Signal Processing With Compressive Measurements

Mark A. Davenport, *Student Member, IEEE*, Petros T. Boufounos, *Member, IEEE*, Michael B. Wakin, *Member, IEEE*, and Richard G. Baraniuk, *Fellow, IEEE*

***Abstract*—The recently introduced theory of compressive sensing enables the recovery of sparse or compressible signals from a small set of nonadaptive, linear measurements. If properly chosen, the number of measurements can be much smaller than the number of Nyquist-rate samples. Interestingly, it has been shown that *random projections* are a near-optimal measurement scheme. This has inspired the design of hardware systems that directly implement random measurement protocols. However, despite the intense focus of the community on signal recovery, many (if not most) signal processing problems do not require full signal recovery. In this paper, we take some first steps in the direction of solving *inference* problems—such as *detection*, *classification*, or *estimation*—and *filtering* problems using only compressive measurements and without ever reconstructing the signals involved. We provide theoretical bounds along with experimental results.**

***Index Terms*—Compressive sensing (CS), compressive signal processing, estimation, filtering, pattern classification, random projections, signal detection, universal measurements.**

I. Introduction

A. From DSP to CSP

In recent decades, the digital signal processing (DSP) community has enjoyed enormous success in developing algorithms for capturing and extracting information from signals. Capitalizing on the early work of Whitaker, Nyquist, 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.

Manuscript received February 28, 2009; revised November 12, 2009. Current version published March 17, 2010. The work of M. A. Davenport and R. G. Baraniuk was supported by the Grants NSF CCF-0431150, CCF-0728867, CNS-0435425, and CNS-0520280, DARPA/ONR N66001-08-1-2065, ONR N00014-07-1-0936, N00014-08-1-1067, N00014-08-1-1112, and N00014-08-1-1066, AFOSR FA9550-07-1-0301, ARO MURI W311NF-07-1-0185, ARO MURI W911NF-09-1-0383, and by the Texas Instruments Leadership University Program. The work of M. B. Wakin was supported by NSF Grants DMS-0603606 and CCF-0830320, and DARPA Grant HR0011-08-1-0078. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Rick Chartrand.

M. A. Davenport and R. G. Baraniuk are with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005 USA (e-mail: md@rice.edu; richb@rice.edu).

P. T. Boufounos is with Mitsubishi Electric Research Laboratories, Cambridge, MA 02139 USA (e-mail: petrosb@merl.com).

M. B. Wakin is with the Division of Engineering, Colorado School of Mines, Golden, CO 80401 USA (e-mail: mwakin@mines.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSTSP.2009.2039178

As a result of this success, the amount of data generated by sensing systems has grown from a trickle to a torrent. We are thus confronted with the following challenges:

1) acquiring signals at ever higher sampling rates;
2) storing the resulting large amounts of data;
3) processing/analyzing large amounts of data.

Until recently, the first challenge was largely ignored by the signal processing community, with most advances being made by hardware designers. Meanwhile, the signal processing community has made tremendous progress on the remaining two challenges, largely via research in the fields of modeling, compression, and dimensionality reduction. However, the solutions to these problems typically rely on having a complete set of digital samples. Only after the signal is acquired—presumably using a sensor tailored to the signals of interest—could one distill the necessary information ultimately required by the user or the application. This requirement has placed a growing burden on analog-to-digital converters [1]. As the required sampling rate is pushed higher, these devices move inevitably toward a physical barrier, beyond which their design becomes increasingly difficult and costly [2].

Thus, in recent years, the signal processing community has also begun to address the challenge of signal acquisition more directly by leveraging its successes in addressing the second two. In particular, *compressive sensing* (CS) has emerged as a framework that can significantly reduce the acquisition cost at a sensor. CS builds on the work of Candès, Romberg, and Tao [3], and Donoho [4], 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.

A fundamental difference between CS and classical sampling is the manner in which the two frameworks deal with *signal recovery*, i.e., the problem of recovering the signal from the measurements. In the Shannon–Nyquist framework, signal recovery is achieved through sinc interpolation—a linear process that requires little computation and has a simple interpretation. In CS, however, signal recovery is achieved using nonlinear and relatively expensive optimization-based or iterative algorithms [3]–[5]. Thus, up to this point, most of the CS literature has focused on improving the speed and accuracy of this process [6]–[9].

However, signal recovery is not actually necessary in many signal processing applications. Very often we are only interested in solving an *inference* problem (extracting certain information from measurements) or in *filtering* out information that is not of interest before further processing. While one could always attempt to recover the full signal from the compressive measurements and then solve the inference or filtering problem using tra-

ditional DSP techniques, this approach is typically suboptimal in terms of both accuracy and efficiency.

This paper takes some initial steps towards a general framework for what we call *compressive signal processing* (CSP), an alternative approach in which signal processing problems are solved directly in the compressive measurement domain *without* first resorting to a full-scale signal reconstruction. In espousing the potential of CSP we focus on four fundamental signal processing problems: detection, classification, estimation, and filtering. The first three enable the extraction of information from the samples, while the last enables the removal of irrelevant information and separation of signals into distinct components. While these choices do not exhaust the set of canonical signal processing operations, we believe that they provide a strong initial foundation.

B. Relevance

In what settings is it actually beneficial to take randomized, compressive measurements of a signal in order to solve an inference problem? One may argue that prior knowledge of the signal to be acquired or of the inference task to be solved could lead to a customized sensing protocol that very efficiently acquires the relevant information. For example, suppose we wish to acquire a length-N signal that is K-sparse (i.e., has K nonzero coefficients) in a known transform basis. If we knew in advance which elements were nonzero, then the most efficient and direct measurement scheme would simply project the signal into the appropriate K-dimensional subspace. As a second example, suppose we wish to detect a known signal. If we knew in advance the signal template, then the optimal and most efficient measurement scheme would simply involve a receiving filter explicitly matched to the candidate signal.

Clearly, in cases where strong *a priori* information is available, customized sensing protocols may be appropriate. However, a key objective of this paper is to illustrate the *agnostic* and *universal* nature of random compressive measurements as a compact signal representation. These features enable the design of exceptionally *efficient* and *flexible* compressive sensing hardware that can be used for the acquisition of a variety of signal classes and applied to a variety of inference tasks.

As has been demonstrated in the CS literature, for example, random measurements can be used to acquire any sparse signal without requiring advance knowledge of the locations of the nonzero coefficients. Thus, compressive measurements are agnostic in the sense that they capture the relevant information for the entire class of possible K-sparse signals. We extend this concept to the CSP framework and demonstrate that it is possible to design agnostic measurement schemes that preserve the necessary structure of large signal classes in a variety of signal processing settings.

Furthermore, we observe that one can select a randomized measurement scheme without any prior knowledge of the signal class. For instance, in conventional CS it is not necessary to know the transform basis in which the signal has a sparse representation when acquiring the measurements. The only dependence is between the complexity of the signal class (e.g., the sparsity level of the signal) and the number of random measurements that must be acquired. Thus, random compressive measurements are universal in the sense that if one designs a measurement scheme at random, then with high probability it will preserve the structure of the signal class of interest, and thus explicit *a priori* knowledge of the signal class is unnecessary. We broaden this result and demonstrate that random measurements can universally capture the information relevant for many CSP applications without any prior knowledge of either the signal class or the ultimate signal processing task. In such cases, the requisite number of measurements scales efficiently with both the complexity of the signal and the complexity of the task to be performed.

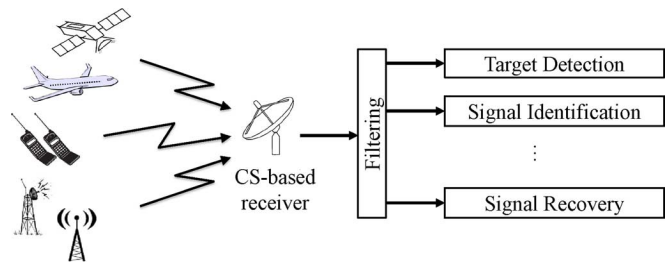

Fig. 1. Example CSP application: broadband signal monitoring.

It follows that, in contrast to the task-specific hardware used in many classical acquisition systems, hardware designed to use a compressive measurement protocol can be extremely flexible. Returning to the binary detection scenario, for example, suppose that the signal template is unknown at the time of acquisition, or that one has a large number of candidate templates. Then what information should be collected at the sensor? A complete set of Nyquist samples would suffice, or a bank of matched filters could be employed. From a CSP standpoint, however, the solution is more elegant: one need only collect a small number of compressive measurements from which many candidate signals can be tested, many signal models can be posited, and many other inference tasks can be solved. What one loses in performance compared to a tailor-made matched filter, one may gain in simplicity and in the ability to adapt to future information about the problem at hand. In this sense, CSP impacts sensors in a similar manner as DSP impacted analog signal processing: expensive and inflexible analog components can be replaced by a universal, flexible, and programmable digital system.

C. Applications

A stylized application to demonstrate the potential and applicability of the results in this paper is summarized in Fig. 1. The figure schematically presents a wide-band signal monitoring and processing system that receives signals from a variety of sources, including various television, radio, and cell-phone transmissions, radar signals, and satellite communication signals. The extremely wide bandwidth monitored by such a system makes CS a natural approach for efficient signal acquisition [10].

In many cases, the system user might only be interested in extracting very small amounts of information from each signal. This can be performed efficiently using the tools we describe in the subsequent sections. For example, the user might be interested in detecting and classifying some of the signal sources, and in estimating some parameters, such as the location, of others.

Full-scale signal recovery might be required for only a few of the signals in the monitored bandwidth.

The detection, estimation, and classification tools we develop in this paper enable the system to perform these tasks much more efficiently in the compressive domain. Furthermore, the filtering procedure we describe facilitates the separation of signals after they have been acquired in the compressive domain so that each signal can be processed by the appropriate algorithm, depending on the information sought by the user.

D. Related Work

In this paper, we consider a variety of estimation and decision tasks. The *data streaming* community, which is concerned with efficient algorithms for processing large streams of data, has examined many similar problems over the past several years. In the data stream setting, one is typically interested in estimating some function of the data stream (such as an ℓ_p norm, a histogram, or a linear functional) based on *sketches*, which in many cases can be thought of as random projections. For a concise review of these results, see [11]. The main differences with our work include the following: 1) data stream algorithms are typically designed to operate in noise-free environments on man-made digital signals, whereas we view compressive measurements as a sensing scheme that will operate in an inherently noisy environment; 2) data stream algorithms typically provide probabilistic guarantees, while we focus on providing deterministic guarantees; and 3) data stream algorithms tend to tailor the measurement scheme to the task at hand, while we demonstrate that it is often possible to use the same measurements for a variety of signal processing tasks.

There have been a number of related thrusts involving detection and classification using random measurements in a variety of settings. For example, in [12] sparsity is leveraged to perform classification with very few random measurements, while in [13], [14] random measurements are exploited to perform manifold-based image classification. In [15], small numbers of random measurements have also been noted as capturing sufficient information to allow robust face recognition. However, the most directly relevant work has been the discussions of classification in [16] and detection in [17]. We will contrast our results to those of [16], [17] below. This paper builds upon work initially presented in [18] and [19].

E. Organization

This paper is organized as follows. Section II provides the necessary background on dimensionality reduction and CS. In Sections III–V, we develop algorithms for signal detection, classification, and estimation with compressive measurements. In Section VI, we explore the problem of filtering compressive measurements in the compressive domain. Finally, Section VII concludes with directions for future work.

II. Compressive Measurements and Stable Embeddings

A. Compressive Sensing and Restricted Isometries

In the standard CS framework, we acquire a signal $x \in \mathbb{R}^N$ via the linear measurements

$$y = \Phi x \tag{1}$$

where Φ is an $M \times N$ matrix representing the sampling system and $y \in \mathbb{R}^M$ is the vector of measurements. For simplicity, we deal with real-valued rather than quantized measurements y. Classical sampling theory dictates that, in order to ensure that there is no loss of information, the number of samples M should be as large as the signal dimension N. The CS theory, on the other hand, allows for $M \ll N$ as long as the signal x is sparse or compressible in some basis [3], [4], [20], [21].

