

# Compressive Sensing of Analog Signals Using Discrete Prolate Spheroidal Sequences

Mark A. Davenport<sup>s</sup> and Michael B. Wakin<sup>c\*</sup>

<sup>s</sup>Department of Statistics, Stanford University,  
390 Serra Mall, Sequoia Hall, Stanford, CA 94305, USA. Email: markad@stanford.edu

<sup>c</sup>Department of Electrical Engineering and Computer Science, Colorado School of Mines,  
1500 Illinois St., Golden, CO 80401, USA. Email: mwakin@mines.edu

September 2011

## Abstract

Compressive sensing (CS) has recently emerged as a framework for efficiently capturing signals that are sparse or compressible in an appropriate basis. While often motivated as an alternative to Nyquist-rate sampling, there remains a gap between the discrete, finite-dimensional CS framework and the problem of acquiring a continuous-time signal. In this paper, we attempt to bridge this gap by exploiting the Discrete Prolate Spheroidal Sequences (DPSS's), a collection of functions that trace back to the seminal work by Slepian, Landau, and Pollack on the effects of time-limiting and bandlimiting operations. DPSS's form a highly efficient basis for sampled bandlimited functions; by modulating and merging DPSS bases, we obtain a dictionary that offers high-quality sparse approximations for most sampled multiband signals. This multiband modulated DPSS dictionary can be readily incorporated into the CS framework. We provide theoretical guarantees and practical insight into the use of this dictionary for recovery of sampled multiband signals from compressive measurements.

**Keywords:** Compressive sensing, multiband signals, Discrete Prolate Spheroidal Sequences, Fourier analysis, sampling, block-sparsity, approximation, signal recovery, greedy algorithms

## 1 Introduction

### 1.1 Compressive sensing of analog signals

In recent decades the digital signal processing community has enjoyed enormous success in developing hardware and algorithms for capturing and extracting information from signals. Capitalizing on the early work of Whitaker, Nyquist, Kotelnikov, and Shannon on sampling and representation of continuous signals, signal processing has moved from the analog to the digital domain and ridden the wave of Moore's law. Digitization has enabled the creation of sensing and processing systems that are more robust, flexible, cheaper and, therefore, more ubiquitous than their analog counterparts.

The foundation of this progress has been the Nyquist sampling theorem, which states that in order to perfectly capture the information in an arbitrary continuous-time signal  $x(t)$  with bandlimit

---

\*Corresponding author. Phone: (303) 273-3607. Fax: (303) 273-3602.

This work was partially supported by NSF grants DMS-1004718 and CCF-0830320, DARPA grant FA8650-08-C-7853, and AFOSR grant FA9550-09-1-0465. A preliminary version of this work was first presented at the 2011 Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS'11) [24].

$\frac{B_{\text{nyq}}}{2}$  Hz, we must sample the signal at its Nyquist rate of  $B_{\text{nyq}}$  samples/sec. This requirement has placed a growing burden on analog-to-digital converters as applications that require processing signals of ever-higher bandwidth lead to ever-higher sampling rates. This pushes these devices toward a physical barrier, beyond which their design becomes increasingly difficult and costly [76].

In recent years, *compressive sensing* (CS) has emerged as a framework that can significantly reduce the acquisition cost at a sensor [2, 10, 27]. CS builds on the work of Candès, Romberg, and Tao [13] and Donoho [27], who showed that a signal that can be compressed using classical methods such as transform coding can also be efficiently acquired via a small set of nonadaptive, linear, and usually randomized measurements.

There remains, however, a prominent gap between the theoretical framework of CS, which deals with acquiring finite-length, discrete signals that are sparse or compressible in a basis or dictionary, and the problem of acquiring a continuous-time signal. Previous work has attempted to bridge this gap by employing two very different strategies. First, works such as [72] have employed a simple *multitone* analog signal model that maps naturally into a finite-dimensional sparse model. Although this assumption allows the reconstruction problem to be formulated directly within the CS framework, the multitone model can be unrealistic for many analog signals of practical interest. Alternatively, other authors have considered a more plausible *multiband* analog signal model that is also amenable to sub-Nyquist sampling [8, 33, 34, 46, 47, 74]. These works, however, have involved customized sampling protocols and reconstruction formulations that fall largely outside of the standard CS framework. Indeed, some of this body of literature and many of its underlying ideas actually pre-date the very existence of CS.

## 1.2 Contributions and paper organization

In this paper, we bridge this gap in a different manner. Namely, we show that when dealing with finite-length windows of samples, it is possible to map the multiband analog signal model—in a very natural way—into a finite-dimensional sparse model. One can then apply many of the standard theoretical tools of CS to develop algorithms for both recovery as well as compressive domain processing of multiband signals.

Our work actually rests on ideas that trace back to the classical signal processing literature and the study of time-frequency localization. The Weyl-Heisenberg uncertainty principle states that a signal cannot be simultaneously localized on a finite interval in both time and frequency. A natural question is to what extent it is possible to concentrate a signal  $x(t)$  and its continuous-time Fourier transform (CTFT)  $X(F)$  near finite intervals. In an extraordinary series of papers from the 1960s and 1970s, Slepian, Landau, and Pollack provide an in-depth investigation into this question [41, 42, 62, 64, 66]. The implications of this body of work have had a tremendous impact across a number of disciplines within mathematics and engineering, particularly in the field of spectral estimation and harmonic analysis (e.g., [68]). Very few of these ideas have appeared in the CS literature, however, and so one goal of this paper is to carefully explain—from a CS perspective—the natural role that these ideas can indeed play in CS.

We begin this paper in Sections 2 and 3 with a description of our problem setup and a survey of the necessary CS background material. In Section 4, we introduce the *multitone* and *multiband* analog signal models. We then discuss how sparse representations for multiband signals can be incorporated into the CS framework through the use of *Discrete Prolate Spheroidal Sequences* (DPSS's) [64]. First described by Slepian in 1978, DPSS's form a highly efficient basis for sampled bandlimited functions. For the sake of clarity and completeness, we provide a self-contained review of the key results of from Slepian's work that are most relevant to the problem of modeling sampled multiband signals. We then explain how, by modulating and merging DPSS bases, one obtains a

dictionary that—to a very high degree of approximation—provides a sparse representation for most finite-length, Nyquist-rate sample vectors arising from multiband analog signals. We also explain why the qualifiers “approximation” and “most” in the preceding sentence are necessary; however, we characterize them formally and justify the use of the multiband modulated DPSS dictionary in practical settings.

In Section 5, we discuss the use of the multiband modulated DPSS dictionary for recovery of sampled multiband signals from compressive measurements. We discuss the implications (in terms of formulating reconstruction procedures and guaranteeing their performance) of the fact that our dictionary is not quite orthogonal; in fact, it may be undercomplete or overcomplete, depending on the setting of a user-defined parameter. We also provide theoretical guarantees for recovery algorithms that exploit the *block-sparse* nature of signal expansions in our dictionary. Ultimately, this allows us to guarantee that most finite-length sample vectors arising from multiband analog signals can—to a high degree of approximation—be recovered from a number of measurements that is proportional to the underlying information level (also known as the Landau rate [40]).

In Section 6, we present the results of a detailed suite of simulations for signal recovery from compressive measurements, illustrating the effectiveness of our proposed approaches on realistic signals. We show that the reconstruction quality achieved using the multiband modulated DPSS dictionary is far better than what is achieved using the discrete Fourier transform (DFT) as a sparsifying basis. These results confirm that a DPSS-based dictionary can provide a very attractive alternative to the DFT for sparse recovery. We conclude in Section 7 with a final discussion and directions for future work.

### 1.3 Relation to existing work

Although customized measurement and reconstruction schemes [8, 33, 34, 46, 47, 74] have previously been proposed for efficiently sampling multiband signals, we believe that our paper is of independent interest from these works, specifically because we restrict ourselves to operating within the finite-dimensional CS framework. There are a variety of plausible CS (and even non-CS) scenarios where a sparse representation of a finite-length Nyquist-rate sample vector would be useful, and it is this problem to which we devote our attention. This work may be of interest, for example, to any practitioner who has struggled with the lack of sparsity that the DFT dictionary provides even for pure sampled tones at “off-grid” frequencies. Moreover, as we discuss more fully in Section 2, several analog CS hardware architectures can be viewed as providing random projections of finite-length, Nyquist-rate sample vectors. Our formulation is compatible with these architectures and does not require a customized sampling protocol.

It is important to mention that we are not the first authors to recognize the potential role that DPSS’s (or their continuous-time counterparts, the Prolate Spheroidal Wave Functions, or PSWF’s [41, 42, 62, 66]) can play in CS. Izu and Lakey [39] have drawn an analogy between sampling bounds for multiband signals and classical results in CS, but not specifically for the purpose of using the finite-dimensional CS framework for sparse recovery of sample vectors from multiband analog signals. Gosse [36] has considered the recovery of smooth functions from random samples; however, this work focuses on a very different setting, employing a PSWF (not DPSS) dictionary, considering only baseband signals, and exploiting sparsity in a different way than our work. Senay et al. [59, 60] have also considered a PSWF dictionary for reconstruction of signals from nonuniform samples; however, this work also focuses on baseband signals and lacks formal approximation and CS recovery guarantees. Oh et al. [52] have employed a modulated DPSS dictionary for sampled bandpass signals; however, this work falls largely outside the standard CS framework and again lacks formal approximation and CS recovery guarantees of the type we provide. Finally, Sejdíć et al. [58]

have proposed a multiband modulated DPSS dictionary very similar to our own and a greedy algorithm for signal decomposition in that dictionary. However, this work is again not devoted to developing sparse approximation guarantees for sampled multiband signals. It focuses not on signal recovery but on identification of a communications channel, and the proposed reconstruction algorithm is not intended to exploit block-sparse structure in the signal coefficients. We hope that our paper will be a valuable addition to this nascent literature and help to encourage much further exploration of the connections between DPSS's, PSWF's, and CS.

## 1.4 Preliminaries

Before proceeding, we first briefly introduce some mathematical notation that we will use throughout the paper. We use bold characters to indicate finite-dimensional vectors and matrices. All such vectors and matrices are indexed beginning at 0, so that the first element of a length- $N$  vector  $\mathbf{x}$  is given by  $\mathbf{x}[0]$  and the last by  $\mathbf{x}[N-1]$ . We denote the Hermetian transpose of a matrix  $\mathbf{A}$  by  $\mathbf{A}^H$ . We use  $\|\cdot\|_p$  to denote the standard  $\ell_p$  norm. We also use  $\|\mathbf{x}\|_0 := |\text{supp}(\mathbf{x})|$  to denote the number of nonzeros of  $\mathbf{x}$ , and we say that  $\mathbf{x}$  is  $S$ -sparse if  $\|\mathbf{x}\|_0 = S$ . We use  $\mathbb{E}[x]$  to denote expected value of a random variable  $x$  and  $\mathbb{P}[E]$  to denote the probability of an event  $E$ . Finally, we adopt the convention  $j = \sqrt{-1}$  throughout the paper.

## 2 Mapping Analog Sensing to the Digital Domain

In the standard CS setting, one is concerned with recovering a finite-dimensional vector  $\mathbf{x} \in \mathbb{C}^N$  from a limited number of measurements. A typical first order assumption is that the vector  $\mathbf{x}$  is *sparse*, meaning that there exists some basis or dictionary  $\Psi$  such that  $\mathbf{x} = \Psi\alpha$  and  $\alpha$  has a small number of nonzeros, i.e.,  $\|\alpha\|_0 \leq S$  for some  $S \ll N$ . One then acquires the measurements

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e} \quad (1)$$

where  $\mathbf{A} \in \mathbb{C}^{M \times N}$  maps  $\mathbf{x}$  to a length- $M$  vector  $\mathbf{y}$  of complex-valued measurements, and where  $\mathbf{e}$  is a length- $M$  vector that represents *measurement noise* generated by the acquisition hardware. In the context of CS, one seeks to design  $\mathbf{A}$  so that  $M$  is on the order of  $S$  (the number of degrees of freedom of the signal) and potentially much smaller than  $N$ .

In the present work, however, we are concerned with the acquisition of a finite-length window of a complex-valued, continuous-time signal, which we denote by  $x(t)$ . Specifically, we suppose that we are interested in a time window of length  $T_w$  seconds and that we acquire the measurements

$$\mathbf{y} = \Phi(x(t)) + \mathbf{e}, \quad (2)$$

where  $\Phi$  is a linear *measurement operator* that maps functions defined on  $[0, T_w)$  to a length- $M$  vector  $\mathbf{y}$  of measurements and  $\mathbf{e}$  again represents measurement noise. We assume throughout this paper that  $x(t)$  is bandlimited with bandlimit  $\frac{B_{\text{nyq}}}{2}$  Hz, i.e., that  $x(t)$  has a continuous-time Fourier transform (CTFT)

$$X(F) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi Ft}dt$$

such that  $X(F) = 0$  for  $|F| > \frac{B_{\text{nyq}}}{2}$ . Additional assumptions on  $x(t)$  will be specified in Section 4.1.2.

Because we assume that  $x(t)$  is bandlimited and that the measurement process (2) takes place over a finite window of time, we restrict our attention to the problem of recovering the Nyquist-rate samples of  $x(t)$  over this time interval. Specifically, we let  $T_s \leq \frac{1}{B_{\text{nyq}}}$  denote a sampling interval (in

seconds) chosen to meet the minimum Nyquist sampling rate, and we let  $x[n]$  denote the infinite-length sequence that would be obtained by uniformly sampling  $x(t)$  with sampling period  $T_s$ , i.e.,  $x[n] = x(nT_s)$ . We are interested in a time window of length  $T_w$  seconds, during which there are  $N = \lceil \frac{T_w}{T_s} \rceil$  samples. We let  $\mathbf{x} = [x[0] \ x[1] \ \cdots \ x[N-1]]^T$  denote  $x[n]$  truncated to the  $N$  samples from 0 to  $N-1$ . This paper is specifically devoted to the problem of recovering  $\mathbf{x}$ , the vector of Nyquist-rate samples of  $x(t)$  on  $[0, T_w]$ ,<sup>1</sup> from compressive measurements  $\mathbf{y}$  of the form (2).

To facilitate this, we first note that the sensing model in (2) is clearly very similar to the standard CS model in (1). We briefly describe conditions under which these models are equivalent. Recall from the Shannon-Nyquist sampling theorem that  $x(t)$  can be perfectly reconstructed from  $x[n]$  since  $\frac{1}{T_s} \geq B_{\text{nyq}}$ . Specifically, we have the formula

$$x(t) = \sum_{n=-\infty}^{\infty} x[n] \text{sinc}(t/T_s - n), \quad (3)$$

where

$$\text{sinc}(t) = \begin{cases} \sin(\pi t)/(\pi t), & t \neq 0 \\ 1, & t = 0. \end{cases}$$

Observe that since  $\Phi$  is linear, we can express each measurement  $\mathbf{y}[m]$  in (2) simply as the inner product between  $x(t)$  and some *sensing functional*  $\phi_m(t)$ , i.e.,<sup>2</sup>

$$\mathbf{y}[m] = \langle \phi_m(t), x(t) \rangle + \mathbf{e}[m]. \quad (4)$$

In this case we can use (3) to reduce (4) to

$$\mathbf{y}[m] = \left\langle \phi_m(t), \sum_{n=-\infty}^{\infty} x[n] \text{sinc}(t/T_s - n) \right\rangle + \mathbf{e}[m] = \sum_{n=-\infty}^{\infty} x[n] \langle \phi_m(t), \text{sinc}(t/T_s - n) \rangle + \mathbf{e}[m]. \quad (5)$$

If we let  $\mathbf{A}$  denote the  $M \times N$  matrix with entries given by

$$\mathbf{A}[m, n] = \langle \phi_m(t), \text{sinc}(t/T_s - n) \rangle$$

and let  $\mathbf{w}$  denote the length- $M$  vector with entries given by

$$\mathbf{w}[m] = \sum_{\substack{n \leq -1 \\ n \geq N}} x[n] \langle \phi_m(t), \text{sinc}(t/T_s - n) \rangle, \quad (6)$$

then (2) reduces to

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w} + \mathbf{e}. \quad (7)$$

If the vector  $\mathbf{w} = \mathbf{0}$ , then (7) is exactly equivalent to the standard CS sensing model in (1). Moreover, if  $\mathbf{w}$  is not zero but is small compared to  $\mathbf{e}$ , then we can simply absorb  $\mathbf{w}$  into  $\mathbf{e}$  and again reduce (7) to (1).

---

<sup>1</sup>Note that our goal is to recover  $\mathbf{x}$ , which of course carries useful information about  $x(t)$ , but recovering  $\mathbf{x}$  may not be sufficient for exactly recovering  $x(t)$  on the entire window  $[0, T_w]$ . (This depends on the exact sampling rate and the decay of the analog interpolation kernel.) In practice, the methods we describe in this paper for digital single-window reconstruction could be implemented in a streaming multi-window setting, and this would allow for a more accurate reconstruction of  $x(t)$  on the entire window.

<sup>2</sup>In our setup, since  $\Phi : [0, T_w] \rightarrow \mathbb{C}^M$ , we are inherently assuming that  $\phi_m(t) = 0$  outside of the interval  $[0, T_w]$ , so that the sensing functionals are time-limited. Although certain acquisition systems (such as the *modulated wideband converter* (MWC) of [47]) do not satisfy this condition, we believe that it is often a reasonable assumption in practice and that many acquisition systems can at least be well-approximated as time-limited.

A precise statement concerning the size of  $\mathbf{w}$  would depend greatly on the choice of the  $\phi_m(t)$ . While a detailed analysis of  $\mathbf{w}$  for various practical choices of  $\phi_m(t)$  is beyond the scope of this paper, we briefly mention some possible strategies for controlling  $\mathbf{w}$ . First, one can easily show that if each  $\phi_m(t)$  consists of any weighted combination of dirac delta functions positioned at times  $0, T_s, \dots, (N-1)T_s$ , then by construction  $\mathbf{w} = \mathbf{0}$ . This should not be surprising, as in this case it is clear that the measurements are simply a linear combination of the Nyquist-rate samples from the finite window. Importantly, it is possible to collect measurements of this type *without first acquiring the Nyquist-rate samples* (see, for example, the architecture proposed in [61]), although there are also plenty of situations in which one might explicitly apply a matrix multiplication to compress data *after* acquiring a length- $N$  vector of Nyquist-rate samples.

For many architectures used in practice, it will not be the case that  $\mathbf{w} = \mathbf{0}$  exactly. However, it may still be possible to ensure that  $\mathbf{w}$  remains very small. There are a number of possible routes to such a guarantee. For example, the  $\phi_m(t)$  could be designed to incorporate a smooth window  $g(t)$  so that we effectively sample  $x(t)g(t)$  instead of  $x(t)$ , where  $g(t)$  is designed to ensure that  $x[n]g[n] \approx 0$  for  $n \leq -1$  or  $n \geq N$ . The reconstruction algorithm could then compensate for the effect of  $g[n]$  on  $0, 1, \dots, N-1$ . Alternatively, by considering a slightly oversampled version of  $x(t)$  (so that  $\frac{1}{T_s}$  exceeds  $B_{\text{nyq}}$  by some nontrivial amount) it is also possible to replace the sinc interpolation kernel with one that decays significantly faster, ensuring that the inner products in (6) decay to zero extremely quickly [17]. Finally, as we will see below, many constructions of  $\phi_m(t)$  often involve a degree of randomness that could also be leveraged to show that with high probability, the inner products in (6) decay even faster. However, since the details will depend greatly on the particular architecture used, we leave such an investigation for future work.

Having argued that the measurement model in (2) can often be expressed in the form (1), we now turn to the central theoretical question of this paper:

Supposing that  $x(t)$  obeys the multiband model described in Section 4.1, how can we recover  $\mathbf{x}$ , i.e., the Nyquist-rate samples of  $x(t)$  on  $[0, T_w)$ , from compressive measurements of the form  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$ ?

In order to answer this question, of course, we will need a dictionary  $\Psi$  that provides a suitably sparse representation for  $\mathbf{x}$ . We devote Section 4 to constructing such a dictionary. In addition to a dictionary  $\Psi$ , however, we will also need a reconstruction algorithm that can efficiently recover  $\mathbf{x}$  from the compressive measurements  $\mathbf{y}$ . While it is certainly possible to apply out-of-the-box CS recovery algorithms to this problem, there are certain properties of our dictionary that make the recovery problem worthy of further consideration. (In particular, the columns of our dictionary  $\Psi$  will typically not be orthogonal, and the sparse coefficient vectors  $\alpha$  that arise will tend to have structured (block-sparse) sparsity patterns.) In light of these nuances, Section 3 now provides additional background on CS that will allow us to formulate a principled recovery technique.

## 3 Compressive Sensing Background

### 3.1 Sensing matrix design

Setting aside the question of how to design the sparsity-inducing dictionary  $\Psi$ , we first address the problem of designing  $\mathbf{A}$ . Although many favorable properties for sensing matrices have been studied in the context of CS, the most common is the *restricted isometry property* (RIP) [14]. We say that the matrix  $\mathbf{A}\Psi$  satisfies the RIP of order  $S$  if there exists a constant  $\delta_S \in (0, 1)$  such that

$$\sqrt{1 - \delta_S} \leq \frac{\|\mathbf{A}\Psi\alpha\|_2}{\|\alpha\|_2} \leq \sqrt{1 + \delta_S} \quad (8)$$

holds for all  $\alpha$  such that  $\|\alpha\|_0 \leq S$ . In words,  $\mathbf{A}\Psi$  preserves the norm of  $S$ -sparse vectors. Note that for any pair of vectors  $\alpha$  and  $\alpha'$  such that  $\|\alpha\|_0 = \|\alpha'\|_0 = S$ , we have that  $\|\alpha - \alpha'\|_0 \leq 2S$ . This gives us an alternative interpretation of (8)—namely that the RIP of order  $2S$  ensures that  $\mathbf{A}\Psi$  preserves Euclidean distances between  $S$ -sparse vectors  $\alpha$ .

A related concept is what we call the  $\Psi$ -RIP (following the notation in [12]). Specifically, we say that the matrix  $\mathbf{A}$  satisfies the  $\Psi$ -RIP of order  $S$  if there exists a constant  $\delta_S \in (0, 1)$  such that

$$\sqrt{1 - \delta_S} \leq \frac{\|\mathbf{A}\Psi\alpha\|_2}{\|\Psi\alpha\|_2} \leq \sqrt{1 + \delta_S} \quad (9)$$

holds for all  $\alpha$  such that  $\|\alpha\|_0 \leq S$ . When  $\Psi$  is an orthonormal basis, (8) and (9) are equivalent. However, we will be concerned in this paper with non-orthogonal (and even non-square) dictionaries  $\Psi$ , in which case the RIP and the  $\Psi$ -RIP are slightly different concepts: the former ensures norm preservation of all sparse coefficient vectors  $\alpha$ , while the latter ensures norm preservation of all signals having a sparse representation  $\mathbf{x} = \Psi\alpha$ . In many problems (such as when  $\Psi$  is an overcomplete dictionary), the RIP is considered to be a stronger requirement.

There are a variety of approaches to constructing matrices that satisfy the RIP or  $\Psi$ -RIP, some of which are better suited to practical architectures than others. From a theoretical standpoint, however, the most fruitful approaches involve the use of random matrices. Specifically, we consider matrices constructed as follows: given  $M$  and  $N$ , we generate a random  $M \times N$  matrix  $\mathbf{A}$  by choosing the entries  $\mathbf{A}[m, n]$  as independent and identically distributed (i.i.d.) random variables. While it is not strictly necessary, for the sake of simplicity we will consider only real-valued random variables, so that  $\mathbf{A} \in \mathbb{R}^{M \times N}$ .

We impose two conditions on the random distribution. First, we require that the distribution is centered and normalized such that  $\mathbb{E}(\mathbf{A}[m, n]) = 0$  and  $\mathbb{E}(\mathbf{A}[m, n]^2) = \frac{1}{M}$ . Second, we require that the distribution is *subgaussian* [9, 75], meaning that there exists a constant  $c_0 > 0$  such that

$$\mathbb{E} \left( e^{\mathbf{A}[m, n]t} \right) \leq e^{c_0^2 t^2} \quad (10)$$

for all  $t \in \mathbb{R}$ . Examples of subgaussian distributions include the Gaussian distribution, the Rademacher distribution, and the uniform distribution. In general, any distribution with bounded support is subgaussian.

