depth comparison of the two approaches is an interesting topic for future work.

3.2. **Example: Fourier measurements and Haar frames of redundancy 2.**

In this subsection, we present a second example of a sampling setup that satisfies the assumptions of incoherence and localization factor of Theorem 3.1. In contrast to the previous example, one needs to precondition and adjust the sampling density to satisfy these assumptions according to Corollary 2.9, which allows us to understand the setup of the Fourier measurement basis and a 1D Haar wavelet frame with redundancy 2, as introduced in the following. Let $n = 2^p$. Recall that the univariate discrete Haar wavelet basis of $\mathbb{C}^{2^p}$ consists of $h^0 = 2^{-p/2}(1, 1, \ldots, 1), h = h_{0,0} = 2^{-p/2}(1, 1, \ldots, 1, -1, -1, \ldots, -1)$ and the frame basis elements $h_{\ell,k}$ given component wise by

$$h_{\ell,k}[j] = h[2^\ell j - k] = \begin{cases} 2^{\frac{\ell-p}{2}} & \text{for} & k2^{p-\ell} \le j < k2^{p-\ell} + 2^{p-\ell-1} \\ -2^{\frac{\ell-p}{2}} & \text{for} & k2^{p-\ell} + 2^{p-\ell-1} \le j < k2^{p-\ell} + 2^{p-\ell} \\ 0 & \text{else,} \end{cases}$$

for any $(\ell, k) \in \mathbb{Z}^2$ satisfying $0 < \ell < p$ and $0 \le k < 2^\ell$. The corresponding basis transformation matrix is denoted by $\boldsymbol{H}$.

One can now create a *wavelet frame of redundancy 2* by considering the union of this basis and a circular shift of it by one index. That is, one adds the vector $\tilde{h}^0 = h^0 = 2^{-p/2}(1, 1, \ldots, 1)$ and vectors of the form $\tilde{h}_{\ell,k}[j] = h_{\ell,k}[j + 1]$ for all (ℓ, k). Here we identify $2^p + 1 = 1$. This is also an orthonormal basis – its basis transformation matrix will be denoted by $\tilde{\boldsymbol{H}}$ in the following, and the matrix $\boldsymbol{D} \in \mathbb{C}^{2^p \times 2^{p+1}}$ with columns

$$\boldsymbol{D}(:, (\ell, 2k-1)) = \frac{1}{\sqrt{2}} h_{\ell,k},$$

$$\boldsymbol{D}(:, (\ell, 2k)) = \frac{1}{\sqrt{2}} \tilde{h}_{\ell,k} \tag{3.4}$$

forms a Parseval frame with redundancy 2.

Corollary 2.9 applies to the example where sparsity is with respect to the redundant Haar frame and where sampling measurements are rows $\{r_k\}$ from the $n \times n$ orthonormal DFT matrix. Indeed, following Corollary 6.4 of [KW14], we have the following coherence estimates:

$$\max_{\ell,j}\{|\langle r_k, h_{\ell,j}\rangle|, |\langle r_k, \tilde{h}_{\ell,j}\rangle|\} \leq \kappa_k := 3\sqrt{2\pi}/\sqrt{k}.$$

Since $\|\kappa\|_2^2 = 18\pi \sum_{k=1}^n k^{-1} \leq 18\pi \log_2(n)$ grows only mildly with n, the Fourier / wavelet frame example is a good fit for Corollary 2.9, provided the localization factor of the Haar frame is also small. We will show that the localization factor of the Haar frame is bounded by $\eta \leq \sqrt{2\log_2(n)}$, leading to the following corollary.

Corollary 3.2. *Fix a sparsity level $s < N$, and constant $0 < \delta < 1$. Let $D \in \mathbb{C}^{n \times N}$ ($N = 2n$) be the redundant Haar frame as defined in* (3.4) *and let $A \in \mathbb{C}^{n \times n}$ with rows $\{r_1, \ldots, r_n\}$ be the orthonormal DFT matrix. Consider the diagonal matrix $W = \text{diag}(w) \in \mathbb{C}^{n \times n}$ with $w_k = C'(\log_2(n))^{-1/2}k^{1/2}$, and construct $\tilde{A} \in \mathbb{C}^{m \times n}$ by sampling rows from A i.i.d. from the probability measure ν on $[n]$ with power-law decay $\nu(k) = Ck^{-1}$.*

As long as the number of measurements satisfies $m \geq C_1\delta^{-2}s\log^3(s\log(n))\log^3(n)$, then with probability $1 - n^{-\log^3 s}$, the following holds for every signal f: the solution $f^\sharp$ of the weighted ℓ_1-analysis problem $(P_{1,w})$ with $y = \tilde{A}f + e$ for noise e with weighted error $\|\frac{1}{\sqrt{m}}We\|_2 \leq \varepsilon$ satisfies

$$\|f^\sharp - f\|_2 \leq C_2\varepsilon + C_3\frac{\|D^*f - (D^*f)_s\|_1}{\sqrt{s}}$$

Here C, C_1, C_2, and C_3 are absolute constants independent of the dimensions and the signal f.

Proof. To derive this result from Corollary 2.9, it suffices to prove that the redundant Haar frame as defined above has localization factor at most $\eta \leq \sqrt{2\log_2(n)}$. We will show that for each s-sparse $z \in \mathbb{R}^N$, D^*Dz is at most $3s\log_2 n$-sparse. Here, D^*D is the $N \times N$ Gramian matrix. For that, consider first either of the two (equal) frame elements $\frac{1}{\sqrt{2}}h^0$ and $\frac{1}{\sqrt{2}}\tilde{h}^0$. Each of these frame elements has non-zero inner product with exactly two frame elements: $\frac{1}{\sqrt{2}}h^0$ and $\frac{1}{\sqrt{2}}\tilde{h}^0$. So the corresponding columns of D^*D have only two non-vanishing entries.

