

COMPRESSIVE SENSING BY RANDOM CONVOLUTION

JUSTIN ROMBERG*

Abstract. This paper outlines a new framework for compressive sensing: convolution with a random waveform followed by random time domain subsampling. We show that sensing by random convolution is a universally efficient data acquisition strategy in that an n -dimensional signal which is S sparse in any fixed representation can be recovered from $m \gtrsim S \log n$ measurements. We discuss two imaging scenarios — radar and Fourier optics — where convolution with a random pulse allows us to seemingly 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 [7, 8, 10, 15, 32]: the number of measurements we need to reconstruct a signal depends on its *sparsity* rather than its bandwidth. These measurements, however, are different than the samples that traditional analog-to-digital converters take. Instead of simple point evaluations, we model the acquisition of a signal x_0 as a series of inner products against different waveforms ϕ_1, \dots, ϕ_m :

$$y_k = \langle \phi_k, x_0 \rangle, \quad k = 1, \dots, m. \quad (1.1)$$

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 results from CS have shown us that recovering x_0 is possible even when there are far fewer measurements than unknowns, $m \ll n$.

The notion of sparsity is critical to compressive sensing — it is the essential measure of signal complexity, and plays roughly the same role in CS that bandwidth plays in the classical Shannon-Nyquist theory. When we say a signal is sparse it always in the context of an orthogonal signal representation. With Ψ as an $n \times n$ orthogonal representation matrix (the columns of Ψ are the basis vectors), 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. In most applications, our signals of interest are not perfectly sparse; a more appropriate model is that they can be accurately approximated using a small number of terms. That is, there is a transform vector $\alpha_{0,S}$ with only S terms such that $\|\alpha_{0,S} - \alpha_0\|_2$ is small. This will be true if the transform coefficients decay like a power law — such signals are called *compressible*.

Sparsity is what makes it possible to recover a signal from undersampled data. Several methods for recovering sparse x_0 from a limited number of measurements have been proposed [12, 16, 29]. Here we will focus on recovery via ℓ_1 minimization. Given the measurements $y = \Phi x_0$, we solve the convex optimization program

$$\min_{\alpha} \|\alpha\|_{\ell_1} \quad \text{subject to} \quad \Phi \Psi \alpha = y. \quad (1.2)$$

* 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.

Submitted to the SIAM Journal on Imaging Science on July 9, 2008.

In words, the above searches for the set of transform coefficients α such that the measurements of the corresponding signal $\Psi\alpha$ agree with y . The ℓ_1 norm is being used to gauge the sparsity of candidate signals.

The number of measurements (rows in Φ) we need to take for (1.2) to succeed depends on the nature of the waveforms ϕ_k . To date, results in CS fall into one of two types of measurement scenarios. Randomness plays a major role in each of them.

1. *Random waveforms.* In this scenario, the shape of the measurement waveforms is random. The ensemble Φ is generated by drawing each element independently from a subgaussian [6, 10, 15, 24] distribution. The canonical examples are generating each entry $\Phi_{i,j}$ of the measurement ensemble as independent Gaussian random variables with zero mean and unit variance, or as independent Bernoulli ± 1 random variables.

The essential result is that if x_0 is S -sparse in any Ψ and we make¹

$$m \gtrsim S \cdot \log n \quad (1.3)$$

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

2. *Random sampling from an incoherent orthobasis.* In this scenario, we select the measurement waveforms ϕ_k from rows of an orthogonal matrix Φ' [7]. By convention, we normalize the rows to have energy $\|\phi_k\|_2^2 = n$, making them the same size (at least on average) as in the random waveform case above. The m rows of Φ' are selected uniformly at random from among all subsets of rows of size m . The principal example of this type of sensing is observing randomly selected entries of a spectrally sparse signal [8]. In this case, the ϕ_k are elements of the standard basis (and $\Phi' = nI$), 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*:

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

When $\mu = 1$, Φ' and Ψ are as different as possible; each element of Ψ is flat in the Φ' domain. If x_0 is S sparse in the Ψ domain, and we take

$$m \gtrsim \mu^2 \cdot S \cdot \log n \quad (1.5)$$

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*.

There are three criteria which we can use to compare strategies for compressive sampling.

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. When Φ is a Gaussian ensemble this is clear, as it will remain Gaussian under any orthogonal transform Ψ . It is less clear but still true [6]

¹Here and below we will use the notation $X \gtrsim Y$ to mean that there exists a known constant C such that $X \geq CY$. For random waveforms, we actually have 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.

that the strategy remains universal when the entries of Φ are independent and subgaussian.

Random sampling from a fixed system Φ' is definitely not universal, as $\mu(\Phi', \Psi)$ cannot be on the order of 1 for every Ψ . We cannot, for example, sample in a basis which is very similar to Ψ ($\mu \approx \sqrt{n}$) and expect to see much of the signal.

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. In general, applying an $m \times n$ matrix to an n -vector requires $O(mn)$ operations; far too many if n and m are only moderately large (in the millions, say).

It is often the case that a Ψ which sparsifies the 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 will ask the same of our measurement system Φ .

The complete lack of structure in measurement ensembles 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 [13] are incoherent with wavelet representations and enjoy a fast transform with the same complexity as the fast Fourier transform.

Physically realizable. In the end, we have to be able to build sensors which take the linear measurements in (1.1). There are architectures for CS where we have complete control over the types of measurements we make; a well-known example of this is the “single pixel camera” of [17]. However, it is often the case that we are exploiting a physical principle to make indirect observations of an object of interest. There may be opportunities for injecting randomness into the acquisition process, but the end result will not be taking inner products against independent waveforms.

In this paper, we introduce a framework for compressive sensing which meets all three of these criteria. The measurement process consists of convolving the signal of interest with a random pulse and then randomly subsampling. This procedure is random enough to be universally incoherent with any fixed representation system, but structured enough to allow fast computations (via the FFT). Convolution with a pulse of our choosing is also a physically relevant sensing architecture. In Section 2 we discuss two applications in particular: radar imaging and coherent imaging using Fourier optics.