To understand how many measurements are required to enable the recovery of a signal x, we must first examine the properties of Φ that guarantee satisfactory performance of the sensing system. In [21], Candès and Tao introduced the *restricted isometry property* (RIP) of a matrix Φ and established its important role in CS. First define Σ_K to be the set of all K-sparse signals, i.e.,

$$\Sigma_K = \{x \in \mathbb{R}^N : \|x\|_0 := |\mathrm{supp}(x)| \leq K\}$$

where $\mathrm{supp}(x) \subset \mathbb{R}^N$ denotes the set of indices on which x is nonzero. We say that a matrix Φ satisfies the RIP of order K if there exists a constant $\delta \in (0,1)$, such that

$$(1-\delta)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1+\delta)\|x\|_2^2 \tag{2}$$

holds for all $x \in \Sigma_K$. In other words, Φ is an approximate isometry for vectors restricted to be K-sparse.

It is clear that if we wish to be able to recover all K-sparse signals x from the measurements y, then a necessary condition on Φ is that $\Phi x_1 \neq \Phi x_2$ for any pair $x_1, x_2 \in \Sigma_K$ with $x_1 \neq x_2$. Equivalently, we require $\|\Phi(x_1 - x_2)\|_2^2 > 0$, which is guaranteed if Φ satisfies the RIP of order $2K$ with constant $\delta < 1$. Furthermore, the RIP also ensures that a variety of practical algorithms can successfully recover any compressible signal from noisy measurements. The following result (Theorem 1.2 of [22]) makes this precise by bounding the recovery error of x with respect to the measurement noise and with respect to the ℓ_1-distance from x to its best K-term approximation denoted x_K

$$x_K = \arg \min_{x' \in \Sigma_K} \|x - x'\|_1.$$

Theorem 1 [Candès]: Suppose that Φ satisfies the RIP of order $2K$ with isometry constant $\delta < \sqrt{2} - 1$. Given measurements of the form $y = \Phi x + e$, where $\|e\|_2 \leq \epsilon$, the solution to

$$\widehat{x} = \arg\min_{x' \in \mathbb{R}^N} \|x'\|_1 \text{ subject to } \|\Phi x' - y\|_2 \leq \epsilon \tag{3}$$

obeys

$$\|\widehat{x} - x\|_2 \leq C_0 \epsilon + C_1 \frac{\|x - x_K\|_1}{\sqrt{K}} \tag{4}$$

where

$$C_0 = 4\frac{\sqrt{1+\delta}}{1-(1+\sqrt{2})\delta}, \quad C_1 = 2\frac{1-(1-\sqrt{2})\delta}{1-(1+\sqrt{2})\delta}.$$

Note that in practice we may wish to acquire signals that are sparse or compressible with respect to a certain *sparsity basis* Ψ, i.e., $x = \Psi\alpha$, where Ψ is represented as a unitary $N \times N$ matrix and $\alpha \in \Sigma_K$. In this case, we would require instead that $\Phi\Psi$ satisfy the RIP, and the performance guarantee would be on $\|\widehat{\alpha} - \alpha\|_2$.

Before we discuss how one can actually obtain a matrix Φ that satisfies the RIP, we observe that we can restate the RIP in a more general form. Let $\delta \in (0,1)$ and $\mathcal{U}, \mathcal{V} \subset \mathbb{R}^N$ be given. We say that a mapping Φ is a *δ-stable embedding* of $(\mathcal{U}, \mathcal{V})$ if

$$(1-\delta)||u-v||_2^2 \leq ||\Phi u - \Phi v||_2^2 \leq (1+\delta)||u-v||_2^2 \quad (5)$$

for all $u \in \mathcal{U}$ and $v \in \mathcal{V}$. A mapping satisfying this property is also commonly called *bi-Lipschitz*. Observe that for a matrix Φ, satisfying the RIP of order $2K$ is equivalent to being a δ-stable embedding of (Σ_K, Σ_K) or of $(\Sigma_{2K}, \{0\})$.[1] Furthermore, if the matrix $\Phi\Psi$ satisfies the RIP of order $2K$ then Φ is a δ-stable embedding of $(\Psi(\Sigma_K), \Psi(\Sigma_K))$ or $(\Psi(\Sigma_{2K}), \{0\})$, where $\Psi(\Sigma_K) = \{\Psi\alpha : \alpha \in \Sigma_K\}$.

B. Random Matrix Constructions

We now turn to the more general question of how to construct linear mappings Φ that satisfy (5) for particular sets $\mathcal{U}$ and $\mathcal{V}$. While it is possible to obtain deterministic constructions of such operators, at present the most efficient designs (i.e., those requiring the fewest number of rows) rely on random matrix constructions. We construct our random matrices as follows: given M and N, we generate random $M \times N$ matrices Φ by choosing the entries ϕ_{ij} as independent and identically distributed (i.i.d.) random variables. We impose two conditions on the random distribution. First, we require that the distribution yields a matrix that is norm-preserving, which requires that

$$\mathbb{E}(\phi_{ij}^2) = \frac{1}{M}. \quad (6)$$

Second, we require that the distribution is a *sub-Gaussian* distribution, meaning that there exists a constant $C > 0$ such that

$$\mathbb{E}\left(e^{\phi_{ij}t}\right) \leq e^{C^2t^2/2} \quad (7)$$

for all $t \in \mathbb{R}$. This says that the moment-generating function of our distribution is dominated by that of a Gaussian distribution, which is also equivalent to requiring that the tails of our distribution decay *at least as fast* as the tails of a Gaussian distribution. Examples of sub-Gaussian distributions include the Gaussian distribution, the Rademacher distribution, and the uniform distribution. In general, any distribution with bounded support is sub-Gaussian. See [23] for more details on sub-Gaussian random variables.

The key property of sub-Gaussian random variables that will be of use in this paper is that for any $x \in \mathbb{R}^N$, the random variable $||\Phi x||_2^2$ is highly concentrated about $||x||_2^2$; that is, there exists a constant $c > 0$ that depends only on the constant C in (7) such that

$$\Pr\left(\left|||\Phi x||_2^2 - ||x||_2^2\right| \geq \delta\, ||x||_2^2\right) \leq 2e^{-cM\delta^2} \quad (8)$$

where the probability is taken over all $M \times N$ matrices Φ (see Lemma 6.1 of [24] or [25]).

[1]In general, if Φ is a δ-stable embedding of $(\mathcal{U}, \mathcal{V})$, this is equivalent to it being a δ-stable embedding of $(\widetilde{\mathcal{U}}, \{0\})$, where $\widetilde{\mathcal{U}} = \{u - v : u \in \mathcal{U}, v \in \mathcal{V}\}$. This formulation can sometimes be more convenient.

C. Stable Embeddings

We now provide a number of results that we will use extensively in the sequel to ensure the stability of our compressive detection, classification, estimation, and filtering algorithms.

We start with the simple case where we desire a δ-stable embedding of $(\mathcal{U}, \mathcal{V})$, where $\mathcal{U} = \{u_i\}_{i=1}^{|\mathcal{U}|}$ and $\mathcal{V} = \{v_j\}_{j=1}^{|\mathcal{V}|}$ are finite sets of points in $\mathbb{R}^N$. In the case where $\mathcal{U} = \mathcal{V}$, this is essentially the Johnson–Lindenstrauss (JL) lemma [26]–[28].

Lemma 1: Let $\mathcal{U}$ and $\mathcal{V}$ be sets of points in $\mathbb{R}^N$. Fix $\delta, \beta \in (0,1)$. Let Φ be an $M \times N$ random matrix with i.i.d. entries chosen from a distribution satisfying (8). If

$$M \geq \frac{\ln\left(|\mathcal{U}||\mathcal{V}|\right) + \ln\left(\frac{2}{\beta}\right)}{c\delta^2} \quad (9)$$

then with probability exceeding $1-\beta$, Φ is a δ-stable embedding of $(\mathcal{U}, \mathcal{V})$.

Proof: To prove the result we apply (8) to the $|\mathcal{U}||\mathcal{V}|$ vectors corresponding to all possible $u_i - v_j$. By applying the union bound, we obtain that the probability of (5) not holding is bounded above by $2|\mathcal{U}||\mathcal{V}|e^{-cM\delta^2}$. By requiring $2|\mathcal{U}||\mathcal{V}|e^{-cM\delta^2} \leq \beta$ and solving for M we obtain the desired result. ■

We now consider the case where $\mathcal{U} = \mathcal{X}$ is a K-dimensional subspace of $\mathbb{R}^N$ and $\mathcal{V} = \{0\}$. Thus, we wish to obtain a Φ that nearly preserves the norm of any vector $x \in \mathcal{X}$. At first glance, this goal might seem very different than the setting for Lemma 1, since a subspace forms an *uncountable* point set. However, we will see that the dimension K bounds the complexity of this space, and thus it can be characterized in terms of a finite number of points. The following lemma is an adaptation of [29, Lemma 5.1].[2]

Lemma 2: Suppose that $\mathcal{X}$ is a K-dimensional subspace of $\mathbb{R}^N$. Fix $\delta, \beta \in (0,1)$. Let Φ be an $M \times N$ random matrix with i.i.d. entries chosen from a distribution satisfying (8). If

$$M \geq 2\frac{K\ln\left(\frac{42}{\delta}\right) + \ln\left(\frac{2}{\beta}\right)}{c\delta^2} \quad (10)$$

then with probability exceeding $1-\beta$, Φ is a δ-stable embedding of $(\mathcal{X}, \{0\})$.

Sketch of Proof: It suffices to prove the result for $x \in \mathcal{X}$ satisfying $||x||_2 = 1$, since Φ is linear. We consider a finite sampling of points $\mathcal{U} \subset \mathcal{X}$ of unit norm and with resolution on the order of $\delta/14$. One can show that it is possible to construct such a $\mathcal{U}$ with $|\mathcal{U}| \leq (42/\delta)^K$ (see [30, Ch. 15]). Applying Lemma 1 and setting M to ensure a $(\delta/\sqrt{2})$-stable embedding of $(\mathcal{U}, \{0\})$, we can use simple geometric arguments to conclude that we must have a δ-stable embedding of $(\{x\}, \{0\})$ for every $x \in \mathcal{X}$ satisfying $||x||_2 = 1$. For details, see [29, Lemma 5.1]. ■

We now observe that we can extend this result beyond a single K-dimensional subspace to all possible K-dimensional subspaces that are defined with respect to an orthonormal basis Ψ, i.e., $\Psi(\Sigma_K)$. The proof follows that of [29, Theorem 5.2].

[2]The constants in [29] differ from those in Lemma 2, but the proof is substantially the same, so we provide only a sketch.

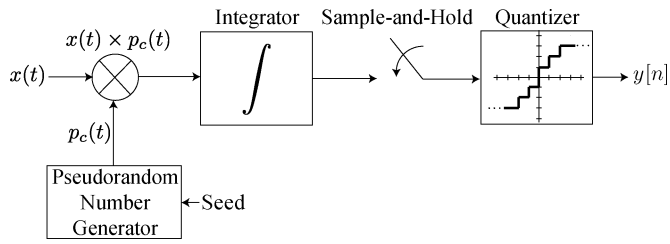

Fig. 2. Random demodulator for obtaining compressive measurements of analog signals.

Lemma 3: Let Ψ be an orthonormal basis for $\mathbb{R}^N$ and fix $\delta, \beta \in (0,1)$. Let Φ be an $M \times N$ random matrix with i.i.d. entries chosen from a distribution satisfying (8). If

$$M \geq 2\frac{K \ln\left(\frac{42eN}{\delta K}\right) + \ln\left(\frac{2}{\beta}\right)}{c\delta^2} \tag{11}$$

with e denoting the base of the natural logarithm, then with probability exceeding $1-\beta$, Φ is a δ-stable embedding of $(\Psi(\Sigma_K), \{0\})$.

Proof: This is a simple generalization of Lemma 2, which follows from the observation that there are $\binom{N}{K} \leq (eN/K)^K$ K-dimensional subspaces aligned with the coordinate axes of Ψ, and so the size of $\mathcal{U}$ increases to $|\mathcal{U}| \leq (42eN/\delta K)^K$. ■

A similar technique has recently been used to demonstrate that random projections also provide a stable embedding of nonlinear manifolds [31]: under certain assumptions on the curvature and volume of a K-dimensional manifold $\mathcal{M} \subset \mathbb{R}^N$, a random sensing matrix with $M = O\left(K\log(N)/\delta^2\right)$ will with high probability provide a δ-stable embedding of $(\mathcal{M}, \mathcal{M})$. Under slightly different assumptions on $\mathcal{M}$, a number of similar embedding results involving random projections have been established [32]–[34].

We will make further use of these connections in the following sections in our analysis of a variety of algorithms for compressive-domain inference and filtering.

D. Stable Embeddings in Practice

Several hardware architectures have been proposed that enable the acquisition of compressive measurements in practical settings. Examples include the random demodulator [35], random filtering and random convolution [36]–[38], and several compressive imaging architectures [39]–[41].

