COMPRESSIVE SENSING WITH APPLICATIONS TO HYPERSPECTRAL IMAGE PROCESSING

By

Bernard Henry Lampe

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, Baltimore County, in partial fulfillment
of the requirements for the degree of
Doctorate of Philosophy
2019

Acknowledgements

This work would not have been possible without the mentorship of my advisor Dr. Chein-I Chang, department of electrical engineering at The University of Maryland Baltimore County. His advice and insights have contributed greatly to the direction of my research. Also the support of my peers Adam Bekit and CJ Della Porta has been invaluable in this endeavor. In addition, I am grateful for the feedback and research ideas given by the dissertation committee members Dr. Kalpakis, Dr. Bradley, Dr. Hu, Dr. Safavi and Dr. Zhu.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BP	Basis Pursuit
BS	Band Selection
BSS	Band Subset Selection
BT	Band Tensor
BT-SVD	Band Tensor – Singular Value Decomposition
BV	Band Vector
CEM	Constrained Energy Minimization Detector
CS	Compressive Sensing
CSBD	Compressively Sensed Band Domain
CSBR	Compressive Sensing Band Rate
CSD	Compressively Sensed Domain
CSSD	Compressively Sensed Sample Domain
CSSR	Compressive Sensing Sampling Rate
DR	Data Reduction
EO	Electro Optical
GSD	Ground Sample Distance
HSI	Hyperspectral Image
IHT	Iterative Hard Thresholding
LCMV	Linearly Constrained Minimum Variance Detector
ODS	Original Data Space
OMP	Orthogonal Match Pursuit
OSP	Orthogonal Subspace Projection
RCP	Restricted Conformal Property
REP	Restricted Entropy Property
RIP	Restricted Isometric Property
RK	Recursive Kurtosis
RSP	Restricted Spectrum Property

RXD	Reed Xu Anomaly Detector
SBS	Selected Band Subset
SCS	Sequential Compressive Sensing
SMMBS	Simultaneous Multi-Band Selection
SNR	Signal-to-Noise Ratio
SQMBS	Sequential Multi-Band Selection
SWaP	Size, Weight and Power

List of Variables

Symbol	Nomenclature
$\mathbf{I} \in \Re^{n_1 x n_2 x L}$	Hyperspectral Image
$\mathbf{r}_i \in \Re^{Lx1}$	Hyperspectral pixel vector $1 \leq i \leq N$
L	Number of spectral bands
$\mathbf{b}_l \in \Re^{Nx1}$	Hyperspectral band vector $1 \leq l \leq L$
N	Number of hyperspectral pixels in an image
$\mathbf{B}_l \in \Re^{n_1 x n_2}$	Hyperspectral band tensor/band image $1 \leq l \leq L$
n_1	The number of rows in a band tensor/band image
n_2	The number of columns in a band tensor/band image
$\boldsymbol{\Psi}$	Basis matrix for sparse transform, DCT Matrix
$\boldsymbol{\psi}$	Single basis vector for sparse transform
k_L	Sparsity of a pixel vector
k_N	Sparsity of a band vector
k_{n_1}	Sparsity of the band column vectors
k_{n_2}	Sparsity of the band row vectors
$\boldsymbol{\Phi}_{k_L} \in \Re^{m_L x L}$	Spectral compressive sampling matrix
$\boldsymbol{\Phi}_{k_N} \in \Re^{m_N x N}$	Spatial compressive sampling matrix for band vector
$\boldsymbol{\Phi}_{k_{n_1}} \in \Re^{m_{n_1} x n_1}$	Spatial compressive sampling matrix for band columns
$\boldsymbol{\Phi}_{k_{n_2}} \in \Re^{m_{n_2} x n_2}$	Spatial compressive sampling matrix for band rows
$\boldsymbol{\phi}_i$	Compressive sampling vector $1 \leq i \leq m$
$\boldsymbol{\varphi}_j$	Column vector of Sampling Matrix $\boldsymbol{\Phi}_k$ $1 \leq j \leq n$
$\mu(\boldsymbol{\Phi}_k, \boldsymbol{\Psi})$	Incoherence measure
$\mathbf{y}_i$	Sensed pixel vector $1 \leq i \leq N$
$\mathbf{y}_l$	Sensed band vector $1 \leq l \leq L$
$\mathbf{Y}_l$	Sensed band tensor $1 \leq l \leq L$
m_L	Optimal number of compressive samples in the spectrum
m_N	Optimal number of compressive samples in band vector

m_{n_1}	Optimal number of compressive samples in band columns
m_{n_2}	Optimal number of compressive samples in band rows
δ_k	Restricted Isometric Constant
ρ	Angle between two vectors in uncompressed domain
α	Angle between two vectors in compressed domain
c	Application constant for optimal sampling size
$\mathbf{X} \in \Re^{L \times N}$	Linearized hyperspectral image in time domain
$\mathbf{Y} \in \Re^{mxN}$	Linearized hyperspectral image in compressed domain
$\mathbf{R}_{LxL} \in \Re^{L \times L}$	Sample correlation matrix in uncompressed domain
$\mathbf{K}_{LxL} \in \Re^{L \times L}$	Sample covariance matrix in uncompressed domain
$\mathbf{u} \in \Re^{L}$	Mean vector
$\mathbf{R}_{\mathbf{\Phi}} \in \Re^{L \times L}$	Sample correlation matrix in compressed domain
$\mathbf{K}_{\mathbf{\Phi}} \in \Re^{L \times L}$	Sample covariance matrix in compressed domain
$\mathbf{U} \in \Re^{Lxp}$	Matrix of selected bands
$\mathbf{P}_{\mathbf{U}}^{\perp} \in \Re^{LxL}$	Projector onto the orthogonal subspace $\mathbf{U}^{\perp}$
$\mathbf{P}_{\mathbf{\Phi U}}^{\perp} \in \Re^{mxm}$	Projector onto the compressed orthogonal subspace $(\mathbf{\Phi U})^{\perp}$
$\delta_{RXD}(\mathbf{r})$	RXD detector function on pixel vector $\mathbf{r}$
$\delta_{CS-RXD}(\mathbf{r})$	Compressed RXD detector function on pixel vector $\mathbf{r}$
$\delta_{LCMV}(\mathbf{r})$	LCMV detector function on pixel vector $\mathbf{r}$
$\delta_{CS-LCMV}(\mathbf{r})$	Compressed LCMV detector function on pixel vector $\mathbf{r}$
$\mathbf{d} \in \Re^{Lx1}$	Desired target signature for detection algorithm
λ	Eigenvalue
$\mathbf{v}$	Eigenvector
$\mathbf{p}, \mathbf{q}$	Spectral probability distribution for hyperspectral pixels

Chapter 1: Introduction

The focus of this dissertation is to utilize and extend compressive sensing (CS) concepts to reduce the computational and memory space requirements of algorithms developed for hyperspectral imaging (HSI) data exploitation [1]–[3]. For a single HSI image, hyperspectral sensors produce hundreds of two dimensional, band images across a narrow and contiguous range of discrete frequencies [3]. The fine spectral resolution enables data processing algorithms to extract significant information which is lost in electro optical sensing (EOS). Enhanced anomaly detection, target detection, material identification, material quantization, and classification all become possible as a result of the increased spectral resolution [4]–[6]. The high spectral resolution of hyperspectral sensors produces large data sets which are hundreds of megabytes per image. Typical images have dimensionality on the order of tens of thousands of pixels and each pixel has hundreds of bands encoded using a dynamic range of 2^{16} bits [7]–[9]. Due to such large amounts of data, it is often not feasible to implement many HSI data processing algorithms on board the collection systems when facing the size weight and power (SWaP) requirements [10]. Therefore, images are typically compressed and cached until further data processing can be off-loaded at the ground station [11]. Also, off-loading the data can be laborious given the limited transmission bandwidth between collection systems and ground stations. While HSI have high bandwidth requirements for collection, storage and transmission, they contain far less information entropy [12]. Due to the high spectral resolution, the spectral samples exhibit high correlation in the sense that knowing one sample gives prediction power over the adjacent samples. This implies that HSI are highly

compressible and data reduction (DR) can be applied before storage or transmission [3]. Common DR methods are reviewed in Chang 2013 [3]. However, this does not solve the costly system requirements of high bandwidth during collection, nor does it address the fact that images must be uncompressed prior to further processing. Ideally, these images could be acquired using a compressive sensor which would avoid high bandwidth during acquisition. Further, it would be advantageous to process these images in compressed form directly and produce comparable results to that using the uncompressed data.

The goal of this dissertation is to take advantage of the fundamental property of CS known as the restricted isometry property (RIP) in order to process data in the compressively sensed sample domain (CSSD) and realize practical improvements in algorithm runtime and memory efficiency with minimal to no loss of accuracy [13]. RIP is a length preserving property of signal vectors when collected compressively. The specific HSI processing algorithms to be addressed in this dissertation are anomaly detection, target detection and band selection [4], [14]. Underpinning each of these algorithms are pixel and band discrimination as well as similarity measures which can be re-derived using RIP. Using these modified measures, algorithm performance can be preserved while processing far less bands and significantly fewer pixel data samples.

The chapters begin with a review of fundamental CS principles followed by Chapter 2 which details the application of CS to HSI images in four variations which include one spectral and three spatial sensing modes. In addition, a new method of estimating the compressive sensing sampling rate (CSSR) for HSI is presented. Chapter 3 utilizes RIP to derive HSI specific properties which are preserved in the CSSD including sample correlations/covariance, orthogonal subspace projection (OSP), and subspace

volumes of the band subsets. Chapter 4 details new CS principles which extend the length preservation property of RIP into new HSI properties, information entropy and spectral information divergence. The following chapters, Chapter 5 and Chapter 6, take advantage of these properties and apply them to HSI data processing algorithms. The first set of algorithms are designed for target detection include the Reed-Xiaoli anomaly detector (RXD) [15] and Linearly Constrained Minimum Variance filter (LCMV) [16]. The key insight into these algorithms is to exploit how the inter-band correlation and covariance matrices can be derived using data in CSSD. Developing these algorithms in CSSD allows significant reduction in computational complexity and memory space requirements while maintaining detection accuracy. A second application is band selection (BS) using orthogonal subspace projection (OSP) and sample subspace volume metrics [17]–[19] where the preservation of OSP and volume under a CS transform is shown to lead to comparable BS results using the data in CSSD with considerable savings in runtime and memory storage.

Compressive Sensing Review

The Shannon-Nyquist rate sets an efficient lower bound on the bandwidth required when using uniform sampling to discretely encode a continuous signal [20]. A sampling rate f_s at twice the highest frequency component f_{max} of the acquired signal, $f_s \geq 2f_{max}$ will ensure that no information is lost for any continuous signal [21]. Recently, a new sampling method for the class of sparse signals was introduced by Emmanuel Candes, Justin Romberg, Terence Tao and David Donoho known as compressive sensing (CS) [22]–[24]. This new acquisition paradigm samples sparse, finite, discrete or continuous signals incoherently via non-adaptive linear projections at a much lower rate given

knowledge of the maximum sparsity of the signal [25]. Using CS implies a finite and sparsely populated signal subspace. The advantage of CS is the ability to under-sample the sparse signals with minimal to no information loss. Sparsity is a necessary condition because the incoherent under-sampling cannot accommodate information across the entire signal subspace simultaneously. Global incoherent projections are required to ensure collection is unconstrained, signal independent, and no prior information is required about where the signals are located in the signal sub-space [26].

The sparsity of a general discrete, finite signal vector $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \Re^{nx1}$ is conceptually measured as the cardinality of the non-zero elements $k = \|\mathbf{\Psi x}\|_0$ where $\mathbf{\Psi}$ is a sparsity transform [27]. This measure is limited to idealized signals because real-world signals are corrupted by background interferers and noise. Noisy signals can be viewed as sparse signals embedded in an unwanted background typically decaying according to a power law [28]. Therefore, a number of other surrogate measures of sparsity have been proposed to determine the sparsity of an underlying signal. The most obvious alternative measure of sparsity utilizes thresholding $k = |\{x_i \mid x_i > t\}|$ for some application-oriented selected threshold t. This can be effective for additive noise models where the sparse components are linearly separable from the unwanted components, but is not appropriate for sparse signals with noise components significantly aliasing the primary signal. More advanced noise removal algorithms such as principal component analysis (PCA) [29], maximum noise fraction (MNF) [30] or independent component analysis (ICA) [31] are often introduced as a post collection or pre-processing step to deal with background and noise. These transforms project the signals into signal representations with different base which mixes the spectral and spatial information and can be computationally prohibitive

on collection systems for larger images. Other sparsity estimation methods measure relative sparsity of signals using the form $k = \|\mathbf{\Psi x}\|_1 \,/\, \|\mathbf{\Psi x}\|_2$. [32]–[34]. These measures can be used to compare sparsity between signals, but reliably measuring the true sparsity of a signal remains an open problem [33]. The work presented in this dissertation does not require the sparsity of the HSI to be estimated directly, but considers the components of the HSI to be sparse and develops methods to estimate the incoherent sampling rate by analysis of the incoming samples.

The global incoherent sampling of CS is done by repeated linear projections using the inner product between the signal vector and normalized incoherent sampling vectors $y_i = \langle \mathbf{x} \mid \boldsymbol{\phi}_i \rangle$ for $1 < i < m$ where $m \ll n$ is the number of compressive samples and m/n is the compressive sensing sampling rate (CSSR) [35]. To perform mathematical analysis, the vectors are arranged as row vectors to form a sampling matrix $\mathbf{\Phi}_k = [\boldsymbol{\phi}_1^T, \boldsymbol{\phi}_2^T, \ldots, \boldsymbol{\phi}_m^T] \in \Re^{mxn}$ and compressive sensing is performed compactly as $\mathbf{y} = \mathbf{\Phi}_k \mathbf{x}$. The subscript k confers the idea that any signal vector with sparsity from zero up to k can be encoded without information loss using the matrix $\mathbf{\Phi}_k$ and an explicit relationship exists between k and m derived from the incoherence.

Compressive Sensing Acquisition

The design of the sampling matrix $\mathbf{\Phi}_k$ is of critical importance in creating efficient CS systems. A sufficient condition of the sampling matrix $\mathbf{\Phi}_k$ is the restricted isometry principle (RIP) as stated in equation (1.1) [13]. The RIP condition was derived from the Johnson-Lindenstrauss (JL) lemma which proves that a relatively small number of points in a high dimension can be encoded in a low dimensional space while preserving distances between the points [36], [37]. Practically, this inequality tells us that the sampling must be

length preserving within a small tolerance. CS takes advantage of the JL lemma through the derivation of the RIP condition in equation (1.1) [38]. The error term in this equation δ_k is known as the restricted isometry constant (RIC) and $\delta_k \to 0$ as $m \to n$. The RIP condition is a sufficient condition to ensure high probability of precise recovery of the sensed signal.

$$(1-\delta_k)\|\mathbf{x}\|_2^2 \leq \|\mathbf{\Phi x}\|_2^2 \leq (1+\delta_k)\|\mathbf{x}\|_2^2 \tag{1.1}$$

A related and useful concept derived from the RIP condition is the restricted conformal property (RCP) [39]. This property, shown in equation (1.2), is a consequence of the RIP condition and shows that the angles between finite signals are preserved when encoded by the same sampling matrix [40]. Here ζ is the angle between any two signal vectors in the original data space (ODS) $\zeta = \angle(\mathbf{x}_i, \mathbf{x}_j)$ and $\rho = \angle(\mathbf{\Phi}_k\mathbf{x}_i, \mathbf{\Phi}_k\mathbf{x}_j)$ is the angle between the identical vector pair in the compressively sensed domain (CSD).

$$\frac{(1-\delta_k)}{(1+\delta_k)}\cos(\zeta) \leq \cos(\rho) \leq \frac{(1+\delta_k)}{(1-\delta_k)}\cos(\zeta) \tag{1.2}$$

Another useful property of CS is the preservation of eigenvalues from ODS to CSD [25]. As $m \to n$, the eigenvalues of the inner product and outer product spaces go to one as shown in equation (1.3). This means that as $\mathbf{\Phi}_k$ becomes closer to full rank $\|\mathbf{\Phi}_k^T\mathbf{\Phi}_k - \mathbf{I}\|_2 < \delta_k$ and $\|\mathbf{\Phi}_k\mathbf{\Phi}_k^T - (n/m)\mathbf{I}\|_2^2 < \delta_k$ are implied [25].

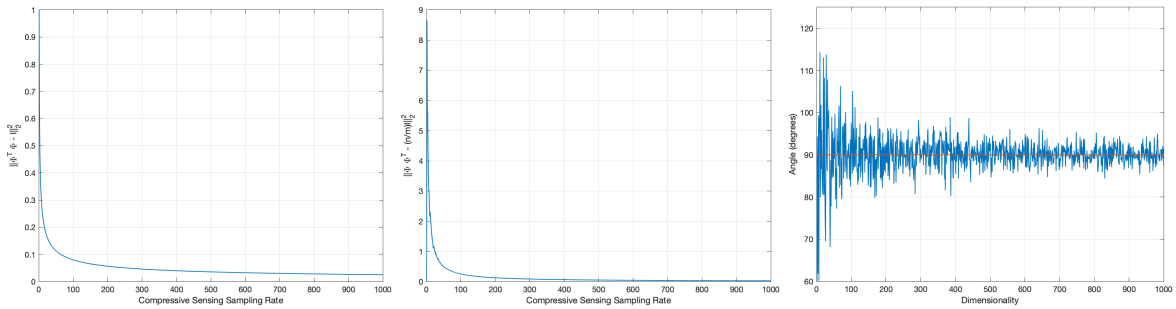
$$1-\delta_k \leq \lambda_{min}(\mathbf{\Phi}_k^T\mathbf{\Phi}_k) \leq \lambda_{max}(\mathbf{\Phi}_k^T\mathbf{\Phi}_k) \leq 1+\delta_k \tag{1.3}$$

It has been shown in Foucart and Rauhut book, page 141, to be a hard problem to construct $\mathbf{\Phi}_k$ deterministically while maintaining the RIP condition [25]. In order to maintain RIP, the entries of $\mathbf{\Phi}_k$ are drawn from a probability distribution creating sampling vectors which are orthogonal in expectation $\mathrm{E}\left[\mathbf{\phi}_i^T\mathbf{\phi}_j\right] = 0$ for $i \neq j$ and have a near certain

probability of incoherence to $\boldsymbol{\Psi}$. The probabilistic construction of the sampling matrix is a key insight to CS. Typically the columns of $\boldsymbol{\Phi}_k$ are normalized to unity explicitly or in expectation. This ensures that the new basis does not distort the vector lengths. Incoherence is measured using equation (1.4) where $\boldsymbol{\varphi}_i$ and $\boldsymbol{\psi}_j$ are column vectors of $\boldsymbol{\Phi}_k = [\boldsymbol{\varphi}_1, \boldsymbol{\varphi}_2, \dots, \boldsymbol{\varphi}_n]$ and $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \dots, \boldsymbol{\psi}_n]$ for $1 \leq i, j \leq n$.

$$\mu(\boldsymbol{\Phi}_k, \boldsymbol{\Psi}) = \max_{1 \leq i,j \leq n} |\langle \boldsymbol{\varphi}_i, \boldsymbol{\psi}_j \rangle| \text{ where } \mu = \left[\frac{1}{\sqrt{n}}, 1\right] \tag{1.4}$$

A simple experiment can be conducted to show that the inner product space approaches identity as $\boldsymbol{\Phi}_k^T \boldsymbol{\Phi}_k \to \mathbf{I} + \mathbf{E}$ with $\mathbf{E} \to \mathbf{0}$ as $m \to n$ and the outer product approaches a scaled identity matrix $\boldsymbol{\Phi}_k \boldsymbol{\Phi}_k^T \to (n/m)\mathbf{I} + \mathbf{E}$ with $\mathbf{E} \to \mathbf{0}$ as $m \to n$. Fig. 1 shows a plot of $\|\boldsymbol{\Phi}_k^T \boldsymbol{\Phi}_k - \mathbf{I}\|_2^2$ and $\|\boldsymbol{\Phi}_k \boldsymbol{\Phi}_k^T - (n/m)\mathbf{I}\|_2^2$ with respect to increasing m for $n = 1000$. Fig. 1(a) shows the exponential decay of the matrix inner product and Fig. 1(b) shows the exponential decay of the matrix outer product. Considering the entries of the inner product matrix $[\boldsymbol{\Phi}_k^T \boldsymbol{\Phi}_k]_{ij}$ for $1 \leq i, j \leq n$, when $i = j$ the inner product of $\boldsymbol{\varphi}_i^T \boldsymbol{\varphi}_j = 1$ due to the column normalization, and when $i \neq j$ the two random vectors rapidly become orthogonal as the dimensionality increases. In addition, consider the entries of the outer product matrix $[\boldsymbol{\Phi}_k \boldsymbol{\Phi}_k^T]_{ij}$ for $1 \leq i, j \leq m$, when $i = j$ the outer product of $\mathrm{E}[\boldsymbol{\phi}_i^T \boldsymbol{\phi}_j] = \mathrm{E}[\|\boldsymbol{\phi}_i\|_2^2] = \mu^2 + \sum_{j=1}^{n} \sigma_j^2 = \frac{n}{m}$ for $\sigma_j^2 = 1/m$ arising from the normalized columns, and when $i \neq j$ the value tends to zero as dimensionality increases. Also notice in the simple experiment, the magnitude does not converge exactly to zero. This residual is due to the probabilistic construction of $\boldsymbol{\Phi}_k$ and the expectation of orthogonality between random vectors. The error will tend to zero as $m \to \infty$, but is clearly not useful when under-



(a) (b) (c)

Fig. 1. Convergence of inner products to identity as CSSR increases

sampling $m < n$. In addition, Fig. 1(c) shows the angle of between two random vectors as the dimensionality increases from $1 \rightarrow n$. The angle approaches $90°$ as the dimensionality increases but some residual error exists. The residual error leads to approximate preservation of the signal lengths via RIP and fluctuations in the results when deriving HSI data processing algorithm using data in the CSSD.

Compressive Sensing Sampling Rate

The number of incoherent samples determines the efficiency of a CS system and the CSSR determines the quality of the sample reconstruction [41]. If $\mathbf{\Phi}_k$ is constructed using a Gaussian distribution, then it has been shown that the number of samples needed follows equation (1.5) where $c = 2$ is an application-dependent quantity [22]. Other distributions such as Bernoulli distribution conforming to RIP is also suitable for CS [42]. While Bernoulli distributions can be used for signal acquisition, this dissertation uses Gaussian distribution for the purpose for sensing through the entire dissertation. However, it does not preclude the use of other distributions to construct $\mathbf{\Phi}_k$.

$$m \geq ck \log(n/k) \tag{1.5}$$

Calculation of CSSR can be accomplished via a sparsity estimation and a direct application of (1.5). An alternative approach to direct estimation is using sequential CS (SCS) [41], [43], [44]. This approach seeks to estimate CSSR directly via a detection problem. The compressive samples are taken one at a time and analyzed to detect when enough samples have been observed. An obvious form of analysis originally proposed for SCS is the reconstruction of the signal. CS is a probabilistically consistent acquisition method which means that increasing the number of samples collected results in a reconstruction converging to the true signal in probability. Therefore, reconstruction can be performed in-line with sequential sampling. Sampling is terminated when the reconstructed signal converges to a consistent vector which is assumed to be the true signal. Despite that, this approach may be effective but is computationally expensive due to the reconstruction process which has a computational complexity on the order of $O(n^3)$ [45]. Introduced in Chapter 2 is a novel light-weight, and recursive analysis method used to improve sequential CS and estimate CSSR.

Compressive Sensing Reconstruction

The reconstruction of CS-encoded signals searches for an optimal $\mathbf{x}^* \approx \mathbf{x}$ by solving the non-linear convex optimization reconstruction problem of equation (1.6). This can be done using the basis pursuit (BP), the matching pursuit optimization (OMP) or the iterative hard thresholding (IHT) developed in [46]–[48].

$$\mathbf{x}^* = \arg\min_{\mathbf{\Psi x'}} \|\mathbf{\Psi x'}\|_1 \text{ subject to } \mathbf{y} = \mathbf{\Phi}_k \mathbf{x'} \tag{1.6}$$

Note that the sparse basis of the signal representation matrix $\mathbf{\Psi}$ is not required to be specified for CS encoding [23], [38]. CS only requires a sparse representation to exist so

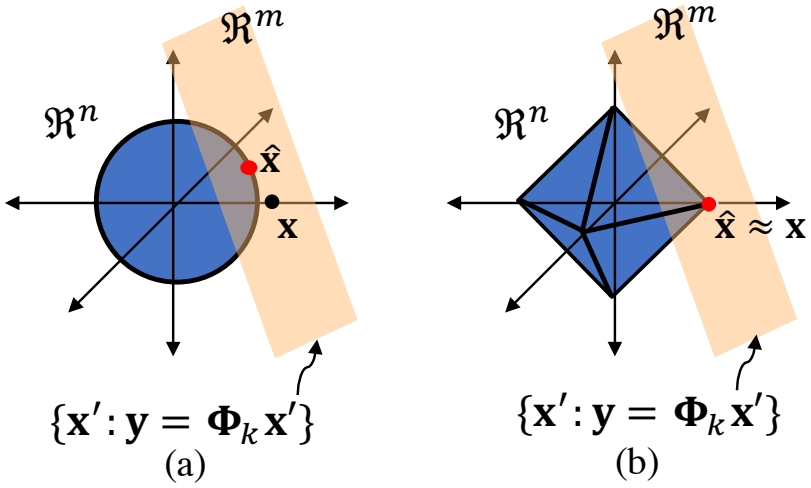


Fig. 2. The notational optimization surface using L_1 and L_2 vector norm

that the decoding can be done by optimizing over the sparse search space while constrained by the sampling matrix. Therefore, CS is a universal encoding scheme and the representation matrices are used for reconstruction where the fidelity of reconstruction is determined by the compact representation. The use of the L_1 norm to reconstruct sparse signals was a key discovery in CS. However, while convex, the L_1 is not differentiable. This leads to iterative numerical solutions using BP, OMP or IHT. A sparse solution to equation (1.6) will exist in the range space of the sampling matrix and simultaneously at a point of the n dimensional L_1 ball [49]. Sparse signals have a small L_0 norm meaning that most values of the vector are zero and the solution point will exist on the axis of the subspace created from the columns of the sampling matrix. Fig. 2 illustrates the location of the solution and why using L_1 results in the an optimal solution. In those diagrams the tan-colored plane is the space spanned by the columns of the m-dimensional sampling matrix $\mathbf{\Phi}_k$. The solution exists in the intersection of this subspace and the axis containing the L_1 measure. Using the L_2-ball results in a minimal energy solution which is not sparse

while using the L_1 will find the sparsest but not minimum energy solution on the point of the L_1-ball.

Conclusion

Chapter 1 introduced the data intensive problems of HSI processing and proposed the use of compressive sensing (CS) as a data reduction (DR) technique. Using CS has a distinct advantage over other DR algorithms in that it can be imposed at the time of collection by the sensor and processing data in the compressively sensed domain (CSD) can be done in such a way as to arrive at the results computed using the original data space (ODS). In addition, this chapter has reviewed the abstract CS principles required for applying CS to HSI. These principles will be exploited to develop HSI processing algorithms in the CSD in subsequent chapters.

