2014

Sparsity and Structure in Hyperspectral Imaging: Sensing, Reconstruction, and Target Detection

Rebecca M. Willett
Marco Duarte, *University of Massachusetts - Amherst*
Mark A. Davenport
Richard G. Baraniuk

SPARSITY AND STRUCTURE IN HYPERSPECTRAL IMAGING: SENSING, RECONSTRUCTION, AND TARGET DETECTION

Rebecca M. Willett, Marco F. Duarte, Mark A. Davenport, and Richard G. Baraniuk

INTRODUCTION

Hyperspectral imaging is a powerful technology for remotely inferring the material properties of the objects in a scene of interest. Hyperspectral images consist of spatial maps of light intensity variation across a large number of spectral bands or wavelengths; alternatively, they can be thought of as a measurement of the spectrum of light transmitted or reflected from each spatial location in a scene. Because chemical elements have unique spectral signatures, observing the spectra at a high spatial and spectral resolution provides information about the material properties of the scene with much more accuracy than is possible with conventional three-color images. As a result, hyperspectral imaging is used in a variety of important applications, including remote sensing, astronomical imaging, and fluorescence microscopy.

While hyperspectral imaging has great potential, acquiring and processing hyperspectral data comes with significant challenges. First, hyperspectral images are extremely high-dimensional: in remote sensing applications we routinely encounter images over 1GB in size. This dimensionality limits our ability to conduct fast and accurate inference (e.g., removing noise or identifying significant spectral signatures). Second, designers of hyperspectral imagers face a myriad of tradeoffs related to photon efficiency, acquisition time, dynamic range, and sensor size, weight, power, and cost.

In this paper, we will review how novel sparse (low-dimensional) image models are enabling sensor designers to tackle many of the above challenges and create new hyperspectral imaging paradigms. We provide an overview the state-of-the-art of hyperspectral image modeling with an emphasis on sparse models that exploit the fact that typical hyperspectral images, while high-dimensional, can usually be represented using just a few elements from a basis or dictionary. We also explain how sparse models facilitate the design of novel hyperspectral imaging hardware for remote sensing applications. We will pay special attention to cameras based on the compressive sensing (CS) framework that achieve sub-Nyquist measurement rates. We then discuss the imaging design tradeoffs among noise performance, temporal/spatial/spectral resolution, and dynamic range that are afforded by the sensor system, the sparse image model, and noise and quantization errors. Finally, we conclude by describing how the combination of sparse image models and CS architectures can enable fast and accurate target detection.
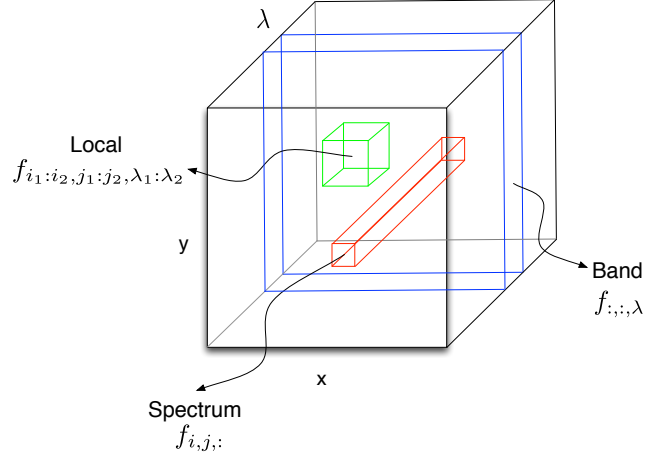

Fig. 1. *Illustration of hyperspectral image partitions. The hyperspectral image spans two spatial dimensions (x,y) and one spectral dimension (λ).*

SPARSE MODELS FOR HYPERSPECTRAL IMAGES

We will consider the problem of acquiring a hyperspectral datacube $\mathbf{f} \in \mathbb{R}^{d_x \times d_y \times d_\lambda}$, where $\mathbf{f}_{i,j,\lambda}$ is the intensity of light in the hyperspectral image at location (i, j) and wavelength λ. For notational simplicity, we will also let f denote a vectorized version of the hyperspectral datacube $\mathbf{f}$, which is just a vector in $\mathbb{R}^d$ where $d \triangleq d_x \cdot d_y \cdot d_\lambda$. We model the hyperspectral imaging process as $y = Af + w$, where $A \in \mathbb{R}^{n \times d}$ represents the imaging process, $y \in \mathbb{R}^n$ is a collection of n measurements generated by our imaging system (where n may be less than d), and $w \in \mathbb{R}^n$ is noise.

Due to the significant structure present in hyperspectral datacubes and the linear nature of the aggregation performed by many hyperspectral imagers, low-dimensional signal models for f have received significant attention in the hyperspectral imaging community in a variety of applications, including image compression, denoising, and processing. Most models operate over a partitioning of the hyperspectral datacube into *patches* along a subset of the dimensions (spatial or spectral) as shown in Figure 1. *Spectrum patches* collect the intensities for a single spatial location and all wavelengths; *band patches* collect the intensities for a single wavelength at all spatial locations; and *local patches* collect the intensities for small intervals of the three dimensions. Denoting the vectorized versions of the patches by the set $\{f_{(1)}, f_{(2)}, \ldots, f_{(l)}\}$, the goal

of a low-dimensional signal model is to represent each one of these patches using a small number of degrees of freedom: we search for a representation dictionary D that yields patch representations $\theta_{(i)}$ with a small number of nonzeros so that we can write $f_{(i)} = D\theta_{(i)}$, $i = 1, \ldots, l$. Below, we discuss two common choices for the dictionary D.

Principal component analysis (PCA) assumes that the data vectors $f_{(i)}$ lie within or very close to a k-dimensional subspace of $\mathbb{R}^{d/l}$ for some $k \ll d/l$. In PCA, one computes the empirical cross correlation matrix for the centered data C; the eigendecomposition $C = U\Sigma U^T$ contains only up to k nonzero (or significant) eigenvalues, and so each patch can be accurately represented as a linear combination of the relevant eigenvectors. In practice, the number k is chosen to obtain sufficiently accurate approximations of the patches.

PCA provides an effective and simple way to approximate hyperspectral data. Consider the case in which the image f corresponds to a scene with only a small number $k \ll d_\lambda$ of different types of spectra present across all pixels. In this case, it is clear that the spectral patches $\{f_{:,:,\lambda}\}$ will lie within a k-dimensional subspace of $\mathbb{R}^{d_x d_y}$. PCA has been applied in this manner for hyperspectral image compression [1], classification, segmentation [2], and denoising under Gaussian [3] and Poisson noise models [4, 5]. Furthermore, PCA models can be estimated directly from a sufficiently large number of compressive measurements given enough training data [6].

