Kronecker Compressive Sensing

Marco F. Duarte, *Member, IEEE,* and Richard G. Baraniuk, *Fellow, IEEE*

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; the construction of sparsifying bases and measurement systems for such signals is complicated by their higher dimensionality. In this paper, 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 structure present in all of the signal dimensions. Second, such matrices can represent the measurement protocols used in distributed settings. Our formulation 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.

I. Introduction

A. CS and multidimensional signals

Compressive sensing (CS) [1, 2] 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 $\mathbf{x}$

Paper submitted December 16, 2009; revised March 17, 2011 and July 28, 2011.

This work was supported by grants NSF CCF-0431150 and CCF-0728867, DARPA/ONR N66001-08-1-2065, ONR N00014-07-1-0936 and N00014-08-1-1112, AFOSR FA9550-07-1-0301, ARO MURIs W911NF-07-1-0185 and W911NF-09-1-0383, and the Texas Instruments Leadership Program.

M. F. Duarte was with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005. He is now with the Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA 01003. E-mail: mduarte@ecs.umass.edu

R. G. Baraniuk is with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005. Email: richb@rice.edu

having a sparse or compressible representation θ in some basis Ψ, i.e., $\mathbf{x} = \Psi\theta$. 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. Many natural signals are sparse or compressible; for example, smooth signals are compressible in the Fourier basis, while piecewise smooth images are compressible in a wavelet basis. In CS the signal is measured not via standard point samples but rather through the projection by a measurement matrix $y = \Phi x$. Such measurements *multiplex* the entries of x when the matrix Φ is dense.

To date the CS literature has mostly focused on problems involving single sensors and one-dimensional (1-D) signal or 2-D image data. However, many 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 measured 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, e.g., environmental sensors, microphone arrays, and camera arrays. These properties of multidimensional data and the corresponding acquisition hardware complicate the design of both the CS measurement matrix Φ and the sparsifying basis Ψ to achieve maximum CS efficiency.

B. CS measurements 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 to the choice of basis Ψ, since dense measurements are needed to capture arbitrary sparsity structure [3]. 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 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, and global multiplexing measurements require aggregation throughout the video acquisition. 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 [4, 5]; such independent multiplexing nature limits

the structure of the measurements obtained.

These application-specific limitations naturally point us in the direction of *partitioned measurements* that depend only on a subset of the entries of the multidimensional signal being acquired. 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.

C. Sparsifying and measurement 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 [6–18] 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 independent signals and impose 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. Additionally, 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. Furthermore, nearly 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 an 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 independent multiplexing operations on each signal dimension. As we will see below, KCS enables the derivation of analytical bounds for recovery of compressible multidimensional signals from randomized or incoherent measurements.

D. Stylized applications

To better motivate the KCS concept, we will consider in this paper three relevant multidimensional CS applications: hyperspectral imaging, video acquisition, and distributed sensing. Consider the hyperspectral imaging application (the other two applications have similar attributes and are discussed in more detail below in Section V). In the single-pixel hyperspectral camera (SPHC) [4, 19], 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 the 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 (e.g., wavelets, DCT) as well as each pixel's spectral signature (e.g., wavelets), which can be integrated into a single Kronecker product sparsifying basis. An example SPHC datacube recovered via KCS is shown in Fig. 1.

E. Contributions

This paper makes three main contributions. First, we propose Kronecker product matrices as sparsifying bases for multidimensional signals to jointly model the signal structure along each of its dimensions. In some cases, such as Kronecker product wavelet bases, we can obtain theoretical bounds for the rate of decay of the signal coefficient magnitudes for certain kinds of data. This rate of decay is dependent on the rates of decay for the coefficient magnitudes of sections of the signals across the different dimensions using the individual bases. When the decay rates using the individual bases for each of the dimensions are different, we show that the Kronecker product basis rate falls between the maximum and minimum

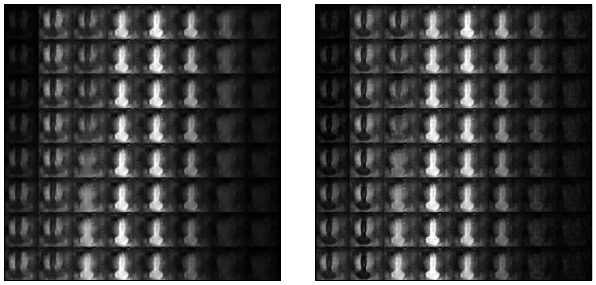

Fig. 1. *Example single-pixel hyperspectral camera [19] capture of the Mandrill test image printed and illuminated by a desk lamp at resolution* $N = 128 \times 128$ *pixels* $\times 64$ *spectral bands (*2^{20} *voxels) covering the 450nm–850nm wavelength range from* $M = 4096$ *CS measurements per band (*$4\times$ *sub-Nyquist). Left: Hyperspectral datacube obtained via independent CS recovery of each spectral band as a independent image. Right: 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 [19, 20].*

of these rates; therefore, when all individual rates of decay are equal, a Kronecker product basis attains the same rate.

