Kronecker Product Matrices for Compressive Sensing

Marco F. Duarte and Richard G. Baraniuk*

Technical Report TREE-1105
Department of Electrical and Computer Engineering
Rice University

March 17, 2011

Abstract

Compressive sensing (CS) is an emerging approach for the acquisition of signals having a sparse or compressible representation in some basis. While the CS literature has mostly focused on problems involving 1-D signals and 2-D images, many important applications involve multidimensional signals; in this case, CS works best with representations that encapsulate the structure of such signals in every dimension. We propose the use of Kronecker product matrices in CS for two purposes. First, such matrices can act as sparsifying bases that jointly model the different types of structure present in the signal. Second, the measurement matrices used in distributed settings can be easily expressed as Kronecker product matrices. The Kronecker product formulation in these two settings enables the derivation of analytical bounds for sparse approximation of multidimensional signals and CS recovery performance as well as a means to evaluate novel distributed measurement schemes.

1 Introduction

1.1 CS and multidimensional signals

Compressive sensing (CS) is a new approach to simultaneous sensing and compression that enables a potentially large reduction in the sampling and computation costs at a sensor for a signal $\mathbf{x}$ having a sparse or compressible representation θ in some basis Ψ (i.e. $\mathbf{x} = \Psi\theta$) [1, 2]. By a sparse representation, we mean that only K out of the N signal coefficients in θ are nonzero, with $K \ll N$. By a compressible representation, we mean that the coefficient's magnitudes, when sorted, have a fast power-law decay, i.e.,

$$|\theta(i)| < Ci^{-1/p} \tag{1}$$

for $p \leq 1$ and $C < \infty$. Many natural signals are sparse or compressible; for example, smooth signals are compressible in the Fourier basis, while piecewise smooth signals and images are compressible in a wavelet basis.

CS builds on the work of Candès, Romberg, and Tao [1] and Donoho [2], who showed that a signal having a sparse or compressible representation in one basis can be recovered from its linear projections onto a small set of $M = \mathcal{O}(K \log(N/K))$ measurement vectors that are incoherent with the sparsifying basis, meaning that the representation of the measurement vectors in this basis is not sparse. When the measurement vectors are stacked as rows of a measurement matrix Φ, the CS measurements can be expressed as a vector

$$\mathbf{y} = \Phi \mathbf{x} \tag{2}$$

of length M. CS acquisition devices multiplex the signal, as they read inner products of the signal vector against the measurement vectors instead of reading the signal vector itself [3]. We can obtain a compressed representation of the signal by obtaining a number of inner products smaller than the signal length. Random vectors play a central role as universal measurements in the sense that they are incoherent with any fixed basis with high probability. The CS measurement process is nonadaptive, and the recovery process is nonlinear; there exist a variety of CS recovery algorithms inspired by sparse approximation techniques [1, 2, 4, 5]. To recover the signal from the measurements, we search for the sparsest signal among all those that yield the observed measurement values.

The CS literature has mostly focused on problems involving single sensors and one-dimensional (1-D) signal or 2-D image data. However, some important applications that hold the most promise for CS involve higher-dimensional signals. The coordinates of these signals may span several physical, temporal, or spectral dimensions. Additionally, these signals are often captured in a progressive fashion, in a sequence of captures corresponding to subsets of the coordinates. Examples include hyperspectral imaging (with spatial and spectral dimensions), video acquisition (with spatial and temporal dimensions), and synthetic aperture radar imaging (with progressive acquisition in the spatial dimensions). Another class of promising applications for CS involves distributed networks or arrays of sensors, including for example environmental sensors, microphone arrays, and camera arrays.

These properties of multidimensional data and the corresponding acquisition hardware complicate the design of both the measurement matrix Φ and the sparsifying basis Ψ to achieve maximum efficiency in CS.

1.2 CS measurement matrices for multidimensional signals

For signals of any dimension, *global* CS measurements that multiplex most or all of the values of the signal together (corresponding to dense matrices Φ) are required for universality, since they are needed to capture arbitrary sparsity structure [6]. However, for multidimensional signals, such measurements require the use of multiplexing sensors that operate simultaneously along all data dimensions, increasing the physical complexity or acquisition time/latency of the CS device. In many settings it can be difficult to implement such multiplex sensors due to the large dimensionality of the signals involved and the ephemeral availability of the data during acquisition. For example, each image frame in a video sequence is available only for a limited time; therefore, any multiplexing sensor that calculates global CS measurements would have to sum of the M partial inner products from each of the frames from the beginning to the end of the video sequence. Similarly, global CS

measurements of a hyperspectral datacube would require simultaneous multiplexing in the spectral and spatial dimensions, which is a challenge with current optical and spectral modulators [7, 8]; such separate multiplexing nature limits the structure of the measurements obtained.

These application-specific limitations naturally point us in the direction of measurement systems Φ that depend only on a portion of the entries of the multidimensional signal being acquired. Many applications and practical hardware designs demand *partitioned measurements* that process only a portion of the multidimensional signal at a time. Each portion usually corresponds to a section of the signal along a given dimension, such as one frame in a video signal or the image of one spectral band of a hyperspectral datacube.

1.3 Sparsifying bases for multidimensional signals

For multidimensional signals, we can often characterize the signal structure present on each of its different dimensions or coordinates in terms of a sparse representation. For example, each image frame in a video sequence is often sparse or compressible in a wavelet basis, since it corresponds to an image obtained at a particular time instant. Simultaneously, the temporal structure of each pixel in a video sequence is often smooth or piecewise smooth, due to camera movement, object motion and occlusion, illumination changes, etc. A similar situation is observed in hyperspectral signals: the reflectivity values at a given spectral band correspond to an image with known structure; additionally, the spectral signature of a given pixel is usually smooth or piecewise smooth, depending on the spectral range and materials present in the observed area.

Initial work on the sparsity and compressibility of multidimensional signals and signal ensembles for CS [9–21] has provided new sparsity models for multidimensional signals. These models consider sections of the multidimensional data corresponding to fixed values for a subset of the coordinates as separate signals and pose correlation models between the values and locations of their sparse representations. To date, the resulting models are rather limited in the types of structures admitted. For almost all models, theoretical guarantees on signal recovery have been provided only for strictly sparse signals, for noiseless measurement settings, or in asymptotic regimes. Additionally, almost all of these models are tied to ad-hoc signal recovery procedures.

Clearly, more generic models for sparse and compressible multidimensional signals are needed in order to leverage the CS framework to a higher degree of effective compression. Ideally, we should be able to formulate a sparsifying basis for the entire multidimensional signal that simultaneously accounts for all the types of structure present in the data.

In this paper, we show that *Kronecker product matrices* offer a natural means to generate both sparsifying bases Ψ and measurement matrices Φ for CS of multidimensional signals, resulting in a formulation that we dub *Kronecker Compressive Sensing* (KCS). Kronecker product sparsifying bases combine the structures encoded by the sparsifying bases for each signal dimension into a single matrix. Kronecker product measurement matrices can be implemented by performing a sequence of separate multiplexing operations on each signal dimension. The Kronecker product formulation for sparsifying bases and measurement matrices enables the derivation of analytical bounds for recovery of compressible multidimensional signals from randomized or incoherent measurements.

1.4 Stylized applications

To better motivate the KCS concept, we consider now in more detail three relevant multidimensional CS applications: hyperspectral imaging, video acquisition, and distributed sensing.
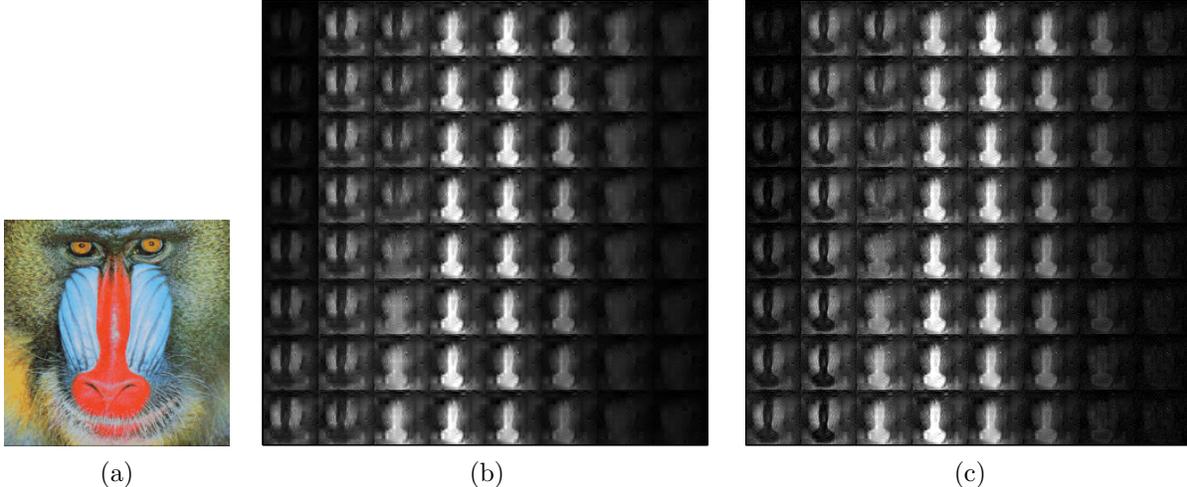
Figure 1: *Example capture from a single-pixel hyperspectral camera [22] at resolution $N = 128 \times 128$ pixels $\times 64$ spectral bands (2^{20} voxels) for the 450nm–850nm wavelength range from $M = 4096$ CS measurements per band (4× sub-Nyquist) [23]. (a) Mandrill test image printed and illuminated by a desk lamp for acquisition. (b) Hyperspectral datacube obtained via independent CS recovery of each spectral band as a separate image. (c) Hyperspectral datacube obtained via KCS; marked improvement is seen in bands with low signal-to-noise ratios. Data courtesy of Kevin Kelly, Ting Sun, and Dharmpal Takhar from Rice University.*

1.4.1 Hyperspectral imaging

Consider the single-pixel hyperspectral camera (SPHC) [7, 22], where the hyperspectral lightfield is focused onto a digital micromirror device (DMD). The DMD acts as a optical spatial modulator and reflects part of the incident lightfield into a scalar spectrometer. In this way, the DMD computes inner products of the image of each spectral band in the hyperspectral lightfield against a measurement vector with 0/1 entries, coded in the orientation of the mirrors. Each spectral band's image is multiplexed by the same binary functions, since the DMD reflects all of the imaged spectra simultaneously. This results in the same measurement matrix Φ being applied to each spectral band image. The resulting measurement matrix applied to the hyperspectral datacube can be represented as a Kronecker product $\mathbf{I}_S \otimes \Phi$, where $\mathbf{I}_S$ is an $S \times S$ identity matrix and S denotes the number of spectral bands recorded. Additionally, there are known sparsifying bases for each spectral band image as well as each pixel's spectral signature, which can be integrated into a single Kronecker product sparsifying basis. An example datacube captured with a SPHC via KCS is shown in Fig. 1 [23].

1.4.2 Video acquisition

Consider the example of compressive video acquisition, where a single-pixel camera applies the same set of measurements to each image frame in the video sequence, resulting once again in a Kronecker product measurement matrix [10]. We can sparsify or compress the temporal structure at each pixel using a Fourier or wavelet transform depending on the video characteristics. Furthermore, we can sparsify each image frame using a standard cosine or wavelet transform. We can then use a Kronecker product of these two bases to sparsify or compress the video sequence.

4

1.4.3 Distributed sensing

In distributed sensing problems, we aim to acquire an ensemble of signals $\mathbf{x}_1, \ldots, \mathbf{x}_J \in \mathbb{R}^N$ that vary in time, space, etc. We assume that each signal's structure can be encoded using sparsity with an appropriate basis Ψ_1. This ensemble of signals can be expressed as a $N \times J$ matrix $\mathbf{X} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \ldots \ \mathbf{x}_J] = [\mathbf{x}^{1^T} \ \mathbf{x}^{2^T} \ \ldots \ \mathbf{x}^{N^T}]^T$, where the individual signals $\mathbf{x}_1, \ldots, \mathbf{x}_j$ corresponding to columns of the matrix, and where the rows $\mathbf{x}^1, \ldots, \mathbf{x}^N$ of the matrix correspond to different *snapshots* of the signal ensemble at different values of time, space, etc. Under this construction, the structure of each signal is observable on each of the columns of the matrix, while the structure of each snapshot (spanning all the signals) is present on each of the rows of the matrix $\mathbf{X}$.

We expect that, in certain applications, the snapshot structure can also be modeled using sparsity; that is, that a basis or frame Ψ_2 can be used to compress or sparsify $\mathbf{x}^1, \ldots \mathbf{x}^N$. For example, in sensor network applications, the structure of each snapshot is determined by the geometry of the sensing deployment, and can also be captured by a sparsity basis [24]. In such cases, we obtain a single sparsifying basis $\Psi_1 \otimes \Psi_2$ for the signal ensemble $\mathbf{x}$ that encodes the structure of both the signals and the snapshots; such representation significantly simplifies the analysis of the signal ensemble sparsity and compressibility. Furthermore, if separate measurements $\mathbf{y}_j = \Phi \mathbf{x}_j$ of each signal are obtained using the same measurement Φ, we can express the resulting measurement matrix acting on the signal ensemble as the Kronecker product $\mathbf{I}_J \otimes \Phi$.