Sparse signal models are able to capture more elaborate structure than PCA alone. Sparse signal models assume that the data vectors $\{f_{(i)}\}$ lie within (or close to) a union of $\binom{d/l}{k}$ subspaces of dimension k, where each subspace is spanned by a different choice of k functions from the transformation D. More precisely, these models rely on a sparsity-inducing orthogonal transform D to obtain coefficient vectors $\theta_{(i)} = D^T f_{(i)}$. In words, the coefficient vector has a small number k of nonzero (or significant) coefficients, and so we can represent the vector $f_{(i)}$ exactly (or approximately) as the linear combination of k components of the transform D. Sparsity models can significantly outperform PCA models in terms of approximation fidelity and are predominant in processing and compression of natural images. Examples of sparsity-inducing transforms include the discrete cosine and wavelet transforms. Such transforms can be applied straightforwardly to band patches, as they correspond to intensity images for different light wavelengths.

An additional contribution from the sparsity literature is the application of *dictionary learning* algorithms to hyperspectral imaging [7]. These methods use a training dataset of image patches to learn a dictionary D which yields sparse (albeit high-dimensional) representations. However, in contrast to the transformations discussed earlier, the dictionaries learned here do not have orthogonal elements and require the application of custom algorithms for sparse approximation, described in the sidebar *Sparse recovery: Methods and guarantees*. In recent years, sparsity has also been studied in

contexts where the types of spectra (called *endmembers*) are known *a priori* and that each particular pixel is a linear combination of only a few of the endmembers [7, 8]. The sparse representation of the spectrum effectively identifies the component endmembers and their concentrations at each pixel, a process referred to as *hyperspectral unmixing* [8].

Various global sparsifying transforms, to be applied to the entire image rather than its patches, have also been proposed [9–11]. Unfortunately, the corresponding increase in dimensionality also increases the computational complexity of the transformation and approximation; furthermore, the improvements in approximation error are often not found to be significant enough to warrant the additional computational load. Nonetheless, it is possible to formulate global transformations with higher computational efficiency using combinations of patch transformations; a common example is to select a spectrum patch transform D_λ and a band patch transform $D_{x,y}$ and combine them using a Kronecker product $D = D_{x,y} \otimes D_\lambda$ [11–13]. PCA models for spectral patches can also be integrated with sparsity models for band patches through the use of Kronecker product matrices [11, 12].

SPARSE MODELS AND HYPERSPECTRAL IMAGERS

The CS framework has received significant attention in the remote sensing community due to the complexities in hyperspectral imaging hardware designs, the high dimensionality of hyperspectral datasets, and the significant degree of structure and redundancy present in hyperspectral images. In this section, we review baseline designs for hyperspectral imagers and describe several approaches for hyperspectral imaging that, inspired by CS, leverage the low-dimensional models introduced earlier to address some of the tradeoffs in traditional designs.

Conventional hyperspectral imagers must address a fundamental problem in their design: the transformation of a 3-D signal (in the spatial and spectral domain) into measurements obtained by optical sensing hardware, which is limited to two spatial dimensions. Thus, a fundamental choice in the design of a hyperspectral imager is the selection of a scheme to translate spatial and spectral dimensions into a 1-D stream of measurements (in time, using a single sensor), a 2-D stream of measurements (in space using optical hardware or in time and space using a sensor array), or a 3-D stream of measurements (in time and space using a sensor array). In this section, we describe the most common designs as illustrated in Figure 2.

Whiskbroom designs feature optics that focus on a specific spatial location and record either a sequence in time of voxel spectral measurements (using a tunable filter and a single sensor) or an array of samples of the spectra (using a diffraction grating and a linear sensor array). The optical components in whiskbroom designs select a single pixel/spatial location at a time. Whiskbroom designs require a raster scan across the entire field of view and have higher capture latency than other designs; their dwell time on each specific pixel is reduced in
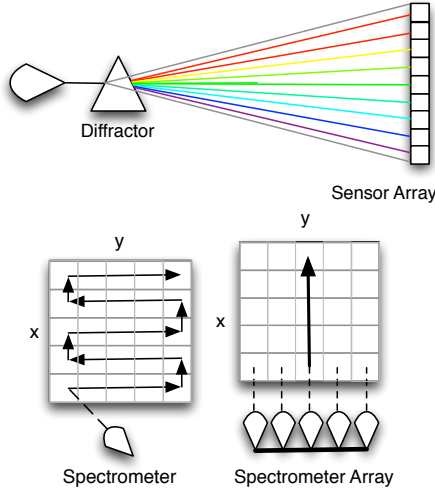
Fig. 2. *Hyperspectral imager architectures. Top: A spectrometer consists of a diffractor (grating or prism) and a sensor array that records light intensities at a variety of wavelengths. Bottom: Whiskbroom designs move the spectrometer spatially throughout the image, scanning one location at at time. Pushbroom designs scan the image along a spatial direction using a spectrometer array.*

comparison with other architectures with matching latency.

Pushbroom designs feature optics that focus along one of the two spatial dimensions (using slit apertures, in comparison with pinhole apertures used by whiskbroom designs) and record a 2-D array of voxels corresponding to a spectral/1-D-spatial cut of the hyperspectral image (using a diffraction grating and a 2-D sensor array). The optical components are usually translated along one spatial dimension to scan the field of view. Although the latency of pushbroom designs is lower than that for whiskbroom designs and their mechanical complexity is comparatively lower, both types of imagers introduce motion in the optics that can result in spatial distortion.

Framing or staring designs feature optics similar to standard imaging cameras that capture 2-D images with additional optics that focus on a single wavelength or band of wavelengths using tunable filters. Their spatial resolution matches that of the sensor array, while spectral resolution is dependent on the tunable filter and latency requirements. The overall design of a staring camera is much simpler than its pushbroom and whiskbroom counterparts. However, the latency due to the tuning of the optical filter is often longer than that of a pushbroom design's scanning system. Furthermore, filtering significantly limits the quantity of light captured at the sensor. **Compressive hyperspectral imagers** address a common theme in the design descriptions above: the high number of samples in the spectral datacube results either in high acquisition latency or in significant requirements for the size of the sensor array in the imager. Thus, it can be desirable to reduce the number of measurements necessary for acquisition of the hyperspectral image at a target spatial and spectral resolution. Since one of the central goals in CS is to minimize the required number of measurements (see the sidebar *Sparse*

recovery: Methods and guarantees for more details), this has naturally led to its application to hyperspectral imaging. In a compressive hyperspectral imager, we continue to model the imaging system as $y = Af + w$ where A is an $n \times d$ matrix, but here we will be specifically interested in the case where n is as small as possible (and hopefully $n \ll d$).