The key property of subgaussian random variables that will be of use in this paper is that for any  $\mathbf{x} \in \mathbb{C}^N$ , the random variable  $\|\mathbf{A}\mathbf{x}\|_2^2$  is highly concentrated about  $\|\mathbf{x}\|_2^2$ . In particular, there exists a constant  $c_1(\eta) > 0$  that depends only on  $\eta$  and the constant  $c_0$  in (10) such that

$$\mathbb{P} \left[ \left| \|\mathbf{A}\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2 \right| \geq \eta \|\mathbf{x}\|_2^2 \right] \leq 4e^{-c_1(\eta)M}, \quad (11)$$

where the probability is taken over all draws of the matrix  $\mathbf{A}$  (see Lemma 6.1 of [26] or [18]).<sup>3</sup> We leave the constant  $c_1(\eta)$  undefined since it will depend both on the particular subgaussian distribution under consideration and on the range of  $\eta$  considered. Importantly, however, for any subgaussian distribution and any  $\eta_{\max}$ , we can write  $c_1(\eta) = \kappa\eta^2$  for  $\eta \leq \eta_{\max}$  with  $\kappa$  being a constant that depends on certain properties of the distribution [18]. This concentration bound has

---

<sup>3</sup>The concentration result in (11) is typically stated for  $\mathbf{x} \in \mathbb{R}^N$  instead of  $\mathbb{C}^N$ . The complex case follows from the real case by handling the real and imaginary parts separately and then applying the union bound, which results in a factor of 4 instead of 2 in front of the exponent.

a number of important consequences. Perhaps most important for our purposes is the following lemma (an adaptation of Lemma 5.1 in [4]).<sup>4</sup>

**Lemma 3.1.** *Let  $\mathcal{X}$  denote any  $S$ -dimensional subspace of  $\mathbb{C}^N$ . Fix  $\delta, \beta \in (0, 1)$ . Let  $\mathbf{A}$  be an  $M \times N$  random matrix with i.i.d. entries chosen from a distribution satisfying (11). If*

$$M \geq \frac{2S \log(42/\delta) + \log(4/\beta)}{c_1(\delta/\sqrt{2})} \quad (12)$$

*then with probability exceeding  $1 - \beta$ ,*

$$\sqrt{1 - \delta} \|\mathbf{x}\|_2 \leq \|\mathbf{A}\mathbf{x}\|_2 \leq \sqrt{1 + \delta} \|\mathbf{x}\|_2 \quad (13)$$

*for all  $\mathbf{x} \in \mathcal{X}$ .*

When  $\Psi$  is an orthonormal basis, one can use this lemma to go beyond a single  $S$ -dimensional subspace to instead consider all possible subspaces spanned by  $S$  columns of  $\Psi$ , thereby establishing the RIP for  $\mathbf{A}\Psi$ . The proof follows that of Theorem 5.2 of [4].

**Lemma 3.2.** *Let  $\Psi$  be an orthonormal basis for  $\mathbb{C}^N$  and fix  $\delta, \beta \in (0, 1)$ . Let  $\mathbf{A}$  be an  $M \times N$  random matrix with i.i.d. entries chosen from a distribution satisfying (11). If*

$$M \geq \frac{2S \log(42eN/\delta S) + \log(4/\beta)}{c_1(\delta/\sqrt{2})} \quad (14)$$

*with  $e$  denoting the base of the natural logarithm, then with probability exceeding  $1 - \beta$ ,  $\mathbf{A}\Psi$  will satisfy the RIP of order  $S$  with constant  $\delta$ .*

*Proof.* This is a simple generalization of Lemma 3.1, which follows from the observation that (8) is equivalent to (13) holding for all  $S$ -dimensional subspaces. There are  $\binom{N}{S} \leq (eN/S)^S$  subspaces of dimension  $S$  aligned with the coordinate axes of  $\Psi$ , and so applying a union bound to Lemma 3.1 we obtain the desired result.  $\square$

From essentially the same argument, we can also prove for more general dictionaries  $\Psi$  that  $\mathbf{A}$  will satisfy the  $\Psi$ -RIP.

**Corollary 3.1.** *Let  $\Psi$  be an arbitrary  $N \times D$  matrix and fix  $\delta, \beta \in (0, 1)$ . Let  $\mathbf{A}$  be an  $M \times N$  random matrix with i.i.d. entries chosen from a distribution satisfying (11). If*

$$M \geq \frac{2S \log(42eD/\delta S) + \log(4/\beta)}{c_1(\delta/\sqrt{2})} \quad (15)$$

*with  $e$  denoting the base of the natural logarithm, then with probability exceeding  $1 - \beta$ ,  $\mathbf{A}$  will satisfy the  $\Psi$ -RIP of order  $S$  with constant  $\delta$ .*

As noted above, the random matrix approach is somewhat impractical to build in hardware. However, several hardware architectures have been implemented and/or proposed that enable compressive samples to be acquired in practical settings. Examples include the random demodulator [72], random filtering [73], the modulated wideband converter [47], random convolution [1, 56], and the compressive multiplexer [61]. In this paper we will rely on random matrices in the development of our theory, but we will see via simulations that the techniques we propose are also applicable to systems that use some of these more practical architectures.

<sup>4</sup>The constants in [4] differ from those in Lemma 3.1, but the proof is substantially the same (see [20]). Note that in [20]  $\mathcal{X}$  is a subspace of  $\mathbb{R}^N$  rather than  $\mathbb{C}^N$ . In our case we incur an additional factor of 2 in the constant which arises as a consequence of the increase in the covering number for a sphere in  $\mathbb{C}^S$  (which can easily be derived from the fact that there is an isometry between  $\mathbb{C}^S$  and  $\mathbb{R}^{2S}$ ).



## 3.2 CS recovery algorithms

### 3.2.1 Greedy and iterative algorithms

Before we return to the problem of designing  $\Psi$ , we first discuss the question of how to recover the vector  $\mathbf{x}$  from measurements of the form  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e} = \mathbf{A}\Psi\boldsymbol{\alpha} + \mathbf{e}$ . The original CS theory proposed  $\ell_1$ -minimization as a recovery technique [10, 27]. Convex optimization techniques are powerful methods for CS signal recovery, but there also exist a variety of alternative greedy or iterative algorithms that are commonly used in practice and that satisfy similar performance guarantees, including iterative hard thresholding (IHT) [6], orthogonal matching pursuit (OMP) [23, 25, 53, 71], and several more recent variations on OMP [15, 16, 28, 49–51].

In this paper we will restrict our attention to two of the most commonly used algorithms in practice—IHT and CoSaMP [6, 49]. We begin with IHT, which is probably the simplest of all CS recovery algorithms. As is the case for most iterative recovery algorithms, a core component of IHT is *hard thresholding*. Specifically, we define the operator

$$\text{hard}(\boldsymbol{\alpha}, S)[n] = \begin{cases} \boldsymbol{\alpha}[n], & |\boldsymbol{\alpha}[n]| \text{ is among the } S \text{ largest elements of } |\boldsymbol{\alpha}|; \\ 0, & \text{otherwise.} \end{cases} \quad (16)$$

In words, the hard thresholding operator sets all but the  $S$  largest elements of a vector to zero (with ties broken according to any arbitrary rule).

To the best of our knowledge, there are no existing papers that specifically discuss how to implement IHT when  $\Psi$  is not an orthonormal basis. Nonetheless, we can envision two natural (and reasonable) ways that the canonical IHT algorithm [6] can be extended to handle a general dictionary. In the first of these variations, the algorithm would consist of iteratively applying the update rule

$$\boldsymbol{\alpha}^{\ell+1} = \text{hard}(\boldsymbol{\alpha}^\ell + \mu(\mathbf{A}\Psi)^H(\mathbf{y} - \mathbf{A}\Psi\boldsymbol{\alpha}^\ell), S) \quad (17)$$

where  $\mu$  is a parameter set by the user. In the second of these variations, the algorithm would consist of iteratively applying the update rule

$$\mathbf{x}^{\ell+1} = \Psi \cdot \text{hard}\left(\Psi^H\left(\mathbf{x}^\ell + \mu\mathbf{A}^H\left(\mathbf{y} - \mathbf{A}\mathbf{x}^\ell\right)\right), S\right). \quad (18)$$

When  $\Psi$  is an orthonormal basis these algorithms are equivalent, but in general they are not. On the whole, IHT is a remarkably simple algorithm, but in practice its performance is greatly dependent on careful selection and adaptation of the parameter  $\mu$ . We refer the reader to [6] for further details.

CoSaMP is a somewhat more complicated algorithm, but can be easily understood as breaking the recovery problem into two separate sub-problems: identifying the  $S$  columns of  $\Psi$  that best represent  $\mathbf{x}$  and then projecting onto that subspace. The former problem is clearly somewhat challenging, but once solved, the latter is relatively straightforward. In particular, if we have identified the optimal columns of  $\Psi$ , indexed by the set  $\Lambda$ , then we can recover  $\mathbf{x}$  via least-squares. In this case, an optimal recovery strategy is to solve the problem:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{z}} \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2 \quad \text{s.t.} \quad \mathbf{z} \in \mathcal{R}(\Psi_\Lambda), \quad (19)$$

where  $\Psi_\Lambda$  denotes the submatrix of  $\Psi$  that contains only the columns of  $\Psi$  corresponding to the index set  $\Lambda$  and  $\mathcal{R}(\Psi_\Lambda)$  denotes the range of  $\Psi_\Lambda$ . If we let  $\tilde{\mathbf{A}} = \mathbf{A}\Psi$ , then one way to obtain the solution to (19) is via the pseudoinverse of  $\tilde{\mathbf{A}}_\Lambda$ , denoted  $\tilde{\mathbf{A}}_\Lambda^\dagger$ . Specifically, we can compute

$$\hat{\boldsymbol{\alpha}}|_\Lambda = \tilde{\mathbf{A}}_\Lambda^\dagger \mathbf{y} = \left(\tilde{\mathbf{A}}_\Lambda^H \tilde{\mathbf{A}}_\Lambda\right)^{-1} \tilde{\mathbf{A}}_\Lambda^H \mathbf{y} \quad \text{and} \quad \hat{\boldsymbol{\alpha}}|_{\Lambda^c} = \mathbf{0} \quad (20)$$

---

**Algorithm 1** Compressive Sampling Matching Pursuit (CoSaMP)

---

**input:**  $\mathbf{A}, \Psi, \mathbf{y}, S$ , stopping criterion  
**initialize:**  $\mathbf{r}^0 = \mathbf{y}, \mathbf{x}^0 = 0, \ell = 0$   
**while** not converged **do**  
  **proxy:**  $\mathbf{h}^\ell = \mathbf{A}^H \mathbf{r}^\ell$   
  **identify:**  $\Omega^{\ell+1} = \text{supp}(\text{hard}(\Psi^H \mathbf{h}^\ell, 2S))$   
  **merge:**  $\Lambda^{\ell+1} = \text{supp}(\Psi^H \mathbf{x}^\ell) \cup \Omega^{\ell+1}$   
  **update:**  $\tilde{\mathbf{x}} = \arg \min_{\mathbf{z}} \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2 \quad \text{s.t.} \quad \mathbf{z} \in \mathcal{R}(\Psi_{\Lambda^{\ell+1}})$   
           $\mathbf{x}^{\ell+1} = \Psi \cdot \text{hard}(\Psi^H \tilde{\mathbf{x}}, S)$   
           $\mathbf{r}^{\ell+1} = \mathbf{y} - \mathbf{A}\mathbf{x}^{\ell+1}$   
           $\ell = \ell + 1$   
**end while**  
**output:**  $\hat{\mathbf{x}} = \mathbf{x}^\ell$

---

and then set  $\hat{\mathbf{x}} = \Psi \hat{\boldsymbol{\alpha}}$ . While this is certainly not the only approach to solving (19) (as we will see in Section 6.1.2), it allows us to easily observe that in the noise-free setting, if the support estimate  $\Lambda$  is correct, then  $\mathbf{y} = \mathbf{A}\mathbf{x} = \tilde{\mathbf{A}}_\Lambda \boldsymbol{\alpha}|_\Lambda$ , and so plugging this into (20) yields  $\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}$  (and hence  $\hat{\mathbf{x}} = \mathbf{x}$ ) provided that  $\tilde{\mathbf{A}}_\Lambda$  has full column rank. Thus, the central challenge in recovery is to correctly identify the set  $\Lambda$ . CoSaMP and related algorithms solve this problem by iteratively identifying likely columns, performing a projection, and then improving the estimate of which columns to use.

Unfortunately, we are again not aware of any papers that specifically discuss how to implement CoSaMP when  $\Psi$  is not an orthonormal basis. Nonetheless, we can envision two natural extensions of the canonical CoSaMP algorithm [49]. One of these is shown in Algorithm 1;<sup>5</sup> in a sense, this formulation is more analogous to (18) than to (17) because it is focused on recovery of  $\mathbf{x}$  rather than  $\boldsymbol{\alpha}$ . However, Algorithm 1 is actually quite flexible and can be invoked in multiple ways. To help distinguish among the different possibilities, it will be helpful to introduce the notation

$$\hat{\mathbf{x}} = \text{CoSaMP}(\mathbf{A}, \Psi, \mathbf{y}, S)$$

to denote the output produced by Algorithm 1 when the arguments  $(\mathbf{A}, \Psi, \mathbf{y}, S)$  are provided as input. Having set this notation, it is also reasonable to consider invoking Algorithm 1 with the input arguments  $(\mathbf{A}\Psi, \mathbf{I}, \mathbf{y}, S)$ . This formulation is more analogous to (17). In this case we will denote the output by  $\hat{\boldsymbol{\alpha}} = \text{CoSaMP}(\mathbf{A}\Psi, \mathbf{I}, \mathbf{y}, S)$  since the algorithm will construct and output an estimate of  $\boldsymbol{\alpha}$  (rather than  $\mathbf{x}$ ).

### 3.2.2 “Model-based” recovery algorithms

Traditional approaches to CS signal recovery, like those described above, place no prior assumptions on  $\text{supp}(\boldsymbol{\alpha})$ . Sparsity on its own implies nothing about the locations of the nonzeros, and hence most approaches to CS signal recovery treat every possible support as equally likely. However, in many practical applications the nonzeros are not distributed completely at random, but rather exhibit a

---

<sup>5</sup>We note that the choice of  $2S$  in the “identify” step is primarily driven by the proof technique, and is not intended to be interpreted as an optimal or necessary choice. For example, in [16] it is shown that the choice of  $S$  is sufficient to establish performance guarantees similar to those for CoSaMP. It is also important to note that when the number of measurements  $M$  is very small (less than  $3S$ ) it is necessary to make suitable modifications as the assumptions of the algorithm are clearly violated in this case. Moreover, a simple extension of CoSaMP as presented here involves including an additional orthogonalization step after pruning  $\tilde{\mathbf{x}}$  down to an  $S$ -dimensional estimate, as is also done in [16]. This can often result in modest performance gains and is a technique that we exploit in our simulations.

degree of structure. In the case of signals exhibiting such *structured sparsity*, it is possible to both reduce the required number of measurements and develop specialized “model-based” algorithms for recovery that exploit this structure [3, 5].

In this paper, we are interested in the model of *block-sparsity* [3, 30]. In a block-sparse vector  $\alpha$ , the nonzero coefficients cluster in a small number of blocks. Specifically, suppose that  $\mathbf{x} = \Psi\alpha$  with  $\Psi$  being an  $N \times D$  matrix and that we decompose  $\Psi$  into  $J$  submatrices of size  $N \times \frac{D}{J}$ , i.e.,

$$\Psi = [\Psi_0 \ \Psi_1 \ \cdots \ \Psi_{J-1}].$$

Then we can write  $\mathbf{x} = \sum_{i=0}^{J-1} \Psi_i \alpha_i$ , where each  $\alpha_i \in \mathbb{C}^{D/J}$ . We say that  $\alpha$  is  $K$ -block-sparse if there exists a set  $\mathcal{I} \subseteq \{0, 1, \dots, J-1\}$  such that  $|\mathcal{I}| = K$  and  $\alpha_i = 0$  for all  $i \notin \mathcal{I}$ . With some abuse of notation, we now let  $\Psi_{\mathcal{I}}$  denote the submatrix of  $\Psi$  that contains only the columns of  $\Psi$  corresponding to the *blocks* indexed by  $\mathcal{I}$ .

We first illustrate how we can exploit block-sparsity algorithmically. Our goal is to generalize IHT and CoSaMP to the block-sparse setting. To do this, we observe that the hard thresholding function plays a key role in both algorithms. One way to interpret this role is that  $\text{hard}(\Psi^H \mathbf{x}, S)$  is actually computing a projection of  $\Psi^H \mathbf{x}$  onto the set of  $S$ -sparse vectors. In the case where  $\Psi$  is an orthonormal basis we can also interpret  $\Psi \cdot \text{hard}(\Psi^H \mathbf{x}, S)$  as projecting  $\mathbf{x}$  onto the set of signals that are  $S$ -sparse with respect to the basis  $\Psi$ .

In the block-sparse case we must replace hard thresholding with an appropriate operator that takes a candidate signal and finds the closest  $K$ -block-sparse approximation. Towards this end, we define

$$\mathcal{S}_{\Psi}(\mathbf{x}, K) := \arg \min_{|\mathcal{I}| \leq K} \min_{\mathbf{z} \in \mathcal{R}(\Psi_{\mathcal{I}})} \|\mathbf{x} - \mathbf{z}\|_2. \quad (21)$$

$\mathcal{S}_{\Psi}(\mathbf{x}, K)$  is analogous to the support of  $\alpha$  in the traditional sparse setting: it tells us which set of  $K$  blocks of  $\Psi$  can best approximate  $\mathbf{x}$ . Along with  $\mathcal{S}_{\Psi}(\mathbf{x}, K)$ , we also define

$$\mathcal{P}_{\Psi}(\mathbf{x}, K) := \arg \min_{\mathbf{z}} \|\mathbf{x} - \mathbf{z}\|_2 \quad \text{s.t.} \quad \mathbf{z} \in \mathcal{R}(\Psi_{\mathcal{S}_{\Psi}(\mathbf{x}, K)}). \quad (22)$$

$\mathcal{P}_{\Psi}$  is simply the projection of the vector  $\mathbf{x} \in \mathbb{C}^N$  onto the set of  $K$ -block-sparse signals. To simplify our notation, we will often write  $\mathcal{S}(\alpha, K) = \mathcal{S}_{\Psi}(\alpha, K)$  and  $\mathcal{P}(\mathbf{x}, K) = \mathcal{P}_{\Psi}(\mathbf{x}, K)$  when  $\Psi$  is clear from the context.

For each of the IHT and CoSaMP algorithms proposed in Section 3.2.1, it is possible to propose a variation of the algorithm designed to exploit block-sparsity simply by replacing the hard thresholding operator with an appropriate block-sparse projection. For example, one block-based version of IHT (which is also a special case of the iterative projection algorithm in [5]), would consist of replacing the core iteration of IHT in (18) with

$$\mathbf{x}^{\ell+1} = \mathcal{P}(\mathbf{x}^{\ell} + \mu \mathbf{A}^H(\mathbf{y} - \mathbf{A}\mathbf{x}^{\ell}), K). \quad (23)$$

Note that the only difference from (18) is that we have replaced hard thresholding with the projection onto the set of block-sparse signals. Similarly, a block-based version of CoSaMP is shown in Algorithm 2.

Algorithm 2 is in differing senses both a special case and a generalization of the model-based CoSaMP algorithm proposed in [3]. Specifically, in [3] an algorithm for block-sparse signal recovery is proposed that is equivalent to Algorithm 2 when  $\Psi = \mathbf{I}$ . The more general case of arbitrary  $\Psi$  is not discussed. However, there are alternative options for handling  $\Psi \neq \mathbf{I}$  besides the one specified in Algorithm 2. Like Algorithm 1, Algorithm 2 is quite flexible and can be invoked in multiple ways. Following our convention in Section 3.2.1, we use the notation  $\hat{\mathbf{x}} = \text{BBCoSaMP}(\mathbf{A}, \Psi, \mathbf{y}, K)$

---

**Algorithm 2** Block-Based CoSaMP (BBCoSaMP)

---

**input:**  $\mathbf{A}, \Psi, \mathbf{y}, K$ , stopping criterion  
**initialize:**  $\mathbf{r}^0 = \mathbf{y}, \mathbf{x}^0 = 0, \ell = 0$   
**while** not converged **do**  
    **proxy:**  $\mathbf{h}^\ell = \mathbf{A}^H \mathbf{r}^\ell$   
    **identify:**  $\tilde{\mathcal{I}}^{\ell+1} = \mathcal{S}(\mathcal{P}(\mathbf{h}^\ell, 2K), 2K)$   
    **merge:**  $\mathcal{I}^{\ell+1} = \mathcal{S}(\mathbf{x}^\ell, K) \cup \tilde{\mathcal{I}}^{\ell+1}$   
    **update:**  $\tilde{\mathbf{x}} = \arg \min_{\mathbf{z}} \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2 \quad \text{s.t.} \quad \mathbf{z} \in \mathcal{R}(\Psi_{\mathcal{I}^{\ell+1}})$   
         $\mathbf{x}^{\ell+1} = \mathcal{P}(\tilde{\mathbf{x}}, K)$   
         $\mathbf{r}^{\ell+1} = \mathbf{y} - \mathbf{A}\mathbf{x}^{\ell+1}$   
         $\ell = \ell + 1$   
**end while**  
**output:**  $\hat{\mathbf{x}} = \mathbf{x}^\ell$

---

to denote the output produced by Algorithm 2 when the arguments  $(\mathbf{A}, \Psi, \mathbf{y}, K)$  are provided as input, and we use the notation  $\hat{\boldsymbol{\alpha}} = \text{BBCoSaMP}(\mathbf{A}\Psi, \mathbf{I}, \mathbf{y}, K)$  to denote the output produced by Algorithm 2 when the arguments  $(\mathbf{A}\Psi, \mathbf{I}, \mathbf{y}, K)$  are provided as input.

### 3.2.3 “Model-based” recovery guarantees

Theoretical guarantees for standard CS recovery algorithms typically rely on the RIP, and since in the standard case any  $S$ -sparse signal is possible, there is little room for improvement. However, the block-sparse model actually rules out a large number of possible signal supports, and so we no longer require the full RIP or  $\Psi$ -RIP, i.e., we no longer need (8) or (9) to hold for all possible  $S$ -sparse signals. Instead we only require that (8) or (9) hold for all  $\boldsymbol{\alpha}$  which are  $K$ -block-sparse. We will refer to these relaxed properties as the block-RIP and  $\Psi$ -block-RIP respectively.

The relaxation to block-sparse signals allows us to potentially dramatically reduce the required number of measurements. Specifically, note that a  $K$ -block-sparse vector  $\boldsymbol{\alpha}$  satisfies  $\|\boldsymbol{\alpha}\|_0 \leq K \frac{D}{J}$ . In the standard sparsity model we would have that the number of possible subspaces is  $\binom{D}{S}$  with  $S = K \frac{D}{J}$ , whereas now the number of possible subspaces is given by  $\binom{J}{K}$ , which can be potentially much smaller. Establishing (8) or (9) for a more general union of subspaces is a problem that has received some attention in the CS literature [7, 31, 44]. In our context it should be clear that we can simply apply Lemma 3.1 just as in the proofs of Lemma 3.2 and Corollary 3.1 to obtain the following improved bounds.

**Corollary 3.2.** *Let  $\Psi$  be an orthonormal basis for  $\mathbb{C}^N$  and fix  $\delta, \beta \in (0, 1)$ . Let  $\mathbf{A}$  be an  $M \times N$  random matrix with i.i.d. entries chosen from a distribution satisfying (11). If*

$$M \geq \frac{2K \left( \frac{N}{J} \log(42/\delta) + \log(eJ/K) \right) + \log(4/\beta)}{c_1(\delta/\sqrt{2})} \quad (24)$$

*with  $e$  denoting the base of the natural logarithm, then with probability exceeding  $1 - \beta$ ,  $\mathbf{A}\Psi$  will satisfy the block-RIP of order  $K$  with constant  $\delta$ .*

**Corollary 3.3.** *Let  $\Psi$  be an arbitrary  $N \times D$  matrix and fix  $\delta, \beta \in (0, 1)$ . Let  $\mathbf{A}$  be an  $M \times N$  random matrix with i.i.d. entries chosen from a distribution satisfying (11). If*

$$M \geq \frac{2K \left( \frac{D}{J} \log(42/\delta) + \log(eJ/K) \right) + \log(4/\beta)}{c_1(\delta/\sqrt{2})} \quad (25)$$

with  $e$  denoting the base of the natural logarithm, then with probability exceeding  $1 - \beta$ ,  $\mathbf{A}$  will satisfy the  $\Psi$ -block-RIP of order  $K$  with constant  $\delta$ .

The measurement requirements in (24) and (25) represent improvements over (14) or (15) in that a straightforward application of (14) or (15) would lead to replacing the  $\log(eJ/K)$  term above with  $\frac{N}{J} \log(eJ/K)$  or  $\frac{D}{J} \log(eJ/K)$  respectively.

We can combine these corollaries with the following theorems to show that we can stably recover block-sparse signals using fewer measurements. Note that the following theorems are simplified guarantees for the case of only approximately block-sparse signals. Both theorems can be generalized to block-compressible signals,<sup>6</sup> but we restrict our attention to the guarantees for sparse signals to simplify discussion.

**Theorem 3.1** (Theorem 2 from [5]). *Suppose that  $\mathbf{x}$  can be written as  $\mathbf{x} = \Psi\boldsymbol{\alpha}$  where  $\boldsymbol{\alpha}$  is  $K$ -block-sparse and that we observe  $\mathbf{y} = \mathbf{A}\Psi\boldsymbol{\alpha} + \mathbf{e} = \mathbf{A}\mathbf{x} + \mathbf{e}$ . If  $\mathbf{A}$  satisfies the  $\Psi$ -block-RIP of order  $2K$  with constant  $\delta_{2K} \leq 0.2$  and  $\frac{1}{\mu} \in [1 + \delta_{2K}, 1.5(1 - \delta_{2K})]$ , then the estimate obtained from block-based IHT (23) satisfies*

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \kappa_1 \|\mathbf{e}\|_2 \quad (26)$$

where  $\kappa_1 > 1$  is a constant determined by  $\delta_{2K}$  and the stopping criterion.