Consider then a non-constant frame element $\frac{1}{\sqrt{2}}h_{\ell,k}$. Because H is an orthonormal basis, it is orthogonal to all $h_{\ell',k'}$ with $(\ell', k') \neq (\ell, k)$, which is why $(D^*D)_{(k,2\ell),(k',2\ell')} = (D^*D)_{(k,2\ell-1),(k',2\ell'-1)} = 0$ for $(\ell', k') \neq (\ell, k)$. So it remains to consider correlations with the $\tilde{h}_{\ell',k'}$. Again by orthogonality, one has for $(\ell', k') \neq (\ell, k)$

$$\langle \tilde{h}_{\ell',k'}, h_{\ell,k}\rangle = \delta_{(\ell',k'),(\ell,k)} + \langle \tilde{h}_{\ell',k'}, h_{\ell,k} - \tilde{h}_{\ell,k}\rangle,$$

where $\delta_{(a,b),(c,d)} = \begin{cases} 1 & \text{if } (a,b) = (c,d) \\ 0 & \text{else.} \end{cases}$ denotes the Kronecker delta. By definition, $h_{\ell,k} - \tilde{h}_{\ell,k}$ has only three non-zero entries. On the other hand, for each level ℓ', the supports of the $\tilde{h}_{\ell',k'}$ are disjoint for different values of k'. So for each ℓ', the supports of at most 3 of the $\tilde{h}_{\ell',k'}$ can intersect the support of $h_{\ell,k} - \tilde{h}_{\ell,k}$. As for $\ell' = \ell$, one of these $\tilde{h}_{\ell',k'}$ must be $\tilde{h}_{\ell,k}$, we conclude that for each ℓ' at most 3 of the $|\langle \tilde{h}_{\ell',k'}, h_{\ell,k}\rangle|$ can be nonzero. As there are $\log_2 n$ levels ℓ', this contributes at most $3\log_2 n$ nonzero entries in the column of D^*D indexed by $(k,2\ell)$. Together with $(D^*D)_{(k,2\ell),(k,2\ell)} = 1$ and noting that a similar analysis holds for columns indexed by $(k, 2\ell-1)$, we obtain that each column of D^*D has at most $3\log_2(n) + 1$ non-zero entries. Now for each s-sparse z, D^*Dz is a linear combination of the s columns corresponding to $\text{supp}\,z$. Consequently, $\|D^*Dz - (D^*Dz)_s\|_0 \leq 3s\log_2(n)$ and thus, by Cauchy-Schwarz

and noting that for a Parseval frame $\boldsymbol{D}$, $\boldsymbol{D}\boldsymbol{D}^*$ is the identity,

$$
\begin{aligned}
\eta &= \sup_{\|\boldsymbol{D}z\|_2=1, \|z\|_0 \le s} \frac{\|\boldsymbol{D}^*\boldsymbol{D}z\|_1}{\sqrt{s}} = \sup_{\|\boldsymbol{D}^*\boldsymbol{D}z\|_2=1, \|z\|_0 \le s} \frac{\|(\boldsymbol{D}^*\boldsymbol{D}z)_s\|_1 + \|\boldsymbol{D}^*\boldsymbol{D}z - (\boldsymbol{D}^*\boldsymbol{D}z)_s\|_1}{\sqrt{s}} \\
&\le 1 + \sup_{\|\boldsymbol{D}^*\boldsymbol{D}z\|_2=1, \|z\|_0 \le s} \frac{\sqrt{3s\log_2 n}\|\boldsymbol{D}^*\boldsymbol{D}z - (\boldsymbol{D}^*\boldsymbol{D}z)_s\|_2}{\sqrt{s}} \\
&\le 1 + \sup_{\|\boldsymbol{D}^*\boldsymbol{D}z\|_2=1} \sqrt{3\log_2 n}\|\boldsymbol{D}^*\boldsymbol{D}z\|_2 = 1 + \sqrt{3\log_2 n}.
\end{aligned}
$$

$\square$

Remark. Note that this proof is closely related to the observation that each signal that is s-sparse in the Haar frame of redundancy 2 is also $O(s\log n)$-sparse in the Haar wavelet basis. So a number of measurements comparable to the one derived here also allows for recovery of the wavelet basis coefficients. In addition, however, a more refined analysis suggests that the entries of $\boldsymbol{D}^*\boldsymbol{D}z$ decay quickly - provided not too many approximate cancellations happen. We conjecture that the number of such approximate cancellations can be controlled using a more sophisticated analysis, but we leave this to follow-up work. We hence conjecture that the logarithmic factor can be removed, making the required number of measurements smaller than for a synthesis approach.

4. PROOF OF MAIN RESULT

Our proof of Theorem 3.1 extends the analysis in [RW13], which extends the analysis in [CGV13] to weighted sparsity and improves on the analysis in [RV08], from orthonormal systems to redundant dictionaries.

Proof of Theorem 3.1. Consider the set

$$
\mathscr{D} \overset{\text{def}}{=} \{\boldsymbol{u} : \boldsymbol{u} \in \mathbb{C}^n, \boldsymbol{u} = \boldsymbol{D}z, \|z\|_{\omega,0} \le s, \|\boldsymbol{u}\|_2 = 1\}.
$$

Let $\tilde{\boldsymbol{A}}$ be the $m \times n$ subsampled matrix of $\boldsymbol{A}$ as in the theorem. Then observe that the smallest value $\delta = \delta_s$ which satisfies the $\boldsymbol{D}$-ωRIP bound (2.8) for $\tilde{\boldsymbol{A}}$ is precisely

$$
\delta_s = \sup_{\boldsymbol{u} \in \mathscr{D}} \boldsymbol{u}^*(\tilde{\boldsymbol{A}}^*\tilde{\boldsymbol{A}} - \mathbf{I}_n)\boldsymbol{u}.
$$

Since $\tilde{\boldsymbol{A}}^*\tilde{\boldsymbol{A}} - \mathbf{I}_n$ is a self-adjoint operator, we may instead define for any self-adjoint matrix $\boldsymbol{B}$ the operator

$$
\|\|\boldsymbol{B}\|\|_s \overset{\text{def}}{=} \sup_{\boldsymbol{u} \in \mathscr{D}} \langle \boldsymbol{B}\boldsymbol{u}, \boldsymbol{u} \rangle.
$$

and equivalently write

$$
\delta_s = \|\|\tilde{\boldsymbol{A}}^*\tilde{\boldsymbol{A}} - \mathbf{I}_n\|\|_s.
$$

Our goal is thus to bound this quantity. To that end, let $\boldsymbol{r}_1, \boldsymbol{r}_2, \dots, \boldsymbol{r}_m \in \mathbb{C}^n$ denote the m randomly selected rows of $\boldsymbol{A}$ that make up $\tilde{\boldsymbol{A}}$. It follows from Definition 2.3 that $\mathbb{E}\boldsymbol{r}_i^*\boldsymbol{r}_i = \mathbf{I}_n$ for each i, where the expectation is taken with respect to the sampling measure ν. We thus have

$$
\delta_s = \|\|\tilde{\boldsymbol{A}}^*\tilde{\boldsymbol{A}} - \mathbf{I}_n\|\|_s = \left\|\left\| \frac{1}{m}\sum_{i=1}^m \boldsymbol{r}_i^*\boldsymbol{r}_i - \mathbf{I}_n \right\|\right\|_s = \frac{1}{m}\left\|\left\| \sum_{i=1}^m \left(\boldsymbol{r}_i^*\boldsymbol{r}_i - \mathbb{E}\boldsymbol{r}_i^*\boldsymbol{r}_i\right) \right\|\right\|_s. \tag{4.1}
$$