Second, we show that several different CS measurements schemes proposed for multidimensional signals can be easily expressed in our Kronecker product framework. KCS also inspires the design of new multi-stage CS acquisition devices where the first stage applies the same measurement matrix on each portion of the signal along a subset of its dimensions and subsequent stages operate on the measurements previously obtained and aligned in a multidimensional fashion.

Third, we provide metrics to evaluate partitioned measurement schemes against Kronecker measurement matrices; these metrics provide guidance on the improvements that may be afforded by the use of such multidimensional structures. In particular, we provide some initial theoretical and practical results for Besov space signals that are compressible in a Kronecker product of wavelet bases. By comparing the recovery error decay rate, we establish conditions under which KCS outperforms standard partitioned CS recovery from independent measurements. Finally, we verify our theoretical findings using experimental results with synthetic and real-world multidimensional signals.

This paper is organized as follows. Section II provides the necessary background on CS. Section III introduces KCS and Section IV provides initial results for wavelet-sparse signals. Section V provides experimental results, Section VI summarizes related work, and Section VII closes with conclusions and suggestions for future work.

II. Background on Compressive Sensing

Let $\mathbf{x} \in \mathbb{R}^N$ be the signal of interest. We say that an orthonormal basis $\Psi \in \mathbb{R}^{N \times N}$ *sparsifies* $\mathbf{x}$ if $\theta = \Psi^T \mathbf{x}$ has only $K \ll N$ nonzero entries, with Ψ^T the transpose of Ψ. We also say that $\mathbf{x}$ is K-sparse or has *sparsity* K in Ψ. We say that Ψ *compresses* $\mathbf{x}$ if the entries of θ, when sorted by magnitude, decay according to $|\theta(i)| < Ci^{-1/p}$ for some $p \leq 1$ and $C < \infty$. In this case, we say that θ is *s-compressible* in Ψ, with $s = 1/p - 1/2$. Such vectors can be compressed using *transform coding* by retaining the K largest coefficients by magnitude and setting the rest to zero; θ_K will denote this approximation. 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 an s-compressible signal, $\sigma_K(\theta) \leq C' K^{-s}$, i.e., the approximation error decays exponentially fast.

The CS acquisition protocol measures inner products of the signal against a set of measurement vectors $\{\phi_1, \ldots, \phi_M\}$. When $M < N$, this 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 succinctly as $\mathbf{y} = \Phi\mathbf{x} + \mathbf{n} = \Phi\Psi\theta + \mathbf{n}$, with the vector $\mathbf{y} \in \mathbb{R}^M$ containing the CS measurements contaminated by measurement noise $\mathbf{n} \in \mathbb{R}^M$.

One important goal of CS is to recover the 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 $\Upsilon = \Phi\Psi$. A key enabler of CS is the fact that, when the signal is sparse enough, it can be accurately recovered from its compressive measurements by solving the quadratic program [1]

$$\widehat{\theta} = \arg\min \|\theta\|_1 \text{ s.t. } \|\mathbf{y} - \Upsilon\theta\|_2 \leq \epsilon, \tag{1}$$

where ϵ is an upper bound on the ℓ_2 norm of the noise vector $\mathbf{n}$, and $\|\cdot\|_1$ denotes the ℓ_1 norm, which is equal to the sum of the absolute values of the vector entries.

Previous contributors have developed conditions on the number and type of measurement vectors necessary for stable signal recovery [1, 21]. The Restricted Isometry Property (RIP) has been proposed to measure the fitness of a matrix Υ for CS.

Definition 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 constants δ_K are determined, they can be used to provide guarantees for CS recovery.

Theorem 1. *[22] If the matrix Υ has $\delta_{2K} < \sqrt{2} - 1$, then the solution $\widehat{\theta}$ to (1) obeys*

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

where C_0 and C_1 depend only on δ_{2K}.

In words, Theorem 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 measurement noise.

Unfortunately, calculating the constants δ_K for a given matrix requires combinatorially complex computation. Consequently, many probabilistic classes of matrices have been advocated. For example, a matrix of size $M \times N$ with independent Gaussian entries with mean zero and variance $1/M$ obeys the condition of Theorem 1 with very high probability if $K \leq \mathcal{O}\left(M/\log(N/M)\right)$ [1, 2]. The same is true of matrices following subgaussian distributions [3].