Chapter 2: Compressive Sensing for Hyperspectral Imagery

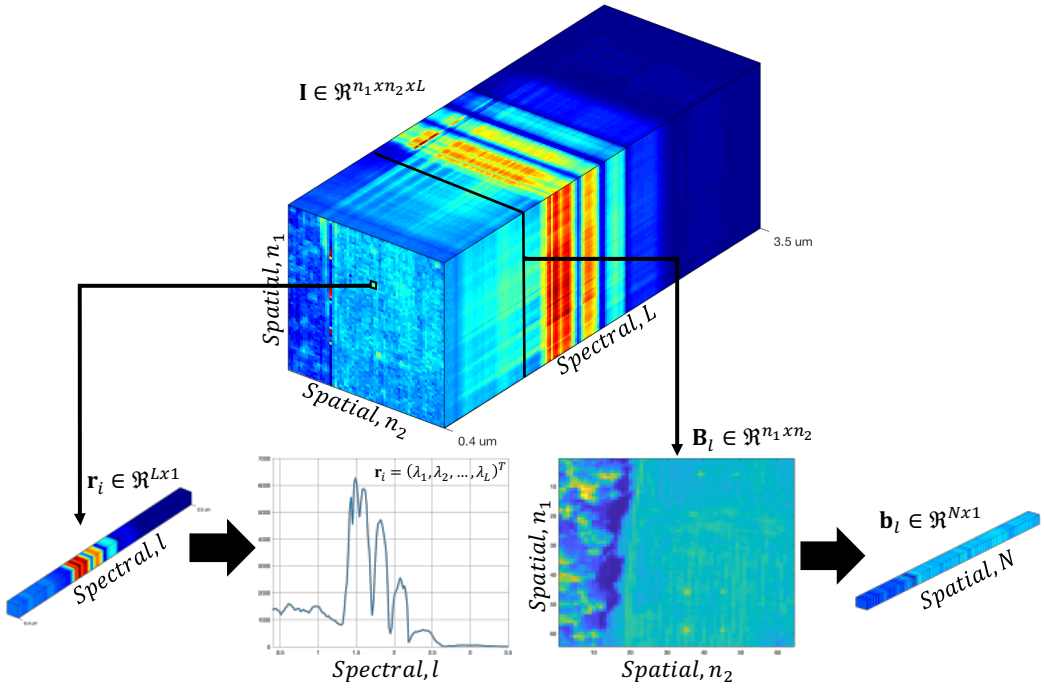


Fig. 3. Hyperspectral image and pixel vectors, band images and band vectors

Applying compressive sensing (CS) to hyperspectral images (HSI) can be done in a number of variations [50]–[52]. The HSI pixels can be considered sparse pixel vectors (PV) and the bands can be interpreted as sparse band tensors (BT) or sparse band vectors (BV) as depicted in Fig. 3. Let a hyperspectral image be denoted $\mathbf{I} \in \Re^{n_1 x n_2 x L}$ consisting of n_1 rows and n_2 columns with $N = n_1 x n_2$ total pixels and L spectral bands.

Spectral Compressive Sensing of Hyperspectral Imagery

A hyperspectral image $\mathbf{I}$ can be expressed as a set of sparse PVs containing the image spectra in a set $\{\mathbf{r}_i\}_{i=1}^{N}$ with $\mathbf{r}_i = (\lambda_1, \lambda_2, \ldots, \lambda_L)^T \in \Re^{Lx1}$ across a range of spectral wavelengths $\{\lambda_j\}_{j=1}^{L}$ [3]. The spectral mode of $\mathbf{I}$ is sensed using the CS sampling matrix

$$\mathbf{y}_i = \mathbf{\Phi}_{k_L} \mathbf{r}_i \tag{2.1}$$

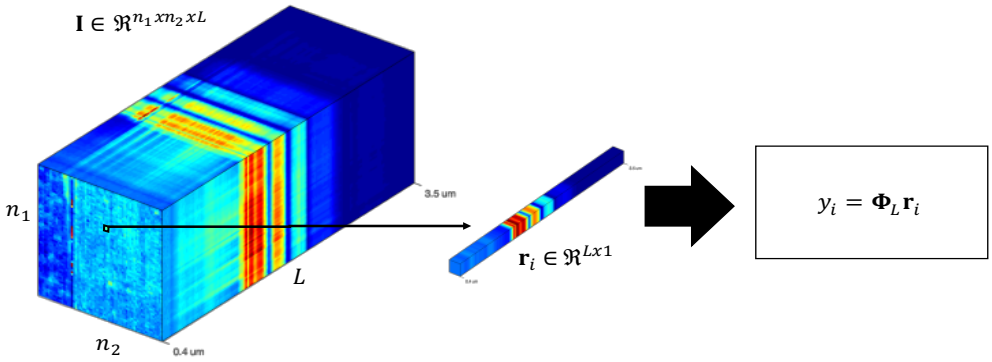


Fig. 4. Applying compressive sensing to hyperspectral pixel vectors

$\mathbf{\Phi}_{k_L} \in \Re^{m_L x L}$ where $k_L < m_L \ll L$ and k_L is the maximum sparsity of the PV set and m_L is compressively sensed band rate (CSBR) used for the spectrum. The application of CS to the spectral PVs is in equation (2.1) and further depicted in Fig. 4. Each PV is sampled using the same sampling matrix requiring k_L to be the maximum sparsity among all PVs.

Spatial Compressive Sensing of Hyperspectral Imagery

Another view of a hyperspectral image is to consider **I** as a set of L two-dimensional sparse band tensors (BT) as a set $\{\mathbf{B}_l\}_{l=1}^{L}$ so that $\mathbf{B}_l \in \Re^{n_1 x n_2}$ [52]. Such tensor modes can be sensed using two sampling matrices $\mathbf{\Phi}_{k_{n_1}} \in \Re^{m_{n_1} x n_1}$ and $\mathbf{\Phi}_{k_{n_2}} \in \Re^{m_{n_2} x n_2}$ where $k_{n_1} < m_{n_1} \ll n_1$ and $k_{n_2} < m_{n_2} \ll n_2$. Here k_{n_1} and k_{n_2} are the maximum sparsity of the rows and columns respectively of all the bands and m_{n_1} and m_{n_2} are the compressive sensing sampling rates (CSSR) for each corresponding mode. The model in (2.2) is used to compress the band tensor modes one at a time and is further depicted in Fig. 5. The two sampling matrices in (2.2) are used to compressively sense each of the bands and is constant for a particular HSI.

$$\mathbf{Y}_l = \mathbf{\Phi}_{k_{n_1}} \mathbf{B}_l \mathbf{\Phi}_{k_{n_2}}^T \tag{2.2}$$

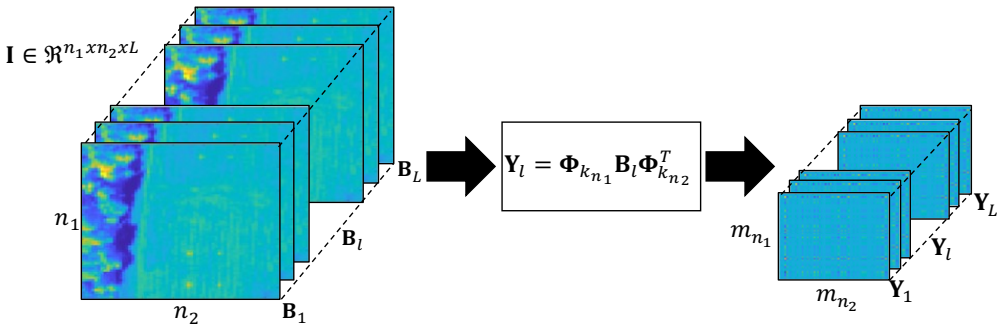


Fig. 5. Applying compressive sensing to hyperspectral band tensors

The BT model can be extended using the singular value decomposition (SVD) of $\mathbf{B}_l = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T$ where $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{n_1}]$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n_2}]$ are the singular vectors and $\mathbf{\Lambda} = diag(\sigma_1, \sigma_2, \dots \sigma_{\min\{n_1, n_2\}})$ are the singular values [53], [54]. Using this, the sampling matrices can be applied to the singular vectors using equation (2.3). This model will be used when referring to compressively sensed band tensors using SVD (BT-SVD) and is depicted in Fig. 6. Here again constant sampling matrices are used to compressively sense each band. Notice that this model implies that the eigenvalues of a square band tensor $\sqrt{\lambda_i} = \sigma_i$ for $1 \le i \le \min\{n_1, n_2\}$ are preserved and the CS transform can be accomplished by solely sensing the eigenvectors.

$$\mathbf{Y}_l = \sum_{i=1}^{\min\{n_1, n_2\}} \sigma_i \left(\mathbf{\Phi}_{k_{n_1}} \mathbf{u}_i\right)\left(\mathbf{\Phi}_{k_{n_2}} \mathbf{v}_i\right)^T \tag{2.3}$$

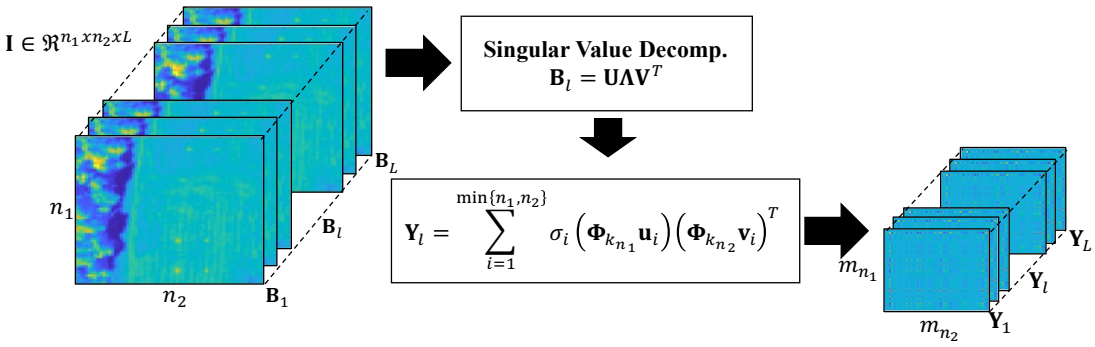


Fig. 6. Applying compressive sensing to SVD of hyperspectral band vectors

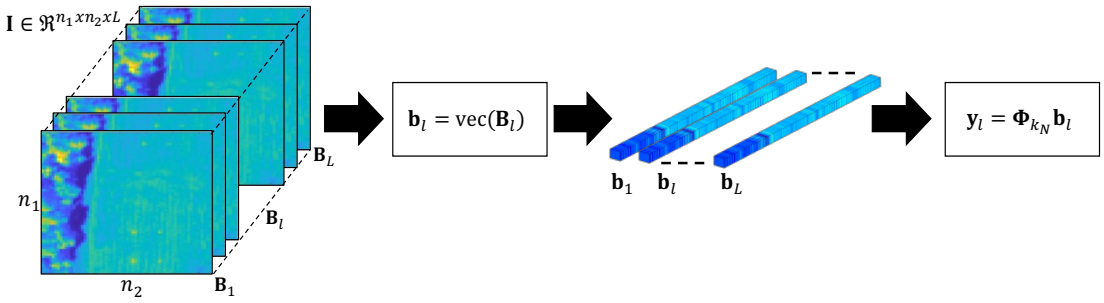


Fig. 7. Applying compressive sensing to hyperspectral band vectors

Finally, an HSI can be expressed as a set of L sparse BVs $\text{vec}(\mathbf{B}_l) = \mathbf{b}_l = (b_{l1}, b_{l2}, \dots, b_{lN})^T$ where $\mathbf{b}_l \in \Re^{Nx1}$ by concatenating each spatial row or column of pixels in a band to form the vector [55]. The band vectors are sensed using the sampling matrix $\mathbf{\Phi}_{k_N} \in \Re^{m_N x N}$ where $k_N < m_N \ll N$. Here k_N is the maximum sparsity of the band vectors and m_N is CSSR. CS applied to the BVs is depicted in Fig. 7 using (2.4).

$$\mathbf{Y}_l = \mathbf{\Phi}_{k_N} \text{vec}(\mathbf{B}_l) = \mathbf{\Phi}_{k_N} \mathbf{b}_l \tag{2.4}$$

Designing the incoherent sampling matrices $\mathbf{\Phi}_k$ for each HSI mode of PV, BT and BV is of critical importance for the performance of CS-derived algorithms applied to HSI [25]. The entries of the matrices $\mathbf{\Phi}_{k_L}$, $\mathbf{\Phi}_{k_N}$, $\mathbf{\Phi}_{k_{n_1}}$, and $\mathbf{\Phi}_{k_{n_2}}$ are obtained using random realizations from Gaussian distributions. The input signal dimensionality from the set $\{L, n_1, n_2\}$ is specified as the subscript to the sparsity k subscript. This explicitly denotes the HSI being sensed. The output number of compressive samples/observations is determined by CSBR and CSSR and is the most critical CS system design parameter.

Compressive Sensing Sampling Rate Estimation

The number of compressive samples created must be large enough to achieve high reconstruction accuracy but low enough to ensure efficient system resource utilization

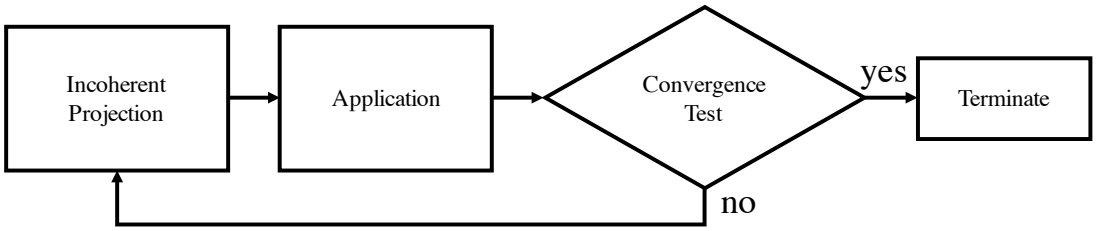


Fig. 8. Sequential compressive sensing (SCS) notional iteration

[41],[56]. A naïve approach to this problem is to explicitly set the signal sparsity as a limiting sensing system parameter. This is analogous to setting the number of pixels in an imaging system which dictates the upper bound on the spatial frequencies which can be acquired set by the Shannon-Nyquist rate [21]. The system created would be capable of acquiring signals up to the specified sparsity. This approach is not in keeping with CS principles in that many signals with lower sparsity acquired would be further compressible. An alternative approach is to compute the signal sparsity dynamically as it is being acquired.

An in-line method to finding the number of compressive samples is to collect incoherent observations to determine if enough samples have been obtained. This paradigm has been the core of sequential CS (SCS) algorithms [44]. SCS algorithms construct a closed loop of collecting samples, running applications and then testing for the resulting application output convergence. The signal reconstructions using basis pursuit (BP), orthogonal matching pursuit (OMP) or iterative hard thresholding (IHT) are the ubiquitous applications used in SCS research [41], [43]. Collection is terminated when the reconstructed signal converges to a consistent solution in a mean squared error sense. A notional flow of SCS is included in Fig. 8.

While reconstruction is a sufficient choice for SCS, BP, OMP, and IHT algorithms run on the order of $O(n^3)$ implying a high computational cost for SCS using reconstruction.

Recently, other algorithms have been proposed which take advantage of the sensed samples in the compressively sensed domain (CSD) [57], [58]. Detection, classification and feature selection have theoretical foundations using length preserving metrics for discrimination and can take advantage of RIP. Convergence of any of these application outputs would suffice in SCS and may reduce the computational burden. The HSI data processing algorithms designed for anomaly and target detection as well as band selection can be utilized in the SCS flow of Fig. 8 to determine an application-specified CSSR in subsequent chapters [3], [4]. While these applications may have lower computational complexity as compared to reconstruction, there is a novel and light-weight method to determine CSSR using SCS by taking advantage of recursive statistical moments of the collected compressive samples, which are analyzed to determine convergence and subsequently an appropriate CSSR. The following derivation will utilize the generalized signal $\mathbf{x} = (x_1, x_2, \dots, x_n)^T \in \Re^{nx1}$ and $\mathbf{\Phi}_k$ but can be extended to include the three modes of the BT set, BT-SVD set, PV set or BV set by using the corresponding sampling matrix.

As stated in Chapter 1, the CS encoding can be expressed as a set of inner products between the m rows of $\mathbf{\Phi}_k = [\boldsymbol{\phi}_1^T, \boldsymbol{\phi}_2^T, \dots, \boldsymbol{\phi}_m^T]$ where $\boldsymbol{\phi}_i^T = (\phi_{i1}, \phi_{i2}, \dots, \phi_{in})^T \in \Re^{1xn}$ then the signal vector $\mathbf{x}$ is encoded into m scalars which are $y_i = \langle \boldsymbol{\phi}_i, \mathbf{x} \rangle$ for $1 \leq i \leq m$ [59]. Furthermore, consider the input pixel vector as constant and the sampling vectors as random. Then each inner product can be expressed as a summation of Gaussian random variables ϕ_{ij}, multiplied by x_j for $1 \leq j \leq n$ as in equation (2.5) [60]. This implies all of the y_i are Gaussian random variables distributed with $\mu = 0$ and $\sigma^2 = \sum_{j=1}^{n} x_j^2$.

$$y_i = \sum_{j=1}^{n} \phi_{ij} x_j \quad \text{where } \phi_{ij} \sim \mathrm{N}(0,1) \tag{2.5}$$

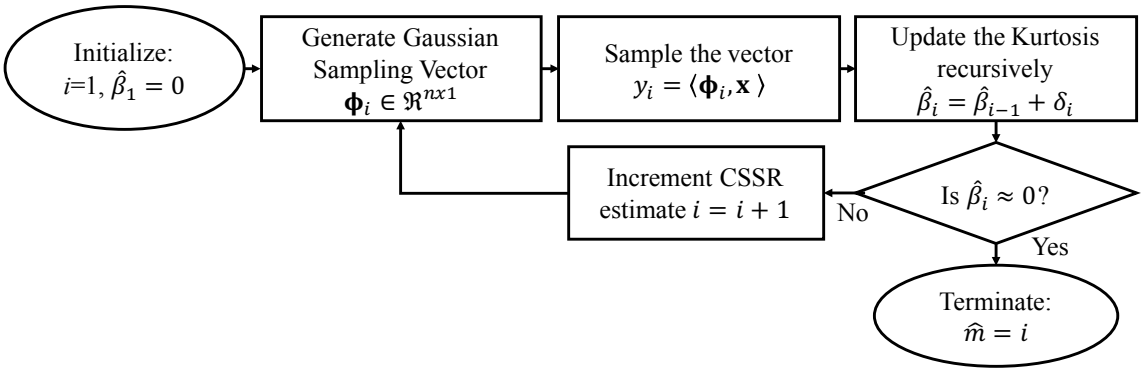


Fig. 9. Flow chart of recursive kurtosis CSSR estimation technique

With the $y_i \sim \mathrm{N}(0, \sum_{j=1}^{n} x_j^2)$, the information conveyed in the encoding process is preserved in the variance of the entries of y_i. In order to find the dimensionality m of $\mathbf{y} = (y_1, y_2, \ldots y_m)^T$, a question arises, "how many samples are needed to estimate the Gaussian distribution of $\mathrm{N}(0, \sum_{j=1}^{n} x_j^2)$?" This is a density estimation problem and has been well studied [61]. While there are many techniques used for this purpose, employing a method that is amenable to the sequential collection and computation is desired. This will minimize the work needed by the encoding process at each iteration. One approach is to estimate the sample variance of the samples $\hat{\sigma}^2(y_1, y_2, \ldots)$ as they are collected and then determine convergence to $\sigma^2 = \sum_{j=1}^{n} x_j^2$. However, this requires knowing the full uncompressed signal to compute $\hat{\sigma}^2$. Another simpler method to determine convergence is to estimate the kurtosis $\hat{\beta}(y_1, y_2, \ldots)$ of the collected samples and then determine when the estimate converges to zero or the value of 3 for excess kurtosis. Fig. 9 gives an overview of the estimation method. Here using $\hat{\beta}_i = \hat{\beta}_{i-1} + \delta_i$ is an on-line kurtosis estimate which relies on a recursive form of the first three moments mean, variance and skewness in equations (2.6), (2.7), (2.8), and (2.9) [62].

$$\mu_i = \mu_{i-1} + \frac{(y_i - \mu_{i-1})}{i} \tag{2.6}$$

$$\sigma_i^2 = \sigma_{i-1}^2 + \frac{(y_i - \mu_{i-1})^2(i-1)}{i} \tag{2.7}$$

$$\gamma_i = \gamma_{i-1} + \frac{(y_i - \mu_{i-1})^3(i-1)(i-2)}{i^2} - \frac{3(y_i - \mu_{i-1})\sigma_{i-1}^2}{i} \tag{2.8}$$

$$\delta_i = \frac{(y_i - \mu_{i-1})^4(i-1)(i^2-3i+3)}{i^3} + \frac{6(y_i - \mu_{i-1})^2\sigma_{i-1}^2}{i^2} - \frac{4(y_i - \mu_{i-1})\gamma_{i-1}}{i} \tag{2.9}$$

The stopping criteria of this algorithm can be implemented as a hard threshold from the excess kurtosis value 3 or using a change point detection method. The initial estimate of all the moments is set to zero at $i = 1$ such as $\mu_1 = 0$, $\sigma_1^2 = 0$, $\gamma_1 = 0$, $\beta_1 = 0$. Due to the variation of the kurtosis, a one dimensional Kalman filter of the recursive kurtosis (RK) is computed [63]. Using this, a convergence point can be selected from the plots.

A simple test was constructed to validate the above mentioned estimation method. A sparse signal was constructed by randomly assigning $k = 20$ values between 1 and 100 to the range [1,1000] and then transformed via the discrete cosine transform (DCT). The sparse signal has an optimal CSSR at $m^* = 2k \log(n/k) = 157$ using equation (1.5) and the raw signal is depicted in Fig. 10(a). Incoherent linear projections were then taken sequentially $y_i = \langle \boldsymbol{\phi}_i \mid \mathbf{x} \rangle$ from $1 \leq i \leq 500$ and the kurtosis was computed using equation (2.9) at each sample. To validate the convergence to the true signal, the reconstruction was performed using OMP at each iteration denoted $\hat{\mathbf{x}}_i$ [47]. The recursive kurtosis as well as the sum of squared error between the reconstruction and original signal was computed $\text{SSE}_i = \|\mathbf{x} - \hat{\mathbf{x}}_i\|_2^2$. The kurtosis β_i and the SSE as well as the 1D Kalman filter are plotted in Fig. 10(b). Salient features of this graph show the convergence of the SSE to zero and the convergence of the kurtosis to 3 just before the optimal CSSR of 157. This result lends credibility to the consistency of the CS reconstruction and effectiveness

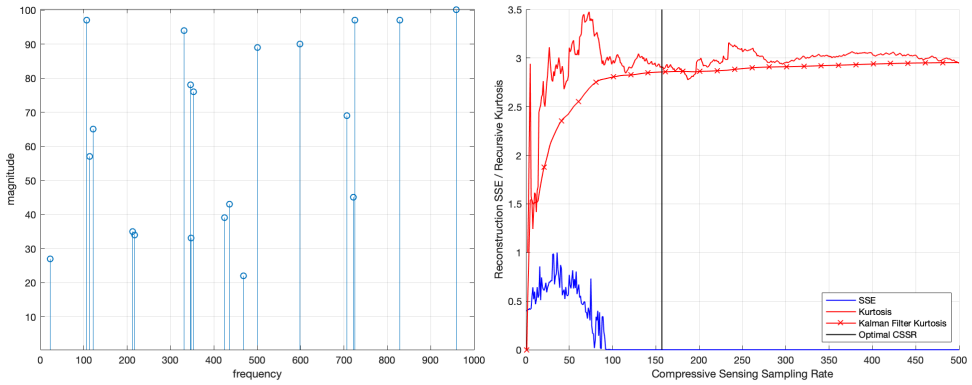


Fig. 10. Recursive kurtosis CSSR experiment signal (a) and result (b)

of using SCS, recursive kurtosis (SCS-RK) to estimate CSSR paradigm in Fig. 8. SCS and SCS-RK will be used in the detection and band selection algorithms discussed in subsequent chapters. The simple test from Fig. 10 was modified by adding additive Gaussian noise to the test signal shown in Fig. 11(a). the SNR is set to 4:1 and the result is shown in Fig. 11(b). This plot displays the convergence. However, the convergence is slower due to the addition of noise. This is paralleled in the reconstruction error which takes considerably longer to converge to zero with the addition of noise. This test implies that finding the convergence point of the signal may require advanced analysis of the curve. Here change-point detection or thresholding could be employed [64]. In this study, CSSR

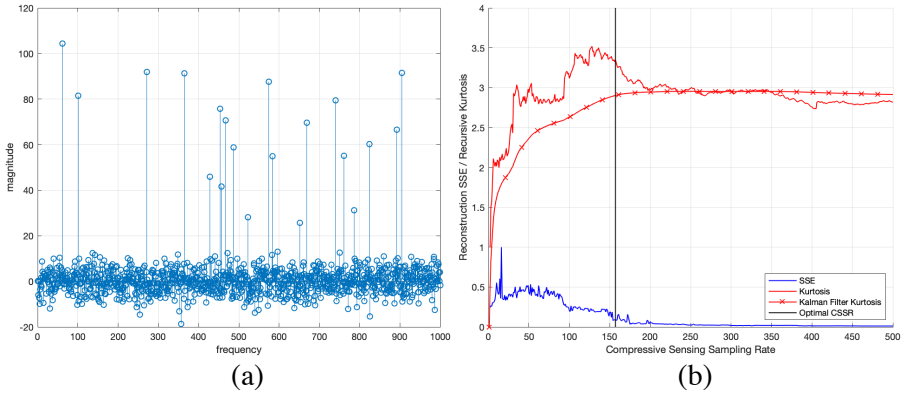


Fig. 11. Recursive kurtosis CSSR experiment with additive Gaussian noise 4:1 SNR

Is estimated by judging the order of magnitude from the SCS-RK curves to determine an appropriate CSSR.

The SCS-RK method of estimating CSBR and CSSR for HSI is cast in terms of a single sparse input signal. The application of CS to HSI requires a single sampling matrix to accommodate multiple PV, BV and BTs. Therefore, the kurtosis measure of each set of HSI vectors is collected and the mean is used to estimate a single CSSR for the set.

Conclusion

This chapter provides mathematical notations needed to precisely identify different portions of the hyperspectral image. Using these notations, CS is applied to the pixel vectors (PV), band vectors (BV) or band tensors (BT). In addition, estimating the compressive sensing band rate (CSBR) and compressive sensing sample rate (CSSR) is discussed and a novel method of computing the CSSR is put forth using sequential compressive sensing (SCS) augmented with recursive kurtosis (SCS-RK) statistics. The SCS and SCS-RK are used for estimating an optimal CSBR and CSSR for the anomaly and target detection as well as band selection applications. The CS models and notations developed in this chapter will be continuously used in the subsequent chapters.

Chapter 3. Preservation of Hyperspectral Properties

Many hyperspectral image (HSI) data processing algorithms rely on the discrimination or similarity measures between HSI pixels or bands [65]–[67]. For example, target detection algorithms utilize the sample correlation and covariance matrices to suppress background effects [2]. Sequential band selection algorithms utilize orthogonal subspace projection (OSP) to find distinct bands [55], and band subset selection algorithms take advantage of measuring the sample subspace volumes [18]. These three fundamental HSI measures are unaffected by compressive sensing (CS) applied to the samples in each band. Specifically, sensing the two-dimensional band tensors (BT) or one dimensional band vectors (BV) using CS does not significantly change the results. This chapter develops a new set of lemmas required to demonstrate the preservation of the sample correlation and covariance matrices as well as the OSP and subspace sample volumes when using data in the compressively sensed sample domain (CSSD). A preliminary study has been done in this area, but is lacking considerably in rigorous development and completeness [58].