We may bound the moments of this quantity using a symmetrization argument (see, e.g., Lemma 6.7 of [Rau10]):

$$\mathbb{E}\left\|\left\|\sum_{i=1}^{m}\left(\boldsymbol{r}_i^*\boldsymbol{r}_i - \mathbb{E}\boldsymbol{r}_i^*\boldsymbol{r}_i\right)\right\|\right\|_s \leq 2\left(\mathbb{E}\left\|\left\|\sum_{i=1}^{m}\epsilon_i\boldsymbol{r}_i^*\boldsymbol{r}_i\right\|\right\|_s\right)$$

$$= 2\mathbb{E}_{\boldsymbol{r}}\mathbb{E}_{\epsilon}\sup_{\boldsymbol{x}\in\mathscr{D}}|\langle\sum_{i=1}^{m}\epsilon_i\boldsymbol{r}_i^*\boldsymbol{r}_i\boldsymbol{x},\boldsymbol{x}\rangle|$$

$$= 2\mathbb{E}_{\boldsymbol{r}}\mathbb{E}_{\epsilon}\sup_{\boldsymbol{x}\in\mathscr{D}}\left|\sum_{i=1}^{m}\epsilon_i|\langle\boldsymbol{r}_i,\boldsymbol{x}\rangle|^2\right| \tag{4.2}$$

where ϵ_i are independent symmetric Bernoulli (± 1) random variables.

Conditional on $(\boldsymbol{r}_i)$, we have a subGaussian process indexed by $\mathscr{D}$. For a set T, a metric d, and a given $t > 0$, the *covering number* $\mathcal{N}(T, d, t)$ is defined as the smallest number of balls of radius t centered at points of T necessary to cover T with respect to d. For fixed $\boldsymbol{r}_i$, we work with the (pseudo-)metric

$$d(\boldsymbol{x},\boldsymbol{z}) = \left(\sum_{i=1}^{m}(|\langle\boldsymbol{r}_i,\boldsymbol{x}\rangle|^2 - |\langle\boldsymbol{r}_i,\boldsymbol{z}\rangle|^2)^2\right)^{1/2} \tag{4.3}$$

Then Dudley's inequality [Dud67] implies

$$\mathbb{E}_{\epsilon}\sup_{\boldsymbol{x}\in\mathscr{D}}|\langle\sum_{i=1}^{m}\epsilon_i\boldsymbol{r}_i^*\boldsymbol{r}_i\boldsymbol{x},\boldsymbol{x}\rangle| \leq 4\sqrt{2}\int_0^{\infty}\sqrt{\log(\mathcal{N}(\mathscr{D},d,t))}\,dt. \tag{4.4}$$

Continuing as in [RW13], inspired by the approach of [CGV13], we estimate the metric d using Hölder's inequality with exponents $p \geq 1$ and $q \geq 1$ satisfying $1/p + 1/q = 1$ to be specified later. Using also the reverse triangle inequality, we have for $\boldsymbol{x}, \boldsymbol{z} \in \mathscr{D}$,

$$d(\boldsymbol{x},\boldsymbol{z}) \leq 2\sup_{\boldsymbol{u}\in\mathscr{D}}\left(\sum_{i=1}^{m}|\langle\boldsymbol{r}_i,\boldsymbol{u}\rangle|^{2p}\right)^{1/(2p)}\left(\sum_{i=1}^{m}|\langle\boldsymbol{r}_i,\boldsymbol{x}-\boldsymbol{z}\rangle|^{2q}\right)^{1/(2q)}. \tag{4.5}$$

We will optimize over the values of p, q later on. To further bound this quantity, we have $|\langle\boldsymbol{r}_i,\boldsymbol{d}_j\rangle| \leq \omega_j$ by assumption. Recall the localization factor in Definition 2.4. Then, fixing $\boldsymbol{u} \in \mathscr{D}$ such that $\boldsymbol{u} = \boldsymbol{D}\boldsymbol{z}$ with $\|\boldsymbol{z}\|_{\omega,0} \leq s$, we have for any realization of $(\boldsymbol{r}_i)$,

$$|\langle\boldsymbol{r}_i,\boldsymbol{u}\rangle| = |\langle\boldsymbol{D}\boldsymbol{D}^*\boldsymbol{r}_i,\boldsymbol{u}\rangle|$$

$$= |\langle\boldsymbol{D}^*\boldsymbol{r}_i,\boldsymbol{D}^*\boldsymbol{D}\boldsymbol{z}\rangle|$$

$$\leq \sum_j|\langle\boldsymbol{r}_i,\boldsymbol{d}_j\rangle|_j|(\boldsymbol{D}^*\boldsymbol{D}\boldsymbol{z})_j|$$

$$\leq \sum_j\omega_j|(\boldsymbol{D}^*\boldsymbol{D}\boldsymbol{z})_j|$$

$$= \|\boldsymbol{D}^*\boldsymbol{D}\boldsymbol{z}\|_{\omega,1}$$

$$\leq \sqrt{s}\eta. \tag{4.6}$$

The first line uses that $\boldsymbol{D}$ is a Parseval frame, while the last line uses the definition of η. The quantity $s\eta^2$ will serve as a rescaled sparsity parameter throughout the remaining proof.