In some applications, the sensing system constrains the types of measurement matrices that are feasible, either due to the computational power needed to take the measurements or due to the limitations of feasible multiplexing. To formalize this notion, we assume that a basis $\Phi \in \mathbb{R}^{N \times N}$ is provided for measurement purposes and that we have the option to choose a subset of the signal's coefficients in this transform as measurements. That is, set $\widetilde{\Phi}$ to 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 new metric arises to evaluate the performance of CS.

Definition 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 determines the number of measurements necessary for accurate CS recovery.

Theorem 2. *[23, Remark 3.5.2] Choose a subset $\Gamma \subseteq \{1, \ldots, N\}$ for the set of observed CS measurements, with $M = |\Gamma|$, uniformly at random. Suppose that*

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

for a fixed value of C. Then with probability at least $1 - 5e^{-t}$ over the choice of Γ, the 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 1 for compressible signals with the number of measurements M dictated by the mutual coherence value $\mu(\Phi, \Psi)$. Since the range of possible mutual coherence values $\mu(\Phi, \Psi)$ is $[N^{-1/2}, 1]$, the number of measurements required by Theorem 2 ranges from $O(K \log(N))$ to $O(K \log(N)\sqrt{N})$.

III. Kronecker Product Matrices for Multidimensional Compressive Sensing

We now develop 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 d^{th} 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 $\mathbf{x}_{i,j,\cdot} := [\mathbf{x}(i,j,1) \ \ldots \ \mathbf{x}(i,j,N_3)]$ is a 3-section of $\mathbf{x}$. The definition can be extended to subsets of the dimensions; for example, $\mathbf{x}_{\cdot,\cdot,i} = [\mathbf{x}(1,1,i) \ \ldots \ \mathbf{x}(N_1,N_2,i)]$ is a $\{1,2\}$-section of $\mathbf{x}$.

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

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

Thus, $A \otimes B$ is a matrix of size $PR \times QS$. The definition has a straightforward extension to the Kronecker product of vectors $a \otimes b$. Additionally, if Ψ_V and Ψ_W are bases for $\mathbb{R}^v$ and $\mathbb{R}^w$, respectively, then $\Psi_{V \otimes W} = \Psi_V \otimes \Psi_W$ is a basis for $\mathbb{R}^{vw}$.

A. Kronecker product sparsifying bases

We can obtain a single sparsifying basis for an entire multidimensional signal as the Kronecker product of 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. A sparsifying basis for $\mathbf{x}$ is then obtained using Kronecker products as $\overline{\Psi} = \Psi_1 \otimes \ldots \otimes \Psi_D$; the coefficient vector Θ for the signal ensemble gives $\mathbf{X} = \overline{\Psi}\Theta$, where $\mathbf{X}$ is a vector-reshaped representation of $\mathbf{x}$.

B. Kronecker product measurement matrices

We can also design measurement matrices using Kronecker products; such matrices correspond to measurement processes that independently multiplex portions of the multidimensional signal. We call

this approach *Kronecker CS* (KCS). 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 V 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 I-D where we obtain *independent measurements*, in the sense that each measurement depends on only one of the signals $\{\mathbf{x}_1\ \mathbf{x}_2\ \ldots\ \mathbf{x}_J\}$. The signals could be collected as a matrix $\overline{\mathbf{x}}$ so that each signal is a 1-section of $\overline{\mathbf{x}}$ and the measurements are distributed along the first dimension of $\overline{\mathbf{x}}$. More formally, for each signal $\mathbf{x}_j,\ 1 \leq j \leq J$ we obtain independent measurements $\mathbf{y}_j = \Phi_j \overline{\mathbf{x}}_{\cdot,j}$ with an independent measurement matrix being applied to each signal. The structure of such measurements can be succinctly captured by Kronecker products. To compactly represent the signal and measurement ensembles, we denote $\mathbf{X} = [\mathbf{x}_1^T\ \mathbf{x}_2^T\ \ldots\ \mathbf{x}_J^T]^T$, $\mathbf{Y} = [\mathbf{y}_1^T\ \mathbf{y}_2^T\ \ldots\ \mathbf{y}_J^T]^T$, and

$$\overline{\Phi} = \begin{bmatrix} \Phi_1 & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \Phi_2 & \ldots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \Phi_J \end{bmatrix}, \tag{3}$$

with $\mathbf{0}$ denoting a matrix of appropriate size with all entries equal to 0. We then have $\mathbf{Y} = \overline{\Phi}\mathbf{X}$. Equation (3) 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 independent 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.