Preservation of Sample Correlation and Covariance

The covariance and correlation matrices capture the inter-band statistics and are a useful estimate of the variation between spectral bands [68]. This variation is predominantly used to characterize the background and noise behaviors of HSI to enable anomaly and target detection. The following derivations show that the covariance and correlation matrices are minimally affected by CS applied to the spatial samples using BV, BT or BT-SVD models. Without loss of generality, assume the pixel vectors (PV) are mean centered, therefore the correlation and covariance matrices are equal $\mathbf{R} = \mathbf{K}$ and for clarity

of derivation, the normalization constants are omitted. Each of the following derivations use CS to sense the band information and utilize the compressed spatial samples to estimate the sample correlation and covariance matrices. There are three variations to compress the spatial bands of BV, BT, and BT-SVD.

Using the BV model of HSI from (2.4), let $\mathbf{X} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_L]$ and then compute the pixel sample correlation matrix as $\mathbf{R} = \mathbf{X}^T\mathbf{X} \in \Re^{LxL}$ [4]. The correlation matrix is computed using bands in CSSD such that $\mathbf{R}_{\mathbf{\Phi}} \to \mathbf{R}$ where $\mathbf{R}_{\mathbf{\Phi}}$ is the correlation computed using compressed samples. Sensing the band vectors in $\mathbf{X}$ yields equation (3.1).

$$\mathbf{X}_{\mathbf{\Phi}} = \left[\mathbf{\Phi}_{k_N}\mathbf{b}_1, \mathbf{\Phi}_{k_N}\mathbf{b}_2, \dots, \mathbf{\Phi}_{k_N}\mathbf{b}_L\right] = \mathbf{\Phi}_{k_N}\mathbf{X} \tag{3.1}$$

The sensed sample correlation matrix is $\mathbf{R}_{\mathbf{\Phi}} = \mathbf{X}_{\mathbf{\Phi}}^T\mathbf{X}_{\mathbf{\Phi}} \in \Re^{LxL}$ given by equation (3.2).

$$\mathbf{R}_{\mathbf{\Phi}} = \left(\mathbf{\Phi}_{k_N}\mathbf{X}\right)^T\mathbf{\Phi}_{k_N}\mathbf{X} = \mathbf{X}^T\mathbf{\Phi}_{k_N}^T\mathbf{\Phi}_{k_N}\mathbf{X} \tag{3.2}$$

Using $\mathbf{\Phi}_{k_N}^T\mathbf{\Phi}_{k_N} \to \mathbf{I} + \mathbf{E}$ the correlation computed using CS applied to BV is in (3.3).

$$\mathbf{R}_{\mathbf{\Phi}} = \mathbf{X}^T(\mathbf{I} + \mathbf{E})\mathbf{X} = \mathbf{R} + \mathbf{X}^T\mathbf{E}\mathbf{X} \approx \mathbf{R} \tag{3.3}$$

Also, the error $\mathbf{E} \to \mathbf{0}$ is appreciably small as $m_N \to N$ compared to $\mathbf{R}$. Therefore, the sample correlation matrix is approximately preserved when computed using bands in the CSSD and the BV model.

Next, consider the BT and BT-SVD models of band sensing in (2.2) or (2.3) [52], it is possible to show that indeed $\mathbf{R}_{\mathbf{\Phi}} \to \mathbf{R}$ as $m_{n_1} \to n_1$ and $m_{n_2} \to n_2$. Begin by representing the image using the BTs to transform $\mathbf{X}_{\mathbf{\Phi}}$ to the CSSD by equation (3.4).

$$\mathbf{X}_{\mathbf{\Phi}} = \left[\text{vec}(\mathbf{\Phi}_{k_{n_1}}\mathbf{B}_1\mathbf{\Phi}_{k_{n_2}}^T), \dots, \text{vec}(\mathbf{\Phi}_{k_{n_2}}\mathbf{B}_L\mathbf{\Phi}_{k_{n_2}}^T)\right] \tag{3.4}$$

If the sampling matrices are broken into m_{n_1} and m_{n_2}vectors $\mathbf{\Phi}_{k_{n_1}}^T = \left[\mathbf{\phi}_{11}, \mathbf{\phi}_{12}, \dots, \mathbf{\phi}_{1m_{n_1}}\right] \in \Re^{n_1 x m_{n_1}}$ and $\mathbf{\Phi}_{k_{n_2}}^T = \left[\mathbf{\phi}_{21}, \mathbf{\phi}_{22}, \dots, \mathbf{\phi}_{2m_{n_2}}\right] \in \Re^{n_2 x m_{n_2}}$, the ij^{th}

entry of $\mathbf{X_\Phi}$ for each band tensor $\mathbf{B}_l$ is written as in equation (3.5) where $1 \leq i \leq m_{n_1}$, $1 \leq j \leq m_{n_2}$ and $1 \leq l \leq L$.

$$\left[\mathbf{\Phi}_{k_{n_1}} \mathbf{B}_l \mathbf{\Phi}_{k_{n_2}}^T\right]_{ij} = \boldsymbol{\phi}_{1i}^T \mathbf{B}_l \boldsymbol{\phi}_{2j} \tag{3.5}$$

Subsequently $\mathbf{X_\Phi}$ can be further expressed using the sensed and vectorized BTs in matrix form as equation (3.6). Each column is a different band.

$$\mathbf{X_\Phi} = \begin{bmatrix} \boldsymbol{\phi}_{11}^T \mathbf{B}_1 \boldsymbol{\phi}_{21} & \cdots & \boldsymbol{\phi}_{11}^T \mathbf{B}_L \boldsymbol{\phi}_{21} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\phi}_{1m_{n_1}}^T \mathbf{B}_1 \boldsymbol{\phi}_{2m_{n_2}} & \cdots & \boldsymbol{\phi}_{1m_{n_1}}^T \mathbf{B}_L \boldsymbol{\phi}_{2m_{n_2}} \end{bmatrix} \tag{3.6}$$

Then the entries of $\mathbf{R_\Phi} = \mathbf{X_\Phi^T X_\Phi}$ are computed using the sum in equation (3.7).

$$[\mathbf{X_\Phi^T X_\Phi}]_{ij} = \left[\sum_{h=1}^{m_{n_1}} \sum_{k=1}^{m_{n_2}} (\boldsymbol{\phi}_{1h}^T \mathbf{B}_i \boldsymbol{\phi}_{2k})(\boldsymbol{\phi}_{1h}^T \mathbf{B}_j \boldsymbol{\phi}_{2k})\right]_{ij} \tag{3.7}$$

Rearranging the summation in (3.7) yields the following summation.

$$[\mathbf{X_\Phi^T X_\Phi}]_{ij} = \left[\sum_{h=1}^{m_{n_1}} \boldsymbol{\phi}_{1h}^T \mathbf{B}_i \left(\sum_{k=1}^{m_{n_2}} \boldsymbol{\phi}_{2k} \boldsymbol{\phi}_{2k}^T\right) \mathbf{B}_j^T \boldsymbol{\phi}_{1h}\right]_{ij} \tag{3.8}$$

Next, using the fact derived from RIP $\sum_{k=1}^{m_{n_2}} \boldsymbol{\phi}_{2k} \boldsymbol{\phi}_{2k}^T = \mathbf{\Phi}_{n_2} \mathbf{\Phi}_{n_2}^T = (\mathrm{n_2}/\mathrm{m}_{n_2})\mathbf{I} + \mathbf{E}$ and $\mathbf{E} \to \mathbf{0}$ as $m_{n_2} \to n_2$ resulting in equation (3.9).

$$[\mathbf{X_\Phi^T X_\Phi}]_{ij} = \left(\frac{n_2}{m_{n_2}}\right)\left[\sum_{h=1}^{m_{n_1}} \boldsymbol{\phi}_{1h}^T \mathbf{B}_i \mathbf{B}_j^T \boldsymbol{\phi}_{1h}\right]_{ij} + [\mathbf{E}]_{ij} \tag{3.9}$$

If the BTs are broken into column vectors $\mathbf{B}_i = \left[\mathbf{b}_{i1}, \mathbf{b}_{i2}, \ldots, \mathbf{b}_{in_2}\right]$ and $\mathbf{B}_j = \left[\mathbf{b}_{j1}, \mathbf{b}_{j2}, \ldots, \mathbf{b}_{jn_2}\right]$ equation (3.9) can be re-written as a summation of dot products as in equation (3.10).

$$[\mathbf{X}_{\mathbf{\Phi}}^T\mathbf{X}_{\mathbf{\Phi}}]_{ij} = \left(\frac{n_2}{m_{n_2}}\right)\left[\sum_{h=1}^{m_{n_1}}\sum_{k=1}^{n_2}(\boldsymbol{\phi}_{1h}^T\mathbf{b}_{ik})(\mathbf{b}_{jk}^T\boldsymbol{\phi}_{1h})\right]_{ij} + [\mathbf{E}]_{ij} \tag{3.10}$$

Transposing the dot products and then rearranging the summation results in the following.

$$[\mathbf{X}_{\mathbf{\Phi}}^T\mathbf{X}_{\mathbf{\Phi}}]_{ij} = \left(\frac{n_2}{m_{n_2}}\right)\left[\sum_{k=1}^{n_2}\mathbf{b}_{ik}^T\left(\sum_{h=1}^{m_{n_1}}\boldsymbol{\phi}_{1h}\boldsymbol{\phi}_{1h}^T\right)\mathbf{b}_{jk}\right]_{ij} + [\mathbf{E}]_{ij} \tag{3.11}$$

The equality of $\sum_{h=1}^{m_{n_1}}\boldsymbol{\phi}_{1h}\boldsymbol{\phi}_{1h}^T = \mathbf{\Phi}_{n_1}\mathbf{\Phi}_{n_1}^T = (n_1/m_{n_1})\mathbf{I} + \mathbf{E}$ and $\mathbf{E} \to \mathbf{0}$ as $m_{n_1} \to n_1$ is invoked again which arrives at equation (3.12) is approximately the scaled covariance $\mathbf{R}$.

$$[\mathbf{R}_{\mathbf{\Phi}}]_{ij} = \left(\frac{n_1}{m_{n_1}}\right)\left(\frac{n_2}{m_{n_2}}\right)\left[\sum_{k=1}^{n_2}\mathbf{b}_{ik}^T\,\mathbf{b}_{jk}\right]_{ij} + [\mathbf{E}]_{ij} \approx c\sigma_{ij}^2 + [\mathbf{E}]_{ij} \tag{3.12}$$

This derivation proves the preservation of the covariance and correlation matrices when computed using the compressed band tensors.

It should be noted that the BV mode requires a considerably larger sampling matrix $\mathbf{\Phi}_{k_N}$ when compared to the dimensionality of the two sampling matrices $\mathbf{\Phi}_{k_{n_1}}$ and $\mathbf{\Phi}_{k_{n_2}}$ using the BT models. The BV and BT models are related by considering $\mathbf{\Phi}_{k_N} \in \Re^{(m_{n_1}m_{n_2})x(n_1n_2)}$. Therefore, the BV sensing matrix can be seen as multiple concatenated copies of the BT sensing matrices. Given the lower dimensionality, the BT models have lower memory requirements than the BV method.

It has been shown that applying CS to the HSI bands using the BV and BT models has little effect on the accuracy of the correlation and covariance matrices. Therefore with adequate CSSR of each spatial mode the matrices of $\mathbf{R}$ and $\mathbf{K}$ can be replaced by $\mathbf{R}_{\mathbf{\Phi}}$ and $\mathbf{K}_{\mathbf{\Phi}}$ and the inverse of these matrices will hold as well $\mathbf{R}^{-1} \approx \mathbf{R}_{\mathbf{\Phi}}^{-1}$ and $\mathbf{K}^{-1} \approx \mathbf{K}_{\mathbf{\Phi}}^{-1}$.

Preservation of Sample Orthogonal Subspace Projection

Orthogonal subspace projection (OSP), developed for HSI applications by Harsanyi and Chang [3], [17], has been used for target detection and band selection. In this derivation, OSP is used to compute the projection distance of one band $\text{vec}(\mathbf{B}_l) = \mathbf{b}_l$ to a subset of bands $\{\text{vec}(\mathbf{B}_j)\}_{j=1}^{p} = \{\mathbf{b}_j\}_{j=1}^{p}$ which does not contain $\mathbf{b}_l$. The set of bands is arranged in a matrix according to (3.13). From the matrix containing the band subset

$$\mathbf{U} = \left[\text{vec}(\mathbf{B}_1) \mid \text{vec}(\mathbf{B}_2) \mid \ \ldots \ \mid \text{vec}\left(\mathbf{B}_p\right)\right] \tag{3.13}$$

OSP constructs an orthogonal projector $\text{P}_{\mathbf{U}}^{\perp} = \mathbf{I} - \mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T$ [69]. The RIP and RCP conditions in (1.1) and (1.2) preserve the length and angle between vectors of the projections which is maintained in CSSD. Compressing the bands such as $\mathbf{\Phi}_{k_N}\mathbf{b}_l$ and $\mathbf{\Phi}_{k_N}\mathbf{U}$ results in a projector in CSSD described in equation (3.14).

$$\text{P}_{\mathbf{\Phi}_{k_N}\mathbf{U}}\mathbf{\Phi}_{k_N}\mathbf{b}_l = \left(\mathbf{\Phi}_{k_N}\mathbf{U}\left((\mathbf{\Phi}_{k_N}\mathbf{U})^T\mathbf{\Phi}_{k_N}\mathbf{U}\right)^{-1}(\mathbf{\Phi}_{k_N}\mathbf{U})^T\right)\mathbf{\Phi}_{k_N}\mathbf{b}_l \tag{3.14}$$

The projector $\text{P}_{\mathbf{\Phi}_{k_N}\mathbf{U}}$ computes distance from the sensed BV $\mathbf{\Phi}_{k_N}\mathbf{b}_l$ to the subspace $\mathbf{\Phi}_{k_N}\mathbf{U}$ in CSSD. Rearrange the symbols using the matrix transpose rules and apply the property $\mathbf{\Phi}_{k_N}^{\text{T}}\mathbf{\Phi}_{k_N} \to \mathbf{I} + \mathbf{E}$ as $m_N \to N$ results in equal to equation (3.15) and (3.16).

$$\text{P}_{\mathbf{\Phi}_{k_N}\mathbf{U}}\mathbf{\Phi}_{k_N}\mathbf{b}_l = \left(\mathbf{\Phi}_{k_N}\mathbf{U}\left(\mathbf{U}^T\mathbf{\Phi}_{k_N}^T\mathbf{\Phi}_{k_N}\mathbf{U}\right)^{-1}\mathbf{U}^T\mathbf{\Phi}_{k_N}^T\right)\mathbf{\Phi}_{k_N}\mathbf{b}_l \tag{3.15}$$

$$\text{P}_{\mathbf{\Phi}_{k_N}\mathbf{U}}\mathbf{\Phi}_{k_N}\mathbf{b}_l \approx \mathbf{\Phi}_{k_N}\mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^{\text{T}}\mathbf{b}_l \tag{3.16}$$

Appling the L_2-norm to (3.16) results in the following equation.

$$\left\|\text{P}_{\mathbf{\Phi}_{k_N}\mathbf{U}}\mathbf{\Phi}_{k_N}\mathbf{b}_l\right\|_2^2 \approx \left\|\mathbf{\Phi}_{k_N}\mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{b}_l\right\|_2^2 \tag{3.17}$$

Now the RIP condition can be used to assert the lengths of the projections are preserved [70]. However, going one step further and expanding the norm in (3.17) yields (3.18), and as $m_N \to N$, $\mathbf{\Phi}_{k_N}^T \mathbf{\Phi}_{k_N} \to \mathbf{I} + \mathbf{E}$ resulting in the approximation in (3.19).

$$\left\| \mathrm{P}_{\mathbf{\Phi}_{k_N}\mathbf{U}} \mathbf{\Phi}_{k_N} \mathbf{b}_l \right\|_2^2 \approx (\mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{b}_l)^T \mathbf{\Phi}_{k_N}^T \mathbf{\Phi}_{k_N} (\mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{b}_l) \tag{3.18}$$

$$\left\| \mathrm{P}_{\mathbf{\Phi}_{k_N}\mathbf{U}} \mathbf{\Phi}_{k_N} \mathbf{b}_l \right\|_2^2 \approx \|\mathbf{U}(\mathbf{U}^T\mathbf{U})^{-1}\mathbf{U}^T\mathbf{b}_l\|_2^2 = \|\mathrm{P}_{\mathbf{U}}\mathbf{b}_l\|_2^2 \tag{3.19}$$

Therefore, the projection length of a vector in CSSD will approximately equal the projection length in ODS. This holds for $\mathrm{P}^{\perp}_{\mathbf{\Phi}_{k_N}\mathbf{U}} = \mathbf{I} - \mathrm{P}_{\mathbf{\Phi}_{k_N}\mathbf{U}}$ as well. OSP for the band vectors (BV) is equivalent to the OSP for band tensors (BT). Using BVs, the OSP of $\mathbf{b}_l$ onto $\mathbf{U}$ is a linear predictor $(\alpha_1, \alpha_1, \dots, \alpha_i)^T$ such that $\alpha_1\mathbf{b}_1 + \alpha_2\mathbf{b}_2 + \cdots + \alpha_i\mathbf{b}_i = \mathbf{b}_l$ [17]. For BT, the linear predictor is $\alpha_1\mathbf{B}_1 + \alpha_2\mathbf{B}_2 + \cdots + \alpha_i\mathbf{B}_i = \mathbf{B}_l$. This view makes it clear that it is possible to vectorize the band tensors and compute the OSP giving the same result. This derivation shows that compressively sampling the BV and BT has little effect on the OSP between bands. This can be used as a band discrimination metric when finding the most distinct bands in terms of the least squared error.

Preservation of Sample Subspace Volume

The volume of the subspace spanned by a band subset is also an important HSI measure [71]. Band selection algorithms attempt to find a minimal but descriptive band subset. Unsupervised algorithms can take advantage of the volume estimates to find subsets which span the HSI data space [72]. The following derivations show that the subspace volume using the band subsets is preserved in CSSD. First arrange the bands in the set

$\{\text{vec}(\mathbf{B}_j)\}_{j=1}^{p} = \{\mathbf{b}_j\}_{j=1}^{p}$ into a matrix $\mathbf{M} = [\text{vec}(\mathbf{B}_1) \mid \text{vec}(\mathbf{B}_2) \mid \ldots \mid \text{vec}(\mathbf{B}_p)]$, the volume of the subspace can be measured using equation (3.20) [73].

$$J(\mathbf{M}) = \frac{1}{(p-1)!}\sqrt{|\mathbf{M}^T\mathbf{M}|} \tag{3.20}$$

Using the BV model of CS from (2.4), the subspace volume becomes as follows.

$$J(\mathbf{\Phi}_{k_N}\mathbf{M}) = \frac{1}{(p-1)!}\sqrt{|(\mathbf{\Phi}_{k_N}\mathbf{M})^T\mathbf{\Phi}_{k_N}\mathbf{M}|} \tag{3.21}$$

Using simple linear algebra manipulations it becomes the follow is derived.

$$J(\mathbf{\Phi}_{k_N}\mathbf{M}) = \frac{1}{(p-1)!}\sqrt{|\mathbf{M}^T(\mathbf{\Phi}_{k_N}^T\mathbf{\Phi}_{k_N})\mathbf{M}|} \tag{3.22}$$

As $m_N \to N$ the matrix $\mathbf{\Phi}_{k_N}^T\mathbf{\Phi}_{k_N} \to \mathbf{I} + \mathbf{E}$ with an exponentially decaying error function. Therefore, $J(\mathbf{\Phi}_{k_N}\mathbf{M}) \to J(\mathbf{M})$ as m_N increases. This implies the subspace volume is identical both in ODS and CSSD.

Using CS BT model in (2.2) or (2.3), the band subset of compressed tensors $\left\{\mathbf{\Phi}_{k_{n_1}}\mathbf{B}_j\mathbf{\Phi}_{k_{n_2}}^T\right\}_{j=1}^{p}$ are arranged into a matrix becoming $\mathbf{M}_{\mathbf{\Phi}} = \left[\text{vec}\left(\mathbf{\Phi}_{k_{n_1}}\mathbf{B}_1\mathbf{\Phi}_{k_{n_2}}^T\right) \mid \ldots \mid \text{vec}\left(\mathbf{\Phi}_{k_{n_1}}\mathbf{B}_p\mathbf{\Phi}_{k_{n_2}}^T\right)\right]$ and the volume of the space spanned by that matrix is given by equation (3.23).

$$J(\mathbf{M}_{\mathbf{\Phi}}) = \frac{1}{(p-1)!}\sqrt{|\mathbf{M}_{\mathbf{\Phi}}^T\mathbf{M}_{\mathbf{\Phi}}|} \tag{3.23}$$

To show that $J(\mathbf{M}_{\mathbf{\Phi}}) \to J(\mathbf{M})$ using the BT models, the inner product can be exploited by applying $\mathbf{\Phi}_{k_N}^T\mathbf{\Phi}_{k_N} \to \mathbf{I} + \mathbf{E}$ to eliminate the sampling matrices. Begin by breaking the sampling matrices into m_{n_1} and m_{n_2} vectors $\mathbf{\Phi}_{k_{n_1}}^T = \left[\mathbf{\phi}_{11}, \mathbf{\phi}_{12}, \ldots, \mathbf{\phi}_{1m_{n_1}}\right] \in \mathcal{R}^{n_1 x m_{n_1}}$ and

$\boldsymbol{\Phi}_{k_{n_2}}^T = \left[\boldsymbol{\phi}_{21}, \boldsymbol{\phi}_{22}, \ldots, \boldsymbol{\phi}_{2m_{n_2}}\right] \in \mathcal{R}^{n_2 x m_{n_2}}$, then the ij^{th} entry of $\mathbf{M}_{\boldsymbol{\Phi}}$ for each band tensor $\mathbf{B}_l$ is given as follows.

$$\left[\boldsymbol{\Phi}_{k_{n_1}} \mathbf{B}_l \boldsymbol{\Phi}_{k_{n_2}}^T\right]_{ij} = \boldsymbol{\phi}_{1i}^T \mathbf{B}_l \boldsymbol{\phi}_{2j} \tag{3.24}$$

Here $1 \leq i \leq m_{n_1}$, $1 \leq j \leq m_{n_2}$ and $1 \leq l \leq L$. Then writing out the vectorized and compressed BTs in $\mathbf{M}_{\boldsymbol{\Phi}}$ yields

$$\mathbf{M}_{\boldsymbol{\Phi}} = \begin{bmatrix} \boldsymbol{\phi}_{11}^T \mathbf{B}_1 \boldsymbol{\phi}_{21} & \cdots & \boldsymbol{\phi}_{11}^T \mathbf{B}_L \boldsymbol{\phi}_{21} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\phi}_{1m_{n_1}}^T \mathbf{B}_1 \boldsymbol{\phi}_{2m_{n_2}} & \cdots & \boldsymbol{\phi}_{1m_{n_1}}^T \mathbf{B}_L \boldsymbol{\phi}_{2m_{n_2}} \end{bmatrix} \tag{3.25}$$

Using this and the inner product, the entries of $\mathbf{M}_{\boldsymbol{\Phi}}^T \mathbf{M}_{\boldsymbol{\Phi}}$ are given by

$$[\mathbf{M}_{\boldsymbol{\Phi}}^T \mathbf{M}_{\boldsymbol{\Phi}}]_{ij} = \left[\sum_{h=1}^{m_{n_1}} \sum_{k=1}^{m_{n_2}} (\boldsymbol{\phi}_{1h}^T \mathbf{B}_i \boldsymbol{\phi}_{2k})(\boldsymbol{\phi}_{1h}^T \mathbf{B}_j \boldsymbol{\phi}_{2k})\right]_{ij} \tag{3.26}$$

Rearranging the summation results in the following.

$$[\mathbf{M}_{\boldsymbol{\Phi}}^T \mathbf{M}_{\boldsymbol{\Phi}}]_{ij} = \left[\sum_{h=1}^{m_{n_1}} \boldsymbol{\phi}_{1h}^T \mathbf{B}_i \left(\sum_{k=1}^{m_{n_2}} \boldsymbol{\phi}_{2k} \boldsymbol{\phi}_{2k}^T\right) \mathbf{B}_j^T \boldsymbol{\phi}_{1h}\right]_{ij} \tag{3.27}$$

Next $\sum_{k=1}^{m_{n_2}} \boldsymbol{\phi}_{2k} \boldsymbol{\phi}_{2k}^T = \boldsymbol{\Phi}_{k_{n_2}} \boldsymbol{\Phi}_{k_{n_2}}^T = (\mathrm{n}_2/\mathrm{m}_2)\mathbf{I} + \mathbf{E}$, $\mathbf{E} \to \mathbf{0}$ as $m_{n_2} \to n_2$ which follows.

$$[\mathbf{M}_{\boldsymbol{\Phi}}^T \mathbf{M}_{\boldsymbol{\Phi}}]_{ij} = \left(\frac{n_2}{m_2}\right)\left[\sum_{h=1}^{m_{n_1}} \boldsymbol{\phi}_{1h}^T \mathbf{B}_i \mathbf{B}_j^T \boldsymbol{\phi}_{1h}\right]_{ij} + [\mathbf{E}]_{ij} \tag{3.28}$$

Breaking the BTs into column vectors $\mathbf{B}_i = \left[\mathbf{b}_{i1}, \mathbf{b}_{i2}, \ldots, \mathbf{b}_{in_2}\right]$ and $\mathbf{B}_j = \left[\mathbf{b}_{j1}, \mathbf{b}_{j2}, \ldots, \mathbf{b}_{jn_2}\right]$ it is possible to re-write as a summation of dot products given in the following.

$$[\mathbf{M}_{\boldsymbol{\Phi}}^T \mathbf{M}_{\boldsymbol{\Phi}}]_{ij} = \left(\frac{n_2}{m_2}\right)\left[\sum_{h=1}^{m_{n_1}} \sum_{k=1}^{n_2} (\boldsymbol{\phi}_{1h}^T \mathbf{b}_{ik})(\mathbf{b}_{jk}^T \boldsymbol{\phi}_{1h})\right]_{ij} + [\mathbf{E}]_{ij} \tag{3.29}$$

Transposing the dot products and then rearranging the summation results in the following.

$$[\mathbf{M}_{\mathbf{\Phi}}^{T}\mathbf{M}_{\mathbf{\Phi}}]_{ij} = \left(\frac{n_2}{m_2}\right)\left[\sum_{k=1}^{n_2}\mathbf{b}_{ik}^{T}\left(\sum_{h=1}^{m_{n_1}}\boldsymbol{\phi}_{1h}\boldsymbol{\phi}_{1h}^{T}\right)\mathbf{b}_{jk}\right]_{ij} + [\mathbf{E}]_{ij} \tag{3.30}$$

It is known that $\sum_{h=1}^{m_{n_1}}\boldsymbol{\phi}_{1h}\boldsymbol{\phi}_{1h}^{T} = \mathbf{\Phi}_{k_{n_1}}\mathbf{\Phi}_{k_{n_1}}^{T} = (n_1/m_1)\mathbf{I} + \mathbf{E}$ and $\mathbf{E} \to \mathbf{0}$ as $m_{n_1} \to n_1$. The following equation can be derived.

$$\left[\mathbf{M}_{\mathbf{\Phi}}^{\mathrm{T}}\mathbf{M}_{\mathbf{\Phi}}\right]_{ij} = \left(\frac{n_1}{m_1}\right)\left(\frac{n_2}{m_2}\right)\left[\sum_{k=1}^{n_2}\mathbf{b}_{ik}^{\mathrm{T}}\,\mathbf{b}_{jk}\right]_{ij} + [\mathbf{E}]_{ij} \tag{3.31}$$

This proves the preservation of the volume using the BT in the CSSD implying that $J(\mathbf{M}_{\mathbf{\Phi}})/\mathrm{c} \to J(\mathbf{M})$ as $m_{n_1} \to n_1$ and $m_{n_2} \to n_2$. The preservation of the volume will be used when developing band subset selection algorithms in later chapters.

