Recovery of Frequency-Sparse Signals from Compressive Measurements

Marco F. Duarte
Program in Applied and Computational Mathematics
Princeton University

Richard G. Baraniuk
Department of Electrical and Computer Engineering
Rice University

***Abstract*—Compressive sensing (CS) is a new approach to simultaneous sensing and compression for sparse and compressible signals. While the discrete Fourier transform has been widely used for CS of frequency-sparse signals, it provides optimal sparse representations only for signals with components at integral frequencies. There exist redundant frames that provide compressible representations for frequency-sparse signals, but such frames are highly coherent and severely affect the performance of standard CS recovery. In this paper, we show that by modifying standard CS recovery algorithms to prevent coherent frame elements from being present in the signal estimate, it is possible to bypass the shortcomings introduced by the coherent frame. The resulting algorithm comes with theoretical guarantees and is shown to perform significantly better for frequency-sparse signal recovery than its standard counterparts. The algorithm can also be extended to similar settings that use coherent frames.**

I. Introduction

The emerging theory of *compressive sensing* (CS) combines digital data acquisition with digital data compression to enable a new generation of signal acquisition systems that operate at sub-Nyquist rates. Rather than acquiring N samples $\mathbf{x} = [\mathbf{x}[1]\ \mathbf{x}[2]\ \ldots\ \mathbf{x}[N]]^T$ of an analog signal at the Nyquist rate, 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]. 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 signals as well as to so-called *compressible* signals that are not exactly sparse but can be closely approximated as such. Sparse signals have coefficients θ that, when sorted, decay according to a power law $|\theta[i]| < Ci^{-1/p}$, $p \leq 1$; the smaller the decay exponent p, the faster the decay and the better the recovery performance we can expect from CS.

A great many applications feature smooth or modulated signals that can be modeled as a linear combination of K sinusoids [3]:

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

where $a_k \in \mathbb{C}$ are their coefficients and ω_k are their frequencies. When the sinusoids are of infinite extent, such signals have an exactly K-sparse representation in terms of the discrete-time Fourier transform (DTFT):

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

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 analysis and CS recovery has been the discrete Fourier transform (DFT):

$$\mathbf{X}[l] = \sum_{n=1}^{N} \mathbf{x}[n] e^{-j2\pi ln/N},\ 1 \leq l \leq N.$$

The DFT $\mathbf{X}[l]$ of N consecutive samples from the smooth signal model (1) can be obtained from the DTFT (2) by first convolving with the Dirichlet kernel and then sampling:

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

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

$$D_N(x) = \sum_{k=0}^{N-1} e^{jkx} = e^{j\omega(N-1)/2} \frac{\sin(Nx/2)}{\sin(x/2)}.$$

Unfortunately, the DFT coefficients in (3) do not share the same sparsity property as the DTFT coefficients in (2), except in the (contrived) case when 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 smooth signals *not sparse* in the

DFT domain, but, owing to the slow asymptotic decay of the Dirichlet kernel, 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.

In this paper, we develop new CS recovery algorithms for practical 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 of sinusoids that we will call a *DFT frame*. Unfortunately, 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 increases recovery performance. However, simultaneously, the frame becomes increasingly *coherent* [4], which decreases recovery performance. In order to optimize this tradeoff, we leverage the last few decades of progress in the field of spectral estimation [5] plus recent progress on model-based CS [6] and marry these techniques with a new class of greedy CS recovery algorithms. We refer to our general approach as *spectral compressive sensing* (SCS).

The primary 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. We will provide an analytical characterization of the number of measurements M required for stable SCS signal recovery under this model and will study the performance of the framework under parameter variations.

This paper is organized as follows. Section II provides the usual background on CS and model-based CS and summarizes existing schemes for parameter estimation of frequency-sparse signals. Section III develops our proposed SCS recovery algorithms, and Section IV presents our experimental results. Section V summarizes related work in this area, and Section VI gives conclusions.