Continuing from (4.5), we may bound

$$\sup_{\boldsymbol{u}\in\mathscr{D}}\left(\sum_{i=1}^{m}|\langle\boldsymbol{r}_i,\boldsymbol{u}\rangle|^{2p}\right)^{1/(2p)} = \sup_{\boldsymbol{u}\in\mathscr{D}}\left(\sum_{i=1}^{m}|\langle\boldsymbol{r}_i,\boldsymbol{u}\rangle|^2|\langle\boldsymbol{r}_i,\boldsymbol{u}\rangle|^{2p-2}\right)^{1/(2p)}$$

$$\leq (s\eta^2)^{(p-1)/(2p)}\left(\sup_{\boldsymbol{u}\in\mathscr{D}}\sum_{i=1}^{m}|\langle\boldsymbol{r}_i,\boldsymbol{u}\rangle|^2\right)^{1/(2p)}. \tag{4.7}$$

We now introduce the (semi-)norm

$$\|\boldsymbol{u}\|_{X,q} := \left(\sum_{i=1}^{m} |\langle \boldsymbol{r}_i, \boldsymbol{u} \rangle|^{2q} \right)^{1/(2q)}.$$

Using basic properties of covering numbers and the bound in (4.4), we obtain

$$\mathbb{E}_{\epsilon} \sup_{\boldsymbol{u} \in \mathscr{D}} |\langle \sum_{i=1}^{m} \epsilon_i \boldsymbol{r}_i^* \boldsymbol{r}_i \boldsymbol{u}, \boldsymbol{u} \rangle| \le C_1 (s\eta^2)^{(p-1)/(2p)} \left(\sup_{\boldsymbol{u} \in \mathscr{D}} \sum_{i=1}^{m} |\langle \boldsymbol{r}_i, \boldsymbol{u} \rangle|^2 \right)^{1/(2p)} \int_0^{\infty} \sqrt{\log(\mathcal{N}(\mathscr{D}, \|\cdot\|_{X,q}, t))} \, dt \quad (4.8)$$

where C_1 is an absolute constant.

We now estimate the key integral above in two different ways, the first bound being better for small values of t, the second bound better for larger values of t.

Small values of t:: Following (4.6) and (4.7) and since $\|\boldsymbol{u}\|_2 = 1$ for all $\boldsymbol{u} \in \mathscr{D}$, we have that for any $\boldsymbol{u} \in \mathscr{D}$,

$$\|\boldsymbol{u}\|_{X,q} \le (s\eta^2)^{1/2} m^{1/(2q)}. \tag{4.9}$$

Standard covering arguments show that for any seminorm $\|\cdot\|$, one has for the unit ball $B_{\|\cdot\|} \subset \mathbb{C}^s$ that $\mathcal{N}(B_{\|\cdot\|}, \|\cdot\|, t) \le (1 + 1/t)^{2s}$. Thus for $\Lambda \subset \{1, \dots, N\}$ with $|\Lambda| = s'$, one can define the seminorm $\|\cdot\|_{\Lambda}$ on $\mathbb{C}^{s'}$ by

$$\|\boldsymbol{y}\|_{\Lambda} \stackrel{\text{def}}{=} \left(\sum_{i=1}^{m} |\langle \boldsymbol{D}_{\Lambda}^* \boldsymbol{r}_i^*, \boldsymbol{y} \rangle|^{2q} \right)^{1/(2q)},$$

and observe that for $\boldsymbol{u}$ with $\|\boldsymbol{u}\|_{X,q} \le 1$ and $\boldsymbol{u} = \boldsymbol{D}_{\Lambda} \boldsymbol{y}$ for $\boldsymbol{y} \in \mathbb{C}^{s'}$ that one has the equivalence $\|\boldsymbol{y}\|_{\Lambda} = \|\boldsymbol{u}\|_{X,q}$. Applying this equivalence for $\Lambda = \text{supp}(\boldsymbol{z})$ and $\boldsymbol{y} = \boldsymbol{z}_{\Lambda}$, we have that

$$\mathcal{N}(\mathscr{D}, \|\cdot\|_{X,q}, t) \le \sum_{\omega(\Lambda) \le s} \mathcal{N}(B_{\|\cdot\|_{\Lambda}}, m^{1/(2q)} \|\cdot\|_2 \sqrt{(s\eta^2)}, t)$$

$$\le \binom{N}{s} (1 + \sqrt{(s\eta^2)} m^{1/(2q)} t^{-1})^{2s}$$

$$\le (eN/s)^s (1 + \sqrt{(s\eta^2)} m^{1/(2q)} t^{-1})^{2s},$$

where we have applied [FR13, Proposition C.3] in the final line.

Large values of t:: To get a bound for larger values of t, we will utilize another embedding, and for that reason we define the auxiliary set

$$\mathscr{D}_{\text{aux}} \stackrel{\text{def}}{=} \{\boldsymbol{u} : \|\boldsymbol{D}^* \boldsymbol{u}\|_{\omega,1} \le 1, \boldsymbol{u} = \boldsymbol{D}\boldsymbol{z} \in \mathbb{C}^n, \|\boldsymbol{z}\|_{\omega,0} \le s\} = \bigcup_{\Lambda : \omega(\Lambda) \le s} \mathscr{B}_1^{\Lambda},$$

where $\mathscr{B}_1^{\Lambda} \stackrel{\text{def}}{=} \{\boldsymbol{u} : \|\boldsymbol{D}^* \boldsymbol{u}\|_{\omega,1} \le 1, \boldsymbol{u} = \boldsymbol{D}\boldsymbol{z} \in \mathbb{C}^n, \text{supp}(\boldsymbol{z}) = \Lambda\}$. Then we have the embedding $\mathscr{D} \subset \sqrt{(s\eta^2)} \mathscr{D}_{\text{aux}}$ since for any $\boldsymbol{u} \in \mathscr{D}$, $\|\boldsymbol{D}^* \boldsymbol{u}\|_{\omega,1} \le \sqrt{(s\eta^2)}$ by definition of the localization factor η.

We now use a variant of Maurey's lemma [Car85], precisely, the variant as stated in Lemma 5.3 in [RW13], in order to deduce a different covering number bound:

Lemma 4.1 (Maurey's lemma). *For a normed space X, consider a finite set $U \subset X$ of cardinality N, and assume that for every $L \in \mathbb{N}$ and $(\boldsymbol{u}_1, \dots, \boldsymbol{u}_L) \in U^L$, $\mathbb{E}_{\epsilon} \|\sum_{j=1}^{L} \epsilon_j \boldsymbol{u}_j\|_X \le A\sqrt{L}$, where ϵ denotes a Rademacher vector. Then for every $t > 0$,*

$$\log \mathcal{N}(\text{conv}(U), \|\cdot\|_X, t) \le c(A/t)^2 \log(N),$$

where $\text{conv}(U)$ is the convex hull of U.

As a corollary, we have the following.

