COMPRESSIVE SENSING BY RANDOM CONVOLUTION

JUSTIN ROMBERG*

Abstract. This paper demonstrates that convolution with random waveform followed by random time-domain subsampling is a universally efficient compressive sensing strategy. We show that an n-dimensional signal which is S-sparse in any fixed orthonormal representation can be recovered from $m \gtrsim S \log n$ samples from its convolution with a pulse whose Fourier transform has unit magnitude and random phase at all frequencies. The time-domain subsampling can be done in one of two ways: in the first, we simply observe m samples of the random convolution, in the second, we break the random convolution into m blocks, and summarize each with a single randomized sum. We also discuss several imaging applications where convolution with a random pulse allows us to super-resolve fine-scale features, allowing us to recover high-resolution signals from low-resolution measurements.

1. Introduction. The new field of *compressive sensing* (CS) has given us a fresh look at data acquisition, one of the fundamental tasks in signal processing. The message of this theory can be summarized succinctly [11,12,14,17]: sparse signals can be recovered from a small number of linear measurements.

We model the acquisition of a signal x_0 as a series of inner products against different waveforms $\phi_1, \ldots, \phi_m$:

$$y_k = \langle \phi_k, x_0 \rangle, \ k = 1, \ldots, m. \tag{1.1}$$

Recovering x_0 from the y_k is then a linear inverse problem. If we assume that x_0 and the ϕ_k are members of some n-dimensional space (the space of n-pixel images, for example), (1.1) becomes a system of $m \times n$ linear equations, $y = \Phi x_0$. In general, we will need more measurements than unknowns, $m \geq n$, to recover x_0 from y. But if the signal of interest x_0 is sparse in a known basis, and the ϕ_k are chosen appropriately, then recovering x_0 is possible even when there are far fewer measurements than unknowns, $m \ll n$.

When we say a signal is "sparse" it is always in the context of an orthonormal signal representation Ψ. Specifically, if Ψ is an $n \times n$ orthogonal matrix (the columns of Ψ form an orthonormal basis for $\mathbb{R}^n$), then we say x_0 is S-sparse in Ψ if we can decompose x_0 as $x_0 = \Psi \alpha_0$, where α_0 has at most S non-zero components.

To recover x_0 from $y = \Phi x_0$, we take explicit advantage of this sparsity. A variety of methods have been recently proposed to do this [8,25,31,38]; in this paper we will focus on ℓ_1 minimization (also called *basis pursuit* [15]). Given the measurements y, we solve

$$\min_{\alpha} \ \|\alpha\|_1 \quad \text{subject to} \quad \Phi\Psi\alpha = y. \tag{1.2}$$

In words, the above searches for the sparsest (as quantified by the ℓ_1 norm) set of transform coefficients α such that the measurements of the corresponding signal $\Psi\alpha$ agree with y.

The effectiveness of (1.2) depends on the nature of the waveforms ϕ_k. The program is guaranteed to succeed from the smallest number of measurements if the ϕ_k are incoherent with respect to Ψ — roughly speaking, this means that the ϕ_k are

* School of Electrical and Computer Engineering, Georgia Tech, Atlanta, Georgia. Email: jrom@ece.gatech.edu. This work was supported by DARPA's Analog-to-Information program and an ONR Young Investigator Award. The author would like to thank Joel Tropp and Andrew Portnoy for conversations related to this paper.
Submitted to the SIAM Journal on Imaging Science on July 9, 2008; revised April 1, 2009.

global and diverse in the Ψ domain. A straightforward way to achieve this diversity is to use random waveforms for the ϕ_k, generating each entry of the measurement matrix Φ independently according to a subgaussian (e.g. normal or Bernoulli with zero mean) distribution [6, 14, 17, 29]. The essential CS result is that if x_0 is S-sparse in Ψ and we make[1]

$$m \gtrsim S \cdot \log n \tag{1.3}$$

random waveform measurements, solving (1.2) will recover x_0 exactly.

Another effective way to generate the measurement ensemble Φ is to select the measurement waveforms ϕ_k from rows of an orthogonal matrix Φ' [11, 35]. By convention, we normalize the rows to have energy $\|\phi_k\|_2^2 = n$, The m rows of Φ' are selected uniformly at random from among all subsets of rows of size m. One example of this type of sensing is observing randomly selected entries of a spectrally sparse signal [12]. In this case, the ϕ_k are elements of the standard basis (and $\Phi' = \sqrt{n}I$), and the representation Ψ is a discrete Fourier matrix.

When we randomly sample from a fixed orthosystem Φ', the number of samples required to reconstruct x_0 depends on the relationship between Φ' and Ψ. One way to quantify this relationship is by using the *mutual coherence*:

$$\mu(\Phi', \Psi) = \max_{\substack{\phi_k \in \mathrm{Rows}(\Phi') \\ \psi_j \in \mathrm{Columns}(\Psi)}} |\langle \phi_k, \psi_j \rangle|. \tag{1.4}$$

It is straightforward to check that this quantity takes values between 1 and $\sqrt{n}$; when $\mu = 1$, Φ' and Ψ are as different as possible, and each element of Ψ is flat in the Φ' domain. If x_0 is S sparse in the Ψ domain, and we take

$$m \gtrsim \mu^2 \cdot S \cdot \log n \tag{1.5}$$

measurements, then (1.2) will succeed in recovering x_0 exactly. Notice that to have the same result (1.3) as the random waveform case, μ will have to be on the order of 1; we call such Φ' and Ψ *incoherent*.

In this paper, we show that bounds similar to (1.3) and (1.4) hold for compressive sensing schemes based on convolution with a random pulse followed by subsampling. The pulse is constructed such that its Fourier transform is a sequence of independent random variables with unit magnitude and uniformly distributed phase. We will consider two types of subsampling schemes, which are discussed in detail in Section 1.2. In the first, we simply sample the convolution at a small number of locations, in the second we apply the "random demodulator" architecture of [22, 40]. This sensing strategy has the following attractive properties:

Universality. A measurement strategy is universal if it is agnostic towards the choice of signal representation. This is the case when sensing with random waveforms [6, 14, 17] (e.g. the entries of Φ are independent Gaussian random variables.) We will see that by convolving the signal with a random pulse, we are essentially transforming it into a random basis which will, with high probability, be incoherent with the sparsity basis Ψ no matter what Ψ is.

[1] Here and below we will use the notation $X \gtrsim Y$ to mean that there exists a known constant C such that $X \geq C\dot{Y}$. For random waveforms, we actually have the the more refined bound of $m \gtrsim S \log(n/S)$. But since we are usually interested in the case $S \ll n$, (1.3) is essentially the same.

Numerical structure. Algorithms for recovering the signal will invariably involve repeated applications of $\Phi\Psi$ and the adjoint $\Psi^*\Phi^*$. Having an efficient method for computing these applications is thus of primary importance. It is often the case that a Ψ which sparsifies signals of interest comes equipped with a fast transform (the wavelet transform for piecewise smooth signals, for example, or the fast Fourier transform for signals which are spectrally sparse). We would like the same from our measurement system Φ.

The complete lack of structure in a measurement ensemble consisting of random waveforms makes a fast algorithm for applying Φ out of the question. There are, however, a few examples of orthobases which are incoherent with sparsity bases of interest and can be applied efficiently. Fourier systems are perfectly incoherent with the identity basis (for signals which are sparse in time), and noiselets [16] are incoherent with wavelet representations and enjoy a fast transform with the same complexity as the fast Fourier transform. Since convolution with a random pulse can be implemented with two FFTs, it gives us a computationally efficient measurement scheme for any sparsity basis.

Physically realizable. In the end, we need to have an acquisition system which collects the linear measurements in (1.1). Collecting samples of an unknown signal convolved with a controlled pulse is a scenario which arises in many applications, including channel estimation in communications and radar, sonar, and seismic imaging.

In the remainder of this section, we carefully define our measurement scheme, state the main theoretical results, and discuss how they relate to other results in the literature. Section 2 briefly overviews how convolution with a random pulse arises in two applications: radar and super-resolved imaging. Section 3 contains proofs of the main results, which draw heavily on the concentration inequalities discussed in the appendix.

1.1. Random Convolution. Our measurement process has two steps. We circularly convolve the signal $x_0 \in \mathbb{R}^n$ with a "pulse" $h \in \mathbb{R}^n$ and then subsample. The pulse is random, global, and broadband in that its energy is distributed uniformly across the discrete spectrum.

In terms of linear algebra, we can write the convolution of x_0 and h as Hx, where

$$H = n^{-1/2} F^* \Sigma F, \tag{1.6}$$

with F as the discrete Fourier matrix whose entries are

$$F_{\omega,t} = e^{-j2\pi(\omega-1)(t-1)/n}, \quad 1 \le t, \omega \le n,$$

and Σ as a diagonal matrix whose non-zero entries are the (renormalized) Fourier transform coefficients of h. We generate h at random by taking

$$\Sigma = \begin{bmatrix} \sigma_1 & 0 & \cdots & \\ 0 & \sigma_2 & \cdots & \\ \vdots & & \ddots & \\ & & & \sigma_n \end{bmatrix},$$

where the σ_ω are unit magnitude complex numbers with random phases. Specifically,

we generate the σ_ω as follows[2]:

$$\begin{aligned}
\omega = 1 \;&:\; \sigma_1 \sim \pm 1 \text{ with equal probability,} \\
2 \le \omega < n/2+1 \;&:\; \sigma_\omega = e^{j\theta_\omega}, \text{ where } \theta_\omega \sim \text{Uniform}([0, 2\pi]), \\
\omega = n/2+1 \;&:\; \sigma_{n/2+1} \sim \pm 1 \text{ with equal probability} \\
n/2+2 \le \omega \le n \;&:\; \sigma_\omega = \sigma^*_{n-\omega+2}, \text{ the conjugate of } \sigma_{n-\omega+2}.
\end{aligned}$$

The action of H on a signal can be broken down into a discrete Fourier transform, followed by a *randomization of the phase* (with constraints that keep the entries of H real), followed by an inverse discrete Fourier transform.

The construction ensures that H will be orthogonal,

$$H^*H = n^{-1}F^*\Sigma^*FF^*\Sigma F = nI,$$

since $FF^* = F^*F = nI$ and $\Sigma^*\Sigma = I$. Thus we can interpret convolution with h as a transformation into a random orthobasis.

1.2. Subsampling. Once we have convolved x_0 with the random pulse, we "compress" the measurements by subsampling. We consider two different methods for doing this. In the first, we simply observe entries of Hx_0 at a small number of randomly chosen locations. In the second, we break Hx_0 into blocks, and summarize each block with a single randomized sum.

1.2.1. Sampling at random locations. In this scheme, we simply keep some of the entries of Hx_0 and throw the rest away. If we think of Hx_0 as a set of Nyquist samples for a bandlimited function, this scheme can be physically realized with an analog-to-digital converter (ADC) that takes non-uniformly spaced samples at an average rate which is appreciably slower than Nyquist. Convolving with the pulse h combines a little bit of information about *all* the samples in x_0 into *each* sample of Hx_0, information which we can untangle using (1.2). Running the ADC at a slower rate gives us the benefit of taking more robust samples, as sample accuracy tends to degrade quickly with sample rate [23, 43], and allows us to acquire signals with bandlimits that are beyond the reach of standard ADCs.

There are two mathematical models for sampling at random locations, and they are more or less equivalent. The first is to set a size m, and select a subset of locations $\Omega \subset \{1, \dots, n\}$ uniformly at random from all $\binom{n}{m}$ subsets of size m. The second is to generate an iid sequence of Bernoulli random variables $\iota_1, \dots, \iota_n$, each of which takes a value of 1 with probability m/n, and sample at locations t where $\iota_t = 1$. In this case, the size of the sample set Ω will not be exactly m. Nevertheless, it can be shown (see [12, Section 2.3] for details) that if we can establish successful recovery with probability $1 - \delta$ for the Bernoulli model, we will have recovery with probability $1 - 2\delta$ in the uniform model. Since the Bernoulli model is easier to analyze, we will use it throughout the rest of the paper.

In either case, the measurement matrix can be written as $\Phi = R_\Omega H$, where R_Ω is the restriction operator to the random set Ω.