1.1. Random Convolution. Our measurement process has two steps. We circularly convolve the signal $x_0 \in \mathbb{R}^n$ with a “pulse” $h \in \mathbb{R}^n$, 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

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

with F as the discrete Fourier matrix

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

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

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

a diagonal matrix whose entries are unit magnitude complex numbers with random phases. We generate the σ_ω as follows²:

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

The action of H on a signal x 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 randomly modulated 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 realized by with an analog-to-digital converter (ADC) that takes non-uniformly spaced samples at an average rate which is appreciably slower than Nyquist. Convolution 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).

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 ι_1, \dots, ι_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 [8] 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 set Ω .

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

1.2.2. Randomly pre-modulated summation. 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 block with a single number. To keep things simple, we will assume that m evenly divides n . With $B_k = \{(k-1)n/m + 1, \dots, kn/m\}$, $k = 1, \dots, m$ 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

$$\phi_k = \sqrt{\frac{m}{n}} \sum_{t \in B_k} \epsilon_t h_t, \quad (1.6)$$

where h_t is the t th row of H . The $\{\epsilon_p\}_{p=1}^n$ 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 norm 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 $\{\epsilon_p\}$, and P sums the result over each block B_k .

In hardware, randomly pre-modulated summation (RPMS) can be implemented by multiplying the incoming signal by a high-rate (Nyquist or above) chipping sequence, effectively changing the polarity across short time intervals. This is followed by an integrator (to compute the analog of the sum in (1.6)) and an ADC taking equally spaced samples at a rate much smaller than Nyquist. This acquisition strategy is analyzed (without the convolution with h) in [33], where it is shown to be effective for capturing spectrally sparse signals. In Section 2.2, we will also see how RPMS can be used in certain imaging applications.

The recovery bounds we derive for random convolution followed by RPMS will essentially be the same (to within a single log factor) as in the random subsampling scheme. But RPMS has one important advantage: it “sees” more of the signal than random subsampling without any amplification. 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 (1.6) (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 . So RPMS is an architecture for subsampling which will on average retain all of the signal energy.

It is important to understand that the summation in (1.6) must be randomly modulated. 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 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 this signal, randomly pre-modulated summation does not. This fact will be especially important in the imaging architecture discussed in Section 2.2, where measurements are taken by averaging over large pixels.

1.3. Main Results. Both random subsampling [8] and RPMS [33] 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*

$$m \geq C_0 \cdot S \cdot \log(n/\delta) \tag{1.7}$$

and also $m \geq C_1 \log^3(n/\delta)$, where C_0 and C_1 are known constants. 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 $1 - \delta$.

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 ($\log^2(n/\delta)$ instead of $\log(n/\delta)$ in (1.7)); 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 RPMS.

THEOREM 1.2. *Let $\Psi, \Gamma, \alpha_0, x_0$, and H be as in Theorem 1.1. Create a random pre-modulated summation matrix $P\Theta$ as described in Section 1.2.1 that outputs a number of samples m with*

$$m \geq C_0 \cdot S \cdot \log^2(n/\delta) \tag{1.8}$$

and also $m \geq C_1 \log^4(n/\delta)$, where C_0 and C_1 are known constants. 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 $1 - \delta$. The form of the bound (1.8) is the same as using RPMS (again, without the random convolution in front) to sense spectrally sparse signals.

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 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 [8]. 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, when we blast the phases by applying Σ , these relationships will no longer exist. A simple example of this phenomena is shown in Figure 1.1.

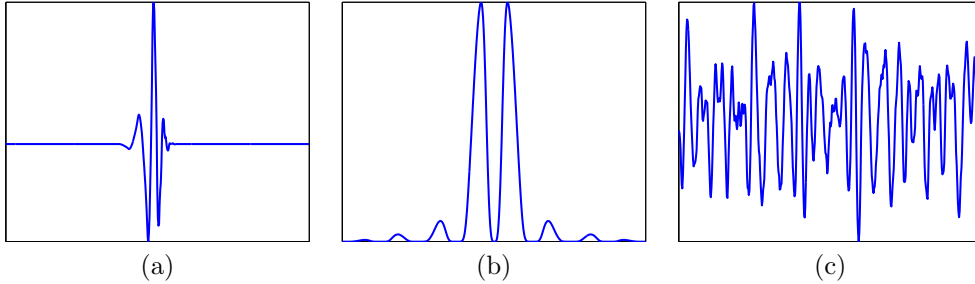


FIG. 1.1. (a) A signal x_0 consisting of a single Daubechies-8 wavelet. Taking samples at random locations of this wavelet will not be an effective acquisition strategy, as very few will be located in its support. (b) Magnitude of the Fourier transform Fx_0 . (c) Inverse Fourier transform after the phase 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.

1.4. Relationship to previous research. 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 [23,35]. Here, we show that a random kernel works for any kind sparsity, and we use an entirely different reconstruction framework.

In [34], numerical results are presented that demonstrate recovery of sparse signals (using orthogonal matching pursuit instead of ℓ_1 minimization) from a small number of samples of the output of a finite length “random filter”. In this paper, we approach things from a more theoretical perspective, deriving bounds on the number of samples need to guarantee sparse reconstruction.

In [4], random convolution is explored in a slightly different context than in this paper. Here, the sensing matrix Φ consists of random selected rows (or modulated sums) of a Toeplitz matrix; in [4], the sensing matrix is itself Toeplitz, corresponding to convolution followed by a small number of consecutive samples. This difference in structure will allow us to derive stronger bounds: (1.7) guarantees recovery from $S \log n$ measurements, while the bound in [4] requires $S^2 \log n$.

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 two imaging scenarios (in a somewhat rarified setting) 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. Radar imaging. Reconstruction from samples of a signal convolved with a known pulse is fundamental in radar imaging. Figure 2.1 illustrates how a scene is measured in spotlight synthetic aperture radar (SAR) imaging (see [25] 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.

The question, then, is how many samples of $y(t)$ are needed to reconstruct the range profile R_θ . A classical reconstruction will require a number of samples proportional to the bandwidth of the pulse h ; in fact, the sampling rate of the analog-to-digital converter is one of the limiting factors in the performance of modern-day radar systems [26]. The results outlined in Section 1.3 suggest that if we have an appropriate representation for the range profiles and we use a broadband random pulse, then the number of samples needed to reconstruct an $R_\theta(r)$ using (1.2) scales linearly with the *complexity* of these range profiles, and only *logarithmically* with the bandwidth of h . We can gain the benefits of a high-bandwidth pulse without paying the cost of an ADC operating at a comparably fast rate.