1.5 Contributions

This paper has three main contributions. First, we propose Kronecker product matrices as sparsifying bases for multidimensional signals to jointly model the signal structure along each one of its dimensions. In some cases, such as Kronecker product wavelet bases, we can to obtain bounds for the magnitude rate of decay of the signal coefficients for certain kinds of data. This rate of decay is dependent on the rates of decay for the coefficients of sections of the signals across the different dimensions using the individual bases. When the rates of decay using the corresponding bases for each of the dimensions are different, we show that the Kronecker product basis rate falls between the maximum and minimum rates among the different dimensions; when the rates of decay are all the same, they are matched by that of the Kronecker product basis.

Second, we show that several different CS measurements schemes proposed for multidimensional signals can be easily expressed in our Kronecker product framework. In particular, when partitioned measurements are used and the same measurement matrix is applied to each portion of the signal, the resulting measurement matrix can be expressed as the Kronecker product of an identity matrix with the measurement matrix. We can also build new D-stage CS acquisition devices that use Kronecker measurement matrices: the first stage applies the same lower-dimensional measurement matrix on each portion of the signal along its first dimension, and each subsequent stage applies additional low-dimensional measurement matrices on previously obtained measurements along the remaining dimensions of the signal. The resulting measurement matrix for the high-dimensional signal is simply the Kronecker product of the low-dimensional matrices used at each stage.

Third, we provide metrics to evaluate partitioned measurement schemes against Kronecker measurement matrices, as well as guidance on the improvements that may be afforded by the use of such multidimensional structures. In particular, we provide some initial results by studying the special case of signals that are compressible in a Kronecker product of wavelet bases. We also compare the rate of decay for the recovery error of KCS to the rate of decay for the recovery error of

standard CS recovery from separate measurements of each portion of the signal. Finally, we verify our theoretical findings using experimental results with synthetic and real-world multidimensional signals.

This paper is organized as follows. Section 2 provides background in compressive sensing and tensor and Kronecker products. Section 3 introduces Kronecker compressive sensing, and Section 4 provides initial results for wavelet-sparse signals. Section 5 provides experimental results, Section 6 summarizes related work, and Section 7 closes the document with conclusions and suggestions for future work.

2 Background

2.1 Compressive sensing

Compressive sensing (CS) is a efficient signal acquisition framework for signals that are sparse or compressible in an appropriate domain. Let $\mathbf{x} \in \mathbb{R}^N$ be the signal of interest. We say that an orthonormal basis[1] $\Psi \in \mathbb{R}^{N \times N}$ sparsifies the signal $\mathbf{x}$ if $\theta = \Psi^T \mathbf{x}$ has only K nonzero entries, with $K \ll N$ and Ψ^T denoting the transpose of Ψ. We then say that $\mathbf{x}$ is K-sparse or has sparsity K in Ψ. Similarly, we say that Ψ compresses $\mathbf{x}$ if the entries of θ, when sorted by magnitude, decay according to (1). In this case, we say that θ is in weak ℓ_p (noted as $\theta \in w\ell_p$) or, alternatively, that θ is s-compressible in Ψ, with $s = 1/p - 1/2$. Such vectors can be compressed using *transform coding* by preserving only the coefficients with largest absolute magnitude; we term by θ_K the approximation with the K largest coefficients of θ. Thus, for a K-sparse signal, the approximation error $\sigma_K(\theta) := \|\theta - \theta_K\|_2 = 0$, where $\| \cdot \|_2$ denotes the ℓ_2 or Euclidean norm. For s-compressible signals, $\sigma_K(\theta) \leq C'K^{-s}$, i.e., the approximation error decays exponentially. Many types of signals of interest are known to be compressible in appropriate bases. For example, smooth signals such as audio recordings are compressible in the Fourier basis, and piecewise smooth signals such as natural images are compressible in a wavelet basis.

The CS acquisition procedure consists of measuring inner products of the signal against a set of measurement vectors $\{\phi_1, \ldots, \phi_M\}$; when $M < N$, the acquisition procedure effectively compresses the signal. By collecting the measurement vectors as rows of a measurement matrix $\Phi \in \mathbb{R}^{M \times N}$, the acquisition procedure can be written as $\mathbf{y} = \Phi\mathbf{x} = \Phi\Psi\theta$, with the vector $\mathbf{y} \in \mathbb{R}^M$ containing the CS measurements.

The goal of CS is to recover the full signal $\mathbf{x}$ from the fewest possible measurements $\mathbf{y}$. Infinitely many vectors $\mathbf{x}$ can yield the recorded measurements $\mathbf{y}$ due to the rank deficiency of the matrix $\Upsilon = \Phi\Psi$. One of the main enablers of CS was the discovery that when the signal being observed is sparse enough, it can be exactly recovered by solving the linear program [1, 2, 25]

$$\widehat{\theta} = \arg\min \|\theta\|_1 \text{ s.t. } \mathbf{y} = \Upsilon\theta. \tag{3}$$

In this case, $\| \cdot \|_1$ denotes the ℓ_1 norm, which is equal to the sum of the absolute values of the vector entries.

In the real world, the CS measurements are corrupted by noise. This provides us with CS measurements $\mathbf{y} = \Phi\mathbf{x} + \mathbf{n}$, with $\mathbf{n}$ denoting the noise vector. In this case, the signal can also be

---

[1]In the sequel, we will use the same notation Ψ to refer to the set of basis vectors and to the matrix having these basis vectors as columns.

successfully recovered using the quadratic program [1]

$$\widehat{\theta} = \arg\min \|\theta\|_1 \text{ s.t. } \|\mathbf{y} - \Upsilon\theta\|_2 \leq \epsilon, \tag{4}$$

where ϵ is an upper bound on the ℓ_2 norm of the noise vector $\mathbf{n}$. The penalty paid is an additional distortion in the recovered version of the signal proportional to ϵ in the worst case.

Previous contributions have posed conditions on the number and type of measurement vectors necessary for signal recovery [1, 26]. The Restricted Isometry Property (RIP) has been proposed to measure the fitness of a matrix Υ for CS.

Definition 2.1 *The K-restricted isometry constant for the matrix Υ, denoted by δ_K, is the smallest nonnegative number such that, for all $\theta \in \mathbb{R}^N$ with $\|\theta\|_0 = K$,*

$$(1 - \delta_K)\|\theta\|_2^2 \leq \|\Upsilon\theta\|_2^2 \leq (1 + \delta_K)\|\theta\|_2^2.$$

Once the RIP constants are determined, they can be used to provide guarantees for CS recovery.

Theorem 2.1 *[27] If the matrix Υ has $\delta_{2K} < \sqrt{2} - 1$, then the solution $\widehat{\theta}$ to (4) obeys*

$$\|\theta - \widehat{\theta}\|_2 \leq C_0 \frac{\|\theta - \theta_K\|_1}{K^{1/2}} + C_1\epsilon,$$

where C_0 and C_1 are fixed constants dependent on δ_{2K}.

In words, Theorem 2.1 guarantees that sparse signals can be recovered perfectly from noiseless measurements; that compressible signals can be recovered to a distortion similar to that of the transform coding compression; and that the recovery process is robust to the presence of noise in the measurements. Unfortunately, calculating the RIP constants for a given matrix requires combinatorially complex computation. Interestingly, many probabilistic classes of matrices have been advocated. For example, a matrix of size $M \times N$ with independent and identically distributed normal entries with variance $1/M$ obeys the condition of Theorem 2.1 with very high probability if $K \leq \mathcal{O}\left(M/\log(N/M)\right)$ [1, 2, 6]. The same is true of matrices following Rademacher or subgaussian distributions.

In some applications, the sensing system constrains the types of measurement matrices that are feasible. This could be due either to the computational power needed to generate the matrix, or due to limitations in the sensing modalities. For example, the single pixel camera [7] uses a subset of the Hadamard transform basis vectors as a measurement matrix. To formalize this framework, we can assume that a basis $\Phi \in \mathbb{R}^{N \times N}$ is provided for measurement purposes, and we have the option to choose a subset of the signal's coefficients in this transform as measurements. That is, we let $\widetilde{\Phi}$ be an $N \times M$ submatrix of Φ that preserves the basis vectors with indices $\Gamma \subseteq \{1, \ldots, N\}$, $|\Gamma| = M$, and $\mathbf{y} = \widetilde{\Phi}^T\mathbf{x}$. In this case, a different metric arises to evaluate the performance of CS.

Definition 2.2 *The* mutual coherence *of the orthonormal bases $\Phi \in \mathbb{R}^{N \times N}$ and $\Psi \in \mathbb{R}^{N \times N}$ is the maximum absolute value for the inner product between elements of the two bases:*

$$\mu(\Phi, \Psi) = \max_{1 \leq i,j \leq N} |\langle \phi_i, \psi_j \rangle|.$$

The mutual coherence then determines the number of measurements necessary for accurate CS recovery:

Theorem 2.2 *[26] Let $\mathbf{x} = \Psi\theta$ be a K-sparse signal in Ψ with support $\Omega \subset \{1, \ldots, N\}$, $|\Omega| = K$, and with entries having signs chosen uniformly at random. Choose a subset $\Gamma \subseteq \{1, \ldots, N\}$ for the set of observed measurements, with $M = |\Gamma|$. Suppose that $M \geq CKN\mu^2(\Phi, \Psi)\log(N/\delta)$ and $M \geq C'\log^2(N/\delta)$ for fixed values of $\delta < 1$, C, C'. Then with probability at least $1 - \delta$, θ is the solution to (3).*

Since the range of possible mutual coherence values $\mu(\Phi, \Psi)$ is $[N^{-1/2}, 1]$, the number of measurements required by Theorem 2.2 ranges from $O(K\log(N))$ to $O(N)$. It is possible to expand the guarantee of Theorem 2.2 to compressible signals by adapting an argument of Rudelson and Vershynin in [28] that links mutual coherence and restricted isometry constants.

Theorem 2.3 *[28, Remark 3.5.2] Choose a subset $\Gamma \subseteq \{1, \ldots, N\}$ for the set of observed measurements, with $M = |\Gamma|$, uniformly at random. Suppose that*

$$M \geq CK\sqrt{N}t\mu(\Phi, \Psi)\log(tK\log N)\log^2 K \tag{5}$$

for a fixed value of C. Then with probability at least $1 - 5e^{-t}$ over the choice of Γ, the resulting matrix $\Phi_\Gamma^T\Psi$ has the RIP with constant $\delta_{2K} \leq \sqrt{2} - 1$, where Φ_Γ denotes the restriction of Φ to the columns indexed by Γ.

Using this theorem, we obtain the guarantee of Theorem 2.1 for compressible signals with the number of measurements M dictated by the mutual coherence value $\mu(\Phi, \Psi)$.

2.2 Tensor and Kronecker products

Let V and W represent Hilbert spaces. The *tensor product* of V and W is a new vector space $V \otimes W$ together with a bilinear map $\mathbb{T} : V \times W \to V \otimes W$ such that for every vector space X and every bilinear map $\mathbb{S} : V \times W \to X$ there is a unique linear map $\mathbb{S}' : V \otimes W \to X$ such that for all $v \in V$ and $w \in W$, $\mathbb{S}(v, w) = \mathbb{S}'(\mathbb{T}(v, w))$.

For example, the *Kronecker product* of two matrices A and B of sizes $P \times Q$ and $R \times S$, respectively, is defined as

$$A \otimes B := \begin{bmatrix} A(1,1)B & A(1,2)B & \ldots & A(1,Q)B \\ A(2,1)B & A(2,2)B & \ldots & A(2,Q)B \\ \vdots & \vdots & \ddots & \vdots \\ A(P,1)B & A(P,2)B & \ldots & A(P,Q)B \end{bmatrix}. \tag{6}$$

Thus, $A \otimes B$ is a matrix of size $PR \times QS$. The definition has a straightforward extension to the Kronecker product of vectors $a \otimes b$. In the case where $V = \mathbb{R}^v$ and $W = \mathbb{R}^w$, it can be shown that $V \otimes W \cong \mathbb{R}^{vw}$, and a suitable map $\mathbb{T} : \mathbb{R}^v \times \mathbb{R}^w \to \mathbb{R}^v \otimes \mathbb{R}^w$ is defined by the Kronecker product as $\mathbb{T}(a, b) := a \otimes b$.