Conclusion

This chapter has shown the preservation of three HSI specific measures when using the sensed samples. By using data in the compressively sensed sample domain (CSSD), the covariance matrix, the correlation matrix as well as the orthogonal subspace projection and subspace volumes are preserved. Therefore, HSI data exploitation algorithms built using these measures can be performed in CSSD. The advantage of using CSSD is to reduce the size of the individual bands. The exponential convergence of the inner product of random Gaussian matrices ensures the measure converges exponentially in the number of compressed samples to the value computed using the original data space (ODS).

Chapter 4: The Restricted Entropy and Spectrum Properties

This chapter develops a new inter-pixel measure by extending the restricted isometric property (RIP) and restricted conformal property (RCP) to a new probabilistic discrimination function useful for hyperspectral (HSI) data processing [13], [39]. The narrow spectral range and high number of contiguous spectral bands are the defining and prominent features of HSI images [3]. This results in spectral data redundancy which implies a lower information rate than the bandwidth and a sparse representation exists so that compressive sensing (CS) can be applied directly to this HSI mode [12], [50]. Each of the N pixels $\mathbf{r}_i = (r_{i1}, r_{i2}, \dots, r_{iL})^T = (\lambda_1, \lambda_2, \dots, \lambda_L)^T \in \Re^{Lx1}$ for the $1 \leq i \leq N$ image pixels and band center wavelengths of $\{\lambda_l\}_{l=1}^{L}$ is sensed using equation (2.1). The sensed pixels are reduced to $\mathbf{y}_i \in \Re^{m_L x1}$ where m_L is the compressive sensing sampling rate (CSSR) of the pixel set. The HSI data exploitation algorithms of classification [74], anomaly detection [4], target detection [16], [67], endmember finding [65] and linear spectral mixture analysis [75] are fundamentally based on measuring the discrepancies and similarities between pixels. Using the length preserving property of RIP in (1.1) and the angle preserving property of RCP in (1.2), inter-pixel measures based on the Euclidean distance (ED) or spectral angle mapper (SAM) between pixels are preserved in the compressively sensed band domain (CSBD) [76]. Algorithms developed using these measures in the original data space (ODS) have been utilized in the hyperspectral community for some time [2], [3]. While ED and SAM measures are geometric inter-pixel measures, an alternative probabilistic measure known as Spectral Information Divergence (SID) was introduced by Chang in 1999 [76]. This measure is based on the Kullback-Leibler (KL) distance between probability distributions [77]. This metric evaluates the

difference in the shape of the probability distributions and incorporates all moments of the distribution using information entropy [78]. Also, it was shown that the SID is effective at accounting for the spectral variability of the pixels due to the scene environment and atmospheric effects [76].

A direct correspondence between these measures in ODS and CSBD can be drawn. RIP is used to preserve ED in both spaces, RCP is used to preserve SAM in both spaces, and the new Restricted Entropy Property (REP) derived here is used to preserve the entropy between spectral pixels in both spaces. The KL distance is derived from the information entropy and REP implies the KL distance is preserved in both spaces. The preservation of the inter-pixel KL distances of an HSI is carried out by Restricted Spectrum Property (RSP).

The KL-distance is a measure of the difference between probably distributions of random variables (RV) not pixel vectors directly [77]. Therefore, an RV is defined for each pixel and a probability mass function (PMF) is estimated using the pixel data known as the spectral information measure (SIM) in equation (4.1) [76].

$$p_{il} = \frac{|r_{il}|}{\sum_{l=1}^{L} |r_{il}|} \tag{4.1}$$

According to (4.1), $\mathbf{r}_i$ is considered the output of a RV defined on a probability space $\mathrm{X} = (\Omega, \Sigma, \mathrm{P})$ where the sample space Ω consists of all individual wavelengths $\{\lambda_l\}_{l=1}^{L}$ of the HSI, the event space Σ is a power set of the wavelengths and the probability measure P is the probability the l^{th} wavelength capture energy [60]. In other words, the PMF of P is the probability that the photons entering the sensor will contain energy in the l^{th} of wavelengths or equally how many photons entering have components of each wavelength.

Restricted Entropy Property Lemma

Using the PMF of the pixel RV defined by the SIM in (4.1), let $\mathbf{p}_i = (p_{i1}, p_{i2}, \dots, p_{iL})^T$ so that the information entropy of the i^{th} pixel $\mathbf{r}_i$ is defined by $H(\mathbf{p}_i) = -\sum_{l=1}^{L} p_{il} \log(p_{il})$ [78]. In this section a lemma called the Restricted Entropy Property (REP) shows that the entropy measure in ODS is approximately preserved in CSBD such that $H(\mathbf{p}_i) \to \boldsymbol{H}(\boldsymbol{\Phi}_{k_L}\mathbf{p}_i)$ as $m_L \to L$. Notice that $H(\mathbf{p}_i)$ is the entropy of the pixel values in $\mathbf{r}_i$ computed using the PMF values of p_{il} and $H(\boldsymbol{\Phi}_{k_L}\mathbf{p}_i)$ is the entropy computed using the sensed pixel values in $\mathbf{y}_i = \boldsymbol{\Phi}_{k_L}\mathbf{r}_i$. Therefore, $\boldsymbol{\Phi}_{k_L}\mathbf{p}_i$ does not imply that the PMF is being transformed and is nonsensical when used outside the entropy operator H and the distance metrics D_{SID} and D_{KL} which uses the SIM and the entropy.

The first step of the proof is to express the entropy in terms of the angles between the L unit basis vectors and the pixel vector $\mathbf{r}_i$ which is denoted $\theta_l = \angle(\mathbf{e}_l, \mathbf{r}_i)$ where $\mathbf{e}_l = (0, 0, \dots, 1, 0, 0)^T$ is the basis vector with the l^{th} position equal to 1 and all other positions equal to 0.

$$H(\mathbf{p}_i) = -\sum_{l=1}^{L} p_{il} \log(p_{il}) = -\sum_{l=1}^{L} \frac{r_{il}}{\bar{r}_i} \log\left(\frac{r_{il}}{\bar{r}_i}\right) \tag{4.2}$$

$$= -\sum_{l=1}^{L} \left(\frac{|\langle \mathbf{e}_l | \mathbf{r}_i \rangle|}{\sum_{l=1}^{L} |\langle \mathbf{e}_l | \mathbf{r}_i \rangle|}\right) \log\left(\frac{|\langle \mathbf{e}_l | \mathbf{r}_i \rangle|}{\sum_{l=1}^{L} |\langle \mathbf{e}_l | \mathbf{r}_i \rangle|}\right) \tag{4.3}$$

$$= -\sum_{l=1}^{L} \left(\frac{|\mathbf{e}_l^T \mathbf{r}_i|}{\sum_{l=1}^{L} |\mathbf{e}_l^T \mathbf{r}_i|}\right) \log\left(\frac{|\mathbf{e}_l^T \mathbf{r}_i|}{\sum_{l=1}^{L} |\mathbf{e}_l^T \mathbf{r}_i|}\right) \tag{4.4}$$

where $\bar{r}_{il} = \sum_{l=1}^{L} |r_{il}|$. Furthermore, (4.4) can be re-expressed as follows to eliminate the vectors norms which are constant and canceled out.

$$H(\mathbf{p}_i) = -\sum_{l=1}^{L}\left(\frac{\|\mathbf{e}_l\|\|\mathbf{r}_i\|\,|\cos\theta_l|}{\sum_{l=1}^{L}\|\mathbf{e}_l\|\|\mathbf{r}_i\|\,|\cos\theta_l|}\right)\log\left(\frac{\|\mathbf{e}_l\|\|\mathbf{r}_i\|\,|\cos\theta_l|}{\sum_{l=1}^{L}\|\mathbf{e}_l\|\|\mathbf{r}_i\|\,|\cos\theta_l|}\right) \tag{4.5}$$

$$= -\sum_{l=1}^{L}\left(\frac{|\cos\theta_l|}{\sum_{l=1}^{L}|\cos\theta_l|}\right)\log\left(\frac{|\cos\theta_l|}{\sum_{l=1}^{L}|\cos\theta_l|}\right) \tag{4.6}$$

Now express the $\cos(\theta_l)$ using the law of cosines to express the entropy algebraically as vector magnitudes as in equation (4.7).

$$\cos(\theta_l) = \frac{\|\mathbf{e}_l\|_2^2 - \|\mathbf{r}_i\|_2^2 - \|\mathbf{e}_l - \mathbf{r}_i\|_2^2}{2\|\mathbf{e}_l\|_2^2\|\mathbf{r}_i\|_2^2} \tag{4.7}$$

By taking advantage of the RIP condition in (1.1), it is possible to replace the vector magnitudes in the law of cosines with the transformed magnitudes $\|\mathbf{r}_i\|_2^2 \approx \left\|\mathbf{\Phi}_{k_L}\mathbf{r}_i\right\|_2^2$.

$$\cos(\theta_l) \approx \cos(\alpha_l) = \frac{\left\|\mathbf{\Phi}_{k_L}\mathbf{e}_l\right\|_2^2 - \left\|\mathbf{\Phi}_{k_L}\mathbf{r}_i\right\|_2^2 - \left\|\mathbf{\Phi}_{k_L}\mathbf{e}_l - \mathbf{\Phi}_{k_L}\mathbf{r}_i\right\|_2^2}{2\left\|\mathbf{\Phi}_{k_L}\mathbf{e}_l\right\|_2^2\left\|\mathbf{\Phi}_{k_L}\mathbf{r}_i\right\|_2^2} \tag{4.8}$$

Now redefine the entropy of (4.5) in CSBD $\mathbf{\Phi}_{k_L}$ where $\alpha_l = \angle(\mathbf{\Phi}_{k_L}\mathbf{e}_l, \mathbf{\Phi}_{k_L}\mathbf{r}_i)$ in equation (4.9). In addition, this redefinition allows the angle to be preserved by the RCP property during CS so that $H\left(\mathbf{\Phi}_{k_L}\mathbf{p}_i\right) =$

$$-\sum_{l=1}^{m_L}\left(\frac{\left\|\mathbf{\Phi}_{k_L}\mathbf{e}_l\right\|\left\|\mathbf{\Phi}_{k_L}\mathbf{r}_i\right\||\cos\alpha_l|}{\sum_{l=1}^{m_L}\left\|\mathbf{\Phi}_{k_L}\mathbf{e}_l\right\|\left\|\mathbf{\Phi}_{k_L}\mathbf{r}_i\right\||\cos\alpha_l|}\right)\log\left(\frac{\left\|\mathbf{\Phi}_{k_L}\mathbf{e}_l\right\|\left\|\mathbf{\Phi}_{k_L}\mathbf{r}_i\right\||\cos\alpha_l|}{\sum_{l=1}^{m_L}\left\|\mathbf{\Phi}_{k_L}\mathbf{e}_l\right\|\left\|\mathbf{\Phi}_{k_L}\mathbf{r}_i\right\||\cos\alpha_l|}\right) \tag{4.9}$$

$$= -\sum_{l=1}^{m_L}\left(\frac{\left|\langle\mathbf{\Phi}_{k_L}\mathbf{e}_l, \mathbf{\Phi}_{k_L}\mathbf{r}_i\rangle\right|}{\sum_{l=1}^{m_L}\left|\langle\mathbf{\Phi}_{k_L}\mathbf{e}_l, \mathbf{\Phi}_{k_L}\mathbf{r}_i\rangle\right|}\right)\log\left(\frac{\left|\langle\mathbf{\Phi}_{k_L}\mathbf{e}_l, \mathbf{\Phi}_{k_L}\mathbf{r}_i\rangle\right|}{\sum_{l=1}^{m_L}\left|\langle\mathbf{\Phi}_{k_L}\mathbf{e}_l, \mathbf{\Phi}_{k_L}\mathbf{r}_i\rangle\right|}\right) \tag{4.10}$$

$$= -\sum_{l=1}^{m_L}\left(\frac{\left|\mathbf{e}_l^T\mathbf{\Phi}_{k_L}^T\mathbf{\Phi}_{k_L}\mathbf{r}_i\right|}{\sum_{l=1}^{m_L}\left|\mathbf{e}_l^T\mathbf{\Phi}_{k_L}^T\mathbf{\Phi}_{k_L}\mathbf{r}_i\right|}\right)\log\left(\frac{\left|\mathbf{e}_l^T\mathbf{\Phi}_{k_L}^T\mathbf{\Phi}_{k_L}\mathbf{r}_i\right|}{\sum_{l=1}^{m_L}\left|\mathbf{e}_l^T\mathbf{\Phi}_{k_L}^T\mathbf{\Phi}_{k_L}\mathbf{r}_i\right|}\right) \tag{4.11}$$

As $m_L \to L$, then $\mathbf{\Phi}_{k_L}^T\mathbf{\Phi}_{k_L} \to \mathbf{I} + \mathbf{E}$ with $\mathbf{E} \to \mathbf{0}$ and the entropy is simplified to the following.

$$H\left(\mathbf{\Phi}_{k_L}\mathbf{p}_i\right) \approx -\sum_{l=1}^{L}\left(\frac{|\mathbf{e}_l^T\mathbf{r}_i|}{\sum_{l=1}^{L}|\mathbf{e}_l^T\mathbf{r}_i|}\right)\log\left(\frac{|\mathbf{e}_l^T\mathbf{r}_i|}{\sum_{l=1}^{L}|\mathbf{e}_l^T\mathbf{r}_i|}\right) \tag{4.12}$$

$$= -\sum_{l=1}^{L}\frac{r_{il}}{\bar{r}_i}\log\left(\frac{r_{il}}{\bar{r}_i}\right) = -\sum_{l=1}^{L}p_{il}\log(p_{il}) = H(\mathbf{p}_i) \tag{4.13}$$

This proves the entropy preservation of $H\left(\mathbf{\Phi}_{k_L}\mathbf{p}_i\right) \to H(\mathbf{p}_i)$ as $m_L \to L$ and the entropy estimates in OSD will coincide with those in CSBD. A corollary to (4.13) is the cross entropy between two pixels. Let $\boldsymbol{q}_j = \left(q_{j1}, q_{j2}, \ldots, q_{jL}\right)^T$ be the PMF computed using the SIM of the j^{th} pixel $\mathbf{s}_j$, then the cross entropy is defined as $H\left(\mathbf{p}_i, \mathbf{q}_j\right) = -\sum_{l=1}^{L}p_{il}\log q_{jl}$. The proof takes advantage of REP by observing the similarities between the entropy and cross entropy equations. Starting with (4.10), the cross entropy in CSBD is as follows.

$$H\left(\mathbf{\Phi}_{k_L}\mathbf{p}_i, \mathbf{\Phi}_{k_L}\mathbf{q}_j\right)$$

$$= -\sum_{l=1}^{m_L}\left(\frac{\left|\langle\mathbf{\Phi}_{k_L}\mathbf{e}_l, \mathbf{\Phi}_{k_L}\mathbf{r}_i\rangle\right|}{\sum_{l=1}^{m_L}\left|\langle\mathbf{\Phi}_{k_L}\mathbf{e}_l, \mathbf{\Phi}_{k_L}\mathbf{r}_i\rangle\right|}\right)\log\left(\frac{\left|\langle\mathbf{\Phi}_{k_L}\mathbf{e}_l, \mathbf{\Phi}_{k_L}\boldsymbol{s}_j\rangle\right|}{\sum_{l=1}^{m_L}\left|\langle\mathbf{\Phi}_{k_L}\mathbf{e}_l, \mathbf{\Phi}_{k_L}\boldsymbol{s}_j\rangle\right|}\right) \tag{4.14}$$

$$= -\sum_{l=1}^{m_L}\left(\frac{\left|\mathbf{e}_l^T\mathbf{\Phi}_{k_L}^T\mathbf{\Phi}_{k_L}\mathbf{r}_i\right|}{\sum_{l=1}^{m_L}\left|\mathbf{e}_l^T\mathbf{\Phi}_{k_L}^T\mathbf{\Phi}_{k_L}\mathbf{r}_i\right|}\right)\log\left(\frac{\left|\mathbf{e}_l^T\mathbf{\Phi}_{k_L}^T\mathbf{\Phi}_{k_L}\boldsymbol{s}_j\right|}{\sum_{l=1}^{m_L}\left|\mathbf{e}_l^T\mathbf{\Phi}_{k_L}^T\mathbf{\Phi}_{k_L}\boldsymbol{s}_j\right|}\right) \tag{4.15}$$

As $m \to L$, then $\mathbf{\Phi}_{k_L}^T\mathbf{\Phi}_{k_L} \to \mathbf{I} + \mathbf{E}$ with $\mathbf{E} \to \mathbf{0}$ then (4.15) becomes as follows.

$$H\left(\mathbf{\Phi}_{k_L}\mathbf{p}_i, \mathbf{\Phi}_{k_L}\mathbf{q}_j\right) \approx -\sum_{l=1}^{L}\left(\frac{|\mathbf{e}_l^T\mathbf{r}_i|}{\sum_{l=1}^{L}|\mathbf{e}_l^T\mathbf{r}_i|}\right)\log\left(\frac{\left|\mathbf{e}_l^T\mathbf{s}_j\right|}{\sum_{l=1}^{L}\left|\mathbf{e}_l^T\mathbf{s}_j\right|}\right) \tag{4.16}$$

$$= -\sum_{l=1}^{L}p_{il}\log\left(q_{jl}\right) = H\left(\mathbf{p}_i, \mathbf{q}_j\right) \tag{4.17}$$

The lemmas in (4.13) for entropy and (4.17) for cross entropy are used to express the KL distance in CSBD.

Restricted Spectrum Property Lemma

The KL distance between the two RV defined by the i^{th} and j^{th} pixels is $\mathrm{D}_{KL}(\mathbf{p}_i \,||\, \mathbf{q}_j) = H(\mathbf{p}_i, \mathbf{q}_j) - H(\mathbf{p}_i)$ where $H(\mathbf{p}_i)$ is the information entropy and $H(\mathbf{p}_i, \mathbf{q}_j)$ is the cross entropy [77], [78]. Using (4.13) and (4.17), the KL distance in the CSBD is defined as $\mathrm{D}_{KL}(\mathbf{\Phi}_{k_L}\mathbf{p}_i \,||\, \mathbf{\Phi}_{k_L}\mathbf{q}_j) = H(\mathbf{\Phi}_{k_L}\mathbf{p}_i, \mathbf{\Phi}_{k_L}\mathbf{q}_j) - H(\mathbf{\Phi}_{k_L}\mathbf{p}_i)$. The REP and cross entropy corollary are directly applied resulting in $\mathrm{D}_{KL}(\mathbf{\Phi}_{k_L}\mathbf{p}_i \,||\, \mathbf{\Phi}_{k_L}\mathbf{q}_j) \rightarrow \mathrm{D}_{KL}(\mathbf{p}_i \,||\, \mathbf{q}_j)$ as $m \rightarrow L$. The KL distance is not symmetric. However, the SID between pixels enforces a symmetric measure and is defined as $\mathrm{D}_{SID}(\mathbf{p}_i \,||\, \mathbf{q}_j) = \mathrm{D}_{KL}(\mathbf{p}_i \,||\, \mathbf{q}_j) + \mathrm{D}_{KL}(\mathbf{q}_j \,||\, \mathbf{p}_i)$. Knowing that $\mathrm{D}_{KL}(\mathbf{\Phi}_{k_L}\mathbf{p}_i \,||\, \mathbf{\Phi}_{k_L}\mathbf{q}_j) \rightarrow \mathrm{D}_{KL}(\mathbf{p}_i \,||\, \mathbf{q}_j)$ as $m \rightarrow L$ implies the RSP property in equation (4.18).

$$\mathrm{D}_{SID}(\mathbf{\Phi}_{k_L}\mathbf{p}_i \,||\, \mathbf{\Phi}_{k_L}\mathbf{q}_j) \rightarrow \mathrm{D}_{SID}(\mathbf{p}_i \,||\, \mathbf{q}_j) \tag{4.18}$$

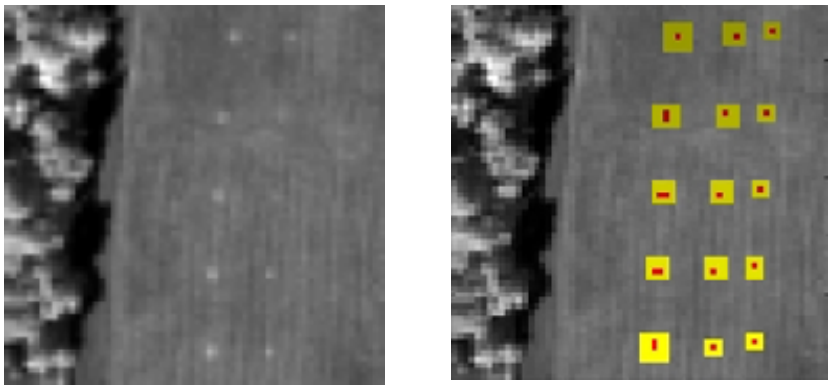
Real Hyperspectral Test Images

The algorithms developed in this study are verified through experiment with real HSI data sets. A common set of benchmarked data sets are used throughout the HSI data processing literature [2], [3], [5], [6]. A fundamental property of HSI exploitation is spectral data redundancy. Therefore, the inter-band correlation is measured as well as CSBR and CSSR using the SCS-RK method developed in the previous chapter.

HYDICE Data Set

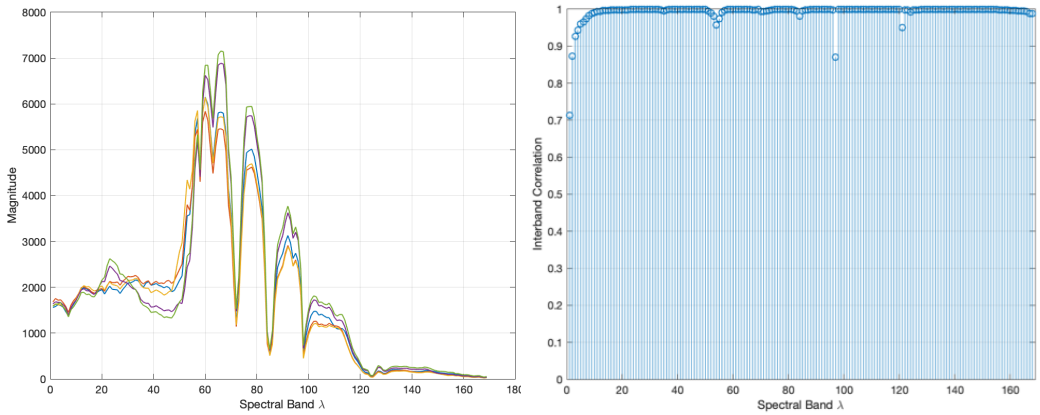
The Hyperspectral Digital Imagery Collection Experiment (HYDICE) data set is often used to validate HSI detection algorithms [7]. Here it will be used to validate the preservation of REP and RSP. It was collected in August 1995 by the airborne visible

infrared imaging spectrometer (AVIRIS) over Aberdeen Proving grounds in Maryland, USA. The sensor has a spectral range of 0.4–2.5 μm and the image consists of 64 x 64 pixels at 1.56 m ground spatial distance (GSD). Low SNR bands of 1-3 and 202-210 were removed. In addition, water vapor absorption bands of 101-112 and 137-153 were removed leaving 169 bands. There scene has 15 painted target panels, a large grass field background, forest left of the field and a road. The set of panels are arranged in a five by three grid. The ground truth map is in Fig. 12(b). The red pixels in the image are considered pure panel pixels and the surrounding yellow pixels are the panel spectra mixed with the background.



(a) (b)

Fig. 12. HYDICE hyperspectral test image and ground truth

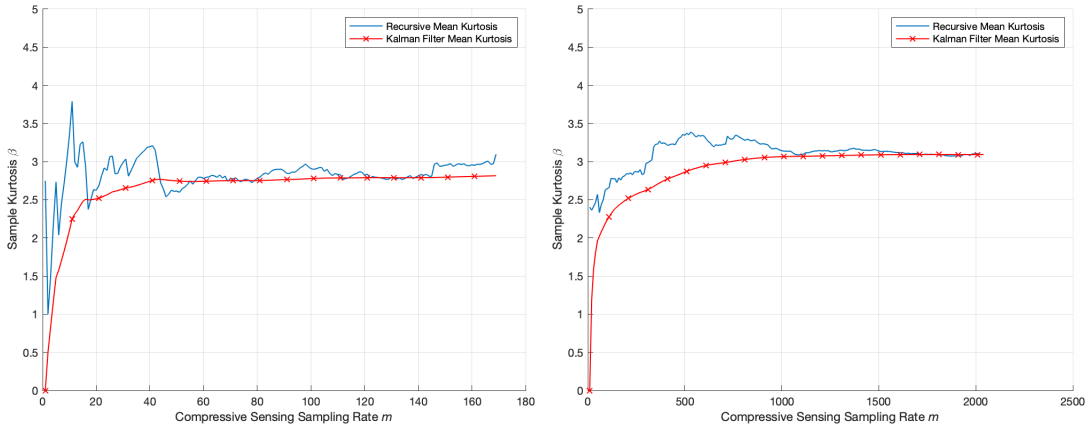


(a) (b)

Fig. 13. HYDICE scene (a) 5 spectral signatures (b) inter-band correlation

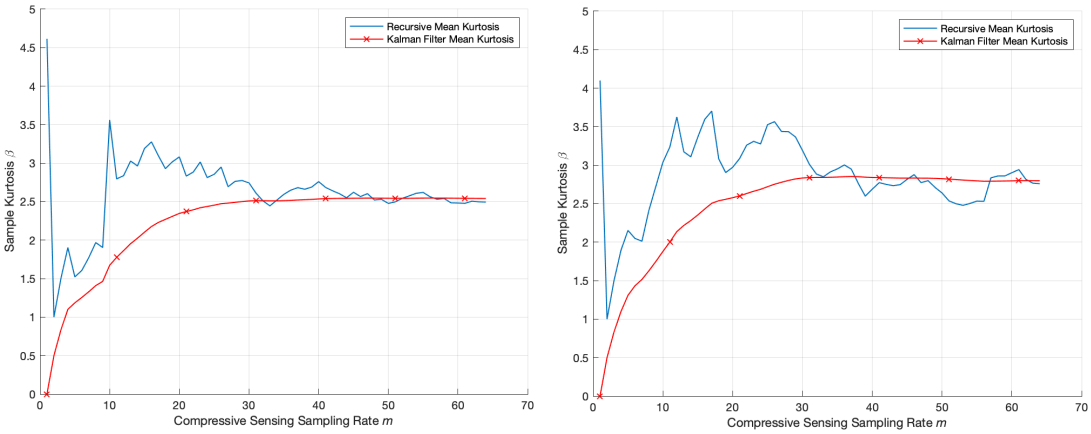
Each row is a unique material and each column represents a different panel size with the left-most column being 3mx3m, the middle is 2mx2m and the right column is 1mx1m. The panels in row 2 and 3 are made with the same material and differing paint. Rows 4 and 5 also are the same material but different paint. The 5 spectra from the panels is plotted in Fig. 13(a) and are used as a-priori information to detection algorithms. The inter-band correlation is plotted in Fig. 13(b) showing the high redundancy between spectral bands.

