Spectral Compressive Sensing[☆]

Marco F. Duarte[a], Richard G. Baraniuk[b]

[a]*Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA 01003*
[b]*Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005*

Abstract

Compressive sensing (CS) is a new approach to simultaneous sensing and compression of sparse and compressible signals based on randomized dimensionality reduction. To recover a signal from its compressive measurements, standard CS algorithms seek the sparsest signal in some discrete basis or frame that agrees with the measurements. A great many applications feature smooth or modulated signals that are *frequency-sparse* and can be modeled as a superposition of a small number of sinusoids; for such signals, the discrete Fourier transform (DFT) basis is a natural choice for CS recovery. Unfortunately, such signals are only sparse in the DFT domain when the sinusoid frequencies live precisely at the centers of the DFT bins; when this is not the case, CS recovery performance degrades significantly. In this paper, we introduce the spectral CS (SCS) recovery framework for arbitrary frequency-sparse signals. The key ingredients are an over-sampled DFT frame and a restricted union-of-subspaces signal model that inhibits closely spaced sinusoids. We demonstrate that SCS significantly outperforms current state-of-the-art CS algorithms based on the DFT while providing provable bounds on the number of measurements required for stable recovery. We also leverage line spectral estimation methods (specifically Thomson's multitaper method and MUSIC) to further improve the performance of SCS recovery.

Keywords: compressive sensing, spectral estimation, redundant frames, structured sparsity

1. Introduction

The emerging theory of *compressive sensing* (CS) [1, 2, 3] combines digital data acquisition with digital data compression to enable a new generation of signal acquisition systems that operate at a signal's intrinsic information rate rather than its ambient data rate. Rather

[☆]MFD and RGB were supported by grants NSF CCF-0431150 and CCF-0728867, DARPA/ONR N66001-08-1-2065, ONR N00014-07-1-0936 and N00014-08-1-1112, AFOSR FA9550-07-1-0301 and FA9550-09-1-0432, ARO MURIs W911NF-07-1-0185 and W911NF-09-1-0383, and the Texas Instruments Leadership Program. MFD was also supported by NSF Supplemental Funding DMS-0439872 to UCLA-IPAM, P.I. R. Caflisch.

Email addresses: mduarte@ecs.umass.edu (Marco F. Duarte), richb@rice.edu (Richard G. Baraniuk)

URL: http://www.ecs.umass.edu/~mduarte (Marco F. Duarte), http://dsp.rice.edu/~richb (Richard G. Baraniuk)

than acquiring N samples $\mathbf{x} = [\mathbf{x}[1]\ \mathbf{x}[2]\ \ldots\ \mathbf{x}[N]]^T$ of a signal, a CS system acquires $M < N$ measurements via the linear dimensionality reduction $\mathbf{y} = \Phi\mathbf{x}$, where Φ is an $M \times N$ measurement matrix. When the signal $\mathbf{x}$ has a *sparse* representation $\mathbf{x} = \Psi\theta$ in terms of an $N \times N$ orthonormal basis matrix Ψ, meaning that only $K \ll N$ out of N signal coefficients θ are nonzero, then the number of measurements required to ensure that $\mathbf{y}$ retains all of the information in $\mathbf{x}$ is just $M = O(K \log(N/K))$ [1, 2, 3]. Moreover, a sparse signal $\mathbf{x}$ can be recovered from its compressive measurements $\mathbf{y}$ via a convex optimization or iterative greedy algorithm. Random matrices play a central role as universal measurements, since they are suitable for signals sparse in any fixed basis with high probability. The theory also extends to noisy measurements as well as to so-called *compressible* signals that are not exactly sparse but can be closely approximated as such. Compressible signals have coefficients θ that, when sorted, decay according to a power law $|\theta[i]| < Ci^{-1/p}$ for some $p \leq 1$; the smaller the decay exponent p, the faster the decay and the better the recovery performance we can expect from CS. The theory can also be extended from orthonormal bases Ψ to more general redundant frames, where we instead require that either the vector of synthesis coefficients θ or the vector of analysis coefficients $\Psi^H\mathbf{x}$ be sparse or compressible [4].

A great many applications feature smooth or modulated signals that can be modeled as a superposition of K sinusoids [5, 6, 7, 8]:

$$\mathbf{x}[n] = \sum_{k=1}^{K} a_k\, e^{-j\omega_k n}, \tag{1}$$

where ω_k are the sinusoid frequencies. When the sinusoids are of infinite extent, such signals have a K-sparse representation in terms of the discrete-time Fourier transform (DTFT),[1] since

$$X(\omega) = \sum_{k=1}^{K} a_k\, \delta(\omega - \omega_k), \tag{2}$$

where δ is the Dirac delta function. We will refer to such signals as *frequency-sparse*.

Practical applications feature signals of finite length N. In this case, the frequency domain tool of choice for both signal analysis and CS recovery has been the discrete Fourier transform (DFT).[2] The DFT $\mathbf{X}[l]$ of N consecutive samples from the signal model (1) can be obtained from the DTFT (2) by first convolving with a Dirichlet kernel and then sampling:

$$\mathbf{X}[l] = \sum_{k=1}^{K} a_k\, D_N\left(\frac{2\pi(l - l_k)}{N}\right), \tag{3}$$

[1]Recall that the DTFT of a signal $\mathbf{x}$ is defined as $X(\omega) = \sum_{n=-\infty}^{\infty} \mathbf{x}[n]e^{-j\omega n}$, with inverse transformation $\mathbf{x}[n] = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(\omega)e^{j\omega n}d\omega$.

[2]Recall that the DFT of a length-N signal $\mathbf{x}$ is defined as $\mathbf{X}[l] = \sum_{n=1}^{N} \mathbf{x}[n]e^{-j2\pi ln/N}$, $1 \leq l \leq N$, with inverse transformation $\mathbf{x}[n] = \frac{1}{N}\sum_{l=1}^{N} \mathbf{X}[l]e^{j2\pi ln/N}$, $1 \leq n \leq N$.

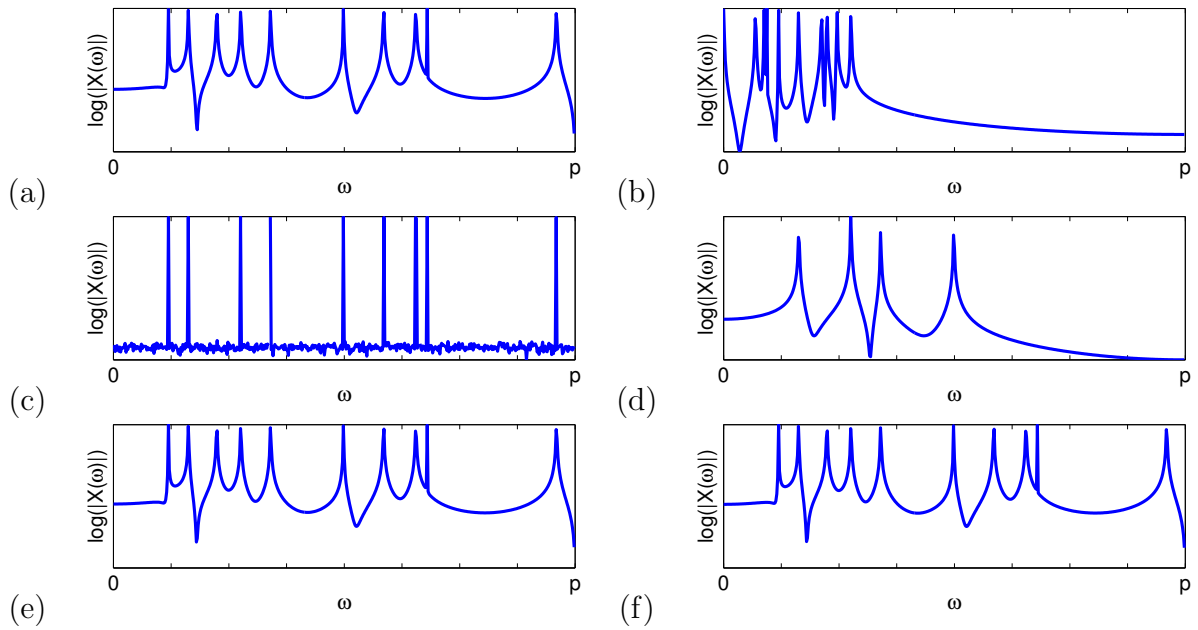

Figure 1: *Compressive sensing (CS) sparse signal recovery from* $M = 300$ *noiseless random measurements of a signal of length* $N = 1024$ *composed of* $K = 20$ *complex-valued sinusoids with arbitrary real-valued frequencies. We compare the frequency spectra obtained from redundant periodograms of (a) the original signal and its recovery using (b) root MUSIC on* M *signal samples (SNR =* $0.65dB$*), (c) standard CS using the orthonormal DFT basis (SNR =* $5.35dB$*), (d) standard CS using a* $10\times$ *redundant DFT frame (SNR =* $-4.45dB$*), (e) spectral CS using SIHT via Algorithm 1 (SNR =* $32.40dB$*), and (f) spectral CS using SIHT via Algorithm 3 (SNR =* $32.03dB$*).*

where $l_k = \frac{N\omega_k}{2\pi}$ and the Dirichlet kernel

$$D_N(x) := \sum_{k=1}^{N} e^{jkx} = e^{jx(N+1)/2} \frac{\sin(Nx/2)}{\sin(x/2)}.$$

Unfortunately, the DFT coefficients in (3) do not share the same sparsity property as the DTFT coefficients in (2), except in the (contrived) case where the sinusoid frequencies in (1) are *integral*, that is, when each and every l_k is equal to an integer. On closer inspection, we see that not only are most signals of the form (3) *not sparse* in the DFT domain, but, owing to the slow asymptotic decay of the Dirichlet kernel away from its peak, they are *just barely compressible*, with a decay exponent of $p = 1$. As a result, practical CS acquisition and recovery of frequency-sparse signals does not perform nearly as well as one might expect (see Fig. 1(c) and the discussions in [8, 9, 10, 11], for example).

The goal of this paper is to develop new recovery algorithms for the standard CS framework (as described in Section 2 and developed in [1, 2, 3]) for general frequency-sparse signals with non-integral frequencies. The naïve first step is to change the signal representation to a zero-padded DFT, which provides samples from the signal's DTFT at a higher rate than the standard DFT. This is equivalent to replacing the DFT basis with a redundant frame [12] of sinusoids that we will call a *DFT frame*. Unfortunately, as we quantify in Section 2,

there exists a tradeoff in the use of these redundant frames for sparse approximation and CS recovery: if we increase the amount of zero-padding / size of the frame, then signals with non-integral frequency components become more compressible, which improves recovery performance. However, simultaneously, the frame becomes increasingly *coherent* [13, 14], which decreases recovery performance (see Fig. 1(d), for example). In order to optimize this tradeoff, we will leverage recent progress on model-based CS [15] (see Section 2 for a summary of these areas) and marry these techniques with a class of greedy CS recovery algorithms. We refer to our general approach as *spectral compressive sensing* (SCS) and describe it in detail in Section 4.

A key novelty of SCS is the concept of taming the coherence of the redundant DFT frame using an inhibition model that ensures the sinusoid frequencies ω_k of (1) are not too closely spaced. Such an assumption is pervasive in the spectrum estimation literature [6, 7, 16]. We provide an analytical characterization of the number of measurements M required for stable SCS signal recovery under the model-based CS approach and study the performance of the framework under parameter variations. As we see from Fig. 1(e-f) and Fig. 4, the performance improvement of SCS over standard DFT-based CS can be substantial (up to 25dB in Fig. 4).

While the model-based SCS recovery algorithm is derived using a periodogram spectral estimate, we also show that more general line spectral estimation methods [5, 6, 7, 16, 17, 18] (described in Section 3) can be integrated into SCS in a straightforward fashion. The resulting recovery algorithms feature reduced computational complexity and increased robustness and noise stability, mirroring the advantages of line spectral estimation over the periodogram. In particular, we showcase two approaches for spectral estimation — Thomson's multitaper method [16] and root MUSIC [19] — and experimentally verify the resulting improvements in SCS recovery performance in Section 5.

Although this paper focuses on frequency-sparse signals, the SCS concept generalizes to other settings featuring signals that are sparse in a parameterized redundant frame, as discussed in Section 6. Examples include the frames underlying localization problems [20, 21, 22, 23], radar imaging [24, 25, 26], and manifold-based signal models [27, 28] to name just a few. Additionally, several alternative frameworks to CS have been proposed for the modeling and acquisition of frequency-sparse signals from a small number of samples, including finite rate of innovation (FROI) sampling [29, 30, 31, 32] and Xampling [9, 10, 33]. We compare and contrast these frameworks with ours in more detail in Section 6.