Let $\Psi_V = \{\psi_{V,1}, \psi_{V,2}, \ldots\}$ and $\Psi_W = \{\psi_{W,1}, \psi_{W,2}, \ldots\}$ be bases for the spaces V and W, respectively. Then one can find a basis for $V \otimes W$ as $\Psi_{V \otimes W} = \{\mathbb{T}(\psi_v, \psi_w) : \psi_v \in \Psi_V, \psi_w \in \Psi_W\}$. Once again, when $V = \mathbb{R}^v$ and $W = \mathbb{R}^w$, we will have $\Psi_{V \otimes W} = \Psi_V \otimes \Psi_W$.

3 Kronecker Product Matrices for Multidimensional Compressive Sensing

We now describe our framework for the use of Kronecker product matrices in multidimensional CS. We term the restriction of a multidimensional signal to fixed indices for all but its dth dimension a *d-section* of the signal. For example, for a 3-D signal $\mathbf{x} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$, the portion $\mathbf{x}_{i,j,\cdot} := [\mathbf{x}(i,j,1) \ \mathbf{x}(i,j,2) \ \ldots \ \mathbf{x}(i,j,N_3)]$ is a 3-section of $\mathbf{x}$. The definition can be extended to subsets of the dimensions; for example, $\mathbf{x}_{\cdot,\cdot,i} = [\mathbf{x}(1,1,i) \ \mathbf{x}(1,2,i) \ \ldots \ \mathbf{x}(N_1,N_2,i)]$ is a $\{1,2\}$-section of $\mathbf{x}$.

3.1 Kronecker product sparsifying bases

We can obtain a single sparsifying basis for an entire multidimensional signal as the Kronecker product sparsifying bases for each of its d-sections. This encodes all of the available structure using a single transformation. More formally, we let $\mathbf{x} \in \mathbb{R}^{N_1} \otimes \ldots \otimes \mathbb{R}^{N_d} = \mathbb{R}^{N_1 \times \ldots \times N_d} \cong \mathbb{R}^{\prod_{d=1}^{D} N_d}$ and assume that each d-section is sparse or compressible in a basis Ψ_d. We then pose a sparsifying basis for $\mathbf{x}$ obtained from Kronecker products as $\overline{\Psi} = \Psi_1 \otimes \ldots \otimes \Psi_D = \{\psi_1 \otimes \ldots \otimes \psi_D, \psi_d \in \Psi_d, 1 \leq d \leq D\}$, and obtain a coefficient vector Θ for the signal ensemble so that $\overline{\mathbf{x}} = \overline{\Psi}\Theta$, where $\overline{\mathbf{x}}$ is a vector-reshaped representation of $\mathbf{x}$.

3.2 Kronecker product measurement matrices

We can also design measurement matrices that are Kronecker products; such matrices correspond to measurement processes that operate individually on portions of the multidimensional signal. For simplicity, we assume in this section that each portion consists of a single d-section of the multidimensional signal, even though other configurations are possible (see Section 5 for examples). The resulting measurement matrix can be expressed as $\overline{\Phi} = \Phi_1 \otimes \ldots \otimes \Phi_D$. Consider the example of distributed sensing of signal ensembles from Section 1.4 where we obtain *separate measurements*, in the sense that each measurement depends on only one of the signals. More formally, for each signal (or 1-section) $\mathbf{x}_{\cdot,j}$, $1 \leq j \leq J$ we obtain separate measurements $\mathbf{y}_j = \Phi_j \mathbf{x}_{\cdot,j}$ with an individual measurement matrix being applied to each 1-section. The structure of such measurements can be succinctly captured by Kronecker products. To compactly represent the signal and measurement ensembles, we denote

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_J \end{bmatrix} \qquad \text{and} \qquad \overline{\Phi} = \begin{bmatrix} \Phi_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \Phi_2 & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \Phi_J \end{bmatrix}, \tag{7}$$

with $\mathbf{0}$ denoting a matrix of appropriate size with all entries equal to 0. We then have $\mathbf{Y} = \overline{\Phi}\overline{\mathbf{x}}$. Equation (7) shows that the measurement matrix that arises from distributed sensing has a characteristic block-diagonal structure when the entries of the sparse vector are grouped by signal. If a matrix $\Phi_j = \Phi'$ is used at each sensor to obtain its individual measurements, then we can express the joint measurement matrix as $\overline{\Phi} = \mathbf{I}_J \otimes \Phi'$, where $\mathbf{I}_J$ denotes the $J \times J$ identity matrix.

9

3.3 CS performance for Kronecker product matrices

We now derive results for metrics of Kronecker product sparsifying and measurement matrices required by the CS recovery guarantees provided in Theorems 2.1 and 2.3. The results obtained provide a link between the performance of the Kronecker product matrix and that of the individual matrices used in the product for CS recovery.

3.3.1 Mutual coherence

Consider a Kronecker sparsifying basis $\overline{\Psi} = \Psi_1 \otimes \ldots \otimes \Psi_D$ and a global measurement basis obtained through a Kronecker product of individual measurement bases: $\overline{\Phi} = \Phi_1 \otimes \ldots \otimes \Phi_D$, with each pair Φ_d and Ψ_d being mutually incoherent for $d = 1, \ldots, D$. The following lemma provides a *conservation of mutual coherence* across Kronecker products (see also [29, 30]).

Lemma 3.1 *Let Φ_d, Ψ_d be bases or frames for $\mathbb{R}^{N_d}$ for $d = 1, \ldots, D$. Then*

$$\mu(\Phi_1 \otimes \ldots \otimes \Phi_D, \Psi_1 \otimes \ldots \otimes \Psi_D) = \prod_{d=1}^{D} \mu(\Phi_d, \Psi_d).$$

Proof. We rewrite the coherence as

$$\mu(\Phi, \Psi) = \|\Phi^T \Psi\|_{\max},$$

where $\| \cdot \|_{\max}$ denotes the matrix max norm, i.e., the largest entry of the matrix. Since

$$(\Phi_1 \otimes \ldots \otimes \Phi_D)^T (\Psi_1 \otimes \ldots \otimes \Psi_D) = \Phi_1^T \Psi_1 \otimes \ldots \otimes \Phi_D^T \Psi_D,$$

and since $\|\Phi \otimes \Psi\|_{\max} = \|\Phi\|_{\max} \|\Psi\|_{\max}$, the theorem follows. $\square$

Since the mutual coherence of each d-section's sparsifying basis and measurement matrix is upper bounded by one, the number of Kronecker product-based measurements necessary for successful recovery of the multidimensional signal is always lower than or equal to the corresponding number of necessary partitioned measurements. This reduction is maximized when the measurement matrix Φ_e for the dimension e along which measurements are to be partitioned is maximally incoherent with the e-section sparsifying basis Ψ_e.

3.3.2 Restricted isometry constants

The restricted isometry constants for a matrix Φ are intrinsically tied to the singular values of all submatrices of Φ of a certain size. The structure of Kronecker product matrices enables simple bounds for their restricted isometry constants.

Lemma 3.2 *Let $\Phi_1, \ldots, \Phi_D$ be matrices with restricted isometry constants $\delta_K(\Phi_1), \ldots, \delta_K(\Phi_D)$, respectively. Then,*

$$\delta_K(\Phi_1 \otimes \Phi_2 \otimes \ldots \otimes \Phi_D) \leq \prod_{d=1}^{D}(1 + \delta_K(\Phi_d)) - 1.$$

Proof. We begin with the case $D = 2$ and denote by $\overline{\Phi}_\Omega$ the K-column submatrix of $\overline{\Phi}$ containing the columns $\overline{\phi}_t$, $t \in \Omega$; its nonzero singular values obey

$$1 - \delta_K(\overline{\Phi}) \leq \sigma_{\min}(\overline{\Phi}_\Omega) \leq \sigma_{\max}(\overline{\Phi}_\Omega) \leq 1 + \delta_K(\overline{\Phi}).$$

Since each $\overline{\phi}_t = \phi_{1,u} \otimes \phi_{2,v}$ for specific u, v, we can build sets Ω_1, Ω_2 of cardinality up to K that contain the values of u, v, respectively, corresponding to the indices $t \in \Omega$. Then, it is easy to see that $\overline{\Phi}_\Omega$ is a submatrix of $\Phi_{1,\Omega_1} \otimes \Phi_{2,\Omega_2}$, which has up to K^2 columns. Furthermore, it is well known that $\sigma_{\min}(\Phi_1 \otimes \Phi_2) = \sigma_{\min}(\Phi_1)\sigma_{\min}(\Phi_2)$ and $\sigma_{\max}(\Phi_1 \otimes \Phi_2) = \sigma_{\max}(\Phi_1)\sigma_{\max}(\Phi_2)$. Additionally, the range of singular values of a submatrix are interlaced inside those of the original matrix [31]. Thus,

$$\sigma_{\min}(\Phi_{1,\Omega_1} \otimes \Phi_{2,\Omega_2}) \leq \sigma_{\min}(\overline{\Phi}_\Omega) \leq \sigma_{\max}(\overline{\Phi}_\Omega) \leq \sigma_{\max}(\Phi_{1,\Omega_1} \otimes \Phi_{2,\Omega_2}),$$
$$\sigma_{\min}(\Phi_{1,\Omega_1})\sigma_{\min}(\Phi_{2,\Omega_2}) \leq \sigma_{\min}(\overline{\Phi}_\Omega) \leq \sigma_{\max}(\overline{\Phi}_\Omega) \leq \sigma_{\max}(\Phi_{1,\Omega_1})\sigma_{\max}(\Phi_{2,\Omega_2}).$$

By using the K-restricted isometry constants for Φ_1 and Φ_2, we obtain the following bounds:

$$(1 - \delta_K(\Phi_1))(1 - \delta_K(\Phi_2)) \leq \sigma_{\min}(\overline{\Phi}_\Omega) \leq \sigma_{\max}(\overline{\Phi}_\Omega) \leq (1 + \delta_K(\Phi_1))(1 + \delta_K(\Phi_2)).$$

For $D > 2$ an inductive argument provides

$$\prod_{d=1}^{D}(1 - \delta_K(\Phi_d)) \leq \sigma_{\min}(\overline{\Phi}_\Omega) \leq \sigma_{\max}(\overline{\Phi}_\Omega) \leq \prod_{d=1}^{D}(1 + \delta_K(\Phi_d)),$$

and so we must have

$$\delta_K(\Phi_1 \otimes \Phi_2 \otimes \ldots \otimes \Phi_D) = \max\left(1 - \prod_{d=1}^{D}(1 - \delta_K(\Phi_d)), \prod_{d=1}^{D}(1 + \delta_K(\Phi_d)) - 1\right).$$

It is simple to show that the second term is always larger than the first, proving the lemma. □

When Φ_1 is an orthonormal basis, it has restricted isometry constant $\delta_K(\Phi_1) = 0$ for all $K \leq N$. Therefore the restricted isometry constant of the Kronecker product of an orthonormal basis and a measurement matrix is equal to that of the measurement matrix. While the bound is in general loose due to the use of a matrix with K^2 columns in the proof, we note that the RIP constant of the Kronecker product matrix is bounded below, by construction, by the largest RIP constant among the individual matrices; that is, $\delta_K(\Phi_1 \otimes \Phi_2 \otimes \ldots \otimes \Phi_D) \geq \max_{1 \leq d \leq D} \delta_K(\Phi_d)$ [32]. Therefore, the resulting pair of bounds is tight in the case where there is a dominant (larger) RIP constant among the matrices $\{\Phi_d\}_{d=1}^{D}$ involved in the product.

3.4 Computational Aspects

We briefly consider the computational complexity of KCS. There exist several solvers of the optimization programs (3-4), such as interior point methods, that have computational complexity $\mathcal{O}(N^3)$, where N denotes the length of the vector θ [33]. Thus, independent recovery of each e-section of a multidimensional dataset yields total complexity $\mathcal{O}\left(N_e^3 \prod_{d\neq e} N_d\right)$. In contrast, the KCS approach relies on solving a single higher-dimensional optimization problem of complexity $\mathcal{O}\left(\prod_{d=1}^{D} N_d^3\right)$, providing a computational overhead of $\mathcal{O}\left(\prod_{d\neq e} N_d^2\right)$ for the improved performance afforded by the Kronecker product sparsity/compressibility basis (as detailed in the next section).

When the measurement matrix (and its transpose) can be applied efficiently to a vector (with complexity $A(N) < \mathcal{O}\left(N^2\right)$), the computational complexity of the optimization solver drops to $O(NA(N))$ and the computational overhead of KCS is reduced to $\mathcal{O}\left(A(N)/A(N_e)\right)$. There exist CS matrices with efficient implementations featuring $A(N) = \mathcal{O}\left(N \log N\right)$, which yield a computational cost for KCS of approximately $\mathcal{O}\left(\prod_{d \neq e} N_d \frac{\sum_{d=1}^{D} \log(N_d)}{\log(N_e)}\right)$. A rough interpretation of this result (for data of similar size among all dimensions) is that the computational cost of KCS is proportional to the dimensionality of the data times the number of data partitions in the eth dimension, i.e., DN/N_e.