We briefly describe the random demodulator as an example of such a system. Fig. 2 depicts the block diagram of the random demodulator. The four key components are a pseudo-random ± 1 "chipping sequence" $p_c(t)$ operating at the Nyquist rate or higher, a low-pass filter, represented by an ideal integrator with reset, a low-rate sample-and-hold, and a quantizer. An input analog signal $x(t)$ is modulated by the chipping sequence and integrated. The output of the integrator is sampled and quantized, and the integrator is reset after each sample.

Mathematically, systems such as these implement a linear operator that maps the analog input signal to a discrete output vector followed by a quantizer. It is possible to relate this operator to a discrete measurement matrix Φ which maps, for example, the Nyquist-rate samples of the input signal to the discrete output vector. The resulting Φ matrices, while randomized, typically contain some degree of structure. For example, a random convolution architecture gives rise to a Φ matrix with a subsampled Toeplitz structure. While theoretical analysis of these matrices remains a topic of active study in the CS community, there do exist guarantees of stable embeddings for such practical architectures [35], [37].

E. Deterministic Versus Probabilistic Guarantees

Throughout this paper, we state a variety of theorems that begin with the assumption that Φ is a δ-stable embedding of a pair of sets and then use this assumption to establish performance guarantees for a particular CSP algorithm. These guarantees are completely deterministic and hold for any Φ that is a δ-stable embedding. However, we use random constructions as our main tool for obtaining stable embeddings. Thus, all of our results could be modified to be probabilistic statements in which we fix M and then argue that with high probability, a random Φ is a δ-stable embedding. Of course, the concept of "high probability" is somewhat arbitrary. However, if we fix this probability of error to be an acceptable constant β, then as we increase M, we are able to reduce δ to be arbitrarily close to 0. This will typically improve the accuracy of the guarantees.

As a side comment, it is important to note that in the case where one is able to generate a new Φ before acquiring each new signal x, then it is often possible to drastically reduce the required M. This is because one may be able to eliminate the requirement that Φ is a stable embedding for an entire class of candidate signals x, and instead simply argue that for each x, a new random matrix Φ with M very small is a δ-stable embedding of $(\{x\}, \{0\})$ (this is a direct consequence of (8)). Thus, if such a probabilistic "for each" guarantee is acceptable, then it is typically possible to place *no assumptions* on the signals being sparse, or indeed having any structure at all. However, in the remainder of this paper we will restrict ourselves to the sort of deterministic guarantees that hold for a class of signals when Φ provides a stable embedding of that class.

III. Detection With Compressive Measurements

A. Problem Setup and Applications

We begin by examining the simplest of detection problems. We aim to distinguish between two hypotheses:

$$\mathcal{H}_0 : y = \Phi n$$
$$\mathcal{H}_1 : y = \Phi(s+n)$$

where $s \in \mathbb{R}^N$ is a known signal, $n \sim \mathcal{N}(0, \sigma^2 I_N)$ is i.i.d. Gaussian noise, and Φ is a known (fixed) measurement matrix. If s is known at the time of the design of Φ, then it is easy to show that the optimal design would be to set $\Phi = s^T$, which is just the *matched filter*. However, as mentioned in the Introduction, we are often interested in universal or agnostic Φ. As an example, if we design hardware to implement the matched filter for a particular s, then we are very limited in what other signal processing tasks that hardware can perform. Even if we are only interested in detection, it is still possible that the signal

s that we wish to detect may evolve over time. Thus, we will consider instead the case where Φ is designed without knowledge of s but is instead a random matrix. From the results of Section II, this will imply performance bounds that depend on how many measurements are acquired and the class $\mathcal{S}$ of possible s that we wish to detect.

B. Theory

To set notation, let

$$P_F = \Pr(\mathcal{H}_1 \text{ chosen when } \mathcal{H}_0 \text{ true}) \text{ and}$$
$$P_D = \Pr(\mathcal{H}_1 \text{ chosen when } \mathcal{H}_1 \text{ true})$$

denote the *false alarm rate* and the *detection rate*, respectively. The *Neyman-Pearson* (NP) detector is the decision rule that maximizes P_D subject to the constraint that $P_F \leq \alpha$. In order to derive the NP detector, we first observe that for our hypotheses, $\mathcal{H}_0$ and $\mathcal{H}_1$, we have the probability density functions[3]

$$f_0(y) = \frac{\exp\left(-\frac{1}{2}y^T(\sigma^2\Phi\Phi^T)^{-1}y\right)}{|\sigma^2\Phi\Phi^T|^{1/2}(2\pi)^{M/2}}$$
$$\text{and } f_1(y) = \frac{\exp\left(-\frac{1}{2}(y-\Phi s)^T(\sigma^2\Phi\Phi^T)^{-1}(y-\Phi s)\right)}{|\sigma^2\Phi\Phi^T|^{1/2}(2\pi)^{M/2}}.$$

It is easy to show (see [42] and [43], for example) that the NP-optimal decision rule is to compare the ratio $f_1(y)/f_0(y)$ to a threshold η, i.e., the *likelihood ratio test*

$$\Lambda(y) = \frac{f_1(y)}{f_0(y)} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \eta$$

where η is chosen such that

$$P_F = \int_{\Lambda(y)>\eta} f_0(y)dy = \alpha.$$

By taking a logarithm we obtain an equivalent test that simplifies to

$$y^T(\Phi\Phi^T)^{-1}\Phi s \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \sigma^2\log(\eta) + \frac{1}{2}s^T\Phi^T(\Phi\Phi^T)^{-1}\Phi s := \gamma.$$

We now define the compressive detector

$$t := y^T(\Phi\Phi^T)^{-1}\Phi s. \tag{12}$$

It can be shown that t is a *sufficient statistic* for our detection problem, and thus t contains all of the information relevant for distinguishing between $\mathcal{H}_0$ and $\mathcal{H}_1$.

We must now set γ to achieve the desired performance. To simplify notation, we define

$$P_{\Phi^T} = \Phi^T(\Phi\Phi^T)^{-1}\Phi$$

[3]This formulation assumes that $\mathrm{rank}(\Phi) = M$ so that $\Phi\Phi^T$ is invertible. If the entries of Φ are generated according to a continuous distribution and $M < N$, then this will be true with high probability for discrete distributions provided that $M \ll N$. In the event that Φ is not full rank, appropriate adjustments can be made.

as the orthogonal projection operator onto $\mathcal{R}(\Phi^T)$, i.e., the row space of Φ. Since $P_{\Phi^T} = P_{\Phi^T}^T$ and $P_{\Phi^T}^2 = P_{\Phi^T}$, we then have that

$$s^T\Phi^T(\Phi\Phi^T)^{-1}\Phi s = \|P_{\Phi^T}s\|_2^2. \tag{13}$$

Using this notation, it is easy to show that

$$t \sim \begin{cases} \mathcal{N}(0, \sigma^2\|P_{\Phi^T}s\|_2^2) & \text{under } \mathcal{H}_0 \\ \mathcal{N}(\|P_{\Phi^T}s\|_2^2, \sigma^2\|P_{\Phi^T}s\|_2^2) & \text{under } \mathcal{H}_1 \end{cases}.$$

Thus, we have

$$P_F = P(t > \gamma|\mathcal{H}_0) = Q\left(\frac{\gamma}{\sigma\|P_{\Phi^T}s\|_2}\right)$$
$$P_D = P(t > \gamma|\mathcal{H}_1) = Q\left(\frac{\gamma - \|P_{\Phi^T}s\|_2^2}{\sigma\|P_{\Phi^T}s\|_2}\right)$$
$$\text{where } Q(z) = \int_z^\infty \exp\left(-\frac{u^2}{2}\right)du.$$

To determine the threshold, we set $P_F = \alpha$, and thus

$$\gamma = \sigma\|P_{\Phi^T}s\|_2 Q^{-1}(\alpha)$$

resulting in

$$P_D(\alpha) = Q\left(Q^{-1}(\alpha) - \frac{\|P_{\Phi^T}s\|_2}{\sigma}\right). \tag{14}$$

In general, this performance could be either quite good or quite poor depending on Φ. In particular, the larger $\|P_{\Phi^T}s\|_2$ is, then the better the performance. Recalling that P_{Φ^T} is the orthogonal projection onto the row space of Φ, we see that $\|P_{\Phi^T}s\|_2$ is simply the norm of the component of s that lies in the row space of Φ. This quantity is clearly at most $\|s\|_2$, which would yield the same performance as the traditional matched filter, but it could also be 0 if s lies in the null space of Φ. As we will see below, however, in the case where Φ is random, we can expect that $\|P_{\Phi^T}s\|_2$ concentrates around $\sqrt{M/N}\|s\|_2$.

Let us now define

$$\text{SNR} := \frac{\|s\|_2^2}{\sigma^2}. \tag{15}$$

We can bound the performance of the compressive detector as follows.

Theorem 2: Suppose that $\sqrt{N/M}P_{\Phi^T}$ provides a δ-stable embedding of $(\mathcal{S}, \{0\})$. Then for any $s \in \mathcal{S}$, we can detect s with error rate

$$P_D(\alpha) \leq Q\left(Q^{-1}(\alpha) - \sqrt{1+\delta}\sqrt{\frac{M}{N}}\sqrt{\text{SNR}}\right) \tag{16}$$

and

$$P_D(\alpha) \geq Q\left(Q^{-1}(\alpha) - \sqrt{1-\delta}\sqrt{\frac{M}{N}}\sqrt{\text{SNR}}\right). \tag{17}$$

Proof: By our assumption that $\sqrt{N/M}P_{\Phi^T}$ provides a δ-stable embedding of $(\mathcal{S},\{0\})$, we know from (5) that

$$\sqrt{1-\delta}\,\|s\|_2 \le \sqrt{\frac{N}{M}}\,\|P_{\Phi^T}s\|_2 \le \sqrt{1+\delta}\,\|s\|_2\,. \tag{18}$$

Combining (18) with (14) and recalling the definition of the SNR from (15), the result follows. ■

For certain randomized measurement systems, one can anticipate that $\sqrt{N/M}P_{\Phi^T}$ will provide a δ-stable embedding of $(\mathcal{S},\{0\})$. As one example, if Φ has orthonormal rows spanning a random subspace (i.e., it represents a *random orthogonal projection*), then $\Phi\Phi^T = I$, and so $P_{\Phi^T} = \Phi^T\Phi$. It follows that $\|P_{\Phi^T}s\|_2 = \left\|\Phi^T\Phi s\right\|_2 = \|\Phi s\|_2$, and for random orthogonal projections, it is known [27] that $\|P_{\Phi^T}s\|_2 = \|\Phi s\|_2$ satisfies

$$(1-\delta)\frac{M}{N}\,\|s\|_2^2 \le \|P_{\Phi^T}s\|_2^2 \le (1+\delta)\frac{M}{N}\,\|s\|_2^2 \tag{19}$$

with probability at least $1-2e^{-cM\delta^2}$. This statement is analogous to (8) but rescaled to account for the unit-norm rows of Φ. As a second example, if Φ is populated with i.i.d. zero-mean Gaussian entries (of any fixed variance), then the orientation of the row space of Φ has random uniform distribution. Thus, $\|P_{\Phi^T}s\|_2$ for a Gaussian Φ has the same distribution as $\|P_{\Phi^T}s\|_2$ for a random orthogonal projection. It follows that Gaussian Φ also satisfy (19) with probability at least $1-2e^{-cM\delta^2}$.

The similarity between (19) and (8) immediately implies that we can generalize Lemmas 1, 2, and 3 to establish δ-stable embedding results for orthogonal projection matrices P_{Φ^T}. It follows that, when Φ is a Gaussian matrix [with entries satisfying (6)] or a random orthogonal projection (multiplied by $\sqrt{N/M}$), the number of measurements required to establish a δ-stable embedding for $\sqrt{N/M}P_{\Phi^T}$ on a particular signal family $\mathcal{S}$ is equivalent to the number of measurements required to establish a δ-stable embedding for Φ on $\mathcal{S}$.

Theorem 2 tells us in a precise way how much information we lose by using random projections rather than the signal samples themselves, not in terms of our ability to recover the signal as is typically addressed in CS, but in terms of our ability to solve a detection problem. Specifically, for typical values of δ

$$P_D(\alpha) \approx Q\left(Q^{-1}(\alpha) - \sqrt{\frac{M}{N}}\sqrt{\text{SNR}}\right) \tag{20}$$

which increases the miss probability by an amount determined by the SNR and the ratio M/N.