2. Background

2.1. Sparse approximation

A signal $\mathbf{x} \in \mathbb{R}^N$ is *K-sparse* ($K \ll N$) in a basis or frame[3] Ψ if there exists a vector θ with $\|\theta\|_0 = K$ such that $\mathbf{x} = \Psi\theta$. Here $\|\cdot\|_0$ denotes the ℓ_0 pseudo-norm, which simply

[3] A discrete-time frame is a matrix $\Psi \in \mathbb{C}^{D\times N}$, $D < N$, such that for all vectors $\mathbf{x} \in \mathbb{R}^D$, $A\|\mathbf{x}\|_2 \leq \|\Psi^H\mathbf{x}\|_2 \leq B\|\mathbf{x}\|_2$ with $0 < A \leq B < \infty$. A frame is a generalization of the concept of a basis to sets of possibly linearly dependent vectors [12].

counts the number of nonzero entries in the vector.

Transform coding is a powerful and hence popular compression approach. In transform coding, there exists a basis or frame Ψ in which the signal of interest $\mathbf{x}$ has a *K-sparse approximation* $\mathbf{x}_K$ in Ψ that yields small approximation error $\|\mathbf{x}-\mathbf{x}_K\|_2$. When Ψ is a basis, the optimal K-sparse approximation of $\mathbf{x}$ in Ψ is trivially found through hard thresholding: we preserve only the entries of θ with the K largest magnitudes and set all other entries to zero. While thresholding is suboptimal when Ψ is a frame, there exist a bevy of *sparse approximation algorithms* that aim to find a good sparse approximation to the signal of interest. Such algorithms include basis pursuit [34] and orthogonal matching pursuit (OMP) [13]. Their approximation performance is directly tied to the *coherence* of the frame Ψ, defined as

$$\mu(\Psi) := \max_{1\leq i,j\leq N} |\langle \psi_i, \psi_j \rangle| ,$$

where ψ_i denotes the i^{th} column of Ψ assumed to have unit norm. For example, orthogonal matching pursuit (OMP) successfully obtains a K-sparse signal representation if [13, 14]

$$\mu(\Psi) \leq \frac{1}{16(K-1)}. \tag{4}$$

2.2. Compressive sensing

Compressive Sensing (CS) is an efficient acquisition framework for signals that are sparse or compressible in a basis or frame Ψ. In this paper, we focus on the development set out in [1, 2, 3], where the signal $\mathbf{x}$ and its representation θ are discrete and finite-dimensional. This framework has successfully been reduced to several practical sensing architectures [8, 35, 36, 37]. To acquire the signal $\mathbf{x}$, we measure inner products of the signal against a set of measurement vectors $\{\phi_1, \ldots, \phi_M\}$; when $M < N$, we effectively compress the signal. By collecting the measurement vectors as rows of a measurement matrix $\Phi \in \mathbb{R}^{M\times N}$, this procedure can be written as $\mathbf{y} = \Phi\mathbf{x} = \Phi\Psi\theta$, with the vector $\mathbf{y} \in \mathbb{R}^M$ containing the CS measurements. We then aim to recover the signal $\mathbf{x}$ from the fewest possible measurements $\mathbf{y}$. Since $\Phi\Psi$ is a dimensionality reduction, it has a null space, and so infinitely many vectors $\mathbf{x}'$ yield the same recorded measurements $\mathbf{y}$. Fortunately, standard sparse approximation algorithms can be employed to recover the signal representation θ by finding a sparse approximation of $\mathbf{y}$ using the frame $\Upsilon = \Phi\Psi$.

Two parallel mathematical frameworks have emerged for the selection of a CS matrix Φ. One can choose to select the matrix Φ by selecting independently at random elements of an orthonormal basis Φ' that is *mutually incoherent* with the basis Ψ [38],[4] i.e., that the maximal magnitude for an inner product between an element of Φ' and an element of Ψ is close to the lower bound of $1/\sqrt{N}$. It is possible to show that as few as $M = \mathcal{O}(K \log N)$ measurements under such a sampling scheme can provide enough information to recover the overwhelming majority of sufficiently sparse signals [38]. For frequency-sparse signals, this measurement

[4]This concept of mutually incoherent bases should not be confused with the prior concept of the coherence of a frame, although they are conceptually related.

selection technique results in a random sampling acquisition scheme [39, 40, 41, 42, 43, 44]. Alternatively, the Restricted Isometry Property (RIP) has been proposed as a measure for the fitness of a matrix Υ for CS [1].

Definition 1. *The* K-restricted isometry constant *for the matrix* Υ*, denoted by* δ_K*, is the smallest nonnegative number such that, for all* $\theta \in \mathbb{C}^N$ *with* $\|\theta\|_0 = K$*,*

$$(1 - \delta_K)\|\theta\|_2^2 \leq \|\Upsilon\theta\|_2^2 \leq (1 + \delta_K)\|\theta\|_2^2. \tag{5}$$

A matrix has the RIP if $\delta_K < 1$. Since calculating δ_K for a given matrix requires a combinatorial amount of computation, random matrices have been advocated [1, 2]. For example, a matrix of size $M \times N$ with independent and identically distributed (i.i.d.) Gaussian entries with variance $1/M$ will have the RIP with very high probability if $K \leq M/\log(N/M)$. The same is true of matrices following Rademacher (± 1) or more general subgaussian distributions. Revisiting our previous example, OMP can recover a K-sparse representation θ from its measurements $\mathbf{y} = \Upsilon\theta$ if the restricted isometry constant $\delta_{K+1} < \frac{1}{3\sqrt{K}}$ [45]. Additional algorithms for signal recovery from CS measurements include CoSaMP [46] and iterative hard thresholding (IHT) [47, 48, 49, 50]. The IHT algorithm can be compactly written in an iterative form:

$$\widehat{\theta}_{i+1} = \text{thresh}_K(\widehat{\theta}_i + \Upsilon^H(\mathbf{y} - \Upsilon\theta)), \tag{6}$$

where the algorithm is initialized to $\widehat{\theta}_0 = 0$, and $\text{thresh}_K(\mathbf{x})$ denotes the hard thresholding operator on $\mathbf{x}$, setting all but the K entries of $\mathbf{x}$ with largest magnitudes to zero. The IHT algorithm can be shown to perfectly recover K-sparse signals when $\delta_{3K} \leq 1/\sqrt{32}$; it also offers performance guarantees in the presence of noise and compressibility.

2.3. Frequency-sparse signals

Recall from the introduction that an infinite-length frequency-sparse signal of the form (1) has a sparse DTFT (2). Unfortunately, however, an N-sample window of such a signal does not necessarily have a sparse DFT. Indeed, the DFT coefficients will be sparse only when the sinusoids in (1) have *integral frequencies* of the form $2\pi n/N$, where n is an integer. Otherwise, the situation is decidedly more complicated due to the spectral leakage induced by the windowing (i.e., convolution by the Dirichlet kernel). To graphically illustrate this issue, Fig. 2(a) plots the average approximation error of signals of length $N = 1024$ containing 20 complex sinusoids of both integral and non-integral frequencies when they are approximated using their best K-term approximation in the DFT basis. As expected, sparse approximation using the DFT basis fails miserably for signals with non-integral frequencies.

The naïve way to combat the spectral leakage caused by nonintegral frequencies is to employ a redundant *DFT frame*. The DFT frame representation provides a finer sampling of the DTFT coefficients for the signal $\mathbf{x}$ observed; it can also be interpreted as an interpolated version of coefficients of the DFT basis. Let $c \in \mathbb{N}$ denote the frequency redundancy factor for the DFT frame, and define the frequency sampling interval $\Delta = 2\pi/cN \in (0, 2\pi/N]$. Let

$$\mathbf{e}(\omega) := \frac{1}{\sqrt{N}}[1 \; e^{j\omega} \; e^{j2\omega} \; \ldots \; e^{j\omega(N-1)}]^T$$

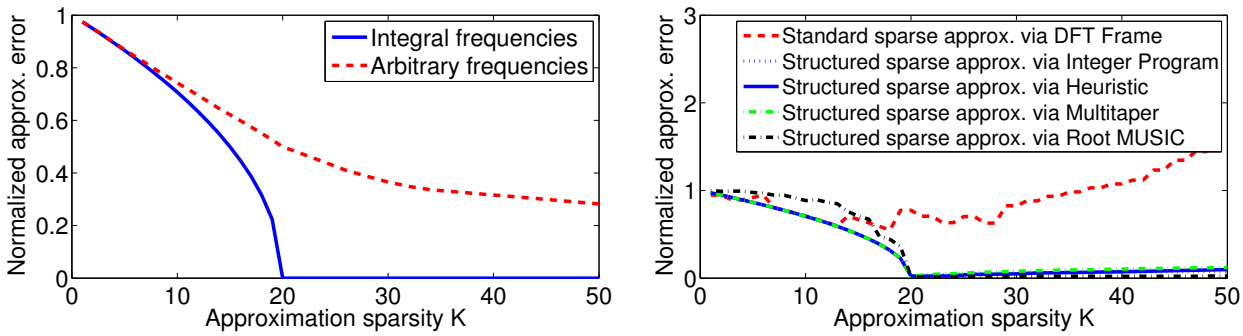

Figure 2: *Performance of standard and structured K-term sparse approximation algorithms on two classes of frequency-sparse signals of length $N = 1024$ and containing 20 sinusoids. Signals in the first class contain sinusoids at only integral frequencies; signals in the second class contain sinusoids at arbitrary integral and nonintegral frequencies. We plot the signal approximation error as a function of the approximation sparsity K. (a) Orthonormal DFT basis approximation performance is ideal for signals with exclusively integral frequencies and atrocious for signals with non-integral frequencies. (b) Five alternative approximation strategies for sinusoids with non-intgeral frequencies. Standard sparse approximation using the DFT frame $\Psi(c)$, $c = 10$, performs even worse than the DFT basis. Structured sparse approximation based on integer programming (Algorithm 1), heuristic (Algorithm 3), Thomson's multitaper method, and Root MUSIC spectral estimation perform much better.*

denote a normalized vector containing regular samples of a complex sinusoid with angular frequency $\omega \in [0, 2\pi)$. The DFT frame with redundancy factor c is then defined as

$$\Psi(c) := [\mathbf{e}(0)\ \mathbf{e}(\Delta)\ \mathbf{e}(2\Delta)\ \ldots\ \mathbf{e}(2\pi - \Delta)],$$

and the corresponding signal representation $\theta = \Psi(c)^H \mathbf{x}$ provides us with cN equispaced samples of the signal's DTFT. Note that $\Psi(1) = \mathbf{F}$, the usual orthonormal DFT basis.

One might presume that we can use the DFT frame $\Psi(c)$ to obtain sparser approximations of frequency-sparse signals with components at nonintegral frequencies, since, as the frequency redundancy factor c increases, the K-sparse approximation provided by $\Psi(c)$ becomes increasingly accurate. The proof of the following lemma is given in Appendix A.

Lemma 1. *Let $\mathbf{x} = \sum_{k=1}^{K} a_k \mathbf{e}(\omega_k)$ be a K-frequency-sparse signal, and let $\mathbf{x}_K = \Psi(c)\theta_K$ be its best K-sparse approximation in the frame $\Psi(c)$, with $\|\theta_K\|_0 = K$. Then the corresponding best K-term approximation error for $\mathbf{x}$ obeys*

$$\|\mathbf{x} - \mathbf{x}_K\|_2 \leq \sqrt{1 - |D_N(\pi/cN)/N|^2}\|\mathbf{a}\|_1, \tag{7}$$

where $\mathbf{a} = [a_1\ \ldots\ a_K]^T$.

The term on the right hand side of (7) goes to zero as $c \to \infty$. Unfortunately, however, standard sparse approximation algorithms for $\mathbf{x}$ in the frame $\Psi(c)$ do not perform well when

c increases, due to the high coherence between the frame vectors, particularly for large values of c (see equation (A.2) in Appendix A):

$$\mu(\Psi(c)) = \frac{|D_N(2\pi/cN)|}{N} \to 1 \text{ as } c \to \infty. \tag{8}$$

Due to this tradeoff, the frequency redundancy factor required by (4) to successfully find the sparse representation of a K-sparse signal is

$$c \leq \frac{\pi}{N\, D_N^{-1}\left(\frac{N}{16(K-1)}\right)},$$

where $D_N^{-1}(y)$ denotes the largest value of x for which $|D_N(x)| \geq y$ for $y \in \mathbb{R}^+$. In words, the sparsity K of the signal limits the maximum size of the redundant DFT frame that we can employ, and vice-versa. Figure 2(b) demonstrates the performance of standard sparse approximation of the same signal with arbitrary frequencies as in Fig. 2(a), but using the redundant frame $\Psi(c)$ instead, with $c = 10$. Due to the high coherence of the frame $\Psi(c)$, the algorithm cannot obtain an accurate sparse approximation of the signal.

2.4. Model-based compressive sensing

While many natural and manmade signals and images can be described to first-order as sparse or compressible, the support of their large coefficients often has an underlying second-order inter-dependency structure. This structure can often be captured by a finite-dimensional *union-of-subspaces* model that enables an algorithmic *model-based CS* framework to exploit signal structure during recovery [15, 51, 52]. We provide a brief review of model-based CS below; in Section 4, we will use this framework to overcome the issues of sparse approximation and CS using coherent frames.