C. CS performance with Kronecker product matrices

The CS recovery guarantees provided in Theorems 1 and 2 induce corresponding constraints on the KCS sparsifying and measurement matrices. These results provide a link between the CS recovery performance of the Kronecker product matrix and that of the individual matrices comprising the product.

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: $\overline{\Phi} = \Phi_1 \otimes \ldots \otimes \Phi_D$, where each pair Φ_d and Ψ_d is mutually incoherent for $d = 1, \ldots, D$. The following lemma provides a *conservation of mutual coherence* across Kronecker products (see [24–26] for a proof).

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

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

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 smaller than or equal to the corresponding necessary number of 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.

2) Restricted isometry constants: The constants δ_K for a matrix Φ are intrinsically tied to the singular values of all column submatrices of a certain size. The convenient structure of Kronecker product matrices yields simple bounds for their restricted isometry constants. The following lemma is proven in [24].

Lemma 2. *Let $\Phi_1, \ldots, \Phi_D$ be matrices with restricted isometry constants $\delta_K(\Phi_1), \ldots, \delta_K(\Phi_D)$, respectively. Then,*

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

When Φ_d is an orthonormal basis, it has restricted isometry constant $\delta_K(\Phi_d) = 0$ 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 this bound is loose in general due to the use of a matrix with K^2 columns in the proof in [24], we note that the RIP constant of the Kronecker product matrix is bounded below, by construction, by the largest RIP constant among its component matrices; that is, $\delta_K(\Phi_1 \otimes \Phi_2 \otimes \ldots \otimes \Phi_D) \geq \max_{1 \leq d \leq D} \delta_K(\Phi_d)$ [27]. 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.

D. Computational considerations

We briefly consider the computational complexity of KCS. There exist several solvers for the optimization program (1), including interior point methods, that have computational complexity $\mathcal{O}\left(N^3\right)$, where N denotes the length of the vector θ [28]. Thus, independent recovery of each d'-section of a

multidimensional dataset incurs a total complexity of $\mathcal{O}\left(N_{d'}^3 \prod_{d \neq d'} N_d\right)$. In contrast, KCS solves a single higher-dimensional optimization problem of complexity $\mathcal{O}\left(\prod_{d=1}^{D} N_d^3\right)$, providing a $\mathcal{O}\left(\prod_{d \neq d'} N_d^2\right)$-times larger computational cost for the improved performance afforded by the Kronecker product sparsity/compressibility basis. 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 $\mathcal{O}\left(A(N)/A(N_{d'})\right)$. There exist CS matrices with efficient implementations requiring just $A(N) = \mathcal{O}\left(N \log N\right)$, which yield a computational cost for KCS of approximately $\mathcal{O}\left(\prod_{d \neq d'} N_d \frac{\sum_{d=1}^{D} \log(N_d)}{\log(N_{d'})}\right)$. 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 d'^{th} dimension, i.e., $DN/N_{d'}$.

IV. Case Study: KCS with Multidimensional Wavelet Bases

Kronecker products play a central role in generating wavelet bases for multidimensional signals [29–31]. There are several different multidimensional wavelet constructions; for each 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 $t_d \in \Omega := [0,1], 1 \leq d \leq D$, such that $\mathbf{x}(n_1, \ldots, n_D) = f(n_1/N_1, \ldots, n_d/N_D)$, with $N = N_1 \times \ldots \times N_D$.

A. Isotropic and hyperbolic wavelets

The wavelet representation of a 1-D signal $f(t) : \Omega \to \mathbb{R}$ with $\Omega = [0,1]$ 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 from 1-D to D-D wavelets can be generated using different combinations of Kronecker products [29–31]. In each case, a D-D wavelet is obtained from the Kronecker product of D 1-D wavelets: $\psi_{i_1,j_1,\ldots,i_D,j_D} = \psi_{i_1,j_1} \otimes \ldots \otimes \psi_{i_D,j_D}$. For example, *isotropic wavelets* are formed from Kronecker products of 1-D wavelets that all live at the same scale i, while *hyperbolic wavelets* are formed from Kronecker products of 1-D wavelets at all possible combinations of different scales. In the sequel, we will identify the isotropic and hyperbolic wavelet bases using Ψ_I and Ψ_H, respectively.

B. Multidimensional Besov spaces

Besov spaces are featured prominently in modern signal compression and processing; they contain signals that are compressible using wavelets. The 1D Besov space $B^s_{p,q}$ contains functions that have (roughly speaking) s derivatives in L_p; the parameter q makes finer distinctions in smoothness [32, 33].