In order to more clearly illustrate the behavior of $P_D(\alpha)$ as a function of M, we also establish the following corollary of Theorem 2.

Corollary 1: Suppose that $\sqrt{N/M}\cdot P_{\Phi^T}$ provides a δ-stable embedding of $(\mathcal{S},\{0\})$. Then for any $s \in \mathcal{S}$, we can detect s with success rate

$$P_D(\alpha) \ge 1 - C_2 e^{-C_1 M/N} \tag{21}$$

where C_1 and C_2 are absolute constants depending only on α, δ, and the SNR.

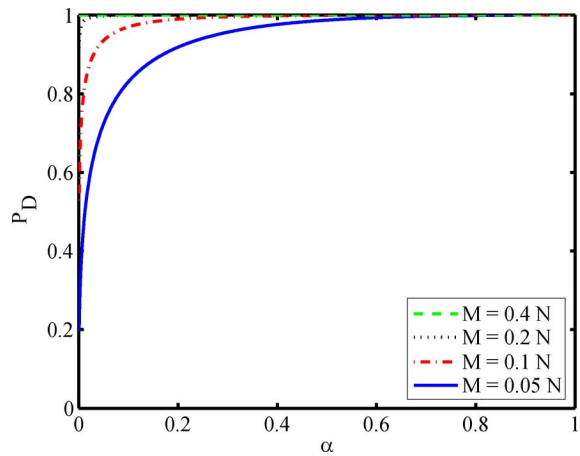

Fig. 3. Effect of M on $P_D(\alpha)$ predicted by (20) (SNR = 20 dB).

Proof: We begin with the following bound from (13.48) of [44]

$$Q(z) \le \frac{e^{-z^2/2}}{2} \tag{22}$$

which allows us to bound P_D as follows. Let $C_1 = (1-\delta)\text{SNR}/2$. Then

$$\begin{aligned} P_D(\alpha) &\ge Q\left(Q^{-1}(\alpha) - \sqrt{\frac{2C_1M}{N}}\right) \\ &= 1 - Q\left(\sqrt{\frac{2C_1M}{N}} - Q^{-1}(\alpha)\right) \\ &\ge 1 - \frac{1}{2}e^{-C_1M/N - \sqrt{2C_1M/N}Q^{-1}(\alpha) + (Q^{-1}(\alpha))^2/2} \\ &\ge 1 - \frac{1}{2}e^{-C_1M/N - \sqrt{2C_1}Q^{-1}(\alpha) + (Q^{-1}(\alpha))^2/2}. \end{aligned}$$

Thus, if we let

$$C_2 = \frac{1}{2}e^{-Q^{-1}(\alpha)\left(Q^{-1}(\alpha)/2 - \sqrt{2C_1}\right)} \tag{23}$$

we obtain the desired result. ■

Thus, for a fixed SNR and signal length, the detection probability approaches 1 exponentially fast as we increase the number of measurements.

C. Experiments and Discussion

We first explore how M affects the performance of the compressive detector. As described above, decreasing M does cause a degradation in performance. However, as illustrated in Fig. 3, in certain cases (relatively high SNR; 20 dB in this example) the compressive detector can perform almost as well as the traditional detector with a very small fraction of the number of measurements required by traditional detection. Specifically, in Fig. 3 we illustrate the *receiver operating characteristic* (ROC) curve, i.e., the relationship between P_D and P_F predicted by (20). Observe that as M increases, the ROC curve approaches the upper-left corner, meaning that we can achieve very high detection rates while simultaneously keeping the false alarm rate very low. As M grows we see that we rapidly reach a regime

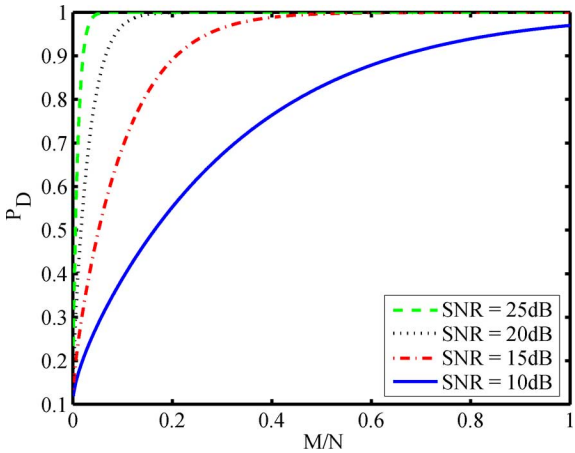

Fig. 4. Effect of M on P_D predicted by (20) at several different SNR levels ($\alpha = 0.1$).

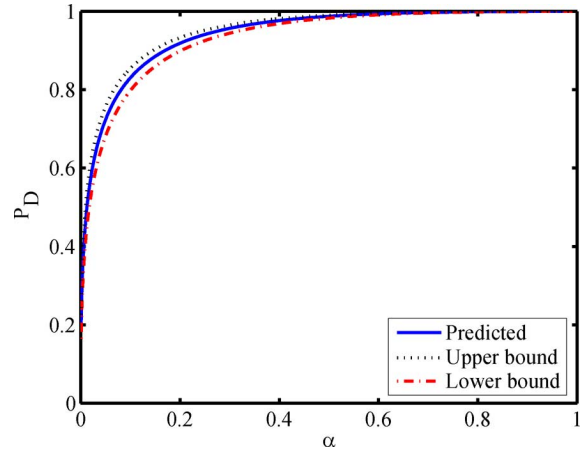

Fig. 5. Concentration of ROC curves for random Φ near the expected ROC curve (SNR $= 20$ dB, $M = 0.05N$, $N = 1000$).

where any additional increase in M yields only marginal improvements in the tradeoff between P_D and P_F.

Furthermore, the exponential increase in the detection probability as we take more measurements is illustrated in Fig. 4, which plots the performance predicted by (20) for a range of SNRs with $\alpha = 0.1$. However, we again note that in practice this rate can be significantly affected by the SNR, which determines the constants in the bound of (21). These results are consistent with those obtained in [17], which also established that P_D should approach 1 exponentially fast as M is increased.

Finally, we close by noting that for any given instance of Φ, its ROC curve may be better or worse than that predicted by (20). However, with high probability it is tightly concentrated around the expected performance curve. Fig. 5 illustrates this for the case where s is fixed; the SNR is 20 dB, Φ has i.i.d. Gaussian entries, $M = 0.05N$, and $N = 1000$. The predicted ROC curve is illustrated along with curves displaying the best and worst ROC curves obtained over 100 independent draws of Φ. We see that our performance is never significantly different from what we expect. Furthermore, we have also observed that these bounds grow significantly tighter as we increase N; so for large problems the difference between the predicted and actual curves will be insignificant. We also note that while some of our theory has been limited to Φ that are Gaussian or random orthogonal projections, we observe that in practice this does not seem to be necessary. We repeated the above experiment for matrices with independent Rademacher entries and observed no significant differences in the results.

IV. Classification With Compressive Measurements

A. Problem Setup and Applications

We can easily generalize the setting of Section III to the problem of binary classification. Specifically, if we wish to distinguish between $\Phi(s_0 + n)$ and $\Phi(s_1 + n)$, then it is equivalent to be able to distinguish $\Phi(s_0+n) - \Phi s_0 = \Phi n$ and $\Phi(s_1 - s_0 + n)$. Thus, the conclusions for the case of binary classification are identical to those discussed in Section III.

More generally suppose that we would like to distinguish between the hypotheses

$$\widetilde{\mathcal{H}_i} : y = \Phi(s_i + n)$$

for $i = 1, 2, \ldots, R$, where each $s_i \in \mathcal{S}$ is one of our known signals and as before, $n \sim \mathcal{N}(0, \sigma^2 I_N)$ is i.i.d. Gaussian noise and Φ is a known $M \times N$ matrix.

It is straightforward to show (see [42] and [43], for example), in the case where each hypothesis is equally likely, that the classifier with minimum probability of error selects the $\widetilde{\mathcal{H}_i}$ that minimizes

$$t_i := (y - \Phi s_i)^T (\Phi\Phi^T)^{-1} (y - \Phi s_i). \tag{24}$$

If the rows of Φ are orthogonal and have equal norm, then this reduces to identifying which Φs_i is closest to y. The $(\Phi\Phi^T)^{-1}$ term arises when the rows of Φ are not orthogonal because the noise is no longer uncorrelated.

As an alternative illustration of the classifier behavior, let us suppose that $y = \Phi x$ for some $x \in \mathbb{R}^N$. Then, starting with (24), we have

$$\begin{aligned} t_i &= (y - \Phi s_i)^T (\Phi\Phi^T)^{-1} (y - \Phi s_i) \\ &= (\Phi x - \Phi s_i)^T (\Phi\Phi^T)^{-1} (\Phi x - \Phi s_i) \\ &= (x - s_i)^T \Phi^T (\Phi\Phi^T)^{-1} \Phi (x - s_i) \\ &= \|P_{\Phi^T} x - P_{\Phi^T} s_i\|_2^2 \end{aligned} \tag{25}$$

where (25) follows from the same argument as (13). Thus, we can equivalently think of the classifier as simply projecting x and each candidate signal s_i onto the row space of Φ and then classifying according to the nearest neighbor in this space.

B. Theory

While in general it is difficult to find analytical expressions for the probability of error even in non-compressive classification settings, we can provide a bound for the performance of the compressive classifier as follows.

Theorem 3: Suppose that $\sqrt{N/M}P_{\Phi^T}$ provides a δ-stable embedding of $(\mathcal{S},\mathcal{S})$, and let $R = |\mathcal{S}|$. Let

$$d = \min_{i,j} \|s_i - s_j\|_2 \tag{26}$$

denote the minimum separation among the s_i. For some $i^* \in \{1,2,\ldots,R\}$, let $y = \Phi(s_{i^*} + n)$, where $n \sim \mathcal{N}(0,\sigma^2 I_N)$ is i.i.d. Gaussian noise. Then with probability at least

$$1 - \left(\frac{R-1}{2}\right) e^{-d^2(1-\delta)M/8\sigma^2 N} \tag{27}$$

the signal can be correctly classified, i.e.,

$$i^* = \arg\min_{i\in\{1,2,\ldots,R\}} t_i. \tag{28}$$

Proof: Let $j \neq i^*$. We will argue that $t_j > t_{i^*}$ with high probability. From (25) we have that

$$\begin{aligned} t_{i^*} &= \|P_{\Phi^T} n\|_2^2 \\ \text{and} \quad t_j &= \|P_{\Phi^T}(s_{i^*} - s_j + n)\|_2^2 \\ &= \|P_{\Phi^T}(s_{i^*} - s_j) + P_{\Phi^T} n\|_2^2 \\ &= \|\tau + P_{\Phi^T} n\|_2^2 \end{aligned}$$

where we have defined $\tau = P_{\Phi^T}(s_{i^*} - s_j)$ to simplify notation. Let us define $P_\tau = \tau\tau^T / \|\tau\|_2^2$ as the orthogonal projection onto the 1-dimensional span of τ, and $P_{\tau^\perp} = (I_N - P_\tau)$. Then we have

$$\begin{aligned} t_{i^*} &= \|P_\tau P_{\Phi^T} n\|_2^2 + \|P_{\tau^\perp} P_{\Phi^T} n\|_2^2 \\ \text{and} \quad t_j &= \|P_\tau(\tau + P_{\Phi^T} n)\|_2^2 + \|P_{\tau^\perp}(\tau + P_{\Phi^T} n)\|_2^2 \\ &= \|\tau + P_\tau P_{\Phi^T} n\|_2^2 + \|P_{\tau^\perp} P_{\Phi^T} n\|_2^2. \end{aligned}$$

Thus, $t_j \leq t_{i^*}$ if and only if

$$\|\tau + P_\tau P_{\Phi^T} n\|_2^2 \leq \|P_\tau P_{\Phi^T} n\|_2^2$$

or equivalently, if

$$\left\| \frac{\tau^T}{\|\tau\|_2}(\tau + P_\tau P_{\Phi^T} n) \right\|_2^2 \leq \left\| \frac{\tau^T}{\|\tau\|_2} P_\tau P_{\Phi^T} n \right\|_2^2$$

or equivalently, if

$$\left| \|\tau\|_2 + \frac{\tau^T}{\|\tau\|_2} P_{\Phi^T} n \right| \leq \left| \frac{\tau^T}{\|\tau\|_2} P_{\Phi^T} n \right|$$

or equivalently, if

$$\frac{\tau^T}{\|\tau\|_2} P_{\Phi^T} n \leq -\frac{\|\tau\|_2}{2}.$$

The quantity $\tau^T/\|\tau\|_2 P_{\Phi^T} n$ is a scalar, zero-mean Gaussian random variable with variance

$$\frac{\tau^T}{\|\tau\|_2} P_{\Phi^T}(\sigma^2 I_N) P_{\Phi^T}^T \frac{\tau}{\|\tau\|_2} = \frac{\tau^T P_{\Phi^T} \tau \sigma^2}{\|\tau\|_2^2} = \sigma^2.$$