1.2.2. Random demodulation. An alternative to simply throwing away most of the samples of Hx_0 is to break them into blocks of size n/m, and summarize each

[2]For simplicity, we assume throughout that n is even; very little changes for the case of n odd.

block with a single number. To keep things manageable, we will assume throughout that m evenly divides n. With

$$B_k = \{(k-1)n/m + 1, \ldots, kn/m\}, \quad k = 1, \ldots, m \tag{1.7}$$

denoting the index set for block k, we take a measurement by multiplying the entries of Hx_0 in B_k by a sequence of random signs and summing. The corresponding row of Φ is then

$$\phi_k = \sqrt{\frac{m}{n}} \sum_{t \in B_k} \epsilon_t h_t, \tag{1.8}$$

where h_t is the tth row of H. The $\{\epsilon_t\}_{t=1}^n$ are independent and take a values of ± 1 with equal probability, and the factor $\sqrt{m/n}$ is a renormalization that makes the norms of the ϕ_k similar to the norms of the h_t. We can write the measurement matrix as $\Phi = P\Theta H$, where Θ is a diagonal matrix whose non-zero entries are the $\{\epsilon_t\}$, and P sums the result over each block B_k; the kth row of P has ones at the column indexes in B_k and zeros elsewhere.

In hardware, the random demodulator can be implemented by multiplying the incoming signal by a high-rate (Nyquist or above) chipping sequence, effectively changing its polarity across short time intervals. This is followed by an integrator (to compute the analog of the sum in (1.8)) and an ADC taking equally spaced samples at a rate well below Nyquist. This acquisition strategy is discussed in [22] and analyzed (without the convolution with h) in [40], where it is shown to be effective for capturing spectrally sparse signals. The advantage of the random demodulator architecture is that we have replaced a high-rate ADC with a less complicated high-rate mixing circuit and a low-rate ADC, thus reducing the complexity of the hardware.

The recovery bounds we derive for random convolution followed by random demodulation will have the same form as those in the random subsampling scheme. Random demodulation has one important advantage: in practice, the sum in (1.8) is implemented without the normalization factor $\sqrt{m/n}$, and as a result it incorporates more of the signal energy into each measurement than random subsampling. To see this, suppose that the energy in Hx_0 is spread out more or less evenly across all locations — this is, after all, one of the purposes of the random convolution. If the energy (ℓ_2 norm squared) in the entire signal is 1, then the energy in each sample will be about $1/n$, and the energy in each block will be about n/m. If we simply take m samples, the total energy in the measurements will be around m/n. If we take a random sum as in (1.8) (without the renormalization factor) over a block B_k of size n/m, the expected squared magnitude of this single measurement will be the same as the energy of Hx_0 in B_k, and the total energy of all the measurements will be (in expectation) the same as the energy in Hx_0. The random demodulator is a subsampling architecture which will, on average, retain all of the signal energy.

It is important to understand that the signs of the terms in the summation (1.8) must be random. Imagine if we were to leave out the $\{\epsilon_t\}$ and simply sum Hx_0 over each B_k. This would be equivalent to putting Hx_0 through a boxcar filter and then subsampling uniformly. Since boxcar filtering Hx_0 is also a convolution, it commutes with H, so this strategy would be equivalent to first convolving x_0 with the boxcar filter, then convolving with h, then subsampling uniformly. But now it is clear how the convolution with the boxcar hurts us: it filters out the high frequency information in x_0, which cannot be recovered no matter what we do next.

We will see that while coherent summation will destroy a sizable fraction of the signal, random demodulation does not. This fact will be especially important in the imaging architectures discussed in Section 2.2, where measurements are taken by averaging over large pixels.

1.3. Main Results. Both random subsampling [12] and random demodulation [40] have been shown to be effective for compressive sampling of spectrally sparse signals. In this paper, we show that preceding either by a random convolution results in a *universal* compressive sampling strategy.

Our first theoretical result shows that if we generate a random pulse as in Section 1.1 and a random sampling pattern as in Section 1.2.1, then with high probability we will be able to sense the vast majority of signals supported on a fixed set in the Ψ domain.

THEOREM 1.1. *Let Ψ be an arbitrary orthonormal signal representation. Fix a support set Γ of size $|\Gamma| = S$ in the Ψ domain, and choose a sign sequence z on Γ uniformly at random. Let α_0 be a set of Ψ domain coefficients supported on Γ with signs z, and take $x_0 = \Psi\alpha_0$ as the signal to be acquired. Create a random convolution matrix H as described in Section 1.1, and choose a set of sample locations Ω of size $|\Omega| = m$ uniformly at random with*

$$m \geq C \cdot (S \log n + \log^3 n), \tag{1.9}$$

where C is a universal constant. Set $\Phi = R_\Omega H$. Then given the set of samples on Ω of the convolution Hx_0, $y = \Phi x_0$, the program (1.2) will recover α_0 (and hence x_0) exactly with probability exceeding $1 - O(n^{-1})$.

Roughly speaking, Theorem 1.1 works because if we generate a convolution matrix H at random, with high probability it will be incoherent with any fixed orthonormal matrix Ψ. Actually using the coherence defined in (1.4) directly would give a slightly weaker bound ($\log^2 n$ instead of $\log n$ in (1.9)); our results will rely on a more refined notion of coherence which is outline in Section 3.1.

Our second result shows that we can achieve similar acquisition efficiency using the random demodulator. The result is slightly weaker in that it also relies on the support Γ of the signal being random.

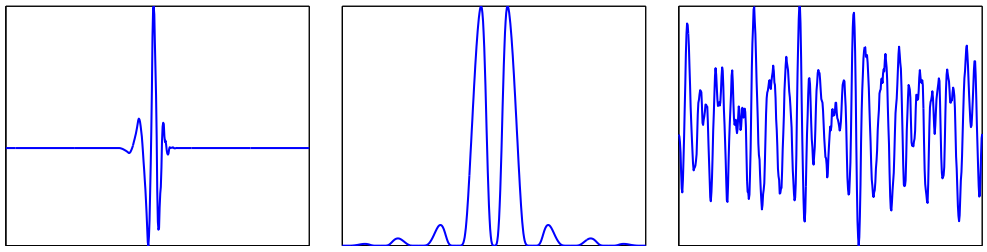
THEOREM 1.2. *Let Ψ and H be as in Theorem 1.1. Choose a support Γ of size S uniformly at random, and then a sign sequence z on Γ uniformly at random. Let α_0 be a coefficient sequence supported on Γ with signs z, and take $x_0 = \Psi\alpha_0$ as the signal to be acquired. Create a random demodulator matrix $P\Theta$, as described in Section 1.2.1, that outputs a number of samples m with*

$$m \;\geq\; C \cdot (S \log n + \log^3 n), \tag{1.10}$$

where C is a constant, and set $\Phi = P\Theta H$. Then given the measurements $y = \Phi x_0$, the program (1.2) will recover α_0 (and hence x_0) exactly with probability exceeding $1 - O(n^{-1})$.

The form of the bound (1.10) is the same as using the random demodulator (without the random convolution in front) to sense spectrally sparse signals [40].

At first, it may seem counterintuitive that convolving with a random pulse and subsampling would work equally well with any sparsity basis. After all, an application of $H = n^{-1/2} F^* \Sigma F$ will not change the magnitude of the Fourier transform, so signals which are concentrated in frequency will remain concentrated and signals which are spread out will stay spread out. For compressive sampling to work, we need Hx_0



(a) (b) (c)

FIG. 1.1. *(a) A signal x_0 consisting of a single Daubechies-8 wavelet. Taking samples at random locations of this wavelet will not be an effective acquisition strategy, as very few will be located in its support. (b) Magnitude of the Fourier transform Fx_0. (c) Inverse Fourier transform after the phase has been randomized. Although the magnitude of the Fourier transform is the same as in (b), the signal is now evenly spread out in time.*

to be "spread out" in the time domain. We already know that signals which are concentrated on a small set in the Fourier domain will be spread out in time [12]. The randomness of Σ will make it highly probable that a signal which is concentrated in time will not remain so after H is applied. Time localization requires very delicate relationships between the phases of the Fourier coefficients, relationships that will no longer exist after applying Σ. A simple example of this phenomena is shown in Figure 1.1.

1.4. Stability. Theorems 1.1 and 1.2 tell us that we can recover a perfectly S-sparse signal from on the order of $S \log n$ samples of its convolution with a random waveform. If we are willing to pay additional log factors, we can also guarantee that the recovery will be stable.

In [13], it is shown that stable recovery of sparse signals follows from Φ having the *restricted isometry property*, which was first introduced in [14]. This property says that all submatrices of Φ created by extracting a certain number of columns are well-conditioned. In particular, if

$$\rho_{2S} := \max_{|\Gamma| \leq 2S} \|m^{-1} \Phi_\Gamma^* \Phi_\Gamma - I\| \leq C_1, \tag{1.11}$$

where $\|\cdot\|$ is the standard operator norm and $C_1 \leq \sqrt{2}-1$ [10], then we can recover the signal about as well as if we had observed the S most significant components directly. Specifically, suppose that (1.11) holds and we make noisy measurements $y = \Phi x_0 + e$ of a (not necessarily sparse) signal $x_0 = \Psi \alpha_0$. If the size of the error can be bounded by $\|e\|_2 \leq E$, then the solution $\alpha^\star$ to a relaxed version of (1.2),

$$\min_\alpha \ \|\alpha\|_1 \quad \text{subject to} \quad \|\Phi\alpha - y\|_2 \ \leq \ E,$$

will obey

$$\|\alpha^\star - \alpha_0\|_2 \ \leq \ C_2 \left(m^{-1/2} E + S^{-1/2} \|\alpha_0 - \alpha_{0,S}\|_1 \right). \tag{1.12}$$

Above, the factor of $m^{-1/2}$ in front of the measurement error E is an artifact of the normalization of the columns of Φ, and $\alpha_{0,S}$ is the best S-term approximation to α_0, formed by keeping the largest S terms in α_0 and setting the rest to zero. In short,

(1.12) shows us that the recovery error will be within a constant times the (scaled) noise error plus an approximation error.

The following theorems show that measurement matrices composed from a random convolution followed by random subsampling or random demodulation will have the desired restricted isometry constants (1.11) for m within a log factor of S. In each, we break the results into two cases: "large-S", which scales linearly with S, and "small-S" which is quadratic in S but scales better with n.

THEOREM 1.3. *Let $\Phi = R_\Omega H \Psi$ be a measurement matrix for random convolution followed by random subsampling as in Theorem 1.1. Then with probability exceeding $1 - O(n^{-1})$, for*

$$m \geq C\rho^{-2} \min(S \log^6 n, S^2 \log^2 n), \tag{1.13}$$

Φ has the restricted isometry constants $\rho_{2S} \leq \rho$.

THEOREM 1.4. *Let $\Phi = P\Theta H \Psi$ be a measurement matrix for random convolution followed by random demodulation as in Theorem 1.2. Then with probability exceeding $1 - O(n^{-1})$, for*

$$m \geq C\rho^{-2} \min(S \log^6 n, S^2 \log n) \tag{1.14}$$

Φ has restricted isometry constants $\rho_{2S} \leq \rho$.

Proofs of these theorems follow quickly from estimates of the spectral norm of the sum of independent matrices developed in [35], and are presented in Section 3.5. We close by noting that if we are willing to accept a constant probability of failure (instead of $O(n^{-1})$), a log factor can be removed from the large-S case in both (1.13) and (1.14) above.

1.5. Relationship to previous research. Random convolution for compressive sensing has been explored in several other places in the literature. In [41], numerical results are presented that demonstrate recovery of sparse signals (using orthogonal matching pursuit instead of ℓ_1 minimization) that have been subsampled after being passed through a finite-length impulse response (FIR) filter whose taps are independent random variables. That paper also demonstrates the trade-off between the number of taps in the filter and the amount of subsampling that can be endured for signals which are sparse in the time domain. In [4,19], theoretical bounds are developed for a slightly different random convolution model. There, the sensing matrix Φ consists of m consecutive rows of a random Toeplitz matrix, corresponding to convolution followed by a small number of *consecutive* samples. It is shown that recovery is guaranteed when $m \gtrsim S^2 \log n$. Very recently, these results were improved in two different ways in [33]: the bound was tightened to $m \gtrsim S \log^3 n$, and was shown to hold for arbitrary (but fixed) sets of samples whose size obeys this bound.

The sensing model used in this paper is different than in both of the above. First, the nature of the pulse is different; rather than consisting of a sequence of independent random variables in the time domain, it is shaped in the Fourier domain. Second, the nature of the subsampling is different. Here, we take a select the sample locations at random, in both works above, the sampling locations are fixed (equally spaced in [41] and consecutive in [19]). These differences are what gives the strategy its universality. Pulses consisting of iid random sequences in the time domain will invariably have small components in the frequency domain, meaning that there are certain spectrally sparse signals which will be nearly impossible to recover. Likewise, if the samples are taken uniformly there will be aliasing effects that cannot be corrected. These problems

completely disappear with the sensing model used in this paper; convolving with a random pulse created as in Section 1.1 and subsampling using one of the two methods in Section 1.2 is an effective strategy for *any* type of sparsity.