**Theorem 3.2** (Theorem 6 from [3]). *Suppose that  $\boldsymbol{\alpha}$  is  $K$ -block-sparse and that we observe  $\mathbf{y} = \mathbf{A}\Psi\boldsymbol{\alpha} + \mathbf{e}$ . If  $\mathbf{A}\Psi$  satisfies the block-RIP of order  $4K$  with constant  $\delta_{4K} \leq 0.1$ , then the output of block-based CoSaMP (Algorithm 2) with  $\hat{\boldsymbol{\alpha}} = \text{BBCoSaMP}(\mathbf{A}\Psi, \mathbf{I}, \mathbf{y}, K)$  satisfies*

$$\|\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}\|_2 \leq \kappa_2 \|\mathbf{e}\|_2 \quad (27)$$

where  $\kappa_2 > 1$  is a constant determined by  $\delta_{4K}$  and the stopping criterion.

Theorems 3.1 and 3.2 show that measurement noise has a controlled impact on the amount of noise in the reconstruction. However, note that Theorems 3.1 and 3.2 are fundamentally different from one another when the matrix  $\Psi$  is not an orthonormal basis. Theorem 3.2 requires the assumption that  $\mathbf{A}\Psi$  satisfies the block-RIP while Theorem 3.1 requires that  $\mathbf{A}$  satisfies the  $\Psi$ -block-RIP. Theorem 3.2 also provides a different guarantee (recovery of  $\boldsymbol{\alpha}$ ) compared to Theorem 3.1 (which guarantees recovery of  $\mathbf{x}$ ). In the case where  $\Psi$  is an orthonormal basis, we could immediately set  $\hat{\mathbf{x}} = \Psi\hat{\boldsymbol{\alpha}}$  so that the recovery guarantee in (27) applies to the error in  $\mathbf{x}$  as well. However, for arbitrary dictionaries  $\Psi$  this equivalence no longer holds.<sup>7</sup> We conjecture that were we to instead consider  $\hat{\mathbf{x}} = \text{BBCoSaMP}(\mathbf{A}, \Psi, \mathbf{y}, K)$  we should be able to establish a theorem for block-based CoSaMP analogous to Theorem 3.1, but we leave such an analysis for future work.

In Section 4 we develop a multiband modulated DPSS dictionary  $\Psi$  designed to offer high-quality block-sparse representations for sampled multiband signals. Using this dictionary we establish block-RIP and  $\Psi$ -block-RIP guarantees in Section 5, which allows us to translate Theorems 3.1 and 3.2 into guarantees for recovery of sampled multiband signals from compressive measurements. When we then turn to implement these algorithms, however, it is important to note that, although we can implement the  $\text{BBCoSaMP}(\mathbf{A}\Psi, \mathbf{I}, \mathbf{y}, K)$  version of block-based CoSaMP with no trouble,

<sup>6</sup>By block-compressible, we mean signals that are well approximated by a block-sparse signal. The guarantee on the recovery error for compressible signals is similar to those in Theorems 3.1 and 3.2 but includes an additional additive component that quantifies the error incurred by approximating  $\boldsymbol{\alpha}$  with a block-sparse signal. If a signal is close to being block-sparse, then this error is negligible, but if a signal is not block-sparse at all, then this error can potentially be large.

<sup>7</sup>It is also worth noting that in the context of traditional (as opposed to model-based) CS, there do exist guarantees on  $\|\hat{\mathbf{x}} - \mathbf{x}\|_2$  for general  $\Psi$  when using an alternative optimization-based approach combined with the  $\Psi$ -RIP [12].

there is an important caveat to the results for block-based IHT and the BBCoSaMP( $\mathbf{A}, \mathbf{\Psi}, \mathbf{y}, K$ ) version of block-based CoSaMP. Specifically, both of those latter algorithms require that we be able to compute  $\mathcal{P}(\mathbf{x}, K)$  as defined in (22). Unfortunately, because our dictionary  $\mathbf{\Psi}$  is not orthonormal we are aware of no algorithms that are guaranteed to solve this problem. However, we will see empirically in Section 6 that we can attempt to solve this problem by applying many of the same algorithms commonly used for CS recovery. In other words, we can perfectly implement an algorithm (BBCoSaMP( $\mathbf{A}\mathbf{\Psi}, \mathbf{I}, \mathbf{y}, K$ )) that has a provable guarantee on the recovery of  $\boldsymbol{\alpha}$ , and we can approximately implement algorithms (block-based IHT and BBCoSaMP( $\mathbf{A}, \mathbf{\Psi}, \mathbf{y}, K$ )) that are intended to accurately recover  $\mathbf{x}$ . We have implemented both variations of block-based CoSaMP, and while both perform well in practice, the empirical performance of BBCoSaMP( $\mathbf{A}, \mathbf{\Psi}, \mathbf{y}, K$ ) turns out to be superior. In Section 6 we present a suite of simulations demonstrating the remarkable effectiveness of BBCoSaMP( $\mathbf{A}, \mathbf{\Psi}, \mathbf{y}, K$ ) (when combined with the multiband modulated DPSS dictionary) in recovering sampled multiband signals from compressive measurements.

## 4 A Sparse Dictionary for Sampled Multiband Signals

### 4.1 Analog signal models

We now confront the problem of designing a dictionary  $\mathbf{\Psi}$  in which  $\mathbf{x}$ , a length- $N$  vector of Nyquist-rate samples of  $x(t)$ , will be sparse or compressible. This is typically a significant challenge to anyone applying CS techniques to analog signals, since many natural analog signal models cannot be obviously represented via a simple basis  $\mathbf{\Psi}$ . We now describe two of the most appealing analog signal models and discuss the degree to which these models can fit within the CS framework.

#### 4.1.1 Multitone signals

There are a variety of possibilities for quantifying the notion of sparsity in a continuous-time signal  $x(t)$ . Perhaps the simplest, dating back at least to the work of Prony [55], is the *multitone* model. At its simplest, it assumes that  $x(t)$  can be expressed as

$$x(t) = \sum_{k=0}^S \alpha_k e^{j2\pi f_k t}, \quad (28)$$

i.e.,  $x(t)$  is simply a weighted combination of  $S$  complex exponentials of arbitrary frequencies. A related model is given by

$$x(t) = \sum_{n=0}^{N-1} \alpha[n] e^{j2\pi n t / NT_s}, \quad (29)$$

where  $\|\boldsymbol{\alpha}\|_0 = S$ . This model is considered in [72], which provided one of the first extensions of the CS framework to the case of continuous-time signals. The advantage of this model is that (29) implies that  $\mathbf{x} = \mathbf{\Psi}\boldsymbol{\alpha}$ , where  $\mathbf{x}$  is the vector of discrete-time samples of  $x(t)$  on  $[0, T_w]$  and  $\mathbf{\Psi}$  is the non-normalized  $N \times N$  DFT matrix. Thus, the model (29) immediately fits into the standard CS framework when the vector of coefficients  $\boldsymbol{\alpha}$  is sparse.

In practice, however, this approach is inadequate for two main reasons. First, the model in (29) assumes that any tones in the signal have a frequency that lies exactly on the so-called *Nyquist-grid*, i.e., the tones are *bounded harmonics*. When dealing with tones of arbitrary frequencies, the corresponding ‘‘DFT leakage’’ results in  $\boldsymbol{\alpha}$  that are not sparse and are not even well-approximated as sparse. In this case it can be useful to either incorporate a smooth windowing function into

the measurement system, as in [72], or to consider the less restrictive model in (28), as in [29]. However, these approaches do not address the second main objection, which is that many (if not most) real-world signals are not mere combinations of a few pure tones. For a variety of reasons, it is typically more realistic to assume that each of the  $S$  signal components has some non-negligible bandwidth, which leads us to instead consider the following extension of the multitone model.

#### 4.1.2 Multiband signals

For the remainder of this paper, we will focus on *multiband* signals, or signals whose Fourier transform is concentrated on a small number of continuous intervals or bands. Towards this end, for a general continuous-time signal  $x(t)$ , we define  $\mathcal{F}(x)$  as the support of  $X(F)$ , i.e.,  $\mathcal{F}(x) = \{F \in \mathbb{R} : X(F) \neq 0\}$ . We will be interested in signals for which we can decompose  $\mathcal{F} = \mathcal{F}(x)$  into a union of  $K$  continuous intervals, so that we can write

$$\mathcal{F} \subseteq \bigcup_{i=1}^K [a_i, b_i].$$

In the most general setting, we would allow the endpoints of each interval to be arbitrary but subject to a restriction on the total Lebesgue measure of their union. See, for example, [8, 33, 34, 74]. In this paper we restrict ourselves to a simpler model. Specifically, we divide the range of possible frequencies from  $-\frac{B_{\text{nyq}}}{2}$  to  $\frac{B_{\text{nyq}}}{2}$  into  $J = \frac{B_{\text{nyq}}}{B_{\text{band}}}$  equal bands of width  $B_{\text{band}}$  and require  $X(F)$  to be supported on at most  $K$  of these bands. An example is illustrated in Figure 1. More formally, we define the  $i^{\text{th}}$  band as

$$\Delta_i = \left[ -\frac{B_{\text{nyq}}}{2} + iB_{\text{band}}, -\frac{B_{\text{nyq}}}{2} + (i+1)B_{\text{band}} \right].$$

We then define

$$\Gamma(K, B_{\text{band}}) = \left\{ \mathcal{F} : \mathcal{F} \subseteq \bigcup_{i \in \mathcal{I}} \Delta_i \text{ for some } \mathcal{I} \subseteq \{0, 1, \dots, J-1\} \text{ with } |\mathcal{I}| \leq K \right\}$$

as the set of all possible supports. Using this, we define

$$\mathcal{M}(K, B_{\text{band}}) = \{x(t) : \mathcal{F}(x) \subseteq \mathcal{F}' \text{ for some } \mathcal{F}' \in \Gamma(K, B_{\text{band}})\} \quad (30)$$

as the set of multiband signals. Note that the total occupied bandwidth is at most  $KB_{\text{band}}$ . Our interest is in the setting where  $KB_{\text{band}} \ll B_{\text{nyq}}$ , so that if we knew *a priori* where the  $K$  bands were located, we could acquire  $x(t)$  in a streaming setting with only  $KB_{\text{band}}$  samples per second. (This is the so-called *Landau rate* [40].) Our goal is to acquire finite windows of multiband signals without such *a priori* knowledge while keeping the measurement rate as close as possible to the Landau rate.

Note, however, that the set  $\mathcal{M}(K, B_{\text{band}})$  is defined for infinite-length signals  $x(t)$ . Indeed, any signal with a Fourier transform supported on a finite range of frequencies cannot also be supported on a finite range of time. This would seem to be somewhat at odds with the finite-dimensional CS framework described above. As a result, previous efforts aimed at sampling multiband signals have developed largely outside the framework of CS [8, 33, 34, 46, 47, 74]. It is our goal in this paper to show that it is possible to recover a finite-length window of samples of a multiband signal using many of the standard tools from CS. To do this, we will need to construct an appropriate dictionary  $\Psi$  for capturing the structure of this set, which we do in the following section.

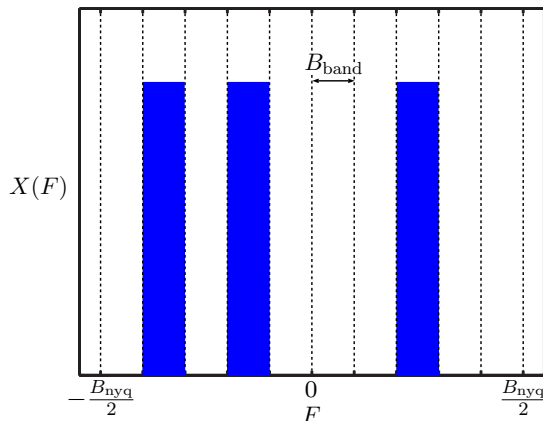


Figure 1: Illustration showing the CTFT of a multiband signal where  $J = \frac{B_{\text{nyq}}}{B_{\text{band}}} = 10$  and  $K = 3$ .

Finally, we also note that while our signal model breaks up the spectrum into  $J = \frac{B_{\text{nyq}}}{B_{\text{band}}}$  bands with fixed boundaries and bandwidth, it actually encompasses the broader class of signals where the bandwidth and center frequency of each band is arbitrary. For example, a signal with  $K$  bands of width  $B_{\text{band}}$  but with arbitrary center frequencies will automatically lie within  $\mathcal{M}(2K, B_{\text{band}})$ . Since we are primarily interested in the case where  $KB_{\text{band}} \ll B_{\text{nyq}}$ , this factor of 2 will not be significant in the development of our theoretical results. However, we do note that in practice it may be possible to achieve a significant gain by exploiting a more flexible signal model.

In the remainder of this section we demonstrate that it is possible to construct discrete-time bases using *Discrete Prolate Spheroidal Sequences* (DPSS's) that efficiently capture the structure of sampled multiband signals. We first review DPSS's and their key properties as first developed in [64], and we then discuss some of the consequences of these properties in terms of their utility in representing sampled continuous-time signals. Ultimately, we demonstrate how to use DPSS's to construct a dictionary  $\Psi$  which sparsely represents windows of sampled multiband signals.

## 4.2 Discrete Prolate Spheroidal Sequences (DPSS's)

Our goal in this subsection is to provide a self-contained review of the concepts from Slepian's work on DPSS's [64] that are most relevant to the problem of modeling sampled multiband signals. We refer the reader to [41, 42, 62–66] for more complete overviews of DPSS's, PSWF's, and their implications in time-frequency localization.

### 4.2.1 DPSS's

Let  $N$  be an integer, and let  $0 < W < \frac{1}{2}$ . Given  $N$  and  $W$ , the DPSS's are a collection of  $N$  discrete-time sequences that are strictly bandlimited to the digital frequency range  $|f| \leq W$  yet highly concentrated in time to the index range  $n = 0, 1, \dots, N - 1$ . The DPSS's are defined to be the eigenvectors of a two-step procedure in which one first time-limits the sequence and then bandlimits the sequence. Before we can state a more formal definition, let us note that for a given discrete-time signal  $x[n]$ , we let

$$X(f) = \sum_{n=-\infty}^{\infty} x[n]e^{-j2\pi fn}$$



denote the discrete-time Fourier transform (DTFT) of  $x[n]$ .<sup>8</sup> Next, we let  $\mathcal{B}_W$  denote an operator that takes a discrete-time signal, bandlimits its DTFT to the frequency range  $|f| \leq W$ , and returns the corresponding signal in the time domain. Additionally, we let  $\mathcal{T}_N$  denote an operator that takes an infinite-length discrete-time signal and zeros out all entries outside the index range  $\{0, 1, \dots, N-1\}$  (but still returns an infinite-length signal). With these definitions, the DPSS's are defined as follows.

**Definition 4.1.** [64] *Given  $N$  and  $W$ , the Discrete Prolate Spheroidal Sequences (DPSS's) are a collection of  $N$  real-valued discrete-time sequences  $s_{N,W}^{(0)}, s_{N,W}^{(1)}, \dots, s_{N,W}^{(N-1)}$  that, along with the corresponding scalar eigenvalues*

$$1 > \lambda_{N,W}^{(0)} > \lambda_{N,W}^{(1)} > \dots > \lambda_{N,W}^{(N-1)} > 0,$$

*satisfy*

$$\mathcal{B}_W(\mathcal{T}_N(s_{N,W}^{(\ell)})) = \lambda_{N,W}^{(\ell)} s_{N,W}^{(\ell)} \quad (31)$$

*for all  $\ell \in \{0, 1, \dots, N-1\}$ . The DPSS's are normalized so that*

$$\|\mathcal{T}_N(s_{N,W}^{(\ell)})\|_2 = 1 \quad (32)$$

*for all  $\ell \in \{0, 1, \dots, N-1\}$ .*

As we discuss in more detail below in Section 4.2.4, the eigenvalues  $\lambda_{N,W}^{(0)}, \lambda_{N,W}^{(1)}, \dots, \lambda_{N,W}^{(N-1)}$  have a very distinctive behavior: the first  $2NW$  eigenvalues tend to cluster extremely close to 1, while the remaining eigenvalues tend to cluster similarly close to 0.

Before proceeding, let us briefly mention several key properties of the DPSS's that will be useful in our subsequent analysis. First, it is clear from (31) that the DPSS's are, by definition, strictly bandlimited to the digital frequency range  $|f| \leq W$ . Second, the DPSS's are also approximately time-limited to the index range  $n = 0, 1, \dots, N-1$ . Specifically, it can be shown that [64]

$$\|s_{N,W}^{(\ell)}\|_2 = \frac{1}{\sqrt{\lambda_{N,W}^{(\ell)}}}. \quad (33)$$

Comparing (32) with (33), we see that for values of  $\ell$  where  $\lambda_{N,W}^{(\ell)} \approx 1$ , nearly all of the energy in  $s_{N,W}^{(\ell)}$  is contained in the interval  $\{0, 1, \dots, N-1\}$ . Third, the DPSS's are orthogonal over  $\mathbb{Z}$  [64], i.e., for any  $\ell, \ell' \in \{0, 1, \dots, N-1\}$  with  $\ell \neq \ell'$ ,  $\langle s_{N,W}^{(\ell)}, s_{N,W}^{(\ell')} \rangle = 0$ .

#### 4.2.2 Time-limited DPSS's

While each DPSS actually has infinite support in time, several very useful properties hold for the collection of signals one obtains by time-limiting the DPSS's to the index range  $n = 0, 1, \dots, N-1$ . First, the time-limited DPSS's  $\mathcal{T}_N(s_{N,W}^{(0)}), \mathcal{T}_N(s_{N,W}^{(1)}), \dots, \mathcal{T}_N(s_{N,W}^{(N-1)})$  are approximately bandlimited to the digital frequency range  $|f| \leq W$ . Specifically, from (31) and (33), one can deduce that

$$\|\mathcal{B}_W(\mathcal{T}_N(s_{N,W}^{(\ell)}))\|_2 = \sqrt{\lambda_{N,W}^{(\ell)}}. \quad (34)$$

---

<sup>8</sup>Note that we use lower-case  $f$  to indicate the digital frequency (so that  $X(f)$  represents the DTFT of a discrete-time sequence  $x[n]$ , while  $X(F)$  represents the CTFT of a continuous-time signal  $x(t)$ ).

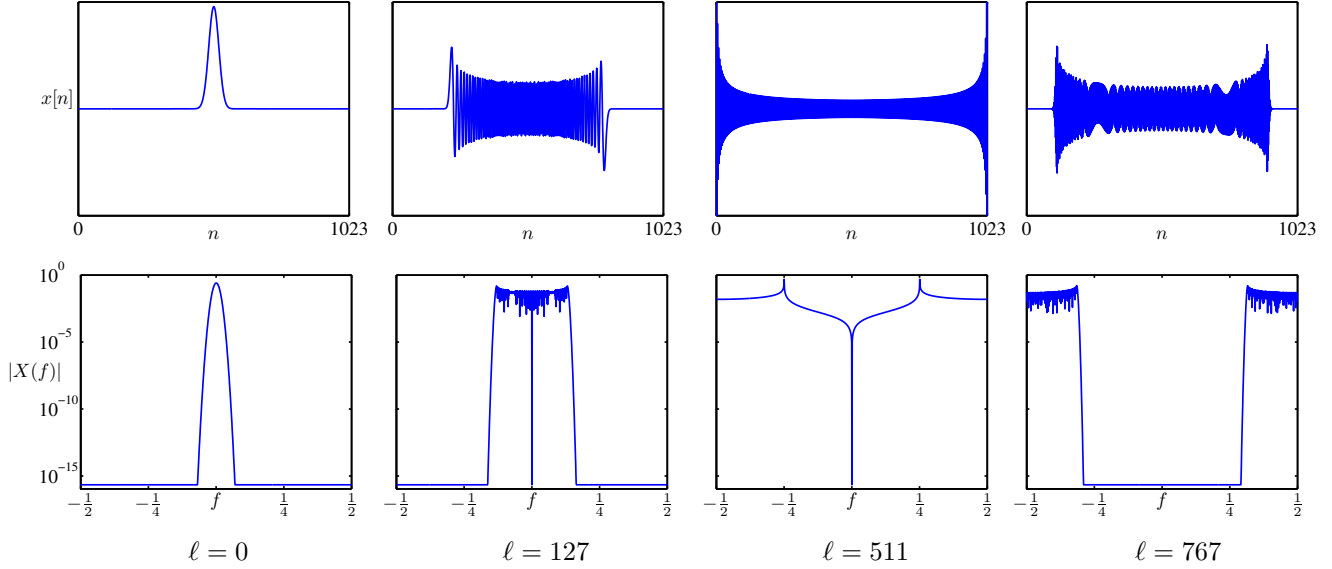


Figure 2: Illustration of four example DPSS's time-limited to the interval  $[0, 1, \dots, N-1]$  and the magnitude of their DTFT's. In this example,  $N = 1024$  and  $W = \frac{1}{4}$ . Note that when  $\ell < 2NW - 1$  the energy of the spectrum is highly concentrated on the interval  $[-W, W]$  and when  $\ell > 2NW - 1$  the energy of the spectrum is concentrated almost entirely outside the interval  $[-W, W]$ .

Comparing (32) with (34), we see that for values of  $\ell$  where  $\lambda_{N,W}^{(\ell)} \approx 1$ , nearly all of the energy in  $\mathcal{T}_N(s_{N,W}^{(\ell)})$  is contained in the frequencies  $|f| \leq W$ . An illustration of four representative time-limited DPSS's and their DTFT's is provided in Figure 2. While by construction the DTFT of any DPSS is perfectly bandlimited, the DTFT of the corresponding time-limited DPSS will only be concentrated in the bandwidth of interest for the first  $2NW$  DPSS's. As a result, we will frequently be primarily interested in roughly the first  $2NW$  DPSS's.

Second, the time-limited DPSS's are also orthogonal [64] so that for any  $\ell, \ell' \in \{0, 1, \dots, N-1\}$  with  $\ell \neq \ell'$ ,

$$\langle \mathcal{T}_N(s_{N,W}^{(\ell)}), \mathcal{T}_N(s_{N,W}^{(\ell')}) \rangle = 0. \quad (35)$$

Finally, like the DPSS's, the time-limited DPSS's have a special eigenvalue relationship with the time-limiting and bandlimiting operators. In particular, if we apply the operator  $\mathcal{T}_N$  to both sides of (31), we see that the sequences  $\mathcal{T}_N(s_{N,W}^{(n)})$  are actually eigenfunctions of the two-step procedure in which one first bandlimits a sequence and then time-limits the sequence.

### 4.2.3 DPSS vectors

Because our focus in this paper is primarily on representing and reconstructing finite-length vectors, we will find the following restriction of the time-limited DPSS's to be useful, where we restrict the domain exclusively to the index range  $n = 0, 1, \dots, N-1$  (discarding the zeros outside this range).

**Definition 4.2.** Given  $N$  and  $W$ , the DPSS vectors  $\mathbf{s}_{N,W}^{(0)}, \mathbf{s}_{N,W}^{(1)}, \dots, \mathbf{s}_{N,W}^{(N-1)} \in \mathbb{R}^N$  are defined by restricting the time-limited DPSS's to the index range  $n = 0, 1, \dots, N-1$ :

$$\mathbf{s}_{N,W}^{(\ell)}[n] := \mathcal{T}_N(s_{N,W}^{(\ell)})[n] = s_{N,W}^{(\ell)}[n]$$

for all  $\ell, n \in \{0, 1, \dots, N-1\}$ .

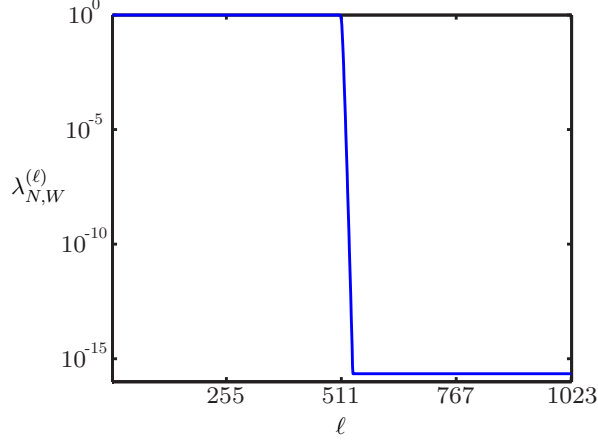


Figure 3: *Eigenvalue concentration for  $N = 1024$  and  $W = \frac{1}{4}$ . Note that the first  $2NW = 512$  eigenvalues cluster near 1 and the remaining eigenvalues rapidly approach the level of machine precision.*

Following from our discussion in Section 4.2.2, the DPSS vectors obey several favorable properties. First, combining (32) and (35), it follows that the DPSS vectors form an orthonormal basis for  $\mathbb{C}^N$  (or for  $\mathbb{R}^N$ ). However, as we discuss in subsequent sections, bases constructed using just the first  $\approx 2NW$  DPSS vectors can be remarkably effective for capturing the energy in our signals of interest. Second, if we define  $\mathbf{B}_{N,W}$  to be the  $N \times N$  matrix with entries given by

$$\mathbf{B}_{N,W}[m, n] := 2W \operatorname{sinc}(2W(m - n)), \quad (36)$$

we see that the eigenvectors of  $\mathbf{B}_{N,W}$  are given by the DPSS vectors  $\mathbf{s}_{N,W}^{(0)}, \mathbf{s}_{N,W}^{(1)}, \dots, \mathbf{s}_{N,W}^{(N-1)}$ , and the corresponding eigenvalues are  $\lambda_{N,W}^{(0)}, \lambda_{N,W}^{(1)}, \dots, \lambda_{N,W}^{(N-1)}$  [64]. Thus, if we concatenate the DPSS vectors into an  $N \times N$  matrix

$$\mathbf{S}_{N,W} := \begin{bmatrix} \mathbf{s}_{N,W}^{(0)} & \mathbf{s}_{N,W}^{(1)} & \cdots & \mathbf{s}_{N,W}^{(N-1)} \end{bmatrix} \in \mathbb{R}^{N \times N}, \quad (37)$$

and let  $\mathbf{\Lambda}_{N,W}$  denote an  $N \times N$  diagonal matrix with the DPSS eigenvalues  $\lambda_{N,W}^{(0)}, \lambda_{N,W}^{(1)}, \dots, \lambda_{N,W}^{(N-1)}$  along the main diagonal, then we can write the eigendecomposition of  $\mathbf{B}_{N,W}$  as

$$\mathbf{B}_{N,W} = \mathbf{S}_{N,W} \mathbf{\Lambda}_{N,W} \mathbf{S}_{N,W}^H. \quad (38)$$

This decomposition will prove useful in our analysis below.