The set Σ_K of all length-N, K-sparse signals is the union of the $\binom{N}{K}$ K-dimensional subspaces aligned with the coordinate axes in $\mathbb{C}^N$. A *structured sparsity model* endows the K-sparse signal $\mathbf{x}$ with additional structure that allows only certain K-dimensional subspaces from Σ_K and disallows others. The signal model $\mathcal{M}_K$ is defined by the set of m_K allowed supports $\{\Omega_1, \ldots, \Omega_{m_K}\}$. Signals from $\mathcal{M}_K$ are called K-structured sparse. Signals that are well-approximated as K-structured sparse are called *structured compressible*.

If we know that the signal $\mathbf{x}$ being acquired is K-structured sparse or structured compressible, then we can relax the RIP constraint on the CS measurement matrix Υ to require isometry only for those signals in $\mathcal{M}_K$ and still achieve stable recovery from the compressive measurements $\mathbf{y} = \Upsilon\theta$. The *model-based* RIP requires for (5) to hold *only* for signals with sparse representations $\theta \in \mathcal{M}_K$ [51, 53]; we denote this new property as $\mathcal{M}_K$-RIP to specify the dependence on the chosen signal model and change the model-based RIP constant from δ_K to $\delta_{\mathcal{M}_K}$ for clarity. This *a priori* knowledge reduces the number of random measurements required for model-based RIP with high probability to $M = \mathcal{O}(K + \log m_K)$ [51]. For some models, the reduction from $M = O(K \log(N/K))$ can be significant [15].

The $\mathcal{M}_K$-RIP property is sufficient for robust recovery of structured sparse signals using specially tailored algorithms [15]. These model-based CS recovery algorithms replace the

standard optimal K-sparse approximation — performed via thresholding — with a *structured sparse approximation* algorithm $\mathbb{M}(\mathbf{x}, K)$ that returns the best K-term approximation of the signal $\mathbf{x}$ belonging in the signal model $\mathcal{M}_K$:

$$\mathbb{M}(\mathbf{x}, K) = \arg \min_{\mathbf{x}' \in \mathcal{M}_K} \|\mathbf{x} - \mathbf{x}'\|_2. \tag{9}$$

Greedy and thresholding-based algorithms are particularly amenable to structured sparsity. For example, the IHT algorithm (6) yields the corresponding model-based IHT algorithm:

$$\widehat{\theta}_{i+1} = \mathbb{M}(\widehat{\theta}_i + \Upsilon^H(\mathbf{y} - \Upsilon\theta), K). \tag{10}$$

Other examples include orthogonal matching pursuit, CoSaMP, and subspace pursuit [46, 54, 55].

To summarize, model-based CS consists of the combination of: (i) a *structured signal model* that allows only some supports for sparse signals, with a reduction in the number of measurements proportional to the amount of pruning; and (ii) a *structured sparse approximation algorithm* that provides the best approximation in the pruned subset of sparse signals for an arbitrary vector. These two components enable us to design model-based greedy recovery algorithms that achieve substantial reductions in the number of measurements required for stable signal recovery.

3. Parameter Estimation for Frequency-Sparse Signals

The goal of CS is to identify the values and locations of the large coefficients of a discrete-time sparse/compressible signal from a small set of linear measurements. For frequency-sparse signals, such an identification can be interpreted as a parameter estimation problem, since each coefficient index corresponds to a sinusoid of a certain frequency. Thus, in this case, CS aims to estimate the frequencies and amplitudes of the largest sinusoids present in the signal. In practice, most CS recovery algorithms iterate through a sequence of increasing-quality estimates of the signal coefficients by distinguishing the signal's actual nonzero coefficients from spurious estimates; spurious coefficients are often modeled as recovery noise.

We now briefly review the extensive prior work in parameter estimation for frequency-sparse signals embedded in noise [5, 6]. We start with the simple sinusoid signal model, expressed as $\mathbf{x} = A\mathbf{e}(\omega) + \mathbf{n}$, where $\mathbf{n} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$ denotes a white noise vector with i.i.d. entries. The model parameters are A and ω, the complex amplitude and frequency of the sinusoid, respectively.

3.1. Periodogram-based methods

The maximum likelihood estimator (MLE) of the amplitude A when the frequency ω is known is given by the DTFT of $\overline{\mathbf{x}}$, the zero-padded, infinite length version of the length-N signal $\mathbf{x}$, at frequency ω: $\widehat{A} = \frac{1}{N}\overline{\mathbf{X}}(\omega) = \langle \mathbf{e}(\omega), \mathbf{x} \rangle$ [5, 6]. Furthermore, since only a single sinusoid is present, the MLE for the frequency ω is given by the frequency of the

largest-magnitude DTFT coefficient of $\overline{\mathbf{x}}$: $\widehat{\omega} = \arg\sup_\omega |\overline{\mathbf{X}}(\omega)| = \arg\sup_\omega |\langle \mathbf{e}(\omega), \mathbf{x}\rangle|$ [5, 6]. This simple estimator can be extended to the multiple sinusoid setting by performing combinatorial hypothesis testing [6].

For frequency-sparse signals with components at integral frequencies, the signal's DFT basis coefficients provide sufficient information to compute the MLE above; in this case, the parameter estimation problem above is equivalent to a 1-sparse approximation in the DFT basis. This approach is known in the spectral analysis literature as the *periodogram method* [6]. The periodogram approach can easily be extended to frequency-sparse signals whose component frequencies are in the set of frequencies sampled by the DFT frame (i.e., frequencies $\frac{2\pi n}{cN}$, where c is the redundancy factor and n is any integer).

From the spectral analysis point of view, we can argue that the coherence of the DFT frame $\Psi(c)$ is simply another manifestation of the spectral leakage problem. The classical way to combat spectral leakage is to apply a tapered window function to the signal before computing the DFT [6, 7]. However, windowing degrades spectral resolution, making it more difficult to identify frequency-sparse signal components with similar frequencies.

3.2. Thomson's multi-taper method

A revolutionary multitaper approach to spectral estimation proposed by Thomson [16] forms a weighted average of windowed DTFTs using a special set of windows $\mathbf{v}_j,\ j = 1, \ldots, J$ known as discrete prolate spheroidal wave functions (DPSWFs). The DPSWF windows $\mathbf{v}_j$ are unit-norm vectors with DTFTs $V_j(f)$ that solve the eigenvector/eigenvalue problem

$$\int_{-W}^{W} \frac{\sin(N\pi(f-f'))}{\sin(\pi(f-f'))} V_j(f')df' = \lambda_j V_j(f), \tag{11}$$

where the parameter W controls the bandwidth of the window. By construction, DPSWF windows are orthogonal, time-limited, and optimally concentrated in the frequency interval $[-W, W]$; in fact, a fraction λ_j of their unit energy is concentrated in this interval, and so one can sort the DPSWFs according to their corresponding eigenvalues. Hence, they are a natural tool for optimizing the resolution of the frequency analysis, trading estimation bias vs. variance [16].

In this paper, we are primarily interested in Thomson's line spectrum estimation technique [16], which computes a weighted sum of windowed periodograms

$$F(\mathbf{x}, \omega) = \frac{\sum_{j=1}^{J} V_j(0)\mathbf{X}_j(\omega)}{\sum_{j=1}^{J} V_j^2(0)}, \tag{12}$$

where V_j is the DTFT of the DPSWF window $\mathbf{v}_j$ and $\mathbf{X}_j$ is the DFT of $\mathbf{x}_j(n) := \mathbf{x}(n)\mathbf{v}_j(n)$. Assuming an additive white Gaussian background noise model, Thomson forms a statistical test for whether a sinusoid $\mathbf{e}(\omega)$ is present in the data using the score function

$$S(\omega) = \frac{(J-1)|F(\mathbf{x}, \omega)|^2 \sum_{j=1}^{J} V_j(0)^2}{\sum_{j=1}^{J} |\mathbf{X}_j(\omega) - F(\mathbf{x}, \omega)V_j(0)|^2}. \tag{13}$$

If $S(\omega)$ exceeds a significance threshold, then we say that a sinusoid exists at frequency ω. The probability of missing a sinusoid increases with the threshold [16].

We can re-formulate the multitaper method as a K-sparse approximation algorithm $\{\widehat{\omega}_k, \widehat{a}_k\}_{k=1}^K = \mathbb{T}_t(\mathbf{x}, K)$ in the frequency domain. First, we obtain the K frequencies within the oversampled frequency grid $\{\widehat{\omega}_k\}_{k=1}^K$ with the top statistical scores (13); second, we estimate the values $\{\widehat{a}_k\}_{k=1}^K$ of the corresponding DTFT coefficients for the signal via (12). Figure 2(b) demonstrates the clear advantages of this approach over a naïve periodogram (DFT frame).

3.3. Eigenanalysis-based methods

A modern alternative to classical periodogram-based spectral estimates are line spectral estimation algorithms based on eigenanalysis of the signal's correlation matrix [6]. Such algorithms estimate the principal components of the signal's autocorrelation matrix in order to find the dominant signal modes in the frequency domain. Eigenanalysis-based methods provide improved resolution of the parameters of a frequency-sparse signal. Example algorithms include Pisarenko's method [56], multiple signal classification (MUSIC) [17], and estimation of signal parameters via rotationally invariant techniques (ESPRIT) [18]. A line spectral estimation algorithm $\mathbb{L}(\mathbf{x}, K)$ returns a set of dominant K frequencies for the input signal $\mathbf{x}$, with K being a controllable parameter.

As a concrete but certainly non-exhaustive example of an $\mathbb{L}(\mathbf{x}, K)$, consider the MUSIC algorithm [17], which estimates the parameters of a frequency-sparse signal embedded in noise. We revisit the model $\mathbf{x} = \mathbf{s} + \mathbf{n}$, where $\mathbf{s}$ is now of the form (1) and $\mathbf{n} \sim \mathcal{N}(0, \sigma_n^2 \mathbf{I})$ denotes a noise vector. By defining the matrix $\mathbf{\Gamma} = [\mathbf{e}(\omega_1)\ \mathbf{e}(\omega_2)\ \ldots\ \mathbf{e}(\omega_K)]$ and the vector $\mathbf{a} = [a_1\ a_2\ \ldots\ a_K]^T$, we can rewrite the signal as

$$\mathbf{x} = \mathbf{\Gamma}\mathbf{a} + \mathbf{n},$$

with the autocorrelation matrix

$$\mathbf{R}_{\mathbf{xx}} = \mathbb{E}[\mathbf{x}\mathbf{x}^H] = \mathbf{\Gamma}\mathbf{A}^2\mathbf{\Gamma}^H + \sigma_n^2\mathbf{I}, \tag{14}$$

where $\mathbf{A} = \mathrm{diag}(\mathbf{a})$. Note that as long as $K < N$ and all frequencies $\{\omega_k\}_{k=1}^K$ are distinct, the matrix $\mathbf{\Gamma}\mathbf{A}^2\mathbf{\Gamma}^H$ has $\mathrm{rank}(\mathbf{\Gamma}\mathbf{A}^2\mathbf{\Gamma}^H) = K$ and K positive sorted eigenvalues $\{\widetilde{\lambda}_k\}_{k=1}^K$, with all other eigenvalues being equal to zero. Therefore, for the sorted eigenvalues $\{\lambda_n\}_{n=1}^N$ of $\mathbf{R}_{\mathbf{xx}}$, we have [7]

$$\lambda_n = \begin{cases} \widetilde{\lambda}_n + \sigma_n^2, & n \leq K, \\ \sigma_n^2, & K < n \leq N. \end{cases} \tag{15}$$

Now consider the matrix $\mathbf{G}$ that contains the eigenvectors of $\mathbf{R}_{\mathbf{xx}}$ corresponding to the $N-K$ smallest eigenvalues. It follows that

$$\mathbf{R}_{\mathbf{xx}}\mathbf{G} = \mathbf{G}\begin{bmatrix} \lambda_{K+1} & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{bmatrix} = \sigma_n^2\mathbf{G} = \mathbf{\Gamma}\mathbf{A}^2\mathbf{\Gamma}^H\mathbf{G} + \sigma_n^2\mathbf{G},$$

where the last two equalities result from (14) and (15), respectively. It follows then that $\mathbf{\Gamma}^H\mathbf{G} = 0$, and so the frequency values $\{\omega_k\}_{k=1}^K$ are the only solutions to the problem $\mathbf{e}(\omega)^H\mathbf{G}\mathbf{G}^H\mathbf{e}(\omega) = 0$.

The MUSIC algorithm [17] searches for the peaks of the pseudospectrum score function

$$P_{\mathsf{MUSIC}}(\omega) = \frac{1}{\mathbf{e}(\omega)^H\mathbf{G}\mathbf{G}^H\mathbf{e}(\omega)}, \tag{16}$$

for $\omega \in [0, 2\pi]$, while the Root MUSIC algorithm [19] searches for the roots of the polynomial $\mathbf{p}^H(z)\mathbf{G}\mathbf{G}^H\mathbf{p}(z)$ for $z \in \mathbb{C}$, $|z| = 1$, where $\mathbf{p}(z) = [1 \; z \; z^2 \; \ldots \; z^{N-1}]$. The frequencies can then be established through the relationship $\mathbf{e}(\omega) = \mathbf{p}(e^{j\omega})$. In practice, MUSIC and Root MUSIC operate on the sampled autocorrelation matrix

$$\widehat{\mathbf{R}_{\mathbf{xx}}} = \frac{1}{P}\sum_{i=1}^{P}\mathbf{x}_i\mathbf{x}_i^T$$

of size $L \times L$, where $L \in [K, N]$ denotes a window length and $\mathbf{x}_i = [\mathbf{x}[i] \; \mathbf{x}[i+1] \; \ldots \; \mathbf{x}[i + W - 1]]^T$ denotes the i^{th} windowed version of the signal $\mathbf{x}$ for $i = 1, \ldots, N - L + 1$. This sampling only requires that $L > K$.