In all of the cases below, the reduction in measurements is achieved through the multiplexing of the voxels of the datacube during acquisition through the optical path. The reduction in measurements can potentially translate to a reduction in acquisition latency; this will be in tradeoff with the additional latency of nonlinear recovery, which in general is costlier than linear methods applied to fully sampled data. Although the sequel focuses on staring designs, the architectures can be modified to pushbroom or whiskbroom designs in a straightforward fashion [14].

The single pixel camera [14–17], like whiskbroom designs, relies on a single spectrometer. However, the measurements do not focus on a single spatial location; rather, each measurement aggregates the intensities from a randomly selected subset of pixels of the image. Such selection is performed by programming an optical modulator (such as a digital micromirror device) to reflect light from a subset of the pixels into the spectrometer while masking the light reflected from the rest of the pixels away from the spectrometer. Choosing this configuration for the optical modulator effectively causes the measurement at the single sensor at instance i to correspond to the projection of each spectral band $f_{:,:,\lambda}$ onto a vector $A_{s,i}$, where $A_{s,i}$ is a binary 0/1 pattern encoding the masking sequence applied by the modulator. By stacking the m vectors as rows of a matrix A_s, the resulting measurement matrix can be expressed as the Kronecker product $A = I \otimes A_s$, where I is the identity matrix; this measurement operator acts separately on each band.

The compression achieved by the single pixel camera can significantly reduce the acquisition latency compared to whiskbroom designs; however, depending on the number of measurements required for recovery (which is dependent on the complexity of the scene), this design may not outperform pushbroom designs in terms of latency. The spatial resolution of this camera design is given by the resolution of the spatial light modulator, while the spectral resolution of this architecture is given by the characteristics of the single spectrometer.

The coded aperture snapshot spectral imager (CASSI) [18] employs a combination of diffraction prisms, coded apertures, and an optical sensor array to perform multiplexing of the voxels in the hyperspectral image. A dispersive element shears the hyperspectral datacube by enacting a distinct spatial translation for the light field at each wavelength; a coded aperture then masks certain pixels (spatial locations) of the sheared datacube, and a second dispersive element reverses the shearing caused by the spatial translation to result in a modified datacube with masked voxels. This masked datacube is acquired using an optical

sensor that effectively flattens the hyperspectral image into a single snapshot. The imager is a completely static, single-shot design, resulting in a mechanically robust and inexpensive system.

The spatial resolution of this design is affected by the sensor array and the coded aperture (which should have matching resolutions), while the spectral resolution is governed by the degree of dispersion and feature size of the coded aperture. A simplified version of CASSI requires only a single dispersive element and captures the sheared datacube, but requires the sensor array size to be $d_x \times (d_y + d_\lambda)$ [19]. This linear acquisition system can be effectively represented by a highly structured $d_x(d_y + d_\lambda) \times d_x d_y d_\lambda$ matrix with binary entries. CASSI is discussed in additional detail in a companion paper in this issue [20], including coded aperture design and extensions in hyperspectral image sensing and modeling.

CMOS-based compressive sensing approaches have recently emerged for optical imaging [11, 21, 22]. In addition to the aforementioned optics-based designs, it is possible to combine these CMOS-based approaches with standard pushbroom or framing designs to reduce the number of measurements taken with respect to the number of voxels. The resulting designs, however, do not change the latency, spectral, or spatial resolution of the resulting compressive camera (when compared to a standard sensor-array camera of the same size and count). Existing implementations of compressive optical sensor arrays perform the computation of the required projections using metal-oxide-semiconductor electronics and are based on random convolution [22], separable transformations [21], block-based transforms [22], structured incoherent transforms like noiselets [11], and randomized integration via Sigma-Delta ADCs [23]. The resulting measurement matrices are expressed in terms of a Kronecker product $I \otimes A_{CMOS}$, where A_{CMOS} denotes the measurement operator implemented by the CMOS design and the Kronecker product represents the replication of the measurement process among the snapshots required by the particular camera design (e.g., across spectral bands for a staring camera or across shifts in a spatial dimension for a pushbroom camera).

PERFORMANCE LIMITS AND TRADEOFFS FOR RECONSTRUCTING HYPERSPECTRAL IMAGES

The compressive hyperspectral imagers described above enable a range of design tradeoffs among noise performance, temporal/spatial/spectral resolution, and dynamic range. These tradeoffs take different forms depending upon what assumptions we can reasonably make about the sensing matrix A, the sparse or low-dimesional structure of the hyperspectral image f, and the distribution of the noise w. We will first consider the classical CS setting with white Gaussian noise, and then discuss effects such as nonnegativity, quantization, and photon-counting noise.

Limits of CS in Gaussian noise. We begin with the simple observation model $y = Af + w$ but where the noise w,

instead of being arbitrary, is i.i.d. Gaussian with mean zero and variance σ^2. This leads to slightly different results than those described in the sidebar *Sparse Recovery: Methods and Guarantees*. Specifically, since the noise w is now random, we consider the *expected* recovery error. While we could directly apply (2) and replace $\|w\|_2$ with $\mathbb{E}[\|w\|_2] = \sqrt{n}\sigma$, it is possible to get a somewhat tighter result (that does not increase if we take more measurements). In particular, under the assumption that $\|Af\|_2 \approx \beta\|f\|_2$, one can show that most standard sparse recovery algorithms yield an estimate satisfying a guarantee of the form

$$\mathbb{E}\left[\|\theta - \widehat{\theta}\|_2\right] \leq C_1' \sqrt{\frac{k \log d}{\beta}}\sigma + C_2\frac{\|\theta - \theta_k\|_1}{\sqrt{k}}, \quad (3)$$

where C_1' and C_2 are absolute constants. Note that we have replaced the standard RIP assumption (that $\|Af\|_2 \approx \|f\|_2$) with the more relaxed assumption that $\|Af\|_2 \approx \beta\|f\|_2$ for some constant β, which is equivalent to saying that A/β satisfies the RIP. This can be quite useful since the RIP induces a particular scaling of the matrix A (unit-norm columns), while other scalings of A may be more natural in practice. Naturally, either an increase in β or a decrease in σ (which are essentially equivalent) leads to improved estimation of θ.

One might wonder whether the first term in (3), which represents the impact of the noise w on the recovery error, can be substantially improved. It turns out that this dependence is essentially optimal. In fact, one can show that given the freedom to pick *any* matrix A (not necessarily satisfying the RIP, but with the same energy as above, i.e., $\|A\|_F^2 = \beta d$) and use *any* recovery procedure, there is no method that can improve on (3) by more than a constant factor [28]. In other words, when it comes to sensing a sparse signal in the presence of Gaussian noise, standard CS algorithms are operating at the limit of what *any* system could achieve given a fixed set of nonadaptive, linear measurements (subject to some energy/SNR constraint on the sensing system A). Moreover, at least if we wish to have an error bound that holds for arbitrary sparse f, we cannot substantially improve this situation even if we pick the rows of this sensing matrix A in a sequential or adaptive fashion [29, 30].