Because $\sqrt{N/M}P_{\Phi^T}$ provides a δ-stable embedding of $(\mathcal{S},\mathcal{S})$, and by our assumption that $\|s_{i^*} - s_j\|_2 \geq d$, we have that $\|\tau\|_2^2 \geq d^2(1-\delta)M/N$. Thus, using also (22), we have

$$\begin{aligned} P(t_j \leq t_{i^*}) &= P\left(\frac{\tau^T}{\|\tau\|_2} P_{\Phi^T} n \leq -\frac{\|\tau\|_2}{2}\right) \\ &= Q\left(\frac{\|\tau\|_2}{2\sigma}\right) \\ &\leq \frac{1}{2} e^{-\|\tau\|_2^2/8\sigma^2} \\ &\leq \frac{1}{2} e^{-d^2(1-\delta)M/8\sigma^2 N}. \end{aligned}$$

Finally, because t_{i^*} is compared to $R-1$ other candidates, we use a union bound to conclude that (28) holds with probability exceeding that given in (27). ■

We see from the above that, within the M-dimensional measurement subspace (as mapped to by P_{Φ^T}), we will have a compaction of distances between points in $\mathcal{S}$ by a factor of approximately $\sqrt{M/N}$. However, the variance of the additive noise in this subspace is unchanged. In other words, the noise present in the test statistics does not decrease, but the relative sizes of the test statistics do. Hence, just as in detection [see (20)], the probability of error of our classifier will increase upon projection to a lower-dimensional space in a way that depends on the SNR and the number of measurements. However, it is again important to note that in a high-SNR regime, we may be able to successfully distinguish between the different classes with very few measurements.

C. Experiments and Discussion

In Fig. 6, we display experimental results for classification among $R = 3$ test signals of length $N = 1000$. The signals s_1, s_2, and s_3 are drawn according to a Gaussian distribution with mean 0 and variance 1 and then fixed. For each value of M, a single Gaussian Φ is drawn and then P_E is computed by averaging the results over 10^6 realizations of the noise vector n. The error rates are very similar in spirit to those for detection (see Fig. 4). The results agree with Theorem 3, in which we demonstrate that, as was the case for detection, as M increases the probability of error decays expoentially fast. This also agrees with the related results of [16].

V. Estimation With Compressive Measurements

A. Problem Setup and Applications

While many signal processing problems can be reduced to a detection or classification problem, in some cases we cannot reduce our task to selecting among a finite set of hypotheses. Rather, we might be interested in *estimating* some function of the data. In this section we will focus on estimating a *linear* function of the data from compressive measurements.

Suppose that we observe $y = \Phi s$ and wish to estimate $\langle \ell, s \rangle$ from the measurements y, where $\ell \in \mathbb{R}^N$ is a fixed test vector. In the case where Φ is a random matrix, a natural estimator is essentially the same as the compressive detector. Specifically,

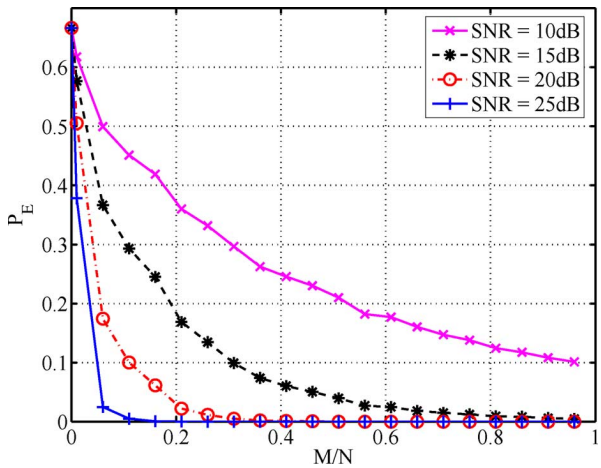

Fig. 6. Effect of M on P_E (the probability of error of a compressive domain classifier) for $R = 3$ signals at several different SNR levels, where SNR $= 10\log_{10}(d^2/\sigma^2)$.

suppose we have a set $\mathcal{L}$ of $|\mathcal{L}|$ linear functions we would like to estimate from y. Example applications include computing the coefficients of a basis or frame representation of the signal, estimating the signal energy in a particular linear subspace, parametric modeling, and so on. One potential estimator for this scenario, which is essentially a simple generalization of the compressive detector in (12), is given by

$$\frac{N}{M} y^T (\Phi\Phi^T)^{-1} \Phi \ell_i \tag{29}$$

for $i = 1, 2, \ldots, |\mathcal{L}|$. While this approach, which we shall refer to as the *orthogonalized* estimator, has certain advantages, it is also enlightening to consider an even simpler estimator, given by

$$\langle y, \Phi \ell_i \rangle . \tag{30}$$

We shall refer to this approach as the *direct* estimator since it eliminates the orthogonalization step by directly correlating the compressive measurements with $\Phi\ell_i$. We will provide a more detailed experimental comparison of these two approaches below, but in the proof of Theorem 4 we focus only on the direct estimator.

B. Theory

We now provide bounds on the performance of our simple estimator.[4] This bound is a generalization of Lemma 2.1 of [22] to the case where $\langle \ell, s \rangle \neq 0$.

Theorem 4: Suppose that $\ell \in \mathcal{L}$ and $s \in \mathcal{S}$ and that Φ is a δ-stable embedding of $(\mathcal{L}, \mathcal{S} \cup -\mathcal{S})$, then

$$|\langle \Phi\ell, \Phi s \rangle - \langle \ell, s \rangle| \leq \delta \|\ell\|_2 \|s\|_2. \tag{31}$$

[4]Note that the same guarantee can be established for the orthogonalized estimator under the assumption that $\sqrt{N/M} P_{\Phi^T}$ is a δ-stable embedding of $(\mathcal{L}, \mathcal{S} \cup -\mathcal{S})$.

Proof: We first assume that $\|\ell\|_2 = \|s\|_2 = 1$. Since

$$\|\ell \pm s\|_2^2 = \|\ell\|_2^2 + \|s\|_2^2 \pm 2\langle \ell, s \rangle = 2 \pm 2\langle \ell, s \rangle$$

and since Φ is a δ-stable embedding of both $(\mathcal{L}, \mathcal{S})$ and $(\mathcal{L}, -\mathcal{S})$, we have that

$$1 - \delta \leq \frac{\|\Phi\ell \pm \Phi s\|_2^2}{2 \pm 2\langle \ell, s \rangle} \leq 1 + \delta.$$

From the parallelogram identity we obtain

$$\begin{aligned} \langle \Phi\ell, \Phi s \rangle &= \frac{1}{4} \left| \|\Phi\ell + \Phi s\|_2^2 - \|\Phi\ell - \Phi s\|_2^2 \right| \\ &\leq \frac{(1 + \langle \ell, s \rangle)(1 + \delta) - (1 - \langle \ell, s \rangle)(1 - \delta)}{2} \\ &= \langle \ell, s \rangle + \delta. \end{aligned}$$

Similarly, one can show that $\langle \Phi\ell, \Phi s \rangle \geq \langle \ell, s \rangle - \delta$. Thus,

$$-\delta \leq \langle \Phi\ell, \Phi s \rangle - \langle \ell, s \rangle \leq \delta.$$

From the bilinearity of the inner product the result follows for ℓ, s with arbitrary norm. ■

One way of interpreting our result is that the angle between two vectors can be estimated accurately; this is formalized as follows.

Corollary 2: Suppose that $\ell \in \mathcal{L}$ and $s \in \mathcal{S}$ and that Φ is a δ-stable embedding of $(\mathcal{L} \cup \{0\}, \mathcal{S} \cup -\mathcal{S} \cup \{0\})$. Then

$$|\cos \measuredangle(\Phi\ell, \Phi s) - \cos \measuredangle(\ell, s)| \leq 2\delta$$

where $\measuredangle(\cdot, \cdot)$ denotes the angle between two vectors.

Proof: Using the standard relationship between inner products and angles, we have

$$\cos \measuredangle(\ell, s) = \frac{\langle \ell, s \rangle}{\|\ell\|_2 \|s\|_2}$$

$$\text{and} \quad \cos \measuredangle(\Phi\ell, \Phi s) = \frac{\langle \Phi\ell, \Phi s \rangle}{\|\Phi\ell\|_2 \|\Phi s\|_2}.$$

Thus, from (31) we have

$$\left| \frac{\langle \Phi\ell, \Phi s \rangle}{\|\ell\|_2 \|s\|_2} - \cos \measuredangle(\ell, s) \right| \leq \delta. \tag{32}$$

Now, using (5), we can show that

$$\frac{(1 - \delta)}{\|\Phi\ell\|_2 \|\Phi s\|_2} \leq \frac{1}{\|\ell\|_2 \|s\|_2} \leq \frac{(1 + \delta)}{\|\Phi\ell\|_2 \|\Phi s\|_2}$$

from which we infer that

$$\left| \frac{\langle \Phi\ell, \Phi s \rangle}{\|\ell\|_2 \|s\|_2} - \frac{\langle \Phi\ell, \Phi s \rangle}{\|\Phi\ell\|_2 \|\Phi s\|_2} \right| \leq \delta \frac{\langle \Phi\ell, \Phi s \rangle}{\|\Phi\ell\|_2 \|\Phi s\|_2} \leq \delta. \tag{33}$$

Therefore, combining (32) and (33) using the triangle inequality, the desired result follows. ■

While Theorem 4 suggests that the absolute error in estimating $\langle \ell, s \rangle$ must scale with $\|\ell\|_2 \|s\|_2$, this is probably the best we can expect. If the $\|\ell\|_2 \|s\|_2$ terms were omitted on the right hand side of (31), then one could estimate $\langle \ell, s \rangle$ with arbitrary

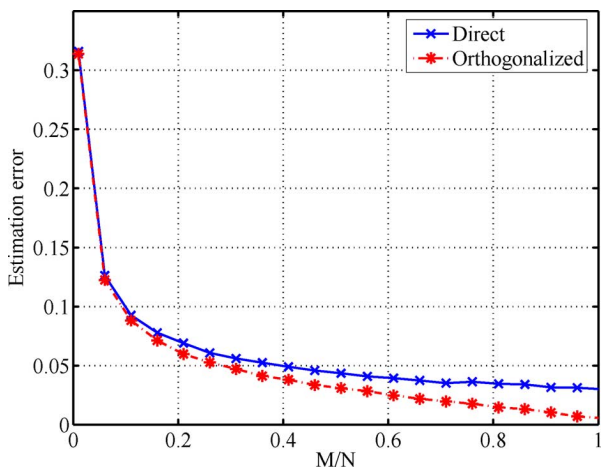

Fig. 7. Average error in the estimate of the mean of a fixed signal s.

accuracy using the following strategy: 1) choose a large positive constant C_{big}; 2) estimate the inner product $\langle C_{\text{big}}\ell, C_{\text{big}}s\rangle$, obtaining an accuracy δ; and then 3) divide the estimate by C_{big}^2 to estimate $\langle \ell, s\rangle$ with accuracy δ/C_{big}^2. Similarly, it is not possible to replace the right hand side of (31) with an expression proportional merely to $\langle \ell, s\rangle$, as this would imply that $\langle \Phi\ell, \Phi s\rangle = \langle \ell, s\rangle$ exactly when $\langle \ell, s\rangle = 0$, and unfortunately this is not the case. (Were this possible, one could exploit this fact to immediately identify the non-zero locations in a sparse signal by letting $\ell_i = e_i$, the ith canonical basis vector, for $i = 1, 2, \ldots, N$.)