#### 4.2.4 Eigenvalue concentration

As mentioned above, the eigenvalues  $\lambda_{N,W}^{(0)}, \lambda_{N,W}^{(1)}, \dots, \lambda_{N,W}^{(N-1)}$  have a very distinctive and important behavior: the first  $2NW$  eigenvalues tend to cluster extremely close to 1, while the remaining eigenvalues tend to cluster similarly close to 0. This behavior—which will allow us to construct efficient bases using small numbers of DPSS vectors—is illustrated in Figure 3 and captured more formally in the following results.

**Lemma 4.1** (Eigenvalues that cluster near one [64]). *Suppose that  $W$  is fixed, and let  $\epsilon \in (0, 1)$  be fixed. Then there exist constants  $C_1, C_2$  (where  $C_2$  may depend on  $W$  and  $\epsilon$ ) and an integer  $N_0$  (which may also depend on  $W$  and  $\epsilon$ ) such that*

$$\lambda_{N,W}^{(\ell)} \geq 1 - C_1 e^{-C_2 N} \quad \text{for all } \ell \leq 2NW(1 - \epsilon) \text{ and all } N \geq N_0. \quad (39)$$

**Lemma 4.2** (Eigenvalues that cluster near zero [64]). *Suppose that  $W$  is fixed, and let  $\epsilon \in (0, \frac{1}{2W} - 1)$  be fixed. Then there exist constants  $C_3, C_4$  (where  $C_4$  may depend on  $W$  and  $\epsilon$ ) and an integer  $N_1$  (which may also depend on  $W$  and  $\epsilon$ ) such that*

$$\lambda_{N,W}^{(\ell)} \leq C_3 e^{-C_4 N} \quad \text{for all } \ell \geq 2NW(1 + \epsilon) \text{ and all } N \geq N_1. \quad (40)$$

*Alternatively, suppose that  $W$  is fixed, and let  $\alpha > 0$  be fixed. Then there exist constants  $C_5, C_6$  and an integer  $N_2$  (where  $N_2$  may depend on  $W$  and  $\alpha$ ) such that*

$$\lambda_{N,W}^{(\ell)} \leq C_5 e^{-C_6 \alpha} \quad \text{for all } \ell \geq 2NW + \alpha \log(N) \text{ and all } N \geq N_2.$$

On occasion, we will have a need to compute bounds on sums of the eigenvalues. First, we note the following.

**Lemma 4.3.**

$$\sum_{\ell=0}^{N-1} \lambda_{N,W}^{(\ell)} = \text{trace}(\mathbf{B}_{N,W}) = 2NW. \quad (41)$$

The following results will also prove useful.

**Corollary 4.1.** *Suppose that  $W$  is fixed, let  $\epsilon \in (0, 1)$ , and let  $k = 2NW(1 - \epsilon)$ . Then for  $N \geq N_0$ ,*

$$\sum_{\ell=k}^{N-1} \lambda_{N,W}^{(\ell)} \leq 2NW(\epsilon + C_1 e^{-C_2 N}),$$

where  $C_1, C_2$ , and  $N_0$  are as specified in Lemma 4.1.

*Proof.* It follows from (39) and (41) that

$$\begin{aligned} \sum_{\ell=k}^{N-1} \lambda_{N,W}^{(\ell)} &= 2NW - \sum_{\ell=0}^{k-1} \lambda_{N,W}^{(\ell)} \\ &\leq 2NW - 2NW(1 - \epsilon)(1 - C_1 e^{-C_2 N}) \\ &= 2NW(1 - (1 - \epsilon)(1 - C_1 e^{-C_2 N})) \\ &= 2NW(\epsilon + C_1 e^{-C_2 N} - \epsilon C_1 e^{-C_2 N}) \\ &\leq 2NW(\epsilon + C_1 e^{-C_2 N}) \end{aligned} \quad (42)$$

for  $N \geq N_0$ . □

**Corollary 4.2.** *Suppose that  $W$  is fixed, let  $\epsilon \in (0, \frac{1}{2W} - 1)$ , and let  $k = 2NW(1 + \epsilon)$ . Then for  $N \geq \max(N_0, N_1)$ ,*

$$\sum_{\ell=k}^{N-1} \lambda_{N,W}^{(\ell)} \leq 2NW \min \left( \epsilon + C_1 e^{-C_2 N}, \frac{C_3}{2W} e^{-C_4 N} \right), \quad (43)$$

where  $C_1, C_2, C_3, C_4, N_0$ , and  $N_1$  are as specified in Lemma 4.1 and Lemma 4.2.

*Proof.* This result follows simply from Corollary 4.1 and from (40). □

### 4.3 DPSS bases for sampled bandpass signals

Using the DPSS vectors, it is possible to construct remarkably efficient bases for representing the discrete-time vectors that arise when collecting finite numbers of samples from most multiband signals (e.g., signals in  $\mathcal{M}(K, B_{\text{band}})$ ). Before presenting our full construction, however, we illustrate the basic concepts using sampled bandpass signals (e.g., signals in  $\mathcal{M}(1, B_{\text{band}})$ ).

#### 4.3.1 A bandpass modulated DPSS basis

For the moment, we restrict our attention to vectors of samples acquired from a continuous-time bandpass signal  $x(t)$ . We assume that  $\mathcal{F}(x)$ , the support of  $X(F)$ , is restricted to some interval  $[F_c - \frac{B_{\text{band}}}{2}, F_c + \frac{B_{\text{band}}}{2}]$ , where the center frequency  $F_c$  and width  $B_{\text{band}}$  are known. We define  $\mathbf{x} = [x(0) \ x(T_s) \ \cdots \ x((N-1)T_s)]^T$  as in Section 3 to be a finite-length vector of  $N$  samples of  $x(t)$  collected uniformly with a sampling interval  $T_s \leq 1/(2 \max\{|F_c \pm \frac{B_{\text{band}}}{2}|\})$ .

As a basis for efficiently representing many such vectors  $\mathbf{x}$ , we propose the following. First, let  $W = \frac{B_{\text{band}}T_s}{2}$ , and as in (37), let  $\mathbf{S}_{N,W}$  denote the  $N \times N$  matrix containing the  $N$  DPSS vectors (constructed with parameters  $N$  and  $W$ ) as columns. Next, define  $f_c = F_cT_s$  and let  $\mathbf{E}_{f_c}$  denote an  $N \times N$  diagonal matrix with entries

$$\mathbf{E}_{f_c}[m, n] := \begin{cases} e^{j2\pi f_c m}, & m = n, \\ 0, & m \neq n. \end{cases} \quad (44)$$

Multiplying a vector by  $\mathbf{E}_{f_c}$  simply modulates that vector by a frequency  $f_c$ . Finally, consider the  $N \times N$  matrix  $\mathbf{E}_{f_c}\mathbf{S}_{N,W}$ , whose columns are given by the DPSS vectors, each modulated by  $f_c$ . One can easily check that  $\mathbf{E}_{f_c}\mathbf{S}_{N,W}$  forms an orthonormal basis for  $\mathbb{C}^N$ , since  $(\mathbf{E}_{f_c}\mathbf{S}_{N,W})^H \mathbf{E}_{f_c}\mathbf{S}_{N,W} = (\mathbf{S}_{N,W})^H (\mathbf{E}_{f_c})^H \mathbf{E}_{f_c} \mathbf{S}_{N,W} = (\mathbf{S}_{N,W})^H \mathbf{S}_{N,W} = \mathbf{I}$ . For a given integer  $k \in \{1, 2, \dots, N\}$ , we let

$$[\mathbf{E}_{f_c}\mathbf{S}_{N,W}]_k$$

denote the  $N \times k$  matrix formed by taking the first  $k$  columns of  $\mathbf{E}_{f_c}\mathbf{S}_{N,W}$ . We will see that taking  $[\mathbf{E}_{f_c}\mathbf{S}_{N,W}]_k$  with  $k \approx 2NW$  forms an efficient basis that, to a high degree of accuracy, captures most sample vectors  $\mathbf{x}$  that can arise from sampling bandpass signals.

#### 4.3.2 Approximation quality for sampled complex exponentials

To best illustrate one of our key points regarding the approximation of bandpass signals, let us first restrict our focus to the simplest possible bandpass signals: pure complex exponentials. Specifically, consider a continuous-time signal of the form  $x(t) = e^{j2\pi Ft}$  where the frequency  $F$  belongs to the interval  $[F_c - \frac{B_{\text{band}}}{2}, F_c + \frac{B_{\text{band}}}{2}]$ . We define  $\mathbf{x} = [x(0) \ x(T_s) \ \cdots \ x((N-1)T_s)]^T$ ,  $f_c = F_cT_s$ , and  $W = \frac{B_{\text{band}}T_s}{2}$  as in Section 4.3.1. Also, defining

$$\mathbf{e}_f := \begin{bmatrix} e^{j2\pi f 0} \\ e^{j2\pi f T_s} \\ \vdots \\ e^{j2\pi f (N-1)T_s} \end{bmatrix}$$

for all  $f \in [f_c - W, f_c + W]$ , we note that the sample vector  $\mathbf{x}$  will equal  $\mathbf{e}_f$  for  $f = FT_s$ . Without knowing the exact carrier frequency  $F$  in advance, we ask whether it is possible to define an efficient low-dimensional basis for capturing the energy in any sample vector  $\mathbf{x}$  that could arise in this model.

At first glance, this problem may appear difficult or impossible. The infinite set of possible sample vectors  $\{\mathbf{e}_f\}_{f \in [f_c - W, f_c + W]}$  traverses a 1-dimensional manifold (parameterized by  $f$ ) within  $\mathbb{C}^N$ . Technically speaking, these vectors collectively span all of  $\mathbb{C}^N$ . What is remarkable, however, is that to a high degree of accuracy, these vectors are approximately concentrated along a very low-dimensional subspace of  $\mathbb{C}^N$ . Moreover, for any  $k \in \{1, 2, \dots, N\}$ , it is possible to analytically find the best  $k$ -dimensional subspace that minimizes the average squared distance from the vectors  $\mathbf{e}_f$  to the subspace, and this subspace is spanned by the first  $k$  modulated DPSS vectors.

To see this, we let  $\mathbf{Q}$  denote a subspace of  $\mathbb{C}^N$  and  $\mathbf{P}_\mathbf{Q}$  denote the orthogonal projection operator onto  $\mathbf{Q}$ . We would like to minimize

$$\frac{1}{B_{\text{band}}} \cdot \int_{F_c - \frac{B_{\text{band}}}{2}}^{F_c + \frac{B_{\text{band}}}{2}} \|\mathbf{e}_{FT_s} - \mathbf{P}_\mathbf{Q} \mathbf{e}_{FT_s}\|_2^2 dF = \frac{1}{2W} \cdot \int_{f_c - W}^{f_c + W} \|\mathbf{e}_f - \mathbf{P}_\mathbf{Q} \mathbf{e}_f\|_2^2 df \quad (45)$$

over all subspaces  $\mathbf{Q}$  of some prescribed dimension  $k$ . As we show below in Theorem 4.1, this minimization problem can be solved by relating it to the Karhunen-Loeve (KL) transform.<sup>9</sup> For the benefit of the reader, we briefly review the relevant concepts from the KL transform in Appendix A.

**Theorem 4.1.** *For any  $k$  with  $k \in \{1, 2, \dots, N\}$ , the  $k$ -dimensional subspace which minimizes (45) is spanned by the columns of  $\mathbf{Q} = [\mathbf{E}_{f_c} \mathbf{S}_{N,W}]_k$ , i.e., by the first  $k$  DPSS vectors modulated to center frequency  $f_c$ . Furthermore, with this choice of  $\mathbf{Q}$ , we will have*

$$\frac{1}{2W} \cdot \int_{f_c - W}^{f_c + W} \|\mathbf{e}_f - \mathbf{P}_\mathbf{Q} \mathbf{e}_f\|_2^2 df = \frac{1}{2W} \sum_{\ell=k}^{N-1} \lambda_{N,W}^{(\ell)}.$$

For a point of comparison, each sampled sinusoid has energy  $\|\mathbf{e}_f\|_2^2 = N$ .

*Proof.* See Appendix B. □

Recall that the first  $\approx 2NW$  DPSS eigenvalues are very close to 1 and the rest are small, so we are guaranteed a high degree of approximation to the sampled sinusoids if we choose  $k \approx 2NW$ . In particular, if we choose  $k = 2NW(1 - \epsilon)$ , it follows from Corollary 4.1 that

$$\frac{1}{2W} \sum_{\ell=k}^{N-1} \lambda_{N,W}^{(\ell)} \leq N(\epsilon + C_1 e^{-C_2 N})$$

for  $N \geq N_0$ . Alternatively, if we choose  $k = 2NW(1 + \epsilon)$ , it follows from Corollary 4.2 that

$$\frac{1}{2W} \sum_{\ell=k}^{N-1} \lambda_{N,W}^{(\ell)} \leq N \min \left( \epsilon + C_1 e^{-C_2 N}, \frac{C_3}{2W} e^{-C_4 N} \right)$$

for  $N \geq \max(N_0, N_1)$ . Since each sampled sinusoid has energy  $\|\mathbf{e}_f\|_2^2 = N$ , for the subspace we have chosen (say with  $k = 2NW(1 - \epsilon)$ ), the average relative approximation error across the sampled sinusoids is bounded by a small fraction  $\epsilon + C_1 e^{-C_2 N}$  of the energy of a given sinusoid.

It is important to note that, while we are guaranteed a very high degree of approximation accuracy in an MSE sense, we are not guaranteed such accuracy uniformly over *all* sampled sinusoids in our band of interest. A relatively small number of sinusoids may have higher values of  $\|\mathbf{e}_f - \mathbf{P}_\mathbf{Q} \mathbf{e}_f\|_2^2$ , and in practice this diminished approximation performance tends to occur for those sinusoids with frequencies near the edge of the band (i.e., for  $f$  near  $f_c \pm W$ ).

---

<sup>9</sup>The observation that a connection exists between DPSS's and the KL transform is not a new one (see, for example, [32, 38, 48, 57]).

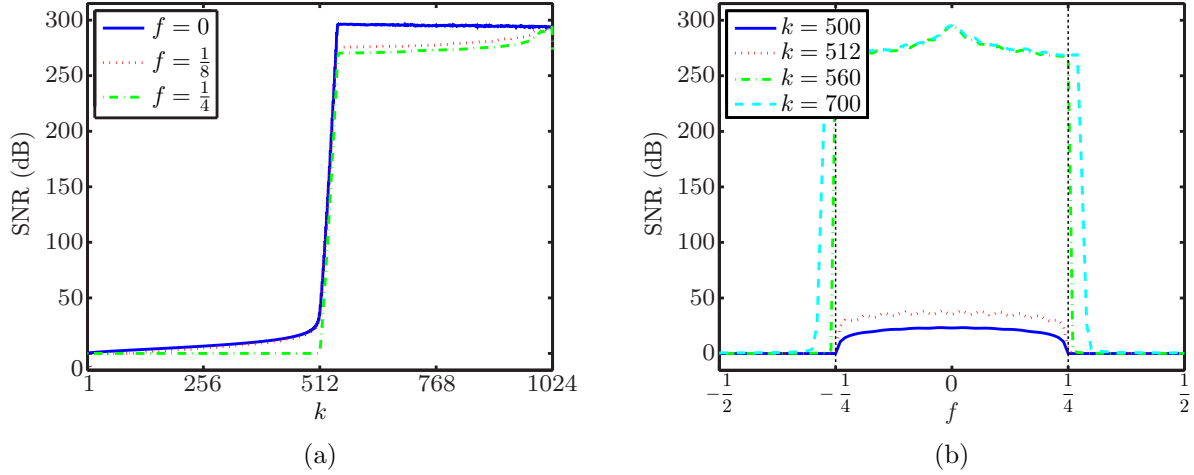


Figure 4: DPSS bases efficiently capture the energy in sampled complex exponentials. (a) SNR captured from three sampled complex exponentials (with differing frequencies  $f$ ) as a function of the number  $k$  of DPSS basis elements. (b) SNR captured as a function of  $f$  for four fixed values of  $k$ . In both plots,  $N = 1024$  and  $W = \frac{1}{4}$ .

This behavior is illustrated in Figure 4. In Figure 4(a) we consider three possible frequencies ( $f = 0$ ,  $\frac{1}{8}$ , or  $\frac{1}{4}$ ) and show the ability of the baseband DPSS basis (with  $f_c = 0$  and  $W = \frac{1}{4}$ ) to capture the energy in these sinusoids as a function of how many DPSS vectors are added to the basis. The ability of a basis to capture a given signal is quantified through

$$\text{SNR} = 20 \log_{10} \left( \frac{\|e_f\|_2}{\|e_f - \mathbf{P} \mathbf{Q} e_f\|_2} \right) \text{dB}.$$

Overall, we observe broadly similar behavior for each frequency in that by adding slightly more than  $2NW$  DPSS vectors to our basis, we capture essentially all of the energy in the original signal. However, we do observe slightly different behavior for the sinusoid with a frequency exactly at  $W$ . In this case we capture very little of the energy in the signal until we have added more than  $2NW$  vectors, while for lower frequencies we begin to do well before this point.

To illustrate this phenomenon, Figure 4(b) considers four different sizes for the DPSS basis and shows the SNR as a function of the frequency of the sinusoid. In the cases where  $k = 500 < 2NW$  and  $k = 2NW$  we see somewhat similar behavior—we are capturing a good fraction, but not all, of any sinusoid whose frequency is not too close to  $W$ . We see a dramatic difference when we increase  $k$  slightly to 560, at which point we are capturing virtually all of the energy in any sinusoid within the band of interest. However, this eventually has a potentially problematic side-effect, as we can see by further increasing  $k$  to 700. Specifically, as we continue to increase the size of the DPSS basis we begin to capture energy of sinusoids that lie outside the targeted band as well. This tradeoff will play an important role in the selection of the appropriate  $k$  in the algorithms we propose below.

#### 4.3.3 Approximation quality for sampled bandpass signals

The above analysis shows that sampled sinusoids, on average, are well approximated by modulated DPSS bases. This is a strong indication that such bases might also be useful for approximating more general sampled bandpass signals, since the vectors  $\{e_f\}_{f \in [f_c - W, f_c + W]} \subset \mathbb{C}^N$  themselves act as “building blocks” for representing sampled, bandpass signals in  $\mathbb{C}^N$ . Formally, for any continuous-time bandpass signal  $x(t)$  with frequency content restricted to the interval  $[F_c - \frac{B_{\text{band}}}{2}, F_c + \frac{B_{\text{band}}}{2}]$ ,

one can show that for each  $n \in \{0, 1, \dots, N-1\}$ ,

$$x[n] := x(nT_s) = \int_{F_c - \frac{B_{\text{band}}}{2}}^{F_c + \frac{B_{\text{band}}}{2}} X(F) e^{j2\pi F n T_s} dF = \int_{f_c - W}^{f_c + W} X(f) \mathbf{e}_f[n] df,$$

where we recall that  $X(F)$  denotes the CTFT of  $x(t)$  and  $X(f)$  denotes the DTFT of  $x[n]$ . So, informally,  $\mathbf{x}$  can be expressed as a linear combination of (infinitely many) sampled complex exponentials  $\mathbf{e}_f$ , where  $f$  ranges from  $f_c - W$  to  $f_c + W$ .

Our analysis from Section 4.3.2 allows us to show that in a certain sense, most continuous-time bandpass signals, when sampled and time-limited, are well-approximated by the modulated DPSS basis. In particular, the following result establishes that bandpass random processes, when sampled and time-limited, are in expectation well-approximated.

**Theorem 4.2.** *Let  $x(t)$  denote a continuous-time, zero-mean, wide sense stationary random process with power spectrum*

$$P_x(F) = \begin{cases} \frac{1}{B_{\text{band}}}, & F \in [F_c - \frac{B_{\text{band}}}{2}, F_c + \frac{B_{\text{band}}}{2}], \\ 0, & \text{otherwise,} \end{cases}$$

and let  $\mathbf{x} = [x(0) \ x(T_s) \ \dots \ x((N-1)T_s)]^T \in \mathbb{C}^N$  denote a finite vector of samples acquired from  $x(t)$  with a sampling interval  $T_s \leq 1/(2 \max\{|F_c \pm \frac{B_{\text{band}}}{2}|\})$ . Then over all  $k$ -dimensional subspaces of  $\mathbb{C}^N$ ,  $\mathbf{x}$  is best approximated (in an MSE sense) by the subspace spanned by the columns of  $\mathbf{Q} = [\mathbf{E}_{f_c} \mathbf{S}_{N,W}]_k$ , where  $f_c = F_c T_s$  and  $W = \frac{B_{\text{band}} T_s}{2}$ . The corresponding MSE is given by

$$\mathbb{E} [\|\mathbf{x} - \mathbf{P}_\mathbf{Q} \mathbf{x}\|_2^2] = \frac{1}{2W} \sum_{\ell=k}^{N-1} \lambda_{N,W}^{(\ell)}, \quad (46)$$

while  $\mathbb{E} [\|\mathbf{x}\|_2^2] = N$ .

*Proof.* See Appendix C. □

As in our discussion following Theorem 4.1, we can ensure that the MSE in (46) is small compared to  $\mathbb{E} [\|\mathbf{x}\|_2^2]$  by choosing  $k \approx 2NW$ . This suggests that in a probabilistic sense, most bandpass signals, when sampled and time-limited, will be well approximated by a small number of modulated DPSS vectors. Again, however, we are not guaranteed such accuracy uniformly over *all* sampled bandpass signals in our band of interest. A relatively small number of bandpass signals  $x(t)$  could lead to sample vectors  $\mathbf{x}$  with higher values of  $\|\mathbf{x} - \mathbf{P}_\mathbf{Q} \mathbf{x}\|_2^2$ . In particular, recalling that the baseband DPSS's are themselves strictly bandlimited, it follows that there exist strictly bandpass signals that when sampled and time-limited yield the modulated DPSS vectors. If we restrict  $\mathbf{Q}$  to the first  $k$  columns of  $\mathbf{E}_{f_c} \mathbf{S}_{N,W}$ , then any bandpass signal producing a sample vector  $\mathbf{x} = \mathbf{E}_{f_c} \mathbf{s}_{N,W}^{(\ell)}$  with  $\ell \geq k$  will have  $\mathbf{P}_\mathbf{Q} \mathbf{x} = \mathbf{0}$  and  $\|\mathbf{x} - \mathbf{P}_\mathbf{Q} \mathbf{x}\|_2^2 = \|\mathbf{x}\|_2^2$ . Fortunately, Theorem 4.2 confirms that such bandpass signals are relatively uncommon: at the risk of belaboring this important point, most bandpass signals, when sampled and time-limited, produce sample vectors approximately in the span of just the first  $k \approx 2NW$  modulated DPSS vectors.

On a related note, signal processing engineers often have a sense for how much “spectral leakage” to anticipate when collecting a finite window of samples of a continuous-time signal. (Frequently, this leakage is reduced via a smooth windowing function [45].) Such practitioners can rest assured that, in every case where the spectral leakage is small outside a bandpass range of frequencies, the resulting sample vector can be well approximated by a small number of modulated DPSS vectors.



**Theorem 4.3.** Let  $x[n] = \mathcal{T}_N(x[n])$  be a time-limited sequence, and suppose that  $x[n]$  is approximately bandlimited to the frequency range  $f \in [f_c - W, f_c + W]$  such that for some  $\delta$ ,

$$\|\mathcal{B}_{f_c, W} x\|_2^2 \geq (1 - \delta) \|x\|_2^2,$$

where  $\mathcal{B}_{f_c, W}$  denotes an orthogonal projection operator that takes a discrete-time signal, bandlimits its DTFT to the frequency range  $f \in [f_c - W, f_c + W]$ , and returns the corresponding signal in the time domain. Let  $\mathbf{x} \in \mathbb{C}^N$  denote the vector formed by restricting  $x[n]$  to the indices  $n = 0, 1, \dots, N - 1$ . Set  $k = 2NW(1 + \epsilon)$  and let  $\mathbf{Q} = [\mathbf{E}_{f_c} \mathbf{S}_{N, W}]_k$ . Then for  $N \geq N_1$ ,

$$\|\mathbf{x} - \mathbf{P}_\mathbf{Q} \mathbf{x}\|_2^2 \leq (\delta + NC_3 e^{-C_4 N}) \|\mathbf{x}\|_2^2, \quad (47)$$

where  $C_3$ ,  $C_4$ , and  $N_1$  are as specified in Lemma 4.2.