Corollary 4.2. *For every $t > 0$,*

$$\log(\mathcal{N}(\mathcal{D}, \|\cdot\|_{X,q}, t)) \leq C\left(2e^{-1/2}\sqrt{2q}\,m^{1/(2q)}\sqrt{2(s\eta^2)}\,t^{-1}\right)^2 \log(N).$$

Proof of Corollary 4.2. Consider Lemma 4.1 with norm $\|\cdot\|_{X,q}$, parameter $A = 2e^{-1/2}\sqrt{2q}\,m^{1/(2q)}$, and $U = \{\pm\omega_j^{-1}\boldsymbol{D}^*(\boldsymbol{e}_j), \pm i\omega_j^{-1}\boldsymbol{D}^*(\boldsymbol{e}_j), j \in [N]\}$. First observe that

$$\mathcal{D} \subset \sqrt{(s\eta^2)}\mathcal{D}_{\text{aux}} \subset \sqrt{2(s\eta^2)}\text{conv}(U).$$

For a Bernoulli random vector $\epsilon = (\epsilon_1, \dots, \epsilon_L)$ and $\boldsymbol{u}_1, \dots, \boldsymbol{u}_L \in U$ we have

$$\begin{aligned}
\mathbb{E}_\epsilon\|\sum_{j=1}^L \epsilon_j \boldsymbol{u}_j\|_{X,q} &\leq \left(\mathbb{E}\|\sum_{j=1}^L \epsilon_j \boldsymbol{u}_j\|_{X,q}^{2q}\right)^{1/(2q)} \\
&= \left(\mathbb{E}\sum_{\ell=1}^m |\langle \boldsymbol{r}_\ell, \sum_{j=1}^L \epsilon_j \boldsymbol{u}_j\rangle|^{2q}\right)^{1/(2q)} \\
&= \left(\sum_{\ell=1}^m \mathbb{E}|\langle \boldsymbol{r}_\ell, \sum_{j=1}^L \epsilon_j \boldsymbol{u}_j\rangle|^{2q}\right)^{1/(2q)} \\
&\leq 2e^{-1/2}\sqrt{2q}\left(\sum_{\ell=1}^m \|(\langle \boldsymbol{r}_\ell, \boldsymbol{u}_j\rangle)_{j=1}^L\|_2^{2q}\right)^{1/(2q)}, \qquad (4.10)
\end{aligned}$$

where we applied Khintchine's inequality [Khi23] in the last step. Since each $\boldsymbol{u}_j$ consists of a multiple of a single column of $\boldsymbol{D}$, we also have for each j and i that $|\langle \boldsymbol{r}_i, \boldsymbol{u}_j\rangle| = |\langle \boldsymbol{D}^*\boldsymbol{r}_i^*, \omega_j^{-1}\boldsymbol{e}_j\rangle|$ for some coordinate vector $\boldsymbol{e}_j$ consisting of all zeros with a 1 in the jth position. Thus $|\langle \boldsymbol{r}_i, \boldsymbol{u}_j\rangle| \leq 1$ for each j and i, which means that $\|(\langle \boldsymbol{r}_i, \boldsymbol{u}_j\rangle)_{k=1}^L\|_2 \leq \sqrt{L}$ for any L and

$$\mathbb{E}_\epsilon\|\sum_{j=1}^L \epsilon_j \boldsymbol{u}_j\|_{X,q} \leq 2e^{-1/2}\sqrt{2q}\,m^{1/(2q)}\sqrt{L} = A\sqrt{L}.$$

The corollary then follows from Lemma 4.1. $\qquad\qquad\square$

We have thus obtained the two bounds for $t > 0$:

$$\sqrt{\log(\mathcal{N}(\mathcal{D}, \|\cdot\|_X, t))} \leq \sqrt{s\log(eN/s) + 2s\log\left(1 + 2\sqrt{(s\eta^2)}\,m^{1/(2q)}\,t^{-1}\right)}$$

$$\sqrt{\log(\mathcal{N}(\mathcal{D}, \|\cdot\|_X, t))} \leq C\sqrt{q}\,m^{1/(2q)}\sqrt{(s\eta^2)}\,t^{-1}\sqrt{\log(N)}.$$

We may now bound the integral in (4.4). Without loss, we take the upper integration bound as $t_0 = \sqrt{(s\eta^2)}\,m^{1/(2q)}$ because, for $t > t_0$, we have $\mathcal{N}(\mathcal{D}, \|\cdot\|_{X,q}, t) = 1$ by (4.9). Splitting the integral in two parts and using our covering number bounds, we have for $\alpha \in (0, t_0)$,

$$\int_0^\infty \sqrt{\log(\mathcal{N}(\mathcal{D}, \|\cdot\|_{X,q}, t))}\,dt$$

$$\leq \int_0^\alpha \sqrt{s\log(eN/s) + 2s\log\left(1 + 2\sqrt{(s\eta^2)}\,m^{1/(2q)}\,t^{-1}\right)}\,dt + C_1\sqrt{q}\,m^{1/(2q)}\sqrt{(s\eta^2)}\int_\alpha^{t_0} t^{-1}\,dt$$

$$\leq \alpha\sqrt{s\log(eN/s)} + \sqrt{2s}\alpha\sqrt{\log(e(1 + \sqrt{(s\eta^2)}\,m^{1/(2q)})\alpha^{-1}} + C_2\sqrt{q\,m^{1/q}(s\eta^2)\ln(4N)}\log(\sqrt{(s\eta^2)}\,m^{1/(2q)}/\alpha)$$

where in the last step we have utilized that for $a > 0$, $\int_0^a \sqrt{\ln(1 + t^{-1})}\, dt \le a\sqrt{\ln(e(1 + a^{-1}))}$ (see Lemma 10.3 of [Rau10]). Choosing $\alpha = m^{1/(2q)}$ yields

$$\int_0^\infty \sqrt{\log(\mathcal{N}(\mathcal{D}, \|\cdot\|_{X,q}, t))}\, dt \le C_3 \sqrt{q(s\eta^2)m^{1/q}\log(N)\log^2(s\eta^2)}.$$

Combining this with (4.1), (4.2) and (4.8), we have

$$\mathbb{E}\delta_s \le \frac{C_3(s\eta^2)^{(p-1)/(2p)}\sqrt{qm^{1/q}(s\eta^2)\log(N)\log^2(s\eta^2)}}{m} \mathbb{E}\sup_{x\in\mathcal{D}}\left(\sum_{i=1}^m |\langle r_i, x\rangle|^2\right)^{1/(2p)}$$

$$\le \frac{C_3(s\eta^2)^{1/2+(p-1)/(2p)}\sqrt{q\log(N)\log^2(s\eta^2)}}{m^{1-1/(2q)}m^{-1/(2p)}}\mathbb{E}\left(\frac{1}{m}\left\|\left\|\sum_{i=1}^m r_i^* r_i - \mathbf{I}_n\right\|\right\|_s + \||\mathbf{I}_n\||_s\right)^{1/(2p)}$$

$$\le \frac{C_3(s\eta^2)^{1/2+(p-1)/(2p)}\sqrt{q\log(N)\log^2(s\eta^2)}}{m^{1/2}}\sqrt{\mathbb{E}\delta_s + 1}.$$