4   Case Study: CS with Multidimensional Wavelet Bases

Kronecker products are prevalent in the extension of wavelet transforms to multidimensional settings. There are several different multidimensional wavelet basis constructions depending on the choice of basis vectors involved in the Kronecker products. For these constructions, our interest is in the relationship between the compressibility of the multidimensional signal in the Kronecker product wavelet basis vs. the compressibility of a partitioned version of the same multidimensional signal in "partial" wavelet bases that cover fewer data dimensions. In this section we assume that the N-length, D-D signal $\mathbf{x}$ is a sampled representation of a continuous-indexed D-D signal $f(t_1, ... t_D)$, with $t_d \in \Omega := [0, 1], 1 \le d \le D$, such that $\mathbf{x}(n_1, \ldots, n_D) = f(n_1/N_1, \ldots, n_d/N_D)$, with $N = N_1 \times \ldots \times N_D$.

4.1   Isotropic, anisotropic, and hyperbolic wavelets

Consider a 1-D signal $f(t) : \Omega \to \mathbb{R}$ with $\Omega = [0, 1]$; its wavelet representation is given by

$$f = v_0 \nu + \sum_{i \geq 0} \sum_{j=0}^{2^i - 1} w_{i,j} \psi_{i,j},$$

where ν is the scaling function and $\psi_{i,j}$ is the wavelet function at scale i and offset j. The wavelet transform consists of the scaling coefficient v_0 and wavelet coefficients $w_{i,j}$ at scale i, $i \geq 0$, and position j, $0 \leq j < 2^i$; the support of the corresponding wavelet $\psi_{i,j}$ is roughly $[2^{-i}j, 2^{-i}(j+1)]$. In terms of our earlier matrix notation, the sampled signal $\mathbf{x}$ has the representation $\mathbf{x} = \Psi\theta$, where Ψ is a matrix containing the sampled scaling and wavelet functions for scales $1, \ldots, L = \log_2 N$ as columns, and $\theta = [v_0,\ w_{0,0},\ w_{1,0},\ w_{1,1},\ w_{2,0}, \ldots]^T$ is the vector of corresponding scaling and wavelet coefficients. We are, of course, interested in sparse and compressible θ.

Several different extensions exist for the construction of D-D wavelets as a Kronecker product of 1-D wavelet functions [34–36]. In each case, a D-D wavelet is obtained from the Kronecker product of D 1-D wavelets: $\psi_{i_1,j_1,\ldots,i_D,j_D} = \psi_{i_1,j_1} \otimes \ldots \otimes \psi_{i_D,j_D}$. Different bases for the multidimensional space can then be obtained through the use of appropriate combinations of 1-D wavelets in the Kronecker product. For example, *isotropic wavelets* arise when the same scale $i = i_1 = \ldots = i_D$ is selected for all wavelet functions involved, while *anisotropic wavelets* force a fixed factor between any two scales, i.e. $a_{d,d'} = i_d/i_{d'}$, $1 \leq d, d' \leq D$. Additionally, *hyperbolic wavelets* result when no restriction is placed on the scales $i_1, \ldots, i_D$. Therefore, a hyperbolic wavelet basis for $\mathbb{R}^{N_1} \otimes \ldots \otimes \mathbb{R}^{N_D}$ is obtained as the Kronecker product of the individual wavelet bases for $\mathbb{R}^{N_d}$, $1 \leq d \leq D$. In the sequel, we identify the isotropic, anisotropic, and hyperbolic wavelet bases as Ψ_I, Ψ_A, and Ψ_H,
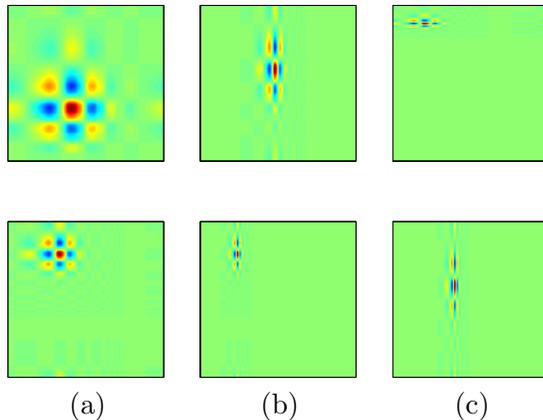
Figure 2: *Example basis elements from 2-D wavelet bases. In each case, green (light) pixels represent zeros, while blue and red (dark) pixels represent large positive and negative values, respectively. (a) Isotropic wavelets have the same degree of smoothness on all dimensions, and are obtained from the Kronecker product of two 1-D wavelets of the same scale; (b) Anisotropic wavelets have different degrees of smoothness in each dimension, but with a constant ratio, and are obtained from the Kronecker product of two 1-D wavelets at ratio-matching scales; (c) Hyperbolic wavelets have different degrees of smoothness in each dimension without restrictions and are obtained from the Kronecker product of two 1-D wavelets of all scales.*

respectively; example basis elements for each type of multidimensional wavelet basis are shown in Fig. 2.

4.2   Isotropic Besov spaces

Isotropic wavelets have been popularized by their suitability for analysis of 2-D signals (images). Significant study has been devoted to identify the types of signals that are sparse or compressible in an isotropic wavelet basis. A fundamental result in this direction states that the discretizations of signals in isotropic Besov spaces are compressible in an isotropic wavelet basis using a sufficiently smooth wavelet function. Such signals have the same degree of smoothness in all dimensions. We begin by providing a brief formal definition of Besov spaces; see [35–37] for details.

We define the *directional derivative* of f in the direction h as $(\Delta_h f)(t) := f(t + h) - f(x)$, with higher-degree derivatives defined as $(\Delta_h^m f)(t) := (\Delta_h(\Delta_h^{m-1} f))(t)$, $m \geq 2$. Here and later we define $(\Delta_h f)(t) = 0$ if $t + h \notin \Omega^D$. For $r \in \mathbb{R}^+$, $m \in \mathbb{N}$ and $0 < p < \infty$, we define the *modulus of smoothness* as

$$\omega_m(f, r, \Omega^D)_p = \sup_{|h| \leq r} \|\Delta_h^m f\|_{p, \Omega^D}.$$

It is easy to see that $\omega_m(f, r, \Omega^D)_p \to 0$ as $r \to 0$; smoother functions have faster decay in this asymptotic behavior.

A signal can be classified according to its smoothness simply by imposing conditions on the rate of decay of its moduli of smoothness. The resulting classes are known as *Besov spaces.* A Besov space $B_{p,q}^s$ contains D-D functions that have (roughly speaking) s derivatives in $L_p(\Omega^D)$; this smoothness is measured by the rate of decay of the modulus of smoothness as a function of

the step size r. The Besov quasi-seminorm is then defined as

$$|f|_{B_{p,q}^s} = \left(\int_0^1 \left[r^{-s} \omega_m(f, r, \Omega^D)_p \right]^q \frac{dr}{r} \right)^{1/q}.$$

Here the parameter q provides finer distinctions of smoothness. Thus, we say that a signal $f \in B_{p,q}^s$ if it has finite Besov norm, defined as $\|f\|_{B_{p,q}^s} = \|f\|_p + |f|_{B_{p,q}^s}$.

Similarly to the discrete signal case, we define the best K-term approximation error in the basis Ψ as

$$\sigma_K(f, \Psi)_p = \min \left\{ \|f - g\|_p, g = \sum_{k=1}^K c_j \psi_{i_k}, \ \psi_{i_k} \in \Psi \text{ for each } i = 1, \ldots, K \right\}.$$

Such isotropic wavelet-based nonlinear approximations provide provable decay rates for the approximation error.

Theorem 4.1 *[37] If the scaling function $\nu \in B_{p,q}^s$, ν has at least s vanishing moments, and $f \in B_{p,q}^r$, with $r \geq D/p - D/2$ and $0 < r < s$, then $\sigma_K(f, \Psi_I)_p < CK^{-r}$.*

In words, Theorem 4.1 states that Besov-smooth signals are compressible in a sufficiently smooth isotropic wavelet transform.

4.3 Anisotropic Besov spaces

In many applications outside of natural image processing, the type of structure present is different in each of the signal's dimensions [12, 14, 38, 39]. For example, a video sequence has different degrees of smoothness in its spatial and temporal dimensions, while a hyperspectral datacube can have different degrees of smoothness in the spatial and spectral dimensions. In these cases, anisotropic and hyperbolic wavelets can be used to achieve sparse and compressible representations for signals of this type. Similarly to isotropic Besov spaces, signals in anisotropic Besov spaces have discretizations that are compressible in an anisotropic wavelet basis. We first provide a formal definition of anisotropic Besov spaces, which closely mirrors that of isotropic Besov spaces, except that the smoothness in each dimension is specified separately.

We define the d-directional derivative of f as $(\Delta_{h,d} f)(t) := f(t + he_d) - f(t)$, $1 \leq d \leq D$, where e_d is the dth canonical vector, i.e., its dth entry is one and all others are zero. This corresponds to the standard directional derivative in which the direction h is a multiple of the canonical vector e_d. We also define higher-degree directional derivatives as $(\Delta_{h,d}^m f)(t) := (\Delta_{h,d}(\Delta_{h,d}^{m-1} f))(t)$, $m \geq 2$. For $r \in \mathbb{R}^+$, $m_d \in \mathbb{N}$ and $0 < p < \infty$, we define the d-directional moduli of smoothness as

$$\omega_{m_d,d}(f, r, \Omega^D)_p = \sup_{|h| \leq r} \|\Delta_{h,d}^{m_d} f\|_{p, \Omega^D}.$$

By defining the anisotropy parameter $\bar{s} = (s_1, \ldots, s_D)$, we define the anisotropic Besov quasi-seminorm as [35, 36]

$$|f|_{B_{p,q}^{\bar{s}}} = \left(\int_0^1 \left[\sum_{d=1}^D r^{-s_d} \omega_{m_d,d}(f, r, \Omega^D)_p \right]^q \frac{dr}{r} \right)^{1/q}.$$

Thus, we say that a signal $f \in B_{p,q}^{\bar{s}}$ if it has finite anisotropic Besov norm, defined as $\|f\|_{B_{p,q}^{\bar{s}}} = \|f\|_p + |f|_{B_{p,q}^{\bar{s}}}$.

An anisotropic Besov space $B_{p,q}^{\bar{s}}$ contains functions of D continuous variables that have (roughly speaking) s_d derivatives in $L_p(\Omega)$ for any d-section of the D-D function; once again, the parameter q provides finer distinctions of smoothness. An example is a multidimensional signal that is expressed as the Kronecker product of individual signals that are compressible in wavelet bases.

We now study the conditions for compressibility of a signal in an anisotropic wavelet basis as a function of the smoothness of the signal in its different dimensions. We will observe that the rate of decay for the wavelet coefficients will depend on the characteristics of the anisotropic Besov space in which the signal lives. Some conditions must be imposed on the wavelets used for compressibility. We denote by $\nu = \{\nu_{\bar{i},\bar{j}}\}_{\bar{i},\bar{j}}$ the family of scaling functions for each scale $\bar{j}$ and offset $\bar{i}$.

Definition 4.1 *A scaling function family ν is $B_{p,q}^{\bar{s}}$-smooth, $\bar{s} > 0$ (i.e. $s_d > 0$, $1 \leq d \leq D$), if for some $(m_1, \ldots, m_D) > \bar{s}$, for each $i_1, \ldots, i_D \in \mathbb{N}_0^D$ and for a finite constant $C > 0$ there exist $\bar{j}_d \in \mathbb{N}_0$, $0 \leq \bar{j}_d < 2^{i_d}$, $1 \leq d \leq D$ such that for each $0 \leq j_d < 2^{i_d}$, $d = 1, \ldots, D$, and $k \in \mathbb{N}_0$,*

$$\omega_{m_d,d}(\nu_{i_1,j_1,\ldots,i_D,j_D}, 2^{-k}, \Omega^D)_p < C\omega_{m_d,d}(\nu_{i_1,\bar{j}_1,\ldots,i_D,\bar{j}_D}, 2^{-k}, \Omega^D)_p,$$

and

$$|\nu_{i_1,\bar{j}_1,\ldots,i_D,\bar{j}_D}|_{B_{p,q}^{\bar{s}}} < C2^{(i_1+\ldots+i_D)(1/2-1/p)} \sum_{d=1}^{D} 2^{i_d s_d}.$$

It can be shown that the scaling functions formed from tensor or Kronecker products of regular 1-D scaling functions $\nu_{i_1,j_1,\ldots,i_D,j_D} = \nu_{i_1,j_1} \otimes \ldots \otimes \nu_{i_D,j_D}$ has this smoothness property when the component scaling functions are smooth enough [35, 36]. This condition suffices to obtain results on approximation rates for the different types of Kronecker product wavelet bases. The following theorem is an extension of a result from [36] to the D-D setting, and is proven in [40, Appendix K].