We can also interpret the line spectral estimation process $\mathbb{L}$ as a K-sparse approximation algorithm $\{\widehat{\omega}_k, \widehat{a}_k\}_{k=1}^K = \mathbb{T}_m(\mathbf{x}, K)$ in the frequency domain: first, we obtain the K frequencies $\{\widehat{\omega}_k\}_{k=1}^K = \mathbb{L}(\mathbf{x}, K)$; second, we estimate the values $\{\widehat{a}_k\}_{k=1}^K$ of the corresponding DTFT coefficients for the signal as shown in Section 3.1. MUSIC provides a tradeoff between estimation accuracy and computational complexity via the selection of the window size W used to estimate the autocorrelation matrix $\mathbf{R}_{\mathbf{xx}}$. Figure 2(b) demonstrates the performance of the sparse approximation algorithm $\mathbb{T}_m(\mathbf{x}, K)$ for a signal with arbitrary frequencies, once again improving over the sparse approximation obtained via the periodogram.

4. Spectral Compressive Sensing

We are now in a position to develop new SCS recovery algorithms for the discrete-time CS framework of [1, 2] that are especially tailored to arbitrary frequency-sparse signals. We will develop two sets of algorithms based on the periodogram and line spectral estimation algorithms from Section 3.

4.1. SCS recovery via structured sparsity

To alleviate the performance-sapping coherence of the redundant DFT frame, we marry it with the model-based CS framework of Section 2.4 that forces the signal approximation to contain linear combinations only of incoherent frame elements. In this section, we propose a structured signal model $\mathcal{T}_{K,c,\nu}$ and a structured sparse approximation algorithm $\mathbb{T}(\mathbf{x}, K, c, \nu)$ that enables recovery of frequency-sparse signals using a coherent DFT frame. We assume initially that the components of the frequency-sparse signal $\mathbf{x}$ have frequencies in the oversampled grid of the redundant frame $\Psi(c)$; we will then extend our treatment to signals with components at arbitrary frequencies at the end of the subsection.

Algorithm 1 Coherence-inhibiting structured sparse approximation $\mathbb{T}(\mathbf{x}, K, c, \nu)$

Inputs: Signal vector $\mathbf{x}$, target sparsity K, frequency redundancy factor c, maximum coherence ν

Outputs: K-structured sparse approximation $\widehat{\mathbf{x}}$

Initialize: $\theta = \Psi(c)^H \mathbf{x}$, $\mathbf{c}_\theta[i] = |\theta[i]|^2,\ i = 0, \ldots, cN-1$ {calculate entry-wise energy}

Solve $\mathbf{s} = \arg\max_{\mathbf{s}\in\{0,1\}^{cN}} \mathbf{c}_\theta^T \mathbf{s}$ such that $\mathbf{D}_\nu \mathbf{s} \leq \mathbf{1}, \mathbf{s}^T\mathbf{1} \leq K$ {obtain optimal approximation support via integer program}

$\widehat{\theta}[i] \leftarrow \theta[i]\mathbf{s}[i], i = 0, \ldots, cN-1$ {mask coefficient vector}

return $\widehat{\mathbf{x}} = \Psi(c)\widehat{\theta}$

4.1.1. Structured signal model

We begin by defining the following structured signal model for frequency-sparse signals requiring that the components of the signal are incoherent with each other:

$$\mathcal{T}_{K,c,\nu} = \left\{ \sum_{k=1}^{K} a_k \mathbf{e}(d_k\Delta) \text{ s. t. } d_k \in \{0, \ldots, cN-1\}, |\langle \mathbf{e}(d_k\Delta), \mathbf{e}(d_j\Delta)\rangle| \leq \nu,\ 1 \leq k \neq j \leq K \right\}, \tag{17}$$

where $\nu \in [0, 1]$ is the maximal coherence allowed and $\Delta = 2\pi/cN$ as before. The union of subspaces contained in $\mathcal{T}_{K,c,\nu}$ corresponds to all linear combinations of K elements from the DFT frame $\Psi(c)$ that are pairwise sufficiently incoherent. The coherence restriction in (17) imposes a resolution limit on the recovery (in the sense of the minimum spacing between discernible sinusoids) in a manner similar to the classical estimators in Section 3. To guarantee a separation of κ Hz between frequencies present in a signal $x \in \mathcal{T}_{K,c,\nu}$, one should set $\nu = |D_N(\kappa)|/N$, cf. (8).

4.1.2. Structured sparse approximation algorithm

Following the coherence-inhibiting model $\mathcal{T}_{K,c,\nu}$ above, we modify a standard sparse approximation algorithm to avoid selecting highly coherent pairs of elements of the DFT frame $\Psi(c)$. Our structured sparse approximation algorithm is an adaptation of the refractory model-based algorithm of [52] and can be implemented as an integer program.

The algorithm $\widehat{\mathbf{x}} = \mathbb{T}(\mathbf{x}, K, c, \nu)$, shown as Algorithm 1, solves the structured sparse approximation problem (9) for the structured sparsity model $\mathcal{T}_{K,c,\nu}$. The algorithm relies on an integer program that employs a cost vector $\mathbf{c}$ and a constraint matrix $\mathbf{D}_\nu \in \mathbb{R}^{cN\times cN}$. This matrix has binary entries that indicate whether each pair of elements from the DFT frame $\Psi(c)$ are coherent:

$$\mathbf{D}_\nu[i,j] = \begin{cases} 1 & \text{if } |\langle \mathbf{e}(i\Delta), \mathbf{e}(j\Delta)\rangle| \geq \nu, \\ 0 & \text{if } |\langle \mathbf{e}(i\Delta), \mathbf{e}(j\Delta)\rangle| < \nu. \end{cases}$$

Since the algorithm operates on the vector of periodogram coefficients of the signal, we say that Algorithm 1 is a periodogram-based sparse approximation algorithm. Figure 2(b) demonstrates the performance of $\mathbb{T}(\mathbf{x}, K)$ for a signal with arbitrary frequencies, improving over the standard sparse approximations obtained via the DFT frame.

Algorithm 2 Spectral iterative hard thresholding (SIHT)

Inputs: CS Matrix Φ, measurements $\mathbf{y}$, structured sparse approx. algorithm $\mathbb{T}(\cdot, K, c, \nu)$.
Outputs: K-sparse signal estimate $\widehat{\mathbf{x}}$.
initialize: $\widehat{\mathbf{x}}_0 = 0$, $\mathbf{r} = \mathbf{y}$, $i = 0$
while halting criterion false **do**
  $i \leftarrow i + 1$
  $\widehat{\mathbf{x}}_i \leftarrow \mathbb{T}(\widehat{\mathbf{x}}_{i-1} + \Phi^T \mathbf{r}, K, c, \nu)$ {prune signal estimate using structured sparsity model}
  $\mathbf{r} \leftarrow \mathbf{y} - \Phi\widehat{\mathbf{x}}_i$ {update measurement residual}
end while
return $\widehat{\mathbf{x}} \leftarrow \Psi\widehat{\theta}$

When the matrix $\mathbf{D}_\nu$ is totally unimodular, the integer program within Algorithm 1 has the same solution as its noninteger relaxation

$$\mathbf{s} = \arg\max_{\mathbf{s}\in\mathbb{R}^{cN}} \mathbf{c}_\theta^T \mathbf{s} \text{ such that } \mathbf{D}_\nu \mathbf{s} \leq \mathbf{1}, \mathbf{s}^T \mathbf{1} \leq K,$$

which is a linear program [57]. One class of totally unimodular matrices are *interval matrices*, which are binary matrices in which the ones appear consecutively in each row. While the matrix $\mathbf{D}_\nu$ we use in our case is not an interval matrix — since each row of $\mathbf{D}_\nu$ contains several intervals — it is possible to relax the integer program by using a modified matrix $\overline{\mathbf{D}_\nu}$. To obtain this new matrix, we decompose each row $\mathbf{s}_n$ of $\mathbf{D}_\nu$ into a set of rows $\mathbf{s}_{n,1}$, $\mathbf{s}_{n,2}, \ldots$ that contain only one interval each and for which $\sum_i \mathbf{s}_{n,i} = \mathbf{s}_n$. The number of rows of $\overline{\mathbf{D}_\nu}$ is then at most $\frac{cN}{\pi\nu}$. If there is conflict within the vector obtained by the modified constraints, then the expansion from $\mathbf{D}_\nu$ to $\overline{\mathbf{D}_\nu}$ can be reversed accordingly to remove the conflicts by merging the intervals onto a single connected interval containing the conflicting smaller intervals.[5]

4.1.3. Recovery algorithm

The model-based IHT algorithm (10) is particularly amenable to modification to incorporate our frequency-sparse approximation algorithms. The modified algorithm, which we dub *spectral iterative hard thresholding* (SIHT), is unfurled in Algorithm 2 and uses the structured sparse approximation algorithm $\mathbb{T}(\cdot)$ introduced in Algorithm 1.

SIHT inherits a strong performance guarantee from standard IHT; we apply the result of [15, Theorem 4] to obtain the following.

Theorem 1. *If the matrix $\Phi\Psi$ has $\mathcal{T}_{3K,c,\nu}$-RIP with $\delta_{\mathcal{T}_{3K,c,\nu}} < 0.1$, the signal $\mathbf{x} = \Psi(c)\theta \in \mathcal{T}_{K,c,\nu}$ (i.e., $\|\theta\|_0 \leq K$ and θ invokes the coherence-inhibiting structure described earlier), and $\mathbf{y} = \Phi\mathbf{x} + \mathbf{n}$, then the estimate $\widehat{\mathbf{x}}_i$ from the i^{th} iteration of Algorithm 2 meets the guarantee*

$$\|\mathbf{x} - \widehat{\mathbf{x}}_i\|_2 \leq 2^{-i}\|\theta\|_2 + 4\|\mathbf{n}\|_2. \tag{18}$$

[5] In our implementation of the algorithm, and for simplicity, we implemented the integer program with the original matrix $\mathbf{D}_\nu$ instead of performing the expansion/merger provided here. The experimental results show that the resulting structured sparse recovery algorithm still has good recovery performance.

4.1.4. Required number of CS measurements

To calculate the number of random CS measurements needed for stable signal recovery using Algorithm 2, we exploit the incoherence of the elements composing a signal $x \in \mathcal{T}_{K,c,\nu}$ and the count of the number of subspaces t_K that compose the signal model $\mathcal{T}_{K,c,\nu}$. In a slight abuse of notation, we express the signal model $\mathcal{T}_{K,c,\nu}$ as the set of the signal supports $\Omega = \text{supp}(\theta)$ that are allowed for the coefficient vector θ so that $x = \Psi(c)\theta \in \mathcal{T}_{K,c,\nu}$ and $t_K = |\mathcal{T}_{K,c,\nu}|$ is the total number of possible supports.

To begin, we adapt [14, Lemma 2.3] to our coherence-inhibiting model.

Lemma 2. *For a support $\Omega \subseteq \{1, \ldots, cN\}$, define the subdictionary $\Psi(c)_\Omega$ as the submatrix of $\Psi(c)$ with columns indexed by Ω. Further, define the* isometry constant *of $\Psi(c)_\Omega$, denoted δ_Ω, as the smallest value such that for all $x \in \mathbb{C}^K$, $(1-\delta_\Omega)\|x\|_2^2 \leq \|\Psi(c)_\Omega x\|_2^2 \leq (1+\delta_\Omega)\|x\|_2^2$, and define the structured isometry constant for Ψ and the model $\mathcal{T}$ as $\delta_{\Psi,\mathcal{T}} = \max_{\Omega\in\mathcal{T}} \delta_\Omega$. Then we have $\delta_{\Psi,\mathcal{T}_{K,c,\nu}} \leq (K-1)\nu$.*

Lemma 2 is proved in Appendix B and can be combined with a modified version of [14, Theorem 2.2], reproduced below for completeness.

Theorem 2. *Let Ψ be a redundant frame, $\mathcal{M}_K$ a structured sparsity model, and $\Phi \in \mathbb{R}^{M\times N}$ be a matrix with i.i.d. Gaussian entries. Assume that*

$$M \geq C\delta^{-2}\left(\log|\mathcal{M}_K| + K\log(2e(1+12/\delta)) + \rho\right)$$

for some $\delta \in (0,1)$ and $\rho > 0$. Then with probability at least $1-e^{-\rho}$, the matrix $\Phi\Psi$ has $\mathcal{M}_K$-RIP with constant $\delta_{\mathcal{M}_K} \leq \delta_{\Psi,\mathcal{M}_K} + \delta(1+\delta_{\Psi,\mathcal{M}_K})$.