There is some question about whether or not generating the pulse in the frequency domain is as convenient as doing so in the time domain. It is true that it would be slightly harder to generate the pulse "on the fly", but this will not be much of an issue for sensing devices which simply use a pre-programmed pulse.

The mathematics in this paper is closely related to two previous works. The random subsampling scheme, where the measurement waveforms ϕ_k are chosen randomly from a fixed orthosystem Φ', was studied in [11]; the main result of that work is summarized in the bound (1.5), which relates the number of samples required to the sparsity of the signal and the incoherence between the representation Φ' and the representation Ψ. The random convolution matrix in (1.6) is a special type of random orthosystem which will, with high probability, be incoherent with any fixed Ψ. Establishing Theorem 1.1, however, will require a slightly more refined notion of coherence. A full analysis is given of the random demodulator's performance on spectrally sparse signals is given in [40]. Theorem 1.2 shows that the same bounds are possible for any representation if the demodulation is preceded by a random convolution.

In [1], Ailon and Chazelle propose the idea of a randomized Fourier transform followed by a random projection as a "fast Johnson-Lindenstrauss transform" (FJLT). The transform is decomposed as $QF\Sigma$, where Q is a sparse matrix with non-zero entries whose locations and values are chosen at random locations. They show that this matrix $QF\Sigma$ behaves like a random waveform matrix in that with extremely high probability, it will not change the norm of an arbitrary vector too much. However, this construction requires that the number of non-zero entries in each row of Q is commensurate with the number of rows m of Q. Although ideal for dimensionality reduction of small point sets, this type of subsampling does not translate well to compressive sampling, as it would require us to randomly combine on the order of m samples of Hx_0 from arbitrary locations to form a single measurement — taking m measurements would require on the order of m^2 samples. We show here that more structured projections, consisting either of one randomly chosen sample per row or a random combination of *consecutive* samples, are adequate for CS. This is in spite of the fact that our construction results in much weaker concentration than the FJLT.

The idea that the sampling rate for a sparse ("spikey") signal can be significantly reduced by first convolving with a kernel that spreads it out is one of the central ideas of sampling signal with finite rates of innovation [26,42]. Here, we show that a random kernel works for any kind sparsity, and we use an entirely different reconstruction framework.

2. Applications. The fact that random convolution is universal and allows fast computations makes it extremely attractive as a theoretical sensing strategy. In this section, we briefly discuss a few imaging scenarios (in a somewhat rarified stetting) in which convolution with a random pulse can be implemented naturally.

We begin by noting that while Theorems 1.1 and 1.2 above deal explicitly with circular convolution, what is typically implemented is linear convolution. One simple way to translate our results to linear convolution would be to repeat the pulse h; then the samples in the midsection of the linear convolution would be the same as samples from the circular convolution.

2.1. Active imaging. There are many imaging applications in which reconstruction of a signal from its convolution with a known pulse is a fundamental task.

A common framework, used for sensing modalities as varied as radar, sonar, and seismic could be described as follows. A transmitter sends out a pulse, the pulse interacts with the environment, and the result is measured by a receiver (either at the same or a different location). The pulse's interaction with the environment can be modeled as a continuous series of reflections, so the receiver is measuring the pulse convolved with a reflectivity profile. This process takes place for a number of different source/receiver locations, and after the reflectivity profiles are recovered from the received signals, they are assembled into an image of the environment.

As a particular example, Figure 2.1 illustrates how a scene is measured in spotlight synthetic aperture radar (SAR) imaging (see [30] for a more in-depth discussion). The radar, located at point p_1, focusses its antenna on the region of interest with reflectivity function $I(x_1, x_2)$ whose center is at orientation θ relative to p_1, and sends out a pulse $h(t)$. If the radar is far enough away from the region of interest, this pulse will arrive at every point along one of the parallel lines at a certain range r at approximately the same time. The net reflectivity from this range is then the integral $R_\theta(r)$ of $I(x_1, x_2)$ over the line at $l_{r,\theta}$,

$$R_\theta(r) = \int_{l_{r,\theta}} I(x_1, x_2) dx_1 dx_2,$$

and the return signal $y(t)$ is thus the pulse convolved with R_θ

$$y(t) = h * R_\theta,$$

where it is understood that we can convert R_θ from a function of range to a function of time by dividing by the speed at which the pulse travels. These R_θ are slices from the Radon transform of $I(x_1, x_2)$; the image $I(x_1, x_2)$ is created by recovering the profiles R_θ over a range of θ, and applying an inverse Radon transform.

Suppose that we want to reconstruct $R_\theta(r)$ to a certain resolution n. There are a number of ways to discretize this problem, but since the support of $R_\theta(r)$ is limited to the region of interest, our goal will be to reconstruct the first n lowpass Fourier series coefficients. The radar creates a pulse whose length is the same as the region of interest, and whose n lowpass Fourier coefficients are non-zero. Two copies of the pulse are sent out, and they get convolved with R_θ to form $y(t)$.

The question, then, is how many samples of $y(t)$ are needed to reconstruct the range profile R_θ. A classical reconstruction will require n equally spaced samples, meaning that the sampling rate will scale linearly with the desired resolution. If, however, R_θ is S sparse in some basis Ψ, and the n non-zero Fourier series coefficients of the pulse have unit magnitude and random phase, then Theorems 1.1 and 1.2 say that we will only require $\sim S \log n$ samples to reconstruct the profile using (1.2). For compressive sampling, the rate increases linearly with the *complexity* of the range profiles, and only *logarithmically* with the resolution. This is significant, as the sampling rate of the analog-to-digital converter is one of the limiting factors in the performance of modern-day radar systems [34, Chapter 1].

A preliminary investigation of using ideas from compressive sensing in radar imaging can be found in [5]. There has also been some recent work on implementing low-cost radars which use random waveforms [2, 3] and traditional image reconstruction techniques. In [20], it is shown that compressive sensing can be used to super-resolve point targets when the radar sends out an incoherent pulse. Related work can also be found in [32], which demonstrates how this type of problem is related to sparse

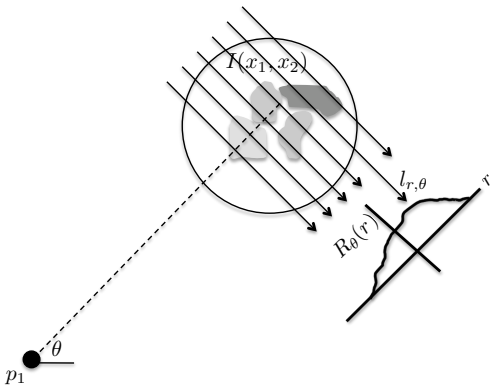


FIG. 2.1. *Geometry for the "spotlight SAR" imaging problem. The return signal from a pulse $h(t)$ emitted from point p_1 will be the range profile $R_\theta(r)$ (collection of line integrals at an angle θ) convolved with h.*

approximation in a dictionary composed of different shifts and modulations of the same waveform.

2.2. Super-resolved imaging. Convolution with a random pulse can also be used to overcome the resolution limitations of an optical imaging system (see also the work in [27, 28]). We consider the scenario where the object we want to image is illuminated by a monochromatic coherent light source; one might think of the object of interest as a pattern on a transparency, and the image we want to acquire as the intensity of the light field exiting this transparency. The intensity of the light field is measured using a detector array. Each measurement is an average of the intensity over a certain region in the image, and we have the standard trade-offs between the size of the "pixels", the signal-to-noise ratio of the measurements (bigger pixels mean higher SNR) and the resolution of the final image. Mathematically, the detector array is applying the matrix P to the light field.

The resolution of this system can be enhanced by convolving the image intensity with a random pattern and then randomly demodulating the result. The random demodulator can be implemented with a high-resolution spatial light modulator (SLM) placed in front of the detector array. The SLM gives us a way to multiply the light intensity pointwise, and implements the action of the diagonal matrix Θ. As the detectors can only measure the magnitudes, we can keep everything positive by using random patterns consisting of 0s and 1s rather than -1s and 1s. If we make two sets of measurements, one with the SLM configured to pass all of the light and one with a random pattern, we can simply subtract them to emulate the effect of a Θ populated with ± 1.

There are a number of ways the convolution can be implemented. The first is through movement, another SLM coded with the pulse h is introduced and "dragged" across the field of view while the sensing array remains stationary. Another way is by using an *optical correlator*, as shown in Figure 2.2. After being illuminated, the image passes through a lens, which takes a Fourier transform of the image, passes through another SLM, which pointwise multiplies the Fourier transform, then passes through another lens which takes an inverse Fourier transform. Again, since in the end only measure the magnitude of the result of the convolution, we ensure that the "pulse" (pattern we are convolving with) is positive by shifting it upwards, convolving

with $h^+ = h + c$, where c is some number larger than the magnitude of the largest negative entry in h. The convolution will now yield $x_0 * h^+ = x_0 * h + \text{constant}$, where the constant depends on the mean of x_0, a quantity that may be possible to observe directly.

Without the spatial light modulators, the resolution of this system scales with the size of the detector array. The big-pixel detectors simply average the light field over a relatively large area, and the result is a coarsely pixellated image. Adding the spatial light modulators allows us to effectively super-resolve this coarse image. With the SLMs in place, the detector is taking incoherent measurements in the spirit of Theorem 1.2 above. The resolution (for sparse images and using the reconstruction (1.2)) is now determined by the SLMs: the finer the grid over which we can effectively manipulate the phases, the more effective pixels we will be able to reconstruct.

Figure 2.3 illustrates the potential of this type of architecture with a simple numerical experiment[3]. A 256×256 synthetic image, was created by placing 40 ellipses — with randomly chosen orientations, positions, and intensities — and adding a modest amount of noise. Measuring this image with a 64×64 "low resolution" detector array produces the image in Fig. 2.3(b), where we have simply averaged the image in part (a) over 4×4 blocks. Figure 2.3(c) is the result when the low resolution measurements are augmented with measurements from the architecture in Figure 2.2. With x_0 as the underlying image, we measure $y = \Phi x_0$, where

$$\Phi = \begin{bmatrix} P \\ P\Theta H \end{bmatrix}.$$

From these measurements, the image is recovered using total variation minimization, a variant of ℓ_1 minimization that tends to give better results on the types of images being considered here. Given y, we solve

$$\min_x \ \mathrm{TV}(x) \quad \text{subject to} \quad \|\Phi x - y\|_2 \leq \epsilon,$$

where ϵ is a relaxation parameter set at a level commensurate with the noise. The result is shown in Figure 2.3(c). As we can see, the incoherent measurements have allowed us to super-resolve the image; the boundaries of the ellipses are far clearer than in part (b).

3. Theory.

3.1. Coherence bounds. First, we will establish a simple bound on the coherence parameter between a random convolution system and a given orthonormal representation.

LEMMA 3.1. *Let Ψ be an arbitrary fixed orthonormal matrix, and create H at random as above with $H = n^{-1/2} F^* \Sigma F$. Choose $0 < \delta < 1$. Then with probability exceeding $1 - \delta$, the coherence $\mu(H, \Psi)$ will obey*

$$\mu(H, \Psi) \leq 2\sqrt{\log(2n^2/\delta)}. \tag{3.1}$$

[3]Matlab code that reproduces this experiment can be downloaded at users.ece.gatech.edu/∼justin/randomconv/ .