C. Experiments and Discussion

In Fig. 7, we display the average estimation error for the orthogonalized and direct estimators, i.e., $\left|(N/M)s^T\Phi^T(\Phi\Phi^T)^{-1}\Phi\ell - \langle \ell, s\rangle\right| / \|s\|_2\|\ell\|_2$ and $\left|\langle \Phi\ell, \Phi s\rangle - \langle \ell, s\rangle\right| / \|s\|_2\|\ell\|_2$ respectively. The signal s is a length $N = 1000$ vector with entries distributed according to a Gaussian distribution with mean 1 and unit variance. We choose $\ell = [1/N\ 1/N \cdots 1/N]^T$ to compute the mean of s. The result displayed is the mean error averaged over 10^4 different draws of Gaussian Φ with s fixed. Note that we obtain nearly identical results for other candidate ℓ, including ℓ both highly correlated with s and ℓ nearly orthogonal to s. In all cases, as M increases, the error decays because the random matrices Φ become δ-stable embeddings of $\{s\}$ for smaller values of δ. Note that for small values of M, there is very little difference between the orthogonalized and direct estimators. The orthogonalized estimator only provides notable improvement when M is large, in which case the computational difference is significant. In this case one must weigh the relative importance of speed versus accuracy in order to judge which approach is best, so the proper choice will ultimately be dependent on the application.

In the case where $|\mathcal{L}| = 1$, Theorem 4 is a deterministic version of Theorem 4.5 of [45] and Lemma 3.1 of [46], which both show that for certain random constructions of Φ, with probability at least $1 - \beta$

$$|\langle \Phi\ell, \Phi s\rangle - \langle \ell, s\rangle| \leq \delta \|\ell\|_2 \|s\|_2 . \tag{34}$$

In [45] $\beta = 2\delta^2/M$ while in [46] more sophisticated methods are used to achieve a bound on β of the form $\beta \leq 2e^{-cM\delta^2}$ as in (8). Our result extends these results to a wider class of random matrices. Furthermore, our approach generalizes naturally to simultaneously estimating multiple linear functions of the data.

Specifically, it is straightforward to extend our analysis beyond the estimation of scalar-valued linear functions to more general linear operators. Any finite-dimensional linear operator on a signal $x \in \mathbb{R}^N$ can be represented as a matrix multiplication Lx, where L has size $Z \times N$ for some Z. Decomposing L in terms of its rows, this computation can be expressed as

$$Lx = \begin{bmatrix} \ell_1^T \\ \ell_2^T \\ \vdots \\ \ell_Z^T \end{bmatrix} x = \begin{bmatrix} \langle \ell_1, x\rangle \\ \langle \ell_2, x\rangle \\ \vdots \\ \langle \ell_Z, x\rangle \end{bmatrix}.$$

From this point, the bound (31) can be applied to each component of the resulting vector. It is also interesting to note that by setting $L = I$, we can observe that

$$\|\Phi^T\Phi x - x\|_\infty \leq \delta\|x\|_2.$$

This could be used to establish deterministic bounds on the performance of the thresholding signal recovery algorithm described in [46], which simply thresholds $\Phi^T y$ to keep only the K largest elements.

We can also consider more general estimation problems in the context of parameterized signal models. Suppose, for instance, that a K-dimensional parameter θ controls the generation of a signal $s = s_\theta$, and we denote by Θ the K-dimensional space to which the parameter θ belongs. Common examples of such articulations include the angle and orientation of an edge in an image ($K = 2$), or the start and end frequencies and times of a linear radar chirp ($K = 4$). In cases such as these, the set of possible signals of interest

$$\mathcal{M} := \{s_\theta : \theta \in \Theta\} \subset \mathbb{R}^N$$

forms a nonlinear K-dimensional manifold. The actual position of a given signal on the manifold reveals the value of the underlying parameter θ. It is now understood [47] that, because random projections can provide δ-stable embeddings of nonlinear manifolds using $M = O\left(K \log N/\delta^2\right)$ measurements [31], the task of estimating position on a manifold can also be addressed in the compressive domain. Recovery bounds on $\|\theta - \widehat{\theta}\|_2$ akin to (4) can be established; see [47] for more details.

VI. Filtering With Compressive Measurements

A. Problem Setup and Applications

In practice, it is often the case that the signal we wish to acquire is contaminated with interference. The universal nature of compressive measurements, while often advantageous, can also increase our susceptibility to interference and significantly affect the performance of algorithms such as those described in Sections III–V. It is therefore desirable to remove unwanted

signal components from the compressive measurements before they are processed further.

More formally, suppose that the signal $x \in \mathbb{R}^N$ consists of two components

$$x = x_S + x_I$$

where x_S represents the *signal of interest* and x_I represents an unwanted signal that we would like to reject. We refer to x_I as *interference* in the remainder of this section, although it might be the signal of interest for a different system module. Supposing we acquire measurements of both components simultaneously

$$y = \Phi(x_S + x_I) \tag{35}$$

our goal is to remove the contribution of x_I from the measurements y while preserving the information about x_S. In this section, we will assume that $x_S \in \mathcal{S}_S$ and that $x_I \in \mathcal{S}_I$. In our discussion, we will further assume that Φ is a δ-stable embedding of $(\widetilde{\mathcal{S}}_S, \mathcal{S}_I)$, where $\widetilde{\mathcal{S}}_S$ is a set with a simple relationship to $\mathcal{S}_S$ and $\mathcal{S}_I$.

While one could consider more general interference models, we restrict our attention to the case where either the interfering signal or the signal of interest lives in a known subspace. For example, suppose we have obtained measurements of a radio signal that has been corrupted by narrow band interference such as a TV or radio station operating at a known carrier frequency. In this case, we can project the compressive measurements into a subspace orthogonal to the interference, and hence eliminate the contribution of the interference to the measurements. We further demonstrate that provided that the signal of interest is orthogonal to the set of possible interference signals, the projection operator maintains a stable embedding for the set of signals of interest. Thus, the projected measurements retain sufficient information to enable the use of efficient compressive-domain algorithms for further processing.

B. Theory

We first consider the case where $\mathcal{S}_I$ is a K_I-dimensional subspace, and we place no restrictions on the set $\mathcal{S}_S$. We will later see that by symmetry the methods we develop for this case will have implications for the setting where $\mathcal{S}_S$ is a K_S-dimensional subspace and where $\mathcal{S}_I$ is a more general set.

We filter out the interference by constructing a linear operator P that operates on the measurements y. The design of P is based solely on the measurement matrix Φ and knowledge of the subspace $\mathcal{S}_I$. Our goal is to construct a P that maps Φx_I to zero for any $x_I \in \mathcal{S}_I$. To simplify notation, we assume that Ψ_I is an $N \times K_I$ matrix whose columns form an orthonormal basis for the K_I-dimensional subspace $\mathcal{S}_I$, and we define the $M \times K_I$ matrix $\Omega = \Phi\Psi_I$. Letting $\Omega^\dagger := (\Omega^T\Omega)^{-1}\Omega^T$ denote the Moore–Penrose pseudoinverse of Ω, we define

$$P_\Omega = \Omega\Omega^\dagger \tag{36}$$

and

$$P_{\Omega^\perp} = I - P_\Omega = I - \Omega\Omega^\dagger. \tag{37}$$

The resulting $P_{\Omega^\perp}$ is our desired operator P: it is an orthogonal projection operator onto the orthogonal complement of $\mathcal{R}(\Omega)$, and its nullspace equals $\mathcal{R}(\Omega)$.

Using Theorem 4, we now show that the fact that Φ is a stable embedding allows us to argue that $P_{\Omega^\perp}$ preserves the structure of $\widetilde{\mathcal{S}}_S = P_{\mathcal{S}_I^\perp}\mathcal{S}_S$ (where $\mathcal{S}_I^\perp$ denotes the orthogonal complement of $\mathcal{S}_I$ and $P_{\mathcal{S}_I^\perp}$ denotes the orthogonal projection onto $\mathcal{S}_I^\perp$), while simultaneously cancelling out signals from $\mathcal{S}_I$.[5] Additionally, P_Ω preserves the structure in $\mathcal{S}_I$ while nearly cancelling out signals from $\widetilde{\mathcal{S}}_S$.

Theorem 5: Suppose that Φ is a δ-stable embedding of $(\widetilde{\mathcal{S}}_S \cup \{0\}, \mathcal{S}_I)$, where $\mathcal{S}_I$ is a K_I-dimensional subspace of $\mathbb{R}^N$ with orthonormal basis Ψ_I. Set $\Omega = \Phi\Psi_I$ and define P_Ω and $P_{\Omega^\perp}$ as in (36) *and* (37). *For any $x \in \mathcal{S}_S \oplus \mathcal{S}_I$ we can write $x = \widetilde{x}_S + \widetilde{x}_I$, where $\widetilde{x}_S \in \widetilde{\mathcal{S}}_S$ and $\widetilde{x}_I \in \mathcal{S}_I$. Then*

$$P_{\Omega^\perp}\Phi x = P_{\Omega^\perp}\Phi\widetilde{x}_S \tag{38}$$

and

$$P_\Omega\Phi x = \Phi\widetilde{x}_I + P_\Omega\Phi\widetilde{x}_S. \tag{39}$$

Furthermore,

$$1 - \frac{\delta}{1-\delta} \le \frac{\|P_{\Omega^\perp}\Phi\widetilde{x}_S\|_2^2}{\|\widetilde{x}_S\|_2^2} \le 1 + \delta \tag{40}$$

and

$$\frac{\|P_\Omega\Phi\widetilde{x}_S\|_2^2}{\|\widetilde{x}_S\|_2^2} \le \delta^2\frac{1+\delta}{(1-\delta)^2}. \tag{41}$$

Proof: We begin by observing that since $\widetilde{\mathcal{S}}_S$ and $\mathcal{S}_I$ are orthogonal, the decomposition $x = \widetilde{x}_S + \widetilde{x}_I$ is unique. Furthermore, since $\widetilde{x}_I \in \mathcal{S}_I$, we have that $\Phi\widetilde{x}_I \in \mathcal{R}(\Omega)$ and hence by the design of $P_{\Omega^\perp}$, $P_{\Omega^\perp}\Phi\widetilde{x}_I = 0$ and $P_\Omega\Phi\widetilde{x}_I = \Phi\widetilde{x}_I$, which establishes (38) and (39).

In order to establish (40) and (41), we decompose $\Phi\widetilde{x}_S$ as $\Phi\widetilde{x}_S = P_\Omega\Phi\widetilde{x}_S + P_{\Omega^\perp}\Phi\widetilde{x}_S$. Since P_Ω is an orthogonal projection we can write

$$\|\Phi\widetilde{x}_S\|_2^2 = \|P_\Omega\Phi\widetilde{x}_S\|_2^2 + \|P_{\Omega^\perp}\Phi\widetilde{x}_S\|_2^2. \tag{42}$$

Furthermore, note that $P_\Omega^T = P_\Omega$ and $P_\Omega^2 = P_\Omega$, so that

$$\langle P_\Omega\Phi\widetilde{x}_S, \Phi\widetilde{x}_S\rangle = \|P_\Omega\Phi\widetilde{x}_S\|_2^2. \tag{43}$$

Since P_Ω is a projection onto $\mathcal{R}(\Omega)$ there exists a $z \in \mathcal{S}_I$ such that $P_\Omega\Phi\widetilde{x}_S = \Phi z$. Since $\widetilde{x}_S \in \widetilde{\mathcal{S}}_S$, we have that $\langle\widetilde{x}_S, z\rangle = 0$,

[5]Note that we do not claim that $P_{\Omega^\perp}$ preserves the structure of $\mathcal{S}_S$, but rather the structure of $\widetilde{\mathcal{S}}_S$. This is because we do not restrict $\mathcal{S}_S$ to be orthogonal to the subspace $\mathcal{S}_I$ which we cancel. Clearly, we cannot preserve the structure of the component of $\mathcal{S}_S$ that lies within $\mathcal{S}_I$ while simultaneously eliminating interference from $\mathcal{S}_I$.

and since $\mathcal{S}_I$ is a subspace, $\mathcal{S}_I = \mathcal{S}_I \cup -\mathcal{S}_I$, and so we may apply Theorem 4 to obtain

$$|\langle P_\Omega \Phi \widetilde{x}_S, \Phi \widetilde{x}_S \rangle| = |\langle \Phi z, \Phi \widetilde{x}_S \rangle| \leq \delta \|z\|_2 \|\widetilde{x}_S\|_2.$$

Since $0 \in \mathcal{S}_I$ and Φ is a δ-stable embedding of $(\widetilde{\mathcal{S}}_S \cup \{0\}, \mathcal{S}_I)$, we have that

$$\|z\|_2 \|\widetilde{x}_S\|_2 \leq \frac{\|\Phi z\|_2 \|\Phi \widetilde{x}_S\|_2}{1-\delta}.$$