The CSBR and CSSR plots using SCS-RK are depicted in Fig. 14 and Fig. 15. These plots show variation of the kurtosis as a function of increasing CSSR. The convergence point of each of these graphs is used to construct the sampling matrices such



(a) (b)

Fig. 14. HYDICE CSSR convergence (a) PV plot and (b) BV plot



(a) (b)

Fig. 15. HYDICE CSSR convergence of BT (a) columns and (b) rows

$\mathbf{\Phi}_{k_L}$ using the CSBR from Fig. 14(a), $\mathbf{\Phi}_{k_N}$ using CSSR from Fig. 14(b), and $\mathbf{\Phi}_{k_{n_1}}$ and $\mathbf{\Phi}_{k_{n_2}}$ from the CSSR extracted from the two plots in Fig. 15.

Empirical Results

Using RIP, RCP and RSP the following pixel discrimination metrics of Euclidean distance (ED) in (4.19), spectral angle mapper (SAM) (4.20), spectral information divergence (SID) (4.21) and a combination of SID and SAM called SIDAM in (4.22) are defined using pixels $\mathbf{r}_i$ and $\mathbf{r}_j$ and corresponding PMFs of $\mathbf{p}_i$ and $\mathbf{q}_j$.

$$\mathrm{D}_{ED}(\mathbf{r}_i, \mathbf{r}_j) \approx \mathrm{D}_{ED}(\mathbf{\Phi}_{k_L}\mathbf{r}_i, \mathbf{\Phi}_{k_L}\mathbf{r}_j) = \|\mathbf{y}_i - \mathbf{y}_j\|_2^2 \tag{4.19}$$

$$\mathrm{D}_{SAM}(\mathbf{r}_i, \mathbf{r}_j) \approx \mathrm{D}_{SAM}(\mathbf{\Phi}_{k_L}\mathbf{r}_i, \mathbf{\Phi}_{k_L}\mathbf{r}_j) = \cos^{-1}\left(\frac{|\langle \mathbf{y}_i, \mathbf{y}_j \rangle|}{\|\mathbf{y}_i\|_2^2 \|\mathbf{y}_j\|_2^2}\right) \tag{4.20}$$

To empirically validate the preservation of each discrimination metric of D_{ED}, D_{SAM}, D_{SID}, and D_{SIDAM} is computed using the data in ODS and CSBD for the set of pixels from the HYDICE data set $\{\mathbf{r}_i\}_{i=1}^{4096}$ of Fig. 12(a).

$$\begin{gathered} \mathrm{D}_{SID}(\mathbf{p}_i, \mathbf{q}_j) \approx \mathrm{D}_{SID}(\mathbf{\Phi}_{k_L}\mathbf{p}_i, \mathbf{\Phi}_{k_L}\mathbf{q}_j) \\ \mathrm{D}_{SID}(\mathbf{\Phi}_{k_L}\mathbf{p}_i, \mathbf{\Phi}_{k_L}\mathbf{q}_j) = \mathrm{D}_{KL}(\mathbf{\Phi}_{k_L}\mathbf{p}_i || \mathbf{\Phi}_{k_L}\mathbf{p}_i) + \mathrm{D}_{KL}(\mathbf{\Phi}_{k_L}\mathbf{p}_i || \mathbf{\Phi}_{k_L}\mathbf{p}_i) \end{gathered} \tag{4.21}$$

$$\begin{gathered} \mathrm{D}_{SIDAM}(\mathbf{r}_i, \mathbf{r}_j) \approx \mathrm{D}_{SIDAM}(\mathbf{\Phi}_{k_L}\mathbf{r}_i, \mathbf{\Phi}_{k_L}\mathbf{r}_j) \\ \mathrm{D}_{SIDAM}(\mathbf{\Phi}_{k_L}\mathbf{r}_i, \mathbf{\Phi}_{k_L}\mathbf{r}_j) = \mathrm{D}_{SID}(\mathbf{\Phi}_{k_L}\mathbf{p}_i, \mathbf{\Phi}_{k_L}\mathbf{q}_j) \cdot \mathrm{D}_{SAM}(\mathbf{\Phi}_{k_L}\mathbf{r}_i, \mathbf{\Phi}_{k_L}\mathbf{r}_j) \end{gathered} \tag{4.22}$$

A reference pixel is chosen and fixed as the first input argument. The target spectral signatures in Fig. 13(a) will be used as a reference. Each target pixel denoted $\{\mathbf{s}_j\}_{j=1}^{5}$ is used and a CSBR of $m_L = 64$ was extracted from Fig. 14(a) using SCS-RK based method

detailed in Chapter 2. Fig. 16(a) shows the metrics computed for $D_{ED}(\mathbf{s}_1, \mathbf{r}_i)$, $D_{SAM}(\mathbf{s}_1, \mathbf{r}_i)$, $D_{SID}(\mathbf{s}_1, \mathbf{r}_i)$, and $D_{SIDAM}(\mathbf{s}_1, \mathbf{r}_i)$ in both ODS and CSBD. Fig. 16(b) shows a zoomed in portion of the measures so the correlation between the metrics is evident. Fig. 17 displays the measures for $D_{ED}(\mathbf{s}_2, \mathbf{r}_i)$, $D_{SAM}(\mathbf{s}_2, \mathbf{r}_i)$, $D_{SID}(\mathbf{s}_2, \mathbf{r}_i)$, and $D_{SIDAM}(\mathbf{s}_2, \mathbf{r}_i)$ in ODS and CSBD using the second target pixel from Fig. 13(a). Fig. 18 plots the measures for $D_{ED}(\mathbf{s}_3, \mathbf{r}_i)$, $D_{SAM}(\mathbf{s}_3, \mathbf{r}_i)$, $D_{SID}(\mathbf{s}_3, \mathbf{r}_i)$, and $D_{SIDAM}(\mathbf{s}_3, \mathbf{r}_i)$ in ODS and CSBD.

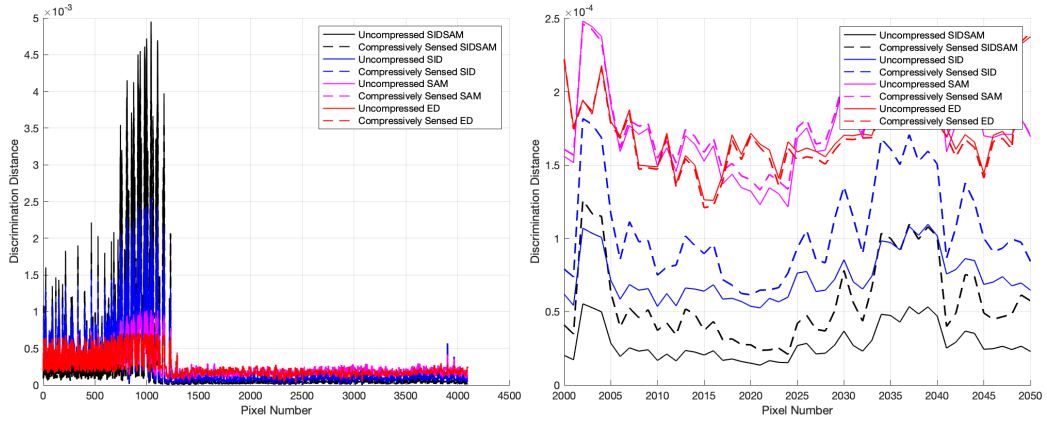


Fig. 16. All four inter-pixel metrics in the ODS and CSBD for reference pixel 1

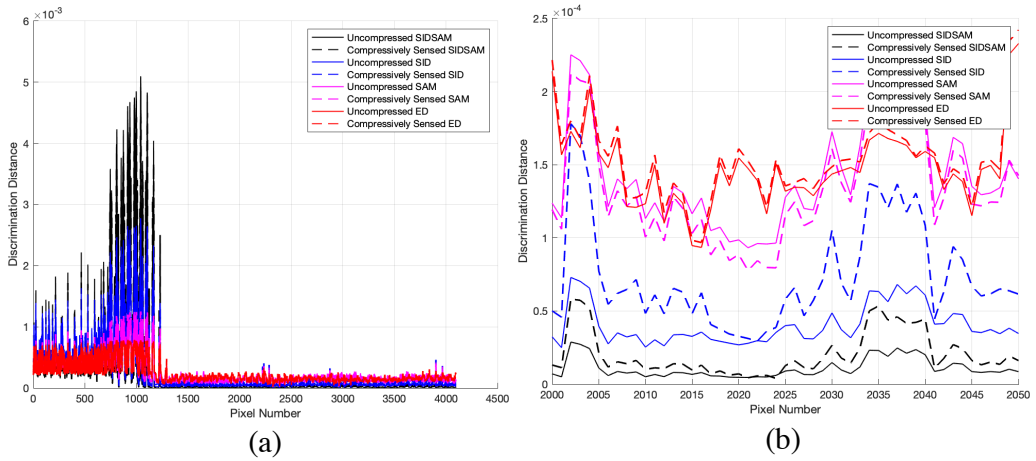


Fig. 17. All four inter-pixel metrics in the ODS and CSBD for reference pixel 2

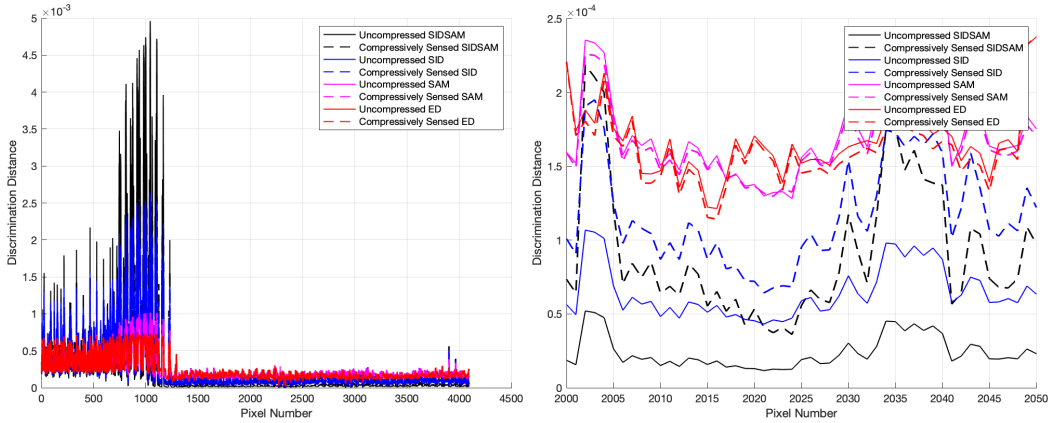


Fig. 18. All four inter-pixel metrics in the ODS and CSBD for reference pixel 3

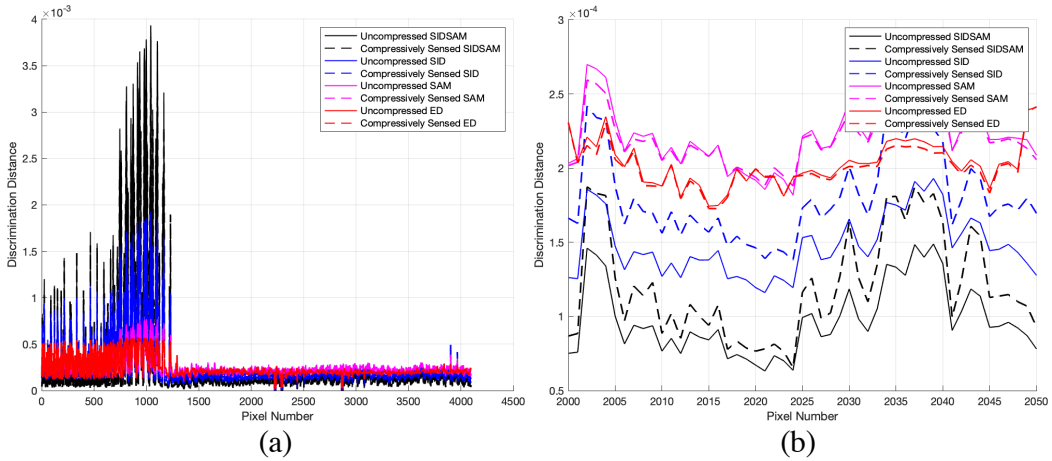


Fig. 19. All four inter-pixel metrics in the ODS and CSBD for reference pixel 4

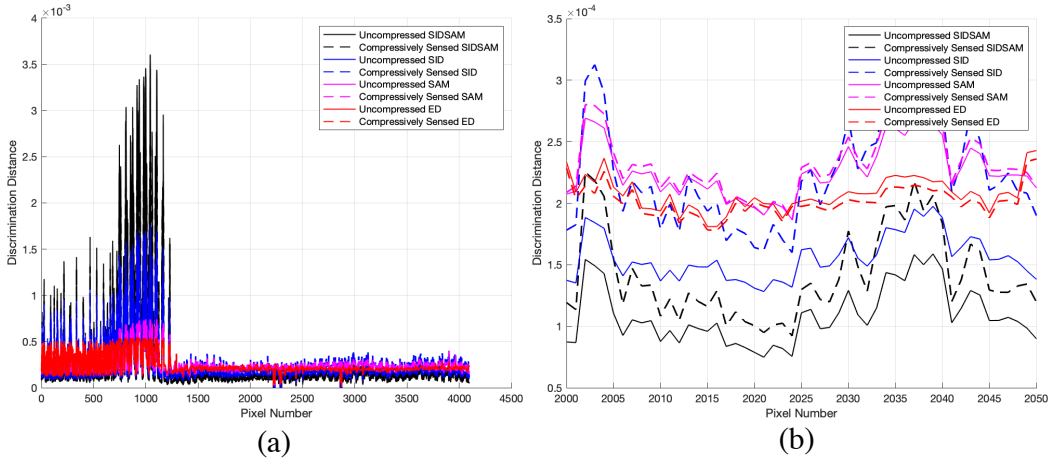


Fig. 20. All four inter-pixel metrics in the ODS and CSBD for reference pixel 5.

Fig. 19 and Fig. 20 show the plots of the measures for D_{ED}, D_{SAM}, D_{SID}, and D_{SIDAM} for reference target pixels $\mathbf{s}_4$ and $\mathbf{s}_5$ from Fig. 13(a) in ODS and CSBD. A measure of accuracy for the discrimination metrics is computed using the Pearson correlation between the measures in ODS and CSBD. Equation (4.23) shows the correlation computed for each metric $\eta = \{ED, SAM, SID, SIDAM\}$ and each reference pixel $\{\mathbf{s}_j\}_{j=1}^{5}$ across the entire pixel set $\{\mathbf{r}_i\}_{i=1}^{4096}$. The correlation bounded between $[-1,1]$ are specified in Table 1. These correlations indicate that indeed the measures are being preserved approximately giving further relevance to the correspondence of RIP and ED, RCP and SAM, RSP and SID as well as SIDAM.

$$\rho\left(\mathrm{D}_{\eta}\left(\mathbf{s}_j, \mathbf{r}_i\right), \mathrm{D}_{\eta}\left(\mathbf{\Phi}_{k_L}\mathbf{s}_j, \mathbf{\Phi}_{k_L}\mathbf{r}_i\right)\right) \text{ for } \eta = \{ED, SAM, SID, SIDAM\} \tag{4.23}$$

Table 1. Correlations between measure curves for each reference pixel

Measure	ED	SAM	SID	SIDAM
$\mathbf{s}_1$	0.9996	0.9939	0.9619	0.9845
$\mathbf{s}_2$	0.9998	0.9987	0.9708	0.9906
$\mathbf{s}_3$	0.9971	0.9995	0.9781	0.9908
$\mathbf{s}_4$	0.9962	0.9970	0.9815	0.9826
$\mathbf{s}_5$	0.9992	0.9996	0.9628	0.9738

Conclusions

This chapter extended the fundamental length and angle preserved CS properties of RIP and RCP to the information entropy and further derived the novel restricted entropy property (REP). In addition, REP is further extended to restricted spectrum property (RSP) by taking advantage of preserving the Kullback-Leibler distance in CSBD. Using these CS properties, four new inter-pixel discrimination measures are defined which can be utilized

in the original data space (ODS) or compressively sensed band domain (CSBD). Therefore, algorithms based on comparing pixels such as classification, anomaly detection, target detection, endmember finding and unmixing can benefit from CS while maintaining the same accuracy.

Chapter 5: Compressive Sensing Hyperspectral Target Detection

Chapters 3 and 4 introduced the idea of using the restricted isometric property (RIP), restricted conformal property (RCP) and restricted spectrum property (RSP) as inter-pixel and inter-band discrimination metrics. This chapter builds on these concepts to demonstrate hyperspectral (HSI) sub-pixel anomaly and target detection algorithms can be applied in the spectral mode of the compressively sensed band domain (CSBD) and spatial modes of the compressively sensed sample domain (CSSD) achieving minimal loss of accuracy given an appropriate compressively sensed band rate (CSBR) and a compressively sensed sample rate (CSSR). The majority of HSI pixel-based detectors characterize the background interferers and noise using the correlation or covariance matrices [2], [4], [14], [79]. Given the large number of pixels in each image, calculating these matrices are computationally expensive [69]. The covariance and correlation matrix derivation in CSSD of Chapter 3 can be used in to reduce the computing time. This chapter presents a spectral-spatial compressive sensing (CS) approach to target detection which improves the computational and memory space complexity of the background characterization as well as the sub-pixel detection algorithms.

The goal of HSI sub-pixel detection algorithms is to automatically model and suppress the image background including of noise, atmospheric effects and interferers [14]. In doing so, effective thresholds can be established which minimize the false positive rates using the Neyman-Pearson lemma as an example [80]. Target detection algorithms applied to HSI can take advantage of typically high signal-to-noise ratios (SNR). However, the low ground sample distances (GSD) between adjacent image pixels results in targets often embedded in a single pixel. Therefore, pixel-based detection algorithms cannot rely on

spatial pixel analysis and must exploit the fine spectral resolution of contiguous bands [67]. Characterizing the spectra of the background can be done effectively using the second order statistics of the inter-band covariance or correlation matrices. As is done with principal component analysis (PCA) [29], these matrices can be seen as a model of the spectral variation between bands in terms of the eigen-space with the dominating eigenvalue, eigenvector pair representing the largest component of the background [69]. Targets and anomalies are defined by their infrequent occurrence in the image and therefore can be seen as the pixels which align with the smallest eigenvalues/vector pairs.

Two ubiquitous detection algorithms are considered in this study. The first developed by Reed and Xiaoli, referred to as RX detector (RXD) [15] which utilizes the Mahalanobis distance to discriminate pixel from the BKG. This is an anomaly detector and does not require apriori target information. Another is the constrained energy minimization (CEM) target detector developed by Chang and Harsanyi [67]. This detector requires prior information in the form of a target spectral signature, and discriminates pixels from the background by minimizing the background energy which is not in the direction of the target pixel signature. CEM is a single target detection algorithm and use the more general, multi-target subspace CEM detector known as linearly constrained minimum variance (LCMV) detector [16]. The RXD and LCMV algorithms are representative of the many variations of HSI detectors and this work can be directly extended to those algorithms.

Hyperspectral Anomaly Detection

The RX anomaly detector computes the Mahalanobis distance from the background for each pixel $\mathbf{r}_i$ for $1 < i < N$ as in equation (5.1). The background is quantified using the sample covariance matrix $\mathbf{K} = 1/(N-1)[\sum_{i=1}^{N}(\mathbf{r}_i - \boldsymbol{\mu})(\mathbf{r}_i - \boldsymbol{\mu})^T] \in \Re^{LxL}$ where $\boldsymbol{\mu}$ is the spectral pixel mean computed using $1/N[\sum_{i=1}^{N}\mathbf{r}_i]$. The sample covariance matrix quantifies the variance between each spectral band.

$$\delta_{RXD}(\mathbf{r}_i) = (\mathbf{r}_i - \boldsymbol{\mu})^T \mathbf{K}^{-1}(\mathbf{r}_i - \boldsymbol{\mu}) \tag{5.1}$$

The key insight to applying CS to RXD is that the samples can be sensed within the band using the models in (2.2), (2.3) or (2.4). Mixing the information between the samples within the band using CS does not affect the computation of the sample covariance matrix. This was proven in Chapter 3 using the BV, BT and BT-SVD models of CS. Therefore, it is advantageous to compress the bands, transforming them into CSSD, then compute the covariance matrix. This reduces the computational cost of finding the covariance while generating the same inter-band statistics. Therefore, $\mathbf{K}^{-1}$ in (5.1) can be replaced with the covariance in CSSD, $\mathbf{K}^{-1} \approx \mathbf{K}_{\boldsymbol{\Phi}}^{-1} \in \Re^{LxL}$.

In addition to using the covariance computed using the compressed samples, the spectral pixels can be compressed using (2.1) with little effect during the computation of the RXD values. Using CSBD in (2.1) imposes CS on both the covariance $\boldsymbol{\Phi}_{k_L}\mathbf{K}_{\boldsymbol{\Phi}}\boldsymbol{\Phi}_{k_L}^T$ and the data samples $\left(\boldsymbol{\Phi}_{k_L}\mathbf{r}_i - \boldsymbol{\Phi}_{k_L}\boldsymbol{\mu}\right)$. Substituting both of these into (5.1) results in CS-RXD in (5.2).

$$\delta_{CS-RXD}(\mathbf{r}_i) = \left(\boldsymbol{\Phi}_{k_L}(\mathbf{r}_i - \boldsymbol{\mu})\right)^T \left(\boldsymbol{\Phi}_{k_L}\mathbf{K}_{\boldsymbol{\Phi}}\boldsymbol{\Phi}_{k_L}^T\right)^{-1} \left(\boldsymbol{\Phi}_{k_L}(\mathbf{r}_i - \boldsymbol{\mu})\right) \tag{5.2}$$

Considering the spectral decomposition of the positive semi-definite covariance matrix $\mathbf{K_\Phi} = \mathbf{U\Lambda U}^T$ such that $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_L]$ are the eigenvectors with $\mathbf{\Lambda} = diag(\lambda_1, \lambda_2, \ldots, \lambda_L)$ eigenvalues the result in (5.3) is found.

$$\left(\mathbf{\Phi}_{k_L}\mathbf{K_\Phi}\mathbf{\Phi}_{k_L}^T\right)^{-1} = \left(\mathbf{\Phi}_{k_L}\mathbf{U\Lambda U}^T\mathbf{\Phi}_{k_L}^T\right)^{-1} \tag{5.3}$$

This can be interpreted using the BT-SVD model showing that when compressing a 2-tensor the eigenvectors are sensed and the eigenvalues are unaffected. Therefore, the inverse of $\left(\mathbf{\Phi}_{k_L}\mathbf{K_\Phi}\mathbf{\Phi}_{k_L}^T\right)^{-1}$ is given in equation (5.4).

$$\left(\mathbf{\Phi}_{k_L}\mathbf{U}\right)\mathbf{\Lambda}^{-1}\left(\mathbf{U}^T\mathbf{\Phi}_{k_L}^T\right) = \mathbf{\Phi}_{k_L}\mathbf{K}_\mathbf{\Phi}^{-1}\mathbf{\Phi}_{k_L}^T \tag{5.4}$$

Using the covariance inverse of (5.4), the matrix transpose rules and the CS-RXD form in (5.2) RXD is augmented resulting in the following final form.

$$\delta_{CS-RXD}(\mathbf{r}_i) = (\mathbf{r}_i - \boldsymbol{\mu})^T(\mathbf{\Phi}_{k_L}^T\mathbf{\Phi}_{k_L})\mathbf{K}_\mathbf{\Phi}^{-1}(\mathbf{\Phi}_{k_L}^T\mathbf{\Phi}_{k_L})(\mathbf{r}_i - \boldsymbol{\mu}) \approx \delta_{RXD}(\mathbf{r}_i) \tag{5.5}$$

With $\mathbf{\Phi}_{k_L}^T\mathbf{\Phi}_{k_L} = \mathbf{I} + \mathbf{E}$ and $\mathbf{E} \to \mathbf{0}$ as $m_L \to L$ and this implies that $\delta_{\mathrm{CS-RXD}} \to \delta_{\mathrm{RXD}}$ as $m_{n_1} \to n_1$, $m_{n_2} \to n_2$ and $m_L \to L$. Under sufficient CSSR for each spatial and spectral mode of the image RXD using data in CSBD will converge to CS-RXD. Note that this convergence is exponential decaying as the dimensionality increases according to Fig. 1. Therefore, CS-RXD can be implemented by computing the sample correlation from the sensed samples in each band. This is followed by using the spectrally sensed PVs and further applying $\mathbf{\Phi}_{k_L}$ to the covariance matrix.

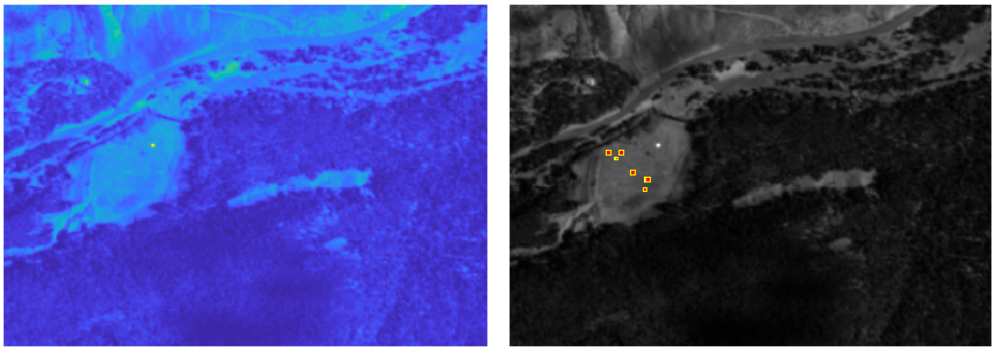
Real Hyperspectral Test Images

The accuracy and runtime reduction of CS-RXD is demonstrated using two, well-known data sets of HYDICE and HyMap collected using the AVIRIS sensor which are

depicted in Fig. 12 and Fig. 21 respectively. The HYDICE and HyMap images are used for testing detection algorithms due to the availability of accurate ground truth information. Accuracy is quantified via comparison of RXD and CS-RXD results while timing is assessed by measuring runtime during covariance and correlation computation as CSSR increases.