Proof sketch. The proof is nearly identical to that of [14, Theorem 2.2], except that (i) we use the structured isometry constant of Ψ instead of its standard isometry constant, and (ii) we change the number of supports/subspaces from $\binom{cN}{K}$ to $|\mathcal{M}_K|$. □

We can then combine these two results to obtain an analog of [14, Corollary 2.4]:

Corollary 1. *Let $\Phi \in \mathbb{R}^{M\times N}$ be a matrix with i.i.d. Gaussian entries. Assume that*

$$M \geq C_1(\log t_K + C_2 K + \rho), \tag{19}$$

for some $\rho > 0$ and fixed constants C_1, C_2, and that $K - 1 \leq \frac{1}{16\nu}$. Then with probability at least $1-e^{-\rho}$ the matrix $\Phi\Psi$ has $\mathcal{T}_{K,c,\nu}$-RIP with constant $\delta_{\mathcal{T}_{K,c,\nu}} \leq 1/3$.

We can leverage this measurement bound by counting the number $t_K = |\mathcal{T}_{K,c,\nu}|$ of K-dimensional subspaces generated by subsets of the frame $\Psi(c)$ where no two vectors in a subspace have frequencies closer than $\kappa = \left|D_N^{-1}(\nu N)\right|$ radians/sample. From [52], we know the number of subspaces to be

$$t_K < \binom{cN - (K-1)(c\kappa - 1)}{K};$$

plugging this count into (19), we obtain

$$M = \mathcal{O}\left(K \log\left(\frac{c(N - K\kappa)}{K}\right)\right). \tag{20}$$

The measurement bound (20) states that the number of measurements required for stable recovery using the redundant DFT frame (containing cN elements) is essentially of the same order as the number of measurements required for stable recovery using an orthonormal basis with cN elements. In other words, the coherence inhibition effectively neutralizes the influence of the frame's coherence on the required number of measurements for stable recovery. We will demonstrate below in Section 5 that, in practice, SCS offers significant reductions in the number of measurements required for accurate recovery of frequency-sparse signals compared to standard CS using both the orthonormal DFT basis and DFT frames.

4.2. Alternatives to structured sparsity

4.2.1. Computationally efficient heuristics

The computational complexity of the structured sparse approximation in Algorithm 1 is $\mathcal{O}(c^3N^3)$ due to the linear program for finding $\mathbf{s}$. This complexity is significantly higher than the $\mathcal{O}(cN \log(cN))$ complexity of sorting-based thresholding. A heuristic alternative to structured sparse approximation $\widehat{\mathbf{x}} = \mathbb{T}_h(\mathbf{x}, K, c, \nu)$ is given in Algorithm 3. To obtain the heuristic structured sparse approximation to the coefficient vector $\theta = \Psi(c)^H\mathbf{x}$, we greedily search for the vector entry $\theta(d_{\max})$ with the largest magnitude. Once a coefficient is selected, we inhibit all coefficients corresponding to coherent sinusoids (i.e., indices d for which $|\langle \mathbf{e}(d\Delta), \mathbf{e}(d_{\max}\Delta)\rangle| > \nu$) by setting those coefficients to zero. This will include all coefficients for frequencies within κ radians/sample of the one selected. We then repeat the process by searching for the next largest coefficient in magnitude until K coefficients are selected or all coefficients are zero. This heuristic has computational complexity $O(cKN \log(cN))$ and offers very good average performance for sparse approximation of arbitrary frequency-sparse signals, as shown in Section 5.

We subsequently can obtain a more computationally efficient version of SIHT simply of swapping $\mathbb{T}(\cdot)$ (Algorithm 1) with $\mathbb{T}_h(\cdot)$ (Algorithm 3) in Algorithm 2. This modified heuristic algorithm, although more computationally efficient, does not inherit the performance guarantee of Theorem 1.

4.2.2. Frequency interpolation

Up to this point, we have focused on recovery algorithms that return estimates having sinusoidal components with frequencies belonging to a fixed grid. We now address the case where the components have frequencies ω that are not exactly on that grid; that is, cases where the ratios ω/Δ are not necessarily integer. We can modify the structured sparse approximation algorithm $\mathbb{T}(\mathbf{x}, K, c, \nu)$ used within Algorithm 2 to include frequency and magnitude estimation steps. In this case, the modified approximation algorithm $\widehat{\mathbf{x}} = \mathbb{T}_i(\mathbf{x}, K, c, \nu)$ uses the frequency value estimates given by the sparse support of $\widehat{\theta}$ within Algorithms 1 and 3. The indices in the support $\{d_1, \ldots, d_K\} \subseteq \{1, \ldots, cN\}$ identify the grid

Algorithm 3 Heuristic coherence-inhibiting sparse approximation $\mathbb{T}_h(\mathbf{x}, K, c, \nu)$

Inputs: Signal vector $\mathbf{x}$, target sparsity K, frequency redundancy factor c, maximum coherence ν
Outputs: K-structured sparse approximation $\widehat{\mathbf{x}}$
Initialize: $\theta = \Psi(c)^H \mathbf{x}$, $\widehat{\theta}[d] = 0,\ d = 0, \ldots, cN-1$
while $\|\widehat{\theta}\|_0 \leq K$ and $\|\theta\|_2 > 0$ **do**
  $d_{\max} = \arg\max_{0 \leq d < cN} |\theta(d)|$ {search for entry with largest magnitude}
  $\widehat{\theta}[d_{\max}] \leftarrow \theta[d_{\max}]$ {copy largest magnitude entry to output estimate}
  for $d = 0$ to $cN-1$ **do**
    if $|\langle \mathbf{e}(d\Delta), \mathbf{e}(d_{\max}\Delta)\rangle| \geq \nu$ **then**
      $\theta[d] \leftarrow 0$ {inhibit entries for coherent dictionary elements}
    end if
  end for
end while
return $\widehat{\mathbf{x}} = \Psi(c)\widehat{\theta}$

frequencies closest to the frequencies of the components of the signal $\mathbf{x}$. It is then possible to refine these component frequency estimates by performing a least squares fit: for each index d_k selected, we fit the frame analysis coefficients $\widehat{\theta}' = \Psi^H \widehat{\mathbf{x}}$ for a set of indices neighboring d_k to the functional form of the Dirichlet kernel-shaped frequency response of a windowed sinusoid. By considering the Taylor series expansion of a translated Dirichlet kernel

$$D_N(\omega - \overline{\omega}) \approx 1 - \frac{(N^3 - N)(\omega - \overline{\omega})^2}{6} + \frac{(3N^4 - 10N^2 + 7)(\omega - \overline{\omega})^4}{360} - \ldots,$$

we see that a quadratic fit works well for frequencies near the peak of its main lobe [16]; see Fig. 3 for an example. The least squares fit then proceeds as follows. For each index d_k in the support of $\widehat{\theta}$, we find the coefficients $(\widehat{c}_{k,1}, \widehat{c}_{k,2}, \widehat{c}_{k,3})$ that minimize

$$(\widehat{c}_{k,1}, \widehat{c}_{k,2}, \widehat{c}_{k,3}) = \arg \min_{c_1, c_2, c_3 \in \mathbb{R}} \sum_{l=-1}^{1} (c_1 \Delta^2 (d_k - l)^2 + c_2 \Delta (d_k - l) + c_3 - \widehat{\theta}'(d_k - l))^2,$$

which provides us with the coefficients for the least-squares quadratic fit to the samples $[\widehat{\theta}'(d_k - 1)\ \widehat{\theta}'(d_k)\ \widehat{\theta}'(d_k + 1)]$ in the neighborhood of the sample $\widehat{\theta}'(d_k)$. We then identify a new estimate of the corresponding frequency value as the frequency returning the maximum value of the quadratic fit

$$\widehat{\omega}_k = \arg \max_{\omega \in \mathbb{R}} \widehat{c}_{k,1}\omega^2 + \widehat{c}_{k,2}\omega + \widehat{c}_{k,3} = -\frac{\widehat{c}_{k,2}}{2\widehat{c}_{k,1}}.$$

The corresponding sinusoid's amplitude can be estimated using the DTFT as described in Section 3.1.

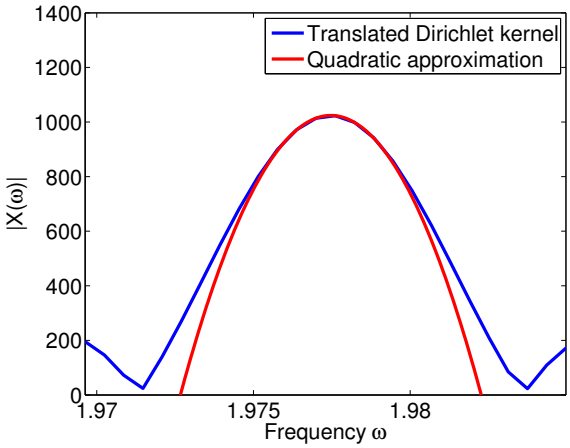

Figure 3: *Approximation of windowed sinusoid's frequency response (Dirichlet kernel) in the main lobe by a quadratic function. The Dirichlet kernel's maximum/peak is located at* $\omega = 1.9775$ *radians/second. The maximum/peak of the quadratic least-squares fit to three samples of a redundant DFT of the signal (with* $N = 1024$ *and* $c = 10$*) around its peak is located at* $\widehat{\omega} = 1.9775$*, i.e., it is an accurate estimate of the sinusoid's frequency with precision to four decimals.*

4.2.3. SCS using alternative spectral estimation methods

While the combination of a redundant frame and a coherence-inhibiting structured sparsity model yields an improvement in the performance of SIHT over standard CS recovery techniques, the algorithm still suffers from a limitation in the resolution of neighboring frequencies that it can distinguish. This limitation is inherited from the frequency and coefficient estimation methods used by SIHT (in Algorithms 1 and 3), which are based on the periodogram.

Fortunately, we can leverage the alternative multitaper and eigenanalysis-based spectral estimation methods described in Sections 3.2 and 3.3, respectively; recall that these methods return a set of detected dominant K frequencies for the input signal, where K is a controllable parameter. Since these methods do not rely on redundant frames, we do not need to leverage the features of SIHT that control the effect of coherence. We simply employ the structured sparse approximation algorithms $\mathbb{T}_t(\mathbf{x}, K)$ (alternatively, $\mathbb{T}_m(\mathbf{x}, K)$) within IHT, resulting in Algorithm 4.

5. Experimental Results

In this section, we perform a range of computational experiments to test the limits of SCS and validate the theoretical guarantees developed above. We compare the performance of the SIHT algorithm variants (based on the periodogram in Algorithm 2 and on line spectral estimation in Algorithm 4) to the standard CS recovery paradigm of [1, 2] implemented using the IHT algorithm (6). We probe the robustness of the algorithms to varying amounts of measurement noise and varying frequency redundancy factors c. We also test the algorithms on a real-world communications signal. Throughout this section, the two metrics of performance we use are the normalized error $E = \|\mathbf{x} - \widehat{\mathbf{x}}\|_2 / \|\mathbf{x}\|_2$ and the signal to noise

Algorithm 4 SIHT using Line Spectral Estimation

Inputs: CS Matrix Φ, measurements $\mathbf{y}$, line approximation algorithm $\mathbb{T}_t(\cdot, K)$.
Outputs: K-sparse approximation $\{\widehat{\omega}_k, \widehat{a}_k\}_{k=1}^K$, signal estimate $\widehat{\mathbf{x}}$.
initialize: $\widehat{\mathbf{x}}_0 = 0$, $\mathbf{r} = \mathbf{y}$, $i = 0$
while halting criterion false **do**
  $i \leftarrow i + 1$
  $\{\widehat{\omega}_k, \widehat{a}_k\}_{k=1}^K \leftarrow \mathbb{T}_t(\widehat{\mathbf{x}}_{i-1} + \Phi^T(\mathbf{y} - \Phi\widehat{\mathbf{x}}_{i-1}), K)$ {obtain parameter estimates}
  $\widehat{\mathbf{x}}_i \leftarrow \sum_{k=1}^K \widehat{a}_k \mathbf{e}(\widehat{\omega}_k)$ {form signal estimate}
end while
return $\widehat{\mathbf{x}} \leftarrow \widehat{\mathbf{x}}_i$, $\{\widehat{\omega}_k, \widehat{a}_k\}_{k=1}^K$

ratio SNR $= -20\log_{10} E$, averaged over all independent iterations of each experiment. A Matlab toolbox containing implementations of the SCS recovery algorithms, together with scripts that generate all figures in this paper, is available at http://dsp.rice.edu/scs.