Recalling that $\Phi z = P_\Omega \Phi \widetilde{x}_S$, we obtain

$$\frac{|\langle P_\Omega \Phi \widetilde{x}_S, \Phi \widetilde{x}_S \rangle|}{\|P_\Omega \Phi \widetilde{x}_S\|_2 \|\Phi \widetilde{x}_S\|_2} \leq \frac{\delta}{1-\delta}.$$

Combining this with (43), we obtain

$$\|P_\Omega \Phi \widetilde{x}_S\|_2 \leq \frac{\delta}{1-\delta} \|\Phi \widetilde{x}_S\|_2.$$

Since $\widetilde{x}_S \in \widetilde{\mathcal{S}}_S$, $\|\Phi \widetilde{x}_S\|_2 \leq \sqrt{1+\delta} \|\widetilde{x}_S\|_2$, and thus we obtain (41). Since we trivially have that $\|P_\Omega \Phi \widetilde{x}_S\|_2 \geq 0$, we can combine this with (42) to obtain

$$\left(1 - \left(\frac{\delta}{1-\delta}\right)^2\right) \|\Phi \widetilde{x}_S\|_2^2 \leq \|P_{\Omega^\perp} \Phi \widetilde{x}_S\|_2^2 \leq \|\Phi \widetilde{x}_S\|_2^2.$$

Again, since $\widetilde{x}_S \in \widetilde{\mathcal{S}}_S$, we have that

$$\left(1 - \left(\frac{\delta}{1-\delta}\right)^2\right)(1-\delta) \leq \frac{\|P_{\Omega^\perp} \Phi \widetilde{x}_S\|_2^2}{\|\widetilde{x}_S\|_2^2} \leq 1+\delta$$

which simplifies to yield (40). ■

Corollary 3: Suppose that Φ is a δ-stable embedding of $(\widetilde{\mathcal{S}}_S \cup \{0\}, \mathcal{S}_I)$, where $\mathcal{S}_I$ is a K_I-dimensional subspace of $\mathbb{R}^N$ with orthonormal basis Ψ_I. Set $\Omega = \Phi\Psi_I$ and define P_Ω and $P_{\Omega^\perp}$ as in (36) and (37). Then $P_{\Omega^\perp}\Phi$ is a $\delta/(1-\delta)$-stable embedding of $(\widetilde{\mathcal{S}}_S, \{0\})$ and $P_\Omega \Phi$ is a δ-stable embedding of $(\mathcal{S}_I, \{0\})$.

Proof: This follows from Theorem 5 by picking $x \in \widetilde{\mathcal{S}}_S$, in which case $x = \widetilde{x}_S$, or picking $x \in \mathcal{S}_I$, in which case $x = \widetilde{x}_I$. ■

Theorem 5 and Corollary 3 have a number of practical benefits. For example, if we are interested in solving an inference problem based only on the signal x_S, then we can use P_Ω or $P_{\Omega^\perp}$ to filter out the interference and then apply the compressive domain inference techniques developed above. The performance of these techniques will be significantly improved by eliminating the interference due to x_I. Furthermore, this result also has implications for the problem of signal recovery, as demonstrated by the following corollary.

Corollary 4: Suppose that Ψ is an orthonormal basis for $\mathbb{R}^N$ and that Φ is a δ-stable embedding of $(\Psi(\Sigma_{2K_S}), \mathcal{R}(\Psi_I))$, where Ψ_I is an $N \times K_I$ submatrix of Ψ. Set $\Omega = \Phi\Psi_I$ and define P_Ω and $P_{\Omega^\perp}$ as in (36) and (37). Then $P_{\Omega^\perp}\Phi$ is a $\delta/(1-\delta)$-stable embedding of $(P_{\mathcal{R}(\Psi_I)^\perp}\Psi(\Sigma_{2K_S}), \{0\})$.

Proof: This follows from the observation that $P_{\mathcal{R}(\Psi_I)^\perp}\Psi(\Sigma_{2K_S}) \subset \Psi(\Sigma_{2K_S})$ and then applying Corollary 3. ■

We emphasize that in the above Corollary, $P_{\mathcal{R}(\Psi_I)^\perp}\Psi(\Sigma_{2K_S})$ will simply be the original family of sparse signals but with zeros in positions indexed by Ψ_I. One can easily verify that if $\delta \leq (\sqrt{2}-1)/\sqrt{2}$, then $\delta/(1-\delta) \leq \sqrt{2}-1$, and thus Corollary 4 is sufficient to ensure that the conditions for Theorem 1 are satisfied. We therefore conclude that under a slightly more restrictive bound on the required RIP constant, we can directly recover a sparse signal of interest x_S that is orthogonal to the interfering x_I without actually recovering x_I. Note that in addition to filtering out true interference, this framework is also relevant to the problem of signal recovery when the support is partially known, in which case the known support defines a subspace that can be thought of as interference to be rejected prior to recovering the remaining signal. Thus, our approach provides an alternative method for solving and analyzing the problem of CS recovery with partially known support considered in [48]. Furthermore, this result can also be useful in analyzing iterative recovery algorithms (in which the signal coefficients identified in previous iterations are treated as interference) or in the case where we wish to recover a slowly varying signal as it evolves in time, as in [49].

This *cancel-then-recover* approach to signal recovery has a number of advantages. Observe that if we attempt to first recover x and then cancel x_I, then we require the RIP of order $2(K_S + K_I)$ to ensure that the *recover-then-cancel* approach will be successful. In contrast, filtering out x_I followed by recovery of x_S requires the RIP of order only $2K_S + K_I$. In certain cases (when K_I is significantly larger than K_S), this results in a substantial decrease in the required number of measurements. Furthermore, since all recovery algorithms have computational complexity that is at least linear in the sparsity of the recovered signal, this can also result in substantial computational savings for signal recovery.

C. Experiments and Discussion

In this section, we evaluate the performance of the cancel-then-recover approach suggested by Corollary 4 . Rather than ℓ_1-minimization we use the iterative CoSaMP greedy algorithm [7] since it more naturally naturally lends itself towards a simple modification described below. More specifically, we evaluate three interference cancellation approaches.

1) **Cancel-then-recover:** This is the approach advocated in this paper. We cancel out the contribution of x_I to the measurements y and directly recover x_S using the CoSaMP algorithm.
2) **Modified recovery:** Since we know the support of x_I, rather than cancelling out the contribution from x_I to the measurements, we modify a greedy algorithm such as CoSaMP to exploit the fact that part of the support of x is known in advance. This modification is made simply by forcing CoSaMP to always keep the elements of J in the active set at each iteration. After recovering $\widehat{x}$, we then set $\widehat{x}_n = 0$ for $n \in J$ to filter out the interference.
3) **Recover-then-cancel:** In this approach, we ignore the fact that we know the support of x_I and try to recover the signal x using the standard CoSaMP algorithm, and then set $\widehat{x}_n = 0$ for $n \in J$ as before.

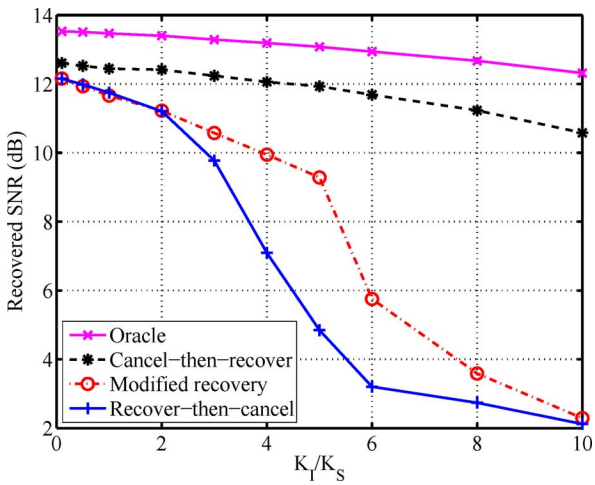

Fig. 8. SNR of x_S recovered using the three different cancellation approaches for different ratios of K_I to K_S compared to the performance of an oracle.

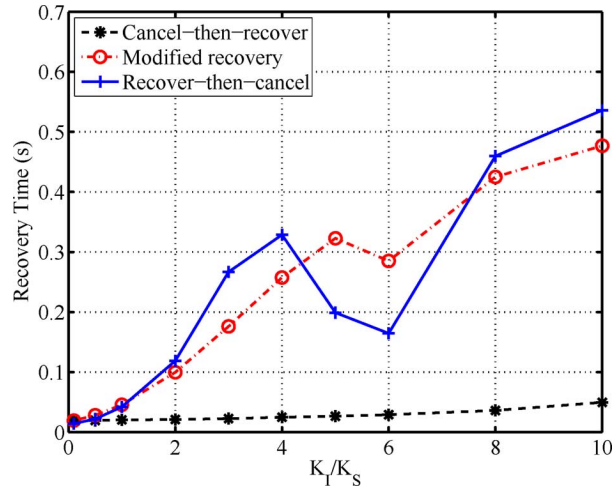

Fig. 9. Recovery time for the three different cancellation approaches for different ratios of K_I to K_S.

In our experiments, we set $N = 1000$, $M = 200$, and $K_S = 10$. We then considered values of K_I from 1 to 100. We choose $\mathcal{S}_S$ and $\mathcal{S}_I$ by selecting random, non-overlapping sets of indices, so in this experiment, $\mathcal{S}_S$ and $\mathcal{S}_I$ are orthogonal (although they need not be in general, since $\widetilde{\mathcal{S}}_S$ will always be orthogonal to $\mathcal{S}_I$). For each value of K_I, we generated 2000 test signals where the coefficients were selected according to a Gaussian distribution and then contaminated with an N-dimensional Gaussian noise vector. For comparison, we also considered an oracle decoder that is given the support of both x_I and x_S and solves the least-squares problem restricted to the known support set.

We considered a range of signal-to-noise ratios (SNRs) and signal-to-interference ratios (SIRs). Fig. 8 shows the results for the case where x_S and x_I are normalized to have equal energy (an SIR of 0 dB) and where the variance of the noise is selected so that the SNR is 15 dB. Our results were consistent for a wide range of SNR and SIR values, and we omit the plots due to space limitations.

Our results show that the cancel-then-recover approach performs significantly better than both of the other methods as K_I grows larger than K_S, in fact, the cancel-then-recover approach performs almost as well as the oracle decoder for the entire range of K_I. We also note that while the modified recovery method did perform slightly better than the recover-then-cancel approach, the improvement is relatively minor.

We observe similar results in Fig. 9 for the recovery time (which includes the cost of computing P in the cancel-then-recover approach). The cancel-then-recover approach performs significantly faster than the other approaches as K_I grows larger than K_S.

We also note that in the case where Φ admits a fast transform-based implementation (as is often the case for the constructions described in Section II-D) the projections P_Ω and $P_{\Omega^\perp}$ can leverage the structure of Φ in order to ease the computational cost of applying P_Ω and $P_{\Omega^\perp}$. For example, Φ may consist of random rows of a discrete Fourier transform or a permuted Hadamard Transform matrix. In such a scenario, there are fast transform-based implementations of Φ and Φ^T. By observing that

$$P_\Omega = \Phi\Psi_I(\Psi_I^T\Phi^T\Phi\Psi_I)^{-1}\Psi_I^T\Phi^T$$

we see that one can use the conjugate gradient method or Richardson iteration to efficiently compute $P_\Omega y$ and by extension $P_{\Omega^\perp} y$ [7].

VII. Conclusion

In this paper, we have taken some first steps towards a theory of compressive signal processing (CSP) by showing that compressive measurements can be effective for a variety of detection, classification, estimation, and filtering problems. We have provided theoretical bounds backed up by experimental results that indicate that in many applications it can be more efficient and accurate to extract information directly from a signal's compressive measurements than first recover the signal and then extract the information. It is important to reemphasize that our techniques are *universal* and *agnostic* to the signal structure and provide *deterministic* guarantees for a wide variety of signal classes.

In the future we hope to provide a more detailed analysis of the classification setting and consider more general models, as well as consider detection, classification, and estimation settings that utilize more specific models, such as sparsity or manifold structure.

References

[1] D. Healy, Analog-to-Information 2005, BAA #05-35 [Online]. Available: http://www.darpa.mil/mto/solicitations/baa05-35/s/index.html

[2] R. Walden, "Analog-to-digital converter survey and analysis," *IEEE J. Sel. Areas Commun.*., vol. 17, no. 4, pp. 539–550, Apr. 1999.