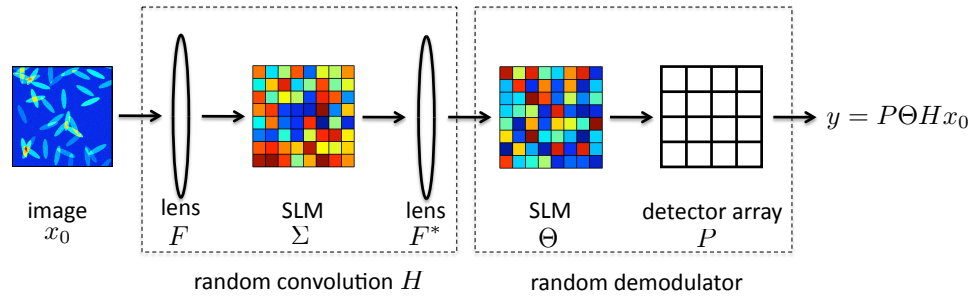


FIG. 2.2. *Fourier optics imaging architecture implementing random convolution followed by RPMS. The computation $y = P\Theta H x_0$ is done entirely in analog; the lenses move the image to the Fourier domain and back, and spatial light modulators (SLMs) in the Fourier and image planes randomly modulate the amplitude and phase.*

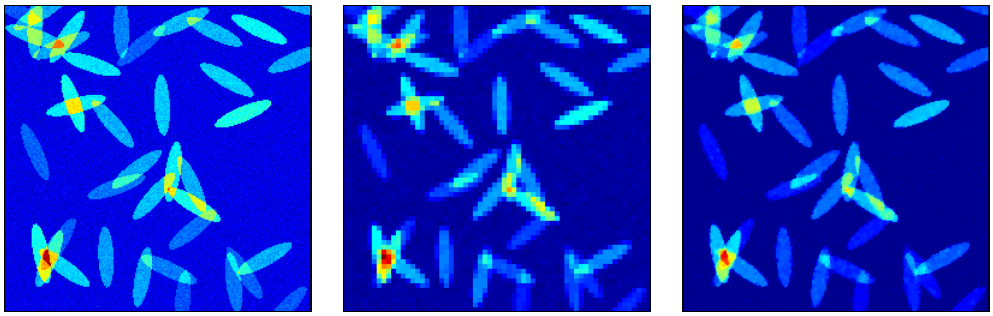


FIG. 2.3. *Super-resolved imaging experiment. (a) The "high-resolution" image x_0 we wish to acquire. The false-color image represents an intensity profile. (b) The high-resolution image pixellated by averaging over 4×4 blocks, Px_0. (c) The image restored from the pixellated version in (b), plus a set of incoherent measurements taken using the architecture from Figure 2.2. The incoherent measurements allow us to effectively super-resolve the image in (b).*

Proof.

Using h_t to denote the tth row of H and ψ_s the sth column of Ψ, we will find a uniform upper bound, holding with the required probability, on $|\langle h_t, \psi_s \rangle|$ over all pairs of t, s. To accomplish this, we show how each of these inner products can be written as a sum of independent random variables, apply Hoeffding's inequality ((A.1) from the Appendix), and use a union bound over all n^2 choices for t and s.

The inner product $\langle h_t, \psi_s \rangle$ is

$$\langle h_t, \psi_s \rangle = \sum_{\omega=1}^{n} e^{j2\pi(t-1)(\omega-1)/n} \sigma_\omega \hat{\psi}_s(\omega),$$

where $\hat{\psi}_s$ is the normalized discrete Fourier transform $n^{-1/2} F \psi_s$. Since h_t and ψ_s are

real-valued and the σ_ω are conjugate symmetric, we can rewrite this sum as

$$\langle h_t, \psi_s \rangle = \sigma_1 \hat{\psi}_s(1) + 2 \sum_{\omega=2}^{n/2} \mathrm{Re}\left[F^*_{t,\omega} \sigma_\omega \hat{\psi}_s(\omega) \right] + (-1)^{t-1} \sigma_{n/2+1} \hat{\psi}_s(n/2+1)$$
$$= \sum_{\omega=1}^{n/2+1} Y_\omega$$

where

$$Y_\omega = \begin{cases} \sigma_1 \hat{\psi}_s(1), & \omega = 1 \\ 2\,\mathrm{Re}\left[F^*_{t,\omega} \sigma_\omega \hat{\psi}_s(\omega) \right], & 2 \le \omega \le n/2 \\ (-1)^{t-1} \sigma_{n/2+1} \hat{\psi}_s(n/2+1), & \omega = n/2+1. \end{cases}$$

As each of the σ_ω in the sum above are independent, $\{Y_\omega\}$ is a sequence of independent, zero-mean random variables, each of which are bounded:

$$|Y_\omega| \le a_\omega, \quad \text{with} \quad a_\omega = \begin{cases} |\hat{\psi}_s(1)|, & \omega = 1 \\ 2|\hat{\psi}_s|, & 2 \le \omega \le n/2 \\ |\hat{\psi}_s(n/2+1)|, & \omega = n/2+1 \end{cases}.$$

Applying Hoeffding's inequality (A.1) yields

$$\mathrm{P}\left(|\langle h_t, \psi_s \rangle| > \lambda \right) \;\le\; 2 \exp\left(-\frac{\lambda^2}{2 \sum_{\omega=1}^{n/2+1} a_\omega^2} \right). \tag{3.2}$$

For the sum-of-squares in the denominator of the exponent above, notice that

$$\|\hat{\psi}_s\|_2^2 = |\hat{\psi}_s(1)|^2 + 2 \sum_{\omega=2}^{n/2} |\hat{\psi}_s(\omega)|^2 + |\hat{\psi}_s(n/2+1)|^2,$$

and so

$$\sum_{\omega=1}^{n/2+1} a_\omega^2 \le 2 \|\hat{\psi}_s\|_2^2 = 2,$$

and (3.2) becomes

$$\mathrm{P}\left(|\langle h_t, \psi_s \rangle| > \lambda \right) \;\le\; 2e^{-\lambda^2/4}.$$

Taking $\lambda = 2\sqrt{\log(2n^2/\delta)}$ and applying the union bound over all n^2 choices of t and s gives us

$$\mathrm{P}\left(\max_{t,s} |\langle h_t, \psi_s \rangle| > 2\sqrt{\log(2n^2/\delta)} \right) \;\le\; \delta,$$

establishing the lemma. □

Applying Lemma 3.1 directly to the recovery result (1.5) guarantees recovery from

$$m \gtrsim S \cdot \log^2 n$$

randomly chosen samples, a log factor off of (1.9). We are able to get rid of this extra log factor by using a more subtle property of the random measurement ensemble H.

Fix a subset Γ of the Ψ domain of size $|\Gamma| = S$, and let Ψ_Γ be the $n \times S$ matrix consisting of the columns of Ψ indexed by Γ. In place of the coherence, our quantity of interest will be

$$\nu := \nu(\Gamma) = \max_{k=1,\dots,n} \|r^k\|_2 \tag{3.3}$$

where the r^k are rows in the matrix $H\Psi_\Gamma$; we will call $\nu(\Gamma)$ the *cumulative coherence* Γ (this quantity was also used in [37]). We will show below that we can bound the number of samples needed to recover (with high probability) a signal on Γ by

$$m \gtrsim \nu^2 \cdot \log n.$$

Since we always have the bound $\nu \le \mu\sqrt{S}$, the result (1.5) is still in effect. However, Lemma 3.2 below will show that in the case where $U = H\Psi$ with H as a random convolution matrix, we can do better, achieving $\nu \lesssim \sqrt{S}$.

LEMMA 3.2. *Fix an orthobasis Ψ and a subset of the Ψ-domain $\Gamma = \{\gamma_1, \dots, \gamma_S\}$ of size $|\Gamma| = S$. Generate a random convolution matrix H at random as described in Section 1.1 above, and set $U = H\Psi$. Let $r^k,\ k = 1, \dots, n$ be the rows of $H\Psi_\Gamma = U_\Gamma$. Then with probability exceeding $1 - \delta$*

$$\nu(\Gamma) = \max_{k=1,\dots,n} \|r^k\|_2 = \max_{k=1,\dots,n} \left(\sum_{\gamma \in \Gamma} U_{k,\gamma}^2 \right)^{1/2} \le \sqrt{8S}, \tag{3.4}$$

for $S \ge 16 \log(2n/\delta)$.

Proof. Let $\hat{\Psi}_\Gamma = n^{-1/2} F \Psi_\Gamma$ be the matrix formed by taking the Fourier transform of the columns of Ψ_Γ. We can then write r^k as a random sum of the *rows* of $\hat{\Psi}_\Gamma$:

$$r^k = \sum_{\omega=1}^{n} F_{\omega,k} \sigma_\omega^* g_\omega,$$

where $g_\omega \in \mathbb{C}^S$ is a row of $\hat{\Psi}_\Gamma$ (column of $\hat{\Psi}_\Gamma^*$):

$$g_\omega = \begin{pmatrix} \hat{\psi_{\gamma_1}}(\omega)^* \\ \hat{\psi_{\gamma_2}}(\omega)^* \\ \cdots \\ \hat{\psi_{\gamma_S}}(\omega)^* \end{pmatrix}.$$

By conjugate symmetry, we can rewrite this as a sum of random vectors in $\mathbb{R}^S$,

$$r^k = \sigma_1 g_1 + \sigma_{n/2+1} (-1)^{k-1} g_{n/2+1} + 2 \sum_{\omega=2}^{n/2} \mathrm{Re}\left[F_{\omega,k} \sigma_\omega^* g_\omega \right]$$

(note that g_1 and $g_{n/2+1}$ will be real-valued). Because the σ_ω are uniformly distributed over the unit circle, the random vector $\mathrm{Re}[F_{k,\omega}\sigma_\omega^* g_\omega]$ is symmetric; it has a distribution identical to $\epsilon_\omega \mathrm{Re}[F_{k,\omega}\sigma_\omega^* g_\omega]$, where ϵ_ω is an independent Rademacher random variable. The random variable of interest is

$$Z = \left\| \sum_{\omega=1}^{n/2+1} \epsilon_\omega v_\omega \right\|_2, \quad \text{where} \quad v_\omega = \begin{cases} \sigma_1 g_1 & \omega = 1 \\ 2\,\mathrm{Re}\left[F_{k,\omega} \sigma_\omega^* g_\omega \right] & 2 \le \omega \le n/2\,, \\ \sigma_{n/2+1} g_{n/2+1} & \omega = n/2+1 \end{cases}$$

which depends on two random sequences $\{\epsilon_\omega\}$ and $\{\sigma_\omega\}$. With the $\{\sigma_k\}$ fixed and for an appropriate value of λ, we will develop a bound on the conditional probability $\mathrm{P}_\epsilon(Z > \lambda)$ that is independent of the values of the $\{\sigma_k\}$ (the notation P_ϵ denotes a probability when only the $\{\epsilon_\omega\}$ are treated as random). Since

$$\mathrm{P}(\|r^k\|_2 > \lambda) = \mathrm{P}_{\epsilon,\sigma}(Z > \lambda) = \int \mathrm{P}_\epsilon(Z > \lambda) d\sigma,$$

this will immediately gives us the same bound for $\|r^k\|_2$.

We can bound the mean (with respect to the $\{\epsilon_\omega\}$) of Z as follows:

$$\begin{aligned}
\left(\mathrm{E}_\epsilon[Z]\right)^2 &\le \mathrm{E}_\epsilon\left[Z^2\right] \\
&= \sum_{\omega=1}^{n/2+1} \|v_\omega\|_2^2 \\
&= \|g_1\|_2^2 + 4\sum_{\omega=2}^{n/2} \|\,\mathrm{Re}\left[F_{k,\omega}\sigma_\omega^* g_\omega\right]\|_2^2 + \|g_{n/2+1}\|_2^2 \\
&\le \|g_1\|_2^2 + 4\sum_{\omega=2}^{n/2} \|g_\omega\|_2^2 + \|g_{n/2+1}\|_2^2 \\
&\le 2\sum_{\omega=1}^{n} \|g_\omega\|_2^2 \qquad (\text{since } \|g_\omega\|_2^2 = \|g_{n-\omega+2}\|_2^2 \text{ for } \omega = 2,\ldots,n/2) \\
&= 2\,\mathrm{Trace}(\hat{\Psi}_\Gamma^* \hat{\Psi}_\Gamma) \\
&= 2S \qquad (\text{since each of the } S \text{ rows of } \hat{\Psi}_\Gamma^* \text{ has unit norm}).
\end{aligned}$$

To show that Z is concentrated about its mean, we use a concentration inequality on the norm of a sum of random vector, specifically (A.4) which is detailed in the Appendix.