Our first experiment compares the performance of standard IHT using the orthonormal DFT basis against that of the SIHT algorithms. Our experiments use signals of length $N = 1024$ samples (chosen for computational efficiency) containing $K = 20$ complex-valued sinusoids. For varying M, we executed 100 independent trials using random measurement matrices Φ of size $M \times N$ with i.i.d. Gaussian entries and signals $\mathbf{x} = \sum_{k=1}^K \mathbf{e}(\omega_k)$, where each pair of frequencies $\omega_i,\ \omega_j,\ 1 \le i, j \le K,\ i \ne j$ are spaced by at least $\kappa = 10\pi/1024$ radians/sample (i.e., two sidelobes away from one another in the Dirichlet kernel). For each CS matrix/sparse signal pair we obtain the measurements $\mathbf{y} = \Phi\mathbf{x}$ and calculate estimates of the signal $\widehat{\mathbf{x}}$ using IHT with the orthonormal DFT basis, SIHT using the periodogram (Algorithm 2) via both integer programming (Algorithm 1) and heuristic approximation (Algorithm 3) with frequency redundancy factor $c = 10$, maximum allowed coherence $\nu = 0.1$ (so that $\nu > |D_N(\kappa)|/N$), and quadratic parametric frequency interpolation as described in Section 4.2.2; and SIHT using line spectral estimation (Algorithm 4) via both Root MUSIC and Thomson's multitaper method. We use a window size $W = N/10$ in Root MUSIC to estimate the autocorrelation matrix $\mathbf{R_{xx}}$ and set $W = 5/2N$ in the multitaper method. For reference, we also evaluate the performance of the standard root MUSIC spectral estimation algorithm applied to M regular samples of the signal obtained by reducing the sampling rate by a factor of M/N, i.e., we obtain fewer samples from the same signal duration to match the equivalent sampling rate obtained in CS. We study the performance of the IHT algorithm with the DFT basis in three different regimes: (*i*) the *average* case, in which the frequencies are selected randomly to machine precision; (*ii*) the *best* case, in which the frequencies are randomly selected and rounded to the closest integral frequency, resulting in zero spectral leakage; and (*iii*) the *worst* case, in which each sinusoid frequency is midway between two consecutive integral frequencies, resulting in maximal spectral leakage. We focus on the average case analysis for root MUSIC and for all four SIHT algorithms.

The results of this experiment are summarized in Fig. 4 and show first that *the average-case performance of IHT with the DFT basis is very close to its worst-case performance*, and second that all SIHT algorithms perform significantly better on the same signals. Note

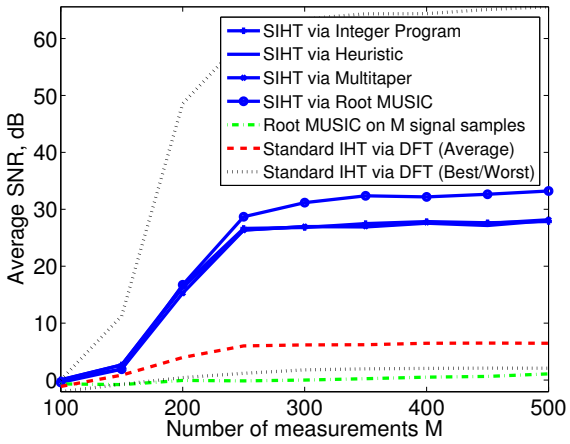

Figure 4: *Performance of standard CS signal recovery via IHT with the orthonormal DFT basis (6), SIHT Algorithm 2 (periodogram) implemented via integer program (Algorithm 1) and heuristic (Algorithm 3), and SIHT Algorithm 4 (line spectral estimation) via Root MUSIC and Thomson's multitaper method. We use signals of length* $N = 1024$ *containing* $K = 20$ *complex-valued sinusoids. The dotted lines indicate the performance of IHT via the orthonormal DFT basis for the best case (when the frequencies of the sinusoids are integral) and the worst case (when each frequency is half way in between two consecutive integral frequencies). We also compare against the performance of the root-MUSIC algorithm applied to* M *regular signal samples. The performance of IHT for arbitrary frequencies is close to its worst-case performance, while all SIHT algorithms perform significantly better for arbitrary frequencies, with the Root MUSIC-based approach providing best performance. Recovery from low-rate sampled versions of the signals performs poorly due to aliasing. All quantities are averaged over 100 independent trials.*

that the SIHT algorithms work well in the average case even though the resulting signals do not exactly match the sparse-in-DFT-frame assumption. Thus, our proposed algorithms are robust to mismatch in the values for the frequencies in the signal model (1). Note also that when we operate on M signal samples directly, the performance of spectral estimation for signal recovery suffers greatly due to the resulting aliasing of higher frequencies (also known as Nyquist folding, evident in Fig. 1(b)), with the performance improving as M increases but never matching that of the CS-based methods.

We repeat the same experimental setup in the rest of this section, but restrict it only to the average case regime. Since Figs. 1, 2, and 4 show that the performances of SIHT via integer programming (Algorithm 1) and heuristic approximation (Algorithm 3) are roughly equivalent, we focus in the sequel on the computationally simpler heuristic approach.

Our second experiment tests the robustness of the SIHT algorithms to additive noise in the measurements. We set the experiment parameters to $N = 1024$, $K = 20$ and $M = 300$, and we add i.i.d. Gaussian noise of variance σ to each measurement. For each value of σ, we fix the matrix Φ (drawn randomly as before) and perform 1000 independent trials; in each trial, we generate the signals $\mathbf{x}$ randomly as in the previous experiment. Figure 5 shows the average norm of the recovery error as a function of the noise variance σ; the linear relationship indicates stability to additive noise in the measurements, confirming the

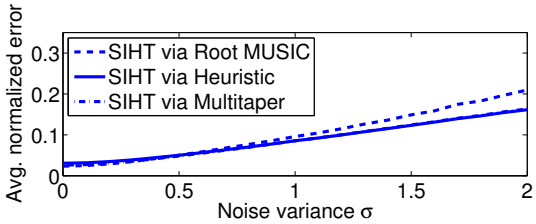

Figure 5: *Performance of SIHT Algorithm 2 (via heuristic) and SIHT Algorithm 4 (via Root MUSIC and Thomson's multitaper method) for CS signal recovery from noisy measurements. We use signals of length* $N = 1024$ *containing* $K = 20$ *complex-valued sinusoids and take* $M = 300$ *measurements. We add noise of varying variances* σ *and calculate the average normalized error magnitude over 1000 independent trials. The linear relationship between the noise variance and the recovery error indicates the robustness of the recovery algorithm to noise.*

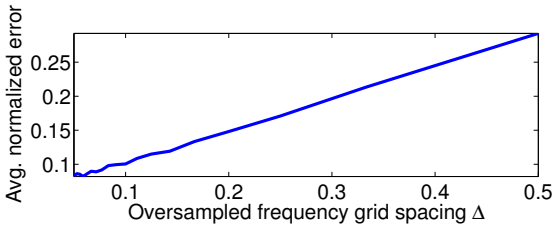

Figure 6: *Performance of SIHT Algorithm 2 via heuristic under varying grid spacing resolutions* $\Delta = 2\pi/cN$. *We use signals of length* $N = 1024$ *containing* $K = 20$ *complex-valued sinusoids and take* $M = 300$ *measurements. We average the recovery error over 10000 independent trials. There is a linear dependence between the granularity of the DFT frame and the norm of the recovery error.*

guarantee given in Theorem 1.

Our third experiment studies the impact of the frequency redundancy factor c on the performance of SIHT (Algorithm 2). We use the same matrix Φ and signal setup as in the previous experiment and execute 10000 independent trials for each value of c. The results, shown in Fig. 6, indicate a linear dependence between the granularity of the DFT frame Δ and the norm of the recovery error. This sheds light on the tradeoff between the computational complexity and the performance of the recovery algorithm, as well as between the redundancy factor M/K (dependent on $\log c$) and the recovery performance. These results also experimentally confirm Lemma 1.

Our fourth experiment tests the capacity of standard CS and SCS recovery algorithms to resolve closely spaced frequencies in frequency-sparse signals. For this experiment, the signal consists of two real-valued sinusoids (i.e., $K = 4$) of length $N = 1024$ with frequencies that are separated by a value δ_ω varying between $0.1 - 5$ cycles/sample ($2\pi/100N - 10\pi/N$ rad/sample); we obtain $M = 100$ measurements for each signal. We measure the performance of standard IHT via DFT (6), SIHT via heuristic (Algorithm 3) with frequency redundancy factor $c = 10$ and maximum allowed coherence $\nu = 0.1$, and SIHT via Thomson's multitaper method and Root MUSIC (Algorithm 4), all as a function of the frequency spacing δ_ω. For this experiment, we modify the window size parameter of the Root MUSIC algorithm to $W = N/3$ to improve its estimation accuracy at the cost of higher computational complexity.

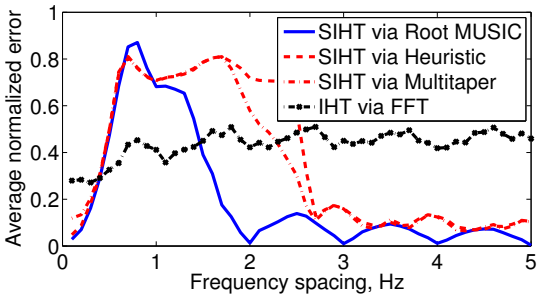

Figure 7: *Performance of standard CS recovery via IHT versus SIHT via heuristic (Algorithm 2) and SIHT via Root MUSIC and Thomson's multitaper method (Algorithm 4) for frequency-sparse signals with components at closely spaced frequencies. We use signals of length $N = 1024$ containing two real sinusoids ($K = 4$) with frequencies separated by δ_ω, and measure the signal recovery performance of IHT (6) and the SIHT algorithms (Algorithms 2 and 4) from $M = 100$ measurements as a function of δ_ω. The results verify the limitations of periodogram-based methods and the markedly improved performance of line spectral estimation methods used by the different versions of SIHT. Additionally, we see that standard IHT outperforms the SIHT algorithms only when the observed frequency-sparse signal contains highly coherent components (that is, very small frequency spacing δ_ω).*

For each value of δ_ω, we execute 100 independent trials as detailed in previous experiments. The results, shown in Fig. 7, verify the limitation of periodogram-based methods as well as the improved resolution performance afforded by line spectral estimation methods like Root MUSIC. Standard IHT outperforms the SIHT algorithms only in the case where the signal does not belong in the class of frequency-sparse signals with incoherent components (that is, for very small frequency spacing δ_ω).

Our fifth and last experiment tests the performance of standard CS and SCS recovery algorithms on a real-world signal. We use the amplitude modulated (AM) signal from [8, Figure 7] that was digitized in the lab at its Nyquist rate to create the signal $\mathbf{x}$. Rather than a Gaussian measurement matrix Φ, we employ a completely discrete-time version of the random demodulator from [8] that measures $\mathbf{x}$ using a banded, lower triangular CS matrix Φ. The signal $\mathbf{x}$ has length $N = 32768$ samples; for computational expediency, we recover the signal in half-overlapping blocks of length $N' = 1024$. We compare five different recovery algorithms as a function of the number of measurements M: standard CS via IHT (6) in the DFT basis, standard CS via ℓ_1-norm minimization in the DFT basis (so that we can directly compare our results with those in [8]), and SIHT via heuristic (Algorithm 3) with frequency redundancy factor $c = 10$ and maximum allowed coherence $\nu = 0.1$, Thomson's multitaper method, and Root MUSIC (Algorithm 4). We set the target signal sparsity to $K = 10$ in the IHT and SIHT algorithms. The AM signal estimates are then demodulated, and the recovery performance is measured in terms of the distortion against the message signal obtained by demodulating $\mathbf{x}$. We average the performance over 20 trials for the random demodulator chipping sequence. The results in Fig. 8 shows that Algorithm 4 consistently outperforms its standard CS counterparts. For example, at a measurement rate of $M = N/10$, SIHT provides a 8dB improvement in performance over the ℓ_1-norm minimization approach of [8].

To summarize, our experiments have shown that SCS achieves significantly improved

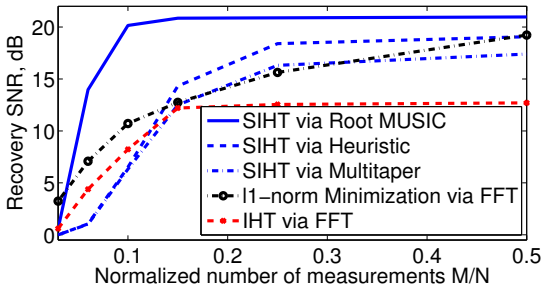

Figure 8: *Performance of ℓ_1-norm minimization, IHT via DFT basis, and SIHT via heuristic, Thomsons' multitaper method, and Root MUSIC (Algorithms 2 and 4) on a real-world AM communications signal of length $N = 32768$ for a varying number of measurements M. The SIHT algorithms (in particular, SIHT via Root MUSIC) significantly outperform their standard CS counterparts.*

signal recovery performance for the overwhelming majority of frequency-sparse signals when compared with standard CS recovery approaches. Our SCS recovery algorithms inherit some attractive properties from their standard counterparts, including robustness to model mismatch and measurement noise.

6. Related Work and Extensions

We now review several avenues of related work on the intersection of sparse approximation, compressive sensing, and spectral estimation. We also summarize new results that have appeared since the original distribution of this manuscript as a preprint [58, 59].