II. Background

A. Sparse approximation

A signal $\mathbf{x} \in \mathbb{R}^N$ is *K-sparse* ($K \ll N$) in a basis or frame[1] 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 counts the number of nonzero entries in the vector. Signal compression often relies on the existence of a known basis or frame Ψ such that for the signal of interest $\mathbf{x}$ there exists 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 [7], orthogonal matching pursuit (OMP) [4], and iterative thresholding [8]. The approximation performance of these algorithms is directly tied to the *coherence* of the frame Ψ, defined as $\mu(\Psi) = \arg\max_{1 \le i,j \le N} |\langle \psi_i, \psi_j \rangle|$, where ψ_i denotes the i^{th} column of Υ. For example, OMP successfully obtains a K-sparse signal representation if $\mu(\Psi) \le \frac{1}{16(K-1)}$ [4].

B. Compressive sensing

Compressive Sensing (CS) is an efficient acquisition framework for signals that are sparse or compressible in a basis or frame Ψ. Rather than uniformly sampling 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 smallest possible measurement vector $\mathbf{y}$. 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 $\Upsilon = \Phi\Psi$.

The Restricted Isometry Property (RIP) [1] has been proposed to measure the fitness of a matrix Υ for CS.

Definition 2.1: The *K-restricted isometry constant* for the matrix Υ, denoted by δ_K, is the smallest nonnegative number such that, for all $\theta \in \mathbb{R}^N$ with $\|\theta\|_0 = K$,

$$(1 - \delta_K)\|\theta\|_2^2 \le \|\Upsilon\theta\|_2^2 \le (1 + \delta_K)\|\theta\|_2^2.$$

A matrix has the RIP if $\delta_K > 0$. Since calculating δ_K for a given matrix Φ requires a combinatorial amount of computation, random matrices have been advocated. 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 $K \le M/\log(N/M)$. The same is true of matrices following Rademacher (± 1) or more general subgaussian distributions.

C. Frequency-sparse signals

Recall from the introduction that frequency-sparse signals of the form (1) have a sparse DTFT (2). How-

[1]Recall that a frame is a matrix $\Psi \in \mathbb{R}^{D \times N}$, $D < N$, such that for all vectors $\mathbf{x} \in \mathbb{R}^D$, $A\|\mathbf{x}\|_2 \le \|\Psi^T\mathbf{x}\|_2 \le B\|\mathbf{x}\|_2$ with $0 < A \le B < \infty$. A frame is a generalization of the concept of a basis to sets of possibly linearly dependent vectors.

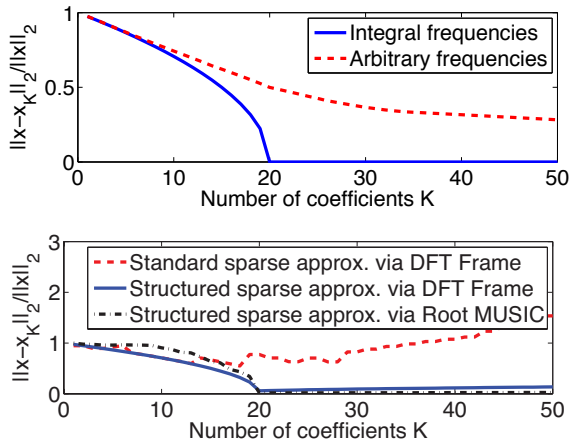

Fig. 1. *Performance of K-term sparse approximation algorithms on a signal containing 20 complex sinusoids. We plot the approximation error vs. the number of terms in the approximation.* Top: *Orthonormal DFT basis approximation performance is perfect for signals with exclusively integral frequencies and atrocious for signals with non-integral frequencies.* Bottom: *Three potential approximation strategies for sinusoids with arbitrary frequencies. Standard sparse approximation using a DFT frame $\Psi(c)$, $c = 10$, performs even worse than the DFT basis. Structured sparse approximations based on a DFT frame and Root MUSIC (described in Section III) perform much better.*

ever, to exploit sparsity in CS, we require a discrete signal representation; thus, the DFT has been the tool of choice for frequency-sparse signals. Additionally, the fast Fourier transform (FFT) provides a very efficient algorithm to calculate the DFT coefficients of a signal. The DFT of an length-N signal $\mathbf{x}$ can be obtained as its representation in the orthonormal *DFT basis* $\mathbf{F}$, which has entries $\mathbf{F}[p,q] = e^{-j2\pi pq/N}/\sqrt{N}$, $1 \le p, q \le N$.