Theorem 4.2 *Assume the scaling function ν that generates the anisotropic wavelet basis Ψ_A with anisotropy parameter $\bar{s} = (s_1, \ldots, s_D)$ is $B_{p,q}^{\bar{s}}$-smooth and $f \in B_{p,q}^{\bar{r}}$, with $\bar{r} = (r_1, \ldots, r_D)$ and $0 < \bar{r} < \bar{s}$. Define $\rho = \min_{1 \leq d \leq D} r_d$ and*

$$\lambda = \frac{D}{\sum_{d=1}^{D} 1/r_d}. \tag{8}$$

If $\rho > D/p + D/2$ then the approximation rate for the function f in an isotropic wavelet basis is $\sigma_K(f, \Psi_I)_p < CK^{-\rho}$. Similarly, if $\lambda > D/p + D/2$, then the approximation rate for the function f in both an anisotropic and a hyperbolic wavelet basis is $\sigma_K(f, \Psi_A)_p < C_A K^{-\lambda}$ and $\sigma_K(f, \Psi_H)_p < C_H K^{-\lambda}$.

To give some perspective to this theorem, we study two example cases: isotropy and extreme anisotropy. In the isotropic case, all the individual rates $r_d = r, 1 \leq d \leq D$, and the approximation rate under anisotropic and hyperbolic wavelets matches that of isotropic wavelets: $\lambda = \rho = r$. In the extreme anisotropic case, we have that one of the approximation rates is much smaller than all others: $r_e \ll r_d$ for all $e \neq d$. In contrast, in this case we obtain a rate of approximation under anisotropic and hyperbolic wavelets of $\lambda \approx Dr_e$, which is D times larger than the rate for isotropic wavelets, $\rho = r_e$. Thus, the approximation rate with anisotropic and hyperbolic wavelets is in the range $\lambda \in [1, D] \min_{1 \leq d \leq D} r_d$. We also note the dependence of the result on the dimensionality of the signal: as D increases, the requirements on the smoothnesses ρ, λ of the function f become more strict.

The disadvantage of anisotropic wavelets, as compared with hyperbolic wavelets, is that they must have an anisotropy parameter that matches that of the anisotropic smoothness of the signal in order to achieve the optimal approximation rate [35]. Additionally, the hyperbolic wavelet basis is the only one out of the three basis types described that can be expressed as the Kronecker product of lower dimensional wavelet bases. Therefore, we use hyperbolic wavelets in the sequel and in the experiments of Section 5.

4.4   Performance of KCS with multidimensional hyperbolic wavelet bases

Since KCS uses Kronecker product matrices for measurement and compression of multidimensional signals, it is possible to compare the rates of approximation that can be obtained by using independent measurements of each d-section of the multidimensional signal against those obtained by KCS. The following Theorem is obtained by amalgamating the results of Theorems 2.1, 2.3, and 4.2 and Lemma 3.1.

Theorem 4.3 *Assume that a D-D signal $\mathbf{x} \in \mathbb{R}^{N_1 \times \ldots \times N_D}$ is the sampled version of a continuous-time signal $f \in B_{1,q}^{\overline{s}}$, with $\overline{s} = s_1, \ldots, s_D$, under the conditions of Theorem 4.2 with $p = 1$. In particular, $\mathbf{x}$ has s_d-compressible d-sections in sufficiently smooth wavelet bases Ψ_d, $1 \leq d \leq D$. Denote by Φ_d, $1 \leq d \leq D$ a set of CS measurement bases that can be applied along each dimension of $\mathbf{x}$. If M total measurements are obtained using a random subset of the columns of $\Phi_1 \otimes \ldots \otimes \Phi_D$, then with high probability the recovery error from these measurements has the property*

$$\|\mathbf{x} - \widehat{\mathbf{x}}\|_2 \leq C \left(\frac{M}{\sqrt{N} \prod_{d=1}^{D} \mu(\Phi_d, \Psi_d)} \right)^{-\beta}, \tag{9}$$

where $\beta = \frac{D}{2 \sum_{d=1}^{D} 1/s_d} + \frac{1}{4}$, while the recovery from M measurements equally distributed among the eth dimension of the signal using the basis $\Phi_1 \otimes \ldots \otimes \Phi_{e-1} \otimes \Phi_{e+1} \otimes \ldots \otimes \Phi_D$ on each $\{1, \ldots, e-1, e+1, \ldots, D\}$-section of $\mathbf{x}$ has the property

$$\|\mathbf{x} - \widehat{\mathbf{x}}\|_2 \leq C N_e^{1/2} \left(\frac{M}{\sqrt{N/N_e} \prod_{d \neq e} \mu(\Phi_d, \Psi_d)} \right)^{-\beta_e}, \tag{10}$$

where $\beta_e = \frac{D-1}{2 \sum_{d \neq e} 1/s_d} + \frac{1}{4}$.

Proof sketch. For recovery we pair the Kronecker product measurement matrix $\Phi_P := \Phi_1 \otimes \ldots \otimes \Phi_D$ with the hyperbolic wavelet basis $\Psi_H = \Psi_1 \otimes \ldots \otimes \Psi_D$. From Lemma 3.1, we have that the mutual coherence of these two bases is $\mu(\Phi_P, \Psi_H) = \prod_{d=1}^{D} \mu(\Phi_d, \Psi_d)$. Plugging this value into Theorem 2.3, the number of measurements needed to achieve the RIP having $\delta_{2K} = \sqrt{2} - 1$ with high probability is

$$M = C_M K \sqrt{N} \prod_{d=1}^{D} \mu(\Phi_d, \Psi_d) \log(tK \log N) \log^2(K). \tag{11}$$

The RIP, in turn, guarantees that the recovery error obeys $\|\mathbf{x} - \widehat{\mathbf{x}}\|_2 = \|\theta - \widehat{\theta}\|_2 \leq CK^{-1/2}\|\theta - \theta_K\|_1$, as given in Theorem 2.1, with θ and $\widehat{\theta}$ denoting the coefficients of $\mathbf{x}$ and $\widehat{\mathbf{x}}$, respectively; note that

the output of the recovery algorithms (3-4) is given by $\widehat{\theta}$, with $\widehat{\mathbf{x}} = \Psi_H \widehat{\theta}$. The approximation error is then bounded using Theorem 4.2 by $\|\theta - \theta_K\|_1 = \sigma_K(f, \Psi_H)_1 \leq C_H K^{-\lambda}$, where we use the fact that θ is the discrete wavelet transform coefficient vector for the vector $\mathbf{x}$, which itself contains N uniform samples of the continuous signal $f \in B_{1,q}^{\bar{s}}$. We therefore obtain

$$\|\mathbf{x} - \widehat{\mathbf{x}}\|_2 \leq CC_H K^{-\lambda-1/2}. \tag{12}$$

At this point we solve for K in (11) to obtain

$$C_K K^2 \geq K \log(tK \log N) \log^2(K) = \frac{M}{C_M \sqrt{N}} \prod_{d=1}^{D} \mu(\Phi_d, \Psi_d)^{-1},$$

$$K \geq C_K' \left(\frac{M}{\sqrt{N}} \right)^{1/2} \prod_{d=1}^{D} \mu(\Phi_d, \Psi_d)^{-1/2}.$$

Plugging into (12), we obtain

$$\|\mathbf{x} - \widehat{\mathbf{x}}\|_2 \leq CC_H C_K'^{\lambda+1/2} \left(\frac{M}{\sqrt{N}} \right)^{-\lambda/2-1/4} \prod_{d=1}^{D} \mu(\Phi_d, \Psi_d)^{\lambda/2+1/4}.$$

By noticing that $\beta = \lambda/2 + 1/4$, we have obtained (9).

The proof of (10) proceeds similarly by partitioning along the eth dimension and adjusting the number of measurements per $\{1, \ldots, e-1, e+1, \ldots, D\}$-section to M/N_e, so that the term $M/\sqrt{N}$ in (9) is replaced by $(M/N_e)/\sqrt{N/N_e} = M/\sqrt{NN_e}$, introducing a new multiplicative term $N_e^{\beta_e/2}$. A triangle inequality to assemble the error for the entire D-D signal from the different e-sections introduces a new additional factor of $\sqrt{N_e}$. $\qquad\square$

To put Theorem 4.3 in perspective, we study the bases and the exponents of the bounds separately. With regards to the bases, the denominators in (9)–(10) provide a scaling for the number of measurements needed to achieve a target recovery accuracy. This scaling is dependent on the measurement matrices via mutual coherences; the denominators take values in the ranges $[1, \sqrt{N}]$ and $[1, \sqrt{N/N_e}]$, respectively. With regards to the exponents, the rates of decay for the recovery error match those of the signal's compressibility approximation error rates λ from (8) for the entire signal and its partitions, respectively. The error decay rate for KCS recovery is higher than that for independent recovery from partitioned measurements when $s_e > \frac{D-1}{\sum_{d \neq e} 1/s_d}$, i.e., when the compressibility exponent of the e-sections is larger than the harmonic mean of the compressibility exponents of all other sections. Thus, KCS provides the most significant improvement in the error rate of decay when the measurement partitioning is applied along the dimension(s) that feature highest compressibility or smoothness. Note also the $\sqrt{N_e}$ cost in (10) of partitioning measurements, which comes from the triangle inequality.

5   Experimental Results

In this section, we perform experiments to verify the compressibility properties of two different classes of signals in a Kronecker product wavelet basis. We also perform KCS sparse recovery experiments that illustrate the advantages of KCS over standard CS schemes. For the multidimensional

signals, in addition to synthetic data, we choose 3D hyperspectral imagery and video sequences, since they can be compressed effectively by well-studied, albeit data-dependent, compression algorithms (the Karhunen-Loève transform (KLT) and motion compensation, respectively). Our intent is to see how close we can come to this nonlinear compression performance using the simpler linear Kronecker product wavelet basis for compression and CS recovery. We will also experimentally verify the tradeoffs provided in Sections 3 and 4 and contrast the recovery performance to that reached by integrating such task-specific compression schemes to distributed CS recovery.

The experiments use the basis pursuit solvers from [41] and [42] for the hyperspectral and video data, respectively. The experiments were executed on a Linux workstation with an Intel Xeon CPU at 3.166 GHz and 4 GB of memory. A Matlab toolbox containing the scripts that generate the results and figures provided in this section is available for download at `http://dsp.rice.edu/kcs`. Additional experimental results are also available in [40].

5.1   Empirical Performance of KCS

Our first experiment considers synthetically generated signals of size $N = 8 \times 8 \times 8$ that are $K = 10$-sparse in a Kronecker product (hyperbolic) wavelet basis and compares three CS recovery schemes: the first uses a single recovery from dense, *global* measurements; the second uses a single *KCS recovery* from the set of measurements obtained independently from each 8×8 1-section; and the third one uses *independent recovery* of each 8×8 1-section from its individual measurements. We let the number of measurements M vary from 0 to N with the measurements evenly split among the 1-sections in the independent and KCS recovery cases. For each value of M, we average 100 iterations by generating K-sparse signals $\mathbf{x}$ with independent and identically distributed (i.i.d.) Gaussian entries and with support following a uniform distribution among all supports of size K, and generating measurement matrices with i.i.d. Gaussian entries. We then measure the probability of successful recovery for each value of M, where a success is declared if the signal estimate $\widehat{\mathbf{x}}$ obeys $\|\mathbf{x} - \widehat{\mathbf{x}}\|_2 \leq 10^{-3}\|\mathbf{x}\|_2$. The results are shown in Fig. 3, which shows that KCS outperforms separate section-by-section recovery while achieving lower success probabilities than recovery from global measurements. In fact, the measurement-to-sparsity ratio M/K required for 95% success rate are 6, 15, and 30 for global measurements, KCS, and independent recovery, respectively.