*Proof.* See Appendix D. □

Note that, in contrast to Theorem 4.2, Theorem 4.3 requires  $k$  to be slightly larger than  $2NW$ .

## 4.4 DPSS dictionaries for sampled multiband signals

### 4.4.1 A multiband modulated DPSS dictionary

In order to construct an efficient dictionary for sampled multiband signals, we propose simply to concatenate an ensemble of modulated DPSS bases, each one modulated to the center of a band in our model. In particular, in light of the multiband signal model discussed in Section 4.1.2, where the bandwidth  $[-\frac{B_{\text{nyq}}}{2}, \frac{B_{\text{nyq}}}{2}]$  is partitioned into bands  $\Delta_i$  of size  $B_{\text{band}}$ , let us define the midpoint of  $\Delta_i$  as

$$F_i = -\frac{B_{\text{nyq}}}{2} + \left(i + \frac{1}{2}\right) B_{\text{band}}, \quad i \in \{0, 1, \dots, J - 1\},$$

where  $J = \frac{B_{\text{nyq}}}{B_{\text{band}}}$ . Let  $W = \frac{B_{\text{band}} T_s}{2}$  (we assume a sampling interval  $T_s \leq \frac{1}{B_{\text{nyq}}}$ ), and for each  $i$ , let  $f_i = F_i T_s$  and define

$$\Psi_i = [\mathbf{E}_{f_i} \mathbf{S}_{N, W}]_k \quad (48)$$

for some value of  $k$  that we can choose as desired. We construct the multiband modulated DPSS dictionary  $\Psi$  by concatenating all of the  $\Psi_i$ :

$$\Psi = [\Psi_0 \ \Psi_1 \ \dots \ \Psi_{J-1}]. \quad (49)$$

The matrix  $\Psi$  will have size  $N \times kJ$  (note that if  $k = 2NW$  and  $T_s = \frac{1}{B_{\text{nyq}}}$ ,  $\Psi$  will be square).

### 4.4.2 Approximation quality for sampled multiband signals

In a probabilistic sense, most multiband signals, when sampled and time-limited, will be well approximated by a small number of vectors from the multiband modulated DPSS dictionary. In particular, *there exists a block-sparse approximation* for such sample vectors using only the modulated DPSS vectors corresponding to the active signal bands.

**Theorem 4.4.** Let  $\mathcal{I} \subseteq \{0, 1, \dots, J - 1\}$  with  $|\mathcal{I}| = K$ . Suppose that for each  $i \in \mathcal{I}$ ,  $x_i(t)$  is a continuous-time, zero-mean, wide sense stationary random process with power spectrum

$$P_{x_i}(F) = \begin{cases} \frac{1}{KB_{\text{band}}}, & F \in \Delta_i, \\ 0, & \text{otherwise,} \end{cases}$$

and furthermore suppose the  $\{x_i\}_{i \in \mathcal{I}}$  are independent and jointly wide sense stationary. Let  $x(t) = \sum_{i \in \mathcal{I}} x_i(t)$ , and let  $\mathbf{x} = [x(0) \ x(T_s) \ \cdots \ x((N-1)T_s)]^T \in \mathbb{C}^N$  denote a finite vector of samples acquired from  $x(t)$  with a sampling interval of  $T_s \leq \frac{1}{B_{\text{nyq}}}$ . Let  $\Psi_{\mathcal{I}}$  denote the concatenation of the  $\Psi_i$  over all  $i \in \mathcal{I}$ , where the  $\Psi_i$  are as defined in (48). Then

$$\mathbb{E} [\|\mathbf{x} - \mathbf{P}_{\Psi_{\mathcal{I}}} \mathbf{x}\|_2^2] \leq \frac{K}{2W} \sum_{\ell=k}^{N-1} \lambda_{N,W}^{(\ell)}, \quad (50)$$

whereas  $\mathbb{E} [\|\mathbf{x}\|_2^2] = N$ .

*Proof.* See Appendix E. □

Theorem 4.4 confirms the existence of a high-quality block-sparse approximation for most sampled multiband signals; in particular, the signal approximation vector specified in (50) can be written as  $\mathbf{P}_{\Psi_{\mathcal{I}}} \mathbf{x} = \Psi \boldsymbol{\alpha}$  for a  $K$ -block-sparse coefficient vector  $\boldsymbol{\alpha}$  given by  $\boldsymbol{\alpha}|_{\mathcal{I}} = \Psi_{\mathcal{I}}^\dagger \mathbf{x}$  and  $\boldsymbol{\alpha}|_{\mathcal{I}^c} = \mathbf{0}$ . As in our previous discussions for bandpass signals, we can ensure the MSE in (50) is small compared to  $\mathbb{E} [\|\mathbf{x}\|_2^2]$  by choosing  $k \approx 2NW$ . Compared to previous analysis, however, the MSE appearing in Theorem 4.4 is larger by a factor of  $K$  (though this quantity may still be quite small). Although it may be possible to improve upon this figure, the reader should keep in mind that the multiband modulated DPSS dictionary is not necessarily the *optimal* basis for representing sampled multiband signals with a given sparsity pattern  $\mathcal{I}$ ; it is merely a generic (and easily computable) dictionary that provides highly accurate approximations for most multiband signals having any possible sparsity pattern.

Although we omit the details here, one could also consider generalizing Theorem 4.3 to multiband signals that have small spectral leakage when windowed in time.

## 5 Recovering Sampled Multiband Signals from Random Measurements

In this section, we proceed to develop theoretical guarantees for signal recovery using the multiband modulated DPSS dictionary  $\Psi$  as defined in (49). Throughout our theoretical discussion and the subsequent experiments, we pay special attention to the role played the number  $k$  of DPSS vectors per band. We begin in Section 5.1 with a collection of RIP guarantees, and we extend these to signal recovery guarantees in Section 5.2. Throughout this section and the remainder of the paper, we assume that  $T_s \leq \frac{1}{B_{\text{nyq}}}$ .

### 5.1 Embedding guarantees

We can actually immediately establish  $\Psi$ -RIP and  $\Psi$ -block-RIP guarantees for any value of  $k$ . The theorem below follows as a direct consequence of Corollaries 3.1 and 3.3.

**Theorem 5.1.** *Let  $k \in \{1, 2, \dots, N\}$ , set  $D = kJ = \frac{kB_{\text{nyq}}}{B_{\text{band}}}$ , and let  $\Psi$  be the  $N \times D$  multiband modulated DPSS dictionary defined in (49). The following statements hold:*

1. *Fix  $\delta, \beta \in (0, 1)$ , and let  $\mathbf{A}$  be an  $M \times N$  subgaussian matrix with  $M$  satisfying (15). Then with probability exceeding  $1 - \beta$ ,  $\mathbf{A}$  satisfies the  $\Psi$ -RIP of order  $S$  with constant  $\delta$ .*
2. *Fix  $\delta, \beta \in (0, 1)$ , and let  $\mathbf{A}$  be an  $M \times N$  subgaussian matrix with  $M$  satisfying (25). Then with probability exceeding  $1 - \beta$ ,  $\mathbf{A}$  satisfies the  $\Psi$ -block-RIP of order  $K$  with constant  $\delta$ .*

In order to establish RIP and block-RIP bounds for  $\mathbf{A}\Psi$ , however, we must restrict our attention to values of  $k$  that are not too large (this ensures the matrix  $\Psi$  is not overcomplete). To be specific, we note that for any  $k$ , the columns of  $\Psi$  have unit norm. In addition, when  $k$  is suitably small, the columns of  $\Psi$  are approximately orthogonal.

**Lemma 5.1.** *Let  $k = 2NW(1 - \epsilon)$ , set  $D = kJ = N(1 - \epsilon)B_{\text{nyq}}T_s < N$ , and let  $\Psi$  be the  $N \times D$  multiband modulated DPSS dictionary defined in (49). Then for any pair of distinct columns  $\mathbf{q}_1, \mathbf{q}_2$  in  $\Psi$ ,*

$$|\langle \mathbf{q}_1, \mathbf{q}_2 \rangle| \leq 3C_1^{1/2} e^{-\frac{C_2 N}{2}} \quad (51)$$

if  $N \geq N_0$ , where  $C_1, C_2$ , and  $N_0$  are as specified in Lemma 4.1.

*Proof.* See Appendix F.  $\square$

Using this fact, we can ensure that whenever  $k = 2NW(1 - \epsilon)$ ,  $\Psi$  must act as an approximate isometry between any coefficient vector  $\alpha \in \mathbb{C}^D$  and the corresponding signal vector  $\mathbf{x} = \Psi\alpha \in \mathbb{C}^N$ .

**Lemma 5.2.** *Let  $k = 2NW(1 - \epsilon)$ , set  $D = kJ = N(1 - \epsilon)B_{\text{nyq}}T_s < N$ , and let  $\Psi$  be the  $N \times D$  multiband modulated DPSS dictionary defined in (49). Then*

$$\sqrt{1 - 3NC_1^{1/2} e^{-\frac{C_2 N}{2}}} \leq \frac{\|\Psi\alpha\|_2}{\|\alpha\|_2} \leq \sqrt{1 + 3NC_1^{1/2} e^{-\frac{C_2 N}{2}}} \quad (52)$$

for all  $\alpha \in \mathbb{C}^D$ .

*Proof.* The sharpest possible lower and upper bounds in (52) are given by the smallest and largest singular values of  $\Psi$ , respectively. Using standard results from linear algebra,  $\sigma_{\min}(\Psi) = \sqrt{\lambda_{\min}(\Psi^H \Psi)}$  and  $\sigma_{\max}(\Psi) = \sqrt{\lambda_{\max}(\Psi^H \Psi)}$ . The Gram matrix  $\Psi^H \Psi$  has size  $D \times D$ . All entries on the main diagonal of  $\Psi^H \Psi$  are equal to 1, and all entries off of the main diagonal can be bounded using (51). From the Geršgorin circle theorem [35], it follows that all eigenvalues of  $\Psi^H \Psi$  must fall in the interval  $[1 - 3DC_1^{1/2} e^{-\frac{C_2 N}{2}}, 1 + 3DC_1^{1/2} e^{-\frac{C_2 N}{2}}]$ , which for simplicity we note is contained within the interval  $[1 - 3NC_1^{1/2} e^{-\frac{C_2 N}{2}}, 1 + 3NC_1^{1/2} e^{-\frac{C_2 N}{2}}]$  since by assumption  $D < N$ .  $\square$

Lemma 5.2 is the key fact we need to establish RIP and block-RIP bounds for  $\mathbf{A}\Psi$ .

**Theorem 5.2.** *Let  $k = 2NW(1 - \epsilon)$ , set  $D = kJ = N(1 - \epsilon)B_{\text{nyq}}T_s < N$ , and let  $\Psi$  be the  $N \times D$  multiband modulated DPSS dictionary defined in (49). The following statements hold:*

1. *If  $\mathbf{A}$  satisfies the  $\Psi$ -RIP of order  $S$  with constant  $\delta$ , then  $\mathbf{A}\Psi$  satisfies the RIP of order  $S$  with constant  $\delta + 6NC_1^{1/2} e^{-\frac{C_2 N}{2}}$ .*
2. *If  $\mathbf{A}$  satisfies the  $\Psi$ -block-RIP of order  $K$  with constant  $\delta$ , then  $\mathbf{A}\Psi$  satisfies the block-RIP of order  $K$  with constant  $\delta + 6NC_1^{1/2} e^{-\frac{C_2 N}{2}}$ .*

*Proof.* For any  $\alpha \in \mathbb{C}^D$  such that (9) holds, we can use (52) to conclude that

$$\sqrt{1 - \delta} \cdot \sqrt{1 - 3NC_1^{1/2} e^{-\frac{C_2 N}{2}}} \leq \frac{\|\mathbf{A}\Psi\alpha\|_2}{\|\alpha\|_2} \leq \sqrt{1 + \delta} \cdot \sqrt{1 + 3NC_1^{1/2} e^{-\frac{C_2 N}{2}}}.$$

For the upper bound, note that

$$\begin{aligned} (1 + \delta) \left( 1 + 3NC_1^{1/2} e^{-\frac{C_2 N}{2}} \right) &= 1 + \delta + 3NC_1^{1/2} e^{-\frac{C_2 N}{2}} + \delta 3NC_1^{1/2} e^{-\frac{C_2 N}{2}} \\ &\leq 1 + \delta + 6NC_1^{1/2} e^{-\frac{C_2 N}{2}}, \end{aligned}$$

where the second line follows from the assumption that  $\delta < 1$ . The lower bound follows similarly.  $\square$

## 5.2 Recovery guarantees

In this section we prove that with a sufficient number of measurements and an appropriately constructed multiband modulated DPSS dictionary, most sample vectors of multiband signals can be accurately reconstructed. Our proof of this fact relies on two principles.

### 5.2.1 Recovery of exactly block-sparse signals

The first of these principles is that, as a consequence of our RIP results in Section 5.1, signal vectors having representations that are exactly  $K$ -block-sparse in the dictionary  $\Psi$  can be accurately reconstructed from compressive samples. The following two results follow from combining Theorems 3.1, 3.2, 5.1, and 5.2.

**Theorem 5.3.** *Let  $k \in \{1, 2, \dots, N\}$ , set  $D = kJ = \frac{kB_{\text{nyq}}}{B_{\text{band}}}$ , and let  $\Psi$  be the  $N \times D$  multiband modulated DPSS dictionary defined in (49). Fix  $\delta \in (0, 0.2)$  and  $\beta \in (0, 1)$ , and let  $\mathbf{A}$  be an  $M \times N$  subgaussian matrix with*

$$M \geq \frac{4K \left( \frac{D}{J} \log(42/\delta) + \log(eJ/2K) \right) + \log(4/\beta)}{c_1(\delta/\sqrt{2})}. \quad (53)$$

*Then with probability exceeding  $1 - \beta$ , the following statement holds: For any  $\mathbf{x}' \in \mathbb{C}^N$  that has a  $K$ -block-sparse representation in the dictionary  $\Psi$  (i.e., that can be written as  $\mathbf{x}' = \Psi \boldsymbol{\alpha}'$  for some  $K$ -block-sparse vector  $\boldsymbol{\alpha}' \in \mathbb{C}^D$ ), if we use block-based IHT (23) with  $\frac{1}{\mu} \in [1 + \delta, 1.5(1 - \delta)]$  to recover an estimate  $\hat{\mathbf{x}}$  of  $\mathbf{x}'$  from the observations  $\mathbf{y} = \mathbf{A}\mathbf{x}' + \mathbf{e}$ , the resulting  $\hat{\mathbf{x}}$  will satisfy*

$$\|\mathbf{x}' - \hat{\mathbf{x}}\|_2 \leq \kappa_1 \|\mathbf{e}\|_2, \quad (54)$$

where  $\kappa_1 > 1$  is as specified in Theorem 3.1.

**Theorem 5.4.** *Let  $k = 2NW(1 - \epsilon)$ , set  $D = kJ = N(1 - \epsilon)B_{\text{nyq}}T_s < N$ , and let  $\Psi$  be the  $N \times D$  multiband modulated DPSS dictionary defined in (49). Fix  $\delta \in (0, 0.1 - 6NC_1^{1/2}e^{-\frac{C_2N}{2}})$  and  $\beta \in (0, 1)$ . Let  $\mathbf{A}$  be an  $M \times N$  subgaussian matrix with*

$$M \geq \frac{8K \left( \frac{D}{J} \log(42/\delta) + \log(eJ/4K) \right) + \log(4/\beta)}{c_1(\delta/\sqrt{2})} \quad (55)$$

*Then with probability exceeding  $1 - \beta$ , the following statement holds: For any  $\mathbf{x}' \in \mathbb{C}^N$  that has a  $K$ -block-sparse representation in the dictionary  $\Psi$  (i.e., that can be written as  $\mathbf{x}' = \Psi \boldsymbol{\alpha}'$  for some  $K$ -block-sparse vector  $\boldsymbol{\alpha}' \in \mathbb{C}^D$ ), if we use block-based CoSaMP (Algorithm 2) with  $\hat{\boldsymbol{\alpha}} = \text{BBCoSaMP}(\mathbf{A}\Psi, \mathbf{I}, \mathbf{y}, K)$  to recover an estimate  $\hat{\boldsymbol{\alpha}}$  of  $\boldsymbol{\alpha}'$  from the observations  $\mathbf{y} = \mathbf{A}\mathbf{x}' + \mathbf{e}$ , the resulting  $\hat{\boldsymbol{\alpha}}$  will satisfy*

$$\|\boldsymbol{\alpha}' - \hat{\boldsymbol{\alpha}}\|_2 \leq \kappa_2 \|\mathbf{e}\|_2 \quad (56)$$

where  $\kappa_2 > 1$  is as specified in Theorem 3.2.

### 5.2.2 Approximating sampled multiband signals with exactly block-sparse signals

The second of our principles is that most multiband signals with  $K$  occupied bands, when sampled and time-limited, have a high-quality  $K$ -block-sparse representation in the dictionary  $\Psi$ . Let  $x(t)$  denote a continuous-time multiband signal with nonzero support on blocks indexed by  $\mathcal{I} \subseteq \{0, 1, \dots, J - 1\}$ , where  $|\mathcal{I}| = K$ . Let  $\mathbf{x} = [x(0) \ x(T_s) \ \dots \ x((N - 1)T_s)]^T \in \mathbb{C}^N$  denote a finite

vector of samples acquired from  $x(t)$  with a sampling interval of  $T_s \leq \frac{1}{B_{\text{nyq}}}$ . Let  $\alpha'$  be a  $K$ -block-sparse coefficient vector given by  $\alpha'|_{\mathcal{I}} = \Psi_{\mathcal{I}}^\dagger \mathbf{x}$  and  $\alpha'|_{\mathcal{I}^c} = \mathbf{0}$ . Defining  $\mathbf{x}' := \Psi \alpha'$  and  $\mathbf{e}_x := \mathbf{x} - \Psi \alpha' = \mathbf{x} - \mathbf{x}'$ , we can then write

$$\mathbf{x} = \mathbf{x}' + \mathbf{e}_x, \quad (57)$$

where  $\mathbf{x}'$  has an exactly  $K$ -block-sparse representation in the dictionary  $\Psi$  and we expect  $\mathbf{e}_x$  to be small.

We can more formally bound the size of  $\mathbf{e}_x$ . For example, under the multiband random process model for  $x(t)$  described in Theorem 4.4, we will have  $\mathbb{E} [\|\mathbf{e}_x\|_2^2] \leq \frac{K}{2W} \sum_{\ell=k}^{N-1} \lambda_{N,W}^{(\ell)}$ . By setting the number of columns per band  $k$  to be on the order of  $2NW$ , we can make this error small relative to  $\mathbb{E} [\|\mathbf{x}\|_2^2] = N$ . For example, if we take  $k = 2NW(1 - \epsilon)$  for some  $\epsilon \in (0, 1)$ , Corollary 4.1 allows us to conclude that

$$\mathbb{E} [\|\mathbf{e}_x\|_2^2] \leq \frac{K}{2W} \sum_{\ell=k}^{N-1} \lambda_{N,W}^{(\ell)} \leq KN(\epsilon + C_1 e^{-C_2 N}). \quad (58)$$

for  $N \geq N_0$ . The rightmost upper bound in (58) can be made as small as desired (relative to  $\mathbb{E} [\|\mathbf{x}\|_2^2] = N$ ) by choosing  $\epsilon$  sufficiently small and  $N$  sufficiently large.

For any value of  $k$ , we can also establish a tail bound on  $\|\mathbf{e}_x\|_2^2$ , guaranteeing that it is unlikely to significantly exceed the quantity  $\frac{K}{2W} \sum_{\ell=k}^{N-1} \lambda_{N,W}^{(\ell)}$ . Using standard concentration of measure arguments for subgaussian and subexponential random variables (see [75] for a thorough discussion), one can make the following guarantee on the squared norm of a Gaussian random vector.

**Lemma 5.3.** *Let  $\mathbf{z} \in \mathbb{C}^N$  be a complex-valued Gaussian random vector with mean zero. Then*

$$\mathbb{P} [|\|\mathbf{z}\|_2^2 - \mathbb{E} [\|\mathbf{z}\|_2^2]| \geq \gamma \mathbb{E} [\|\mathbf{z}\|_2^2]] \leq 2 \exp \left\{ -\min \left( \frac{\gamma^2 (\sum_n \lambda_n)^2}{c_2^2 \sum_n \lambda_n^2}, \frac{\gamma \sum_n \lambda_n}{c_2 \max_n \lambda_n} \right) \right\},$$

where  $c_2$  is a universal constant and  $\{\lambda_n\}$  denote the eigenvalues of the autocorrelation matrix of the length- $2N$  random vector  $[\text{Re}(\mathbf{z})^T \text{Im}(\mathbf{z})^T]^T$ .

If we assume in our multiband random process model that  $x(t)$  is a Gaussian random process, this will imply that  $\mathbf{x}$  is a Gaussian random vector, and since  $\mathbf{e}_x$  is a linear transformation of  $\mathbf{x}$ ,  $\mathbf{e}_x$  will be Gaussian as well. This allows us to apply Lemma 5.3 to the quantity  $\|\mathbf{e}_x\|_2^2$ . Using the fact that  $\|\boldsymbol{\lambda}\|_1 \geq \|\boldsymbol{\lambda}\|_2 \geq \|\boldsymbol{\lambda}\|_\infty$  for any vector  $\boldsymbol{\lambda}$ , we can pessimistically simplify this bound to read:

$$\mathbb{P} \left[ \|\mathbf{e}_x\|_2^2 \geq (1 + \gamma) \frac{K}{2W} \sum_{\ell=k}^{N-1} \lambda_{N,W}^{(\ell)} \right] \leq 2 \exp \left\{ -\min \left( \frac{\gamma^2}{c_2^2}, \frac{\gamma}{c_2} \right) \right\}. \quad (59)$$

### 5.2.3 Combined guarantees

To put these two principles together, we note that when taking noise-free compressive measurements  $\mathbf{A}\mathbf{x}$  of a signal vector  $\mathbf{x}$  obeying (57), we will have  $\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x}' + \mathbf{A}\mathbf{e}_x$ . This allows us to invoke Theorems 5.3 and 5.4 with  $\mathbf{e} := \mathbf{A}\mathbf{e}_x$ ,<sup>10</sup> and when the number of columns per band  $k$  is chosen so that we expect  $\mathbf{e}_x$  to be small, the concentration of measure phenomenon tells us that  $\mathbf{e}$  should

<sup>10</sup>One could easily incorporate actual measurement noise into  $\mathbf{e}$  as well (and thus extend Theorems 5.5 and 5.6 to the case of noisy measurements). However, for the sake of clarity we simply set  $\mathbf{e} := \mathbf{A}\mathbf{e}_x$  in order to highlight the impact of modeling error in our main results.

be small as well. In particular, note that for fixed  $\mathbf{e}_x$  and random subgaussian  $\mathbf{A}$ , (11) guarantees that

$$\mathbb{P} \left[ \left| \|\mathbf{e}\|_2^2 - \|\mathbf{e}_x\|_2^2 \right| \geq \eta \|\mathbf{e}_x\|_2^2 \right] \leq 4e^{-c_1(\eta)M}. \quad (60)$$

Then, for example, if we let  $\hat{\mathbf{x}}$  denote the estimated signal vector recovered via block-based IHT, (60) allows us to write

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \leq \|\mathbf{x} - \mathbf{x}'\|_2 + \|\mathbf{x}' - \hat{\mathbf{x}}\|_2 \leq \|\mathbf{e}_x\|_2 + \kappa_1 \|\mathbf{e}\|_2 \leq (1 + \kappa_1(1 + \eta)^{1/2})\|\mathbf{e}_x\|_2,$$

where the second inequality follows from Theorem 5.3 and the third inequality holds with probability at least  $1 - 4e^{-c_1(\eta)M}$ . Combining this fact with the tail bound (59), we can establish the following guarantee.

**Theorem 5.5.** *Let  $k \in \{1, 2, \dots, N\}$ , set  $D = kJ = \frac{kB_{\text{nyq}}}{B_{\text{band}}}$ , and let  $\Psi$  be the  $N \times D$  multiband modulated DPSS dictionary defined in (49). Fix  $\delta \in (0, 0.2)$  and  $\beta \in (0, 1)$ , and let  $\mathbf{A}$  be an  $M \times N$  subgaussian matrix with  $M$  satisfying (53). If  $x(t)$  is a Gaussian random process obeying the  $K$ -band model described in Theorem 4.4,  $\mathbf{x} \in \mathbb{C}^N$  is generated by sampling  $x(t)$  as in Theorem 4.4, and we use block-based IHT (23) with  $\frac{1}{\mu} \in [1 + \delta, 1.5(1 - \delta)]$  to recover an estimate of  $\mathbf{x}$  from the observations  $\mathbf{y} = \mathbf{A}\mathbf{x}$ , then the resulting estimate  $\hat{\mathbf{x}}$  will satisfy*

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \leq (1 + \kappa_1(1 + \eta)^{1/2})^2(1 + \gamma) \frac{K}{2W} \sum_{\ell=k}^{N-1} \lambda_{N,W}^{(\ell)} \quad (61)$$

with probability at least  $1 - \beta - 4e^{-c_1(\eta)M} - 2e^{-\min(\gamma^2/c_2^2, \gamma/c_2)}$ , where  $W = \frac{B_{\text{band}}T_s}{2}$  as specified in the dictionary construction (49),  $\kappa_1 > 1$  is a constant as specified in Theorem 3.1,  $c_1$  is a constant as specified in (11), and  $c_2$  is a universal constant.