[3] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[4] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[5] J. Tropp and A. Gilbert, "Signal recovery from random information via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[6] D. Needell and R. Vershynin, "Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit," *Found. Comput. Math.*, vol. 9, no. 3, pp. 317–334, Jun. 2009.

[7] D. Needell and J. Tropp, "CoSaMP: Iterative signal recovery from incomplete and inaccurate samples," *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, May 2009.

[8] A. Cohen, W. Dahmen, and R. DeVore, Instance optimal decoding by thresholding in compressed sensing IGPM Rep., Nov. 2008, RWTH Aachen, Germany.

[9] T. Blumensath and M. Davies, "Iterative hard thresholding for compressive sensing," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, Nov. 2009.

[10] J. Treichler, M. Davenport, and R. Baraniuk, "Application of compressive sensing to the design of wideband signal acquisition receivers," in *Proc. U.S./Australia Joint Work. Defense Apps. Signal Process. (DASP)*, Lihue, HI, Sep. 2009.

[11] S. Muthukrishnan, *Data Streams: Algorithms and Applications*. Delft, The Netherlands: Now, 2005.

[12] M. Duarte, M. Davenport, M. Wakin, and R. Baraniuk, "Sparse signal detection from incoherent projections," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Toulouse, France, May 2006, pp. 305–308.

[13] M. Davenport, M. Duarte, M. Wakin, J. Laska, D. Takhar, K. Kelly, and R. Baraniuk, "The smashed filter for compressive classification and target recognition," in *Proc. SPIE Symp. Electron. Imaging: Comput. Imaging*, San Jose, CA, Jan. 2007.

[14] M. Duarte, M. Davenport, M. Wakin, J. Laska, D. Takhar, K. Kelly, and R. Baraniuk, "Multiscale random projections for compressive classification," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, San Antonio, TX, Sep. 2007, pp. 161–164.

[15] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[16] J. Haupt, R. Castro, R. Nowak, G. Fudge, and A. Yeh, "Compressive sampling for signal classification," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, Nov. 2006.

[17] J. Haupt and R. Nowak, "Compressive sampling for signal detection," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Honolulu, HI, Apr. 2007, pp. 1509–1512.

[18] M. Davenport, M. Wakin, and R. Baraniuk, Detection and estimation with compressive measurements Elect. Comput. Eng. Dept., Rice Univ., Houston, TX, 2006, Tech. Rep. TREE0610.

[19] M. Davenport, P. Boufounos, and R. Baraniuk, "Compressive domain interference cancellation," in *Proc. Structure et Parcimonie Pour la Représentation Adaptative de Signaux (SPARS)*, Saint-Malo, France, Apr. 2009.

[20] E. Candès, "Compressive sampling," in *Proc. Int. Congr. Math.*, Madrid, Spain, Aug. 2006.

[21] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[22] E. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus de l'Académie des Sciences, Série I*, vol. 346, no. 9–10, pp. 589–592, May 2008.

[23] V. Buldygin and Y. Kozachenko, *Metric Characterization of Random Variables and Random Processes*. Providence, RI: AMS, 2000.

[24] R. DeVore, G. Petrova, and P. Wojtaszczyk, "Instance-optimality in probability with an ℓ_1 -minimization decoder," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 275–288, Nov. 2009.

[25] M. Davenport, "Concentration of measure and sub-Gaussian distributions," 2009 [Online]. Available: http://cnx.org/content/m32583/latest/

[26] W. Johnson and J. Lindenstrauss, "Extensions of Lipschitz mappings into a Hilbert space," in *Proc. Conf. Modern Anal. Prob.*, New Haven, CT, Jun. 1982.

[27] S. Dasgupta and A. Gupta, An elementary proof of the Johnson-Lindenstrauss lemma Berkeley, CA, 1999, Tech. Rep. TR-99-006.

[28] D. Achlioptas, "Database-friendly random projections," in *Proc. Symp. Principles of Database Systems (PODS)*, Santa Barbara, CA, May 2001.

[29] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Const. Approx.*, vol. 28, no. 3, pp. 253–263, Dec. 2008.

[30] G. Lorentz, M. Golitschek, and Y. Makovoz, *Constructive Approximation: Advanced Problems*. Berlin, Germany: Springer-Verlag, 1996.

[31] R. Baraniuk and M. Wakin, "Random projections of smooth manifolds," *Found. Comput. Math.*, vol. 9, no. 1, pp. 51–77, Feb. 2009.

[32] P. Indyk and A. Naor, "Nearest-neighbor-preserving embeddings," *ACM Trans. Algorithms*, vol. 3, no. 3, Aug. 2007.

[33] P. Agarwal, S. Har-Peled, and H. Yu, "Embeddings of surfaces, curves, and moving points in euclidean space," in *Proc. Symp. Comput. Geometry*, Gyeongju, South Korea, Jun. 2007.

[34] S. Dasgupta and Y. Freund, "Random projection trees and low dimensional manifolds," in *Proc. ACM Symp. Theory Comput. (STOC)*, Victoria, BC, Canada, May 2008.

[35] J. Tropp, J. Laska, M. Duarte, J. Romberg, and R. Baraniuk, "Beyond Nyquist: Efficient sampling of sparse, bandlimited signals," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 505–519, Jan. 2010.

[36] 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, Signal Process. (ICASSP)*, Toulouse, France, May 2006, pp. 872–875.

[37] W. Bajwa, J. Haupt, G. Raz, S. Wright, and R. Nowak, "Toeplitz-structured compressed sensing matrices," in *Proc. IEEE Workshop Statist. Signal Process. (SSP)*, Madison, WI, Aug. 2007, pp. 294–298.

[38] J. Romberg, "Compressive sensing by random convolution," *SIAM J. Imaging Sci.*, vol. 2, no. 4, pp. 1098–1128, Nov. 2009.

[39] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 83–91, Mar. 2008.

[40] R. Robucci, L. Chiu, J. Gray, J. Romberg, P. Hasler, and D. Anderson, "Compressive sensing on a CMOS separable transform image sensor," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Las Vegas, NV, Apr. 2008, pp. 5125–5128.

[41] R. Marcia, Z. Harmany, and R. Willett, "Compressive coded aperture imaging," in *Proc. SPIE Symp. Electron. Imaging: Comput. Imaging*, San Jose, CA, Jan. 2009.

[42] S. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1998, vol. 2.

[43] L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. London, U.K.: Addison-Wesley, 1991.

[44] N. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*. New York: Wiley, 1994, vol. 1.

[45] N. Alon, P. Gibbons, Y. Matias, and M. Szegedy, "Tracking join and self-join sizes in limited storage," in *Proc. Symp. Principles Database Syst. (PODS)*, Philadelphia, PA, May 1999.

[46] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 2210–2219, May 2008.

[47] M. Wakin, *Manifold-Based Signal Recovery and Parameter Estimation From Compressive Measurements*, 2008 [Online]. Available: http://arxiv.org/abs/1002.1247, preprint

[48] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Seoul, Korea, Jun. 2009, pp. 488–492.

[49] N. Vaswani, "Analyzing least squares and Kalman filtered compressed sensing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 3013–3016.

Mark A. Davenport (S'04) received the B.S.E.E. degree in electrical and computer engineering and the B.A. degree in managerial studies in 2004 and the M.S. degree in electrical and computer engineering in 2007, all from Rice University, Houston, TX. He is currently pursuing the Ph.D. degree in the Department of Electrical and Computer Engineering, Rice University.

His research interests include compressive sensing, nonlinear approximation, and the application of low-dimensional signal models to a variety of problems in signal processing and machine learning. He is also cofounder and an editor of *Rejecta Mathematica*.

Mr. Davenport shared the Hershel M. Rich Invention Award from Rice in 2007 for his work on the single-pixel camera and compressive sensing.

Petros T. Boufounos (S'02–M'06) received the S.B. degree in economics, the S.B. and M.Eng. degrees in electrical engineering and computer science, and the Sc.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge in 2000, 2002 and 2006, respectively.

Since January 2009, he has been with Mitsubishi Electric Research Laboratories (MERL), Cambridge. He is also a Visiting Scholar in the Department of Electrical and Computer Engineering, Rice University, Houston, TX.

Between September 2006 and December 2008, he was a Postdoctoral Associate with the Digital Signal Processing Group, Rice University, doing research in compressive sensing. In addition to compressive sensing, his immediate research interests include signal processing, data representations, frame theory, and machine learning applied to signal processing. He is also looking into how compressed sensing interacts with other fields that use sensing extensively, such as robotics and mechatronics.

Dr. Boufounos received the Ernst A. Guillemin Master Thesis Award for his work on DNA sequencing and the Harold E. Hazen Award for Teaching Excellence, both from the MIT Electrical Engineering And Computer Science Department. He has also been an MIT Presidential Fellow. He is a member of Sigma Xi, Eta Kappa Nu, and Phi Beta Kappa.

Michael B. Wakin (M'07) received the B.S. degree in electrical engineering and the B.A. degree in mathematics in 2000 (*summa cum laude*), the M.S. degree in electrical engineering in 2002, and the Ph.D. degree in electrical engineering in 2007, all from Rice University, Houston, TX.

He was an NSF Mathematical Sciences Postdoctoral Research Fellow at the California Institute of Technology, Pasadena, from 2006 to 2007 and an Assistant Professor at the University of Michigan, Ann Arbor, from 2007 to 2008. He is now an Assistant Professor in the Division of Engineering at the Colorado School of Mines, Golden. His research interests include sparse, geometric, and manifold-based models for signal and image processing, approximation, compression, compressive sensing, and dimensionality reduction.

Dr. Wakin shared the Hershel M. Rich Invention Award from Rice University in 2007 for the design of a single-pixel camera based on compressive sensing, and in 2008, Dr. Wakin received the DARPA Young Faculty Award for his research in compressive multi-signal processing for environments such as sensor and camera networks.

Richard G. Baraniuk (M'93–SM'98–F'01) received the B.Sc. degree from the University of Manitoba, Winnipeg, MB, Canada, in 1987, the M.Sc. degree from the University of Wisconsin-Madison, in 1988, and the Ph.D. degree from the University of Illinois at Urbana-Champaign, in 1992, all in electrical engineering.

After spending 1992–1993 with the Signal Processing Laboratory of Ecole Normale Supérieure, Lyon, France, he joined Rice University, Houston, TX, where he is currently the Victor E. Cameron Professor of Electrical and Computer Engineering. He spent sabbaticals at Ecole Nationale Supérieure de Télécommunications in Paris in 2001 and Ecole Fédérale Polytechnique de Lausanne in Switzerland in 2002. His research interests lie in the area of signal and image processing. In 1999, he founded Connexions (cnx.org), a nonprofit publishing project that invites authors, educators, and learners worldwide to "create, rip, mix, and burn" free textbooks, courses, and learning materials from a global open-access repository.

Prof. Baraniuk has been a Guest Editor of several special issues of the *IEEE Signal Processing Magazine*, IEEE JOURNAL OF SPECIAL TOPICS IN SIGNAL PROCESSING, and the *Proceedings of the IEEE* and has served as technical program chair or on the technical program committee for several IEEE workshops and conferences. He received a NATO postdoctoral fellowship from NSERC in 1992, the National Young Investigator award from the National Science Foundation in 1994, a Young Investigator Award from the Office of Naval Research in 1995, the Rosenbaum Fellowship from the Isaac Newton Institute of Cambridge University in 1998, the C. Holmes MacDonald National Outstanding Teaching Award from Eta Kappa Nu in 1999, the Charles Duncan Junior Faculty Achievement Award from Rice in 2000, the University of Illinois ECE Young Alumni Achievement Award in 2000, the George R. Brown Award for Superior Teaching at Rice in 2001, 2003, and 2006, the Hershel M. Rich Invention Award from Rice in 2007, the Wavelet Pioneer Award from SPIE in 2008, and the Internet Pioneer Award from the Berkman Center for Internet and Society at Harvard Law School in 2008. He was selected as one of *Edutopia Magazine*'s Daring Dozen educators in 2007. Connexions received the Tech Museum Laureate Award from the Tech Museum of Innovation in 2006. His work with Kevin Kelly on the Rice single-pixel compressive camera was selected by MIT Technology Review Magazine as a TR10 Top 10 Emerging Technology in 2007. He was coauthor on a paper with M. Crouse and R. Nowak that won the IEEE Signal Processing Society Junior Paper Award in 2001 and another with V. Ribeiro and R. Riedi that won the Passive and Active Measurement (PAM) Workshop Best Student Paper Award in 2003. He was elected a Fellow of the IEEE in 2001 and a Fellow of the AAAS in 2009.