Just as there are several different wavelet basis constructions in higher dimensions, there are several different types of Besov spaces. The isotropic wavelet basis compresses signals from *isotropic Besov spaces*, which contain signals having the same degree of smoothness along each of their dimensions. The hyperbolic wavelet basis compresses signals from *anisotropic Besov spaces*, which contain signals with different degrees of smoothness along each of their dimensions.Our interest will be primarily in the latter spaces. A video sequence, for instance, has different degrees of smoothness along its spatial and temporal dimensions, while a hyperspectral datacube can have different degrees of smoothness in the spatial and spectral dimensions.

We provide a brief overview of anisotropic Besov spaces for completeness [30, 31]. Consider the D-D function $f(t) := f(t_1, \ldots, t_D) : \Omega^D \to \mathbb{R}$. Define the d-directional derivative of f in the direction h as $(\Delta^1_{h,d} f)(t) := f(t + he_d) - f(t),\ 1 \le d \le D$, where e_d is the d^{th} canonical vector, i.e., its d^{th} entry is one and all others are zero. Recursively define the higher-degree directional derivatives as $(\Delta^m_{h,d} f)(t) := (\Delta_{h,d}(\Delta^{m-1}_{h,d} f))(t),\ m \ge 2$. For $r \in \mathbb{R}^+$, $m_d \in \mathbb{N}$ and $0 < p < \infty$, define the d-directional moduli of smoothness as

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

By setting the anisotropy parameter $\overline{s} = (s_1, \ldots, s_D)$, define the anisotropic Besov quasi-seminorm as

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

We say that a signal $f \in B^{\overline{s}}_{p,q}$ if it has finite anisotropic Besov norm, defined as $\|f\|_{B^{\overline{s}}_{p,q}} = \|f\|_p + |f|_{B^{\overline{s}}_{p,q}}$. Signals in an anisotropic Besov space $B^{\overline{s}}_{p,q}$ are λ-compressible by hyperbolic wavelet bases, where

$$\lambda = \frac{D}{\sum_{d=1}^{D} 1/s_d}; \tag{4}$$

see [24, 29–31, 34] for details. This property is particularly relevant since the hyperbolic wavelet basis can be expressed as a D-term Kronecker product matrix, a fact that we leverage to characterize the performance of KCS in the next subsection.

Isotropic Besov spaces are a special case of anisotropic Besov spaces where the smoothness s_d along each dimension d is the same, and so $\lambda = s_d = s$.

C. KCS with hyperbolic wavelets

We now compare the approximation rate obtained by KCS vs. independent measurements of each d-section of a multidimensional signal. The following Theorem is proven in [24].

Theorem 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 $f \in B^{\overline{s}}_{p,q}$, with $\overline{s} = (s_1, \ldots, s_D)$. In particular, $\mathbf{x}$ has s_d-compressible d-sections in sufficiently smooth wavelet bases $\Psi_d,\ 1 \le d \le D$. Denote by $\Phi_d,\ 1 \le d \le D$ 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 is such that*

$$\|\mathbf{x} - \widehat{\mathbf{x}}\|_2 \le C \left(\frac{M}{\sqrt{N} \prod_{d=1}^{D} \mu(\Phi_d, \Psi_d)} \right)^{-\beta}, \tag{5}$$

where $\beta = \frac{D}{2\sum_{d=1}^{D} 1/s_d} + \frac{1}{4}$. In contrast, the recovery error from M measurements equally distributed among the e^{th} dimension of the signal using a subsampling of 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 $\mathbf{x}$ is such that

$$\|\mathbf{x} - \widehat{\mathbf{x}}\|_2 \le C\sqrt{N_e} \left(\frac{M}{\sqrt{N/N_e} \prod_{d \ne e} \mu(\Phi_d, \Psi_d)} \right)^{-\beta_e}, \tag{6}$$

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

To put Theorem 3 in perspective, we study the bases and the exponents of the bounds separately. With regards to the bases, the denominators in (5)–(6) 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 (4) 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 (6) of partitioning measurements, which comes from the triangle inequality.

V. Experimental Results

In this section, we experimentally verify the performance of KCS. In addition to synthetic data, we use 3-D 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 III and IV and contrast the recovery performance to that reached by integrating such task-specific compression schemes to distributed CS recovery.

Our experiments use the ℓ_1-norm minimization solvers from [35] and [36] for the hyperspectral and video data, respectively. All experiments were executed on a Linux workstation with a 3.166 GHz Intel Xeon CPU with 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 available in [24, 34].