While the bulk of the CS literature has focused on the cases of bounded noise, as in (2), or white Gaussian noise, as in (3), these may not necessarily be the most natural in the context of hyperspectral imaging. In particular, Gaussian noise is not a particularly realistic model for photon noise in low-light imaging. We will address this more realistic noise model below. But first, we discuss an important difference between the standard CS framework and the problem of compressive hyperspectral imaging that arises due to the fact that our measurements are constrained to be nonnegative.

Effects of nonnegativity. Consider the mechanism described in the sidebar *Sparse recovery: Methods and guarantees* for constructing the sensing matrix A, where we set each element

[SIDEBAR] SPARSE RECOVERY: METHODS AND GUARANTEES

There are a number of algorithmic approaches to the problem of sparse signal recovery from compressive measurements. We will not provide a complete overview of the possible recovery algorithms here. Instead, we will merely provide a rough outline of what is possible. For further details, we refer the reader to [24] and references therein.

Perhaps the most popular method for sparse recovery is the basis pursuit algorithm (also known as ℓ_1-norm minimization):

$$\widehat{\theta} = \arg \min_{\theta} \|\theta\|_1 \text{ subject to } \|y - AD\theta\|_2 \leq \epsilon, \tag{1}$$

where $\|\theta\|_1 = \sum_i |\theta_i|$ denotes the sum of the magnitudes of the entries of θ and ϵ denotes the tolerable approximation distortion. In addition to (1), there are also a variety of greedy or iterative strategies, including state-of-the-art methods like CoSaMP [25] or iterative hard thresholding (IHT) [26], that treat the vector $A^T y$ as a rough estimate of f and obtain $\widehat{\theta}$ by iteratively identifying likely nonzeros. In general, any standard sparse recovery algorithm can be applied to reconstruct a hyperspectral data cube from compressive measurements. However, as we will see, it is important to exercise some caution due to the challenges posed by certain characteristics of practical compressive hyperspectral imagers.

Together with the development of efficient sparse recovery algorithms, there has also been significant recent progress on conditions that ensure that these algorithms obtain provably accurate estimates of the original signal f. One of the more common assumptions is that the sensing matrix A satisfies the *restricted isometry property (RIP)*, which essentially requires that $\|Af\|_2 \approx \|f\|_2$ for any k-sparse f (i.e., for any f such that we can write $f = D\theta$ where θ has at most k nonzeros). Directly constructing a matrix A that satisfies this property turns out to be rather difficult, but it is possible to show that if we construct A at *random*, then with high probability it will satisfy the RIP. While a variety of random constructions exist, perhaps the simplest (and most relevant to practical compressive hyperspectral imaging systems) is the so-called "Rademacher ensemble," where each entry of A is set to be either $1/\sqrt{n}$ or $-1/\sqrt{n}$ with equal probability. Constructing A in this fashion will, with high probability, lead to a matrix satisfying the RIP, provided that $n = O(k \log(d/k))$ [27]. Given such an A and measurements $y = Af + w$, both the approach in (1) as well as methods like CoSaMP and IHT satisfy a performance guarantee of the form

$$\|\theta - \widehat{\theta}\|_2 \leq C_1 \|w\|_2 + C_2 \frac{\|\theta - \theta_k\|_1}{\sqrt{k}}, \tag{2}$$

where θ_k is the best possible k-sparse approximation to the original θ and C_1, C_2 are absolute constants. From $\widehat{\theta}$, we can then obtain the estimate $\widehat{f} = D\widehat{\theta}$, and when D is an orthonormal basis we can translate this guarantee on $\widehat{\theta}$ into one on $\widehat{f}$. Further discussion regarding what can be proven for more specific noise models and in the specific context of compressive hyperspectral imagers is provided below.

---

of A to be $\pm 1/\sqrt{n}$ with equal probability. Unfortunately, in the context of compressive imaging, such a sensing matrix cannot be implemented. In particular, we can think of A as describing how light is propagated through a linear optical system, so that $A_{i,j}$ denotes the fraction of the total amount of light from the jth voxel in the hyperspectral image that contributes to the ith measurement. Clearly, the fractions cannot have negative values, so $A_{i,j} \geq 0$. Furthermore, the total amount of light sensed cannot be greater than the amount of light incident upon the system (*i.e., photon flux* must be preserved); mathematically, this has several consequences. Most generally, this means that if a_j denotes the jth column of A, then we must have $\|a_j\|_1 \leq 1$, since the entries in a_j correspond to fractions of f_j. This ensures that the total photon flux is preserved, i.e., $\|Af\|_1 \leq \|f\|_1$ for all f (where f, denoting the intensity of light at different locations and wavelengths, also consists solely of nonnegative elements). In particular imaging systems, there are additional constraints on the entries $A_{i,j}$. For instance, in the single pixel camera ar-

chitecture, if we assume that each measurement is allocated an equal amount of time, then the maximum possible value for $A_{i,j}$ is $1/n$ (since only $1/n$ of the total amount of light is available during each measurement period).

These restrictions lead to a small gap between the hyperspectral imaging setting and the standard theoretical treatment of CS. While it is possible to develop a specially tailored theory for certain classes of matrices with nonnegative entries, and ultimately obtain bounds similar to (2) or (3), it is perhaps more instructive to consider how to relate the desired RIP matrix A with $\pm 1/\sqrt{n}$ entries to a physically realizable matrix $\widetilde{A}$ with entries of zero or $1/n$ (with equal probability). Specifically, one can imagine constructing $\widetilde{A}$ by adding $1/\sqrt{n}$ to each element of A to make each element either zero or $2/\sqrt{n}$, and then rescaling by $1/(2\sqrt{n})$ to obtain a matrix with entries of zero or $1/n$. In the i.i.d. Gaussian measurement noise model from above, the impact of this shifting and

renormalization is that we can write our measurements as

$$y = \widetilde{A}f + w = \frac{Af}{2\sqrt{n}} + \frac{\|f\|_1}{2n} + w; \qquad (4)$$

that is, we observe a scaled version of what we would ideally like to measure (Af) plus a constant (DC) offset proportional to the total amount of light in the scene. The constant offset introduces some unique and non-trivial challenges. As we describe below, it has a significant impact on the noise variance in photon-limited settings. However, even in photon-rich settings, where we may accurately adopt a Gaussian noise assumption, the constant offset may cause challenges.