Although the above result holds for any  $k \in \{1, 2, \dots, N\}$ , it is perhaps most interesting when the number of columns per band  $k$  is chosen to be on the order of  $2NW$ . For example, if we take  $k = 2NW(1 - \epsilon)$ , (61) will read

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \leq (1 + \kappa_1(1 + \eta)^{1/2})^2(1 + \gamma)(\epsilon + C_1 e^{-C_2 N})KN. \quad (62)$$

Supposing that  $K$ ,  $\kappa_1$ ,  $\eta$ , and  $\gamma$  are fixed, one can make the upper bound appearing in (62) as small as desired (relative to  $\mathbb{E}[\|\mathbf{x}\|_2^2] = N$ ) by choosing  $\epsilon$  sufficiently small and  $N$  sufficiently large.<sup>11</sup> Specifically, the term  $\epsilon + C_1 e^{-C_2 N}$  can be made arbitrarily close to zero by choosing  $\epsilon$  sufficiently small and  $N$  sufficiently large. Of course, as one increases the length  $N$  of the signal vector, the requisite number  $M$  of measurements will increase proportionally (this is captured in (53) via the dependence on  $D$ ). What is important about the measurement bound is that  $\frac{M}{N}$ , the permitted undersampling ratio relative to the Nyquist sampling rate (supposing  $T_s = \frac{1}{B_{\text{nyq}}}$ ), will scale like

$$\frac{M}{N} \sim \frac{KB_{\text{band}}}{B_{\text{nyq}}}.$$

In other words, we need only collect compressive samples at a rate proportional to  $KB_{\text{band}}$ , the total amount of occupied bandwidth (i.e., the so-called Landau rate [40]).

<sup>11</sup>One could also use Lemma 5.3 to guarantee that, with high probability,  $\|\mathbf{x}\|_2^2$  will not be too small compared to  $N$ . We omit these details in the interest of conciseness.

For sufficiently small  $k$ , we can establish a similar guarantee for block-based CoSaMP. Let  $\hat{\alpha}$  be the recovered coefficient vector, and define  $\hat{\mathbf{x}} := \Psi\hat{\alpha}$ . Then we can write

$$\begin{aligned}\|\mathbf{x} - \hat{\mathbf{x}}\|_2 &\leq \|\mathbf{x} - \mathbf{x}'\|_2 + \|\Psi\alpha' - \Psi\hat{\alpha}\|_2 \\ &\leq \|\mathbf{e}_{\mathbf{x}}\|_2 + \sqrt{1 + 3NC_1^{1/2}e^{-\frac{C_2N}{2}}}\|\alpha' - \hat{\alpha}\|_2 \\ &\leq \|\mathbf{e}_{\mathbf{x}}\|_2 + \kappa_2\sqrt{1 + 3NC_1^{1/2}e^{-\frac{C_2N}{2}}}\|\mathbf{e}\|_2 \\ &\leq (1 + \kappa_2(1 + 3NC_1^{1/2}e^{-\frac{C_2N}{2}})^{1/2}(1 + \eta)^{1/2})\|\mathbf{e}_{\mathbf{x}}\|_2,\end{aligned}$$

where the second line follows from Lemma 5.2, the third line follows from Theorem 5.4, and the last line holds with probability at least  $1 - 4e^{-c_1(\eta)M}$ . Combining this fact with Corollary 4.1 and the tail bound (59), we can establish the following guarantee.

**Theorem 5.6.** *Let  $k = 2NW(1 - \epsilon)$ , set  $D = kJ = N(1 - \epsilon)B_{\text{nyq}}T_s < N$ , and let  $\Psi$  be the  $N \times D$  multiband modulated DPSS dictionary defined in (49). Fix  $\delta \in (0, 0.1 - 6NC_1^{1/2}e^{-\frac{C_2N}{2}})$  and  $\beta \in (0, 1)$ , and let  $\mathbf{A}$  be an  $M \times N$  subgaussian matrix with  $M$  satisfying (55). If  $x(t)$  is a Gaussian random process obeying the  $K$ -band model described in Theorem 4.4,  $\mathbf{x} \in \mathbb{C}^N$  is generated by sampling  $x(t)$  as in Theorem 4.4, we use block-based CoSaMP (Algorithm 2) with  $\hat{\alpha} = \text{BBCoSaMP}(\mathbf{A}\Psi, \mathbf{I}, \mathbf{y}, K)$  to recover an estimate of  $\alpha$  from the observations  $\mathbf{y} = \mathbf{A}\mathbf{x}$ , and we set  $\hat{\mathbf{x}} := \Psi\hat{\alpha}$ , then the resulting estimate will satisfy*

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \leq (1 + \kappa_2(1 + 3NC_1^{1/2}e^{-\frac{C_2N}{2}})^{1/2}(1 + \eta)^{1/2})^2(1 + \gamma)(\epsilon + C_1e^{-C_2N})KN \quad (63)$$

with probability at least  $1 - \beta - 4e^{-c_1(\eta)M} - 2e^{-\min(\gamma^2/c_2^2, \gamma/c_2)}$ , where  $\kappa_2 > 1$  is a constant as specified in Theorem 3.2,  $C_1$  and  $C_2$  are constants as specified in Lemma 4.1,  $c_1$  is a constant as specified in (11), and  $c_2$  is a universal constant.

Once again, supposing that  $K$ ,  $\kappa_2$ ,  $\eta$ , and  $\gamma$  are fixed, one can make the upper bound appearing in (63) as small as desired (relative to  $\mathbb{E}[\|\mathbf{x}\|_2^2] = N$ ) by choosing  $\epsilon$  sufficiently small and  $N$  sufficiently large. Specifically, the terms  $3NC_1^{1/2}e^{-\frac{C_2N}{2}}$  and  $\epsilon + C_1e^{-C_2N}$  can be made arbitrarily close to zero by choosing  $\epsilon$  sufficiently small and  $N$  sufficiently large. Thus, we have again guaranteed that most finite-length sample vectors arising from multiband analog signals can—to a very high degree of approximation—be recovered from a number of compressive measurements that is proportional to the underlying information level.

## 6 Simulations

In this section we present the results of a suite of simulations that demonstrate the effectiveness of our proposed approaches to multiband signal recovery from compressive measurements. In doing so, we address various practical considerations that arise within our framework. All of our simulations were performed via a MATLAB software package that we have made available for download at <http://www.mines.edu/~mwakin/software/>. This software package contains all of the code and results necessary to reproduce the experiments and figures described below, but should additionally serve as a platform upon which other researchers can test and develop their own extensions of these ideas.

### 6.1 Implementation and experimental setup

We begin with a brief discussion regarding our implementation and experimental setup.

### 6.1.1 Computing $\mathcal{P}(\mathbf{x}, K)$

We first recall that a key step in solving either block-based IHT (specifically, the variation in (23)) or block-based CoSaMP (specifically, the variation  $\hat{\mathbf{x}} = \text{BBCoSaMP}(\mathbf{A}, \Psi, \mathbf{y}, K)$ ) is the computation of  $\mathcal{P}(\mathbf{x}, K)$ , i.e., the projection of  $\mathbf{x}$  onto the set of  $K$ -block sparse signals in the dictionary  $\Psi$  (as defined in (22)). If  $\Psi$  were an orthonormal basis, this projection could be computed simply by taking  $\Psi^H \mathbf{x}$  and setting to zero all but the  $K$  blocks of this vector with the highest energy. Unfortunately, the columns of the multiband modulated DPSS dictionary  $\Psi$  (as defined in (49)) are not orthogonal, and so this thresholding approach is not guaranteed to even approximate the correct solution. In fact, when the number of columns per band  $k$  exceeds  $2NW(1-\epsilon)$ ,  $\Psi$  will fail to act as an isometry (recall Lemma 5.2), and when  $k$  exceeds  $2NW$ ,  $\Psi$  will actually be overcomplete. As we will see in our experiments, it can often be desirable to choose  $k$  to be larger than  $2NW$ , but this lack of isometry and this overcompleteness prevent us from applying any of the standard arguments from sparse approximation to solving the projection problem.

It is important to note that in (22) we are seeking a vector  $\mathbf{z}$  that minimizes  $\|\mathbf{x} - \mathbf{z}\|_2$  and has a block-sparse representation in  $\Psi$ . Although  $\mathbf{z}$  will be unique, the corresponding representation in  $\Psi$  need not be unique if  $\Psi$  fails to satisfy the block-RIP. Unfortunately, there seems to be relatively little known about solving this type of sparse approximation problem. Nevertheless, since we are ultimately interested in  $\mathbf{z}$  (not its representation with respect to  $\Psi$ ) this should not necessarily stop us from proceeding under the hope that standard sparse approximation algorithms can still succeed even when  $\Psi$  is moderately overcomplete. Thus, in our simulations, we use block-OMP to obtain an approximation to  $\mathcal{P}(\mathbf{x}, K)$ . Block-OMP is a straightforward generalization of the classical OMP algorithm. It proceeds by: (i) initializing the residual vector  $\mathbf{r} = \mathbf{x}$  and the set  $\mathcal{I} = \emptyset$ , (ii) computing the proxy vector  $\mathbf{h} = \Psi^H \mathbf{r}$ , (iii) identifying the block in  $\mathbf{h}$  that has the largest energy and adding this block to the set  $\mathcal{I}$ , and (iv) orthogonalizing  $\mathbf{r}$  against the columns in  $\Psi_{\mathcal{I}}$ . This last step is equivalent to the “update” step in CoSaMP in Algorithms 1 and 2. Steps (ii) through (iv) are repeated until termination. The final output can be computed from  $\mathbf{x} - \mathbf{r}^K$ , where  $\mathbf{r}^K$  is the residual after  $K$  iterations.

We close by noting that the lack of a provable technique for computing  $\mathcal{P}(\mathbf{x}, K)$  represents an important gap between some of our theory and practice. Specifically, one of our main results (Theorem 5.5) pertains to block-based IHT (23), but this is an algorithm that we can only implement approximately. As noted at the end of Section 3.2.3, we conjecture that one could also establish theoretical guarantees for the  $\text{BBCoSaMP}(\mathbf{A}, \Psi, \mathbf{y}, K)$  version of block-based CoSaMP, but this too relies on being able to compute  $\mathcal{P}(\mathbf{x}, K)$  exactly. In light of this, we point out the following. First, we are able to implement the  $\text{BBCoSaMP}(\mathbf{A}\Psi, \mathbf{I}, \mathbf{y}, K)$  version of block-based CoSaMP exactly because this problem can be solved with simple block thresholding of the vector  $\boldsymbol{\alpha}$ , and one of our main results (Theorem 5.6) does pertain specifically to  $\text{BBCoSaMP}(\mathbf{A}\Psi, \mathbf{I}, \mathbf{y}, K)$ . Second, our simulations in Sections 6.2 and beyond, which rely on block-OMP for approximating  $\mathcal{P}(\mathbf{x}, K)$ , indicate that we are clearly finding high quality solutions to this problem despite the lack of provable guarantees. What is interesting is that our experimental results seem most favorable when  $k$  exceeds  $2NW$  by a nontrivial amount, and that is a regime where the dictionary  $\Psi$  is overcomplete. We hope to further address the implications of overcomplete  $\Psi$  in future work.

### 6.1.2 Regularized least-squares

In our simulations below, we focus exclusively on testing the two versions of block-based CoSaMP. One could conceivably expect similar experimental results using a properly-tuned version of block-based IHT, but since we lack theory concerning  $\mathcal{P}(\mathbf{x}, K)$  we find it more worthwhile to devote



our space to exploring the various practical issues surrounding block-based CoSaMP. Specifically, let us note that when solving block-based CoSaMP, the key steps in this algorithm consist of solving the least-squares problem in the “update” step and of computing  $\mathcal{P}(\mathbf{x}, K)$  (twice), which also involves solving (potentially multiple) least-squares problems. Thus, it is crucial that we solve these least-squares problems efficiently and accurately.

While there has been much work in the CS community on solving these problems efficiently (for example, see [49]), we leave aside the issue of speed for the moment and focus instead on the issue of stability. In our context, we must take some care in solving these problems, since when the number of columns  $k$  per band becomes much larger than  $2NW$ , the matrices involved can become potentially rank deficient and thus standard approaches can be numerically unstable. Fortunately, this instability can be easily addressed using relatively simple techniques.

We first consider the least-squares problem that must be solved in the computation of  $\mathcal{P}(\mathbf{x}, K)$ . In this case we will not focus specifically on the block-OMP algorithm that we implement, but simply assume that we have some technique for obtaining an estimate  $\mathcal{I}$  of  $\mathcal{S}(\mathbf{x}, K)$ . At that point, solving (22) is relatively straightforward. One could solve this via  $\mathcal{P}(\mathbf{x}, K) = \Psi_{\mathcal{I}} \Psi_{\mathcal{I}}^\dagger \mathbf{x}$ , but when  $\mathcal{I}$  contains indices corresponding to adjacent bands,  $\Psi_{\mathcal{I}}$  can be nearly rank deficient. In this case, a better approach is to construct a reduced basis  $\mathbf{U}$  for  $\mathcal{R}(\Psi_{\mathcal{I}})$ . One can then compute the projection simply via  $\mathcal{P}(\mathbf{x}, K) = \mathbf{U} \mathbf{U}^H \mathbf{x}$ . In our context, we use this reduced basis approach to perform the orthogonalization in each step of the block-OMP approach to computing  $\mathcal{P}(\mathbf{x}, K)$ .

We now turn to the main least-squares problem in the “update” step of CoSaMP. In this step we are given a set of blocks  $\mathcal{I}$  and wish to solve a problem of the form

$$\tilde{\mathbf{x}} = \arg \min_{\mathbf{z}} \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2 \quad \text{s.t.} \quad \mathbf{z} \in \mathcal{R}(\Psi_{\mathcal{I}}). \quad (64)$$

Note that in this case we require more than simply the projection of  $\mathbf{y}$  onto  $\mathcal{R}(\mathbf{A}\Psi_{\mathcal{I}})$ —we wish to actually calculate the vector  $\mathbf{z}$  corresponding to this projection. This requires a different approach than the simple reduced basis approach described above (although constructing a reduced basis for  $\mathcal{R}(\Psi_{\mathcal{I}})$  can still be useful in this context). In our simulations we regularize (64) via *Tikhonov regularization* [54, 69, 70]. The key idea is to replace (64) with

$$\tilde{\mathbf{x}} = \arg \min_{\mathbf{z}} \|\mathbf{y} - \mathbf{A}\mathbf{z}\|_2 \quad \text{s.t.} \quad \mathbf{z} \in \mathcal{R}(\Psi_{\mathcal{I}}), \quad \|\mathbf{z}\|_2 \leq \gamma. \quad (65)$$

The constraint  $\|\mathbf{z}\|_2 \leq \gamma$  significantly improves the conditioning of this problem when  $\mathbf{A}\Psi_{\mathcal{I}}$  is ill-conditioned. In our simulations below, we use the toolbox of [37] to efficiently solve (65). Note also that in our setting, the parameter  $\gamma$  can be easily set using the norm of the original signal  $\mathbf{x}$  that we are acquiring. In all simulations below (using either variation of block-based CoSaMP), we assume that an upper bound on  $\|\mathbf{x}\|_2$  is known. In practical settings such an upper bound will usually be available, but if necessary one could also estimate this upper bound from  $\|\mathbf{y}\|_2$ .

### 6.1.3 Experimental setup

In all of the experiments below, we assume that  $T_s = \frac{1}{B_{\text{nyq}}}$  and that there are  $J = \frac{B_{\text{nyq}}}{B_{\text{band}}} = 256$  possible bands. For each value of  $k$  that we consider, we set  $D = kJ$  and let  $\Psi$  be the  $N \times D$  multiband modulated DPSS dictionary defined in (49). The digital half-bandwidth parameter  $W$  is set to be  $W = \frac{B_{\text{band}} T_s}{2} = \frac{1}{512}$ , and we consider sample vectors with length  $N = 4096$ , so that  $2NW = 16$ .

In our experiments we generate our sampled multiband signal vectors  $\mathbf{x}$  by selecting the positions of the  $K$  occupied bands uniformly at random from the  $J$  possibilities, and then within each

band adding together 50 complex exponentials with frequencies selected uniformly at random from within the frequency band (not aligned with the “Nyquist grid”). Each complex exponential in this summation is given a random amplitude and phase via multiplication by a complex Gaussian random variable. There are a variety of other possibilities for generating test signals which we considered and for which we have observed essentially the same results as presented below.

We will typically report our results in terms of the SNR of the recovery, which in this context is defined as

$$\text{SNR} = 20 \log_{10} \left( \frac{\|\mathbf{x}\|_2}{\|\mathbf{x} - \hat{\mathbf{x}}\|_2} \right) \text{dB},$$

where  $\mathbf{x}$  is the original input to the measurement matrix and  $\hat{\mathbf{x}}$  is the recovered estimate of  $\mathbf{x}$  provided by CoSaMP. In all cases where we plot the SNR, what we actually show is the result of 50 independent trials (with a new signal for each trial). Rather than plotting the mean SNR, we plot the contour for which 95% of trials result in an SNR *at least* as large as the level indicated (so that only 5% of trials result in a worse performance than indicated in the curves below).

## 6.2 BBCoSaMP( $\mathbf{A}, \Psi, \mathbf{y}, K$ ) versus BBCoSaMP( $\mathbf{A}\Psi, \mathbf{I}, \mathbf{y}, K$ )

We begin with a comparison between the two versions of block-based CoSaMP discussed in Section 3.2—specifically, BBCoSaMP( $\mathbf{A}, \Psi, \mathbf{y}, K$ ) and BBCoSaMP( $\mathbf{A}\Psi, \mathbf{I}, \mathbf{y}, K$ ). Recall that the key difference between these approaches is that the former essentially tries to recover  $\mathbf{x}$  directly, while the latter instead attempts to first recover  $\boldsymbol{\alpha}$  and then estimates  $\hat{\mathbf{x}}$  as  $\Psi\hat{\boldsymbol{\alpha}}$ . Since less is known about the theoretical properties of BBCoSaMP( $\mathbf{A}, \Psi, \mathbf{y}, K$ ) (both in and of itself and with respect to the computation of  $\mathcal{P}(\mathbf{x}, K)$ ), we have focused more on BBCoSaMP( $\mathbf{A}\Psi, \mathbf{I}, \mathbf{y}, K$ ) in our theoretical development. However, all of our theorems regarding BBCoSaMP( $\mathbf{A}\Psi, \mathbf{I}, \mathbf{y}, K$ ) (e.g., Theorems 5.4 and 5.6) are limited to the case where the number  $k$  of DPSS vectors per band used to construct the multiband modulated DPSS dictionary  $\Psi$  satisfies  $k < 2NW$ , and as we will see shortly, in practice we obtain significant improvements in performance by considering  $k > 2NW$ .

In our experiments we evaluate both versions of block-based CoSaMP as a function of  $k$ . For the purposes of this experiment, we assume that there are  $K = 5$  active bands, we set the number of measurements  $M = 512$ , and we construct the measurement matrix  $\mathbf{A}$  to be  $M \times N$  with i.i.d. Gaussian entries with mean zero and variance  $\frac{1}{M}$ . In this experiment, we take  $\mathbf{y} = \mathbf{A}\mathbf{x}$  and do not add noise to the measurements. In Figure 5(a) we see that when  $k$  is small the two approaches perform similarly, but when  $k$  exceeds  $2NW$  by more than a small amount, the performance of BBCoSaMP( $\mathbf{A}, \Psi, \mathbf{y}, K$ ) is far superior to that of BBCoSaMP( $\mathbf{A}\Psi, \mathbf{I}, \mathbf{y}, K$ ). In fact, when  $k$  becomes large we observe that for at least 5% of trials BBCoSaMP( $\mathbf{A}\Psi, \mathbf{I}, \mathbf{y}, K$ ) results in an SNR of almost 0dB (which can be achieved simply via the trivial estimate  $\hat{\mathbf{x}} = \mathbf{0}$ .) This gap is further illustrated in Figure 5(b), which shows the probability that the performance of BBCoSaMP( $\mathbf{A}\Psi, \mathbf{I}, \mathbf{y}, K$ ) is within 3dB of BBCoSaMP( $\mathbf{A}, \Psi, \mathbf{y}, K$ ) as a function of  $k$ . This also begins to rapidly decay when  $k$  exceeds  $2NW$ . For a point of reference, in Figure 5(c) we plot the eigenvalue  $\lambda_{N,W}^{(k+1)}$  corresponding to the first DPSS basis vector which is not used in constructing our dictionary  $\Psi$ . This value is the dominant term in the approximation error that we can expect when using  $\Psi$  to represent multiband signals. We see that BBCoSaMP( $\mathbf{A}, \Psi, \mathbf{y}, K$ ) continues to improve with increasing  $k$ , roughly until  $\lambda_{N,W}^{(k+1)}$  approaches the level of machine precision.

Our experimental results suggest that in our analysis of BBCoSaMP( $\mathbf{A}\Psi, \mathbf{I}, \mathbf{y}, K$ ) we are correct in requiring that  $k$  be relatively small, as the algorithm does indeed break down when  $k$  becomes too large. However, before this breakdown the performance can be quite favorable, with recovery SNR exceeding 88dB, and Theorem 5.6 does guarantee even better performance were  $N$  to increase.

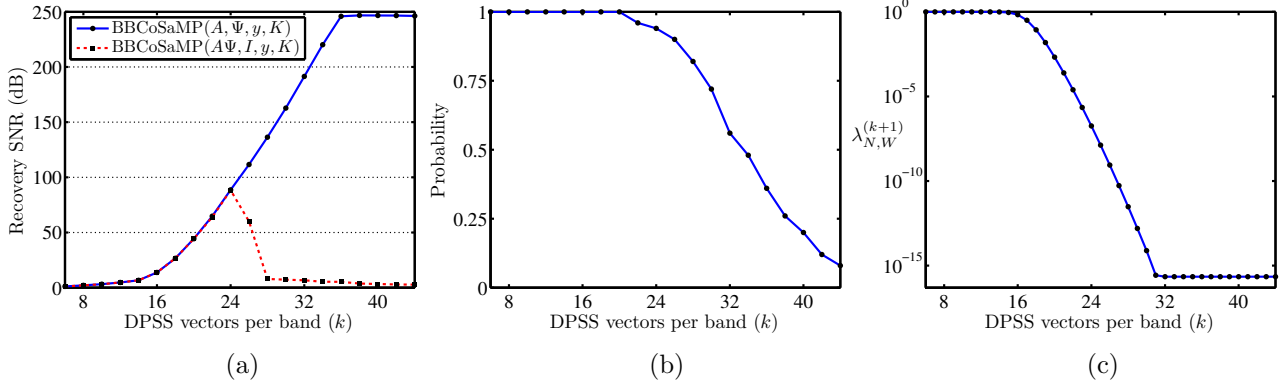


Figure 5: Comparison of  $\text{BBCoSaMP}(\mathbf{A}, \Psi, \mathbf{y}, K)$  to  $\text{BBCoSaMP}(\mathbf{A}\Psi, \mathbf{I}, \mathbf{y}, K)$ . (a) The recovery SNR of  $\text{BBCoSaMP}(\mathbf{A}, \Psi, \mathbf{y}, K)$  and  $\text{BBCoSaMP}(\mathbf{A}\Psi, \mathbf{I}, \mathbf{y}, K)$  as a function of the number  $k$  of DPSS vectors used per band in constructing the multiband modulated DPSS dictionary  $\Psi$ . (b) The probability that the performance of  $\text{BBCoSaMP}(\mathbf{A}\Psi, \mathbf{I}, \mathbf{y}, K)$  is within 3dB of  $\text{BBCoSaMP}(\mathbf{A}, \Psi, \mathbf{y}, K)$  as a function of  $k$ . (c) The eigenvalue corresponding to the first DPSS basis vector that is not used.

We speculate that the breakdown in the performance of  $\text{BBCoSaMP}(\mathbf{A}\Psi, \mathbf{I}, \mathbf{y}, K)$  as  $k$  grows is likely due to the fact that, for large enough  $k$ , the dictionary  $\Psi$  begins to contain highly coherent columns, so that any method that attempts to recover  $\alpha$  itself is likely to encounter significant problems. However, the strong performance of  $\text{BBCoSaMP}(\mathbf{A}, \Psi, \mathbf{y}, K)$  (with no clear limitation on the size of  $k$ ) seems to suggest that this latter approach likely satisfies the kinds of guarantees provided for block-based IHT in Theorems 5.3 and 5.5, which provide for arbitrarily large  $k$  and hence arbitrarily accurate recovery. In light of these results, for the remainder of our experiments we focus exclusively on  $\text{BBCoSaMP}(\mathbf{A}, \Psi, \mathbf{y}, K)$ .

### 6.3 Impact of measurement noise

In most realistic scenarios, the compressive measurements  $\mathbf{y}$  will be subject to various sources of noise (including noise in the signal itself, noise within the sensing hardware, quantization, etc.) As noted in Section 5.2.3, our approach to signal recovery inherits the same robustness to noise that is exhibited by traditional CoSaMP. To illustrate this, we consider the case where the measurements  $\mathbf{y}$  are corrupted by additive white Gaussian noise,<sup>12</sup> i.e.,  $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}$  where each  $e \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$ . In our experiments we consider three different noise levels, quantified by the “measurement SNR” (MSNR), which is defined as

$$\text{MSNR} = 20 \log_{10} \left( \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{e}\|_2} \right) \text{ dB},$$

In general, the theoretical guarantees for this scenario suggest that the SNR of the recovery should be roughly comparable to the MSNR.