A. 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 CS recovery from: (i) dense, *global measurements*; (ii) a single *dense KCS* recovery via a Kronecker product measurement matrix obtained from two dense matrices of sizes 64×64 (for each 1-section) and 8×8 (for each $\{2,3\}$-section); (iii) a single *distributed KCS* recovery from the set of measurements obtained independently from each 8×8 1-section; (iv) *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

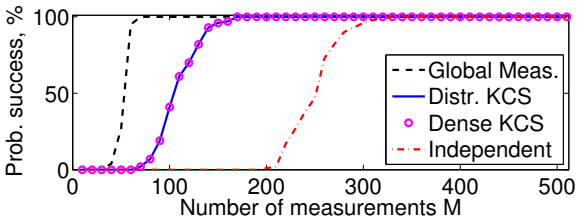

Fig. 2. *Empirical performance of KCS. We generate 3-D signals of size* $N = 8 \times 8 \times 8$ *that are* $K = 10$*-sparse in a hyperbolic wavelet basis. While distributed and dense KCS do not match the performance of global measurements, they perform significantly better than independent recovery.*

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 $\|\mathbf{x} - \widehat{\mathbf{x}}\|_2 \leq 10^{-3}\|\mathbf{x}\|_2$. We see in Fig. 2 that the above approaches perform best to worst in the order (i)–(iv) presented above. In particular, the measurement-to-sparsity ratio M/K required for 95% successful recovery are 6, 15, and 30 for global measurements, KCS, and independent recovery, respectively.

B. Hyperspectral data

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 a $N = 128 \times 128 \times 128$ voxel portion of the AVIRIS Moffett Field database [37]. We process the signal through six different transforms. The first three (*Space Wavelet, 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[1] $\mathbf{P}_1$ for the spectral dimension, respectively). The fourth (*Isotropic Wavelet*) transforms the entire datacube with a 3-D isotropic wavelet basis. The fifth and 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

[1]A KLT basis is learned from a datacube of the same size extracted from a different spatial region of the original AVIRIS dataset [17, 38, 39]. 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.

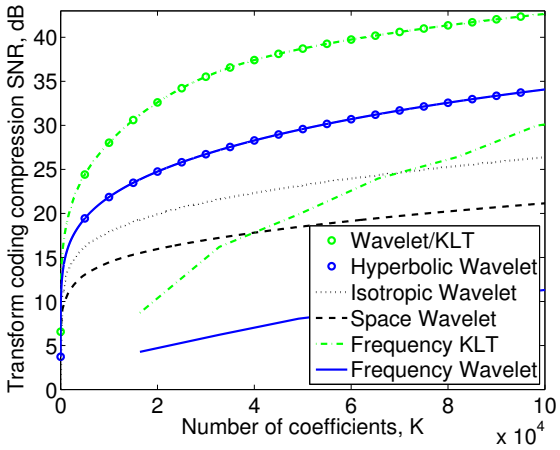

Fig. 3. *Empirical performance of Kronecker product sparsity for hyperspectral imaging. The figure shows the compression SNR as a function of the number of coefficients K. Kronecker products performs better than either component basis independently.*

Daubechies-8 wavelet was used. For each one of these transforms, we measured the signal-to-noise ratio (SNR) when the various representations are thresholded to K terms for varying values of K. As we see in Fig. 3, the Kronecker (hyperbolic wavelet) transform provides the best compression of the signal, outperforming the partial transforms in terms of SNR. However, the decay rate for the normalized error of the Kronecker (hyperbolic wavelet) transform is only slightly better than the minimum rate of decay among the partial (spatial and frequency) wavelet transforms [24]. Our analysis indicates that this result is due to the difference between the degrees of smoothness among the various signal dimensions.

2) KCS: We compare the performance of KCS to CS using the 2-D basis $\mathbf{W}_2$ to yield compressible coefficient vectors for individual spectral band images. In our experiments we take CS measurements using the subsampled permuted Hadamard transform of [4] on each spectral band image with a matrix Φ_2. As a baseline, we also take *global* CS measurements that multiplex all the voxels of the datacube; such measurements result in a fully dense measurement matrix Φ and therefore are difficult to obtain in real-world applications. We operate with two "flattened" datacubes of sizes $128 \times 128 \times 16$ 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 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