First, consider recovering f from y using the standard sparse recovery methods described in the sidebar *Sparse recovery: Methods and guarantees*. The nonnegativity of A can lead to some important algorithmic challenges when the recovery algorithm has been specifically designed under the assumption that A satisfies the RIP. In particular, one of the consequences of the RIP is that $A^T A$ acts like an isometry when applied to sparse vectors. This fact is explicitly exploited by greedy algorithms which make decisions based on $A^T y$, and sometimes implicitly exploited by some ℓ_1-minimization solvers to speed convergence. Unfortunately, this is no longer the case when the entries of A are nonnegative, since in this case all the columns of A are correlated with each other. For the algorithms that rely on this fact, simply plugging y into the algorithm without any modifications will yield inaccurate reconstructions and/or slow convergence.

Fortunately, in many cases it is possible to sidestep this issue. For example, in the context of (4), if we can use the data y to accurately estimate $\|f\|_1$ (or can directly obtain an estimate of this value in advance), then we can set $y' = y - \|f\|_1/(2n)$ and then feed y' into standard sparse recovery methods. This fix can significantly improve the speed and accuracy of reconstruction (although this approach can have significant noise implications in the low-light regime, see [31] for details).

Alternatively, it is also often relatively straightforward to modify the algorithm to rely less heavily on the RIP assumption. For example, greedy algorithms can be modified by replacing A^T with the pseudoinverse $A^\dagger = A^T(AA^T)^{-1}$. More generally, this can be viewed as a special case of *preconditioning* the data y, which is shown to significantly improve reconstruction accuracy [32, 33]. Note that it is also possible to modify standard sparse recovery methods to enforce nonnegativity in f as well [34].

Quantization and dynamic range. A more significant challenge posed by nonnegativity arises due to the fact that physical systems must ultimately also quantize the measurements y. Typically a quantizer will have a fixed number of quantization levels arranged to cover the entire range of different values that elements of y may take. When this range is precisely known in advance, each quantization level corresponds to a small interval of different values, yielding accurate measurements. In the context of the model in (4), however, note that
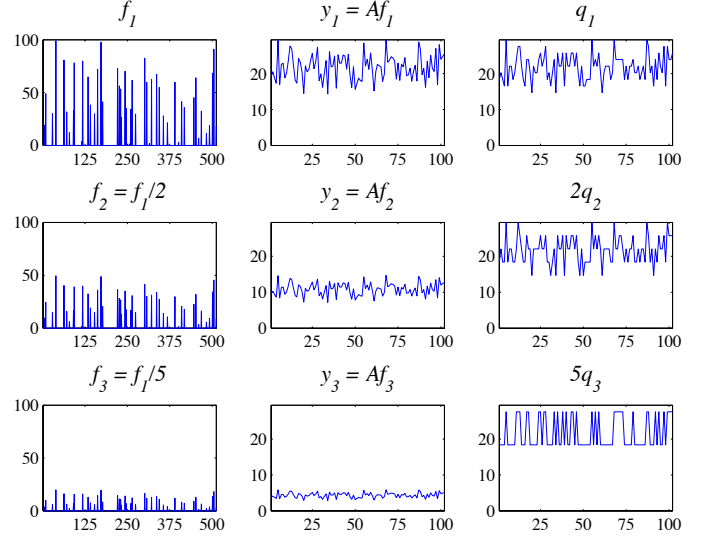

Fig. 3. *Illustration of dynamic range and quantization challenges in compressive hyperspectral imaging. Left: Depiction of the same sparse signal at three different intensity levels (brightnesses). Middle: Depiction of unquantized compressive measurements of the signals on the left using the sensing matrix construction in* (4). *Right: The quantized measurements, rescaled for easy visual comparison. We apply the same 4-bit uniform quantizer, designed to quantize values between 0 and 30, to each set of measurements. Clearly designing a quantizer capable of quantizing measurements from bright sources limits the accuracy of quantized measurements at lower intensities.*

we are actually trying to quantize small fluctuations (determined by Af) around a constant offset (determined by $\|f\|_1$) that will, in general, be unknown *a priori*. This poses a challenge when using a traditional quantizer since, if the range of the quantizer is set to be too small, the elements of y may fall outside the range of the quantizer, but if the range is too large, the small fluctuations determined by Af will fail to use the full quantization range and the system will lose precision. This is especially problematic when using a quantizer with low bit-depth. Thus, in the context of compressive hyperspectral imaging, quantization noise can be a significant source of error. A toy example illustrating this effect is presented in Figure 3, which demonstrates the challenge associated with designing a *single* mechanism for uniformly quantizing a signal with an unknown intensity or brightness, and hence an unknown constant offset.

We would like to be able to address this challenge in an automatic fashion, without resorting to manual tuning of the quantizer range for each scene of interest. One approach is to simply use very high bit-depth quantizers, but this can be costly and ultimately fails to fully address the challenge. A more robust approach is to compensate for the offset in hardware before quantizing [23, 35]. While this requires specialized sensor circuitry and can be somewhat costly, when designing a system which will be used to image scenes of widely varying brightness the improvement in performance may be

worth this increased cost.

Yet another approach to this problem relies on some of the rather unique properties of randomized measurements. In particular, the randomized measurements typically used in CS are *democratic*, roughly meaning that they each contain roughly the same amount of information, and hence by taking additional measurements we can be robust to having large errors (or even erasures) on a subset of the measurements [36]. This has a number of consequences in the context of quantization. First, while classical systems typically try to set the quantizer range to ensure that saturation occurs with extremely low probability, it has been shown empirically that in CS systems one can obtain improved performance by allowing a nontrivial number of saturation events (*e.g.,* on the order of 5–10%) [36]. Second, it allows for a particularly elegant method for automatically adjusting a quantizer to mitigate the problem described above. In particular, if the measurements are obtained sequentially in time (as in the single pixel camera architecture) then one can perform *automatic gain control* to dynamically adjust the pre-quantization gain to ensure that some desired fraction of the measurements saturate the quantizer (on both ends of the quantization range). This approach ensures that the full range of the quantizer is exploited without the need to manually measure the offset in (4), but it has the drawback of requiring a certain amount of "burn-in time" before stabilizing.

Finally, it is worth noting that as long as we can compensate for the unknown constant offset in (4), CS actually has the potential to result in significant gains over non-compressive systems in terms of quantization error and dynamic range. In particular, in a non-compressive system, we typically would quantize each voxel using the same fixed quantization range, but voxel intensity can vary *dramatically* both spatially and across spectra. This causes saturation and loss of detail in bright and dark regions of the datacube. In contrast, by combining random combinations of voxels into a single measurement, compressive systems dramatically reduce the dynamic range over which the measurements that we must quantize can fluctuate. This has been studied in the context of analog-to-digital conversion in [37] and is apparent in comparing the first and second columns of Figure 3. For a given bit-depth, this reduced range can allow for reduced quantization error in the compressive case. Exploiting this, along with the fact that by taking fewer measurements in a given time window we can use a lower-rate quantizer with a higher bit depth, there is potential for compressive systems to be *more* effective at mitigating quantization error than traditional systems.