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. Also, in [19], it is shown that compressive sensing can be used to super-resolve point targets when the radar sends out an incoherent pulse.

2.2. Fourier optics. Convolution with a pulse of our choosing can also be implemented optically. Figure 2.2 sketches a possible compressive sensing imaging architecture. The object is illuminated by a 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 light field exiting this transparency. The first lens takes an optical Fourier transform of the image, the phase of the Fourier transform is then manipulated using a spatial light modulator. The next lens inverts the Fourier transform, and then a second spatial light modulator and a low-resolution detector array with “big pixels” affect the RPMS subsampling scheme. In this particular situation, we will assume

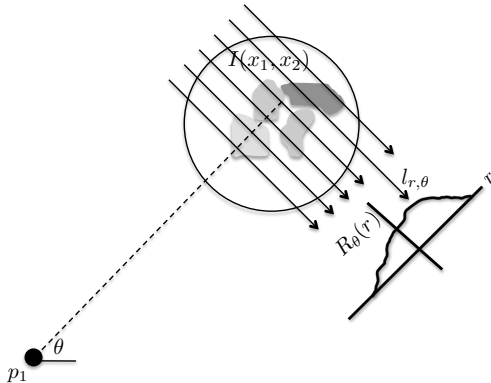


FIG. 2.1. Geometry for the “spotlight SAR” imaging problem. The return signal from a pulse $h(t)$ emitted from point p_1 will be the range profile $R_\theta(r)$ (collection of line integrals at an angle θ) convolved with h .

that our detectors can observe both the magnitude and the phase of the final light field (modern heterodyne detectors have this capability).

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 architecture with a simple numerical experiment³. 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

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

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

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

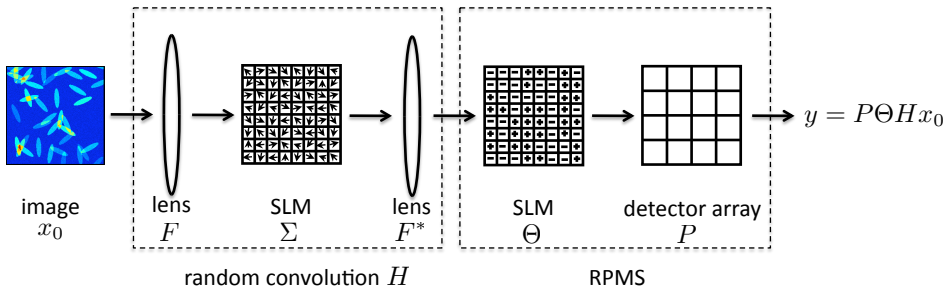


FIG. 2.2. *Fourier optics imaging architecture implementing random convolution followed by RPMS. The computation $y = P\Theta Hx_0$ is done entirely in analog; the lenses move the image to the Fourier domain and back, and spatial light modulators (SLMs) in the Fourier and image planes randomly change the phase.*

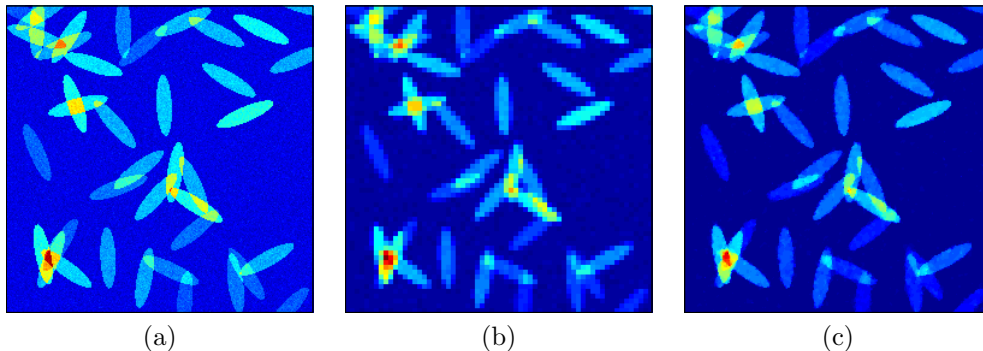


FIG. 2.3. *Fourier optics imaging experiment. (a) The “high-resolution” image we wish to acquire. (b) The high-resolution image pixellated by averaging over 4×4 blocks. (c) The image restored from the pixellated version in (b), plus a set of incoherent measurements taken using the architecture from Figure 2.2. The incoherent measurements allow us to effectively super-resolve the image in (b).*

allowed us to super-resolve the image; the boundaries of the ellipses are far clearer than in part (b).

The architecture is suitable for imaging with incoherent illumination as well, but there is a twist. Instead of convolution with the random pulse h (the inverse Fourier transform of the mask Σ), the lens-SLM-lens system convolves with $|h|^2$. While $|h|^2$ is still random, and so the spirit of the device remains the same, convolution with $|h|^2$ is no longer a unitary operation, and thus falls slightly outside of the mathematical framework we develop in this paper.

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 representation system.

LEMMA 3.1. *Let Ψ be an arbitrary fixed orthogonal 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*

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

Proof. The proof is a simple application of Hoeffding's inequality (see Appendix A). If h_t is the t th row of H (measurement vector) and ψ_s is the s th column of Ψ (representation vector), then

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

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

$$\langle h_t, \psi_s \rangle = \sigma_1 \hat{\psi}_s(1) + (-1)^{t-1} \sigma_{n/2+1} \hat{\psi}_s(n/2+1) + 2 \sum_{\omega=2}^{n/2} \text{Re} \left[F_{t,\omega}^* \sigma_\omega \hat{\psi}_s(\omega) \right], \quad (3.2)$$

where we have assumed without loss of generality that n is even; it will be obvious how to extend to the case where n is odd. Now each of the σ_ω in the sum above are independent. Because the σ_ω are uniformly distributed on the unit circle, $\text{Re}[F_{t,\omega}^* \sigma_\omega \hat{\psi}_s(\omega)]$ has a distribution identical to $\epsilon_\omega |\hat{\psi}_s| \cos(\hat{\theta}(\omega))$, where $\hat{\theta}(\omega)$ is the phase of $\hat{\psi}_s(\omega) F_{t,\omega}^*$ and $\{\epsilon_\omega\}$ is an independent random sign sequence. Thus,