Above we applied Hölder's inequality and used that $1/q + 1/p = 1$ as well as $p \ge 1$. We now choose $p = 1 + 1/\log(s\eta^2)$ and $q = 1 + \log(s\eta^2)$ to give $(s\eta^2)^{(p-1)/(2p)} \le (s\eta^2)^{(p-1)/2} = (s\eta^2)^{1/(2\log(s\eta^2))} = e^{1/2}$ and

$$\mathbb{E}\delta_s \le C_4\sqrt{2\log(N)\log^2(s\eta^2)/m}\sqrt{\mathbb{E}\delta_s + 1}.$$

Squaring this inequality and completing the square finally shows that

$$\mathbb{E}\delta_s \le C_5\sqrt{\frac{(s\eta^2)\log(N)\log^3(s\eta^2)}{m}}, \tag{4.11}$$

provided the term under the square root is at most 1. Then $\mathbb{E}\delta_s \le \delta/2$ for some $\delta \in (0,1)$ if

$$m \ge C_6\delta^{-2}(s\eta^2)\log^3(s\eta^2)\log(N). \tag{4.12}$$

It remains to show that δ_s does not deviate much from its expectation. For this *probability bound*, we may write δ_s as the supremum of an empirical process as in [[Rau10], Theorem 6.25] and apply the following Bernstein inequality for the supremum of an empirical process:

Theorem 4.3 ([Bou03], Theorem 6.25 of [Rau10]). *Let $\mathscr{F}$ be a countable set of functions $f : \mathbb{C}^n \to \mathbb{R}$. Let $Y_1, \ldots, Y_m$ be independent copies of a random vector Y on $\mathbb{C}^n$ such that $\mathbb{E}f(Y) = 0$ for all $f \in \mathscr{F}$, and assume $f(Y) \le 1$ almost surely. Let Z be the random variable $Z = \sup_{f\in\mathscr{F}} \sum_{\ell=1}^m f(Y_\ell)$, and $\mathbb{E}Z$ its expectation. Let $\sigma^2 > 0$ such that $\mathbb{E}[f(Y)^2] \le \sigma^2$ for all $f \in \mathscr{F}$. Then, for all $t > 0$,*

$$\mathbb{P}(Z \ge \mathbb{E}Z + t) \le \exp\left(-\frac{t^2}{2(m\sigma^2 + 2\mathbb{E}Z) + 2t/3}\right) \tag{4.13}$$

We apply this theorem to provide a probability bound. Let $f_{z,w}(r) = \mathrm{Re}(\langle (r_i^* r_i - \mathbf{I})z, w\rangle)$ so that

$$m\delta_s = \left\|\left\|\sum_{i=1}^m (r_i^* r_i - \mathbb{E}r_i^* r_i)\right\|\right\|_s = \sup_{(z,w)\in\mathcal{D},\, i=1}\sum_{i=1}^m f_{z,w}(r_i).$$

Since $\mathbb{E}r_i^* r_i = \mathbf{I}$ we have $\mathbb{E}f_{z,w}(r) = 0$. Moreover, $|f_{z,w}(r)| \le \max_{z\in\mathcal{D}} |\langle r_i, z\rangle|^2 + 1 \le s\eta^2 + 1$. For the variance term, we have $\mathbb{E}|f_{z,w}(r_i)|^2 = \mathbb{E}\|(r_i^* r_i - \mathbf{I})z\|_2^2 \le (s\eta^2 + 1)^2$.

Fix $\delta \in (0,1)$, and suppose the number of measurements m satisfies (4.12) so that $\mathbb{E}\delta_s \leq \delta/2$. Then it follows

$$
\begin{aligned}
\mathbb{P}(\delta_s \geq \delta) \quad &\leq \quad \mathbb{P}(\delta_s \geq \mathbb{E}\delta_s + \delta/9) \\
&= \quad \mathbb{P}\left(\left\| \sum_{i=1}^{m} \left(r_i^* r_i - \mathbb{E}r_i^* r_i \right) \right\|_s \geq \mathbb{E}\left\| \sum_{i=1}^{m} \left(r_i^* r_i - \mathbb{E}r_i^* r_i \right) \right\|_s + \delta m/9 \right) \\
&= \quad \mathbb{P}\left(\frac{1}{s\eta^2+1}\left\| \sum_{i=1}^{m} \left(r_i^* r_i - \mathbb{E}r_i^* r_i \right) \right\|_s \geq \frac{1}{s\eta^2+1}\mathbb{E}\left\| \sum_{i=1}^{m} \left(r_i^* r_i - \mathbb{E}r_i^* r_i \right) \right\|_s + \frac{\delta m/9}{s\eta^2+1} \right) \\
&\leq \quad \exp\left(-\frac{(\frac{\delta m/9}{s\eta^2+1})^2}{2m(1+\frac{\delta}{s\eta^2+1}) + \frac{2}{3}(\frac{\delta m/9}{s\eta^2+1})} \right) \\
&\leq \quad \exp\left(-\frac{\delta^2 m}{C_7(s\eta^2)} \right). \qquad\qquad (4.14)
\end{aligned}
$$

The last term is bounded by $\gamma \in (0,1)$ if $m \geq C\delta^{-2}(s\eta^2)\log(1/\gamma)$. Together, we have $\delta_s \leq \delta$ with probability at least $1 - \gamma$ if

$$
m \geq C_8 \delta^{-2} s\eta^2 \max\left\{ \log^3(s\eta^2)\log(N), \log(1/\gamma) \right\}.
$$

This completes the proof.   □

5. CONCLUSION

We have introduced a coherence-based analysis of compressive sensing when the signal to be recovered is approximately sparse in a redundant dictionary. Whereas previous theory only allowed for unstructured random sensing measurements, our coherence-based analysis extends to structured sensing measurements such as subsampled uniformly bounded bases, bringing the theory closer to the setting of practical applications. We also extend the theory of variable density sampling to the dictionary setting, permitting some coherence between sensing measurements and sparsity dictionary. We further extend the analysis to allow for weighted sparse expansions. Still, several open questions remain. While we provided two concrete examples of dictionaries satisfying the bounded localization factor condition required by our analysis – the oversampled DFT frame and redundant Haar wavelet frame – these bounds can almost certainly be extended to more general classes of dictionaries, and improved considerably in the case of the oversampled DFT frame. We have also left several open problems related to the full analysis for variable density sampling in this setting, including the removal of a weighted noise assumption in the ℓ_1-analysis reconstruction method. Finally, we believe that the D-RIP assumption used throughout our analysis can be relaxed, and that a RIPless analysis [CP12] should be possible and permit non-uniform signal recovery bounds at a reduced number of measurements. It would also be useful to extend our results to measurement matrices constructed in a deterministic fashion for those applications in which randomness is not admissible; of course this is a challenge even in the classical setting [DeV07].