Photon counting noise. Up to this point, we have considered the impact of noise, nonnegativity, and quantization, but only when the noise vector w is signal-independent. However, in many hyperspectral imaging contexts we are in fact photon-limited, so that the total number of photons detected by our system is small relative to the desired resolution. In photon-limited settings, we may model the observations as obeying a Poisson distribution, which has a mean equal to its variance. This effect introduces serious limitations. In particular, in (4) we saw that the signal of interest was added to a constant offset. Since the mean and variance of the noise are equal, this offset plays a critical role in controlling the noise variance.

Some of the major theoretical challenges associated with the application of CS to linear optical systems in the presence of Poisson noise have been addressed in the recent literature [38, 39]. These works considered two novel sensing paradigms, based on either pseudo-random dense sensing matrices (akin to the shifted and scaled dense sensing matrix described above) or expander graph constructions, both of which satisfy the nonnegativity and flux preservation constraints. In these settings, for a fixed signal intensity (*i.e.,* fixed $\|f\|_1$), the error bound actually *grows* linearly with the number of measurements or sensors, n, since a limited amount of light is spread across an increasing number of detectors, each with a decreasing signal-to-noise ratio (SNR). In other words, keeping n as small as possible (a central goal in CS) helps maximize SNR and reconstruction accuracy in a way not reflected in conventional CS bounds. Thus incorporating real-world constraints into the measurement model has a significant impact on the expected performance of a compressive hyperspectral imager, and these constraints should be considered carefully throughout any design process.

Imperfect system models. A major challenge in the design of compressive hyperspectral imagers is accurate knowledge of the projection operator A. While we might design a system to have a particular sensing matrix A, calibration errors and optical effects will always introduce inaccuracies. Even if we had the ability to estimate A precisely, there are settings where using an approximation of A has advantages; for instance, when we can approximately compute Af using fast Fourier transforms, conducting sparse recovery is much faster than with a dense matrix representation of A.

When we run a sparse recovery algorithm with an inaccurate sensing matrix A, it corresponds to the observation model $y_i = Af + Ef + w$, where Ef represents the difference between the *true* projections collected by hyperspectral imager and the *assumed* projections in A. The term Ef can be thought of as signal-dependent noise. Analysis of the theoretical ramifications of these kinds of errors allow the designers of spectral imagers to accurately assess tradeoffs between accurate calibration of A and computational efficiency [40].

Additional tradeoffs. One of the advantages of compressive methods for hyperspectral imaging is that they also enable a range of new design tradeoffs. For example, the single pixel camera architecture allows us to achieve high spectral resolution while trading off between spatial resolution and temporal resolution by adjusting the resolution of the patterns used by the optical modulator (a higher-resolution pattern will also require a larger total number of measurements, increasing spatial resolution at a cost of lower temporal resolution). Alternatively, the CASSI system allows for high temporal resolution

while trading off between spatial and spectral resolution. For all architectures, however, we have a fundamental tradeoff between resolution and the SNR. If we fix the temporal resolution (*i.e.,* the total acquisition time, and hence the total amount of light incident upon a hyperspectral imager), then increasing either spatial or spectral resolution means decreasing the amount of light measured for each voxel in the hyperspectral image. As resolution increases, measurements become more photon-limited and hence noisy.

HYPERSPECTRAL TARGET DETECTION FROM PROJECTION MEASUREMENTS

In addition to enabling the design of new hyperspectral imaging hardware and acquisition methods, sparsity and other low-dimensional structures provide for new ways to efficiently process the data produced by these new sensors, in some cases *without ever explicitly estimating the high-dimensional hyperspectral image* [41, 42].

In this section, we address the question of whether projections of hyperspectral images of the form $y = Af + w$ can be used to accurately and efficiently infer whether f belongs to some target class without estimating f directly. As a motivating example, consider the CASSI system discussed earlier: it collects *one* coded projection of each spectrum in the scene. One projection per spectrum is sufficient for reconstructing spatially homogeneous spectral images, since projections of neighboring locations can be combined to infer each spectrum. Significantly more projections are required for detecting targets of unknown strengths without the benefit of spatial homogeneity. One might ask how several such systems can be used in *parallel* to reliably detect spectral targets and anomalies from different coded projections.

Hyperspectral imaging introduces several unique challenges. For instance, in remote sensing applications each measured spectrum reflects the mixing of multiple spectra across a relatively large physical area – so that the spectrum of interest may be mixed with other spectra in unknown proportions. A *mixed pixel* model accounts for such interferences by modeling every spatial location as either a target material corrupted by background, or just background [43]. This background may be modeled using a multivariate Gaussian distribution: $b \sim \mathcal{N}(0, \Sigma_b)$, so that we have mixed observations according to

$$y_m = A(f + b) + w = y + Ab. \quad (5)$$

Thus in the mixed pixel setting our ideal compressive observations are contaminated by Ab, which suggests that the statistics of b must be considered when choosing A.

One approach to this challenge is to apply a pre-whitening filter $P \in \mathbb{R}^{n \times n}$ to the mixed observations y_m, with the goal of mitigating the effects of the background b. The pre-whited observations can be expressed as $z = Py_m = \widetilde{A}f + \widetilde{w}$, where $\widetilde{w}$ is white Gaussian noise with variance one and $\widetilde{A} = PA$.

This suggests choosing the hyperspectral camera optical design, described by A, in a way that depends on the background covariance Σ_b, so that the product PA facilitates accurate compressive signal classification and detection (*e.g.,* a random $n \times d$ matrix with i.i.d. $\mathcal{N}(0,1)$ entries, commonly considered in the CS literature) [44]. This approach naturally leads to constraints on the amount of background tolerated by a target detection method.

Target dictionaries. The goal of hyperspectral target detection is, in the context of mixed observations, to determine whether $f = 0$ (*i.e.,* no target and only background is present) or which f in a dictionary of target spectral signatures $\mathcal{D}$ corresponds to the observations.

Theoretical performance bounds provide key insight into how error rates scale with the number of measurements collected, the spectral resolution of targets, the amount of background signal present, the signal-to-noise ratio, and properties of $\mathcal{D}$. In particular, let ρ denote the minimum Euclidean distance between any two target spectra in the target class $\mathcal{D}$, and let $|\mathcal{D}|$ denote the size of the dictionary. Performance can be characterized in terms of a method's *positive false discovery rate (pFDR)*, which measures the fraction of declared targets that are false alarms and is a useful metric in multiple testing scenarios such as this.