$$\sum_{\omega=1}^{n/2+1} \epsilon_\omega a_\omega, \quad \text{with} \quad a_\omega = \begin{cases} \hat{\psi}_s(1), & \omega = 1 \\ 2|\hat{\psi}_s| \cos(\hat{\theta}(\omega)), & 2 \leq \omega \leq n/2 \\ \hat{\psi}_s(n/2+1), & \omega = n/2+1 \end{cases}$$

has a distribution identical to $\langle h_t, \psi_s \rangle$. Since

$$\sum_{\omega=1}^{n/2+1} a_\omega^2 \leq 2\|\psi_s\|_2^2 = 2,$$

applying the bound (A.1) gives us

$$\text{P} \left(\left| \sum_{\omega=1}^{n/2+1} a_\omega^2 \right| > \lambda \right) \leq 2e^{-\lambda^2/4}.$$

Taking $\lambda = 2\sqrt{\log(2n^2/\delta)}$ and applying the union bound over all n^2 choices of (t, s) establishes the lemma. \square

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 (1.7). 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

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

with the r^k as rows in the matrix $H\Psi_\Gamma$; we will call $\nu(\Gamma)$ the *cumulative coherence* Γ (this quantity was also used in [32]). We will show below that we can bound the number of measurements needed to recover (with high probability) a signal on Γ by

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

Since we always have the bound $\nu \leq \mu\sqrt{S}$, 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 let r^k , $k = 1, \dots, n$ be the rows of $H\Psi_\Gamma$. Then with probability exceeding $1 - \delta$*

$$\nu(\Gamma) = \max_{k=1, \dots, n} \|r^k\|_2 \leq \sqrt{8S}, \quad (3.4)$$

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

Proof. of Lemma 3.2. We can write r^k as the following sum of random vectors in \mathbb{C}^S :

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

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

$$g_\omega = \begin{pmatrix} \hat{\psi}_{\gamma_1}(\omega)^* \\ \hat{\psi}_{\gamma_2}(\omega)^* \\ \dots \\ \hat{\psi}_{\gamma_S}(\omega)^* \end{pmatrix}.$$

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

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

(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 $\text{Re}[F_{k,\omega} \sigma_\omega^* g_\omega]$ has a distribution identical to $\epsilon_\omega \text{Re}[F_{k,\omega} \sigma_\omega^* g_\omega]$, where ϵ_ω is an independent Rademacher random variable. We set

$$Y = \sum_{\omega=1}^{n/2+1} \epsilon_\omega v_\omega, \quad \text{where } v_\omega = \begin{cases} g_1 & \omega = 1 \\ 2 \text{Re} [F_{k,\omega} \sigma_\omega^* g_\omega] & 2 \leq \omega \leq n/2 \\ g_{n/2+1} & \omega = n/2 + 1 \end{cases}.$$

and will show that with high probability $\|Y\|_2$ will be within a fixed constant of \sqrt{S} .

We can bound the mean of $\|Y\|_2$ as follows:

$$\mathbb{E}[\|Y\|_2]^2 \leq \mathbb{E}[\|Y\|_2^2] = \sum_{\omega=1}^{n/2+1} \|v_\omega\|_2^2 \leq 2 \sum_{\omega=1}^n \|g_\omega\|_2^2 = 2S,$$

where the last inequality come from the facts that $\|v_\omega\|_2^2 \leq 4\|g_\omega\|_2^2$ and $\|g_\omega\|_2^2 = \|g_{n-\omega+2}\|_2^2$ for $\omega = 2, \dots, n/2$.

To show that $\|Y\|$ is concentrated about its mean, we use a concentration inequality on the norm of a sum of random vector, specifically (A.3) which is detailed in the Appendix.

To apply this bound first note that for any vector $\xi \in \mathbb{R}^s$

$$|\langle \xi, \text{Re}[F_{k,\omega} \sigma_\omega^* g_\omega] \rangle|^2 \leq |\langle \xi, g_\omega \rangle|^2.$$

Thus

$$\begin{aligned} \sup_{\|\xi\|_2 \leq 1} \sum_{\omega=1}^{n/2+1} |\langle \xi, v_\omega \rangle|^2 &\leq 2 \sup_{\|\xi\|_2 \leq 1} \sum_{\omega=1}^n |\langle \xi, g_\omega \rangle|^2 \\ &= 2 \sup_{\|\xi\|_2 \leq 1} \xi^T \Psi_\Gamma^T \Psi_\Gamma \xi \\ &= 2. \end{aligned}$$

We now apply (A.3) to get

$$\mathbb{P}(\|Y\|_2 \geq \sqrt{2S} + \lambda) \leq \mathbb{P}(\|Y\|_2 \geq \mathbb{E}\|Y\|_2 + \lambda) \leq 2e^{-\lambda^2/32}$$

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

$$\mathbb{P}(\|r^k\|_2 \geq C\sqrt{S}) = \mathbb{P}(\|Y\|_2 \geq C\sqrt{S}) \leq 2e^{-(C-\sqrt{2})^2 S/32} \leq \delta/n$$

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

3.2. Sparse Recovery. The following two results extends the main results of [7] and [33] to take advantage of our more refined bound on the norm coherence M . Theorems 3.3 and 3.4 are stated for general measurement systems U with $U^*U = nI$, as they may be of independent interest. Theorems 1.1 and 1.2 are then simple applications with $U = H\Psi$ and the bound (3.4) on $\nu(\Gamma)$.

Here and below we will use U_Γ for the $n \times S$ matrix consisting of the columns of U indexed by the set Γ .

THEOREM 3.3. *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 $\nu := \max_{k=1, \dots, n} \|r^k\|_2$. Choose a subset Ω of the measurement domain of size $|\Omega| = m$ and a sign sequence z on Γ uniformly at random. Set $\Phi = R_\Omega U$, the matrix constructed from the rows of U indexed by Ω . Suppose that*

$$m \geq C_0 \cdot \nu^2 \cdot \log(n/\delta),$$

and also $m \geq C'_0 \cdot \mu^2 \log^2(n/\delta)$ where C_0 and C'_0 are known constants. Then with probability exceeding $1 - O(\delta)$, every vector α_0 supported on Γ with sign sequence z can be recovered from $y = \Phi\alpha_0$ by solving (1.2).