5.2   Hyperspectral data

5.2.1   Compressibility

We first evaluate the compressibility of a real-world hyperspectral datacube using independent spatial and spectral sparsifying bases and compare it with a Kronecker product basis. The datacube for this experiment is obtained from the AVIRIS Moffett Field database [43]. A $N = 128 \times 128 \times 128$ voxel portion is used. We then process the signal through six different transforms. The first three (*Space, Frequency Wavelet, Frequency KLT*) perform transforms along a subset of the dimensions of the data (a 1-D wavelet basis $\mathbf{W}_1$ for the spectral dimension, a 2-D wavelet basis $\mathbf{W}_2$ for the spatial dimensions, and a 1-D KLT basis[2] $\mathbf{P}_1$ for the spectral dimension, respectively). The fourth (*Isotropic Wavelet*) transforms the entire datacube with a 3-D isotropic wavelet basis. The fifth and

---

[2]A KLT basis is learned from a datacube of the same size extracted from a different spatial region of the original AVIRIS dataset [20, 44, 45]. The resulting transformation provides a linear approximation scheme that preserves the coefficients for the most significant principal components, rather than the nonlinear approximation scheme used in sparse approximation.
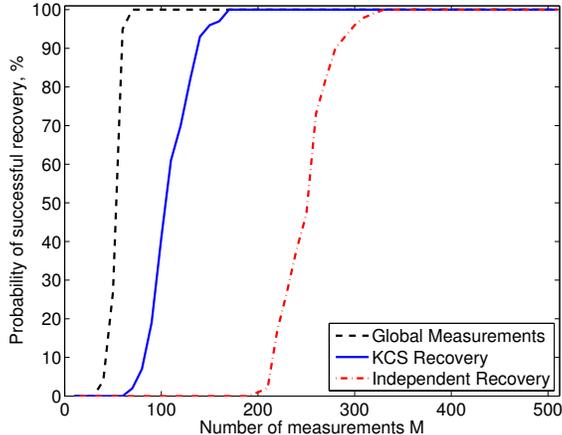
Figure 3: *Empirical performance of KCS. We generate 3-D signals of size $N = 8 \times 8 \times 8$ that are $K = 10$-sparse in a hyperbolic wavelet basis. We compare the performance of three approaches:* KCS recovery *using the measurement matrix $\mathbf{I}_8 \otimes \Phi$, with Φ denoting an $M/8 \times N$ random matrix with i.i.d. Gaussian entries; standard CS recovery using* global measurements *with a dense random $M \times N$ matrix; and* independent recovery *of each 1-section using standard CS with an $M/8 \times N$ random matrix. While KCS does not match the performance of global measurements, it performs significantly better than independent recovery using the same measurements.*

sixth (*Hyperbolic Wavelet* and *Wavelet/KLT*) transform the entire datacube with a basis formed from the Kronecker products $\mathbf{W}_1 \otimes \mathbf{W}_2$ of a 1-D wavelet basis in frequency and a 2-D isotropic wavelet basis in space, and $\mathbf{P}_1 \otimes \mathbf{W}_2$ of a a 1-D KLT basis in frequency and a 2-D isotropic wavelet basis in space, respectively. In all cases the Daubechies-8 wavelet was used. For each one of these transforms, we measured the signal-to-noise ratio (SNR) when transform coding is used to preserve K coefficients of the data for varying values of K. The results are shown in Figs. 4 and 5; the Kronecker (anisotropic wavelet) transform provides the best compression of the signal, as shown in Fig. 4, outperforming the partial transforms in terms of SNR. However, Fig. 5(b) shows that the rate of decay for the normalized error of the Kronecker (anisotropic wavelet) transform is only slightly higher than the minimum rate of decay among the partial (spatial andfrequency) wavelet transforms. Our analysis indicates that this result is due to the difference between the degrees of smoothness among the signal dimensions.

5.2.2 KCS

We also compare the performance of KCS to CS using the 2-D basis $\mathbf{W}_2$ to yield compressible/approximately sparse coefficient vectors for individual spectral band images. In our experiments we obtain CS measurements using the subsampled permuted Hadamard transform of [7] on each spectral band image with a matrix Φ_2. We also obtain *global* CS measurements that depend on all the voxels of the datacube as a baseline; such measurements result in a fully dense measurement matrix Φ and therefore are difficult to obtain in real-world applications. We operate with two datacubes: the original $128 \times 128 \times 128$ voxel version from previous experiments, and "flattened" datacubes of sizes $128 \times 128 \times 16$, $128 \times 128 \times 32$, and $128 \times 128 \times 64$ voxels. The flattening was performed by aggregating the intensities among the bands in each spectral neighborhood for each of the pixels in the image.
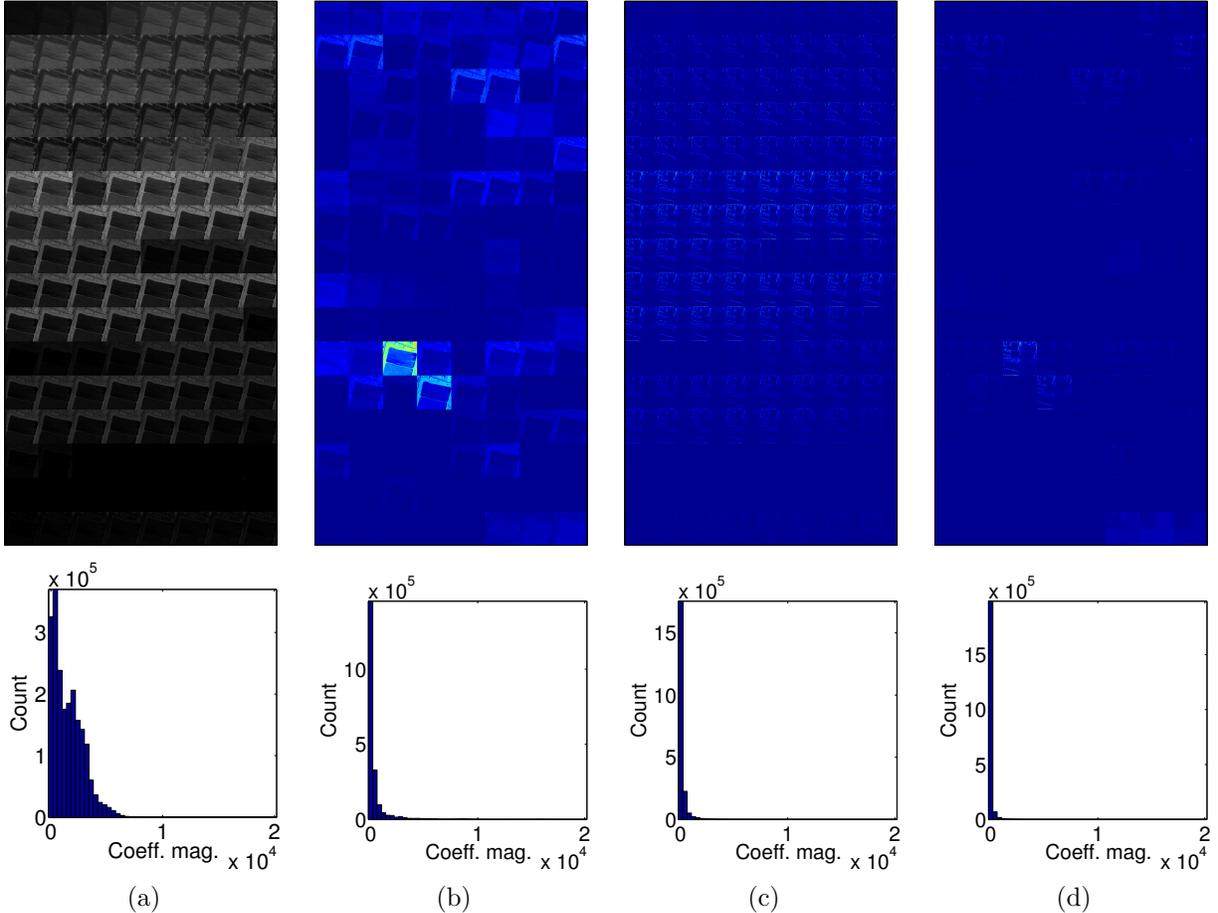
Figure 4: *Examples of transform coding of a hyperspectral datacube of size* $128 \times 128 \times 128$. *(a) Original data; (b) Coefficients in a 1-D wavelet basis* $\mathbf{W}_1$ *applied at each pixel in the spectral domain; (c) Coefficients in a 2-D isotropic wavelet basis* $\mathbf{W}_2$ *applied at each pixel in the spatial domain; (d) Coefficients in a Kronecker product basis* $\mathbf{W}_1 \otimes \mathbf{W}_2$. *The top row shows the datacube or coefficients flattened to 2-D by concatenating each spectral band's image, left to right, top to bottom. In (b-d), blue (dark) pixels represent coefficients with small magnitudes. The bottom row shows histograms for the coefficient magnitudes, showing the highest concentrations of small coefficients for the Kronecker product basis.*

Figure 6 shows the recovery error for each datacube from several different recovery setups: *Independent* recovery operates on each spectral band independently with the measurement matrix Φ_2 using the basis $\mathbf{W}_2$ to sparsify each spectral band. KCS employs the Kronecker product measurement matrix $\mathbf{I} \otimes \Phi_2$ to perform joint recovery. We test two different Kronecker product sparsifying bases: *KCS-Wavelet* uses a Kronecker products of wavelet bases $\mathbf{W}_1 \otimes \mathbf{W}_2$, and *KCS-KLT* uses a Kronecker product $\mathbf{P}_1 \otimes \mathbf{W}_2$ of a KLT basis $\mathbf{P}_1$ in the spectral dimension and a 2-D wavelet basis $\mathbf{W}_2$ in the spatial dimensions. We also show results using these two Kronecker product sparsifying bases together with *Global* measurements Φ that depend on all voxels of the datacube.

We see an improvement on recovery from distributed over global measurements when the number of measurements M obtained for each band is small; as M increases, this advantage vanishes due to the availability of sufficient information. We also see that the performance of independent recovery improves as the number of spectral bands increases and eventually matches the performance of
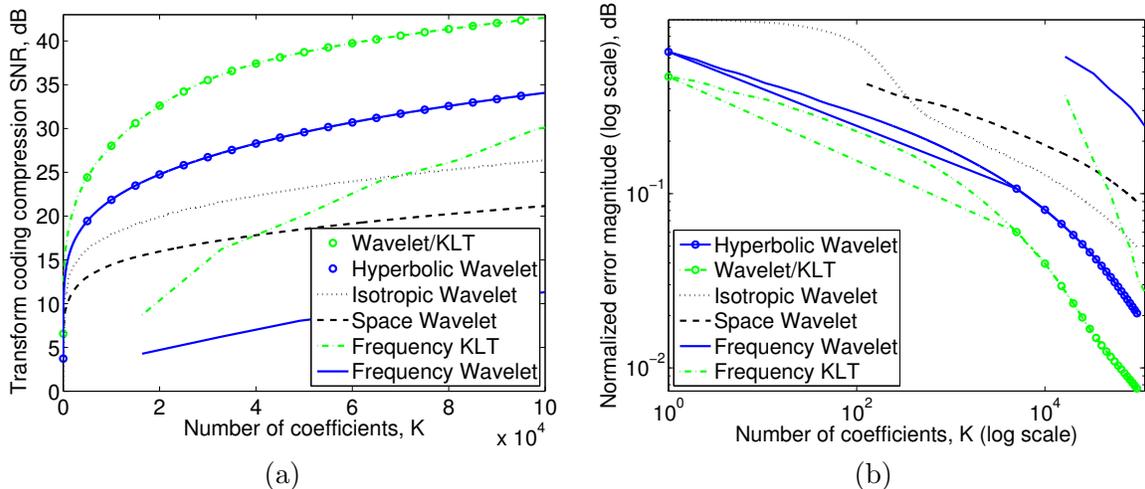
Figure 5: *Empirical performance of Kronecker product sparsity for hyperspectral imaging. A $N = 128 \times 128 \times 128$ voxel datacube is transform coded using a 2-D wavelet basis $\mathbf{W}_2$ for each spectral section (Space Wavelet), a 1-D wavelet basis $\mathbf{W}_1$ and a 1-D KLT basis $\mathbf{P}_1$ for each pixel (Frequency Wavelet, Frequency KLT), a 3-D isotropic wavelet basis for the entire datacube (Isotropic Wavelet), and Kronecker products bases $\mathbf{W}_1 \otimes \mathbf{W}_2$ and $\mathbf{P}_1 \otimes \mathbf{W}_2$ for the entire datacube (Hyperbolic Wavelet, Wavelet/KLT). In each case a number K of wavelet coefficients are preserved to obtain a compressed version of the datacube. (a) Compression SNR and (b) normalized error magnitude as a function of the number of coefficients K. The Kronecker products performs better than either component basis independently. However, the rate of decay of the compression error using the Kronecker product bases are approximately the same as the lower rate obtained from the component bases.*

global measurements. In other words, the performance of the Kronecker-based approaches, which involve the same CS measurement matrix and spatial transform, fails to improve in a similar fashion as the number of spectral bands increases. We conjecture that such penalty is due to the localized nature (in the spectral dimension) of the elements used in the sparsity bases (wavelets and KLT basis functions). Since the measurements used in KCS are localized, the measurement and sparsity bases become increasingly coherent as the spectral dimension resolution increases.

We finish by examining the computational complexity of the recovery algorithms for the $128 \times 128 \times 128$ voxel datacube problem. The average execution time for independent recovery of all spectral bands is approximately 9 minutes, while the average execution times for KCS and recovery from global measurements using the anisotropic wavelet basis for sparsity are approximately 25 minutes and 31 minutes, respectively. Similarly, the average execution times for KCS and recovery from global measurements using the wavelet/KLT Kronecker product basis for sparsity are approximately 35 minutes and 36 minutes, respectively. These increases are much more modest than what is anticipated by the theoretical discussion in Section 3.4.