A target detection method based on a nearest-neighbor approach applied to pre-whited measurements z yields the bound

$$\mathrm{pFDR} = O\left[\frac{1}{|\mathcal{D}|} \left(\left[1 + \frac{\rho^2}{4n} \right]^{n/2} - \frac{1}{|\mathcal{D}|} \right)^{-1} \right], \quad (6)$$

which decays with the number of measurements n and the size of the target dictionary, but increases with ρ. Thus introducing new candidate targets which are very similar to existing candidate targets can significantly deteriorate performance, regardless of the spectral resolution d. Experimental results show that using these theoretically-supported designs of A yields significantly better target detection accuracy than simply measuring low-resolution hyperspectral images [44].

Target manifolds. The fixed-dictionary hyperspectral target detection problem formulation above is accurate if the signals in the dictionary are faithful representations of the target signals that we observe. In reality, however, the target signals will differ from those in the dictionary due to the differences in the experimental conditions under which they are collected. For instance, in remote sensing applications, the observed spectrum of a material will not match the reference spectrum observed in a laboratory due to differences in atmospheric and illumination conditions. In this case, one could reasonably model the target signals observed under different experimental conditions as lying in a low-dimensional submanifold of the high-dimensional ambient signal space; this has been shown to be true for hyperspectral images in [45].

Thus in many practical settings, rather than differentiate

among a finite collection of candidate spectra, we must differentiate among a collection of candidate target manifolds. Target detection in this setting has two key components: (a) a search for the closest point in each candidate target manifold to the observation, followed by (b) a minimum distance-based detection step controlled by the desired false alarm probability. It is now known that the randomized projections common in compressive sensing also preserve the structure of the manifold; this can be shown by adapting the earlier Johnson-Lindenstrauss lemma argument to a sufficiently dense sampling of the manifold [46]. In this case, nearest-neighbor target detection can be executed in the compressive domain. This approach has been dubbed *smashed filtering* in [41].

Anomaly detection. While in many settings target dictionaries can be formed in a laboratory or using "ground truth" data (usually collected at considerable expense and time), at times target dictionaries are simply unavailable. In such settings, one might be interested in detecting objects *not* in the dictionary. Here, the target signals of interest are anomalous and are not known *a priori* to the user. The target detection methods discussed above can be extended to anomaly detection by exploiting the distance preservation property of the sensing matrix A to detect anomalous targets from projection measurements, as detailed in [44, 47, 48].

CONCLUSIONS AND FUTURE DIRECTIONS

Due to the enormous size of hyperspectral images with high spatial and spectral resolution, approaches that enable efficient data collection, signal reconstruction, and target detection tasks have enormous practical potential. The good news is that typical hyperspectral images have significant structure that can be exploited within the context of sparse models and CS. Armed with such models, we can engineer novel compressive sensors and reconstruction algorithms.

On the surface, the application of the CS theory and algorithms to hyperspectral imaging appears very promising. However, one of the central themes of this article is that these theories and methods cannot be applied blindly to this application arena. For a compressive hyperspectral imaging design to be truly effective, it must account for the physical constraints of the measurements system, use appropriate quantization methods, accommodate realistic noise models (including photon noise, background signal effects, and calibration errors), and use reconstruction algorithms that specifically account for all of these effects. None of these aspects can be considered in isolation, and any system design which ignores these issues has limited potential.

Despite these caveats, researchers are pushing the boundaries of our collective knowledge of how to exploit signal structure for improved sensing and inference. For example, while sequentially selecting the rows of A in an adaptive fashion is of limited benefit in some of the hardest possible sparse recovery problems [29, 30], in high SNR regimes or settings where we have structured or group sparsity (common in hy-

perspectral imaging), adaptivity can potentially yield significant gains. Exploring the applications of these ideas to practical imaging systems is an important area of ongoing research.

AUTHORS

Rebecca Willett is an Associate Professor in the Electrical and Computer Engineering Department at the University of Wisconson-Madison. She was an assistant then associate professor of Electrical and Computer Engineering at Duke University between 2005 and 2013. She received the NSF CAREER award in 2007 and an AFOSR Young Investigator award in 2010.

Marco F. Duarte is an Assistant Professor in the Department of Electrical and Computer Engineering at the University of Massachusetts Amherst. His research interests include machine learning, compressed sensing, sensor networks, and optical image coding. Prof. Duarte received the Presidential Fellowship and the Texas Instruments Distinguished Fellowship in 2004 and the Hershel M. Rich Invention Award in 2007, all from Rice University. He coauthored (with C. Hegde and V. Cevher) the Best Student Paper at the 2009 International Workshop on Signal Processing with Adaptive Sparse Structured Representations (SPARS).

Mark A. Davenport is an Assistant Professor with the School of Electrical and Computer Engineering, Georgia Institute of Technology. His research interests include compressive sensing, low-rank matrix recovery, nonlinear approximation, and the application of low-dimensional signal models in signal processing and machine learning. In 2007 he received the Hershel M. Rich Invention Award and in 2010 he was awarded the Ralph Budd Thesis Award, both from Rice University.

Richard G. Baraniuk is the Victor E. Cameron Professor of Electrical and Computer Engineering at Rice University. His research interests lie in the areas of signal processing, sensors, and machine learning. He was elected a Fellow of the IEEE in 2001 and of AAAS in 2009.

ACKNOWLEDGEMENTS

REFERENCES

[1] S. Lim, K. H. Sohn, and C. Lee, "Principal component analysis for compression of hyperspectral images," in *IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Sydney, Australia, July 2001.

[2] C. Rodarmel and J. Shan, "Principal component analysis for hyperspectral image classification," *Surveying Land Inf. Sci.*, vol. 62, no. 2, pp. 115–122, 2002.

[3] M. Maggioni, V. Katkovnik, K. Egiazarian, and A. Foi, "A nonlocal transform-domain filter for volumetric data denoising and reconstruction," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 119–133, 2013.

[4] J. Salmon, Z. Harmany, C. Deledalle, and R. Willett, "Poisson noise reduction with non-local PCA," *to appear in J. Math. Imag. Vision*, 2013.

[5] M. Makitalo and A. Foi, "Optimal inversion of the Anscombe transformation in low-count Poisson image denoising," *IEEE Trans. Image Process.*, vol. 20, no. 1, pp. 99–109, 2011.

[6] J. Fowler, "Compressive-projection principal component analysis," *IEEE Trans. Image Proc.*, vol. 18, no. 10, pp. 2230–2242, Oct. 2009.

[7] A. Charles, B. Olshausen, and C. Rozell, "Learning sparse codes for hyperspectral imagery," *IEEE J. Select. Top. Signal Process.*, vol. 5, no. 5, pp. 963–978, 2011.