To apply this bound first note that for any real-valued vector $\xi \in \mathbb{R}^S$

$$|\langle \xi, \mathrm{Re}[F_{k,\omega}\sigma_\omega^* g_\omega]\rangle|^2 \;\le\; |\langle \xi, F_{k,\omega}\sigma_\omega^* g_\omega\rangle|^2 \;=\; |\langle \xi, g_\omega\rangle|^2.$$

Thus

$$\begin{aligned}
\sup_{\|\xi\|_2\le 1} \sum_{\omega=1}^{n/2+1} |\langle \xi, v_\omega\rangle|^2 &\le \sup_{\|\xi\|_2\le 1} \sum_{\omega=1}^{n/2+1} |\langle \xi, g_\omega\rangle|^2 \\
&\le 2 \sup_{\|\xi\|_2\le 1} \sum_{\omega=1}^{n} |\langle \xi, g_\omega\rangle|^2 \\
&= 2 \sup_{\|\xi\|_2\le 1} \xi^T \hat{\Psi}_\Gamma^* \hat{\Psi}_\Gamma \xi \\
&= 2,
\end{aligned}$$

where the last equality follows from the fact that the columns of $\hat{\Psi}_\Gamma$ are orthonormal. We now apply (A.4) to get

$$\mathrm{P}_\epsilon\left(Z \ge \sqrt{2S} + \lambda\right) \;\le\; \mathrm{P}_\epsilon\left(Z \ge \mathrm{E}_\epsilon[Z] + \lambda\right) \;\le\; 2e^{-\lambda^2/32}$$

and so for any $C > \sqrt{2}$,

$$\mathrm{P}\left(\|r^k\|_2 \geq C\sqrt{S}\right) = \mathrm{P}_\epsilon\left(Z \geq C\sqrt{S}\right) \leq 2e^{-(C-\sqrt{2})^2 S/32} \leq \delta/n$$

when $S \geq 32(C-\sqrt{2})^{-2}\log(2n/\delta)$. Taking $C = 2\sqrt{2}$ and applying the union bound over $k = 1, \ldots, n$ establishes the lemma. □

In the establishing the random demodulator result, Theorem 1.2, we will also find it useful to bound the norm of the entries in every column of $H\Psi$ restricted to each block B_k.

LEMMA 3.3. *For Ψ, H as above, set $U = H\Psi$, and denote the individual entries in U as $U_{p,\gamma}$. Let B_k be the index set for the kth block of the random demodulator as in* (1.7)*. With probability exceeding $1-\delta$,*

$$\max_{\substack{1\leq\gamma\leq n\\ 1\leq k\leq m}} \left(\sum_{p\in B_k} U_{p,\gamma}^2\right)^{1/2} \quad \leq \quad \sqrt{8n/m} \tag{3.5}$$

for

$$m \leq \frac{n}{16\log(2n^2/\delta)}.$$

Proof. Since U is circulant, $U_{k+p,\gamma} = U_{k,(\gamma-p)_n}$, where the subtraction $(\gamma-p)_n$ is modulo n, and so

$$\sum_{p\in B_k} U_{p,\gamma}^2 \;=\; \sum_{q=1}^{n/m} U_{(k-1)n/m+q,\gamma}^2 \;=\; \sum_{q=1}^{n/m} U_{(k-1)n/m,(\gamma-q)_n}^2.$$

From here, we proceed exactly as in Lemma 3.2 with $\Gamma = (\gamma - B_k)_n$, $|\Gamma| = n/m$ and take the union bound over all $\gamma = 1, \ldots, n$ and $k = 1, \ldots, m$. □

3.2. Sparse Recovery. The proofs of Theorems 1.1 and 1.2 follow the same general outline put forth in [11] and [40], but with a few important modifications (namely Lemmas 3.4 and 3.7 below) that take advantage of the refined notions of coherence discussed in Section 3.1 above. As detailed in [12,18,39], sufficient conditions for the successful recovery of a vector α_0 supported on Γ with sign sequence z are that Φ_Γ has full rank, where Φ_Γ is the $m \times S$ matrix consisting of the columns of Φ indexed by Γ, and that

$$|\pi(\gamma)| = |\langle(\Phi_\Gamma^*\Phi_\Gamma)^{-1}\Phi_\Gamma^*\varphi_\gamma, z\rangle| < 1 \quad \text{for all} \quad \gamma \in \Gamma^c, \tag{3.6}$$

where φ_γ is the column of Φ at index γ.

There are three essential steps in establishing (3.6):

1. Show that with probability exceeding $1 - O(n^{-1})$, the random matrix $\Phi_\Gamma^*\Phi_\Gamma$ will have a bounded inverse:

$$\|(\Phi_\Gamma^*\Phi_\Gamma)^{-1}\| \leq 2/m, \tag{3.7}$$

where $\|\cdot\|$ is the standard operator norm. This has already been done for us in the random subsampling case; it is essentially shown in [11, Th.1.2] that (3.7) will hold when

$$m \;\geq\; C \cdot \nu^2 \cdot \log n,$$

for a constant C. Combining this with the bound on ν from (3.4) means that (3.7) will hold when $m \geq C \cdot S \log n$. For the random demodulator, we will build on results from [40] (see Proposition 3.6 below), and (3.7) will follow, with the same probability, from a bound on the inner products between the columns of Φ (Lemma 3.7) for

$$m \;\geq\; C(S \log n + \log^3 n).$$

2. Establish, again with probability exceeding $1 - O(n^{-1})$, that the norm of each of the vectors $\Phi_\Gamma^* \varphi_\gamma$ is on the order of $m/\sqrt{\log n}$. For the random subsampling case, the is accomplished in Lemma 3.4. For the random demodulator, it will be another consequence of Lemma 3.7. Combined with step 2, this means that the norm of $(\Phi_\Gamma^* \Phi_\Gamma)^{-1} \varphi_\gamma$ is on the order of $1/\sqrt{\log n}$.
3. Given that $\|(\Phi_\Gamma^* \Phi_\Gamma)^{-1} \varphi_\gamma\|_2 \sim 1/\sqrt{\log n}$, show that with probability exceeding $1 - O(n^{-1})$, $|\pi(\gamma)| < 1$ for all $\gamma \in \Gamma^c$. Taking z as a random sign sequence, this is a straightforward application of the Hoeffding inequality.

3.3. Proof of Theorem 1.1. The bounds we will develop for the random subsampling case depend on only on the mutual coherence μ and norm coherence ν of the system U, allowing us to state a general sampling result, stated at Theorem 3.5 below, which may be of independent interest.

As step 1 is already established, we start with step 2. We will assume without loss of generality that $m \leq \nu^4/\mu^2$, as the probability that we will be able to recover a given signal increases monotonically with m. The following lemma shows that $\|\Phi_\Gamma^* \varphi_\gamma\|_2 \sim \nu/\sqrt{m}$ for each $\gamma \in \Gamma^c$.

LEMMA 3.4. *Let U be a $n \times n$ orthogonal matrix with $U^*U = nI$ with mutual coherence μ. Fix a subset Γ, let r^k be the rows of U_Γ, and set $\nu := \max_{k=1,\ldots,n} \|r^k\|_2$. Choose a subset Ω of the measurement domain of size $|\Omega| = m$, and let $\Phi = R_\Omega U$, the matrix constructed from the rows of U indexed by Ω. Fix $\gamma \in \Gamma^c$, and consider the random vector $\Phi_\Gamma^* \varphi_\gamma = U_\Gamma^* R_\Omega^* \varphi_\gamma$. Assume that $\sqrt{m} \leq \nu^2/\mu$. Then for any $a \leq 2m^{1/4}\mu^{-1/2}$,*

$$\mathrm{P}\left(\|\Phi_\Gamma^* \varphi_\gamma\|_2 \geq \nu\sqrt{m} + a\mu^{1/2} m^{1/4} \nu\right) \leq 3e^{-Ca^2},$$

where C is a universal constant.

Proof. We show that the mean $\mathrm{E}\,\|\Phi_\Gamma^* \varphi_\gamma\|_2$ is less than $\nu\sqrt{m}$, and then apply the Talagrand concentration inequality, (A.5) in the Appendix, to show that $\|\Phi_\Gamma^* \varphi_\gamma\|_2$ is concentrated about its mean.

We will denote the individual entries of the matrix U as $U_{k,\gamma}$. Using the Bernoulli sampling model, $\Phi_\Gamma^* \varphi_\gamma$ can be written as a sum of independent random vectors,

$$\Phi_\Gamma^* \varphi_\gamma = \sum_{k=1}^{n} \iota_k U_{k,\gamma} r^k = \sum_{k=1}^{n} (\iota_k - m/n) U_{k,\gamma} r^k,$$

where the second equality follows from the orthogonality of the columns of U_Γ. To

bound the mean, we use

$$\begin{aligned}
\mathrm{E}[\|\Phi_\Gamma^*\varphi_\gamma\|_2]^2 &\le \mathrm{E}[|\langle\Phi_\Gamma^*\varphi_\gamma, \Phi_\Gamma^*\varphi_\gamma\rangle|^2] \\
&= \sum_{k_1,k_2=1}^{n} \mathrm{E}[(\iota_{k_1} - m/n)(\iota_{k_2} - m/n)]U_{k_1,\gamma}U_{k_2,\gamma}\langle r^{k_1}, r^{k_2}\rangle \\
&= \sum_{k=1}^{n} \frac{m}{n}\left(1 - \frac{m}{n}\right) U_{k,\gamma}^2 \|r^k\|_2^2 \\
&\le \frac{m\nu^2}{n}\sum_{k=1}^{n} U_{k,\gamma}^2 \\
&= m\nu^2.
\end{aligned}$$

We will apply the concentration inequality (A.5) with $\eta_k = (\iota_k - m/n)U_{k,\gamma}$ and $v_k = r^k$. We have

$$B = \mu \cdot \max_k \|v_k\|_2 \le \mu\nu$$

and

$$\begin{aligned}
\bar\sigma^2 &= \mathrm{E}[|\eta_k|^2] \cdot \sup_{\|\xi\|_2\le 1} \sum_{k=1}^{n} |\langle\xi, v_k\rangle|^2 \\
&= \frac{m}{n}\left(1 - \frac{m}{n}\right)\mu^2 \cdot \sup_{\|\xi\|_2\le 1} \sum_{k=1}^{n} |\langle\xi, r^k\rangle|^2 \\
&\le m\mu^2.
\end{aligned}$$

Plugging the bounds for $\mathrm{E}\,\|\Phi_\Gamma^*\varphi_\gamma\|_2$, B, and $\bar\sigma^2$ into (A.5) yields

$$\mathrm{P}\left(\|\Phi_\Gamma^*\varphi_\gamma\|_2 > \nu\sqrt{m} + \lambda\right) \le 3\exp\left(-\frac{\lambda}{K\mu\nu}\log\left(1 + \frac{\mu\nu\lambda}{m\mu^2 + \mu\nu^2\sqrt{m}}\right)\right).$$

Using our assumption that $\sqrt{m} \le \nu^2/\mu$ and the fact that $\log(1+x) > 2x/3$ when $0 \le x \le 1$, this becomes

$$\mathrm{P}\left(\|\Phi_\Gamma^*\varphi_\gamma\|_2 > \nu\sqrt{m} + \lambda\right) \le 3\exp\left(-\frac{\lambda^2}{3K\mu\sqrt{m}\nu^2}\right)$$

for all $0 \le \lambda \le 2\nu\sqrt{m}$. Thus

$$\mathrm{P}\left(\|\Phi_\Gamma^*\varphi_\gamma\|_2 > \nu\sqrt{m} + a\mu^{1/2}m^{1/4}\nu\right) \le 3e^{-Ca^2}$$

for $a \le 2m^{1/4}\mu^{-1/2}$, where $C = 1/(3K)$.
□

To finish off the proof of the Theorem, let $\mathcal{A}$ be the event that (3.7) holds; step 1 tells us that $\mathrm{P}(\mathcal{A}^c) \le n^{-1}$ when $m \ge C\nu^2\log n$. Let $\mathcal{B}_\lambda$ be the event that

$$\max_{\gamma\in\Gamma^c} \|(\Phi_\Gamma^*\Phi_\Gamma)^{-1}\Phi_\Gamma^*\varphi_\gamma\|_2 \le \lambda.$$

By Lemma 3.4, and taking the union bound over all $\gamma \in \Gamma^c$, we have

$$\mathrm{P}(\mathcal{B}_\lambda^c \mid \mathcal{A}) \le 3ne^{-Ca^2}.$$

By the Hoeffding inequality,

$$\mathrm{P}\left(\max_{\gamma\in\Gamma^c}|\pi(\gamma)|>1 \mid \mathcal{B}_\lambda,\mathcal{A}\right)\le 2ne^{-1/2\lambda^2}.$$

Our final probability of success can then be bounded by

$$\begin{aligned}\mathrm{P}\left(\max_{\gamma\in\Gamma^c}|\pi(\gamma)|>1\right) &\le \mathrm{P}\left(\max_{\gamma\in\Gamma^c}|\pi(\gamma)|>1 \mid \mathcal{B}_\lambda,\mathcal{A}\right)+\mathrm{P}\left(\mathcal{B}_\lambda^c|\mathcal{A}\right)+\mathrm{P}(\mathcal{A}^c)\\ &\le 2ne^{-1/2\lambda^2}+3ne^{-Ca^2}+n^{-1}. \end{aligned} \tag{3.8}$$

Choose $\lambda = 2\nu m^{-1/2} + 2a\mu^{1/2}m^{-3/4}\nu$. Then we can make the second term in (3.8) less than n^{-1} by choosing $a = C^{-1/2}\sqrt{\log(3n^2)}$; for $m \ge 16C^{-2}\mu^2\log^2(3n^2)$ we will have $a \le (1/2)m^{1/4}\mu^{-1/2}$. This choice of a also ensures that $\lambda \le 3\nu m^{-1/2}$. For the first term in (3.8) to be less than n^{-1}, we need $\lambda^2 \le (2\log(2n^2))^{-1}$, which holds when

$$m \ge 18\cdot\nu^2\cdot\log(2n^2).$$

Theorem 1.1 is now a direct consequence (in combination with the coherence bounds (3.4) and (3.1)) of the following sampling theorem for general orthogonal systems U. The following statement, which we have proven above, can be viewed as an extension of the main result in [11].