5.3   Single-pixel hyperspectral camera

Our next experiment uses real-world data obtained from the SPHC [22] described in Section 1.4 using the separate measurements of (7). Figure 1(a) shows an example capture from the SPHC. The target is a printout of the *Mandrill* test image (illuminated by a desk lamp), for which 64 spectral bands spanning the 450–850 nm wavelength range at a resolution of 128×128 pixels were
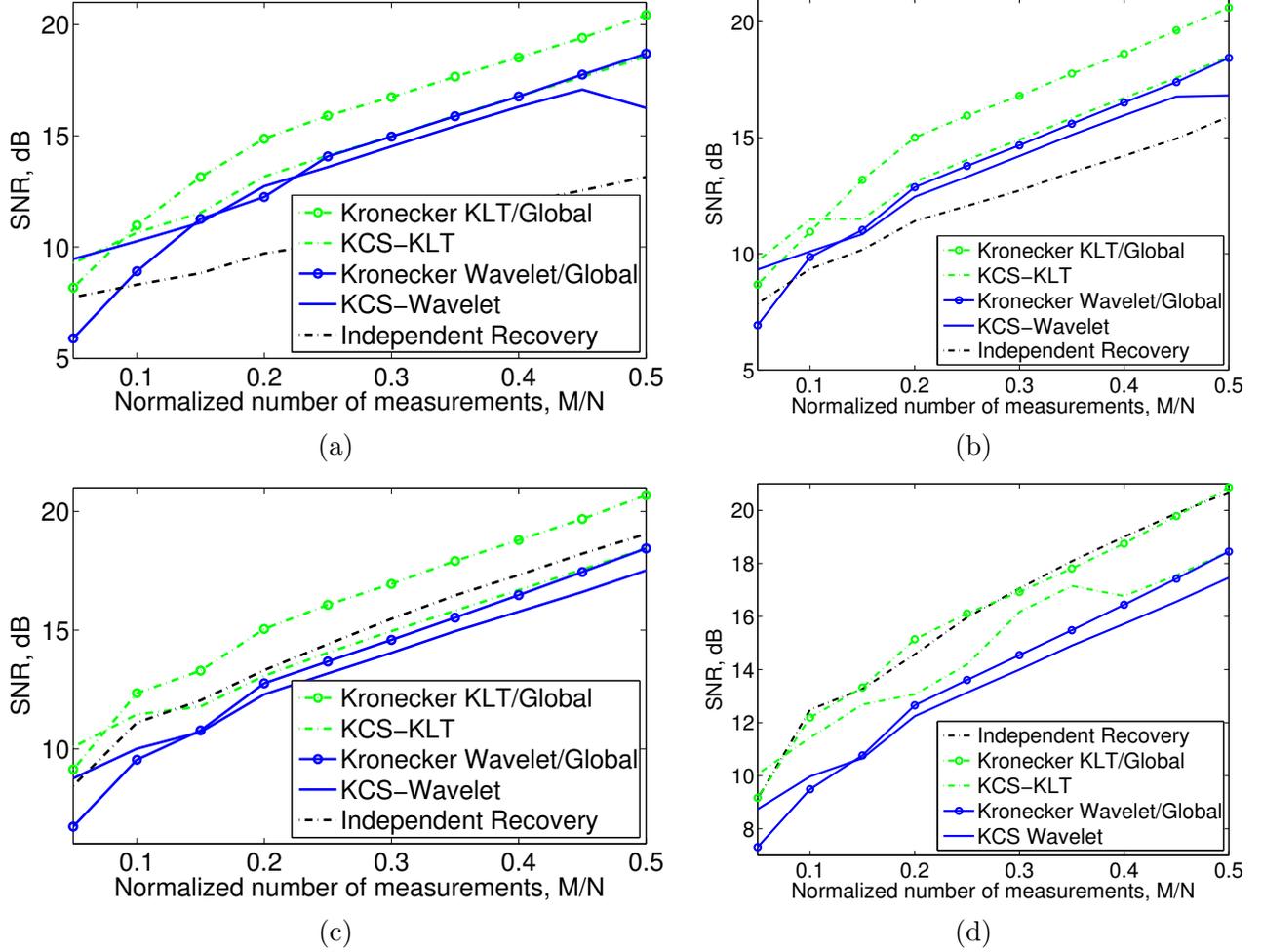
Figure 6: *Empirical performance of KCS and standard CS for hyperspectral imaging for datacubes of sizes (a) $N = 128 \times 128 \times 16$, (b) $N = 128 \times 128 \times 32$, (c) $N = 128 \times 128 \times 64$, and (d) $N = 128 \times 128 \times 128$ voxels. Each datacube is recovered from CS measurements of each spectral band image from a matrix Φ_2 using separate CS recovery of each spectral band image using the measurement matrix Φ_2 and a sparsifying 2-D wavelet basis $\mathbf{W}_2$ (Independent Recovery); joint CS recovery of all spectral bands using a global measurement matrix Φ and sparsifying Kronecker product bases $\mathbf{W}_1 \otimes \mathbf{W}_2$ and $\mathbf{P}_1 \otimes \mathbf{W}_2$ (Kronecker Wavelet/Global and Kronecker KLT/Global, respectively); and KCS recovery using the measurement matrix $\mathbf{I} \otimes \Phi_2$ and the sparsifying bases $\mathbf{W}_1 \otimes \mathbf{W}_2$ and $\mathbf{P}_1 \otimes \mathbf{W}_2$ (KCS-Wavelet and KCS-KLT, respectively). Recovery using the Kronecker product sparsifying bases outperforms separate recovery. Additionally, there is an advantage to applying distributed rather than global measurements when the number of measurements M is low. Furthermore, as the resolution of the spectral dimension increases, the Kronecker sparsity and Kronecker measurement bases become increasingly coherent, hampering the performance of joint recovery techniques.*

obtained. In Fig. 1(b), each spectral band was recovered separately. In Fig. 1(c), the spectral bands were recovered jointly with KCS using the measurement structure of (7) and a hyperbolic wavelet basis. The results show a considerable quality improvement over independent recovery, particularly for spectral frames with low signal power.
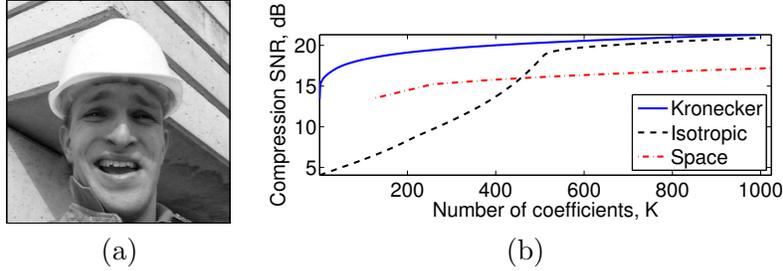
Figure 7: *(a) Example cropped frame from the QCIF-format* Foreman *video sequence, size* 128×128. *(b) Empirical performance of Kronecker product sparsifying basis for transform coding of the* Foreman *video sequence,* $N = 128 \times 128 \times 128 = 2^{21}$ *voxels. We subject it to transform coding using a spatial 2D wavelet basis* $\mathbf{W}_2$ *for each frame (*Space*), a 3-D* Isotropic *wavelet bases* $\mathbf{W}_3$ *for the entire sequence, and a Kronecker product basis* $\mathbf{W}_1 \otimes \mathbf{W}_2$ *for the entire sequence. The Kronecker product performs better in distortion than the alternative bases.*

5.4 Video data

5.4.1 Compressibility

We evaluate the compressibility of video sequences in an independent spatial (per frame) sparsifying basis and compare it with a standard isotropic wavelet basis and a Kronecker product wavelet basis. We use the standard *Foreman* video sequence, which we crop around the center to have frames of size 128×128 pixels, as shown in Fig. 7(a). We select 128 frames to obtain a signal of length $N = 2^{21}$ samples. We then process the signal through three different transforms: the first (*Space*) applies the 2-D wavelet basis $\mathbf{W}_2$ along the spatial dimensions of the data, with no compression on the temporal dimension; the second (*Isotropic*) applies the standard isotropic 3D wavelet basis $\mathbf{W}_3$ on the entire video sequence, and the third (*Kronecker*) transforms the entire sequence with the Kronecker product basis $\mathbf{W}_1 \otimes \mathbf{W}_2$, providing a hyperbolic wavelet basis. For each one of these transforms, we measured the compression signal-to-noise ratio (SNR) when transform coding is used to preserve K coefficients of the data for varying values of K. The results are shown in Fig. 7(b) and closely resemble those obtained for hyperspectral data. Additionally, the Kronecker product outperforms isotropic wavelets due to the difference in smoothness between the temporal and spatial dimensions.

5.4.2 KCS

We compare the performance of KCS to that of CS using the low-dimensional basis $\mathbf{W}_2$ to yield compressible/approximately sparse coefficient vectors for individual frames. In our experiments we obtain CS measurements on each video frame using a matrix Φ_2 obtained from a subsampled permuted Hadamard transform [7]. For KCS we use a single Kronecker product measurement matrix as shown in (7), while for standard CS we perform independent recovery of each frame using the measurement matrix Φ_2. We also use a *global* CS measurement matrix Φ, where the measurements depend on all the pixels of the video sequence, as a baseline. Figure 8 shows the recovery error from several different setups. *Independent* recovery uses CS on each video frame independently with the sparsifying basis $\mathbf{W}_2$. KCS employs the Kronecker product measurement and sparsity/compressibility transform matrices $\mathbf{I} \otimes \Phi_2$ and $\mathbf{W}_1 \otimes \mathbf{W}_2$, respectively, to perform joint recovery of all frames. We also show results using the Kronecker product sparsity/compressibility
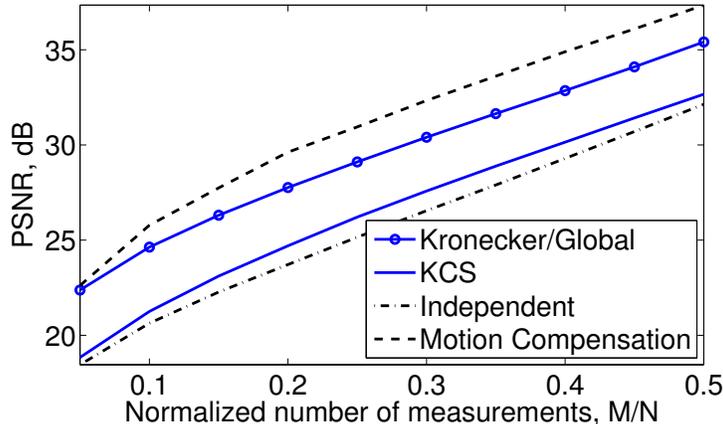
Figure 8: *Empirical performance of KCS for the* Foreman *video sequence. We recover the video sequence using* independent *recovery of each frame using a measurement matrix* Φ_2 *and sparsifying basis* $\mathbf{W}_2$*; independent block-based CS and recovery with* motion compensation *post-processing on individual frames with GOP size 8, block-wise random measurement matrix and block-wise 2-D discrete cosine transform (DCT) sparsifying basis; KCS with measurement matrix* $\mathbf{I} \otimes \Phi_2$ *and sparsifying basis* $\mathbf{W}_1 \otimes \mathbf{W}_2$*; and joint recovery of all frames in the sequence using the Kronecker sparsifying basis* $\mathbf{W}_1 \otimes \mathbf{W}_2$ *and a* global *measurement matrix* Φ*. While KCS does not perform as well as CS using global measurements, it shows an improvement over separate recovery of each frame in the video sequence using the same measurements. The motion compensation-aided approach outperforms the generic approaches.*

transform basis $\mathbf{W}_1 \otimes \mathbf{W}_2$ paired with the *Global* measurement matrix Φ.

Finally, we compare the above linear approaches to a state-of-the-art recovery algorithm based on nonlinear motion compensated block-based CS (MC-BCS) [21]. In MS-BCS disjoint blocks of each video frame are separately measured using both a random measurement matrix and a 2-D discrete cosine transform (DCT) for sparsity/compressibility. The blocks of a reference frame are recovered using standard CS recovery algorithms. MC-BCS then calculates measurements for the difference with the subsequent frame by subtracting the corresponding measurement vectors, and recovers the blocks of the frame difference using standard CS algorithms. The frame difference is then refined using motion compensation (MC); the MC output is used to obtain a new frame difference and the process is repeated iteratively for each frame, and again for each subsequent frame in the group of pictures (GOP). Further refinements enable additional improvements in the quality of the recovered video sequence. A toolbox implementing MC-BCS was made available while this paper was under review [21]. We set the GOP size to 8 and use blocks of size 16×16, following the parameter values of the toolbox implementation. In contrast to [21], we set the number of measurements for each of the frames to be equal to match the KCS partitioning of measurements.

The *Foreman* sequence features camera movement, which is reflected in sharp changes in the value of each pixel across frames. We see, once again, that while KCS does not perform as well as CS with global measurements, it does outperform independent recovery of each frame in the sequence operating on the same measurements. Furthermore, the quality of KCS recovery comes within 5 dB of that of MC-BCS, which may be surprising considering that the motion compensation performed in MC-BCS is especially designed for video coding and compression.