The DFT basis preserves the sparsity of the DTFT for frequency-sparse signals only when the signal components have *integral frequencies* of the form $2\pi l/N$, where l is an integer. Otherwise, the situation is decidedly more complicated due to the spectral leakage induced by windowing (convolution by the Dirichlet kernel). To illustrate the difficulty, Fig. 1(a) plots the sparse approximation error of signals of length $N = 1024$ containing 20 complex sinusoids of both integral and non-integral frequencies using the DFT basis. As expected, sparse approximation using a DFT basis fails miserably for signals with non-integral frequencies.

The naïve way to combat spectral leakage is to employ a redundant frame, which we term a *DFT frame*, that provides a finer sampling of the DTFT coefficients of the signal $\mathbf{x}$. Let $c \in \mathbb{N}$ denote the frequency oversampling factor for the DFT frame, and define the frequency sampling interval $\Delta = 2\pi/cN \in (0, 2\pi/N]$. Also let $\mathbf{e}(\omega) = N^{-1/2}[1 \;\; e^{j\omega} \;\; \dots \;\; e^{j\omega(N-1)}]^T$ denote a normalized vector containing regular samples of a complex sinusoid with angular frequency $\omega \in [0, 2\pi)$. The DFT frame with oversampling factor c is then defined as $\Psi(c) = [\mathbf{e}(0) \; \mathbf{e}(\Delta) \; \dots \; \mathbf{e}(2\pi - \Delta)]^T$; the corresponding signal representation $\theta = \Psi(c)^T\mathbf{x}$ provides cN equispaced samples of the signal's DTFT.

We can use the DFT frame $\Psi(c)$ to obtain sparse approximations for frequency-sparse signals with components at arbitrary frequencies; as the frequency oversampling factor c increases, the K-sparse approximation provided by $\Psi(c)$ becomes increasingly accurate. The following lemma is proven in [9].

Lemma 2.1: 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 $\|\theta_K\|_0 = K$. Denote $\mathbf{a} = [a_1 \;\; \dots \;\; a_K]^T$. The best K-term approximation error for $\mathbf{x}$ in the frame $\Psi(c)$ obeys

$$\|\mathbf{x} - \mathbf{x}_K\|_2 \le \sqrt{1 - |D_N(\pi/cN)/N|^2}\|\mathbf{a}\|_1.$$

Unfortunately, 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:

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

Due to this tradeoff, the maximum frequency oversampling factor that still supports sparse representation of K-sparse signals is

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

where $D_N^{-1}(\cdot)$ denotes the inverse of the Dirichlet kernel within the interval $[0, 2\pi/N]$. 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 1(b) demonstrates the performance of standard sparse approximation of the same signal with arbitrary frequencies as in Fig. 1(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.

D. 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 be leveraged by an algorithmic *model-based CS* framework to exploit signal structure during recovery [6]. The framework has three components:

1) Structured sparsity model: The set Σ_K of all N-length, K-sparse signals is the union of the $\binom{N}{K}$, K-dimensional subspaces aligned with the coordinate axes in $\mathbb{R}^N$. A *structured sparsity model* $\mathcal{M}_K$ endows the K-sparse signal $\mathbf{x}$ with additional structure that allows only certain K-dimensional subspaces from Σ_K and disallows others. Signals from $\mathcal{M}_K$ are called K-structured sparse.

2) Measurement reduction: If we know that the signal $\mathbf{x}$ being acquired is K-structured sparse, then we can relax the RIP constraint on the CS measurement matrix Υ so that it holds only for signals in $\mathcal{M}_K$. We denote this new property as $\mathcal{M}_K$-RIP; it is sufficient to achieve stable recovery from the measurements $\mathbf{y} = \Upsilon\theta$. We can thus reduce the number of random measurements required to $M = \mathcal{O}(\log m_K)$, where m_K is the number of subspaces in $\mathcal{M}_K$.

3) Structured sparse approximation: The $\mathcal{M}_K$-RIP is sufficient for robust recovery of structured sparse signals using *model-based CS recovery algorithms*. These algorithms replace the standard optimal K-sparse approximation performed by 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 $\mathcal{M}_K$.