THEOREM 3.5. *Let U, μ, ν, Γ and Ω be as in Lemma 3.4. Choose a sign sequence z on Γ uniformly at random, and set $\Phi = R_\Omega U$. Suppose that*

$$m \ge C(\nu^2\log n + \mu^2\log^2 n), \tag{3.9}$$

where C is a constant. Then with probability exceeding $1 - O(n^{-1})$, every vector α_0 supported on Γ with sign sequence z can be recovered from $y = \Phi\alpha_0$ by solving (1.2).

To finish of the proof of Theorem 1.1, we apply Theorem 3.5 with $U = H\Psi$. Lemmas 3.2 and Lemma 3.1 show that for this U, $\mu^2 \le C\log n$ and $\nu^2 \le C\cdot S$, both with probability exceeding $1-n^{-1}$. The bound (3.9) from Theorem 3.5 becomes

$$m \;\ge\; C(S\log n + \log^3 n),$$

which is the desired result.

3.4. Proof of Theorem 1.2. The following proposition is contained in [40, Thms. 11 and 13].

PROPOSITION 3.6. *Let U be an orthogonal matrix with $U^*U = nI$. Create a random demodulator matrix $P\Theta$ as in Section 1.2.2 with*

$$m \;\ge\; C_1\left(S\log n + \log^3 n\right),$$

for a fixed constant C_1, set $\Phi = P\Theta U$, and denote the columns of Φ as φ_γ. Let $\mathcal{A}$ be the event that both

$$\max_{1\le\gamma\le n}\;\left|\|\varphi_\gamma\|_2^2 - m\right| \;\le\; 0.15,$$

and

$$\max_{1\le\gamma,\alpha\le n}|\langle\varphi_\gamma,\varphi_\alpha\rangle| \;\le\; C_2 m^{1/2}\sqrt{\log n},$$

for a fixed constant C_2. Choose the support Γ uniformly at random as in Theorem 1.2. Then conditioned on $\mathcal{A}$,

$$\mathrm{P}\left(\|(\Phi_\Gamma^*\Phi_\Gamma)^{-1}\| \;\geq\; 2/m \mid \mathcal{A}\right) \;\leq\; n^{-1}$$

and

$$\mathrm{P}\left(\max_{\gamma\in\Gamma^c}\|\Phi_\Gamma^*\varphi_\gamma\|_2 \;\geq \frac{m}{4\sqrt{\log n}} \;\middle|\; \mathcal{A}\right) \;\leq\; n^{-1}.$$

The key to establishing Theorem 1.2 is to estimate the probability that the event $\mathcal{A}$ occurs. This is accomplished with the following lemma.

LEMMA 3.7. *Let H be a random convolution matrix and Ψ be a fixed orthosystem. Set $U = H\Psi$, and assume that Lemma 3.3 holds. Draw a random demodulator system with*

$$m \;\geq\; C_1\mu^2\log(2n^2/\delta),$$

for a constant C_1, and set $\Phi = P\Theta U$. Then with probability exceeding $1-\delta$,

$$\max_{\gamma,\alpha}|\langle\varphi_\gamma,\varphi_\alpha\rangle| \;\leq\; C_2 m^{1/2}\sqrt{\log(2n^2/\delta)},$$

and

$$\max_{\gamma}\left|\|\varphi_\gamma\|_2^2 - m\right| \;\leq\; C_2 m^{1/2}\sqrt{\log(2n^2/\delta)},$$

for a constant C_2.

Proof. The quantities of interest are the magnitudes of the random variables

$$Z_{\gamma,\alpha} = \frac{m}{n}\sum_{k=1}^{m}\sum_{\substack{p_1,p_2\in B_k\\ p_1\neq p_2}}\epsilon_{p_1}\epsilon_{p_2}U_{p_1,\gamma}U_{p_2,\alpha},$$

for all $1\leq\gamma,\alpha\leq n$, where B_k is the index set for the kth sample of the random demodulator, as in (1.7). For $\gamma\neq\alpha$, $Z_{\gamma,\alpha} = \langle\varphi_\gamma,\varphi_\alpha\rangle$, and for $\gamma=\alpha$, $Z_{\gamma,\alpha} = \|\varphi_\gamma\|_2^2 - m$. The quantity $Z_{\gamma,\alpha}$ is a *Rademacher chaos* — it is a sum of random variables that are not independent. Nevertheless, strong concentration inequalities for this type of sum have been developed; in particular, we will apply (A.3) from the Appendix. To apply (A.3), we first need to make the sum symmetric:

$$Z_{\gamma,\alpha} = \frac{m}{2n}\sum_{k=1}^{m}\sum_{\substack{p_1,p_2\in B_k\\ p_1\neq p_2}}\epsilon_{p_1}\epsilon_{p_2}(U_{p_1,\gamma}U_{p_2,\alpha,} + U_{p_2,\gamma}U_{p_1,\alpha})$$

For convenience, we will sometimes use the notation $a_{p_1,p_2} = U_{p_1,\gamma}U_{p_2,\alpha} + U_{p_2,\gamma}U_{p_1,\alpha}$.

First we bound the variance of $Z_{\gamma,\alpha}$, which we will use to bound the mean magnitude:

$$\mathrm{E}[Z_{\gamma,\alpha}^2] = \frac{m^2}{4n^2}\sum_{k=1}^{m}\sum_{\ell=1}^{m}\sum_{\substack{p_1,p_2\in B_k\\ p_1\neq p_2}}\sum_{\substack{q_1,q_2\in B_\ell\\ q_1\neq q_2}}E[\epsilon_{p_1}\epsilon_{p_2}\epsilon_{q_1}\epsilon_{q_2}]a_{p_1,p_2}a_{q_1,q_2}.$$

Since we have the constraints $p_1 \neq p_2$ and $q_1 \neq q_2$, the terms in the sum above are nonzero only when $k = \ell$ and

$$(p_1 = q_1 \text{ and } p_2 = q_2) \quad \text{or} \quad (p_1 = q_2 \text{ and } p_2 = q_1)$$

In both of these cases, $\mathrm{E}[\epsilon_{p_1}\epsilon_{p_2}\epsilon_{q_1}\epsilon_{q_2}] = 1$ and $a_{p_1,p_2}a_{q_1,q_2} = a_{p_1,p_2}^2$, since $a_{p_1,p_2} = a_{p_2,p_1}$. Thus

$$\mathrm{E}[Z_{\gamma,\alpha}^2] = \frac{m^2}{2n^2}\sum_{k=1}^{m}\sum_{\substack{p_1,p_2\in B_k\\ p_1\neq p_2}} U_{p_1,\gamma}^2U_{p_2,\alpha}^2 + 2U_{p_1,\gamma}U_{p_2,\alpha}U_{p_2,\gamma}U_{p_1,\alpha} + U_{p_1,\alpha}^2U_{p_2,\gamma}^2.$$

We will bound each of the three terms in the sum above in turn.

Assuming that (3.5) holds, we can bound the first and third terms using

$$\begin{aligned}\frac{m^2}{2n^2}\sum_{k=1}^{m}\sum_{\substack{p_1,p_2\in B_k\\ p_1\neq p_2}} U_{p_1,\gamma}^2U_{p_2,\alpha}^2 &\leq \frac{m^2}{2n^2}\sum_{k=1}^{m}\sum_{p_1\in B_k} U_{\gamma,p_1}^2\sum_{p_2\in B_k} U_{p_2,\alpha}^2\\ &\leq \frac{8m}{n}\sum_{k=1}^{m}\sum_{p_1\in B_k} U_{p_1,\gamma}^2\\ &= 8m\end{aligned}$$

For the cross term

$$\begin{aligned}\frac{m^2}{2n^2}\sum_{k=1}^{m}\sum_{\substack{p_1,p_2\in B_k\\ p_1\neq p_2}} 2U_{p_1,\gamma}U_{p_2,\alpha}U_{p_2,\gamma}U_{p_1,\alpha} &\leq \frac{m^2}{n^2}\sum_{k=1}^{m}\sum_{p_1\in B_k} |U_{p_1,\gamma}U_{p_1,\alpha}|\sum_{p_2\in B_k} |U_{p_2,\alpha}U_{p_2,\gamma}|\\ &\leq \frac{8m}{n}\sum_{k=1}^{m}\sum_{p_1\in B_k} |U_{p_1,\gamma}U_{p_1,\alpha}|\\ &\leq 8m.\end{aligned}$$

Collecting these results gives us

$$\mathrm{E}[|Z_{\gamma,\alpha}|] \;\leq\; \left(\mathrm{E}[|Z_{\gamma,\alpha}|^2]\right)^{1/2} \;=\; \left(\frac{m^2}{2n^2}\sum_{k=1}^{m}\sum_{p_1\in B_k}\sum_{\substack{p_2\in B_k\\ p_1\neq p_2}} a_{p_1,p_2}^2\right)^{1/2} \;\leq\; \sqrt{24m}. \quad (3.10)$$

To apply the concentration bound (A.3), we need to estimate the variances B and $\bar{\sigma}^2$. The variance B is simply $B = \max_{x\in\{-1,1\}} \|M_1(x)\|$, where M_1 is an $n \times n$ block diagonal matrix whose entries are given by

$$\begin{aligned}[M_1(x)]_{p_1,p_2} &= x\frac{m}{n}a_{p_1,p_2}\\ \text{for } \; p_1,p_2 \in B_k, p_1 \neq p_2, \; k &= 1,\ldots,m.\end{aligned}$$

The spectral norm of $M_1(x)$ can be bounded by its absolute row sum. Each row of

$M_1(x)$ contains at most n/m nonzero terms, and so

$$\begin{aligned}\|M_1(x)\| &\le \frac{m}{n}\max_{k,p_1}\sum_{p_2\in B_k}|a_{p_1,p_2}| \\ &\le \frac{m\mu}{n}\max_{k,p_1}\sum_{p_2\in B_k}|U_{p_2,\alpha}|+|U_{p_2,\gamma}| \\ &\le 4\sqrt{2}\mu\end{aligned}$$

since $\sum_{p_2\in B_k}|U_{p_2,\alpha}|\le\sqrt{8}n/m$ by (3.5) and Cauchy-Schwarz.

Next we develop an estimate for $\bar{\sigma}$:

$$\begin{aligned}\bar{\sigma} &= \mathrm{E}\left(\frac{m^2}{4n^2}\sum_{k=1}^{m}\sum_{p_1\in B_k}\left|\sum_{\substack{p_2\in B_k\\ p_2\neq p_1}}\epsilon_{p_2}a_{p_1,p_2}\right|^2\right)^{1/2} \\ &\le \left(\frac{m^2}{4n^2}\sum_{k=1}^{m}\sum_{p_1\in B_k}\mathrm{E}\left|\sum_{p_2\in B_k}\epsilon_{p_2}a_{p_1,p_2}\right|^2\right)^{1/2} \\ &\le \left(\frac{m^2}{4n^2}\sum_{k=1}^{m}\sum_{p_1\in B_k}\sum_{p_2\in B_k}a_{p_1,p_2}^2\right)^{1/2} \\ &\le \sqrt{12m},\end{aligned}$$

where we have re-used the bound (3.10) for the last inequality.