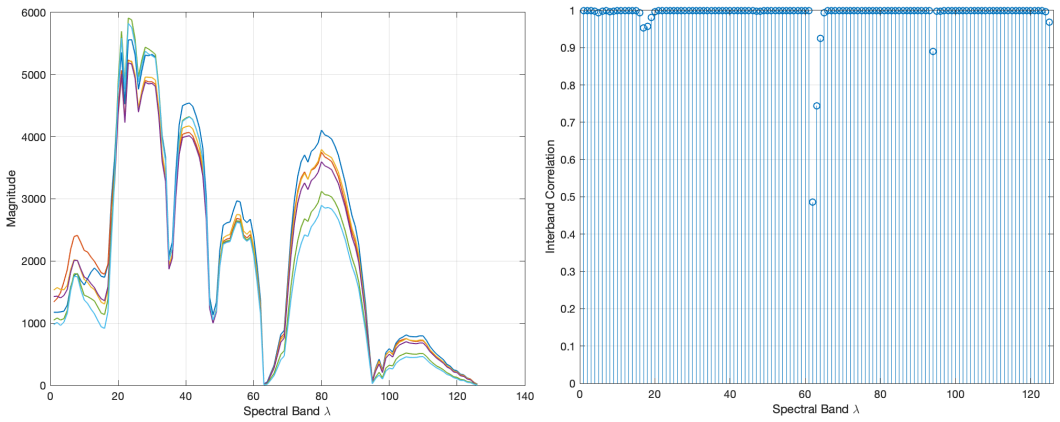
HyMap Data Set

The HyMap image is much larger compared to the HYDICE scene and is used to show the scale of the compute time savings when estimating the sample correlation and covariance matrices [8]. The image was collected with a HyMap sensor over Cooke City, Montana USA in July 2006. The original image is 260 x 800 pixels with 126 bands with 0.45 – 2.5 μm spectral bandwidth and 2 m GSD. This study used half of the image of 260 x 400 pixels which contain four targets of varying materials. The targets are labeled F1 through F4, where F1 is a 3 meter red cotton material, F2 is a 3 meter yellow nylon target, F3 is a 1 meter by 2 meter blue cotton target and F4 is a 1 meter by 2 meter red nylon target



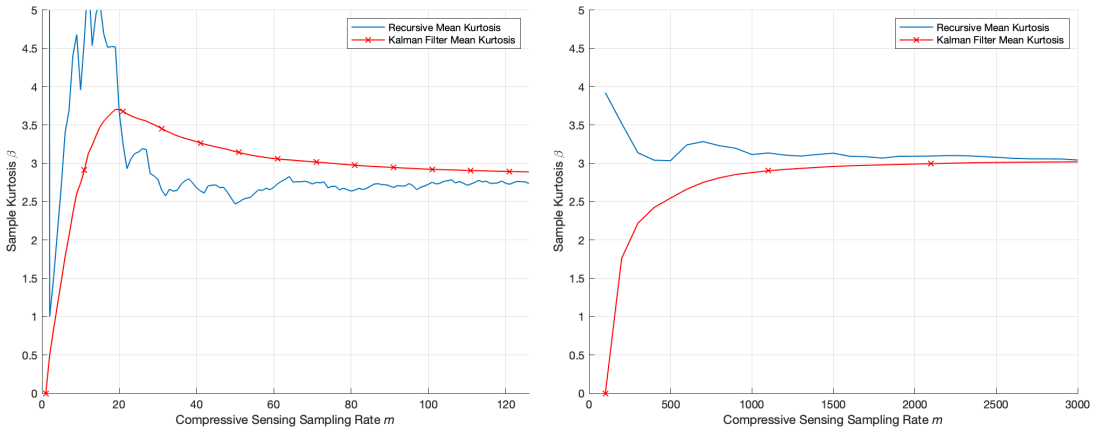
(a) (b)

Fig. 21. HyMap hyperspectral test image and ground truth



(a) (b)

Fig. 22. HyMap scene (a) 6 fabric spectral signatures (b) inter-band correlation



(a) (b)

Fig. 23. HyMap CSSR convergence (a) PV plot and (b) BV plot

all laid out in a field surrounded by mountainous relief. Band 80 of the scene is rendered in Fig. 21 with red ground truth pixels and yellow pixels signifying the surrounding area. Fig. 22(a) plots show the 6 fabric spectral signatures used for target detection algorithms and Fig. 22(b) showing the inter-band correlation. As with the HYDICE image SCS-RK results are plotted in Fig. 23 and Fig. 24 to estimate CSBR and CSSR of the HSI modes when constructing the CS sampling matrices by determining the point of convergence.

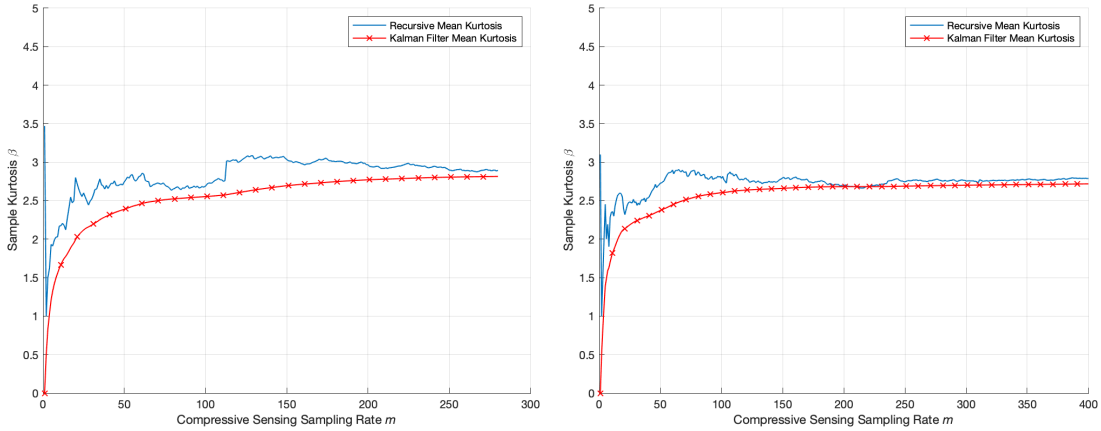


Fig. 24. HyMap CSSR convergence of BT (a) columns and (b) rows

Anomaly Detection Empirical Results

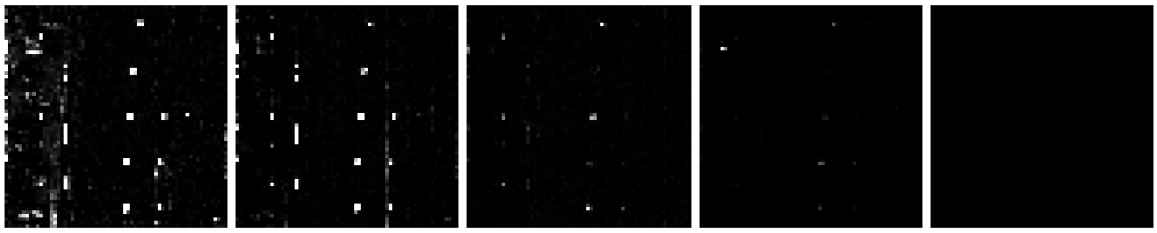
To verify the accuracy of CS-RXD, the output of the detector from (5.5) is computed and the absolute difference is taken between RXD using ODS and the data in CSSD as in equations (5.6) for each pixel location. Each pixel location is assigned a scalar detection value by RXD and the differences are rendered in a 2D image showing convergence as CSBR and CSSR improves.

$$|\delta_{RXD}(\mathbf{r}_i) - \delta_{CS-RXD}(\mathbf{r}_i)| \text{ for } 1 \leq i \leq N \tag{5.6}$$

In addition to the differences at each individual pixel location, an aggregate measure using the sum of squared errors is used across a range of spectral and spatial CSSR as specified in equation (5.7). This metric gives a scalar estimate of convergence at a specific CSSR.

$$\sum_{i=1}^{N} \left(\delta_{RXD}(\mathbf{r}_i) - \delta_{CS-RXD}(\mathbf{r}_i)\right)^2 \tag{5.7}$$

The first experiment implements RXD on the HYDICE data. This experiment is performed in three variations. The first is using the BV spatial representation



(a) (b) (c) (d) (e)

Fig. 25. RXD difference images for HYDICE scene using CS applied to BV

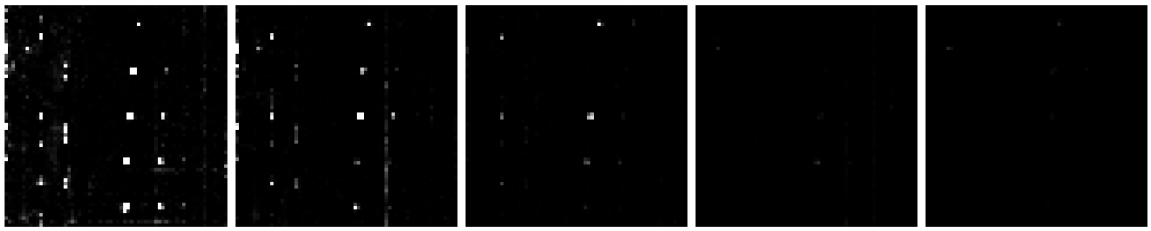
of (2.4) to compress the bands. These sensed bands are then used to compute the covariance matrix $\mathbf{K}_{\mathbf{\Phi}} = \mathbf{X}_{\mathbf{\Phi}}^{T}\mathbf{X}_{\mathbf{\Phi}} \in \Re^{LxL}$ assuming mean centered pixels. The image is 64 x 64 pixels and CSSR for the BV of the HYDICE data has range of $m_N = [1,4096]$. The CSSR has a range of $m_L = [1,169]$. To demonstrate the convergence of the detector output as CSBR and CSSR increases, the error in (5.6) is computed in Fig. 25.

Here CSBR and CSSR was varied from (a) $(m_N, m_L) = (64,16)$, (b) $(m_N, m_L) = (256,32)$, (c) $(m_N, m_L) = (512,64)$, (d) $(m_N, m_L) = (1024,128)$, (e) $(m_N, m_L) = (2048,169)$ out of 4096 spatial samples and 169 bands. The difference images are plotted on a common dynamic range and show the convergence of $\delta_{CS-RXD}(\mathbf{r}_i) \rightarrow \delta_{RXD}(\mathbf{r}_i)$ when using the BV model validating the preservation of the covariance in (3.3) and the CS-RXD in (5.5).

The experiment from Fig. 25 was repeated using the BT model from (2.2) applying CS to the BTs which are then used to compute the sample covariance as in (3.7). Again the CSBR of the bands varied from $m_L = [1,169]$ and CSSR for each spatial mode

(a) (b) (c) (d) (e)

Fig. 26. RXD difference images for HYDICE scene using CS applied to BT

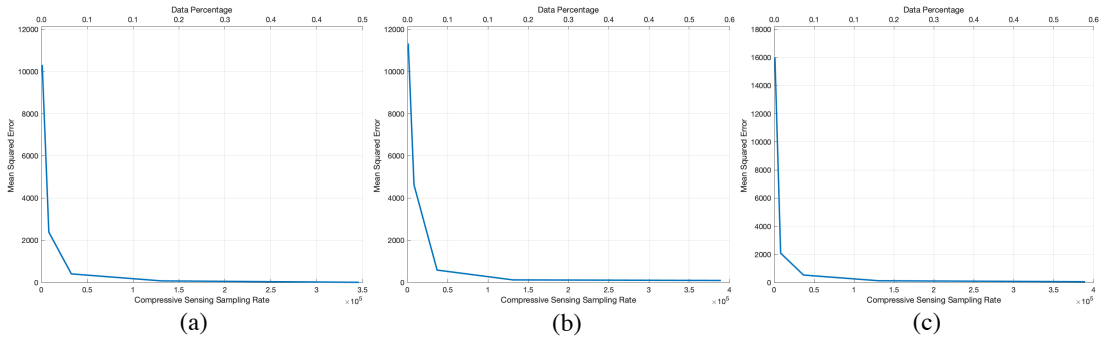


(a) (b) (c) (d) (e)

Fig. 27. RXD difference images for HYDICE scene using CS applied to BT-SVD

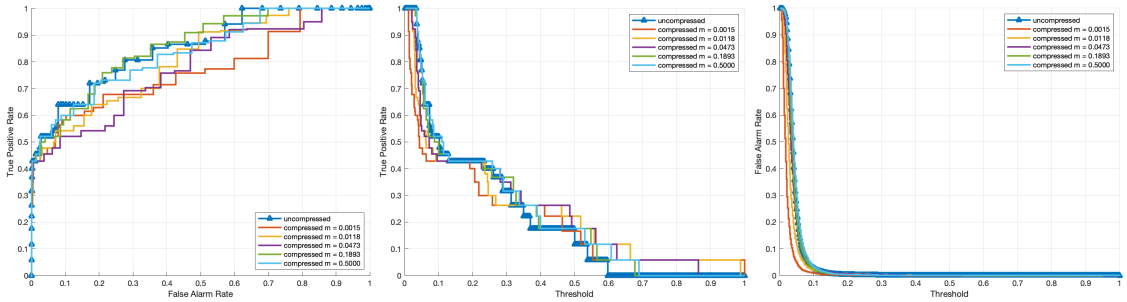
varied from $n_1, n_2 = [1,64]$. The result is in Fig. 26 for (a) $(m_{n_1}, m_{n_2}, m_L) = (8,8,16)$, (b) $(m_{n_1}, m_{n_2}, m_L) = (16,16,32)$, (c) $(m_{n_1}, m_{n_2}, m_L) = (24,24,64)$, (d) $(m_{n_1}, m_{n_2}, m_L) = (32,32,128)$ and (e) $(m_{n_1}, m_{n_2}, m_L) = (48,48,169)$ from 64x64 pixel tensor. The convergence of the detector output is evident from the difference images. This validates using the BT model to compute the covariance matrix using (3.12) and again the CS-RXD of (5.5).

Another RXD experiment using HYDICE applies CS to the singular vector of the BTs using SVD as in (2.3). This approach compresses the singular vectors of SVD and the result is in Fig. 27. As before, CSSR varied from (a) $(m_{n_1}, m_{n_2}, m_L) = (8,8,16)$, (b) $(m_{n_1}, m_{n_2}, m_L) = (16,16,32)$, (c) $(m_{n_1}, m_{n_2}, m_L) = (24,24,64)$, (d) $(m_{n_1}, m_{n_2}, m_L) = (32,32,128)$ and (e) $(m_{n_1}, m_{n_2}, m_L) = (48,48,169)$ from 64 x 64 pixel tensor. This experiment further validates using the BT-SVD model to compute the covariance matrix using (3.12) and again the detector output convergence in



(a) (b) (c)

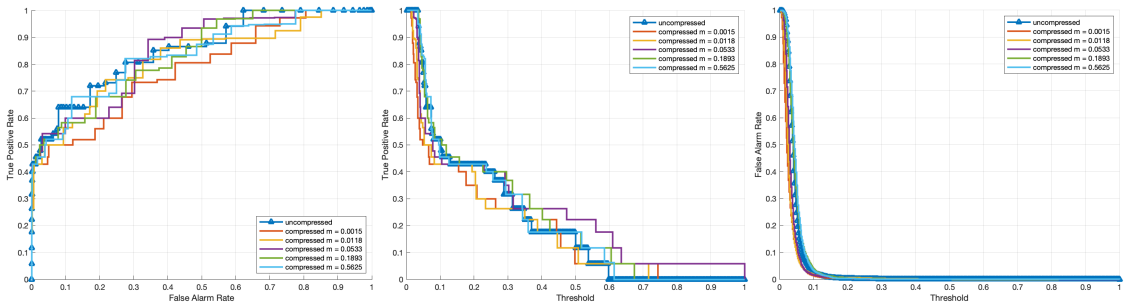
Fig. 28. SSE plots of the RXD difference images using HYDICE scene



(a) (b) (c)

Fig. 29. Detection performance resulting using HYDICE and RXD with BV

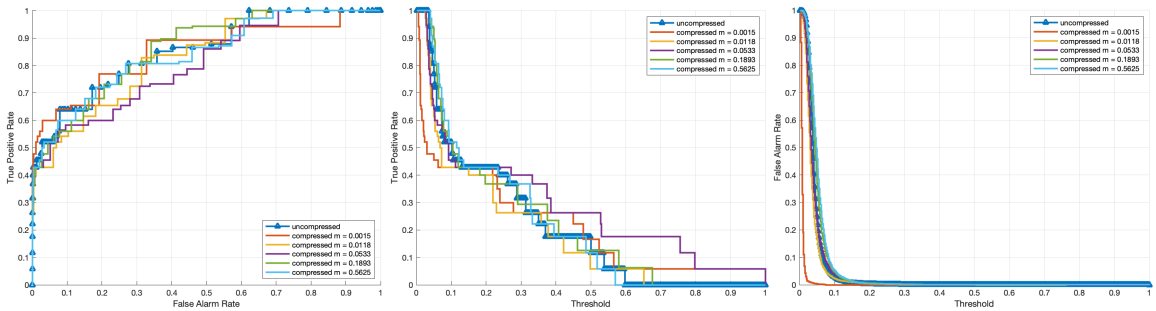
(5.5). The aggregate SSE metric in (5.7) is plotted in Fig. 28. These three graphs for (a) BV, (b), BT and (c) BT-SVD CS model demonstrate the exponential convergence of the detector output across the images for varying CSSR. Notice the dramatic reduction in size achievable while maintaining accuracy. Less than 20% of the data is required to see close



(a) (b) (c)

Fig. 30. Detection performance using HYDICE and RXD with BT at varying CSSR

to identical results in all cases. The ground truth of the HYDICE scene is depicted in Fig. 12(b). Using the ground truth pixel locations, the receiver-operating characteristic (ROC) curve is plotted by varying CSBR and CSSR at each variation of band CS model [3]. The



(a) (b) (c)

Fig. 31. Detection performance using HYDICE and RXD with BT-SVD

Table 2. AUC of ROC curves for varying CSSR for HYDICE and RXD

	1	2	3	4	5	No CS
BV	0.7708	0.8207	0.8317	0.8287	0.8540	0.8590
BT	0.7944	0.8099	0.7895	0.8468	0.8485	0.8590
BT-SVD	0.7559	0.7876	0.8143	0.8304	0.8496	0.8590

results of CS-RXD using CS applied to the BVs when computing the covariance matrix is in Fig. 29 with (a) ROC curve, (b) the true positive rate (TPR) curve and (c) the false positive rate (FPR) curve. Careful observation of these curves show convergence as CSSR increases to the curve computed using ODS. The performance rates for CS applied using the BT model from (2.2) are plotted in Fig. 30 showing the convergence of the detection rates as CSBR and CSSR increases. The other mode of applying CS to the bands is the BT-SVD model (2.3). The detection performance using the BTSVD model is in Fig. 31 showing again the convergence of the detection performance as CSBR and CSSR improve.

Each of the ROC curves validates the detection performance of CS-RXD using compressed samples to estimate the sample covariance matrix. Another measure derived from the ROC curves is the integral area under the curves (AUC) [80]. Each plot in Fig. 29(a), Fig. 30(a), and Fig. 31(a) has 6 ROC curves. The ROC using data in ODS along with 5 other CSSR of increasing size. Table 2 details the AUC and convergence of the detection power as CSSR increases towards the full data size.

A second data set used to validate the accuracy and runtime of CS-RXD is the HyMap image from Fig. 21. The first experiment computes the absolute difference in the detector output for each pixel location from equation (5.6) as CSBR and CSSR increases. The spatial band CSSR has range of $m_N = [1,80000]$ and the spectral CSBR range is $m_L = [1,126]$.

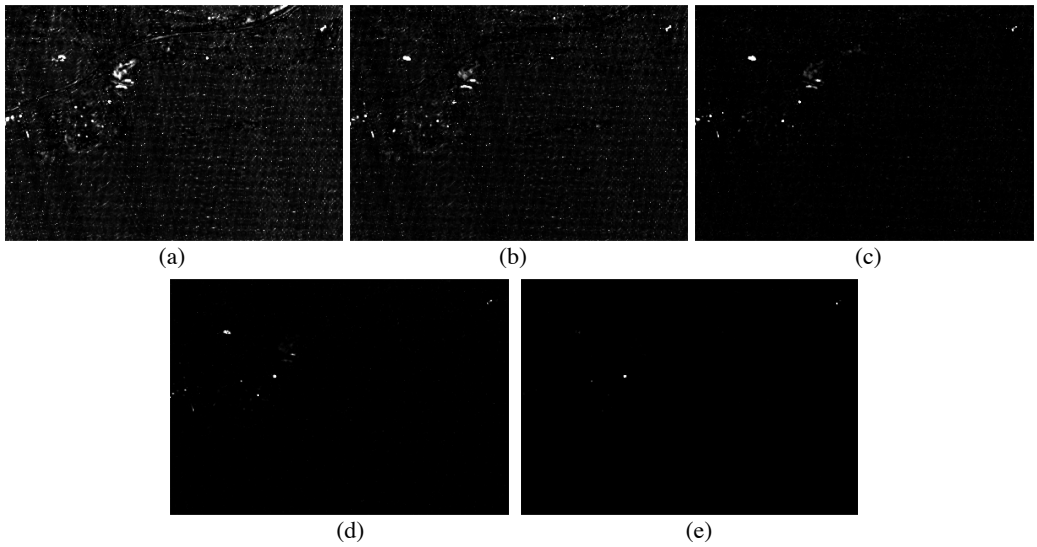


Fig. 32. RXD difference images for HyMap scene using CS applied to BV

Fig. 32 renders the error of each pixel on a common dynamic range for an increasing CSBR and CSSR of (a) $(m_N, m_L) = (64,16)$, (b) $(m_N, m_L) = (256,32)$, (c) $(m_N, m_L) = (512,64)$, (d) $(m_N, m_L) = (1024,100)$, and (e) $(m_N, m_L) = (2048,126)$. The error is decreasing as CSSR increases.

The prior experiment was repeated using the BT model of CS applied in (2.2) and the compressed spatial bands are used to estimate the sample covariance matrix. The CSSR

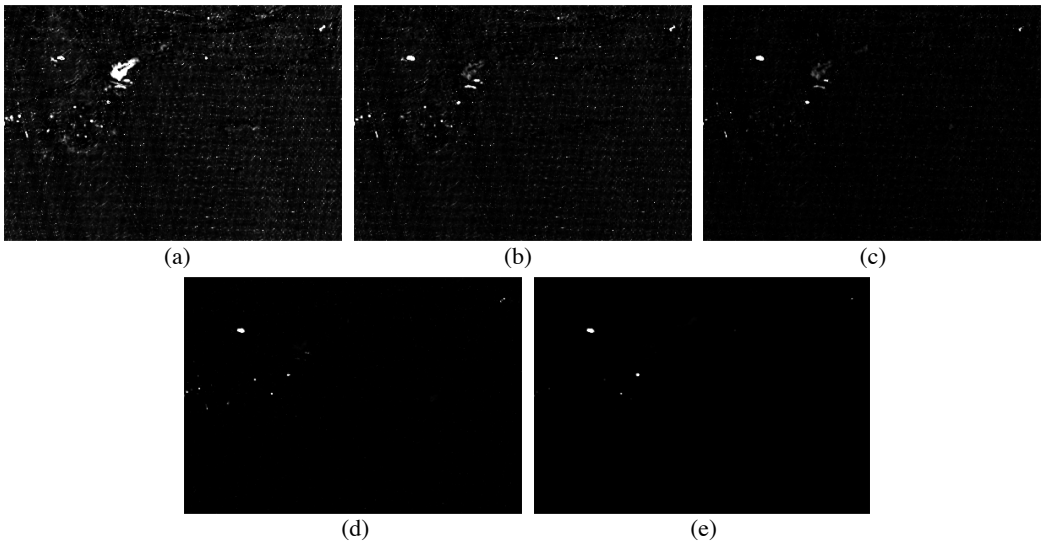


Fig. 33. RXD difference images for HyMap scene using CS applied to BT

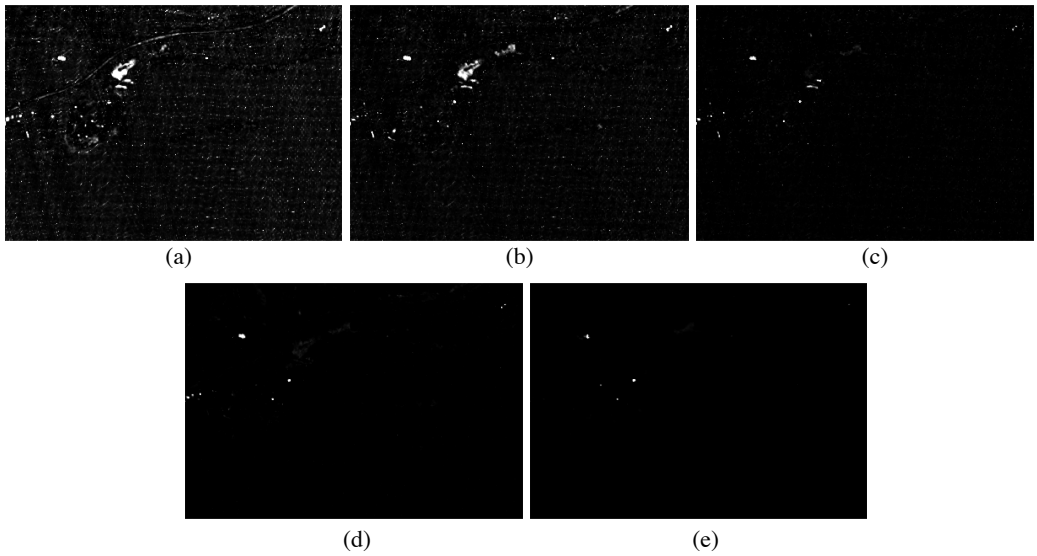


Fig. 34. RXD difference images for HyMap scene using CS applied to BT-SVD

range for the BT modes is $m_{n_1} = [1,200]$ and $m_{n_2} = [1,400]$ and the spectral range is $m_L = [1,126]$. The error in the detector output for each pixel location is rendered in Fig. 33 and the images are at (a) $(m_{n_1}, m_{n_2}, m_L) = (16,16,16)$, (b) $(m_{n_1}, m_{n_2}, m_L) = (32,32,32)$, (c) $(m_{n_1}, m_{n_2}, m_L) = (64,64,64)$, (d) $(m_{n_1}, m_{n_2}, m_L) = (96,96,100)$, and (e) $(m_{n_1}, m_{n_2}, m_L) = (128,128,126)$. The error is decreasing as the CSSR increases. Next, the BT-SVD model is applied to the bands as in (2.3) and the compressed bands are used to estimate the sample covariance matrix as from (3.12). CSSR is incremented using the same rates as done in the previous experiment and results are shown in Fig. 34. The

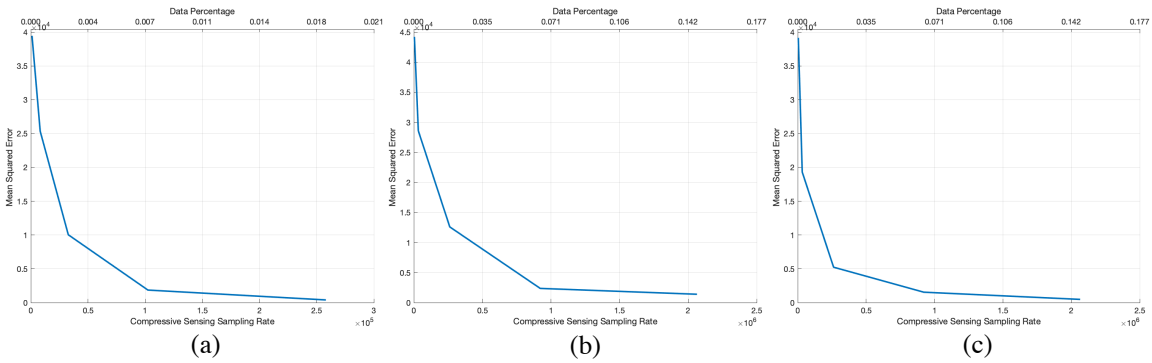


Fig. 35. SSE plots of the RXD difference images using HyMap scene

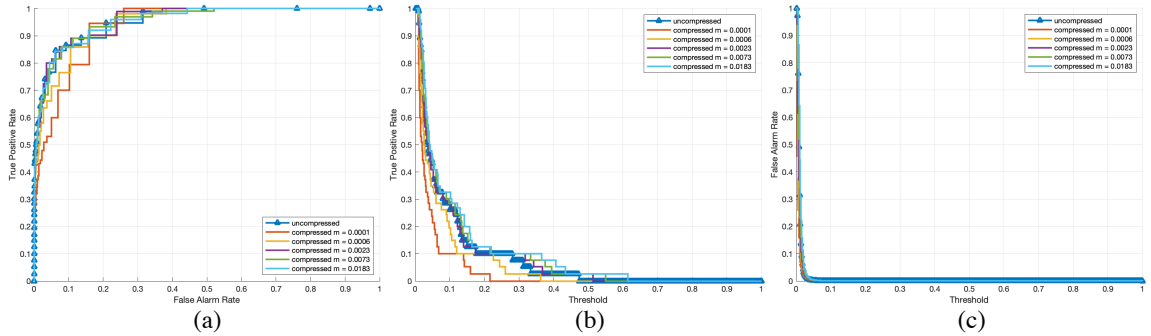


Fig. 36. Detection performance resulting using HyMap and RXD with BV

observed trend is the decreasing error in each pixel as CSSR increases. In addition, the aggregate SSE plotted in Fig. 35 shows an exponential decay in error with increasing CSSR. Also note the percentage of data used from the image is less than 2.1% resulting in a dramatic reduction in data while maintaining the detector output.