The results of our experiment are illustrated in Figure 6. In this case we are still assuming that there are  $K = 5$  active bands, we use an i.i.d. Gaussian matrix  $\mathbf{A}$  with  $M = 512$  rows, and we plot the performance as a function of  $k$ . We observe that the results are essentially the same as in Figure 5(a), but that as we increase  $k$  the recovery SNR will hit a plateau dictated by the best

<sup>12</sup>Note that the case where the signal is corrupted with white noise as opposed to the measurements can be reduced to the case where the signal is noise-free and the measurements are noisy, and thus we restrict our attention to noisy measurements. See [11, 21] for further discussion of this equivalence and the challenges posed by signal noise.

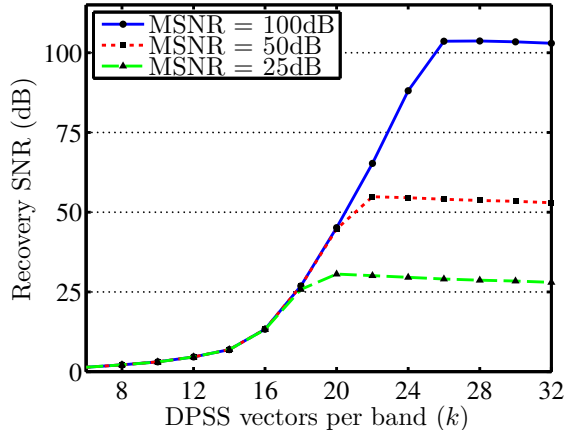


Figure 6: *Impact of measurement noise on the signal recovery SNR as a function of  $k$ .*

possible SNR achievable for a given MSNR. Roughly speaking, when  $k$  is small, performance is being limited by “modeling error”, but as  $k$  increases we eventually reach a regime where measurement noise surpasses modeling error as the limiting factor, and no further gains are possible by increasing  $k$ . Thus, in practice it will typically be the noise level that dictates the optimal choice in  $k$ . For the remainder of our experiments we wish to avoid any assumptions about the noise level, and so we restrict our attention to noise-free measurements. However, the results all translate to the noisy setting roughly as one would expect based on these results.

#### 6.4 Required measurement rate

We now study the performance our approach as a function of the number of measurements  $M$ . Specifically, we consider the cases where  $K = 5, 10$ , and  $15$  bands are active, and for each value of  $K$  we let  $M$  vary from  $2NWK$  (the Landau rate) up to  $14NWK$  (oversampling the Landau rate by a factor of 7). The results are shown in Figures 7(a) and (b), which plot the results in terms of  $\frac{M}{2NWK}$  and  $M$ , respectively. We observe that when the measurement rate is only 3 or 4 times that of the Landau rate, we are already doing extremely well, and we obtain near-perfect recovery at 6 times the Landau rate. We observe very similar behavior for all values of  $K$ . Thus, for the sake of simplicity, we restrict the remainder of our experiments to the case where there are only  $K = 5$  active bands.

Before moving on, we note that an important caveat to these results is that as we vary  $M$  and  $K$ , it is sometimes necessary to adjust the value of  $k$  used in the construction of  $\Psi$ . This can easily be seen by considering the regime where the measurement rate is very close to the Landau rate. Suppose  $M = 2NWK$  and that we knew *a priori* which bands were occupied. In this case we could perform recovery using the appropriate submatrix of  $\Psi$ , which would have  $kK$  columns. Thus, if  $k > 2NW$ , then we would have only  $M = 2NWK$  measurements and would need to estimate  $kK > M$  values—a situation we would clearly like to avoid. Moreover, if we must also estimate the support from the data as well, then it is clearly best to avoid setting  $k$  to be too large when  $M$  is small. Thus, in the experiments for Figure 7 (and for those below) we set  $k$  based on a rule of thumb that we determined based on empirical performance. Specifically, we set  $k = 2NW$  when  $\frac{M}{2NWK} \leq 2$ . For  $\frac{M}{2NWK} \in [2, 6]$ , we linearly increase  $k$  from  $2NW = 16$  to, in our case  $k = 38$  (which represents the rough point at which the first omitted eigenvalue  $\lambda_{N,W}^{(k+1)}$  reaches the level of machine precision). For larger values of  $M$  there is no performance gain by considering  $k$  larger

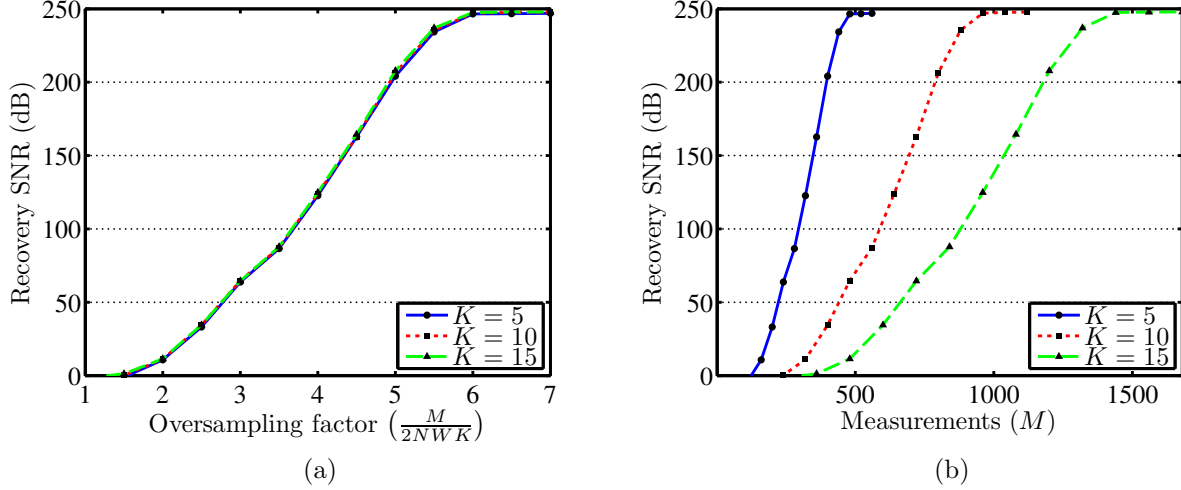


Figure 7: Recovery performance as a function of the number of measurements  $M$  for  $K = 5, 10$ , and  $15$  active bands. (a) Performance in terms of  $\frac{M}{2NWK}$ , which measures oversampling with respect to the Landau rate. (b) Performance in terms of the actual number of measurements  $M$ .

than this level.

## 6.5 Alternative measurement architectures

As promised in Section 3.1, we now evaluate the performance of our approach using more practical measurement schemes. We compare the performance achieved using an i.i.d. Gaussian matrix to that achieved using an  $M \times N$  random demodulator matrix [72] and to a random sampling approach where  $M$  samples are taken uniformly at random on the length- $N$  Nyquist grid. The results are shown in Figure 8. We observe very similar performance among the three approaches, with the Gaussian matrix performing best, then random demodulator, followed by random sampling. Note that while there is certainly a gap between these three approaches, it is also most pronounced in a regime which is likely irrelevant in practice, since it is rare to find applications where a recovery SNR of 200dB or more is feasible.

## 6.6 Comparison with DFT-based approach

Finally, we close with a comparison between the performance achievable using our proposed multi-band modulated DPSS dictionary  $\Psi$  and the performance achievable using the  $N \times N$  DFT as a sparsifying basis. We test both dictionaries using an  $M \times N$  random demodulator measurement matrix [72], which was originally designed with frequency-sparse signals in mind (but under a multitone signal model, not a multiband model). Due to DFT leakage effects, we believe that it would be inappropriate to break the DFT dictionary into bands and use a block block-based recovery algorithm; thus when  $\Psi$  is the DFT basis, we use  $\hat{x} = \text{CoSaMP}(\mathbf{A}, \Psi, \mathbf{y}, S)$  as our recovered signal estimate. In order to give the DFT approach the best possible chance for success, for each value of  $M$  we consider a range of possible values for the sparsity parameter  $S$ , selecting the value of  $S$  that achieves the best possible performance according to an oracle.

The performance gap between these two approaches—illustrated in Figure 9(a)—is monumental. Using the DFT dictionary, we never achieve a recovery SNR better than 20dB over this range of measurements, whereas using the multiband modulated DPSS dictionary with block-based CoSaMP, the recovery SNR can exceed 200dB.

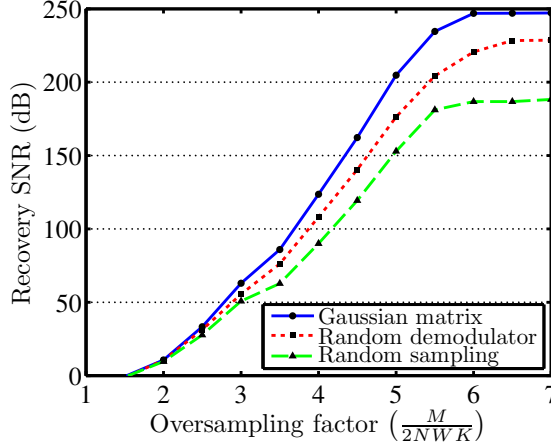


Figure 8: Comparison of recovery SNR using i.i.d. Gaussian measurements, random demodulator measurements, and random samples.

To illustrate the typical difference between these two approaches, in Figure 9(b) we show in blue the DTFT of a representative signal from one trial in this experiment. Plotted in red is the DTFT of the signal vector recovered using the DPSS dictionary with  $\frac{M}{2NWK} = 4$ ; the recovery SNR is 109dB and the recovered signal is visually indistinguishable from the original. Plotted in green are the DFT coefficients of the signal vector recovered using the DFT dictionary with the same measurements; the recovery SNR is now only 13.4dB. While the DFT-based estimate does successfully capture the main peaks of each band, it naturally misses all of the sidelobes of each band (and despite the multiband nature of  $x(t)$ , these sidelobes are important for accurately representing the window  $\mathbf{x}$ ). Moreover, due to modeling error, the DFT-based approach also results in a number of spurious artifacts in regions where there is no significant frequency content in the original signal.

We note that for the DFT-based approach in this experiment, the best estimate produced using CoSaMP( $\mathbf{A}, \Psi, \mathbf{y}, S$ ) was observed when setting  $S = 85$ . On the other hand, the DPSS approach (according to our rule of thumb) selects  $k = 27$ , so that in this approach there are a total of  $kK = 135$  free parameters. It is not surprising that the DPSS approach results in superior performance since the model has so many extra dimensions. What *is* potentially surprising is that if we set  $S = 135$ , the DFT approach exhibits a complete breakdown and is unable to make any practical use of these extra degrees of freedom.

## 7 Conclusions

There are likely to be many ways that one could bridge the gap between the discrete, finite framework of CS and the problem of acquiring a continuous-time signal. In this paper, using a dictionary constructed from a sequence of modulated DPSS basis elements, we have argued that when dealing with finite-length windows of Nyquist-rate samples, it is possible to map the multiband analog signal model—in a very natural way—into a finite-dimensional sparse model. This allows us to apply many of the standard theoretical and algorithmic tools from CS to the problem of recovering such a sample vector from compressive measurements. Moreover, the sparse signals that we encounter in this model actually have a structured sparsity pattern (namely, block-sparsity), which we can exploit through the use of specialized “model-based” CS recovery algorithms. Although our recovery bounds are qualified—we have showed that *most* finite-length sample vectors arising

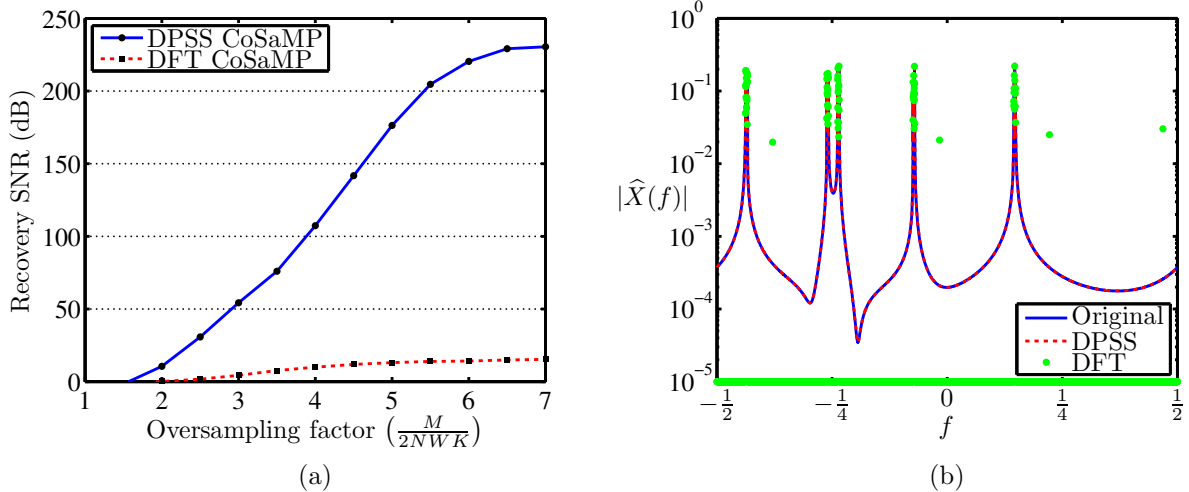


Figure 9: Comparison of recovery performance between block-based CoSaMP with the multiband modulated DPSS dictionary and standard CoSaMP using the DFT basis. (a) Performance comparison as a function of  $\frac{M}{2NWK}$ . (b) The DTFT of a representative signal, and the recovered estimate using each approach.

from multiband analog signals can be *highly accurately* recovered from a number of measurements that is proportional to the Landau rate—these qualifiers are ultimately necessary. Moreover, we have demonstrated through a series of careful experiments that the multiband modulated DPSS dictionary can provide a far superior alternative to the DFT for sparse recovery. Our experiments also confirm that certain practical measurement architectures such as the random demodulator, which was previously viewed only as a mechanism for capturing multitone signals [43], can indeed be used to efficiently capture multiband signals.

As the reader will note from our discussions regarding the selection of the number of columns per band  $k$ , the computation of  $\mathcal{P}(\mathbf{x}, K)$ , etc., there are certain subtle challenges in choosing the proper dictionary design and implementing an effective recovery algorithm. However, our experimental results do confirm that it is possible to navigate these waters. We have also aimed to give some practical insight into the proper choice of design parameters, and we have made all of the software for our simulations available for download at <http://www.mines.edu/~mwakin/software/>. Ultimately, in addition to the open algorithmic questions discussed in Section 6, there are many open questions concerning the most effective way to construct a multiband DPSS dictionary; one could imagine possible advantages to considering multiscale dictionaries (see also [58]), adaptive band sizes, overlapping bands, etc. We leave the consideration of these questions to future work.

It is worth emphasizing that the remarkable efficiency of our DPSS dictionary for sparsely approximating finite-length multiband sample vectors is owed entirely to the eigenvalue concentration behavior described in Section 4.2.4. It would be interesting to explore other possible problem domains (beyond multiband signal models) where similar eigenvalue concentration behavior holds. Again, we leave the consideration of such questions to future work.

We conclude by noting that applications of sparse representations abound, and so there are many possible settings outside of CS where the multiband modulated DPSS dictionary could be useful for processing finite-length sample vectors arising from multiband signals. For example, there are several possible applications in *compressive signal processing* [20], where one can attempt to answer certain questions about a signal vector directly from compressive measurements of that vector without having to actually perform signal recovery. One specific problem in this area could involve cancellation of an interfering signal that has corrupted the measurements. Suppose  $s(t)$  is

an analog signal of interest (not necessarily obeying a multiband model), and  $i(t)$  is a multiband (or even just single-band) interferer supported on bands indexed by the set  $\mathcal{I}$ . Let  $\mathbf{s}$  and  $\mathbf{i}$  denote the sample vectors arising from these two signals, and suppose we collect compressive measurements  $\mathbf{y} = \mathbf{A}(\mathbf{s} + \mathbf{i})$ . Then because  $\mathbf{i}$  is concentrated in  $\mathcal{R}(\Psi_{\mathcal{I}})$ , it follows that  $\mathbf{A}\mathbf{i}$  is concentrated in  $\mathcal{R}(\tilde{\mathbf{A}}_{\mathcal{I}})$ , where  $\tilde{\mathbf{A}} = \mathbf{A}\Psi$ . One can therefore define an orthogonal projection matrix  $\mathbf{P} := \mathbf{I} - \tilde{\mathbf{A}}_{\mathcal{I}}\tilde{\mathbf{A}}_{\mathcal{I}}^{\dagger}$ . By applying  $\mathbf{P}$  to the measurement vector  $\mathbf{y}$ , one can remove virtually all of the influence of the interferer, since  $\mathbf{P}\mathbf{A}\mathbf{i} \approx \mathbf{0}$  even for very strong interferers, and therefore  $\mathbf{P}\mathbf{y} = \mathbf{P}\mathbf{A}\mathbf{s} + \mathbf{P}\mathbf{A}\mathbf{i} \approx \mathbf{P}\mathbf{A}\mathbf{s}$ . From these processed measurements  $\mathbf{P}\mathbf{y}$  one can attempt to recover  $\mathbf{s}$  or answer various compressive-domain questions that do not require signal recovery. It can even be possible to derive RIP bounds for  $\mathbf{P}\mathbf{A}$  and thus guarantee the performance of these techniques. We refer the reader to [19, 22] for additional discussion of the interference cancellation problem and an example using a preliminary version of the modulated DPSS dictionary.

## Acknowledgements

We are very grateful to Richard Baraniuk, Emmanuel Candès, and Justin Romberg for helpful discussions during the formative stages of this work and to Marco Duarte, Armin Eftekhar, Mark Embree, Chinmay Hegde, Deanna Needell, Yaniv Plan, and Borhan Sanandaji for their input on certain technical points.

## A Review of the Karhunen-Loeve (KL) Transform

We briefly review the basic ideas behind the KL transform [67]. Suppose that  $\mathbf{r} \in \mathbb{C}^N$  is a random vector with mean zero and a known autocorrelation matrix denoted by  $\mathbf{R}$  with entries defined as

$$\mathbf{R}[m, n] = \mathbb{E} \left[ \overline{\mathbf{r}[n]} \mathbf{r}[m] \right],$$

where the overline denotes complex conjugation. Next suppose that we would like to find, for some fixed  $k \in \{1, 2, \dots, N\}$ , a  $k$ -dimensional subspace  $\mathbf{Q}$  of  $\mathbb{C}^N$  that best captures the energy of  $\mathbf{r}$  in expectation. That is, we wish to find the  $\mathbf{Q}$  that minimizes

$$\mathbb{E} [\|\mathbf{r} - \mathbf{P}_{\mathbf{Q}}\mathbf{r}\|_2^2]$$

over all  $k$ -dimensional subspaces of  $\mathbb{C}^N$ . The optimal solution to this minimization problem can be found by computing an eigendecomposition of  $\mathbf{R}$ . In particular, let

$$\mathbf{R} = \mathbf{U}\mathbf{\Lambda}_R\mathbf{U}^H$$

denote the eigendecomposition of  $\mathbf{R}$ , with the eigenvalues  $\lambda_0(\mathbf{R}) \geq \lambda_1(\mathbf{R}) \geq \dots \geq \lambda_{N-1}(\mathbf{R}) \geq 0$  along the main diagonal of  $\mathbf{\Lambda}_R$ . Let  $\mathbf{U}_k$  denote the  $N \times k$  matrix resulting by taking the first  $k$  columns of  $\mathbf{U}$ . The columns of  $\mathbf{U}_k$  provide an orthonormal basis for the optimal  $\mathbf{Q}$ , and thus we will have  $\mathbf{P}_{\mathbf{Q}} = \mathbf{U}_k\mathbf{U}_k^H$ . For this optimal choice of  $\mathbf{Q}$ , we will also have

$$\mathbb{E} [\|\mathbf{r} - \mathbf{P}_{\mathbf{Q}}\mathbf{r}\|_2^2] = \sum_{\ell=k}^{N-1} \lambda_{\ell}(\mathbf{R}).$$



## B Proof of Theorem 4.1

*Proof.* Let  $f'$  denote a random variable with uniform distribution on  $[f_c - W, f_c + W]$ , let  $\theta'$  denote a random variable with uniform distribution on  $[0, 2\pi)$ , and let  $\mathbf{r} = \mathbf{r}(f', \theta') := e^{j\theta'} \mathbf{e}_{f'}$  denote a random vector in  $\mathbb{C}^N$  defined in terms of  $f'$  and  $\theta'$ . Then we can write

$$\begin{aligned} \frac{1}{2W} \cdot \int_{f_c-W}^{f_c+W} \|\mathbf{e}_f - \mathbf{P}_Q \mathbf{e}_f\|_2^2 df &= \int_{f_c-W}^{f_c+W} \|\mathbf{e}_{f'} - \mathbf{P}_Q \mathbf{e}_{f'}\|_2^2 p_{f'}(f') df' \\ &= \int_0^{2\pi} \int_{f_c-W}^{f_c+W} \|e^{j\theta'} (\mathbf{e}_{f'} - \mathbf{P}_Q \mathbf{e}_{f'})\|_2^2 p_{f'}(f') \frac{1}{2\pi} df' d\theta' \\ &= \int_0^{2\pi} \int_{f_c-W}^{f_c+W} \|e^{j\theta'} \mathbf{e}_{f'} - \mathbf{P}_Q (e^{j\theta'} \mathbf{e}_{f'})\|_2^2 p_{f'}(f') p_{\theta'}(\theta') df' d\theta' \\ &= \mathbb{E} [\|\mathbf{r} - \mathbf{P}_Q \mathbf{r}\|_2^2]. \end{aligned}$$

We next verify that the random vector  $\mathbf{r}$  has mean zero, i.e., for each  $n \in \{0, 1, \dots, N-1\}$  we have

$$\begin{aligned} \mathbb{E} [\mathbf{r}[n]] &= \int_0^{2\pi} \int_{f_c-W}^{f_c+W} e^{j\theta'} \mathbf{e}_{f'}[n] p_{f'}(f') p_{\theta'}(\theta') df' d\theta' \\ &= \int_0^{2\pi} \int_{f_c-W}^{f_c+W} e^{j\theta'} e^{j2\pi f' n} \frac{1}{4W\pi} df' d\theta' \\ &= \frac{1}{4W\pi} \cdot \int_{f_c-W}^{f_c+W} e^{j2\pi f' n} \int_0^{2\pi} e^{j\theta'} d\theta' df' \\ &= 0. \end{aligned}$$

Next, observe that the random vector  $\mathbf{r}$  has autocorrelation matrix  $\mathbf{R}$  with entries:

$$\begin{aligned} \mathbf{R}[m, n] &= \mathbb{E} [\overline{\mathbf{r}[n]} \mathbf{r}[m]] \\ &= \mathbb{E} [\overline{\mathbf{e}_{f'}[n]} e^{-j\theta'} e^{j\theta'} \mathbf{e}_{f'}[m]] \\ &= \frac{1}{2W} \cdot \int_{f_c-W}^{f_c+W} e^{-j2\pi f' n} e^{j2\pi f' m} df' \\ &= \frac{1}{2W} \cdot e^{j2\pi f_c m} \cdot 2W \operatorname{sinc}(2W(m-n)) \cdot e^{-j2\pi f_c n}. \end{aligned}$$

Therefore, we can write

$$\mathbf{R} = \frac{1}{2W} \cdot \mathbf{E}_{f_c} \mathbf{B}_{N,W} \mathbf{E}_{f_c}^H, \quad (66)$$

where  $\mathbf{E}_{f_c}$  is as defined in (44) and  $\mathbf{B}_{N,W}$  is as defined in (36). Plugging the eigendecomposition for  $\mathbf{B}_{N,W}$  from (38) into (66), we obtain

$$\mathbf{R} = \frac{1}{2W} \cdot \mathbf{E}_{f_c} \mathbf{S}_{N,W} \mathbf{\Lambda}_{N,W} \mathbf{S}_{N,W}^H \mathbf{E}_{f_c}^H.$$

From this eigendecomposition and standard results concerning the KL transform (see Appendix A), it follows that the optimal  $k$ -dimensional subspace  $\mathbf{Q}$  will be spanned by the first  $k$  columns of  $\mathbf{E}_{f_c} \mathbf{S}_{N,W}$ . Furthermore, this choice of  $\mathbf{Q}$  yields

$$\frac{1}{2W} \cdot \int_{f_c-W}^{f_c+W} \|\mathbf{e}_f - \mathbf{P}_Q \mathbf{e}_f\|_2^2 df = \mathbb{E} [\|\mathbf{r} - \mathbf{P}_Q \mathbf{r}\|_2^2] = \sum_{\ell=k}^{N-1} \lambda_{\ell}(\mathbf{R}) = \frac{1}{2W} \sum_{\ell=k}^{N-1} \lambda_{N,W}^{(\ell)},$$

as desired.  $\square$

## C Proof of Theorem 4.2

*Proof.* The sequence  $x[n] := x(nT_s), n \in \mathbb{Z}$  is a discrete-time, zero-mean, wide sense stationary random process with power spectrum equal to

$$P_x(f) = \begin{cases} \frac{1}{2W}, & f_c - W \leq f \leq f_c + W, \\ 0, & \text{otherwise,} \end{cases}$$

and the vector  $\mathbf{x} \in \mathbb{C}^N$  equals the restriction of  $x[n]$  to the indices  $n = 0, 1, \dots, N-1$ . The inverse DTFT of the power spectrum  $P_x(f)$  gives the autocorrelation function for  $x[n]$ :

$$r_x[n] = e^{j2\pi n f_c} \cdot \text{sinc}(2Wn).$$

It follows that the random vector  $\mathbf{x}$  has mean zero and an autocorrelation matrix  $\mathbf{R}$  given by (66). Thus,  $\mathbf{x}$  has exactly the same autocorrelation structure as the random vector  $\mathbf{r}$  we considered in Appendix B, and so the same KL transform analysis applies. Finally, we can also compute

$$\mathbb{E} [\|\mathbf{x}\|_2^2] = \sum_{n=0}^{N-1} \mathbb{E} [|\mathbf{x}[n]|^2] = N r_x[0] = N.$$