A recent paper [11] independently studied the poor performance of DFT-based CS recovery on frequency-sparse signals. The paper provides a generic framework for studying sparsity basis mismatch in which an inaccurate sparsity basis is used for CS recovery and determines a bound for the approximation error as a function of the basis mismatch. The paper shows that in the noiseless setting, CS via the DFT basis provides lower accuracy that linear prediction methods on subsampled sinusoids. However, such linear prediction methods are very sensitive to noise and thus are not suitable for the CS recovery approach in Section 4.

The CS framework was also recently extended to signals having a sparse representation in a redundant and coherent frame through the use of analysis sparsity [4]. While the standard CS framework is based on the sparsity of the synthesis coefficients θ of a signal $\mathbf{x} = \Psi\theta$ in the frame Ψ, it is also possible to recover signals with sparse or compressible analysis coefficient vectors $\theta' = \Psi^H\mathbf{x}$ by making a small modification to the recovery algorithm. For signals that have sparse synthesis coefficients (such as our frequency-sparse signals using DFT frames), one can express the analysis coefficients as $\theta' = \Psi^H\mathbf{x} = \mathbf{G}\theta$, where $\mathbf{G} = \Psi^H\Psi$ is the Gram matrix for the frame Ψ. Under this formulation, the coherence (i.e., the maximum entry of $\mathbf{G}$ in magnitude) can be arbitrarily large, and θ' may still be sparse or approximately sparse as long as the matrix $\mathbf{G}$ has few nonzero entries outside of the diagonal. Unfortunately, for a DFT frame Ψ, the Gram matrix $\mathbf{G}$ is dense — each row and column corresponds to a sampling of the Dirichlet kernel — and so the analysis coefficient vector θ' will not be sparse,

in general. This behavior of the Gram matrix $\mathbf{G}$ can be interpreted as a manifestation of the spectral leakage problem in oversampled spectral analysis.

Sparse approximation algorithms for frames characterized by continuously varying parameters have also been considered [60]. Here, one designs a frame composed of vectors corresponding to a sampling of the parameter space; this sampled frame can be used with a modified greedy algorithm to obtain an initial estimate of the parameter value, followed by a refinement via gradient descent. To date the analysis of such approaches has been limited to the convergence rate of the sparse approximation error $\|\mathbf{y} - \Phi\widehat{\mathbf{x}}\|_2$, which is not exactly relevant to CS applications where we instead seek low error in the sparse representation (i.e. $\|\mathbf{x} - \widehat{\mathbf{x}}\|_2$).

We have focused our efforts in this paper towards frequency-sparse signals consisting of a sparse sum of sinusoids, a model that has been also termed *sparse multitone* in the literature [61] and for which compressive analog-to-digital converters have been developed [8, 36, 37, 62, 63]. A parallel frequency-sparse signal model known as a *sparse multiband* model has emerged as an alternative [9, 33, 61, 64]. In contrast with the sparse multitone model, the sparse multiband model partitions the observable spectrum into a number of bins and assumes that the spectral content of the observed signal occupies a small number of bins in the partition, without making further assumptions as to the contents of each bin. The sparse multiband model has driven the design of recovery algorithms [9, 10, 64] and additional compressive analog-to-digital converters (dubbed Xampling in [65]) that leverage the additional structure of the model. In particular, the approach of [64] builds a dictionary composed of modulated discrete prolate spheroidal sequences (cousins of the DPSWFs used in (12) above by the multitaper method of [16]) that yields group-sparse representations for multiband signals. When applied to multitone signals, the multiband framework is agnostic to the exact values of their frequencies; however, the number of measurements required for successful recovery is proportional to the size of the spectral bins. Comparisons between the benefits and shortcomings of both signal models can be found in [33, 61].

Finite rate of innovation (FROI) sampling [29], which predates the development of CS, enables uniform sampling of analog signals governed by a small number of parameters using a specially designed sampling kernel. These samples are processed to obtain an annihilating filter, which is used to estimate the values of the parameters. The application of FROI to multitone signals results in the linear prediction method used in [11], where the arguments of the complex roots of an annihilating filter reveal the frequencies ω_k of the signal components in (1). In fact, noise-tolerant line spectral estimation algorithms [66, 67] have been proposed to extend FROI to noisy sampling settings [30, 31, 32], albeit without performance guarantees to date.

The coherence inhibition concept behind our SCS framework can be extended to other signal recovery settings where each component of the signal's sparse representation is governed by a small set of parameters. While such classes of signals are well suited for manifold models when the signal consists of a known number of parameterized components [27, 28], they fall short for arbitrary linear combinations of a varying number of components; in this case, we must estimate both the model order (number of components) and the parameter

values (choice of components).

A recent paper proposes sparse approximation in a frame whose elements are drawn from a manifold and correspond to a sampling of the manifold's parameter space [68]. However, when the manifold model is very smooth, the resulting frame will be highly coherent, limiting the performance of standard sparse approximation algorithms. Following the SCS ethos, we can impose a coherence-inhibiting models such as that of (17) to enable accurate recovery of sparse signals with such a coherent frame, as originally discussed in [58, 59] and subsequently developed in [69]. We expect such an approach algorithms to have performance guarantees similar to those given for SIHT. Similarly to [60], we can also refine the parameter estimates obtained from a frame sampling through the use of gradient descent or a least squares fit to a parametric manifold approximation. Immediate applications of this formulation include sparsity-based localization [20, 21, 22, 23, 70], radar imaging [24, 25, 26], and sparse time-frequency representations [12].

7. Conclusions

In this paper we have developed a new framework for CS recovery of frequency-sparse signals, which we have dubbed spectral compressive sensing (SCS). The framework uses a redundant frame of sinusoids corresponding to a redundant frequency grid together with a coherence-inhibiting structured signal model that prevents the usual loss of performance due to the frame coherence. We have provided both performance guarantees for SCS signal recovery and a bound on the number of random measurements needed to enable these guarantees. We have also presented adaptations of standard line spectral estimation methods to achieve recovery of combinations of sinusoids with arbitrarily close frequencies while achieving low computational complexity. As Fig. 4 indicates, SCS recovery significantly outperforms CS recovery based on the orthonormal DFT basis (up to 25dB in the figure).

Further work includes integrating our frequency inhibition and line spectral estimation approaches into more powerful greedy [46], iterative [71], and ℓ_1-norm minimization [72] recovery algorithms, as well as obtaining a full performance characterization for the line spectral estimators used in the algorithms of Section 4. The performance of these algorithms (accuracy, robustness, and resolution) might be different when they are applied to signal estimates obtained from compressive measurements. A very recent contributions in the direction of optimization-based approaches shows promising results [73]. We are also interested in extensions to other CS recovery algorithms to be used in conjunction with parameterized frame models. SCS can also be applied to signal ensembles; when a microphone or antenna array is used and the emitter is static, the dominant frequencies are the same for each of the sensors, following the common sparse supports joint sparsity model [74]. For mobile emitters, the changes in the frequency values can be modeled according to the Doppler effect, which increases the number of parameters for the signal ensemble observed from two (for emitter position) to four (for emitter position and velocity).

Appendix A. Proof of Lemma 1

Proof of Lemma 1. We start with a K-term approximation in the frame $\Psi(c)$:

$$\mathbf{x}' = \sum_{k=1}^{K} a'_k \mathbf{e}(\omega'_k),$$

where $a'_k = a_k b_k$, with b_k to be defined, and $\omega'_k = \Delta \,\mathrm{round}(\omega_k/\Delta)$. We then have

$$\begin{aligned}
\|\mathbf{x} - \mathbf{x}_k\|_2 \le \|\mathbf{x} - \mathbf{x}'\|_2 &= \left\| \sum_{k=1}^{K} a_k \mathbf{e}(\omega_k) - \sum_{k=1}^{K} a'_k \mathbf{e}(\omega'_k) \right\|_2, \\
&= \left\| \sum_{k=1}^{K} \left(a_k \mathbf{e}(\omega_k) - a_k b_k \mathbf{e}(\omega'_k)\right) \right\|_2 \le \sum_{k=1}^{K} |a_k| \left\| \mathbf{e}(\omega_k) - b_k \mathbf{e}(\omega'_k) \right\|_2, \\
&= \sum_{k=1}^{K} |a_k| \sqrt{\|\mathbf{e}(\omega_k)\|_2^2 + |b_k|^2 \|\mathbf{e}(\omega'_k)\|_2^2 - 2|b_k| \, |\langle \mathbf{e}(\omega_k), \mathbf{e}(\omega'_k)\rangle|}, \\
&= \sum_{k=1}^{K} |a_k| \sqrt{\|\mathbf{e}(\omega_k)\|_2^2 + |b_k|^2 \|\mathbf{e}(\omega'_k)\|_2^2 - \frac{2|b_k|}{N} \left| \sum_{n=1}^{N} e^{j(\omega_k - \omega'_k)n} \right|}, \\
&= \sum_{k=1}^{K} |a_k| \sqrt{\|\mathbf{e}(\omega_k)\|_2^2 + |b_k|^2 \|\mathbf{e}(\omega'_k)\|_2^2 - \frac{2|b_k|}{N} |D_N(\omega_k - \omega'_k)|}. \qquad \text{(A.1)}
\end{aligned}$$

Now we replace $\|\mathbf{e}(\omega_k)\|_2 = 1$ and minimize (A.1) by setting $b_k = D_N(\omega_k - \omega'_k)/N$. We then obtain

$$\begin{aligned}
\|\mathbf{x} - \mathbf{x}_k\|_2 &\le \sum_{k=1}^{K} |a_k| \sqrt{1 - |D_N(\omega_k - \omega'_k)/N|^2} \le \sum_{k=1}^{K} |a_k| \sqrt{1 - |D_N(\Delta/2)/N|^2} \\
&\le \sqrt{1 - |D_N(\pi/cN)/N|^2} \sum_{k=1}^{K} |a_k| = \sqrt{1 - |D_N(\pi/cN)/N|^2} \|\mathbf{a}\|_1,
\end{aligned}$$

proving the lemma. □

In the process of proving Lemma 1, we calculated the coherence of the DFT frame:

$$\begin{aligned}
\mu(\Psi(c)) = \arg \max_{1 \le i,j \le cN} |\langle \mathbf{e}(i\Delta), \mathbf{e}(j\Delta)\rangle| = |\langle \mathbf{e}(i\Delta), \mathbf{e}((i+1)\Delta)\rangle| = \left| \frac{1}{N} \sum_{n=0}^{N-1} e^{j\Delta n} \right| = \frac{|D_N(\Delta)|}{N} \\
= \frac{|D_N(2\pi/cN)|}{N}. \qquad \text{(A.2)}
\end{aligned}$$

Appendix B. Proof of Lemma 2

Proof of Lemma 2. Fix $\Omega \in \mathcal{T}_{K,c,\nu}$. We begin by noting that

$$\|\Psi(c)_\Omega x\|_2^2 = x^H \Psi(c)_\Omega^H \Psi(c)_\Omega x = x^H G_\Omega x,$$

where $G_\Omega = \Psi(c)_\Omega^H \Psi(c)_\Omega$ denotes the partial Gram matrix of $\Psi(c)$. Therefore, we must have $1-\delta_\Omega \leq \lambda_{\min}(G_\Omega)$ and $1+\delta_\Omega \geq \lambda_{\max}(G_\Omega)$, where $\lambda_{\min}(G_\Omega)$ and $\lambda_{\max}(G_\Omega)$ are the minimal and maximal eigenvalues of G_Ω. A straightforward application of Geršgorin's circle theorem [75] shows that, since the elements of $\Psi(c)_\Omega$ have inner products bounded in magnitude by ν (due to $\Omega \in \mathcal{T}_{K,c,\nu}$), we must have $1-(K-1)\nu \leq \lambda_{\min}(G_\Omega) \leq \lambda_{\max}(G_\Omega) \leq 1+(K-1)\nu$. Therefore, it follows that we can pick $\delta_\Omega \leq (K-1)\nu$, proving the lemma. □

Acknowledgements

Thanks to Volkan Cevher, Yuejie Chi, Mark Davenport, Yonina Eldar, Chinmay Hegde, Joel Tropp, and Cédric Vonesch for helpful discussions and to Isabel Duarte for reporting a bug in the SCS toolbox. This paper is dedicated to the memory of Dennis M. Healy; his insightful discussions with us inspired much of this project.

References

[1] E. J. Candès, Compressive sampling, in: Int. Congress of Mathematicians, Vol. 3, Madrid, Spain, 2006, pp. 1433–1452.

[2] D. L. Donoho, Compressed sensing, IEEE Trans. Info. Theory 52 (4) (2006) 1289–1306.

[3] R. G. Baraniuk, Compressive sensing, IEEE Signal Proc. Mag. 24 (4) (2007) 118–120, 124.

[4] E. J. Candès, Y. C. Eldar, D. Needell, P. Randall, Compressed sensing with coherent and redundant dictionaries, Appl. Comput. Harmon. Anal. 31 (1) (2011) 59–73.

[5] S. M. Kay, Fundamentals of statistical signal processing: Estimation theory, Prentice-Hall, Englewood Cliffs, NJ, 1998.