E. Parameter estimation for frequency-sparse signals

The goal of CS is to identify the values and locations of the nonzero coefficients of a sparse 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 differentiating the signal's actual nonzero coefficients from spurious estimates; such spurious coefficients are often modeled as recovery noise.

Thus, we now briefly review the extensive prior work in parameter estimation for frequency-sparse signals embedded in noise [5]. 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.

1) Periodogram-based methods: The maximum likelihood estimator (MLE) of the amplitude A of a single sinusoid when the frequency ω is known is given by the DTFT of $\overline{\mathbf{x}}$, the zero-padded, infinite length version of the signal $\mathbf{x}$, at frequency ω: $\widehat{A} = \frac{1}{N}\overline{\mathbf{X}}(\omega) = \langle \mathbf{e}(\omega), \mathbf{x} \rangle$. [5] 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 $\overline{\mathbf{x}}$: $\widehat{\omega} = \arg\sup_\omega |\overline{\mathbf{X}}(\omega)| = \arg\sup_\omega |\langle \mathbf{e}(\omega), \mathbf{x} \rangle|$. This approach is often described as the *periodogram method* for parameter estimation [5]. This simple estimator can be extended to the multiple sinusoid setting by performing combinatorial hypothesis testing [5].

For frequency-sparse signals with components at integral frequencies, the signal's representation in the DFT basis provides the information needed by the MLEs above; in this case, the parameter estimation problem is equivalent to sparse approximation in the DFT basis. This equivalence can also be extended to frequency-sparse signals whose component frequencies are included in an oversampled DTFT sampling grid by using a DFT frame instead.

2) Window-based methods: 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 [5]. However, windowing can also degrade the spectral analysis resolution, making it more difficult to identify frequency-sparse signal components with similar frequencies.

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 [5]. Such algorithms provide improved resolution of the parameters of a frequency-sparse signal by estimating the principal components of the signal's autocorrelation matrix in order to find the dominant signal modes in the frequency domain. Example algorithms include Pisarenko's method, MUSIC, root MUSIC, and ESPRIT. 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.

We can interpret the line spectral estimation process $\mathbb{L}$ as a K-sparse approximation algorithm $\mathbb{T}_l(\mathbf{x}, K)$ 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 coefficient values using the DTFT as described earlier. We note that most line spectral estimation algorithms provide a tradeoff between estimation accuracy and computational complexity in the selection of the window size used to estimate the autocorrelation matrix $\mathbf{R}_{\mathbf{xx}}$.

III. Spectral Compressive Sensing

We are now in a position to develop new SCS recovery algorithms that are especially tailored to frequency-sparse signals of arbitrary frequencies. We will develop two sets of algorithms based on the periodogram and line spectral estimation algorithms from Section II-E.

A. SCS via periodogram

To alleviate the performance-sapping coherence of the redundant DFT frame, we marry it with the model-based CS framework of Section II-D that forces the signal approximation to contain linear combinations of only incoherent frame elements.

1) Structured signal model: Our structured signal model $\mathcal{T}_{K,c,\mu}$ requires that the components of the frequency-sparse signal are incoherent with each other; i.e., $\mathbf{x} = \sum_{i=1}^K a_i \mathbf{e}(d_i\Delta)$, with $d_i \in \{0, \ldots, cN-1\}$

and $|\langle \mathbf{e}(d_i\Delta), \mathbf{e}(d_j\Delta)\rangle| \leq \mu$, the maximal coherence allowed, for all $1 \leq i, j \leq K$. The coherence restriction imposes a lower bound on the frequency spacing between any two sinusoids present in a recoverable signal. We note that the limitations on the sparse approximation algorithms discussed in Sect. II-A require the use of a signal model $\mathcal{T}_{K,c,\mu}$ obeying $\mu \leq \frac{1}{16(K-1)}$. In words, the signals in the model must have components that are mutually incoherent; however, the frame $\mu(\Psi(c))$ used for recovery can be much more coherent.

2) Structured sparse approximation: We can modify standard sparse approximation algorithms to avoid selecting highly coherent pairs of elements of the DFT frame $\Psi(c)$. Our structured sparse approximation algorithm $\widehat{\theta} = \mathbb{T}(\theta, K, c, \mu)$ is an adaptation of that used for the refractory model of [10] and can be implemented as an integer program.