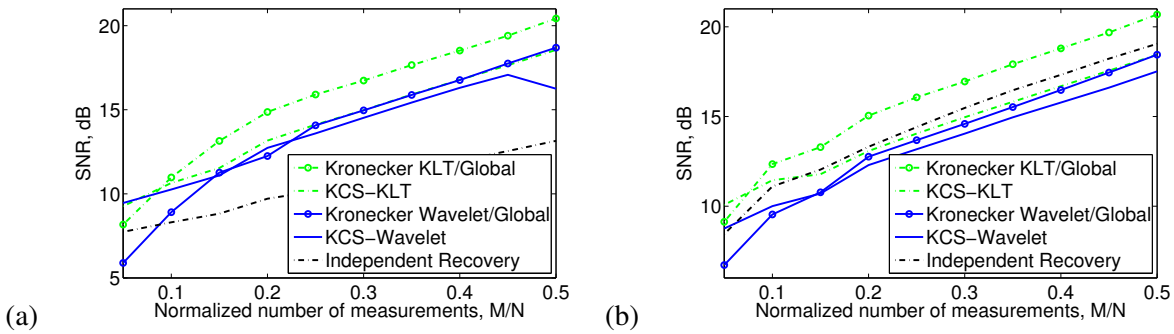

Fig. 4. *Empirical performance of KCS and standard CS for hyperspectral imaging for datacubes of sizes (a)* $N = 128 \times 128 \times 16$ *and (b)* $N = 128 \times 128 \times 64$ *voxels. Recovery using the Kronecker product sparsifying bases outperforms independent recovery in (a). Additionally, there 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 in (b), the Kronecker sparsity and Kronecker measurement bases become increasingly coherent, hampering the performance of joint recovery techniques.*

perform joint recovery. We test two different Kronecker product sparsifying bases: *KCS-Wavelet* uses a Kronecker product 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 for reference.

In Fig. 4, we see an improvement in recovery using KCS 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 global measurements [24]. In other words, the performance of the Kronecker-based approaches, which involve the same CS measurement matrix and spatial transform as independent recovery, fails to improve in a similar fashion as the number of spectral bands increases. We conjecture that this 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. Approximate average execution times were measured as follows: independent recovery of all spectral bands took 9 minutes, while KCS and recovery from global measurements using

the hyperbolic wavelet basis for sparsity took 25 and 31 minutes, respectively; KCS and recovery from global measurements using the wavelet/KLT Kronecker product basis for sparsity took 35 and 36 minutes, respectively. These increases are much more modest than what would be anticipated by the theoretical discussion in Section III-D.

3) Single-pixel hyperspectral camera: Our next experiment uses real-world data obtained from the SPHC [19] described in Section I-D using the independent measurements of (3). Figure 1 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 obtained. In Fig. 1 (left), each spectral band was recovered independently. In Fig. 1 (right), the spectral bands were recovered jointly with KCS using the measurement structure of (3) and a hyperbolic wavelet basis. The results show a considerable quality improvement over independent recovery, particularly for spectral frames with low signal power.

C. Video data

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 produce frames of size 128×128 pixels. We select 128 frames to obtain a 3-D signal of total size $N = 2^{21}$ voxels. We process the signal through three different transforms: (i) (*Space*) applies the 2-D wavelet basis $\mathbf{W}_2$ along the spatial dimensions of the data, with no compression on the temporal dimension; (ii) (*Isotropic*) applies the standard isotropic 3D wavelet basis $\mathbf{W}_3$ on the entire video sequence; (iii) (*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. 5 and closely resemble those obtained for hyperspectral data. In particular, the Kronecker product outperforms isotropic wavelets due to the difference in smoothness between the temporal and spatial dimensions.

2) KCS: We compare the performance of KCS to that of CS using the low-dimensional basis $\mathbf{W}_2$ to yield compressible 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 [4]. For KCS we use a single Kronecker product measurement matrix as shown in (3), while

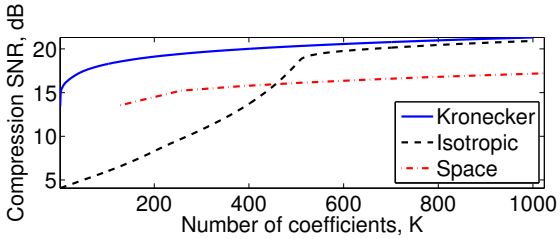

Fig. 5. *Empirical performance of Kronecker product sparsifying basis for transform coding of the* Foreman *video sequence,* $N = 128 \times 128 \times 128 = 2^{21}$ *voxels. The Kronecker product performs better in distortion than the alternative bases.*

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 multiple all the pixels of the video sequence, as a baseline.

Figure 6 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 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) [18]. In MS-BCS disjoint blocks of each video frame are independently 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 released while this paper was under review [18]. 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 [18], 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

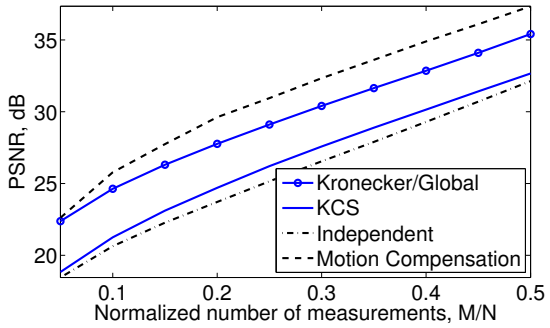

Fig. 6. *Empirical performance of KCS on the* Foreman *video sequence. While KCS does not perform as well as CS using global measurements, it shows an improvement over independent recovery of each frame in the video sequence using the same measurements. The motion compensation-aided approach outperforms the generic approaches.*

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.