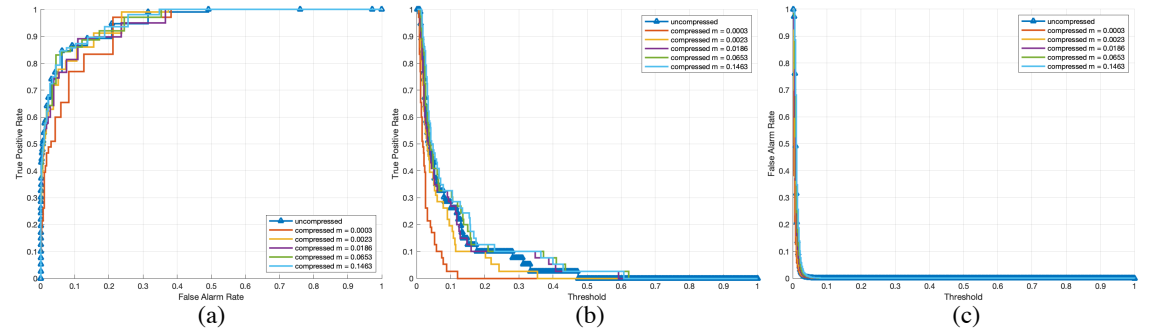


Fig. 37. Detection performance resulting using HyMap and RXD with BT

The ground truth for the HyMap image is in Fig. 21 and shows the pixel locations of the 6 fabric signatures. Using the ground truth, the ROC are plotted using the BV, BT

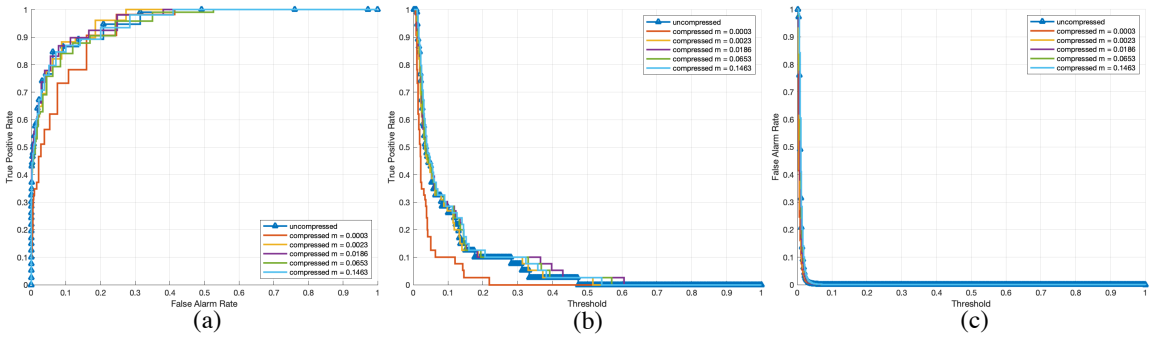


Fig. 38. Detection performance resulting using HyMap and RXD with BTSVD

and BT-SVD CS models when computing the covariance matrices. Fig. 36 has the plots for the (a) ROC, the (b) true positive rate, and (c) false positive rate using the HyMap image, the BV CS model using CS-RXD. Notice the convergence of the curves as the CSSR increases.

The experiment was again repeated using the BT model when computing the covariance matrix. The performance rates are plotted in terms of the ROC in Fig. 37. The curves show a trend of convergence to the true curve as CSSR increases. Finally, the ROC curves are plotted using the HyMap image and the BT-SVD CS model to compute the covariance matrix. The convergence trend remains when examining Fig. 38. The AUC is also computed for the ROC curves using the BV, BT and BT-SVD when estimating the covariance. The AUC values are plotted in Table 3 at six increasingly large CSSR and the detection power tends towards the power when utilizing samples in ODS.

Table 3. AUC of ROC curves for varying CSSR for HyMap and RXD

	1	2	3	4	5	No CS
BV	0.9344	0.9629	0.9611	0.9640	0.9634	0.9610
BT	0.9590	0.9602	0.9581	0.9624	0.9606	0.9610
BT-SVD	0.9509	0.9584	0.9619	0.9625	0.9640	0.9610

Hyperspectral Target Detection

An HSI target detection algorithm is designed in this section to detect known targets embedded in the sub-pixel spectrum given prior spectral profiles [16], [67]. Let $\{\mathbf{m}_i\}_{i=1}^{p}$ be the set of spectral signatures of the p targets $\{\mathbf{t}_i\}_{i=1}^{p}$ in the HSI $\mathbf{I}$, and $\mathbf{M} = [\mathbf{m}_1, \mathbf{m}_2, \ldots, \mathbf{m}_p] \in \Re^{Lxp}$ be the target signature matrix. Therefore, the detectors attempt to find these known signatures in the image while suppressing the background.

As defined in Chang, 2003, the linearly constrained minimum variance (LCMV) detector finds an optimal FIR filter vector $\mathbf{w}_{\text{LCMV}} \in \Re^{Lx1}$ in (5.8) which minimizes the energy from the background [16]. Here $\mathbf{R} = 1/\text{N}[\sum_{i=1}^{N} \mathbf{r}_i \mathbf{r}^T] \in \Re^{LxL}$ is the sample correlation matrix and $\mathbf{M}$ is the target sample matrix.

$$\min_{\mathbf{w}}\{\mathbf{w}^T\mathbf{R}\mathbf{w}\} \text{ subject to } \mathbf{M}^T\mathbf{w} = \mathbf{1} \tag{5.8}$$

The solution is found using p Lagrange multipliers and is given by equation (5.9) [2].

$$\delta_{LCMV}(\mathbf{r}_i) = \mathbf{r}_i^T\mathbf{R}^{-1}\mathbf{M}(\mathbf{M}^T\mathbf{R}^{-1}\mathbf{M})^{-1}\mathbf{1} \tag{5.9}$$

As with the anomaly detector, applying CS to the band samples has minimal effect on the sample correlation matrix. Therefore, the $\mathbf{R}_{\mathbf{\Phi}}$ can be used as a substitute for $\mathbf{R}$ where the sample correlation matrix is computed using the data in the CSSD. The result from Chapter 3 also implies that $\mathbf{R}_{\mathbf{\Phi}}^{-1} \approx \mathbf{R}^{-1}$ and therefore can be used when deriving $\delta_{LCMV}(\mathbf{r}_i)$ in the CSSD as follows.

$$\delta_{LCMV}(\mathbf{r}_i) \approx \mathbf{r}_i^T\mathbf{R}_{\mathbf{\Phi}}^{-1}\mathbf{M}(\mathbf{M}^T\mathbf{R}_{\mathbf{\Phi}}^{-1}\mathbf{M})^{-1}\mathbf{1} \tag{5.10}$$

While the inter-band statistics are preserved by applying CS to the spatial samples, the LCMV algorithm can also benefit from CS applied to the PVs as in (2.1). Applying the spectral CS sampling matrix to the sample correlation matrix results in $\mathbf{R}_{\mathbf{\Phi}}^{-1'} =$

$\mathbf{\Phi}_{k_L}\mathbf{R}_{\mathbf{\Phi}}^{-1}\mathbf{\Phi}_{k_L}^T$. In addition, the PVs can be sensed using $\mathbf{\Phi}_{k_L}\mathbf{r}_i$ and the target spectral signatures can be compressed forming $\mathbf{\Phi}_{k_L}\mathbf{M}$. Using all three of these sensed forms of the correlation, PVs and spectral signatures, the LCMV detector can be derived using spatial-spectral CS as follows.

$$\delta_{CS-LCMV}(\mathbf{r}_i) = \left(\mathbf{\Phi}_{k_L}\mathbf{r}_i\right)^T \mathbf{R}_{\mathbf{\Phi}}^{-1'}\left(\mathbf{\Phi}_{k_L}\mathbf{M}\right)\left(\left(\mathbf{\Phi}_{k_L}\mathbf{M}\right)^T \mathbf{R}_{\mathbf{\Phi}}^{-1'}\mathbf{\Phi}_{k_L}\mathbf{M}\right)^{-1}\mathbf{1} \tag{5.11}$$

Using the matrix transpose rules, and substituting $\mathbf{R}_{\mathbf{\Phi}}^{-1'} = \mathbf{\Phi}_{k_L}\mathbf{R}_{\mathbf{\Phi}}^{-1}\mathbf{\Phi}_{k_L}^T$ the following equality is found of $\delta_{CS-LCMV}(\mathbf{r}_i) =$

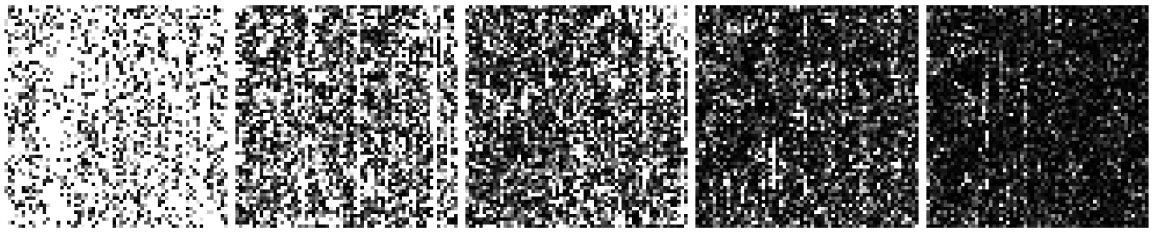
$$\mathbf{r}_i^T\left(\mathbf{\Phi}_{k_L}^T\mathbf{\Phi}_{k_L}\right)\mathbf{R}_{\mathbf{\Phi}}^{-1}\left(\mathbf{\Phi}_{k_L}^T\mathbf{\Phi}_{k_L}\right)\mathbf{M}\left(\mathbf{M}^T\left(\mathbf{\Phi}_{k_L}^T\mathbf{\Phi}_{k_L}\right)\mathbf{R}_{\mathbf{\Phi}}^{-1}\left(\mathbf{\Phi}_{k_L}^T\mathbf{\Phi}_{k_L}\right)\mathbf{M}\right)^{-1}\mathbf{1} \tag{5.12}$$

Using $\mathbf{\Phi}_{k_L}^T\mathbf{\Phi}_{k_L} = \mathbf{I} + \mathbf{E}$ with $\mathbf{E} \to \mathbf{0}$ as $m_L \to L$, equation (5.12) simplifies so that $\delta_{\text{CS-LCMV}} \to \delta_{\text{LCMV}}$ as $m_{n_1} \to n_1$, $m_{n_2} \to n_2$ and $m_L \to L$. Given an adequate CSBR and CSSR for each HSI mode, CS-LCMV will converge to LCMV. This convergence is also exponential as the dimensionality of the sensing matrices increase.

Target Detection Empirical Results

The HYDICE and HyMap datasets in Fig. 12 and Fig. 21 were used to validate CS-LCMV (5.9) when using compressive samples to compute the correlation matrix $\mathbf{R}_{\mathbf{\Phi}} = \mathbf{X}_{\mathbf{\Phi}}^T\mathbf{X}_{\mathbf{\Phi}} \in \Re^{LxL}$. These two datasets have full target ground truth location and the average of these target locations was used to model the target apriori signatures. As was done in the previous section, the absolute output error between LCMV and CS-LCMV for the individual pixels is computed in (5.13) and rendered as CSBR and CSSR increases

$$\left|\delta_{LCMV}(\mathbf{r}_i) - \delta_{CS-LCMV}(\mathbf{r}_i)\right| \text{ for } 1 \le i \le N \tag{5.13}$$



(a) (b) (c) (d) (e)

Fig. 39. LCMV difference images for HYDICE scene using CS applied to BV

to show the convergence of the detector output. In addition, an aggregate measure using the SSE is computed using (5.14) to measure the convergence of the detector output using a single number at each CSBR and CSSR.

$$\sum_{i=1}^{N}\left(\delta_{\mathrm{LCMV}}(\mathbf{r}_i)-\delta_{\mathrm{CS-LCMV}}(\mathbf{r}_i)\right)^2 \tag{5.14}$$

Fig. 39. renders the results of the absolute error of the detector output for each pixel location using the HYDICE image with the BV model for applying CS and the range of CSBR and CSSR from $m_L = [1, 169]$ and $m_N = [1{,}4096]$ respectively. Specifically, the renderings are at CSBR and CSSR of (a) $(m_N, m_L) = (64{,}16)$, (b) $(m_N, m_L) = (256, 32)$, (c) $(m_N, m_L) = (512{,}64)$, (d) $(m_N, m_L) = (1024{,}128)$, (e) $(m_N, m_L) = (2048{,}169)$. The difference images show the convergence of the detector output at CSSR increases. Comparing the difference image of LCMV in Fig. 40 to the difference image computed using CS-RXD in Fig. 25 shows the convergence is slower using LCMV. Fundamentally anomaly detection and target subspace detection are

(a) (b) (c) (d) (e)

Fig. 40. LCMV difference images for HYDICE scene using CS applied to BT

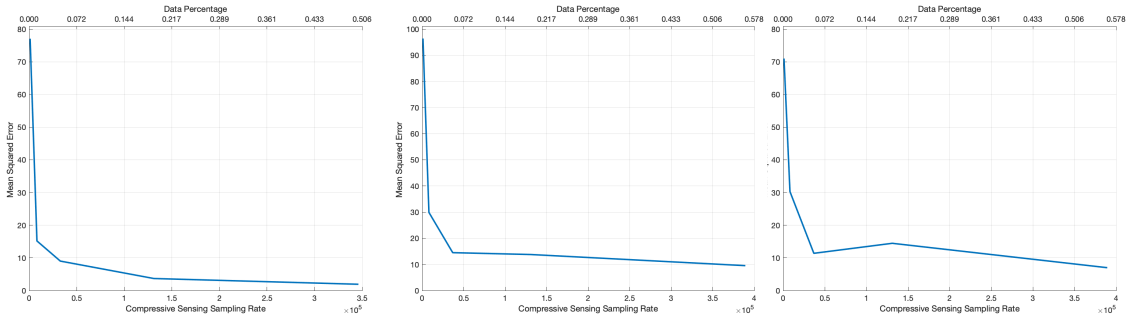
(a) (b) (c) (d) (e)

Fig. 41. LCMV difference images for HYDICE scene using CS applied to BT-SVD

two different problems. LCMV uses known information of five target signatures from Fig. 13 while RXD does not. The weaker convergence of the output is the accumulation of error from the five signatures. More specifically, the error across all five signatures is combined in the detector output. In addition, the dynamic range of the detector is much smaller than RXD leading to images which appear to show a lack of convergence.

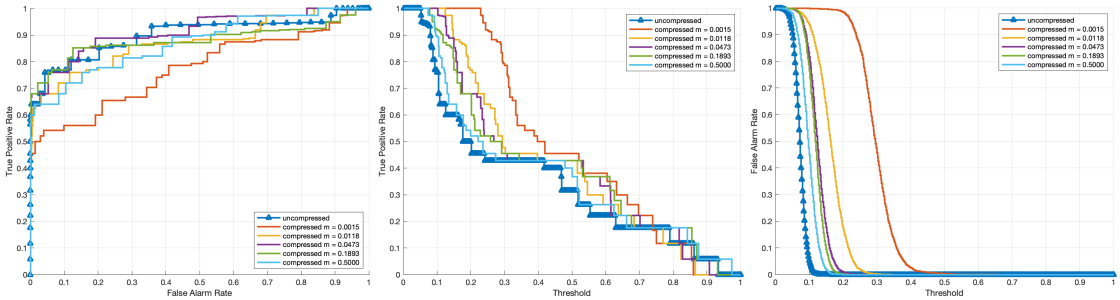
Fig. 40 shows the same experiment and error output of LCMV using the BT model to apply CS to the bands which are then used to estimate the correlation matrix. The trend of convergence is the same from Fig. 39 showing the error from the five targets being detected using LCMV.

The final experiment in Fig. 41 shows the error output of LCMV using the BT-SVD to apply CS to the bands and then used when estimating the sample correlation matrix. The output of this experiment coincides with the other two using HYDICE and LCMV



(a) (b) (c)

Fig. 42. SSE plots of the LCMV difference images using HYDICE scene



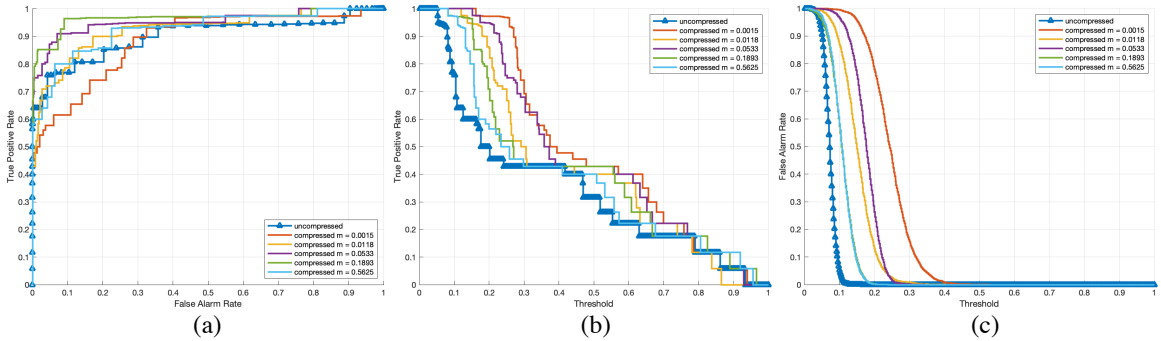
(a) (b) (c)

Fig. 43. Detection performance resulting using HYDICE and LCMV with BV

using compressive samples. The aggregate SSE of the detector error is plotted in Fig. 42 which shows the slower convergence of the error using LCMV with CS and the (a) BV, (b) BT, and (c) BT-SVD CS of the bands.

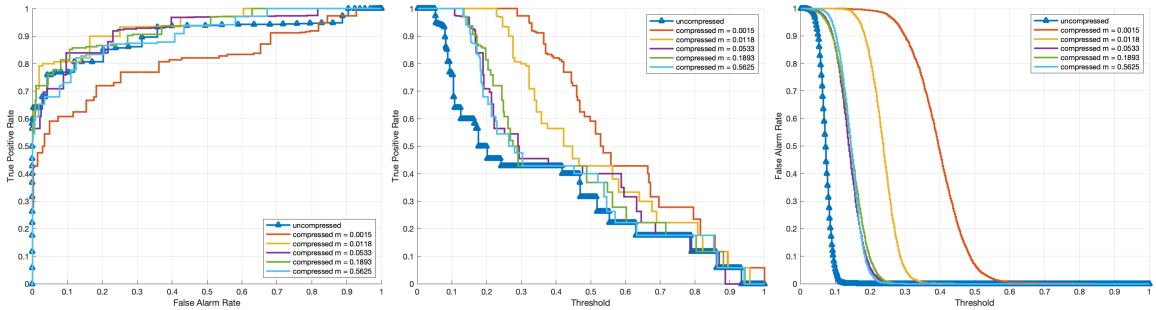
The performance rates of LCMV are also used to verify LCMV accuracy when using CSSD data as compared to using ODS. Fig. 43 shows the (a) ROC, (b) TPR and (c) FPR using the BV model as the CSSR increases. The detection performance converges to that of ODS as CSSR increases.

The detection performance is also evaluated using the BT model and is depicted in Fig. 44 with the (a) ROC, (b) TPR and (c) FPR. These curves show the same trend of convergence using compressed bands to estimate the correlation matrix. Detection performance using the BT-SVD model to compress the bands subsequently used to estimate the correlation matrix are plotted in Fig. 45. The (a) ROC, (b) TPR and (c) FPR



(a) (b) (c)

Fig. 44. Detection performance resulting using HYDICE and LCMV with BT



(a) (b) (c)

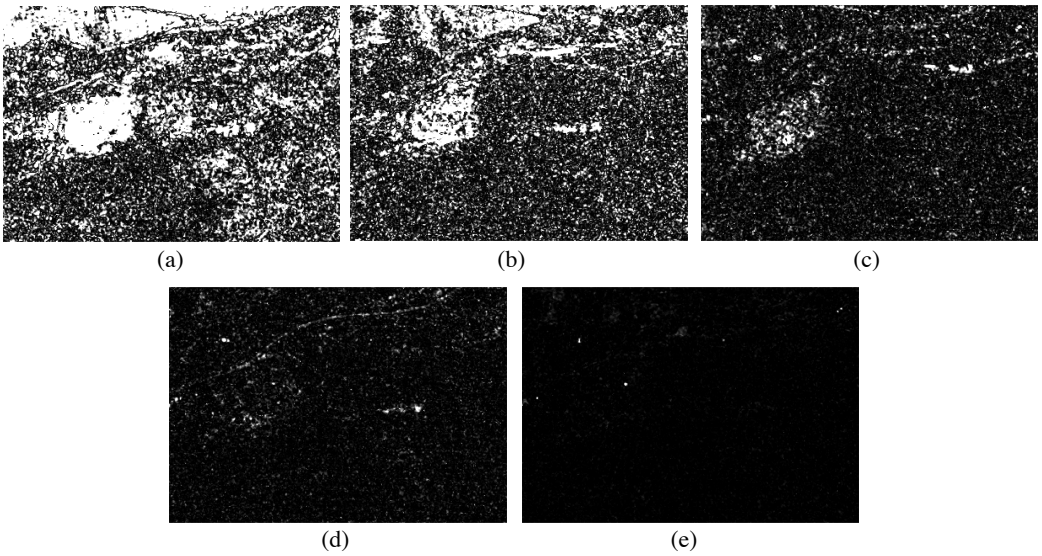
Fig. 45. Detection performance resulting using HYDICE and LCMV with BT-SVD

are all plotted as the threshold increases showing the convergence of the detection rates as CSBR and CSSR increases. The final metric used to quantify LCMV accuracy as CSSR increases is the AUC of ROC curves at the varying CSSR. Shown in Table 4 is the AUC of the ROCs from Fig. 43, Fig. 44, and Fig. 45. The trend of the convergence is less evident using this metric with LCMV compared to RXD due to the accumulated error of the five targets.

Table 4. AUC of ROC curves for varying CSSR for HYDICE and LCMV

	1	2	3	4	5	No CS
BV	0.8051	0.9032	0.8327	0.8582	0.8943	0.9034
BT	0.8364	0.8908	0.8961	0.9550	0.9127	0.9034
BT-SVD	0.8968	0.9349	0.9108	0.8876	0.8868	0.9034

The other dataset used to validate LCMV is the HyMap image from Fig. 21 [8]. This data set has six spectral signatures used for target detection and is described in detail in the prior section. Fig. 46 shows the absolute difference of the detector at each pixel location as CSBR and CSSR increases. This experiment utilized the BV model of CS to compress the bands and then estimate the correlation matrix. Specifically, (a) $(m_N, m_L) = (64,16)$, (b) $(m_N, m_L) = (256,32)$, (c) $(m_N, m_L) = (512,64)$, (d) $(m_N, m_L) = (1024,100)$, and (e) $(m_N, m_L) = (2048,126)$. This shows a convergence

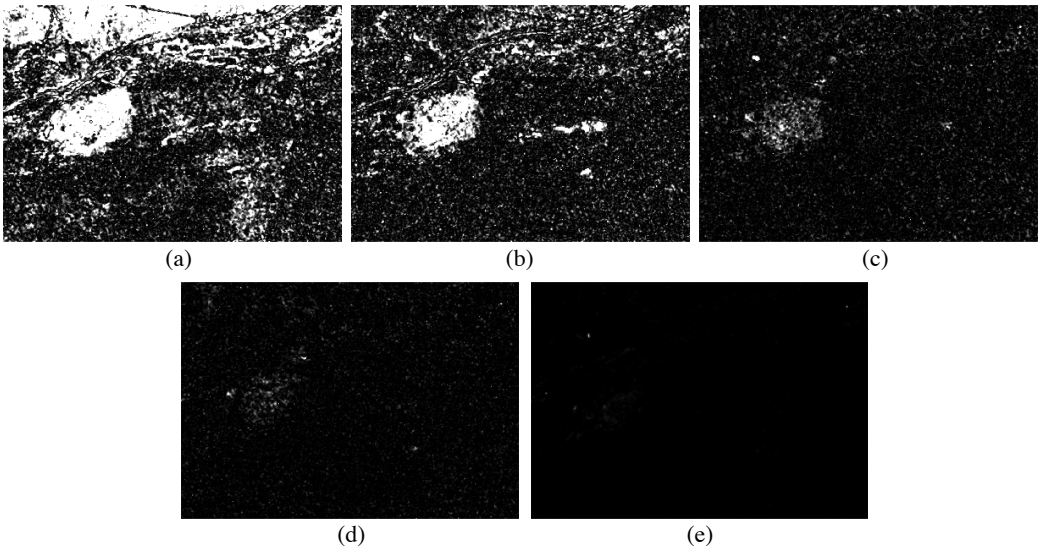


(a) (b) (c)

(d) (e)

Fig. 46. LCMV difference images for HyMap scene using CS applied to BV

of the error as CSBR and CSSR improves. The experiment from Fig. 46 was repeated using the BT model of CS from (2.2) to compress the bands and then estimate the correlation matrix. CSBR and CSSR varied from (a) $(m_{n_1}, m_{n_2}, m_L) = (16,16,16)$, (b) $(m_{n_1}, m_{n_2}, m_L) = (32,32,32)$, (c) $(m_{n_1}, m_{n_2}, m_L) = (64,64,64)$, (d) $(m_{n_1}, m_{n_2}, m_L) = (96,96,100)$, and (e) $(m_{n_1}, m_{n_2}, m_L) = (128,128,126)$. This experiment further validates the used of compressive samples to estimate the correlation



(a) (b) (c)

(d) (e)

Fig. 47. LCMV difference images for HyMap scene using CS applied to BT

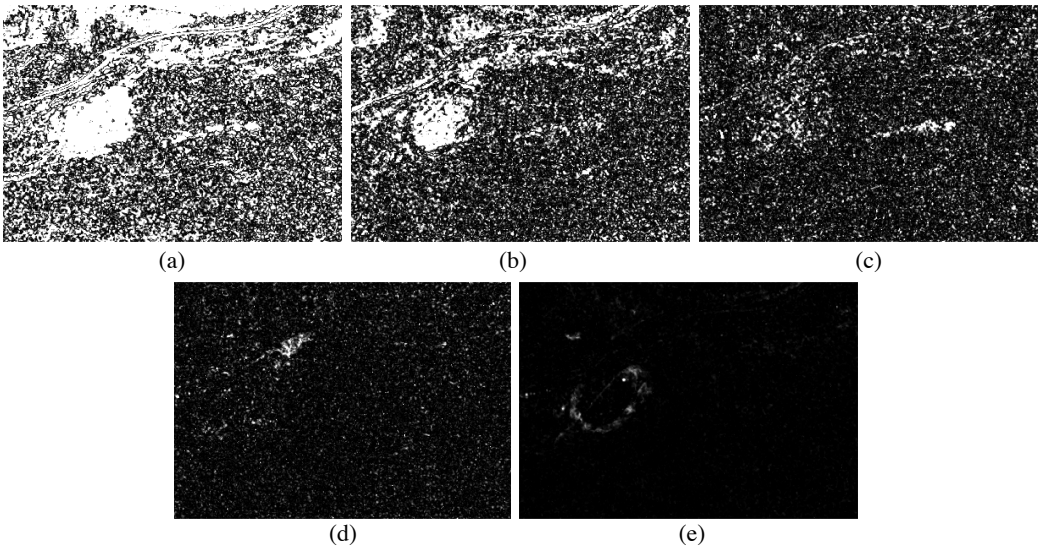


Fig. 48. LCMV difference images for HyMap scene using CS applied to BT-SVD

matrix for LCMV. The final experiment using the HyMap data and LCMV utilized the BT-SVD model of CS to compress the bands before computing the correlation matrix. The absolute difference image in Fig. 48 shows the convergence of the detector output to that of ODS as CSSR increases at the same rates of the last experiment.