THEOREM 3.4. *Let U, μ, Γ, ν, z be as in Theorem 3.3. Create an RPMS matrix $P\Theta$ as in Section 1.2.2, and set $\Phi = P\Theta U$. Suppose that*

$$m \geq C_1 \cdot \nu^2 \cdot \log^2(n/\delta),$$

and also $m \geq C'_1 \cdot \mu^2 \log^3(n/\delta)$ where C_1 and C'_1 are known constants. Then with probability exceeding $1 - O(\delta)$, every vector α_0 supported on Γ with sign sequence z can be recovered from $y = \Phi\alpha_0$ by solving (1.2).

The proofs of Theorems 3.3 and 3.4 follow the same general outline put forth in [7], with with one important modification (Lemmas 3.5 and 3.6 below). As detailed in [8, 18, 31], 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

$$|\pi(\gamma)| = | \langle (\Phi_\Gamma^* \Phi_\Gamma)^{-1} \Phi_\Gamma^* \varphi_\gamma, z \rangle | < 1 \quad \text{for all } \gamma \in \Gamma^c, \quad (3.5)$$

where φ_γ is the column of Φ at index γ .

There are three essential steps in establishing (3.5):

1. Show that with probability exceeding $1 - \delta$, the random matrix $\Phi_\Gamma^* \Phi_\Gamma$ will have a bounded inverse:

$$\|(\Phi_\Gamma^* \Phi_\Gamma)^{-1}\| \leq 2/m, \quad (3.6)$$

where $\|\cdot\|$ is the standard operator norm. This has already been done for us in both the random subsampling and the RPMS cases. For random subsampling, it is essentially shown in [7, Th.1.2] that this is true when

$$m \geq \nu^2 \cdot \max(C_1 \log S, C_2 \log(3/\delta)),$$

for given constants C_1, C_2 . For RPMS, a straightforward generalization of [33, Th. 7] establishes (3.6) for

$$m \geq C_3 \cdot \nu^2 \cdot \log^2(S/\delta). \quad (3.7)$$

2. Establish, again with probability exceeding $1 - O(\delta)$, that the norm of $\Phi_\Gamma^* \varphi_\gamma$ is on the order of $\nu\sqrt{m}$. This is accomplished in Lemmas 3.5 and 3.6 below. Combined with step 2, this means that the norm of $(\Phi_\Gamma^* \Phi_\Gamma)^{-1} \varphi_\gamma$ is on the order of ν/\sqrt{m} .
3. Given that $\|(\Phi_\Gamma^* \Phi_\Gamma)^{-1} \varphi_\gamma\|_2 \sim \nu/\sqrt{m}$, show that $|\pi(\gamma)| < 1$ for all $\gamma \in \Gamma^c$ with probability exceeding $1 - O(\delta)$. Taking z as a random sign sequence, this is a straightforward application of the Hoeffding inequality.

3.3. Proof of Theorem 3.3. As step 1 is already established, we start with step 2. We will assume without loss of generality that $m^{1/2} \leq \nu/\mu$, as the probability of success monotonically increases as m increases. The following lemma shows that $\|\Phi_\Gamma^* \varphi_\gamma\|_2 \sim \nu/\sqrt{m}$ for each $\gamma \in \Gamma^c$.

LEMMA 3.5. *Let Φ, μ, Γ, ν , and m be as in Theorem 3.3. Fix $\gamma \in \Gamma^c$, and consider the random vector $v_\gamma = \Phi_\Gamma^* \varphi_\gamma = U_\Gamma^* R_\Omega^* \varphi_\gamma$. Assume that $\sqrt{m} \leq \nu^2/\mu$. Then for any $a \leq 2m^{1/4} \mu^{-1/2}$,*

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

where C is a known constant.

Proof. We show that the mean $\mathbb{E} \|v_\gamma\|_2$ is less than $\nu\sqrt{m}$, and then apply the Talagrand concentration inequality to show that $\|v_\gamma\|_2$ is concentrated about its mean.

Using the Bernoulli sampling model, $\Phi_\Gamma^* \varphi_\gamma$ can be written as a sum of independent random variables,

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

where r^k is the k th row of $U_\Gamma = H\Psi_\Gamma$, and the second equality follows from the orthogonality of the columns of U_Γ . To bound the mean, we use

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

We will apply Talagrand (see (A.4) in the Appendix) using $Y_k = (I_k - m/n)U_{k,\gamma}r^k$. Note that for all $f \in \mathbb{R}^S$ with $\|f\|_2 \leq 1$,

$$|\langle f, Y_k \rangle| \leq |U_{k,\gamma}| \cdot |\langle f, r^k \rangle| \leq \mu\nu,$$

and so we can use $B = \mu\nu$. For $\bar{\sigma}^2$, note that

$$\mathbb{E}|\langle f, Y_k \rangle|^2 = \frac{m}{n} \left(1 - \frac{m}{n}\right) \cdot |U_{k,\gamma}|^2 \cdot |\langle f, r^k \rangle|^2$$

and so

$$\begin{aligned} \bar{\sigma}^2 &= \sup_f \sum_{k=1}^n \mathbb{E}|\langle f, Y_k \rangle|^2 \\ &= \frac{m}{n} \left(1 - \frac{m}{n}\right) \mu^2 \sum_{k=1}^n |\langle f, r^k \rangle|^2 \\ &\leq m\mu^2. \end{aligned}$$

Plugging the bounds for $\mathbb{E}\|\Phi_\Gamma^* \varphi_\gamma\|_2$, B , and $\bar{\sigma}^2$ into (A.4), we have

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

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

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

for all $0 \leq t \leq 2\nu\sqrt{m}$. Thus

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

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

□

To finish off the proof of the Theorem, let \mathcal{A} be the event that (3.6) holds; step 1 tells us that $P(\mathcal{A}^c) \leq \delta$. Let \mathcal{B}_λ be the event that

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

By Lemma 3.5 and taking the union bound over all $\gamma \in \Gamma^c$, we have

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

By the Hoeffding inequality,

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

Our final probability of success can then be bounded by

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

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

$$m \geq 18 \cdot \nu^2 \cdot \log(2n/\delta).$$

3.4. Proof of Theorem 3.4. As before, we control the norm of $\Phi_\Gamma^* \phi_\gamma$ (now with $\Phi = P\Theta U$), and then use Hoeffding to establish (3.5) with high probability. Lemma 3.6 says that $\|\Phi_\Gamma^* \phi_\gamma\|_2 \sim \sqrt{m\nu} \sqrt{\log n}$ with high probability.