Approximate average execution time were as follows: independent recovery of all video frames took 13 seconds, while KCS and recovery from global measurements using the hyperbolic wavelet basis for sparsity took 104 and 220 minutes, respectively; MC-BCS recovery took 34 minutes. These results agree with the discussion in Section III-D, since the computational time of KCS is increased by a factor of about $128 \times 3 = 384$ over independent recovery.

VI. Related work

Prior work for CS of multidimensional signals focuses on the example applications given in this paper – hyperspectral imaging [8, 38–40] and video acquisition [7, 9–12, 14, 17, 18] – with limited additional work in sensor networks [6, 15, 16] and confocal microscopy [13]. Some formulations employ measurement schemes that act on a partition of the data $\{\mathbf{x}_1, \ldots, \mathbf{x}_J\}$, such as frames of a video sequence [6–14, 17, 18]. For those cases, individual measurement vectors $\{\mathbf{y}_1, \ldots, \mathbf{y}_J\}$ are obtained using a set of matrices $\{\Phi_1, \ldots, \Phi_J\}$, resulting in the measurement matrix structure of (3). While global measurements that multiplex the entire set of data have been proposed [5, 7, 13, 17], practical architectures that provide such measurements are rare [5]. Similarly, partitioned measurements have been proposed for CS of low-dimensional signals for computational purposes [25, 26, 41]. Below we contrast the signal model and

algorithms used in these approaches with those used in KCS.

Several frameworks have been proposed to sparsify multidimensional signals. The most significant class of structures link signal partitions 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 $\mathbf{x} = (\mathbf{I} \otimes \Psi)\mathbf{P}\Theta$. The vector Θ encodes the correlations among partitions and has length lower than the sum of the sparsities of the partitions [6, 10, 12, 15, 16]. Such matrices are very limited in the kinds of structures they can represent. Kronecker product matrices can represent a variety of multidimensional structures by using sparse representations for each of the signal dimensions.

Kronecker product matrices have been proposed for use as sparsifying bases in CS for certain spatiotemporal signals [9, 11, 17]. In other cases, specialized compression bases are combined with custom recovery algorithms [8, 12, 14, 17, 18]; a prime example is motion compensation for video sequences [12, 18]. While such 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 multidimensional CS, such as isotropic wavelets, suffice only for very specific signal classes that do feature similar degrees of smoothness in each dimension [7, 13]; KCS using hyperbolic wavelets is suitable for signals with different degrees of smoothness in each of their dimensions.

In transform coding, hyperbolic wavelet bases have been proposed for compression of hyperspectral datacubes and video sequences [17, 40, 42]; 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 [17, 38, 39]; 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 [25–27]. Our study of mutual coherence matches that independently obtained in [25, 26], while [27] 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.

VII. Conclusions and Further Work

In this paper we have developed the concept of Kronecker compressive sensing (KCS) 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 hyberpolic 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 [43, 44]. 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 have focused on hyperspectral imaging and video acquisition, there exist other interesting applications where KCS is relevant. In sensor networks and arrays, sparsity-based distributed localization [45–47] 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 [9], 3-D magnetic resonance imaging (MRI) [9], and functional MRI.

ACKNOWLEDGEMENTS

Thanks to Ron DeVore for many valuable discussions and for pointing us to [30, 31] and to Justin Romberg for pointing us to the remark in [23]. Special thanks to Kevin Kelly, Ting Sun and Dharmpal Takhar for providing experimental data for the single-pixel hyperspectral camera. And thanks to Don Johnson, Chinmay Hegde, Jason Laska, Shriram Sarvotham, and Michael Wakin for helpful comments. Finally, we thank the anonymous reviewers for many helpful comments.

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] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

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

[12] 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.

[13] 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.

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

[15] 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.

[16] 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.

[17] 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.

[18] 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.

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

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

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

[22] 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.

[23] 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.

[24] M. F. Duarte and R. G. Baraniuk, "Kronecker product matrices for compressive sensing," Tech. Rep. TREE-1105, Rice University Department of Electrical and Computer Engineering, Houston, TX, Mar. 2011.

[25] 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.

[26] 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.

[27] 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.

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

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

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

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

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

[33] A. Cohen, "A primer on Besov spaces," Sep. 2009, http://cnx.org/content/col10679.

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

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

[36] 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.

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

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

[39] 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.

[40] 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.

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

[42] 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.

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

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

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

[46] 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.

[47] 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.