To reduce the computational complexity of $\mathbb{T}(\theta, K, c, \mu)$, we propose a heuristic that relies on *frequency inhibition*. The heuristic $\mathbb{T}_h(\theta, K, c, \mu)$ obtains the projection of $\mathbf{x}$ on the elements of the DFT frame $\theta = \Psi(c)^T\mathbf{x}$ and searches for the coefficient $\theta(d)$ with the largest magnitude. Once a coefficient is selected, the algorithm inhibits all coefficients for neighboring sinusoids (i.e., indices d' for which $D_N(\Delta(d - d')) < \mu$) by setting those coefficients to zero. We then repeat the process by searching for the next largest coefficient in magnitude until K coefficients have been selected or all coefficients are zero. This heuristic has complexity $O(cKN\log(cN))$ and offers very good performance for sparse approximation of arbitrary frequency-sparse signals, as shown in Fig. 1(b). Our experimental results in Sect. IV employ this heuristic.

3) Model-based recovery algorithm: The model-based IHT algorithm of [6] is particularly amenable to modification to incorporate our heuristic frequency-sparse approximation algorithm. Due to the redundancy of the frame $\Psi(c)$, we perform some minor surgery on the algorithm: we replace the matrix Φ by the matrix product $\Phi\Psi$ and multiply the signal estimate $\mathbf{b}$ by the Gramian matrix of the frame each time so that coherent frame elements featuring coefficients of opposing signs can cancel each other. The modified algorithm, which we dub *spectral iterative hard thresholding* (SIHT), is unfurled in Algorithm 1. It inherits a strong performance guarantee from standard IHT [8]: If the matrix Φ has the $\mathcal{T}_{2K,c,\mu}$-RIP, then we have

$$\|\mathbf{x} - \widehat{\mathbf{x}}\|_2 \leq C_0 \frac{\|\mathbf{x} - \Psi\mathbb{T}(\Psi^T\mathbf{x}, K, \mu)\|_2}{K^{1/2}} + C_1\epsilon. \quad (4)$$

We note that for signals that are frequency-sparse and composed of incoherent sinusoids with frequencies of the form $l\Delta$, we have $\Psi\mathbb{T}(\Psi^T\mathbf{x}, K, c, \mu) = \mathbf{x}$, meaning that recovery from noiseless measurements is exact.

Algorithm 1 Spectral Iterative Hard Thresholding

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

4) Required number of measurements: By counting the number of subspaces t_K that compose the signal model $\mathcal{T}_{K,c,\mu}$, we calculate the required number of random measurements to be $M = \mathcal{O}\left(K\log(cN/K)\right)$ [9]. In words, the number of measurements needed corresponds to that required by standard CS for a K-sparse, cN-dimensional signal. That is, SIHT can employ a coherent DFT frame without paying the penalty on the number of measurements required for stable signal recovery. We demonstrate below in Sect. IV that, in practice, SIHT offers a significant reduction in the number of measurements over standard CS recovery algorithms.

B. SCS via line spectral estimation

While the combination of a redundant frame and the coherence-inhibiting structured sparsity model yields an improvement in the performance of SIHT, 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, which are based on the periodogram.

Fortunately, we can leverage the now-classical line spectral estimation methods [5] that return a set of K dominant frequencies for the input signal, with K being a controllable parameter. Since these methods do not rely on redundant frames, we do not need to leverage the features of model-based CS that control the effect of coherence. We simply employ the spectral estimator-based sparse approximation algorithm $\widehat{\mathbf{x}} = \mathbb{T}_l(\mathbf{x}, K)$ from Section II-E3 instead of thresholding in the standard IHT algorithm, and we dub the result *SIHT via line spectral estimation*. While analytical results for this modification have proven difficult to establish, we show experimentally below that its performance matches that of the periodogram-based SIHT algorithm while exhibiting a much simpler implementation. We choose to use the root MUSIC algorithm in this paper.