We now apply (A.3):

$$P\left(|Z_{\gamma,\alpha}|>4\sqrt{12m}+\lambda\right) \;\le\; 2\exp\left(-\frac{1}{K}\min\left(\frac{\lambda^2}{12m},\frac{\lambda}{4\sqrt{2}\mu}\right)\right).$$

Take $\lambda=\sqrt{12Km\log(2n^2/\delta)}$. When

$$m\ge 8K\mu^2\log(2n^2/\delta),$$

we will have $\lambda^2/8m<\lambda/4\sqrt{2}\mu$, and

$$\mathrm{P}\left(|Z_{\alpha,\gamma}|>Cm^{1/2}\log^{1/2}(n^2/\delta)\right) \;\le\; \delta/n^2$$

Applying the union bound over all n^2 choices of γ and α establishes the lemma. □

We now collect results to finish of the proof of Theorem 1.2. After creating the random convolution H, taking $\delta=n^{-1}$ in Lemmas 3.1 and 3.3 tells us that with probability at least $1-2n^{-1}$, the coherence of $U=H\Psi$ is controlled by

$$\mu^2 \;\le\; C\log n$$

and that the norm of each of the columns $U=H\Psi$ restricted to each random demodulator block is controlled by (3.5) for $m\le Cn/\log n$. With these norms controlled Lemma 3.7 shows that with probability at least $1-n^{-1}$, if $m\ge C\log^2 n$ then inner products between the columns of Φ can be bounded by

$$|\langle\varphi_\gamma,\varphi_\alpha\rangle| \;\le\; C\sqrt{m\log n}\quad\text{for all}\;\;\gamma,\alpha=1,\ldots,n \tag{3.11}$$

and the norms of the columns of Φ are close to m:

$$\left| \|\varphi_\gamma\|_2^2 - m \right| \;\leq\; C\sqrt{m \log n} \quad \text{for all} \;\; \gamma = 1, \ldots, n. \tag{3.12}$$

With (3.11) and (3.12) in force, we apply Proposition 3.6, which guarantees us (to within probability $2n^{-1}$) that the norms of the vectors $v_\gamma = (\Phi_\Gamma^* \Phi_\Gamma)^{-1} \Phi_\Gamma^* \varphi_\gamma$ are controlled:

$$\|v_\gamma\|_2^2 \;\leq\; \frac{1}{4 \log n} \quad \text{for all} \;\; \gamma = 1, \ldots, n.$$

Since $\pi(\gamma) = \langle v_\gamma, z \rangle$, the Hoeffding inequality (A.1) tells us that conditioned on all the events above,

$$|\pi(\gamma)| \;<\; 1 \quad \text{for all} \;\; \gamma = 1, \ldots, n,$$

with probability exceeding $2n^{-1}$. Thus the overall probability of success exceeds $1 - 7n^{-1}$, and the theorem is established.

3.5. Proof of Theorems 1.3 and 1.4. The following theorem in contained in [35, Thm. 3.3], and will immediately, in conjunction with the coherence bound from Lemma 3.1, establish the first part of Theorem 1.3.

THEOREM 3.8 (Rudelson and Vershynin). *Let U be an orthogonal matrix, $U^*U = nI$ with coherence μ, and set $\Phi = R_\Omega U$, where the sample set Ω has size $|\Omega| = m$ and is chosen uniformly at random. For*

$$m \;\geq\; C\rho^{-2} \mu^2 S \log^5 n,$$

for a universal constant C, the matrix Φ will have restricted isometry constants

$$\rho_{2S} \;=\; \max_{|\Gamma| = 2S} \|m^{-1} \Phi_\Gamma^* \Phi_\Gamma - I\| \;\leq\; \rho$$

with probability exceeding $1 - O(n^{-1})$.

Lemma 3.1 delivers the bound $\mu^2 \leq C \log n$ with probability $1 - O(n^{-1})$, and so we will have $\rho_{2S} \leq \rho$ for

$$m \;\geq\; C\rho^{-2} S \log^6 n.$$

The small-S case follows directly from the following lemma.

LEMMA 3.9. *Let $\Phi = R_\Omega H \Psi$, as in Theorem 1.1, with*

$$m \;\geq\; C_1 \mu^2 \log(2n^2/\delta),$$

for a constant C_1. Then with probability exceeding $1 - \delta$,

$$\max_{\gamma, \alpha} |\langle \varphi_\gamma, \varphi_\alpha \rangle| \;\leq\; C_2 \mu m^{1/2} \sqrt{\log(2n^2/\delta)}$$

and

$$\max_{\gamma} \;\left| \|\varphi_\gamma\|_2^2 - m \right|_2^2 \;\leq\; C_2 \mu m^{1/2} \sqrt{\log(2n^2/\delta)},$$

for a constant C_2.

Proof. Let $U = H\Psi$. As in Lemma 3.7 above, we are interested in a uniform bound on the random variables

$$Z_{\alpha,\gamma} = \sum_{k=1}^{n} (\iota_k - m/n) U_{k,\gamma} U_{k,\alpha},$$

since $Z_{\alpha,\gamma} = \langle \varphi_\gamma, \varphi_\alpha \rangle$ for $\gamma \neq \alpha$ and $Z_{\gamma,\gamma} = \|\varphi_\gamma\|_2^2 - m$. To do this, we will apply the Bernstein inequality (A.2) from the Appendix. Each $Z_{\alpha,\gamma}$ is a sum of independent random variables

$$Y_k = (\iota_k - m/n) U_{k,\gamma} U_{k,\alpha}$$

with

$$|Y_k| \leq B = \mu^2 \quad \text{and} \quad \sum_{k=1}^{n} \mathrm{E}[Y_k^2] = \frac{m}{n}(1 - m/n) \sum_{k=1}^{n} U_{k,\gamma}^2 U_{k,\alpha}^2 \leq m\mu^2.$$

Invoking (A.2), we have

$$\mathrm{P}\left(|Z_{\alpha,\gamma}| > \lambda\right) \;\leq\; 2\exp\left(-\frac{\lambda^2}{2(m\mu^2 + \mu^2\lambda/3)}\right).$$

Choose $\lambda = (8/3 \cdot \mu^2 m \log(2n^2/\delta))^{1/2}$. Then for $m \geq 8/3 \cdot \mu^2 \log(2n^2/\delta)$,

$$\mathrm{P}\left(|Z_{\alpha,\gamma}| > (8/3 \cdot \mu^2 m \log(2n^2/\delta))^{1/2}\right) \;\leq\; \delta/n^2.$$

The lemma follows by taking the union bound over all n^2 choices of γ, α. □

Lemma 3.9 establishes a uniform bound on the entries of the matrix $(\Phi^*\Phi - mI)$. For any set Γ of size S, we can bound the spectral norm $\|\Phi_\Gamma^*\Phi_\Gamma - mI\|$ by the maximum row sum. Applying this element-wise bound with the bound (3.1) for the coherence with $\delta = n^{-1}$ tells us that

$$\max_{|\Gamma| \leq 2S} \|\Phi_\Gamma^*\Phi_\Gamma - mI\| \;\leq\; C \cdot \sqrt{m} \cdot S \log n$$

will hold with probability exceeding $1 - n^{-1}$. Thus

$$\rho_{2S} = \max_{|\Gamma| \leq 2S} \|m^{-1}\Phi_\Gamma^*\Phi_\Gamma - I\| \;\leq\; C \cdot \frac{S \log n}{\sqrt{m}} \;\leq\; \rho$$

when

$$m \;\geq\; C \cdot \rho^{-2} \cdot S^2 \log^2 n.$$

The proof of Theorem 1.4 is similar. We will work under the assumption that (3.5) holds for our $U = H\Psi$, Lemma 3.3 tells us that this will be true with probability exceeding $1 - O(n^{-1})$. In [40], the technology in [35] for estimating the restricted isometry constants is adapted to random demodulator matrices $\Phi = P\Theta U$, where U is a Fourier matrix. The core of the argument, as formalized in the proposition below, depends only on estimating the maximum magnitude of the entries in Φ. The following is contained in [40, Thm. 16].

PROPOSITION 3.10. *Let U be an orthogonal matrix, $U^*U = nI$, and create the random demodulator matrix $\Phi = P\Theta U$. Let the random variable B be the maximum absolute entry in Φ,*

$$B = \max_{k,\gamma} |\Phi_{k,\gamma}|.$$

If

$$\mathrm{E}[B^2] \le C_1 \log n \qquad \textit{and} \qquad \mathrm{P}\left(B > C_2\sqrt{\log n}\right) \;\le\; n^{-1} \tag{3.13}$$

then for

$$m \;\ge\; C_3 \cdot \rho^{-2} \cdot S \log^6 n \tag{3.14}$$

Φ will have restricted isometry constants $\rho_{2S} \le \rho$ with probability exceeding $1 - O(n^{-1})$.

It remains, then, to show that the estimates for B in (3.13) hold for random convolution systems $U = H\Psi$. We will start with the tail bound. We can write an entry in Φ as

$$\Phi_{k,\gamma} = \sqrt{\frac{m}{n}} \sum_{p \in B_k} \epsilon_p U_{p,\gamma}.$$

The Hoeffding inequality (A.1) immediately yields

$$\mathrm{P}\left(|\Phi_{k,\gamma}| > \lambda\right) \;\le\; 2\exp\left(-\frac{\lambda^2}{2(m/n)\sum_{p\in B_k} U_{p,\gamma}^2}\right).$$

Lemma 3.3 tells us that with probability $1 - O(n^{-1})$, $\sum_{p\in B_k} U_{p,\gamma}^2 \le 8n/m$, so applying the union bound over all mn choices of k and γ gives us

$$\mathrm{P}(B > \lambda) \;\le\; 2mne^{-\lambda^2/16} \;\le\; 2n^2 e^{-\lambda^2/16}. \tag{3.15}$$

Choosing $\lambda = 4\sqrt{\log(2n^3)}$ gives us the desired tail bound:

$$\mathrm{P}\left(B > 4\sqrt{\log(2n^3)}\right) \;\le\; n^{-1}.$$

We can leverage (3.15) to estimate the second moment of B. Let $t_k = \sqrt{32k\log n}$. With $\int dB$ indicating integration with respect to the probability measure that governs B, we can take apart the second moment calculation as

$$\begin{aligned}
\mathrm{E}[B^2] = \int_0^\infty b^2 dB &= \int_0^{t_1} b^2 dB \;+\; \sum_{k=1}^{\infty} \int_{t_k}^{t_{k+1}} b^2 dB \\
&\le t_1^2 + \sum_{k=1}^{\infty} t_{k+1}^2 \,\mathrm{P}(B > t_k) \\
&\le 32\log n + 64n^2 \log n \sum_{k=1}^{\infty} (k+1) n^{-2k} \\
&\le C_1 \log n
\end{aligned}$$

where the last inequality follows from the fact that $\sum_{k\geq 1}(k+1)n^{-2k} \leq \text{Const} \cdot n^{-2}$. Thus the estimates in (3.13) are in force, and Φ has the desired restricted isometry constants for m as in (3.14).

The small-S case in Theorem 1.4 follows directly from Lemma 3.7. As in the random subsampling case, we can bound each of the spectral norms $\|\Phi_\Gamma^*\Phi_\Gamma - mI\|$ by their maximum row sum. The matrices $\Phi_\Gamma^*\Phi_\Gamma - mI$ are just submatrices of $(\Phi^*\Phi - mI)$, and Lemma 3.7 tells us that with probability $1-n^{-1}$, none of the entries in $(\Phi^*\Phi - mI)$ will exceed $\sim \sqrt{m \log n}$. Thus

$$\rho_{2S} = \max_{|\Gamma|\leq 2S} \|m^{-1}\Phi_\Gamma^*\Phi_\Gamma - I\| \;\leq\; C \cdot \frac{S\sqrt{\log n}}{\sqrt{m}} \;\leq\; \rho$$

when

$$m \;\geq\; C \cdot \rho^{-2} \cdot S^2 \log n.$$

Appendix A. Concentration inequalities.

Almost all of the analysis in this paper relies on controlling the magnitude/norm of the sum of a sequence of random variables/vectors. In this appendix, we briefly outline the concentration inequalities that we use in the proofs of Theorems 1.1, 1.2, 1.3, and 1.4.

The Hoeffding inequality [21] is a classical tail bound on the sum of a sequence of real-valued independent random variables. Let $Y_1, Y_2, \ldots, Y_n$ be independent, zero-mean random variables bounded by $|Y_k| \leq a_k$, and let the random variable Z be

$$Z = \left| \sum_{k=1}^{n} Y_k \right|.$$

Then

$$\mathrm{P}\left(Z > \lambda\right) \leq 2\exp\left(-\frac{\lambda^2}{2\|a\|_2^2}\right), \tag{A.1}$$

for every $\lambda > 0$. For random variables that have much smaller standard deviation than maximum value, a slightly more refined bound is given by the Bernstein inequality [7]. If $|Y_k| \leq B$, then

$$\mathrm{P}\left(Z > \lambda\right) \leq 2\exp\left(-\frac{\lambda^2}{2(\sigma^2 + B\lambda/3)}\right), \tag{A.2}$$

where the variance σ^2 is given by $\sigma^2 = \sum_{k=1}^{n} \mathrm{E}[Y_k^2]$. A nice derivation of both the Hoeffding and Bernstein inequalities can be found in the introduction of [9].