ACKNOWLEDGMENTS

REFERENCES

[AC09]   N. Ailon and B. Chazelle. The fast Johnson-Lindenstrauss transform and approximate nearest neighbors. *SIAM J. Comput.*, 39:302–322, 2009.

[Ach03]   D. Achlioptas. Database-friendly random projections: Johnson–lindenstrauss with binary coins. *J.Comp.Sys.Sci.*, 66(4):671–687, 2003. Special issue of invited papers from PODS'01.

[AHPR13]   B. Adcock, A. C. Hansen, C. Poon, and B. Roman. Breaking the coherence barrier: A new theory for compressed sensing. Submitted, 2013.

[BDDW08]   R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constr. Approx.*, 28(3):253–263, 2008.

[BDWZ12]   N. Burq, S. Dyatlov, R. Ward, and M. Zworski. Weighted eigenfunction estimates with applications to compressed sensing. *SIAM Journal on Mathematical Analysis*, 44(5):3481–3501, 2012.

[Bou03]   O. Bousquet. *Concentration inequalities for sub-additive functions using the entropy method*, pages 213–247. Stochastic Inequalities and Applications, Progr. Probable. 56. Birkauser, Basel, 2003.

[BS07]   R. Baraniuk and P. Steeghs. Compressive radar imaging. In *Proc. IEEE Radar Conf*, pages 128–133. IEEE, 2007.

[BSO08]   J. Bobin, J.-L. Starck, and R. Ottensamer. Compressed sensing in astronomy. *IEEE J. Sel. Top. Signa.*, 2(5):718–726, 2008.

[Car85]   B. Carl. Inequalities of Bernstein-Jackson-type and the degree of compactness of operators in Banach spaces. *Ann. Inst. Fourier (Grenoble)*, 35(3):79–118, 1985.

[CCS08]   J.-F. Cai, R. H. Chan, and Z. Shen. A framelet-based image inpainting algorithm. *Appl. Comput. Harmon. A.*, 24(2):131–149, 2008.

[CD04]   E. J. Candès and D. L. Donoho. New tight frames of curvelets and optimal representations of objects with piecewise c^2 singularities. *Commun. Pur. Appl. Math.*, 57(2):219–266, 2004.

[CENR10]   E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Appl. Comput. Harmon. A.*, 31(1):59–73, 2010.

[CFG13]   E. J. Candès and C. Fernandez-Granda. Super-resolution from noisy data. *J. Fourier Anal. Appl.*, 19(6):1229–1254, 2013.

[CFG14]   E. J. Candès and C. Fernandez-Granda. Towards a mathematical theory of super-resolution. *Commun. Pur. Appl. Math.*, 67(6):906–956, 2014.

[CFM+11]   P. Casazza, M. Fickus, D. Mixon, Y. Wang, and Z. Zhou. Constructing tight fusion frames. *Appl. Comput. Harmon. A.*, 30(2):175–187, 2011.

[CGV13]   M. Cheraghchi, V. Guruswami, and A. Velingker. Restricted isometry of Fourier matrices and list decodability of random linear codes. *SIAM J. Comput.*, 42(5):1888–1914, 2013.

[CHKK11]   P. Casazza, A. Heinecke, F. Krahmer, and G. Kutyniok. Optimally sparse frames. *IEEE T. Inform. Theory*, 57(11):7279–7287, 2011.

[CP12]   E. J. Candès and Y. Plan. A probabilistic and ripless theory of compressed sensing. *IEEE T. Inform. Theory*, 2012. to appear.

[CR07]   E. J. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *IEEE T. Inform. Theory*, 23(3):969, 2007.

[CT05a]   E. J. Candès and T. Tao. Decoding by linear programming. *IEEE T. Inform. Theory*, 51:4203–4215, 2005.

[CT05b]   E. J. Candès and T. Tao. Near-optimal signal recovery from random projections: Universal encoding strategies. *IEEE T. Inform. Theory*, 52(12):5406–5425, Nov. 2005.

[DeV07]   R. A. DeVore. Deterministic constructions of compressed sensing matrices, 2007. Unpublished manuscript.

[DG03]   S. Dasgupta and A. Gupta. An elementary proof of a theorem of Johnson and Lindenstrauss. *IEEE T. Inform. Theory*, 22:60–65, 2003.

[DNW12]   M. Davenport, D. Needell, and M. B. Wakin. Signal space CoSaMP for sparse recovery with redundant dictionaries. *IEEE T. Inform. Theory*, 59(10):6820, 2012.

[Don06]   D. L. Donoho. Compressed sensing. *IEEE T. Inform. Theory*, 52(4):1289–1306, Apr. 2006.

[Dud67]   R. M. Dudley. The sizes of compact subsets of Hilbert space and continuity of gaussian processes. *J. Funct. Anal.*, 1(3):290–330, 1967.

[Dut89]   P. Dutilleux. An implementation of the "algorithme à trous" to compute the wavelet transform, 1989. in *Wavelets: Time-Frequency Methods and Phase-Space J.M. Combes, A. Grossmann, and P. Tchamitchian, Eds. New York: Springer*.

[ELL08]   G. Easley, D. Labate, and W.-Q. Lim. Sparse directional image representations using the discrete shearlet transform. *Appl. Comput. Harmon. A.*, 25(1):25–46, 2008.

[EMR07]   M. Elad, P. Milanfar, and R. Rubinstein. Analysis versus synthesis in signal priors. *Inverse Probl.*, 23(3):947–968, 2007.

[Fou15]   S. Foucart. Dictionary-sparse recovery via thresholding-based algorithms. *J. Fourier Anal. Appl.*, pages 1–14, 2015.

[FR13]   S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Appl. Num. Harm. Anal., Birkhauser, 2013.