[8] J. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Select. Top. Appl. Earth Obs. Remote Sens.*, vol. 5, no. 2, pp. 354–379, 2012.

[9] P. Dragotti, G. Poggi, and A. Ragozini, "Compression of multispectral images by three-dimensional SPIHT algorithm," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 1, pp. 416–428, Jan. 2000.

[10] J. Fowler and J. Rucker, "3D wavelet-based compression of hyperspectral imagery," in *Hyperspectral Data Exploitation: Theory and Applications*, pp. 379–407. Wiley, Hoboken, NJ, 2007.

[11] M. Duarte and Y. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053–4085, 2011.

[12] W. Li, S. Prasad, and J. Fowler, "Integration of spectral-spatial information for hyperspectral image reconstruction from compressive random projections," *to appear in IEEE Geosci. Remote Sens. Lett.*, 2013.

[13] M. Golbabaee, S. Arberet, and P. Vandergheynst, "Compressive source separation: Theory and methods for hyperspectral imaging," *arxiv:1208.4505*, 2012.

[14] J. Ma, "A single-pixel imaging system for remote sensing by two-step iterative curvelet thresholding," *IEEE Remote Sensing Letters*, vol. 6, no. 4, pp. 676–680, Oct. 2009.

[15] T. Sun, D. Takhar, J. Laska, M. Duarte, V. Bansal, R. Baraniuk, and K. Kelly, "Realization of confocal and hyperspectral microscopy via compressive sensing," *Bull. Am. Phys. Soc.*, vol. 53, no. 2, 2008.

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

[17] Y. August, C. Vachman, Y. Rivenson, and A. Stern, "Compressive hyperspectral imaging by random separable projections in both the spatial and the spectral domains," *Applied Optics*, vol. 52, no. 10, pp. D46–D54, Apr. 2013.

[18] M. Gehm, R. John, D. Brady, R. Willett, and T. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Opt. Express*, vol. 15, no. 21, pp. 14013–14027, 2007.

[19] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Appl. Opt.*, vol. 47, no. 10, pp. B44–B51, 2008.

[20] G. Arce, D. Brady, L. Carin, H. Arguello, and D. S. Kittle, "An introduction to compressive coded aperture spectral imaging," *IEEE Signal Process. Mag.*, Jan. 2014.

[21] L.-L. Xiao, K. Liu, and D.-P. Han, "CMOS low data rate imaging method based on compressed sensing," *Opt. Laser Tech.*, vol. 44, no. 5, pp. 1338–1345, 2012.

[22] M. Dadkhah, M. Deen, and S. Shirani, "Block-based compressive sensing in a CMOS image sensor," *to appear in IEEE Sensors J.*, 2013.

[23] Y. Oike and A. El Gamal, "CMOS image sensor with per-column $\Sigma\Delta$ ADC and programmable compressed sensing," *IEEE J. Solid-State Circuits*, vol. 48, no. 1, pp. 318–328, Jan. 2013.

[24] J. Tropp and S. Wright, "Computational methods for sparse solution of linear inverse problems," *Proc. IEEE*, vol. 98, no. 6, pp. 948–958, 2010.

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

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

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

[28] E. Candès and M. Davenport, "How well can we estimate a sparse vector?," *Appl. Comput. Harmon. Anal.*, vol. 34, no. 2, pp. 317–323, 2013.

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

[30] E. Arias-Castro, E. Candès, and M. Davenport, "On the fundamental limits of adaptive sensing," *IEEE Trans. Inform. Theory*, vol. 59, no. 1, pp. 472–481, 2013.

[31] R. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–120, 124, 2007.

[32] S. Xiong, B. Dai, and P. Z. G. Qian, "Orthogonalizing penalized regression," arXiv:1108.0185, 2011.

[33] J. Jia and K. Rohe, "Preconditioning to comply with the irrepresentable condition," *arXiv:1208.5584*, 2012.

[34] Z. Harmany, D. Thompson, R. Willett, and R. Marcia, "Gradient projection for linearly constrained convex optimization in sparse signal recovery," in *Proc. IEEE Int. Conf. Image Proc. (ICIP)*, Hong Kong, Sep. 2010.

[35] V. Majidzadeh, L. Jacques, A. Schmid, P. Vandergheynst, and Y. Leblebici, "A (256 × 256) pixel 76.7 mW CMOS imager/compressor based on real-time in-pixel compressive sensing," in *Proc. IEEE Int. Symp. Circuits and Systems (ISCAS)*, Paris, France, May 2010.

[36] J. Laska, P. Boufounos, M. Davenport, and R. Baraniuk, "Democracy in action: Quantization, saturation, and compressive sensing," *Appl. Comput. Harmon. Anal.*, vol. 31, no. 3, pp. 429–443, 2011.

[37] M. Davenport, J. Laska, J. Treichler, and R. Baraniuk, "The pros and cons of compressive sensing for wideband signal acquisition: Noise folding vs. dynamic range," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4628–4642, 2012.

[38] M. Raginsky, R. Willett, Z. Harmany, and R. Marcia, "Compressed sensing performance bounds under Poisson noise," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 3990–4002, 2010.

[39] M. Raginsky, S. Jafarpour, Z. Harmany, R. Marcia, R. Willett, and R. Calderbank, "Performance bounds for expander-based compressed sensing in Poisson noise," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4139–4153, 2011.

[40] P.-L. Loh and M. Wainwright, "High-dimensional regression with noisy and missing data: provable guarantees with non convexity," *Ann. Stat.*, vol. 40, no. 3, pp. 1637–1664, 2012.

[41] M. Davenport, M. Duarte, M. Wakin, J. Laska, D. Takhar, K. Kelly, and R. Baraniuk, "The smashed filter for compressive classification and target recognition," in *Proc. IS&T/SPIE Symp. Elec. Imag.: Comp. Imag.*, San Jose, CA, Jan. 2007.

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

[43] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, 2002.

[44] K. Krishnamurthy, R. Willett, and M. Raginsky, "Target detection performance bounds in compressive imaging," *to appear in EURASIP J. Adv. Signal Processing*, 2013.

[45] G. Healey and D. Slater, "Models and methods for automated material identification in hyperspectral imagery acquired under unknown illumination and atmospheric conditions," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2706–2717, 1999.

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

[47] J. Fowler and Q. Du, "Anomaly detection and reconstruction from random projections," *IEEE Transactions on Image Processing*, vol. 21, no. 1, pp. 184 –195, Jan. 2012.

[48] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Effects of linear projections on the performance of target detection and classification in hyperspectral imagery," *J. Applied Remote Sensing*, vol. 5, no. 1, Nov. 2011.