IV. Experimental Results

In this section, we report experimental results for the performance of the periodogram-based and root MUSIC-based SIHT recovery algorithms as compared to standard CS recovery using the IHT algorithm. This comparison effectively illustrates the performance improvement afforded by the structured sparsity model. We probe the robustness of the algorithms to varying amounts of measurement noise and varying frequency oversampling factors c. We also test the algorithms on a real-world communications signal. A Matlab toolbox containing implementations of the SIHT 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 containing $K = 20$ complex-valued sinusoids. For each 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 $\omega_i,\ \omega_j,\ 1 \leq i, j \leq K,\ i \neq j$ are spaced apart by at least $10\pi/1024$ radians/sample. 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 via periodogram with frequency oversampling factor $c = 10$ and maximum allowed coherence $\mu = 0.1$ (Algorithm 1), and SIHT via Root MUSIC. We use a window size $W = N/10$ in Root MUSIC to estimate the autocorrelation matrix $\mathbf{R}_{\mathbf{xx}}$. We study 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 frequency is half-way in between two consecutive integral frequencies, resulting in maximal spectral leakage. The results are summarized in Fig. 2 and show first that the average performance of standard IHT is very close to its worst-case performance, and second that both SIHT algorithms perform significantly better on the same average case regime. We also note that SIHT works well in the average case regime even though the resulting signals do not exactly match the sparse-in-DFT-frame assumption. Thus, the proposed algorithms are robust to mistmatch in the values for the frequencies in the signal model (1). We use this experimental setup in the rest of this section, but we restrict ourselves to the average case regime.

Our second experiment tests the robustness of SIHT 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^2 to each measurement. For each value of σ, we perform 1000 independent noise trials; in each trial, we generate the matrices Φ and signals $\mathbf{x}$ randomly as in the previous experiment. Fig. 3 shows the average norm of the recovery error as a function of the noise variance σ; the linear relationship indicates stability to additive noise to the measurements, as predicted in (4).

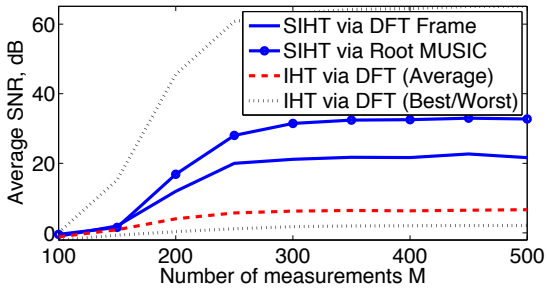

Fig. 2. *Performance of CS signal recovery via IHT using the orthonormal DFT basis, SIHT via periodogram, and SIHT via Root MUSIC. We use signals of length $N = 1024$ containing $K = 20$ complex-valued sinusoids. The dotted lines indicate the performance of IHT via the orthonormal DFT basis for the best 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). The performance of IHT for arbitrary frequencies is close to its worst-case performance, while both SIHT algorithms perform significantly better for arbitrary frequencies. All quantities are averaged over 100 independent trials.*

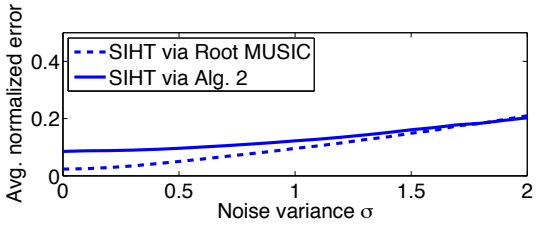

Fig. 3. *Performance of SIHT via periodogram and SIHT via Root MUSIC for signal recovery from noisy measurements. We use signals of length $N = 1024$ containing $K = 20$ complex-valued sinusoids and take $M = 300$ measurements. We add noise of varying variances σ and calculate the average normalized 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.*

Our third experiment studies the impact of the frequency oversampling factor c on the performance of SIHT. We use the same matrix and signal setup as in the previous experiment: we execute 10000 independent trials for each value of c. The results, shown in Fig. 4, 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 recovery performance, as well as between the oversampling factor M/K (dependent on $\log c$) and the recovery performance. These results also experimentally confirm Lemma 2.1.