[6] S. M. Kay, Modern spectral estimation: Theory and application, Prentice Hall, Englewood Cliffs, NJ, 1988.

[7] P. Stoica, R. L. Moses, Introduction to spectral analysis, Prentice Hall, Upper Saddle River, NJ, 1997.

[8] J. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, R. G. Baraniuk, Beyond Nyquist: Efficient sampling of bandlimited signals, IEEE Trans. Info. Theory 56 (1) (2010) 1–26.

[9] Y. Eldar, M. Mishali, Blind multi-band signal reconstruction: Compressed sensing for analog signals, IEEE Trans. Signal Proc. 57 (3) (2009) 993–1009.

[10] M. Mishali, Y. Eldar, From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals, IEEE J. Selected Topics in Signal Proc. 4 (2) (2010) 375–391.

[11] Y. Chi, L. Scharf, A. Pezeshki, R. Calderbank, The sensitivity to basis mismatch of compressed sensing in spectrum analysis and beamforming, IEEE Trans. Signal Proc. 59 (5) (2011) 2182–2195.

[12] S. Mallat, A wavelet tour of signal processing, Academic Press, 1999.

[13] J. A. Tropp, Greed is good: Algorithmic results for sparse approximation, IEEE Trans. Inform. Theory 50 (10) (2004) 2231–2242.

[14] H. Rauhut, K. Schnass, P. Vandergheynst, Compressed sensing and redundant dictionaries, IEEE Trans. Info. Theory 54 (5) (2008) 2210–2219.

[15] R. G. Baraniuk, V. Cevher, M. F. Duarte, C. Hegde, Model-based compressive sensing, IEEE Trans. Info. Theory 56 (4) (2010) 1982–2001.

[16] D. J. Thomson, Spectrum estimation and harmonic analysis, Proceedings of the IEEE 70 (9) (1982) 1055–1094.

[17] R. O. Schmidt, Multiple emitter location and signal parameter estimation, IEEE Trans. Antennas Propagation 34 (3) (1986) 276–280.

[18] R. Roy, T. Kailath, ESPRIT — Estimation of signal parameters via rotational invariance techniques, IEEE Trans. Acoust., Speech, Signal Proc. 37 (7) (1989) 984–995.

[19] A. Barabell, Improving the resolution performance of eigenstructure-based direction-finding algorithms, in: IEEE Int. Conf. Acoustics, Speech and Signal Proc. (ICASSP), Boston, MA, 1983, pp. 336–339.

[20] I. F. Gorodnitsky, B. D. Rao, Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm, IEEE Trans. Signal Proc. 45 (3) (1997) 600–616.

[21] D. Malioutov, M. Cetin, A. S. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays, IEEE Trans. Signal Proc. 53 (8) (2005) 3010–3022.

[22] V. Cevher, M. F. Duarte, R. G. Baraniuk, Localization via spatial sparsity, in: European Signal Proc. Conf. (EUSIPCO), Lausanne, Switzerland, 2008.

[23] V. Cevher, A. C. Gurbuz, J. H. McClellan, R. Chellappa, Compressive wireless arrays for bearing estimation, in: IEEE Int. Conf. Acoustics, Speech and Signal Proc. (ICASSP), Las Vegas, NV, 2008, pp. 2497–2500.

[24] R. G. Baraniuk, P. Steeghs, Compressive radar imaging, in: IEEE Radar Conf., Boston, MA, 2007, pp. 128–133.

[25] K. R. Varshney, M. Cetin, J. W. Fisher, A. S. Willsky, Sparse representation in structured dictionaries with application to synthetic aperture radar, IEEE Trans. Signal Proc. 56 (8) (2008) 3548–3561.

[26] M. Herman, T. Strohmer, High resolution radar via compressive sensing, IEEE Trans. Signal Proc. 57 (6) (2009) 2275–2284.

[27] R. G. Baraniuk, M. B. Wakin, Random projections of smooth manifolds, Foundations of Computational Mathematics 9 (1) (2009) 51–77.

[28] P. Shah, V. Chandrasekaran, Iterative projections for signal identification on manifolds: Global recovery guarantees, in: Allerton Conf. Communication, Control, and Computing, Monticello, IL, 2011, pp. 760–767.

[29] M. Vetterli, P. Marziliano, T. Blu, Sampling signals with finite rate of innovation, IEEE Trans. Signal Proc. 50 (6) (2002) 1417–1428.

[30] I. Maravić, M. Vetterli, Sampling and reconstruction of signals with finite innovation in the presence of noise, IEEE Trans. Signal Proc. 53 (8) (2005) 2788–2805.

[31] A. Ridolfi, I. Maravić, J. Kusuma, M. Vetterli, Sampling signals with finite rate of innovation: the noisy case, Tech. Rep. LCAV-REPORT-2002-001, École Polytechnique Fédérale de Laussane, Lausanne, Switzerland (Dec. 2002).

[32] T. Blu, M. P.-L. Dragotti, M. Vetterli, P. Marziliano, L. Coulot, Sparse sampling of signal innovations, IEEE Signal Proc. Mag. 25 (2) (2008) 31–40.

[33] M. Mishali, Y. C. Eldar, Sub-Nyquist sampling, IEEE Signal Proc. Mag. 28 (6) (2011) 98–124.

[34] S. Chen, D. Donoho, M. Saunders, Atomic decomposition by basis pursuit, SIAM J. on Sci. Comp. 20 (1) (1998) 33–61.

[35] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, R. G. Baraniuk, Single pixel imaging via compressive sampling, IEEE Signal Proc. Mag. 25 (2) (2008) 83–91.

[36] S. Becker, J. Yoo, M. Loh, A. Emami-Neyestanak, E. Candès, Practical design of a random demodulation sub-Nyquist ADC, in: Workshop on Signal Proc. with Adaptive Sparse Structured Representations (SPARS), Edinburgh, Scotland, 2011.

[37] M. B. Wakin, S. Becker, E. Nakamura, M. Grant, E. Sovero, D. Ching, J. Yoo, J. K. Romberg, A. Emami-Neyestanak, E. J. Candès, A non-uniform sampler for wideband spectrally-sparse environments, Preprint (June 2012).

[38] E. J. Candès, J. Romberg, Sparsity and incoherence in compressive sampling, Inverse Problems 23 (3) (2007) 969–985.

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

[40] A. C. Gilbert, S. Guha, P. Indyk, S. Muthukrishnan, M. Strauss, Near-optimal sparse Fourier representations via sampling, in: ACM Symposium on Theory of Computing (STOC), ACM, New York, NY, USA, 2002, pp. 152–161.

[41] A. C. Gilbert, S. Muthukrishnan, M. J. Strauss, Improved time bounds for near-optimal sparse Fourier representations, in: Wavelets XI at SPIE Optics and Photonics, San Diego, CA, 2005.

[42] A. C. Gilbert, M. J. Strauss, J. A. Tropp, A tutorial on fast Fourier sampling, IEEE Signal Proc. Mag. (2008) 57–66.

[43] H. Rauhut, Random sampling of sparse trigonometric polynomials, Appl. Comput. Harmon. Anal. 22 (1) (2007) 16–42.

[44] M. A. Iwen, Combinatorial sublinear-time Fourier algorithms, Foundations of Computational Mathematics 10 (3) (2010) 303–338.

[45] M. A. Davenport, M. B. Wakin, Analysis of orthogonal matching pursuit using the restricted isometry property, IEEE Trans. Info. Theory 56 (9) (2010) 4395–4401.

[46] D. Needell, J. Tropp, CoSaMP: Iterative signal recovery from incomplete and inaccurate samples, Appl. Comput. Harmon. Anal. 26 (3) (2009) 301–321.

[47] M. Figueiredo, R. Nowak, An EM algorithm for wavelet-based image restoration, IEEE Trans. Image Proc. 12 (8) (2003) 906–916.

[48] I. Daubechies, M. Defrise, C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Comm. Pure Appl. Math. 57 (2004) 1413–1457.

[49] E. J. Candès, J. K. Romberg, Signal recovery from random projections, in: Computational Imaging III at SPIE Electronic Imaging, Vol. 5674, SPIE, San Jose, CA, 2005, pp. 76–86.

[50] T. Blumensath, M. E. Davies, Iterative hard thresholding for compressed sensing, Appl. Comput. Harmon. Anal. 27 (3) (2009) 265–274.

[51] T. Blumensath, M. E. Davies, Sampling theorems for signals from the union of finite-dimensional linear subspaces, IEEE Trans. Info. Theory 55 (4) (2009) 1872–1882.

[52] C. Hegde, M. F. Duarte, V. Cevher, Compressive sensing recovery of spike trains using a structured sparsity model, in: Workshop on Signal Proc. with Adaptive Sparse Structured Representations (SPARS), Saint Malo, France, 2009.

[53] Y. M. Lu, M. N. Do, Sampling signals from a union of subspaces, IEEE Signal Proc. Mag. 25 (2) (2008) 41–47.

[54] J. Tropp, A. C. Gilbert, Signal recovery from partial information via orthogonal matching pursuit, IEEE Trans. Info. Theory 53 (12) (2007) 4655–4666.

[55] W. Dai, O. Milenkovic, Subspace pursuit for compressive sensing signal reconstruction, IEEE Trans. Info. Theory 55 (5) (2009) 2230–2249.

[56] V. F. Pisarenko, The retrieval of harmonics from a covariance function, Geophysical J. Royal Astronomical Soc. 33 (3) (1973) 347–366.

[57] G. L. Nemhauser, L. A. Wolsey, Integer and combinatorial optimization, Wiley-Interscience, 1999.

[58] M. F. Duarte, R. G. Baraniuk, Recovery of frequency-sparse signals from compressive measurements, in: Allerton Conf. Communication, Control, and Computing, Monticello, IL, 2010, pp. 599–606.

[59] M. F. Duarte, R. G. Baraniuk, Spectral compressive sensing, ECE Department Tech. Report TREE-1005, Rice University, Houston, TX (Feb. 2010).

[60] L. Jacques, C. De Vleeschouwer, A geometrical study of matching pursuit parametrization, IEEE Trans. Signal Proc. 56 (7) (2008) 2835–2848.

[61] M. A. Lexa, M. E. Davies, J. S. Thompson, Reconciling compressive sampling systems for spectrally-sparse continuous-time signals, IEEE Trans. Signal Proc. 60 (1) (2012) 155–171.

[62] Z. Yu, S. Hoyos, B. M. Sadler, Mixed-signal parallel compressed sensing and reception for cognitive radio, in: IEEE Int. Conf. Acoustics, Speech, and Signal Proc. (ICASSP), Las Vegas, NV, 2008, pp. 3861–3864.

[63] J. P. Slavinsky, J. N. Laska, M. A. Davenport, R. G. Baraniuk, The compressive multiplexer for multi-channel compressive sensing, in: IEEE Int. Conf. Acoustics, Speech, and Signal Proc. (ICASSP), Prague, Czech Republic, 2011.

[64] M. A. Davenport, M. B. Wakin, Reconstruction and cancellation of sampled multiband signals using discrete prolate spheroidal sequences, Appl. Comput. Harmon. Anal. 33 (3) (2012) 438–472.

[65] M. Mishali, Y. C. Eldar, Xampling: Signal acquisition and processing in unions of subspaces, IEEE Trans. Signal Proc. 59 (10) (2011) 4719–4734.

[66] J. A. Cadzow, Signal enhancement — A composite property mapping algorithm, IEEE Trans. Acoust., Speech, Signal Proc. 36 (1) (1988) 49–62.

[67] D. Potts, M. Tasche, Parameter estimation for exponential sums by approximate Prony method, Signal Processing 90 (5) (2010) 1631–1642.

[68] C. D. Austin, R. L. Moses, J. N. Ash, E. Ertin, On the relationship between sparse reconstruction and parameter estimation with model order selection, IEEE J. Selected Topics in Signal Proc. 4 (3) (2010) 560–570.

[69] A. C. Fannjiang, W. Liao, Coherence-pattern guided compressive sensing with unresolved grids, SIAM J. Imaging Sciences 5 (1) (2012) 179–202.

[70] M. F. Duarte, Localization and bearing estimation via structured sparsity models, in: IEEE Statistical Signal Proc. Workshop (SSP), Ann Arbor, MI, 2012.

[71] D. Donoho, A. Maleki, A. Montanari, Message passing algorithms for compressed sensing, Proc. Nat. Acad. Sci. 106 (45) (2009) 18914–18919.

[72] J. Tropp, S. Wright, Computational methods for sparse solution of linear inverse problems, Proc. IEEE 98 (6) (2010) 948 –958.

[73] G. Tang, B. N. Bhaskar, P. Shah, B. Recht, Compressed sensing off the grid, Preprint (July 2012).

[74] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, R. G. Baraniuk, Distributed compressive sensing, ECE Department Tech. Report TREE-0612, Rice University, Houston, TX (Nov. 2006).

[75] S. A. Geršgorin, Über die abgrenzung der eigenwerte einer matrix, Izv. Akad. Nauk SSSR Ser. Fiz.-Mat. 6 (1931) 749–754.