LEMMA 3.6. *Let Φ, μ, Γ, ν , and m be as in Theorem 3.4. Fix $\gamma \in \Gamma^c$. Then for any $\delta > 0$*

$$P\left(\|\Phi_\Gamma^* \phi_\gamma\|_2 > Cm^{1/2} \sqrt{\log(n/\delta)} \left(\nu + 4\mu \sqrt{\log(2n^2/\delta)}\right)\right) = 2\delta/n,$$

where C is a known constant.

Proof. To start, note that we can write $\Phi_\Gamma^* \phi_\gamma$ as

$$\begin{aligned} \Phi_\Gamma^* \phi_\gamma &= \frac{m}{n} \sum_{k=1}^m \sum_{p_1, p_2 \in B_k} \epsilon_{p_1} \epsilon_{p_2} U_{p_2, \gamma} r^{p_1} \\ &= \frac{m}{n} \sum_{k=1}^m \sum_{\substack{p_1, p_2 \in B_k \\ p_1 \neq p_2}} \epsilon_{p_1} \epsilon_{p_2} U_{p_2, \gamma} r^{p_1} \end{aligned}$$

since

$$\sum_{k=1}^m \sum_{p \in B_k} U_{p, \gamma} r^p = 0,$$

by the orthogonality of the columns of U . To decouple the sum above, we must first make it symmetric:

$$\Phi_{\Gamma}^* \phi_{\gamma} = \frac{m}{2n} \sum_{k=1}^m \sum_{\substack{p_1, p_2 \in B_k \\ p_1 \neq p_2}} \epsilon_{p_1} \epsilon_{p_2} (U_{p_2, \gamma} r^{p_1} + U_{p_1, \gamma} r^{p_2}).$$

Now, if we set

$$Z = \left\| \frac{m}{2n} \sum_{k=1}^m \sum_{\substack{p_1, p_2 \in B_k \\ p_1 \neq p_2}} \epsilon_{p_1} \epsilon'_{p_2} (U_{p_2, \gamma} r^{p_1} + U_{p_1, \gamma} r^{p_2}) \right\|_2, \quad (3.9)$$

where the sequence $\{\epsilon'_i\}$ is an independent sign sequence distributed the same as $\{\epsilon_i\}$, we have (see [14, Chapter 3])

$$\mathbb{P}(\|\Phi_{\Gamma}^* \phi_{\gamma}\|_2 > \lambda) < C_D \mathbb{P}(Z > \lambda/C_D) \quad (3.10)$$

for some known constant C_D , and so we can shift our attention to bounding Z .

We will bound the two parts of the random vector separately. Let

$$v_{\gamma} = \frac{m}{n} \sum_{k=1}^m \sum_{\substack{p_1, p_2 \in B_k \\ p_1 \neq p_2}} \epsilon_{p_1} \epsilon'_{p_2} U_{p_2, \gamma} r^{p_1}.$$

We can rearrange this sum to get

$$\begin{aligned} v_{\gamma} &= \frac{m}{n} \sum_{k=1}^m \sum_{p \in B_k} \sum_{q=1}^{n/m-1} \epsilon_p \epsilon'_{(p+q)_k} U_{(p+q)_k, \gamma} r^p \quad (\text{where } (p+q)_k \text{ is addition modulo } B_k) \\ &= \frac{m}{n} \sum_{q=1}^{n/m-1} \epsilon_q \sum_{k=1}^m \sum_{p \in B_k} \frac{\epsilon_p}{\epsilon_q} \epsilon'_{(p+q)_k} U_{(p+q)_k, \gamma} r^p \\ &= \frac{m}{n} \sum_{q=1}^{n/m-1} \epsilon_q \sum_{k=1}^m \sum_{p \in B_k} \epsilon''_{(p+q)_k} U_{(p+q)_k, \gamma} r^p \quad (\text{where } \{\epsilon''_i\} \text{ is an iid sign sequence independent of } \{\epsilon_q\}) \\ &= \frac{m}{n} \sum_{q=1}^{n/m-1} \epsilon_q w_q, \quad \text{where } w_q = \sum_{k=1}^m \sum_{p \in B_k} \epsilon''_{(p+q)_k} U_{(p+q)_k, \gamma} r^p. \end{aligned}$$

We will show that $\|w_q\|_2$ can be controlled properly for each q , and then use the a concentration inequality to control the norm of v_{γ} . We start by bounding the mean of $\|w_q\|_2$:

$$\begin{aligned} (\mathbb{E} \|w_q\|_2)^2 &\leq \mathbb{E} \|w_q\|_2^2 \\ &= \sum_{k=1}^m \sum_{p \in B_k} U_{(p+q)_k, \gamma}^2 \|r^p\|_2^2 \\ &\leq \nu^2 \sum_{i=1}^n U_{i, \gamma}^2 \\ &= n\nu^2. \end{aligned}$$

Using (A.3), we have that

$$\mathbb{P}(\|w_q\|_2 > \sqrt{n} \cdot \nu + \lambda) \leq 2e^{-\lambda^2/16\sigma^2},$$

where

$$\sigma^2 = \sup_{\|\xi\|_2 \leq 1} \xi^T V V^T \xi, \quad \text{with } V := (U_{(1-q)k, \gamma} r^1 \quad U_{(2-q)k, \gamma} r^2 \quad \cdots \quad U_{(n-q)k, \gamma} r^n) = U_\Gamma^* D,$$

and D is a diagonal matrix with the $U_{\cdot, \gamma}$ along the diagonal. Since σ^2 is the largest (squared) singular value of V , and since $\|U_\Gamma^*\| = \sqrt{n}$ and $\|D\| \leq \mu$, we have that $\sigma^2 \leq n\mu^2$. Thus

$$\mathbb{P}(\|w_q\|_2 > \sqrt{n} \cdot \nu + \lambda) \leq 2e^{-\lambda^2/16n\mu^2}.$$