□

## D Proof of Theorem 4.3

*Proof.* First, let us define  $\mathcal{E}_{f_c}$  to be an operator that modulates a discrete-time signal by  $f_c$ ; more specifically,

$$\mathcal{E}_{f_c}(s)[n] = e^{j2\pi f_c n} s[n]$$

for all  $n \in \mathbb{Z}$ . Since the modulated DPSS vectors form an orthonormal basis for  $\mathbb{C}^N$ , and since the discrete-time signal  $x[n]$  is time-limited, we can write  $x[n]$  as

$$x = \sum_{\ell=0}^{N-1} \alpha[\ell] \mathcal{E}_{f_c} \mathcal{T}_N(s_{N,W}^{(\ell)}),$$

where the coefficients  $\alpha$  are given by  $\alpha[\ell] = \langle x, \mathcal{E}_{f_c} \mathcal{T}_N(s_{N,W}^{(\ell)}) \rangle = \langle \mathbf{x}, \mathbf{E}_{f_c} \mathbf{s}_{N,W}^{(\ell)} \rangle$  for  $\ell = 0, 1, \dots, N-1$  and satisfy  $\|\alpha\|_2^2 = \|x\|_2^2 = \|\mathbf{x}\|_2^2$ . By linearity, we can also write

$$\mathcal{B}_{f_c, W} x = \sum_{\ell=0}^{N-1} \alpha[\ell] \mathcal{B}_{f_c, W} \mathcal{E}_{f_c} \mathcal{T}_N(s_{N,W}^{(\ell)}),$$

where the functions  $\mathcal{B}_{f_c, W} \mathcal{E}_{f_c} \mathcal{T}_N(s_{N,W}^{(\ell)})$  are modulated DPSS's and therefore remain orthogonal. Therefore,

$$(1 - \delta) \|x\|_2^2 \leq \|\mathcal{B}_{f_c, W} x\|_2^2 = \sum_{\ell=0}^{N-1} |\alpha[\ell]|^2 \|\mathcal{B}_{f_c, W} \mathcal{E}_{f_c} \mathcal{T}_N(s_{N,W}^{(\ell)})\|_2^2 = \sum_{\ell=0}^{N-1} |\alpha[\ell]|^2 \lambda_{N,W}^{(\ell)}.$$

Now note that if  $k = 2NW(1 + \epsilon)$ , then

$$\begin{aligned}
(1 - \delta)\|\mathbf{x}\|_2^2 &\leq \sum_{\ell=0}^{k-1} |\boldsymbol{\alpha}[\ell]|^2 \lambda_{N,W}^{(\ell)} + \sum_{\ell=k}^{N-1} |\boldsymbol{\alpha}[\ell]|^2 \lambda_{N,W}^{(\ell)} \\
&\leq \sum_{\ell=0}^{k-1} |\boldsymbol{\alpha}[\ell]|^2 + \sum_{\ell=k}^{N-1} \|\mathbf{x}\|_2^2 \lambda_{N,W}^{(\ell)} \\
&\leq \|P_Q \mathbf{x}\|_2^2 + \|\mathbf{x}\|_2^2 N C_3 e^{-C_4 N} \\
&= (\|\mathbf{x}\|_2^2 - \|\mathbf{x} - P_Q \mathbf{x}\|_2^2) + \|\mathbf{x}\|_2^2 N C_3 e^{-C_4 N}
\end{aligned}$$

for  $N \geq N_1$ . Rearranging terms in the above inequality yields (47).  $\square$

## E Proof of Theorem 4.4

*Proof.* For each  $i \in \mathcal{I}$ , the sequence  $x_i[n] := x_i(nT_s), n \in \mathbb{Z}$  is a discrete-time, zero-mean, wide sense stationary random process with power spectrum equal to

$$P_{x_i}(f) = \begin{cases} \frac{1}{2WK}, & f \in \left[-\frac{B_{\text{nyq}}T_s}{2} + i2W, -\frac{B_{\text{nyq}}T_s}{2} + (i+1)2W\right], \\ 0, & \text{otherwise.} \end{cases}$$

These discrete-time sequences will also be independent and jointly wide sense stationary. Defining  $\mathbf{x}_i \in \mathbb{C}^N$  to be the restriction of  $x_i[n]$  to the indices  $n = 0, 1, \dots, N-1$ , we have  $\mathbf{x} = \sum_{i \in \mathcal{I}} \mathbf{x}_i$  and  $P_x(f) = \sum_{i \in \mathcal{I}} P_{x_i}(f)$ . We can compute

$$\mathbb{E} [\|\mathbf{x}\|_2^2] = \sum_{n=0}^{N-1} \mathbb{E} [|\mathbf{x}[n]|^2] = \sum_{n=0}^{N-1} \sum_{i, \ell \in \mathcal{I}} \mathbb{E} [\mathbf{x}_i[n] \overline{\mathbf{x}_\ell[n]}] = \sum_{n=0}^{N-1} \sum_{i \in \mathcal{I}} \mathbb{E} [|\mathbf{x}_i[n]|^2] = N \cdot K \cdot \frac{1}{K} = N.$$

Also, we have

$$\begin{aligned}
\mathbb{E} [\|\mathbf{x} - P_{\Psi_{\mathcal{I}}} \mathbf{x}\|_2^2] &= \mathbb{E} \left[ \left\| \left( \sum_{i \in \mathcal{I}} \mathbf{x}_i \right) - P_{\Psi_{\mathcal{I}}} \left( \sum_{i \in \mathcal{I}} \mathbf{x}_i \right) \right\|_2^2 \right] \\
&= \mathbb{E} \left[ \left\| \sum_{i \in \mathcal{I}} (\mathbf{x}_i - P_{\Psi_{\mathcal{I}}} \mathbf{x}_i) \right\|_2^2 \right] \\
&\leq \mathbb{E} \left[ \left( \sum_{i \in \mathcal{I}} \|\mathbf{x}_i - P_{\Psi_{\mathcal{I}}} \mathbf{x}_i\|_2 \right)^2 \right] \\
&\leq \mathbb{E} \left[ \left( \sum_{i \in \mathcal{I}} \|\mathbf{x}_i - P_{\Psi_i} \mathbf{x}_i\|_2 \right)^2 \right] \\
&\leq \mathbb{E} \left[ K \sum_{i \in \mathcal{I}} \|\mathbf{x}_i - P_{\Psi_i} \mathbf{x}_i\|_2^2 \right] \\
&= K \sum_{i \in \mathcal{I}} \mathbb{E} [\|\mathbf{x}_i - P_{\Psi_i} \mathbf{x}_i\|_2^2] \\
&= K \cdot K \cdot \frac{1}{2WK} \sum_{\ell=k}^{N-1} \lambda_{N,W}^{(\ell)},
\end{aligned}$$

where the third line follows from the triangle inequality, the fourth line follows because the columns of each  $\Psi_i$  belong to  $\Psi_{\mathcal{I}}$  as well, the sixth line follows because  $\|\mathbf{h}\|_1^2 \leq K\|\mathbf{h}\|_2^2$  for any  $K$ -dimensional vector  $\mathbf{h}$ , and the final line follows from Theorem 4.2.  $\square$

## F Proof of Lemma 5.1

*Proof.* Let  $\Psi_{i_1}$  and  $\Psi_{i_2}$  denote the blocks from which  $\mathbf{q}_1$  and  $\mathbf{q}_2$ , respectively, are drawn. If  $i_1 = i_2$ , we have already established at the beginning of Section 4.3.1 that  $\langle \mathbf{q}_1, \mathbf{q}_2 \rangle = 0$ . Thus, we assume henceforth that  $i_1 \neq i_2$ .

Let  $\mathcal{E}_{f_i}$  denote an operator that modulates a discrete-time signal by  $f_i$ ; more specifically,  $\mathcal{E}_{f_i}(s)[n] = e^{j2\pi f_i n} s[n]$  for all  $n \in \mathbb{Z}$ . Let  $\mathcal{B}_{f_i, W}$  denote an orthogonal projection operator that takes a discrete-time signal, bandlimits its DTFT to the frequency range  $f \in [f_i - W, f_i + W]$ , and returns the corresponding signal in the time domain. Finally, define  $\mathcal{B}_{f_{i_1}, f_{i_2}, W}^c := I - \mathcal{B}_{f_{i_1}, W} - \mathcal{B}_{f_{i_2}, W}$ , which is also an orthogonal projection for finite-energy signals because  $|f_{i_1} - f_{i_2}| \geq 2W$ . For some  $\ell, \ell' \in \{0, 1, \dots, 2NW(1 - \epsilon) - 1\}$ , we can write

$$\begin{aligned} \langle \mathbf{q}_1, \mathbf{q}_2 \rangle &= \left\langle \mathbf{E}_{f_{i_1}} s_{N, W}^{(\ell)}, \mathbf{E}_{f_{i_2}} s_{N, W}^{(\ell')} \right\rangle \\ &= \left\langle \mathcal{E}_{f_{i_1}} \mathcal{T}_N s_{N, W}^{(\ell)}, \mathcal{E}_{f_{i_2}} \mathcal{T}_N s_{N, W}^{(\ell')} \right\rangle \\ &= \left\langle \mathcal{B}_{f_{i_1}, W} \mathcal{E}_{f_{i_1}} \mathcal{T}_N s_{N, W}^{(\ell)} + \mathcal{B}_{f_{i_2}, W} \mathcal{E}_{f_{i_1}} \mathcal{T}_N s_{N, W}^{(\ell)} + \mathcal{B}_{f_{i_1}, f_{i_2}, W}^c \mathcal{E}_{f_{i_1}} \mathcal{T}_N s_{N, W}^{(\ell)}, \right. \\ &\quad \left. \mathcal{B}_{f_{i_1}, W} \mathcal{E}_{f_{i_2}} \mathcal{T}_N s_{N, W}^{(\ell')} + \mathcal{B}_{f_{i_2}, W} \mathcal{E}_{f_{i_2}} \mathcal{T}_N s_{N, W}^{(\ell')} + \mathcal{B}_{f_{i_1}, f_{i_2}, W}^c \mathcal{E}_{f_{i_2}} \mathcal{T}_N s_{N, W}^{(\ell')} \right\rangle \\ &= \left\langle \mathcal{B}_{f_{i_1}, W} \mathcal{E}_{f_{i_1}} \mathcal{T}_N s_{N, W}^{(\ell)}, \mathcal{B}_{f_{i_1}, W} \mathcal{E}_{f_{i_2}} \mathcal{T}_N s_{N, W}^{(\ell')} \right\rangle + \left\langle \mathcal{B}_{f_{i_2}, W} \mathcal{E}_{f_{i_1}} \mathcal{T}_N s_{N, W}^{(\ell)}, \right. \\ &\quad \left. \mathcal{B}_{f_{i_2}, W} \mathcal{E}_{f_{i_2}} \mathcal{T}_N s_{N, W}^{(\ell')} \right\rangle + \left\langle \mathcal{B}_{f_{i_1}, f_{i_2}, W}^c \mathcal{E}_{f_{i_1}} \mathcal{T}_N s_{N, W}^{(\ell)}, \mathcal{B}_{f_{i_1}, f_{i_2}, W}^c \mathcal{E}_{f_{i_2}} \mathcal{T}_N s_{N, W}^{(\ell')} \right\rangle. \end{aligned}$$

From (32), (34), and the fact that  $|f_{i_2} - f_{i_1}| \geq 2W$ , it follows that

$$|\langle \mathbf{q}_1, \mathbf{q}_2 \rangle| \leq \sqrt{\lambda_{N, W}^{(\ell)}} \cdot \sqrt{1 - \lambda_{N, W}^{(\ell')}} + \sqrt{1 - \lambda_{N, W}^{(\ell)}} \cdot \sqrt{\lambda_{N, W}^{(\ell')}} + \sqrt{1 - \lambda_{N, W}^{(\ell)}} \cdot \sqrt{1 - \lambda_{N, W}^{(\ell')}}.$$

Finally, (51) follows by bounding the  $\sqrt{\lambda}$  terms in this expression by 1 and by bounding the  $\sqrt{1 - \lambda}$  terms in this expression using (39).  $\square$

## References

- [1] W. Bajwa, J. Haupt, G. Raz, S. Wright, and R. Nowak. Toeplitz-structured compressed sensing matrices. In *Proc. IEEE Work. Stat. Signal Processing*, Madison, WI, Aug. 2007.
- [2] R. Baraniuk. Compressive sensing. *IEEE Signal Processing Mag.*, 24(4):118–120, 124, 2007.
- [3] R. Baraniuk, V. Cevher, M. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Trans. Inform. Theory*, 56(4):1982–2001, 2010.
- [4] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Const. Approx.*, 28(3):253–263, 2008.
- [5] T. Blumensath. Sampling and reconstructing signals from a union of linear subspaces. *IEEE Trans. Inform. Theory*, 57(7):4660–4671, 2011.
- [6] T. Blumensath and M. Davies. Iterative hard thresholding for compressive sensing. *Appl. Comput. Harmon. Anal.*, 27(3):265–274, 2009.
- [7] T. Blumensath and M. Davies. Sampling theorems for signals from the union of finite-dimensional linear subspaces. *IEEE Trans. Inform. Theory*, 55(4):1872–1882, 2009.

- [8] Y. Bresler and P. Feng. Spectrum-blind minimum-rate sampling and reconstruction of 2D multiband signals. In *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Zurich, Switzerland, Sept. 1996.
- [9] V. Buldygin and Y. Kozachenko. *Metric Characterization of Random Variables and Random Processes*. American Mathematical Society, Providence, RI, 2000.
- [10] E. Candès. Compressive sampling. In *Proc. Int. Congress of Math.*, Madrid, Spain, Aug. 2006.
- [11] E. Candès and M. Davenport. How well can we estimate a sparse vector? *Arxiv preprint arXiv:1104.5246*, 2011.
- [12] E. Candès, Y. Eldar, D. Needell, and P. Randall. Compressed sensing with coherent and redundant dictionaries. *Appl. Comput. Harmon. Anal.*, 31(1):59–73, 2011.
- [13] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.
- [14] E. Candès and T. Tao. Decoding by linear programming. *IEEE Trans. Inform. Theory*, 51(12):4203–4215, 2005.
- [15] A. Cohen, W. Dahmen, and R. DeVore. Instance optimal decoding by thresholding in compressed sensing. In *Proc. Int. Conf. Harmon. Anal. and Partial Diff. Eq.*, El Escorial Madrid, Spain, Jun. 2008.
- [16] W. Dai and O. Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Trans. Inform. Theory*, 55(5):2230–2249, 2009.
- [17] I. Daubechies and R. DeVore. Approximating a bandlimited function using very coarsely quantized data: A family of stable sigma-delta modulators of arbitrary order. *Ann. Math.*, 158(2):679–710, 2003.
- [18] M. Davenport. Concentration of measure and sub-gaussian distributions, 2009. available online at <http://cnx.org/content/m32583/latest/>.
- [19] M. Davenport, P. Boufounos, and R. Baraniuk. Compressive domain interference cancellation. In *Proc. Work. Struc. Parc. Rep. Adap. Signaux (SPARS)*, Saint-Malo, France, Apr. 2009.
- [20] M. Davenport, P. Boufounos, M. Wakin, and R. Baraniuk. Signal processing with compressive measurements. *IEEE J. Select. Top. Signal Processing*, 4(2):445–460, 2010.
- [21] M. Davenport, J. Laska, J. Treichler, and R. Baraniuk. The pros and cons of compressive sensing for wideband signal acquisition: Noise folding vs. dynamic range. *Arxiv preprint arXiv:1104.4842*, 2011.
- [22] M. Davenport, S. Schnelle, J.P. Slavinsky, R. Baraniuk, M. Wakin, and P. Boufounos. A wideband compressive radio receiver. In *Proc. Military Communications Conference (MILCOM)*, San Jose, CA, Oct. 2010.
- [23] M. Davenport and M. Wakin. Analysis of Orthogonal Matching Pursuit using the restricted isometry property. *IEEE Trans. Inform. Theory*, 56(9):4395–4401, 2010.
- [24] M. Davenport and M. Wakin. Reconstruction and cancellation of sampled multiband signals using Discrete Prolate Spheroidal Sequences. In *Proc. Work. Struc. Parc. Rep. Adap. Signaux (SPARS)*, Edinburgh, Scotland, Jun. 2011.
- [25] G. Davis, S. Mallat, and Z. Zhang. Adaptive time-frequency decompositions. *SPIE J. Opt. Engin.*, 33(7):2183–2191, 1994.
- [26] R. DeVore, G. Petrova, and P. Wojtaszczyk. Instance-optimality in probability with an  $\ell_1$ -minimization decoder. *Appl. Comput. Harmon. Anal.*, 27(3):275–288, 2009.
- [27] D. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.
- [28] D. Donoho, I. Drori, Y. Tsaig, and J.-L. Starck. Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit. Stanford Technical Report, 2006.
- [29] M. Duarte and R. Baraniuk. Spectral compressive sensing. Preprint, 2010.
- [30] Y. Eldar, P. Kuppinger, and H. Bölcskei. Block-sparse signals: Uncertainty relations and efficient recovery. *IEEE Trans. Signal Processing*, 58(6):3042–3054, 2010.
- [31] Y. Eldar and M. Mishali. Robust recovery of signals from a structured union of subspaces. *IEEE Trans. Inform. Theory*, 55(11):5302–5316, 2009.
- [32] C. Fancourt and J. Principe. On the relationship between the Karhunen-Loeve transform and the prolate spheroidal wave functions. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Istanbul, Turkey, Jun. 2000.
- [33] P. Feng. *Universal spectrum blind minimum rate sampling and reconstruction of multiband signals*. PhD thesis, University of Illinois at Urbana-Champaign, Mar. 1997.

- [34] P. Feng and Y. Bresler. Spectrum-blind minimum-rate sampling and reconstruction of multiband signals. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Atlanta, GA, May 1996.
- [35] S. Geršgorin. Über die Abgrenzung der Eigenwerte einer Matrix. *Izv. Akad. Nauk SSSR Ser. Fiz.-Mat.*, 6:749–754, 1931.
- [36] L. Gousse. Compressed sensing with preconditioning for sparse recovery with subsampled matrices of Slepian prolate functions. Preprint, 2010.
- [37] P. C. Hansen. Regularization Tools Version 4.0 for Matlab 7.3. *Numerical Algorithms*, 46:189–194, 2007.
- [38] S. Hein and A. Zakhor. Theoretical and numerical aspects of an SVD-based method for band-limiting finite-extent sequences. *IEEE Trans. Signal Processing*, 42(5):1227–1230, 1994.
- [39] S. Izu and J. Lakey. Time-frequency localization and sampling of multiband signals. *Acta Applicandae Mathematicae*, 107(1):399–435, 2009.
- [40] H. Landau. Necessary density conditions for sampling and interpolation of certain entire functions. *Acta Math.*, 117:37–52, 1967.
- [41] H. Landau and H. Pollak. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. II. *Bell Systems Tech. J.*, 40(1):65–84, 1961.
- [42] H. Landau and H. Pollak. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. III. *Bell Systems Tech. J.*, 41(4):1295–1336, 1962.
- [43] M. Leta, M. Davies, and J. Thompson. Reconciling compressive sampling systems for spectrally-sparse continuous-time signals. *Arxiv preprint arXiv:1101.4100*, 2011.
- [44] Y. Lu and M. Do. A theory for sampling signals from a union of subspaces. *IEEE Trans. Signal Processing*, 56(6):2334–2345, 2008.
- [45] R. Lyons. *Understanding Digital Signal Processing*. Pearson Education, Inc., Upper Saddle River, NJ, 2004.
- [46] M. Mishali and Y. Eldar. Blind multi-band signal reconstruction: Compressed sensing for analog signals. *IEEE Trans. Signal Processing*, 57(3):993–1009, 2009.
- [47] M. Mishali and Y. Eldar. From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals. *IEEE J. Select. Top. Signal Processing*, 4(2):375–391, 2010.
- [48] D. Mugler, Y. Wu, and S. Clary. Linear prediction of bandpass signals based on past samples. In *Proc. Sampling Theory and Applications (SampTA)*, Loen, Norway, Aug. 1999.
- [49] D. Needell and J. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.*, 26(3):301–321, 2009.
- [50] D. Needell and R. Vershynin. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit. *Found. Comput. Math.*, 9(3):317–334, 2009.
- [51] D. Needell and R. Vershynin. Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. *IEEE J. Select. Top. Signal Processing*, 4(2):310–316, 2010.
- [52] J. Oh, S. Senay, and L. Chaparro. Signal reconstruction from nonuniformly spaced samples using evolutionary Slepian transform-based POCS. *EURASIP J. Adv. Signal Processing*, 2010.
- [53] Y. Pati, R. Rezaifar, and P. Krishnaprasad. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. In *Proc. Asilomar Conf. Signals, Systems, and Computers*, Pacific Grove, CA, Nov. 1993.
- [54] D. Phillips. A technique for the numerical solution of certain integral equations of the first kind. *J. ACM*, 9:84–97, 1962.
- [55] R. Prony. Essai expérimental et analytique sur les lois de la Dilatabilité des fluides élastiques et sur celles de la Force expansive de la vapeur de l’eau et de la vapeur de l’alkool, à différentes températures. *J. de l’École Polytechnique*, Floréal et Prairial III, 1(2):24–76, 1795. R. Prony is Gaspard Riche, baron de Prony.
- [56] J. Romberg. Compressive sensing by random convolution. *SIAM J. Imag. Sci.*, 2(4):1098–1128, 2009.
- [57] L. Scharf. The SVD and reduced rank signal processing. *Signal Processing*, 25(2):113–133, 1991.
- [58] E. Sejdić, M. Luccini, S. Primak, K. Baddour, and T. Willink. Channel estimation using DPSS based frames. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Las Vegas, Nevada, Mar. 2008.
- [59] S. Senay, L. Chaparro, and L. Durak. Reconstruction of nonuniformly sampled time-limited signals using prolate spheroidal wave functions. *Signal Processing*, 89(12):2585–2595, 2009.

- [60] S. Senay, L. Chaparro, M. Sun, and R. Scabassi. Compressive sensing and random filtering of EEG signals using Slepian basis. In *Proc. European Signal Processing Conf. (EUSIPCO)*, Lausanne, Switzerland, Aug. 2008.
- [61] J. P. Slavinsky, J. Laska, M. Davenport, and R. Baraniuk. The compressive mutliplexer for multi-channel compressive sensing. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Prague, Czech Republic, May 2011.
- [62] D. Slepian. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. IV. *Bell Systems Tech. J.*, 43(6):3009–3058, 1964.
- [63] D. Slepian. On Bandwidth. *Proc. IEEE*, 64(3):292–300, 1976.
- [64] D. Slepian. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. V – The discrete case. *Bell Systems Tech. J.*, 57(5):1371–1430, 1978.
- [65] D. Slepian. Some comments on Fourier analysis, uncertainty, and modeling. *SIAM Rev.*, 25(3):379–393, 1983.
- [66] D. Slepian and H. Pollak. Prolate spheroidal wave functions, Fourier analysis, and uncertainty. I. *Bell Systems Tech. J.*, 40(1):43–64, 1961.
- [67] H. Stark and J. Woods. *Probability, random processes, and estimation theory for engineers*. Prentice-Hall, 1986.
- [68] D. Thomson. Spectrum estimation and harmonic analysis. *Proc. IEEE*, 70(9):1055–1096, 1982.
- [69] A. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Dokl. Akad. Nauk. SSSR*, 151:1035–1038, 1963.
- [70] A. Tikhonov and V. Arsenin. *Solutions of Ill-Posed Problems*. Winston & Sons, Washington, D.C., 1977.
- [71] J. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory*, 50(10):2231–2242, 2004.
- [72] J. Tropp, J. Laska, M. Duarte, J. Romberg, and R. Baraniuk. Beyond Nyquist: Efficient sampling of sparse, bandlimited signals. *IEEE Trans. Inform. Theory*, 56(1):520–544, 2010.
- [73] J. Tropp, M. Wakin, M. Duarte, D. Baron, and R. Baraniuk. Random filters for compressive sampling and reconstruction. In *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP)*, Toulouse, France, May 2006.
- [74] R. Venkataramani and Y. Bresler. Further results on spectrum blind sampling of 2D signals. In *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Chicago, IL, Oct. 1998.
- [75] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. *Arxiv preprint arXiv:1011.3027*, 2010.
- [76] R. Walden. Analog-to-digital converter survey and analysis. *IEEE J. Selected Areas Comm.*, 17(4):539–550, 1999.