The average execution time for independent recovery of all video frames is approximately 13 seconds. In contrast, the average execution times for KCS and recovery from global measurements

using the anisotropic wavelet basis for sparsity are approximately 104 minutes and 220 minutes, respectively. These results agree with the discussion in Section 3.4, since the computational time of KCS recovery is increased by a factor of about $128 \times 3 = 384$ over that of independent recovery. The average execution time for motion compensation-aided recovery is 34 minutes.

6 Related work

Prior work for CS of multidimensional signals focuses on the example applications given in this paper – hyperspectral imaging [11, 44–46] and video acquisition [10, 12–15, 17, 20, 21] – with limited additional work dealing with applications such as sensor networks [9, 18, 19] and confocal microscopy [16]. These formulations employ measurement schemes that act on a partition of the data $\{\mathbf{x}_1, \dots, \mathbf{x}_J\}$, such as frames of a video sequence. For those cases, individual measurement vectors $\{\mathbf{y}_1, \dots, \mathbf{y}_J\}$ are obtained using a set of matrices $\{\Phi_1, \dots, \Phi_J\}$ [9–17, 20, 21], resulting in the measurement matrix structure of (7). While global measurements that depend on the entire set of data have been proposed [8, 10, 16, 20], practical architectures that provide such measurements are rare [8]. Similarly, partitioned measurements have been proposed for CS of low-dimensional signals for computational purposes [29, 30, 47]. Below we contrast the signal model and algorithms used in these approaches with those used in KCS.

Several frameworks have been proposed to encode the sparse structure of multidimensional signals. The most significant class of structures link the signals through overlap of nonzero coefficient values and locations. That is, there exists a matrix $\mathbf{P}$ of size $JN \times D$ with binary entries (0 or 1) and a vector Θ of length D such that $\overline{\mathbf{x}} = (\mathbf{I} \otimes \Psi)\mathbf{P}\Theta$. The vector Θ encodes the correlations and has length lower than the sum of the sparsities of the signals [9, 13, 15, 18, 19]. Such matrices are very rigid in the kinds of structures they can represent; in KCS we can represent a variety of multidimensional structures by using sparse representations on each of the signal dimensions.

Kronecker product matrices have been proposed for use as sparsifying bases in CS for certain spatiotemporal signals [12, 14, 20]. In other cases, specialized compression bases are combined with specially tailored recovery algorithms [11, 15, 17, 20, 21]; a prime example is motion compensation for video sequences [15, 21]. While such tailored algorithms often provide superior performance, they seldom come with theoretical tractability and performance guarantees. In contrast, KCS can use a variety of standard CS recovery algorithms and preserves their guarantees, since it relies on standard matrices for measure and sparsity/compressibility transforms. Standard sparsifying bases for CS, such as multidimensional isotropic wavelets, suffice only for very specific classes of signals that do feature similar degrees of smoothness along each of their dimensions [10, 16]; KCS using hyperbolic wavelet bases can be applied to signals with different degrees of smoothness in each of their dimensions.

In transform coding, anisotropic and hyperbolic wavelet bases have been proposed for compression of hyperspectral datacubes and video sequences [20, 38, 46]; however, to date no mathematical analysis of their performance has been provided. Kronecker products involving matrices obtained from principal component analysis and Karhunen-Loève transforms have also been used for this purpose. However, they rely on linear low-dimensional approximations rather than nonlinear sparse representations [20, 44, 45]; thus, the approaches are more data-dependent and more difficult to generalize among different datasets.

Finally, we are aware of two initial studies on the properties of Kronecker product matrices for CS [29, 30, 32]. Our study of their mutual coherence properties matches that independently obtained

in [29, 30], while [32] provides only a lower bound for their restricted isometry constants; we have provide an upper bound based on the properties of the eigendecomposition of their submatrices.

7   Conclusions and Further Work

In this paper we have developed the concept of Kronecker compressive sensing and presented initial analytical and experimental results on its performance. Out theoretical framework is motivated by new sensing applications that acquire multidimensional signals in a progressive fashion, as well as by settings where the measurement process is distributed, such as sensor networks and arrays. We have also provided analytical results for the recovery of signals that live in anisotropic Besov spaces, where there is a well-defined relationship between the degrees of compressibility obtained using lower-dimensional wavelet bases on portions of the signal and multidimensional anisotropic wavelet bases on the entire signal. Furthermore, because the formulation follows the standard CS approach of single measurement and sparsifying matrices, standard recovery algorithms that provide provable guarantees can be used; this obviates the need to develop ad-hoc algorithms to exploit additional signal structure.

Further work remains in finding additional signal classes for which the use of multidimensional structures provides an advantage during compression. Some promising candidates include modulation spaces, which contain signals that can be compressed using Wilson and brushlet bases [48, 49]. Our KCS framework also motivates the formulation of novel structured representations using sparsifying bases in applications where transform coding compression schemes have not been developed.

While we focused on hyperspectral imaging and video acquisition, there exist other interesting applications where Kronecker product sparsifying bases and KCS are relevant. In sensor networks and arrays, sparsity-based distributed localization [24, 50, 51] obtains a sparse estimate of the vector containing the samples obtained in a dictionary that contains the responses of a known source at a set of feasible locations. The sparse vector will encode the location of the source within the feasible set. When the source signal is not known, we can assume that it is sparse in a known basis and employ a Kronecker product matrix that encodes both the propagation physics and the sparse or compressible structure of the source signal. In medical imaging, there are many applications where estimates of high-dimensional data are obtained from highly undersampled measurements, including 3-D computed tomography, angiography [12], 3-D magnetic resonance imaging (MRI) [12], and functional MRI.

References

[1] E. J. Candès, "Compressive sampling," in *Int. Congress of Mathematicians*, Madrid, Spain, 2006, vol. 3, pp. 1433–1452.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, Sept. 2006.

[3] D. J. Brady, "Multiplex sensors and the constant radiance theorem," *Optics Letters*, vol. 27, no. 1, pp. 16–18, Jan. 2002.

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

[5] J. Haupt and R. Nowak, "Signal reconstruction from noisy random projections," *IEEE Trans. Info. Theory*, vol. 52, no. 9, pp. 4036–4048, Sept. 2006.

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

[7] 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.*, vol. 25, no. 2, pp. 83–91, March 2008.

[8] A. Wagadarikar, R. John, R. Willett, and David Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Applied Optics*, vol. 47, no. 10, pp. B44–B51, 2008.

[9] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," Tech. Rep. TREE-0612, Rice University Department of Electrical and Computer Engineering, Houston, TX, Nov. 2006.

[10] M. B. Wakin, J. N. Laska, M. F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "Compressive imaging for video representation and coding," in *Picture Coding Symposium*, Beijing, China, Apr. 2006.

[11] R. M. Willett, M. E. Gehm, and D. J. Brady, "Multiscale reconstruction for computational hyperspectral imaging," in *IS&T/SPIE Symposium on Electronic Imaging: Comp. Imaging*, San Jose, CA, Jan. 2007.

[12] M. Lustig, D. L. Donoho, J. M. Santos, and J. M. Pauly, "Compressed sensing MRI," *IEEE Signal Proc. Mag.*, vol. 25, no. 2, pp. 72–82, Mar. 2008.

[13] R. F. Marcia and R. M. Willett, "Compressive coded aperture video reconstruction," in *European Signal Proc. Conf. (EUSIPCO)*, Lausanne, Switzerland, Aug. 2008.

[14] V. Stankovic, L. Stankovic, and S. Cheng, "Compressive video sampling," in *European Signal Proc. Conf. (EUSIPCO)*, Lausanne, Switzerland, Aug. 2008.

[15] L.-W. Kang and C.-S. Lu, "Distributed compressed video sensing," in *IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 1169–1172.

[16] P. Ye, J. L. Paredes, G. R. Arce, Y. Wu, C. Chen, and D. W. Prather, "Compressive confocal microscopy: 3D reconstruction algorithms," in *SPIE Photonics West*, San Jose, CA, Jan. 2009.

[17] J. Y. Park and M. B. Wakin, "A multiscale framework for compressive sensing of video," in *Picture Coding Simposium*, Chicago, IL, May 2009.

[18] Y. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Info. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.

[19] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Info. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

[20] M Cossalter, M. Tagliasacchi, G. Valenzise, and S. Tubaro, "Joint compressive video coding and analysis," *IEEE Trans. Multimedia*, vol. 12, no. 3, pp. 168–183, Apr. 2010.

[21] S. Mun and J. E. Fowler, "Residual reconstruction for block-based compressed sensing of video," in *IEEE Data Compression Conf. (DCC)*, Snowbird, UT, Mar. 2011.

[22] T. Sun and K. F. Kelly, "Compressive sensing hyperspectral imager," in *Comp. Optical Sensing and Imaging (COSI)*, San Jose, CA, Oct. 2009.

[23] D. Takhar, *Compressed sensing for imaging applications*, Ph.D. thesis, Rice University, Houston, TX, Feb. 2008.

[24] V. Cevher, M. F. Duarte, and R. G. Baraniuk, "Localization via spatial sparsity," in *European Signal Proc. Conf. (EUSIPCO)*, Lausanne, Switzerland, 2008.

[25] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. on Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1998.

[26] E. J. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, 2007.

[27] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de l'Academie des Sciences Series I*, vol. 346, pp. 589–592, 2008.

[28] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Comm. on Pure and Applied Math.*, vol. 61, no. 8, pp. 1025–1171, Aug. 2008.

[29] Y. Rivenson and A. Stern, "Compressed imaging with a separable sensing operator," *IEEE Signal Proc. Letters*, vol. 16, no. 6, pp. 449–452, June 2009.

[30] Y. Rivenson and A. Stern, "An efficient method for multi-dimensional compressive imaging," in *Comp. Optical Sensing and Imaging (COSI)*, San Jose, CA, Oct. 2009.

[31] R. C. Thompson, "Principal submatrices XI: Interlacing inequalities for singular values," *Linear Algebra and Appl.*, vol. 5, pp. 1–12, 1972.

[32] S. Jokar and V. Mehrmann, "Sparse representation of solutions of Kronecker product systems," *Linear Algebra and Appl.*, vol. 431, no. 12, pp. 2437–2447, Dec. 2009.

[33] S. Boyd and L. Vanderberghe, *Convex Optimization*, Cambridge University Press, New York, NY, 2004.

[34] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, Academic Press, San Diego, CA, 3rd. edition, Dec. 2008.

[35] R. Hochmuth, "Wavelet characterizations for anisotropic Besov spaces," *Applied and Comp. Harmonic Analysis*, vol. 12, no. 2, pp. 179–208, Mar. 2002.

[36] R. Hochmuth, "N-term approximation in anisotropic function spaces," *Mathematische Nachrichten*, vol. 244, no. 1, pp. 131–149, Oct. 2002.

[37] R. A. DeVore, "Nonlinear approximation," *Acta Numerica*, vol. 7, pp. 51–150, 1998.

[38] E. Cristophe, C. Mailhes, and P. Duhamel, "Hyperspectral image compression: Adapting SPIHT and EZW to anisotropic 3-D wavelet coding," *IEEE Trans. Image Proc.*, vol. 17, no. 12, pp. 2334–2346, Dec. 2008.

[39] A. Kutil, "Anisotropic 3-D wavelet packet bases for video coding," in *IEEE Int. Conf. on Image Proc. (ICIP)*, Barcelona, Spain, Sept. 2003, vol. 2, pp. II–73–76.

[40] M. F. Duarte, *Compressive sensing for signal ensembles*, Ph.D. thesis, Rice University, Houston, TX, July 2009.

[41] E. J. Candès and J. K. Romberg, "The ℓ_1-magic toolbox," `http://www.l1-magic.org`.

[42] E. van den Berg and M. P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, Nov. 2008.

[43] NASA Jet Propulsion Laboratory, "AVIRIS Homepage," `http://aviris.jpl.nasa.gov`.

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

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

[46] J. E. Fowler and J. T. Rucker, "3D wavelet-based compression of hyperspectral imagery," in *Hyperspectral Data Exploitation: Theory and Applications*, chapter 14, pp. 379–407. John Wiley and Sons, Inc., Hoboken, NJ, 2007.

[47] L. Gan, "Block compressed sensing of natural images," in *Int. Conf. Digital Signal Proc.*, Cardiff, Wales, July 2007, pp. 403–406.

[48] H. G. Feichtinger, K. Gröchenig, and D. Walnut, "Wilson bases and modulation spaces," *Mathematische Nachrichten*, vol. 155, no. 1, pp. 7–17, Jan. 1992.

[49] L. Borup and M. Nielsen, "Nonlinear approximation in α-modulation spaces," *Mathematische Nachrichten*, vol. 279, no. 1–2, pp. 101–120, Jan. 2006.

[50] V. Cevher, A. C. Gurbuz, J. H. McClellan, and R. Chellappa, "Compressive wireless arrays for bearing estimation," in *IEEE Int. Conf. Acoustics, Speech, Signal Proc. (ICASSP)*, Las Vegas, NV, Apr. 2008, pp. 2497–2500.

[51] D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Proc.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.