Our fourth experiment tests the capacity of the IHT and SIHT algorithms to resolve neighboring frequencies

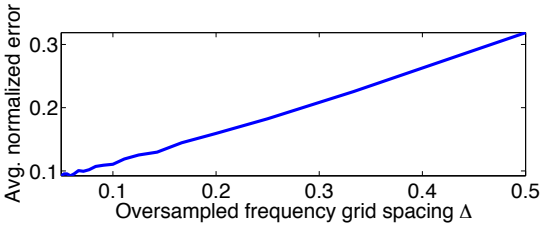

Fig. 4. *Performance of SIHT via periodogram for signal recovery under varying grid spacing resolutions $\Delta = 2\pi/cN$. We use signals of length $N = 1024$ containing $K = 20$ complex-valued sinusoids and take $M = 300$ measurements. We average the recovery error over 10000 independent trials. There is a linear dependence between the granularity of the DFT frame and the norm of the recovery error.*

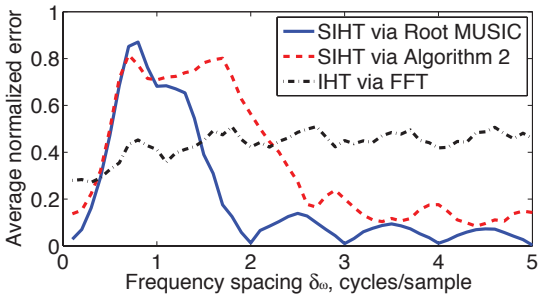

Fig. 5. *Performance of IHT and SIHT algorithms for frequency-sparse signals with components at closely spaced frequencies. The signal has length $N = 1024$ and contains 2 real sinusoids ($K = 4$) with frequencies separated by δ_ω. IHT and SIHT are used to recover the signal from $M = 100$ measurements. The use of line spectral estimation in SIHT significantly improves its performance. Additionally, standard IHT outperforms SIHT only when the signal of interest is not contained in the class of frequency-sparse signals with incoherent components.*

in frequency-sparse signals. For this experiment, the signal consists of 2 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 of the signal. We measure the performance of standard IHT with the DFT basis, SIHT via periodogram with frequency oversampling factor $c = 10$ and maximum allowed coherence $\mu = 0.1$, and SIHT via Root MUSIC, 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. For each value of δ_ω, we execute 100 independent trials as detailed in previous experiments. The results, shown in Fig. 5, 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 only outperforms the SIHT algorithms when the signal does not belong in the class of frequency-sparse signals with incoherent components (that is, very small frequency spacing δ_ω).

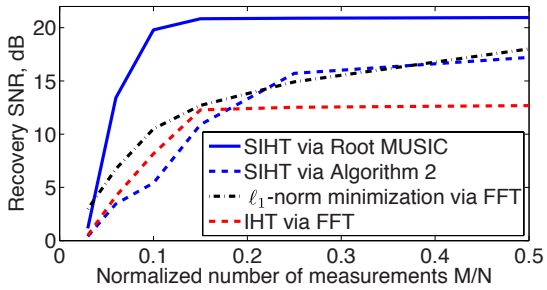

Fig. 6. *Performance of ℓ_1-norm minimization, IHT via orthonormal DFT, and SIHT via Root MUSIC algorithms on a real-world AM communications signal of length $N = 32768$ for a varying number of measurements M. SIHT significantly outperforms its standard counterparts.*

Our last experiment tests the performance of the IHT and SIHT algorithms on a real-world signal. We use the amplitude modulated (AM) signal from [3, Figure 7], measured in a lab, and simulate its acquisition using the random demodulator [3]. The signal has length $N = 32768$ samples. We recover the signal in half-overlapping blocks of length $N' = 1024$ using the IHT and SIHT algorithms; we also use the ℓ_1-norm minimization algorithm used in [3] for reference. We set the target signal sparsity to $K = N/100$ in the IHT and SIHT algorithms. The recovered AM signals are then demodulated, and the recovery performance is measured in terms of the distortion against the message obtained by demodulating the signal at Nyquist rate. We average the performance over 20 trials for the random demodulator chipping sequence. The results in Fig. 6 shows that SIHT consistently outperforms standard CS recovery algorithms.