[Gir14]   R. Giryes. Sampling in the analysis transform domain. *arXiv preprint arXiv:1410.6558*, 2014.

[GN13]   R. Giryes and D. Needell. Greedy signal space methods for incoherence and beyond. *Appl. Comput. Harmon. A.*, 2013. To appear.

[GN14]   R. Giryes and D. Needell. Near oracle performance and block analysis of signal space greedy methods. Submitted, 2014.

[GNE+14] R. Giryes, S. Nam, M. Elad, R. Gribonval, and M. E. Davies. Greedy-like algorithms for the cosparse analysis model. *Linear Algebra Appl.*, 441:22–60, 2014.

[GPV15]   R. Giryes, Y. Plan, and R. Vershynin. On the effective measure of dimension in analysis cosparse models, 2015. in Proc. 11th Int. Conf. SampTA.

[HD15]   J. Hampton and A. Doostan. Compressive sampling of polynomial chaos expansions: convergence analysis and sampling strategies. *Journal of Computational Physics*, 280:363–386, 2015.

[HS09]   M. A. Herman and T. Strohmer. High-resolution radar via compressed sensing. *IEEE T. Signal Proces.*, 57(6):2275–2284, Jun. 2009.

[HV11]   A. Hinrichs and J. Vybiral. Johnson-Lindenstrauss lemma for circulant matrices. *Random Structures and algorithms*, 39:391–398, 2011.

[Khi23]   A. Khintchine. Über dyadische brüche. *Math. Z.*, 18(1):109–116, 1923.

[KL07]   G. Kutyniok and D. Labate. Construction of regular and irregular shear lets. *J. Wavelet Theory and Appl.*, 1:1–10, 2007.

[KMR14]   F. Krahmer, S. Mendelson, and H. Rauhut. Suprema of chaos processes and the restricted isometry property. *Comm. Pure Appl. Math.*, 67(11):1877–1904, 2014.

[KW11]   F. Krahmer and R. Ward. New and improved Johnson Lindenstrauss embeddings via the restricted isometry property. *SIAM J. Math. Anal.*, 43, 2011.

[KW14]   F. Krahmer and R. Ward. Stable and robust sampling strategies for compressive imaging. *IEEE T. Image Process.*, 23:612–622, 2014.

[LDP07]   M. Lustig, D. Donoho, and J. M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnet. Reson. Med.*, 58(6):1182–1195, 2007.

[LLKW05] D. Labate, W.-Q. Lim, G. Kutyniok, and G. Weiss. Sparse multidimensional representation using shearlets. In *Optics & Photonics 2005*, pages 59140U–59140U. International Society for Optics and Photonics, 2005.

[LML12]   Y. Liu, T. Mi, and S. Li. Compressed sensing with general frames via optimal-dual-based-analysis. *IEEE T. Inform. Theory*, 58(7):4201–4214, 2012.

[Mal99]   S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, London, 2nd edition, 1999.

[NDEG13] S. Nam, M. E. Davies, M. Elad, and R. Gribonval. The cosparse analysis model and algorithms. *Appl. Comp. Harm. Anal.*, 34(1):30–56, 2013.

[NW13a]   D. Needell and R. Ward. Near-optimal compressed sensing guarantees for total variation minimization. *IEEE T. Image Process.*, 22(10):3941–3949, 2013.

[NW13b]   D. Needell and R. Ward. Stable image reconstruction using total variation minimization. *SIAM J. Imaging Sciences*, 6(2):1035–1058, 2013.

[PE13]   T. Peleg and M. Elad. Performance guarantees of the thresholding algorithm for the cosparse analysis model. *IEEE T. Inform. Theory*, 59(3):1832–1845, 2013.

[Poo15a] C. Poon. On the role of total variation in compressed sensing. *SIAM J. Imaging Science*, 8(1):682–720, 2015.

[Poo15b] C. Poon. Structure dependent sampling in compressed sensing: theoretical guarantees for tight frames. *Arxiv preprint*, 2015.

[Rau10]   H. Rauhut. Compressive sensing and structured random matrices. In M. Fornasier, editor, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, volume 9 of *Radon Series Comp. Appl. Math.*, pages 1–92. deGruyter, 2010. Corrected Version (April 4, 2011): Noncommutative Khintchine inequality for Rademacher chaos (Theorem 6.22) corrected, Proof of Theorem 9.3 corrected. Proof of Lemma 8.2 corrected.

[RK13]   H. Rauhut and M. Kabanava. Analysis ℓ_1-recovery with frames and Gaussian measurements. *arXiv preprint arXiv:1306.1356*, 2013.

[RRT12]   H. Rauhut, J. Romberg, and J. A. Tropp. Restricted isometries for partial random circulant matrices. *Appl. Comput. Harmon. A.*, 32(2):242–254, 2012.

[RSV08]   H. Rauhut, K. Schnass, and P. Vandergheynst. Compressed sensing and redundant dictionaries. *IEEE T. Inform. Theory*, 54(5):2210–2219, 2008.

[RV08]   M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61:1025–1045, 2008.

[RW11]   H. Rauhut and R. Ward. Sparse recovery for spherical harmonic expansions. *arXiv preprint arXiv:1102.4097*, 2011.

[RW12]   H. Rauhut and R. Ward. Sparse Legendre expansions via ℓ_1-minimization. *J. Approx. Theory*, 164:517=533, 2012.

[RW13]   H. Rauhut and R. Ward. Interpolation via weighted ℓ_1 minimization. *arXiv:1308.0759*, 2013.

[SB12]   P. Stoica and P. Babu. Sparse estimation of spectral lines: Grid selection problems and their solutions. *IEEE T. Signal Proces.*, 60(2):962–967, 2012.

[SED04]   J. L. Starck, M. Elad, and D. L. Donoho. Redundant multiscale transforms and their application for morphological component analysis. *Adv. Imag. Elect. Phys.*, 132, 2004.

[SFM07]   J. L. Starck, J. Fadili, and F. Murtagh. The undecimated wavelet decomposition and its reconstruction. *IEEE T.Sig.Proc.*, 16(2):297–309, 2007.

[TBSR12]   G. Tang, B. N. Bhaskar, P. Shah, and B. Recht. Compressive sensing off the grid. In *Allerton Conf. Communication, Control, and Computing*, pages 778–785. IEEE, 2012.

[VW04]   R. Vale and S. Waldron. Tight frames and their symmetries. *Constr. Approx.*, 21(1):83–112, 2004.