Let \mathcal{M}_λ be the event that all of the $\|w_q\|_2$ are close to their mean, specifically

$$\max_{q=1, \dots, n/m} \|w_q\|_2 \leq \sqrt{n} \cdot \nu + \lambda,$$

and note that

$$\mathbb{P}(\mathcal{M}_\lambda^c) \leq 2nm^{-1}e^{-\lambda^2/n\mu^2}. \quad (3.11)$$

Then

$$\mathbb{P}(\|v_\gamma\|_2 > \alpha) \leq \mathbb{P}(\|v_\gamma\|_2 > \alpha \mid \mathcal{M}_\lambda) + \mathbb{P}(\mathcal{M}_\lambda^c). \quad (3.12)$$

Given that \mathcal{M}_λ has occurred, we will now control $\|v_\gamma\|_2$. Again, we start with the expectation

$$\begin{aligned} \mathbb{E}[\|v_\gamma\|_2^2 \mid \mathcal{M}_\lambda] &= \frac{m^2}{n^2} \sum_{q_1, q_2} \mathbb{E}[\epsilon_{q_1} \epsilon_{q_2} \mid \mathcal{M}_\lambda] \mathbb{E}[\langle w_{q_1}, w_{q_2} \rangle \mid \mathcal{M}_\lambda] \\ &= \frac{m^2}{n^2} \sum_{q_1, q_2} \mathbb{E}[\epsilon_{q_1} \epsilon_{q_2}] \mathbb{E}[\langle w_{q_1}, w_{q_2} \rangle \mid \mathcal{M}_\lambda] \\ &= \frac{m^2}{n^2} \sum_q \mathbb{E}[\|w_q\|_2^2 \mid \mathcal{M}_\lambda] \\ &\leq \frac{m^2}{n^2} \sum_q (\sqrt{n}\nu + \lambda)^2 \\ &= \frac{m}{n} (\sqrt{n}\nu + \lambda)^2. \end{aligned}$$

Using Khintchine-Kahane ((A.2) in the Appendix), we have

$$\mathbb{P}\left(\|v_\gamma\|_2 > C' m^{1/2} n^{-1/2} (n^{1/2} \nu + \lambda) \sqrt{\log(n/\delta)} \mid \mathcal{M}_\lambda\right) \leq \delta/n. \quad (3.13)$$

Choosing $\lambda = 4n^{1/2} \mu \sqrt{\log(2n^2/m\delta)}$, we fold (3.11) and (3.13) into (3.12) to get

$$\mathbb{P}\left(\|v_\gamma\|_2 > C' m^{1/2} \sqrt{\log(n/\delta)} \left(\nu + 4\mu \sqrt{\log(2n^2/m\delta)}\right)\right) \leq 2\delta/n \quad (3.14)$$

By symmetry, we can replace $\|v_\gamma\|_2$ with Z in (3.14), and arrive at the lemma by applying the decoupling result (3.10).

□

We finish up the proof of Theorem 3.4 by again applying the Hoeffding inequality. Let \mathcal{A} be the event that the spectral bound (3.6) holds; we know that $P(\mathcal{A}^c) \leq \delta$ for m as in (3.7). Let \mathcal{B} be the event that

$$\max_{\gamma \in \Gamma^c} \|(\Phi_\Gamma^* \Phi_\Gamma)^{-1} \Phi_\Gamma^* \varphi \gamma\|_2 \leq 2Cm^{-1/2} \sqrt{\log(n/\delta)} \left(\nu + 4\mu \sqrt{\log(2n^2/\delta)} \right),$$

where the constant C is the same as that in Lemma 3.6; we have shown that $P(\mathcal{B}^c | \mathcal{A}) \leq 2\delta$. By the Hoeffding inequality

$$P\left(\sup_{\gamma \in \Gamma^c} |\pi(\gamma)| > 1 \mid \mathcal{B}, \mathcal{A}\right) \leq 2n \exp\left(-\frac{m}{8C^2 \log(n/\delta) (\nu + 4\mu \sqrt{\log(2n^2/m\delta)})}\right). \quad (3.15)$$

When $\nu \geq 4\mu \sqrt{\log(2n^2/m\delta)}$, as will generally be the case, the left-hand side of (3.15) will be less than δ when

$$m \geq \text{Const} \cdot \nu^2 \cdot \log(n/\delta).$$

When $\nu < 4\mu \sqrt{\log(2n^2/m\delta)}$, the left-hand side of (3.15) will be less than δ when

$$m \geq \text{Const} \cdot \mu^2 \cdot \log^3(n/\delta).$$

The theorem then follows from a union bound with \mathcal{A} and \mathcal{B} .

4. Discussion. Theorems 1.1 and 1.2 tell us that we can recover a perfectly S sparse signal from on the order of $S \log n$, in the case of random subsampling, or $S \log^2 n$, in the case of random pre-modulated summation, noiseless measurements. If we are willing to pay additional log factors, we can also guarantee that the recovery will be stable. In [27], it is shown that if Φ is generated by taking random rows from an orthobasis, then with high probability it will obey the *uniform uncertainty principle* when

$$m \gtrsim \mu^2 \cdot S \cdot \log^4 n.$$

The connection between the uniform uncertainty principle and stable recovery via ℓ_1 minimization is made in [9–11]. Along with Lemma 3.1, we have stable recovery from a randomly subsampled convolution from

$$m \gtrsim S \cdot \log^5 n$$

measurements. There is also a uniform uncertainty principle for an orthobasis which has been subsampled using randomly pre-modulated summation [30] with an additional log factor; in this case, stable recovery would be possible from $m \gtrsim S \cdot \log^6 n$ measurements.

The results in this paper depend on the fact that different shifts of the pulse h are orthogonal to one another. But how critical is this condition? For instance, suppose that h is a random sign sequence in the time domain. Then shifts of h are almost orthogonal, but it is highly doubtful that convolution with h followed by subsampling is a universal CS scheme. The reason for this is that, with high probability, some entries of the Fourier transform \hat{h} will be very small, suggesting that there are spectrally sparse signals which we will not be able to recover. Compressive sampling using a pulse with independent entries is a subject of future investigation.

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 Theorem 1.1 and 1.2.

The Hoeffding inequality [20] is a classical tail bound on the sum of a sequence of independent random variables. Let Y_1, Y_2, \dots, 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

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

for every $\lambda > 0$.