Lemma 3.7 depends on bounding a sum of random variables of the form

$$Z = \left| \sum_{i,j} \epsilon_i \epsilon_j v_{ij} \right|$$

where the ϵ_i are independent ± 1 Bernoulli random variables and the v_{ij}, $i, j = 1, \ldots, n$ are real numbers with $v_{ij} = v_{ji}$ and $v_{ii} = 0$. This type of sum is called a *Rademacher chaos*, and its magnitude can be controlled in much the same manner as in (A.2).

In [36, Thm. 1.2], it is shown that Z will not be too much greater than twice its mean:

$$\mathrm{P}\left(Z > 2\,\mathrm{E}[Z] + \lambda\right) \;\leq\; 2\exp\left(-\frac{1}{K}\min\left(\frac{\lambda^2}{\bar{\sigma}^2}, \frac{\lambda}{B}\right)\right), \tag{A.3}$$

where K is an absolute constant, and $B, \bar{\sigma}^2$ are variances defined by spectral norms of matrices constructed from the v_{ij}. The variance B is given by

$$B = \sup_{|x|=1} \|M_1(x)\|,$$

where $M_1(x)$ is the matrix with xv_{ij} as the entry in the ith row and jth column, and $\bar{\sigma}$ is

$$\bar{\sigma} = \mathrm{E}\left(\sum_j \left(\sum_i \epsilon_i v_{ij}\right)^2\right)^{1/2} = \mathrm{E}\,\|M_2\|,$$

where M_2 is the n vector whose jth entry is $\sum_i \epsilon_i v_{ij}$.

Concentration inequalities analogous to (A.1) exist for the norm of a random sum of vectors. Let $v_1, v_2, \ldots, v_n$ be a fixed sequence of vectors in $\mathbb{R}^S$, and again let $\epsilon_1, \ldots, \epsilon_n$ be a sequence of independent random variables taking values of ± 1 with equal probability. Let the random variable Z be the norm of the randomized sum

$$Z = \left\|\sum_{k=1}^{n} \epsilon_k v_k\right\|_2.$$

If we create the $S \times n$ matrix V by taking the v_i as columns, Z is the norm of the result of the action of V on the vector $[\epsilon_1 \;\cdots\; \epsilon_n]^T$. The second moment of Z is easily computed

$$\begin{aligned}
\mathrm{E}[Z^2] &= \sum_{k_1}\sum_{k_2} \mathrm{E}[\epsilon_{k_1}\epsilon_{k_2}]\langle v_{k_1}, v_{k_2}\rangle \\
&= \sum_k \|v_k\|_2^2 \\
&= \|V\|_F^2,
\end{aligned}$$

where $\|\cdot\|_F$ is the Frobenius norm. This gives us an easy bound on the mean of Z:

$$\mathrm{E}[Z] \;\leq\; \sqrt{E[Z^2]} = \|V\|_F.$$

We can also quantify how tightly Z is concentrated around its mean using the maximum singular value (operator norm) of V. In particular, [24, Thm. 7.3] shows that for all $\lambda > 0$

$$\mathrm{P}\left(Z \geq \mathrm{E}[Z] + \lambda\right) \;\leq\; 2\exp\left(-\frac{\lambda^2}{16\sigma^2}\right), \tag{A.4}$$

where the "variance" σ^2 is simply the largest squared singular value of V:

$$\sigma^2 = \sup_{\|\xi\|_2 \leq 1} \sum_{i=1}^{n} |\langle \xi, v_i\rangle|^2 = \|V\|^2.$$

When the random weights in the vector sum have a variance which is much smaller than their maximum possible magnitude (as is the case in Lemma 3.4), an even tighter bound is possible. Now let Z be the norm of the random sum

$$Z = \left\| \sum_{k=1}^{n} \eta_k v_k \right\|_2,$$

where the η_k are zero-mean iid random variables with $|\eta_k| \leq \mu$. A result due to Talagrand [36, Thm. 1.4] gives

$$\mathrm{P}\left(Z \geq \mathrm{E}[Z] + \lambda\right) \;\leq\; 3\exp\left(-\frac{\lambda}{KB}\log\left(1 + \frac{B\lambda}{\bar{\sigma}^2 + B\,\mathrm{E}[Z]}\right)\right), \tag{A.5}$$

where K is a fixed numerical constant,

$$\bar{\sigma}^2 = \mathrm{E}[|\eta_k|^2] \cdot \|V\|^2, \quad \text{and} \quad B = \mu \cdot \max_k \|v_k\|_2.$$

REFERENCES

[1] N. Ailon and B. Chazelle, *Approximate nearest neighbors and the fast Johnson-Lindenstrauss transform*, in Proc. 38th ACM Symp. Theory of Comput., Seattle, WA, 2006, pp. 557–563.

[2] S. R. J. Axelsson, *Analysis of random step frequency radar and comparison with experiments*, IEEE Trans. Geosci. Remote Sens., 45 (2007), pp. 890–904.

[3] ———, *Random noise radar/sodar with ultrawideband waveforms*, IEEE Trans. Geosci. Remote Sens., 45 (2007), pp. 1099–1114.

[4] W. U. Bajwa, J. D. Haupt, G. M. Raz, S. J. Wright, and R. D. Nowak, *Toeplitz-structured compressed sensing matrices*, in Proc. IEEE Stat. Sig. Proc. Workshop, Madison, WI, August 2007, pp. 294–298.

[5] R. Baraniuk and P. Steeghs, *Compressive radar imaging*, in Proc. IEEE Radar Conference, Boston, MA, April 2007, pp. 128–133.

[6] R. G. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, *A simple proof of the restricted isometry property for random matrices*, Constructive Approximation, 28 (2008), pp. 253–263.

[7] S. N. Bernstein, *The Theory of Probabilities*, Gastehizdat Publishing House, Moscow, 1946.

[8] T. Blumensath and M. Davies, *Iterative hard thresholding for compressed sensing*, Submitted manuscript, (2008).

[9] S. Boucheron, O. Bousquet, and G. Lugosi, *Concentration inequalities*, in Advanced Lectures in Machine Learning, O. Bousquet and G. R atsch, eds., Springer, 2004, pp. 208–240.

[10] E. Candès, *The restricted isometry property and its implications for compressed sensing*, C. R. Acad. Sci. Paris, Ser. I, 346 (2008), pp. 589–592.

[11] E. Candès and J. Romberg, *Sparsity and incoherence in compressive sampling*, Inverse Problems, 23 (2007), pp. 969–986.

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

[13] ———, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. on Pure and Applied Math., 59 (2006), pp. 1207–1223.

[14] E. Candès and T. Tao, *Near-optimal signal recovery from random projections and universal encoding strategies?*, IEEE Trans. Inform. Theory, 52 (2006), pp. 5406–5245.

[15] S. S. Chen, D. L. Donoho, and M. A. Saunders, *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comput., 20 (1999), pp. 33–61.

[16] R. Coifman, F. Geshwind, and Y. Meyer, *Noiselets*, Appl. Comp. Harmonic Analysis, 10 (2001), pp. 27–44.

[17] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.

[18] J. J. Fuchs, *Recovery of exact sparse representations in the presence of bounded noise*, IEEE Trans. Inform. Theory, 51 (2005), pp. 3601–3608.

[19] J. Haupt, W. Bajwa, G. Raz, and R. Nowak, *Toeplitz compressed sensing matrices with applications to sparse channel estimation*, Submitted to IEEE Trans. Inform. Theory, (2008).

[20] M. A. Herman and T. Strohmer, *High-resolution radar via compressed sensing*. Submitted to IEEE. Trans. Sig. Proc., 2008.
[21] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, J. American Stat. Assoc., 58 (1963), pp. 13–30.
[22] S. Kirolos, J. Laska, M. Wakin, M. Duarte, D. Baron, T. Ragheb, Y. Massoud, and R. Baraniuk, *Analog-to-information conversion via random demodulation*, in Proc. IEEE Dallas/CAS Workshop on Design, Applications, Integration, and Software, Richardson, TX, October 2006, pp. 71–74.
[23] B. Le, T. W. Rondeau, J. H. Reed, and C. W. Bostian, *Analog-to-digital converters*, IEEE Signal Proc. Mag., 22 (2005), pp. 69–77.
[24] M. Ledoux, *The Concentration of Measure Phenomenon*, American Mathematical Society, 2001.
[25] S. G. Mallat and Z. Zhang, *Matching pursuits with time-frequency dictionaries*, IEEE Trans. Signal Proc., 41 (1993), pp. 3397–3415.
[26] I. Maravic, M. Vetterli, and K. Ramchandran, *Channel estimation and synchornization with sub-nyquist sampling and application to ultra-wideband systems*, in IEEE Symp. on Circuits and Systems, vol. 5, Vancouver, May 2004, pp. 381–384.
[27] R. F. Marcia, Z. T. Harmany, and R. M. Willet, *Compressive coded aperture imaging*, in Proc. SPIE Conference on Computational Imaging VII, January 2009.
[28] R. F. Marcia and R. M. Willet, *Compressive coded aperture superresolution image reconstruction*, in Proc. IEEE Int. Conf. Acoust. Speech Sig. Proc., April 2008, pp. 833–836.
[29] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, *Reconstruction and subgaussian operators in asymptotic geometric analysis*, Geometric and Functional Analysis, 17 (2007), pp. 1248–1282.
[30] D. C. Munson, J. D. O'Brien, and W. K. Jenkins, *A tomographic formulation of spotlight-mode synthetic aperure radar*, Proc. IEEE, 71 (1983), pp. 917–925.
[31] D. Needell and J. Tropp, *COSAMP: Iterative signal recovery from incomplete and inaccurate measurements*, Appl. and Comp. Harmonic Analysis, 26 (2009), pp. 301–321.
[32] G. E. Pfander, H. Rauhut, and J. Tanner, *Identification of matrices having a sparse representation*, IEEE Trans. Signal Proc., 56 (2008), pp. 5376–5388.
[33] H. Rauhut, *Circulant and Toeplitz matrices in compressed sensing*, in SPARS'09 - Signal Processing with Adaptive Sparse Structured Representations, Rémi Gribonval, ed., Saint Malo France, 2009, Inria Rennes - Bretagne Atlantique.
[34] M. Richards, *Fundamentals of Radar Signal Processing*, McGraw-Hill, 2005.
[35] M. Rudelson and R. Vershynin, *On sparse reconstruction from Fourier and Gaussian measurements*, Comm. on Pure and Applied Math., 61 (2008), pp. 1025–1045.
[36] M. Talagrand, *New concentration inequalities in product spaces*, Invent. Math., 126 (1996), pp. 505–563.
[37] J. Tropp, *Norms of random submatrices and sparse approximation*, C. R. Acad. Sci. Paris, Ser. I, 346 (2008), pp. 1271–1274.
[38] J. Tropp and A. Gilbert, *Signal recovery from partial information via orthogonal matching pursuit*, IEEE Trans. Inform. Theory, 53 (2007), pp. 4655–4666.
[39] J. A. Tropp, *Recovery of short, complex linear combinations via ℓ_1 minimization*, IEEE Trans. Inform. Theory, 51 (2005), pp. 1568–1570.
[40] J. A. Tropp, J. N. Laska, M. F. Duarte, J. Romberg, and R. G. Baraniuk, *Beyond Nyquist: efficient sampling of sparse bandlimited signals*. Submitted to IEEE. Trans. Inform. Theory, February 2009.
[41] J. A. Tropp, M. B. Wakin, M. F. Duarte, D. Baron, and R. G. Baraniuk, *Random filters for compressive sampling and reconstruction*, in Proc. IEEE Int. Conf. Acoust. Speech Sig. Proc., Toulouse, France, May 2006.
[42] M. Vetterli, P. Marziliano, and T. Blu, *Sampling signals with finite rate of innovation*, IEEE Trans. Signal Proc., 50 (2002), pp. 1417–1428.
[43] R. H. Walden, *Analog-to-digital converter survey and analysis*, IEEE Journal of Select Areas in Communications, 17 (1999), pp. 539–550.