To summarize, our experiments have shown that SIHT achieves significantly improved signal recovery performance for the overwhelming majority of frequency-sparse signals when compared with standard CS recovery algorithms and sparsity bases. The two SIHT algorithms we introduced and tested inherit some attractive properties from their standard counterparts, including robustness to model mismatch and measurement noise.

V. Related work

A recent paper [11] independently studied the poor performance of standard CS recovery algorithms on frequency-sparse signals when the DFT basis is used. The paper provides a generic framework for 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 use in SCS.

There also exists a related body of related work on compressive acquisition of signals governed by a small number of continuous-valued parameters. The Xampling framework [12] enables recovery of analog signals from sampling rates below those dictated by the Nyquist theorem by using sparsity coupled with sampling theory and analog signal processing tools. In finite rate of innovation (FROI) sampling [13], certain classes of signals governed by a small number of parameters can be acquired by uniformly sampling them using a specially designed kernel; the samples are then processed to obtain an annihilating filter, which in turn is used to estimate the values of the parameters. The application of FROI to frequency-sparse 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, line spectral estimation algorithms have been proposed to extend FROI to noisy sampling settings [14], albeit without performance guarantees.

In contrast to the Xampling and FROI frameworks, which focus on analog signal representations, our work in this paper aims to augment the standard CS framework [1, 2], which works on signal samples directly, to succeed for a wider class of frequency-sparse signals.

VI. 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 fuses a redundant frame of sinusoids corresponding to an oversampled frequency grid 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 the SIHT algorithms 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 recover frequency-sparse signals containing arbitrarily close frequencies with low computational complexity. As Fig. 2 indicates, the SIHT algorithms significantly outperform SIHT using the orthonormal DFT basis (up to 25dB in the figure).

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 signals are well suited for manifold models when only one component is present in the signal [15], they fall short for linear combinations of a varying number of components. Following SCS, we can construct a frame whose elements represent a uniform sampling of the manifold in parameter space. However, when the manifold is very smooth, the resulting frame will be highly coherent, limiting the performance of sparse approximation algorithms. By posing coherence-inhibiting models, we can enable accurate recovery of sparse signals under the redundant frame.

Acknowledgements

Thanks to Jason Laska for facilitating the data and code for the example of Fig. 6, and to Volkan Cevher, Yuejie Chi, Mark Davenport, Yonina Eldar, Chinmay Hegde, Joel Tropp, and Cédric Vonesch for valuable discussions. This paper is dedicated to the memory of Dennis M. Healy; his insightful discussions with us inspired much of this project. This work was 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 MURI W911NF-07-1-0185 and W911NF-09-1-0383, and the TI Leadership Program. MFD was also supported by NSF Supplemental Funding DMS-0439872 to UCLA-IPAM, P.I. R. Caflisch.

References

[1] E. J. Candès, "Compressive sampling," in *Proc. International Congress of Mathematicians*, vol. 3, Madrid, Spain, 2006, pp. 1433–1452.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, Sept. 2006.

[3] J. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond Nyquist: Efficient sampling of bandlimited signals," *IEEE Trans. Info. Theory*, vol. 56, pp. 1–26, Jan. 2010.

[4] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

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

[6] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Info. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[7] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. on Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1998.

[8] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, Nov. 2009.

[9] M. F. Duarte and R. G. Baraniuk, "Spectral compressive sensing," 2010, preprint.

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

[11] Y. Chi, L. Scharf, A. Pezeshki, and R. Calderbank, "The sensitivity to basis mismatch of compressed sensing in spectrum analysis and beamforming," in *Workshop on Defense Applications of Signal Processing (DASP)*, Lihue, HI, Oct. 2009.

[12] Y. Eldar and M. Mishali, "Blind multi-band signal reconstruction: Compressed sensing for analog signals," *IEEE Trans. Signal Processing*, vol. 57, no. 3, pp. 993–1009, Mar. 2009.

[13] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Proc.*, vol. 50, no. 6, pp. 1417–1428, June 2002.

[14] I. Maravić and M. Vetterli, "Sampling and reconstruction of signals with finite innovation in the presence of noise," *IEEE Trans. Signal Proc.*, vol. 53, no. 8, pp. 2788–2805, Aug. 2005.

[15] R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," *Found. Comp. Math.*, vol. 9, pp. 51–77, Feb. 2009.