Concentration inequalities analogous to (A.1) exist for the norm of a random sum of vectors. Let v_1, v_2, \dots, v_n be a fixed sequence of vectors in \mathbb{R}^S , and let $\{\epsilon_1, \dots, \epsilon_n\}$ 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 $[\epsilon_1 \ \dots \ \epsilon_n]^T$.

The second moment of Z is easily computed

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

where $\|\cdot\|_F$ is the Frobenius norm. In fact, the Khintchine-Kahane inequality [22] allows us to bound all of the moments in terms of the second moment; for every $q \geq 2$

$$\mathbb{E}[Z^q] \leq C \cdot q^{q/2} \cdot (\mathbb{E}[Z^2])^{1/2},$$

where $C \leq 2^{1/4}$. Using the Markov inequality, we have

$$\mathbb{P}(Z > \lambda) \leq \frac{\mathbb{E}[Z^q]}{\lambda^q} \leq \left(\frac{C\sqrt{q}\|V\|_F}{\lambda}\right)^q,$$

for any $q \geq 2$. Choosing $q = \log(1/\delta)$ and $\lambda = C \cdot e \cdot \sqrt{q} \cdot \|V\|_F$ gives us

$$\mathbb{P}\left(Z > C' \|V\|_F \sqrt{\log(1/\delta)}\right) \leq \delta, \quad (\text{A.2})$$

with $C' \leq e \cdot 2^{1/4}$.

When the singular values of V are all about the same size, meaning that the operator norm (largest singular value) is significantly smaller than the Frobenius norm (sum-of-squares of the singular values), Z is more tightly concentrated around its mean than (A.2) suggests. In particular, Theorem 7.3 in [21] shows that for all $\lambda > 0$

$$\mathbb{P}(Z \geq \mathbb{E}[Z] + \lambda) \leq 2 \exp\left(-\frac{\lambda^2}{16\sigma^2}\right), \quad (\text{A.3})$$

where

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

The “variance” σ^2 is simply the largest squared singular value of V .

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 3.5), 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 [28] gives

$$\mathbb{P}(Z \geq \mathbb{E}[Z] + \lambda) \leq 3 \exp\left(-\frac{\lambda}{KB} \log\left(1 + \frac{B\lambda}{\bar{\sigma}^2 + B\mathbb{E}[Z]}\right)\right), \quad (\text{A.4})$$

where K is a fixed numerical constant,

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

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*. to appear in Constructive Approximation, 2008.
- [7] E. CANDÈS AND J. ROMBERG, *Sparsity and incoherence in compressive sampling*, Inverse Problems, 23 (2007), pp. 969–986.
- [8] 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.
- [9] ———, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. on Pure and Applied Math., 59 (2006), pp. 1207–1223.

- [10] 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.
- [11] ———, *The Dantzig selector: statistical estimation when p is much smaller than n* , Annals of Stat., 35 (2007), pp. 2313–2351.
- [12] S. S. CHEN, D. L. DONOHO, AND M. A. SAUNDERS, *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comput., 20 (1999), pp. 33–61.
- [13] R. COIFMAN, F. GESHWIND, AND Y. MEYER, *Noiselets*, Appl. Comp. Harmonic Analysis, 10 (2001), pp. 27–44.
- [14] V. H. DE LA PEÑA AND E. GINÉ, *Decoupling: From Dependence to Independence*, Springer, 1999.
- [15] D. L. DONOHO, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.
- [16] D. L. DONOHO, Y. TSAIG, I. DRORI, AND J. L. STARCK, *Sparse solutions of underdetermined linear equations by stagewise orthogonal matching pursuit*, Technical Report, Stanford University, (2006).
- [17] M. F. DUARTE, M. A. DAVENPORT, D. TAKHAR, J. N. LASKA, T. SUN, K. F. KELLY, AND R. G. BARANIUK, *Single-pixel imaging via compressive sampling*, IEEE Signal Proc. Mag., 25 (2008), pp. 83–91.
- [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] M. A. HERMAN AND T. STROHMER, *High-resolution radar via compressed sensing*. submitted to IEEE. Trans. Sig. Proc., 2008.
- [20] W. Hoeffding, *Probability inequalities for sums of bounded random variables*, J. American Stat. Assoc., 58 (1963), pp. 13–30.
- [21] M. LEDOUX, *The Concentration of Measure Phenomenon*, American Mathematical Society, 2001.
- [22] M. LEDOUX AND M. TALAGRAND, *Probability in Banach Spaces*, vol. 23 of Ergebnisse der Mathematik und ihrer Grenzgebiete, Springer-Verlag, 1991.
- [23] I. MARAVIC, M. VETTERLI, AND K. RAMCHANDRAN, *Channel estimation and synchronization 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.
- [24] 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.
- [25] D. C. MUNSON, J. D. O'BRIEN, AND W. K. JENKINS, *A tomographic formulation of spotlight-mode synthetic aperture radar*, Proc. IEEE, 71 (1983), pp. 917–925.
- [26] M. RICHARDS, *Fundamentals of Radar Signal Processing*, McGraw-Hill, 2005.
- [27] M. RUDELSON AND R. VERSHYNIN, *On sparse reconstruction from Fourier and Gaussian measurements*, Comm. on Pure and Applied Math., 61 (2008), pp. 1025–1045.
- [28] M. TALAGRAND, *New concentration inequalities in product spaces*, Invent. Math., 126 (1996), pp. 505–563.
- [29] J. TROPP AND A. GILBERT, *Signal recovery from partial information via orthogonal matching pursuit*, IEEE Trans. Inform. Theory, 53 (2007), pp. 4655–4666.
- [30] J. A. TROPP. Personal communication.
- [31] ———, *Recovery of short, complex linear combinations via ℓ_1 minimization*, IEEE Trans. Inform. Theory, 51 (2005), pp. 1568–1570.
- [32] ———, *Random subdictionaries of general dictionaries*. Submitted manuscript, August 2006.
- [33] 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, May 2008.
- [34] 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.
- [35] M. VETTERLI, P. MARZILIANO, AND T. BLU, *Sampling signals with finite rate of innovation*, IEEE Trans. Signal Proc., 50 (2002), pp. 1417–1428.