The aggregate difference of LCMV using the HyMap dataset is plotted in Fig. 49. These show the SSE between LCMV using data in ODS and the detector using data in CSSD. The convergence here has an exponential trend. Notice the convergence occurs with less than 2% of the data required to

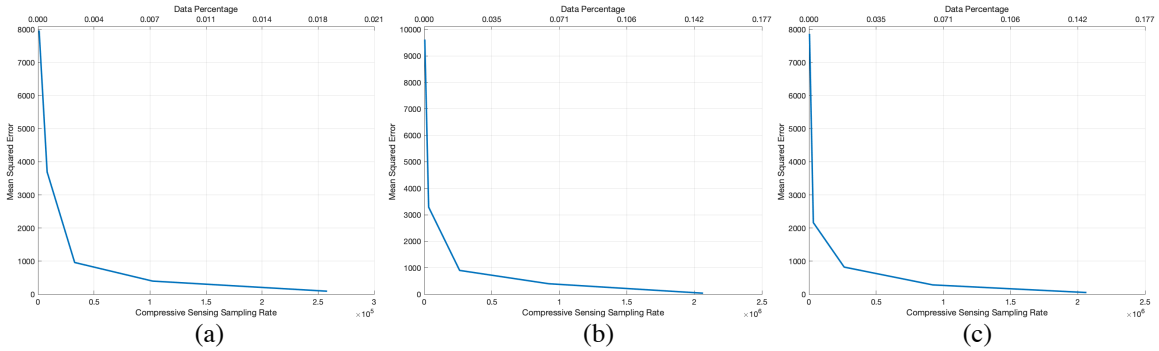
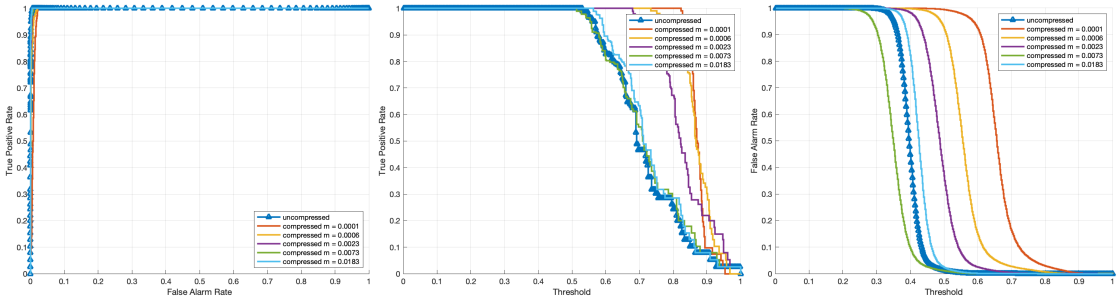


Fig. 49. SSE plots of the LCMV difference images using HyMap scene



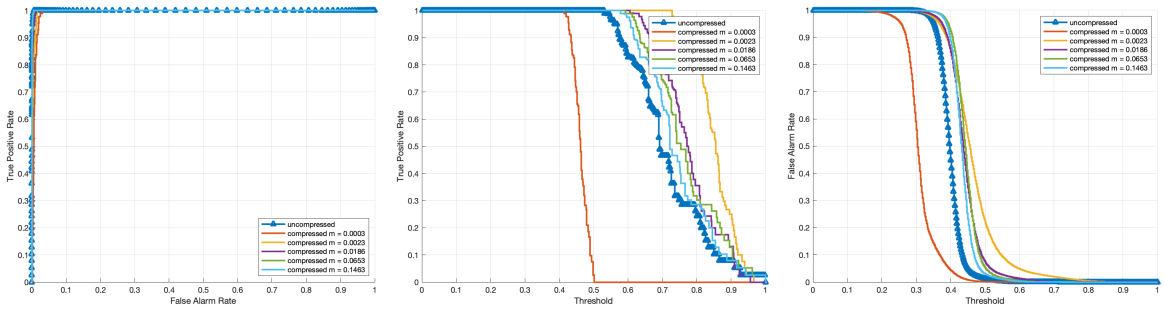
(a) (b) (c)

Fig. 50. Detection performance resulting using HyMap and LCMV with BV

achieve a detector output with a convergence SSE for all pixel spatial locations. The detection performance using CS applied using the BVs which were then used to compute the correlation matrix for CS-LCMV applied to the HyMap dataset is captured in Fig. 50. This figure plots the (a) ROC curves, (b) TPR and (c) FPR as CSBR and CSSR increase over a range of 100 thresholds of the detector output. These graphs show the convergence of the detection algorithm rates as CSSR increases.

Fig. 51 shows the recurring detection performance curves of (a) ROC curves, (b), TPR and (c) FPR using the BT model of CS sensing to assist in computing the sample correlation matrix. The convergence of these curves using CS applied to the BTs compared to that applied to ODS is evident.

The final experiment using CS-LCMV and HyMap data set produced Fig. 52 which shows the performance curves using the BT-SVD model of CS in (a) ROC curves, (b) TPR



(a) (b) (c)

Fig. 51. Detection performance resulting using HyMap and LCMV with BT

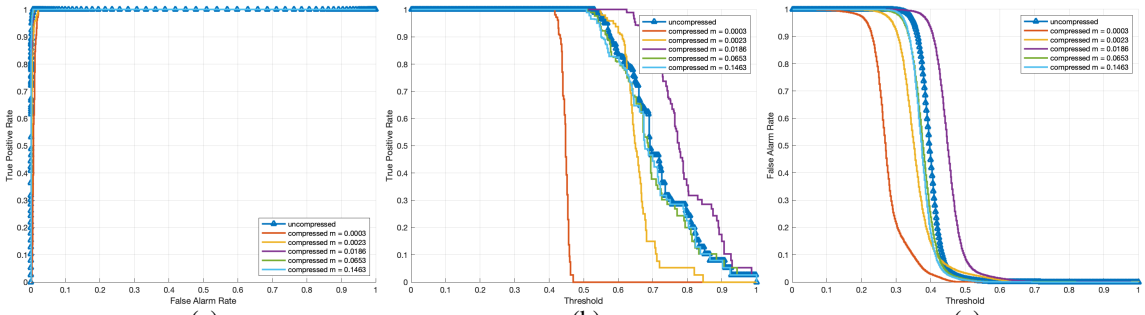


Fig. 52. Detection performance resulting using HyMap and LCMV with BT-SVD

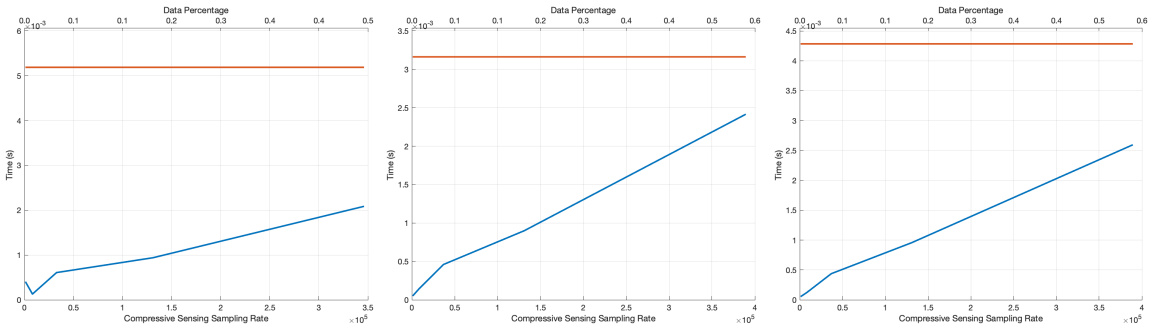
and (c) FPR. The convergence trend is apparent here to showing the convergence of the detector output as CSBR and CSSR improves. Table 5 shows the AUC values of the ROC curves from Fig. 50, Fig. 51, and Fig. 52 showing the convergence trend of the detection power as the CSSR increases.

Table 5. AUC of ROC curves for varying CSSR for HyMap and LCMV

	1	2	3	4	5	No CS
BV	0.9804	0.9971	0.9991	0.9995	0.9997	0.9997
BT	0.9864	0.9941	0.9994	0.9995	0.9997	0.9997
BT-SVD	0.9831	0.9968	0.9991	0.9997	0.9997	0.9997

Detection Algorithm Empirical Runtimes

The compute runtime of the covariance and correlation matrices of the HYDICE and HyMap scenes using CS samples are in Fig. 53 and Fig. 54 for (a) the BV, (b) using BT and (c) for BT-SVD at varying CSSR. These plots show the linear runtime relationship between CSSR and covariance and correlation computing times. The accuracy of the detector is extrapolated from the prior difference images, SSE plots and ROC curves in the previous sections and is exponentially decaying while the runtime is linearly increasing.

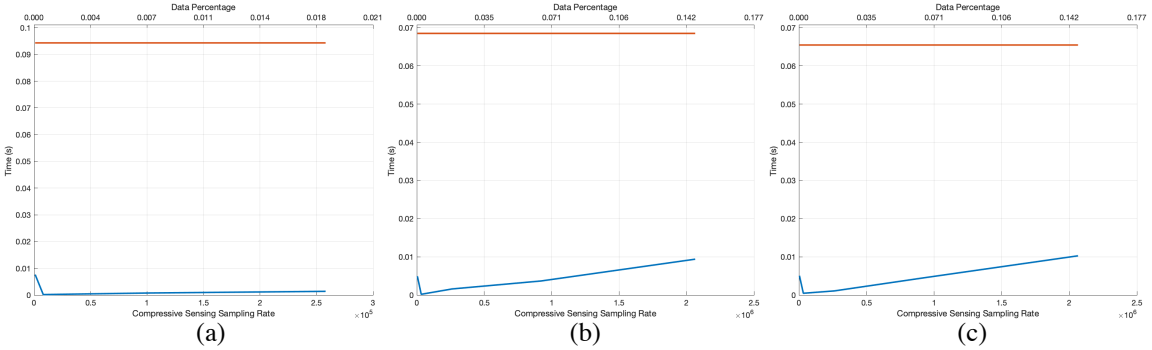


(a) (b) (c)

Fig. 53. The covariance and correlation matrix estimation runtime for HYDICE

This trend highlights the benefit of using compressed band vectors. These times do not include the CS acquisition simulation.

The complexity of each algorithm is dominated by the computation of the covariance and correlation matrices. If assume a compressive sensor, then the acquired samples are already in CSSD. Computing the correlation matrix using the full band data requires L^2 inner products of N multiplication and $N-1$ addition operations. The total number of operations is $(2N-1)\,L^2$. Using the CSBV model, the number of operations is reduced to $(2m_N-1)\,L^2$ operations where $m_N \ll N$. The BT model reduces further to $\left(2m_{n_1}m_{n_2}-1\right)L^2$ with $m_{n_1} \ll n_1$ and $m_{n_2} \ll n_2$. The BT-SVD model is the most costly model with a computational complexity dominated by the $\mathcal{O}(Ln_1n_2^2)$ when computing the SVD of each 2-D band.



(a) (b) (c)

Fig. 54. The covariance and correlation matrix estimation runtime for HyMap

The running time is tracked when computing the correlation and covariance matrices for CS-RXD and CS-LCMV respectively. The results are in Fig. 53 using the HYDICE data and in Fig. 54 for the HyMap data. Each figure shows the increasing linear trend as the CSSR increases. In accordance with CSSR used for ROC computation, CSSR for the BV model is $m_N m_L$ and $m_{n_1} m_{n_2} m_L$ for the BT and BT-SVD models.

Conclusions

This chapter showed that using CS to reduce the size of the bands in hyperspectral images using a band vector, band tensor or band tensor SVD CS models minimally affects the estimation of the sample correlation and covariance matrices. Furthermore, applying compressive sensing to the spectral domain will minimally affect the detector outputs. The experiments conducted showed the convergence of the detector output for each pixel as well as an aggregate SSE error metric for the detector output. These errors decay exponentially as the sampling rate increases implying considerable data reduction using compressive sensing can retain nearly the same performance compared to that using the data in the ODS. This finding was further verified by plotting the performance receiver-operator curves, true positive rates, false positive rates and detection power estimated using the area under the ROC curves. Finally, the computational time of the correlation and covariance matrices was plotted as the sampling rates increased and showed a considerable decrease in computing time with a linearly increasing trend as the sampling rates increased.

Chapter 6: Compressive Sensing Band Selection

Band selection (BS) is another hyperspectral (HSI) application which can benefit from lower computational and memory advantages enabled by applying compressive sensing (CS) to the band samples [81], [82]. Chapter 3 showed the preservation of orthogonal subspace projections (OSP) and subspace volumes in the compressively sensed sample domain (CSSD) using band vectors (BV), band tensor (BT) or band tensor SVD (BT-SVD) sensing models from (2.2), (2.3) and (2.4) detailed in Chapter 2 [19], [83]. This chapter utilizes the preserved metrics to develop HSI BS algorithms in CSSD. BS algorithms select an optimal subset of bands from the image for further processing. This is done by ranking the bands based on importance using band prioritization and then finding the bands which are most distinct from each other using band decorrelation [84]. Band decorrelation relies on inter-band discrimination measures such as correlation, Euclidean distance, spectral angle, spectral information, channel capacity, etc [84], [85]. As was shown in chapter 3, many of these measures can be designed based on the restricted isometric property (RIP), restricted conformal property (RCP) and restricted spectrum property (RSP) to ensure they are preserved in CSSD [13], [39]. Operating on sensed samples mitigates the computation and data intensive problems suffered by most BS optimization algorithms which use the original data space (ODS) where each band can be on the order of mega-bytes in size.

A prominent feature of HSI is large amounts of redundant data per image due to the high number of spectral bands and high spectral correlation. This is a result of the hundreds of bands in a narrow spectral range of the sensors [3]. These contiguous bands exhibit high data redundancy which can be removed by data reduction (DR) algorithms

whose aim is to keep the application specific features for further processing. DR algorithms such as principal component analysis (PCA) [29], maximum noise fraction (MNF) [30] and independent component analysis (ICA) [31] have been explored for HSI data reduction. However, these algorithms project the data into a new data space basis which mixes the spectral and spatial information hindering further interpretation. By applying CS to the individual bands, the samples in a single band can be first reduced without mixing spectral and spatial information. Furthermore, bands can be compared to each other in CSSD using RIP, RCP and RSP principles which preserve the underlying cost function of BS optimization. This maintains the accuracy of BS while utilizing a reduced image.

Hyperspectral Image Band Selection

As explored by Chang 2013 [3] the design of unsupervised BS algorithms must take into account several subproblems. Firstly, the estimation of the band subset size must be addressed. Effective methods to determine the size of the band subspace have been explored and virtual dimensionality (VD) has been shown useful for this purpose [86], [87]. Next, bands are ranked by an importance metric called a band prioritization (BP) measure [84]. BP measures such as signal-to-noise ratio (SNR), variance, entropy or information divergence have been explored in the literature [66], [88]. In addition, the optimal band subset will minimize redundant data. Therefore, a band decorrelation (BD) measure is used to evaluate band subsets [84]. This is any measure comparing multiple bands to determine application fitness [81]. Finally, the set of band subsets is combinatorically large and the cost function is multimodal [82]. This means that an exhaustive search is computationally prohibitive. Therefore, a search strategy must be utilized which efficiently traverses the band subset metric space. Search algorithms will typically select multiple bands

sequentially (SQMBS) using BP and BD measures, or will select and evaluate multiple bands simultaneously (SMMBS) further referred to as band subset selection (BSS) [55], [71]. Sequential algorithms have been proposed in the literature utilizing sequential forward search (SFS) or sequential backward search (SBS) to add bands to a subset or remove them from a subset using BP and BD measures [89]. Simultaneous search strategies in the literature have included particle swarm optimization [90], firefly algorithm [91], band clustering [66], [92], multi-task sparsity pursuit [93], sequential band subset selection (SQ-BSS) and successive band subset selection (SC-BSS) [88], [94]. This study develops both a SQMBS algorithm based on OSP as well as a BSS based on subspace volume. Both metrics are approximately preserved in CSSD.

Using the notation of the previous chapters, let the entire 3-dimensional HSI image be denoted by $\mathbf{I} \in \Re^{n_1 x n_2 x L}$ with $N = n_1 x n_2$ as the spatial dimension with N pixels and L spectral bands. Then let the each band be $\mathbf{B}_l \in \Re^{n_1 x n_2}$ and the full band set of L bands be $\Omega = \{\mathbf{B}_l\}_{l=1}^{L}$ with the selected band set of p bands be $\widetilde{\Omega} = \{\widetilde{\mathbf{B}}_j\}_{j=1}^{p}$. The cost of the selected band subset (SBS) is expressed by the generic objective function $J(\widetilde{\Omega})$. The goal of any BS algorithm is to find the optimal band subset which minimizes or maximizes the cost function as in equation (6.1). There are L choose p combinations resulting in a exponentially large search space of size $L!/(p!\,(L-p)!)$ band subsets to search.

$$\widetilde{\Omega}^* = \arg\left\{\max/\min_{\widetilde{\Omega}\subset\Omega,|\Omega|=p} J(\widetilde{\Omega})\right\} \tag{6.1}$$

A SQMBS algorithm will find an optimal band subset by choosing an initial band set $\widetilde{\Omega}^0$ containing a minimal set of bands typically one or two [55], [72]. Then the algorithm iterates from $1 \le i \le p$ and the i^{th} subsequent iterations of the SQMBS algorithms adds a

single band to the set using the search strategy and inter-band decorrelation measure such as orthogonal subspace projection (OSP) as in equation (6.2).

$$\widetilde{\Omega}^{i+1} = \arg\left\{\max/\min_{\mathbf{B}_l \subset \Omega \setminus \widetilde{\Omega}} J\left(\widetilde{\Omega}^i \mid \mathbf{B}_l\right)\right\} \tag{6.2}$$

In contrast, the BSS algorithms can be seen as a direct application of equation (6.1) which selects and evaluates complete subsets $\widetilde{\Omega}$ where $\left|\widetilde{\Omega}\right| = p$. The largest portion of computing time in (6.1) or (6.2) is evaluating the optimality criteria $J\left(\widetilde{\Omega}\right)$. This function must compute inter-band measures which use 2D band images on the order of mega-bytes. Prior algorithms have used simple measures or pre-processing DR to reduce the computing time of the cost function J [55], [82], [95]. However, these approaches can distort the results by missing important data in ODS. This research takes advantage of the CS acquisition process to reduce the data size while maintaining the accuracy of the result.

Compressive Sensing Sequential Band Selection

The first band selection algorithm developed takes advantage of SFS and OSP in an iterative manner searching for an optimal band subset by selecting bands one at a time sequentially [17], [89]. The bands are first transformed to CSSD and, using the results in Chapter 3, the OSP measure between the sensed bands is approximately preserved [17]. This CS-based SQMBS algorithm, detailed in Algorithm 1, first finds the number of bands to select. The input to the algorithm uses the HFC method to compute the virtual dimensionality (VD) of the image as the number of bands to be selected $p = n_{\mathrm{VD}}$ [96]. The algorithm proceeds by transforming all the band vectors into CSSD using (2.4). Once transformed, the sensed BVs are created in $\left\{\mathbf{\Phi}_{k_N}\mathbf{b}_l\right\}_{l=1}^{L}$. The algorithm proceeds with SFS to select bands one at a time. Therefore, an initial condition band subset is required to begin

the iteration. The bands are all positive vectors and exist in the first quadrant of the L-dimensional cone. Therefore, finding two bands at the extremes of this cone is an effective way to capture bands which are dissimilar. This algorithm implementation chooses the sensed band vector which has the minimal projection length with the smallest L_2-norm. It continues by entering the band selection loop in step 4. Each iteration of the loop will select a single band until $p = n_{\mathrm{VD}}$ bands are selected. The orthogonal projection matrix $\mathrm{P}^{\perp}_{\boldsymbol{\Phi}_{k_N}\mathbf{U}_i}$ is computed by concatenating the selected bands from $\Omega_{BS_i} = \left\{\tilde{\mathbf{b}}_j\right\}_{j=1}^{i}$ into a matrix $\mathbf{U}_i = [\tilde{\mathbf{b}}_1, \tilde{\mathbf{b}}_2, \ldots, \tilde{\mathbf{b}}_i]$. The optimal band to be selected at each iteration is the band which

---

Algorithm 1. Sequential Forward Search Band Selection using Orthogonal Subspace Projection via Compressively Sensed Band Vectors

---

Inputs: a. Number of bands to select found using HFC = $p = n_{VD}$

b. The compressive sensing sampling rate m_N

1. Construct the sampling matrix $\boldsymbol{\Phi}_{\boldsymbol{k_N}}$ using a Gaussian distribution at $m_N x N$ realizations
2. Transform each band from $\{\mathbf{b}_l\}_{l=1}^{L}$ into the compressed domain $\boldsymbol{\Phi}_{k_N}\mathbf{b}_l$ creating a compressed set $\Omega = \left\{\boldsymbol{\Phi}_{k_N}\mathbf{b}_l\right\}_{l=1}^{L}$ of all band vectors
3. Find the initial band vector with the minimal magnitude $\widetilde{\Omega}_{BS_1} = \left\{\boldsymbol{\Phi}_{k_N}\mathbf{b}_1\right\}$ so that $\boldsymbol{\Phi}_{k_N}\tilde{\mathbf{b}}_1 = \min\limits_{1\le l\le L}\left\{\left\|\boldsymbol{\Phi}_{k_N}\mathbf{b}_l\right\|_2^2\right\}$
4. Begin SFS loop, set $i = 1$
   a. Form the projector $\mathbf{P}^{\perp}_{\mathbf{U}_{\boldsymbol{\Phi}_i}} = \mathbf{I} - \mathbf{U}_i(\mathbf{U}_i^T\mathbf{U}_i)^{-1}\mathbf{U}_i^T$ from $\widetilde{\Omega}_{BS_i}$ where $\mathbf{U}_{\boldsymbol{\Phi}_i} = [\boldsymbol{\Phi}_{k_N}\tilde{\mathbf{b}}_1 \mid \boldsymbol{\Phi}_{k_N}\tilde{\mathbf{b}}_2 \mid \ldots \mid \boldsymbol{\Phi}_{k_N}\tilde{\mathbf{b}}_i]$
   b. Choose the next band which has the maximum orthogonal projection error $(\boldsymbol{\Phi}_{k_N}\mathbf{b}_l)^* = \arg\max\limits_{\boldsymbol{\Phi}_{k_N}\mathbf{b}_l\notin\mathbf{U}_i}\left\{\|\mathbf{P}^{\perp}_{\mathbf{U}_i}\boldsymbol{\Phi}_{k_N}\mathbf{b}_l\|_2^2\right\}$
   c. Add the band to selected subset $\widetilde{\Omega}_{BS_{i+1}} = \widetilde{\Omega}_{BS_i} \cup (\boldsymbol{\Phi}_{k_N}\mathbf{b}_l)^*$
   d. Increment the loop variable $i = i + 1$
   e. Test the loop condition, if $i = n_{VD}$ terminate, else go to step 5

Output: Selected band subset $\widetilde{\Omega}_{BS_p} = \left\{\tilde{\mathbf{b}}_j\right\}_{j=1}^{p}$

---

has the maximum orthogonal projection to the selected band subset (SBS) subspace formed by $\mathbf{U}_i$. Thus, OSP is used as a band discrimination or decorrelation measure to find the most unique band at each iteration defined by OSP. In other words, the band which is the worst predictor in the linear combination $\sum_{j=1}^{i} \alpha_j \tilde{\mathbf{b}}_j = \mathbf{b}_l$ where $\mathbf{b}_l \notin \widetilde{\Omega}$ is the most distinct band containing unique information not captures by SBS at the current iteration. The algorithm is terminated when $p = n_{\mathrm{VD}}$ bands are selected in step 9.

Algorithm 2. Sequential Forward Search Band Selection using Orthogonal Subspace Projection via Compressively Sensed Band Tensors

---

Inputs: Number of bands to select found using HFC = $p = n_{VD}$

Compressive sensing sampling rates m_{n_1} and m_{n_2}

1. Construct the random sampling matrices $\mathbf{\Phi}_{k_{n_1}}$and $\mathbf{\Phi}_{k_{n_2}}$ using Gaussian realizations
2. Transform each band tensor $\{\mathbf{B}_l\}_{l=1}^{L}$ into the CSSD using (2.2) or (2.3) creating a compressed band set $\Omega = \left\{\mathbf{\Phi}_{k_{n_1}}\mathbf{B}_l\mathbf{\Phi}_{k_{n_2}}^T\right\}_{l=1}^{L}$
3. Vectorize the compressed tensors using $\left\{\mathrm{vec}(\mathbf{\Phi}_{k_{n_1}}\mathbf{B}_l\mathbf{\Phi}_{k_{n_2}}^T)\right\}_{l=1}^{L}$ and find the initial band $\widetilde{\Omega}_{BS_1} = \left\{\widetilde{\mathbf{B}}_{\mathbf{\Phi}_1}\right\}$ with minimal magnitude $\widetilde{\mathbf{B}}_{\mathbf{\Phi}_1} = \min_{1\le l\le L}\left\{\left\|\mathrm{vec}(\mathbf{\Phi}_{k_{n_1}}\mathbf{B}_l\mathbf{\Phi}_{k_{n_2}}^T)\right\|_2^2\right\}$
4. Begin SFS loop, set $i = 1$
   a. Construct the projector by vectorizing the SBS tensors and computing $\mathbf{P}_{\mathbf{U}_{\mathbf{\Phi}_i}}^{\perp} = \mathbf{I} - \mathbf{U}_{\mathbf{\Phi}_i}\left(\mathbf{U}_{\mathbf{\Phi}_i}^T\mathbf{U}_{\mathbf{\Phi}_i}\right)^{-1}\mathbf{U}_{\mathbf{\Phi}_i}^T$ from $\widetilde{\Omega}_{BS_i}$ where $\mathbf{U}_{\mathbf{\Phi}_i} = \left[\mathrm{vec}\left(\mathbf{\Phi}_{k_{n_1}}\mathbf{B}_1\mathbf{\Phi}_{k_{n_2}}^T\right) | \mathrm{vec}\left(\mathbf{\Phi}_{k_{n_1}}\mathbf{B}_2\mathbf{\Phi}_{k_{n_2}}^T\right) | \ldots | \mathrm{vec}\left(\mathbf{\Phi}_{k_{n_1}}\mathbf{B}_i\mathbf{\Phi}_{k_{n_2}}^T\right)\right]$
   b. Choose the band having maximum OSP error $(\mathbf{\Phi}_{k_{n_1}}\mathbf{B}_l\mathbf{\Phi}_{k_{n_2}}^T)^* = \underset{\mathrm{vec}(\mathbf{\Phi}_{k_{n_1}}\mathbf{B}_l\mathbf{\Phi}_{k_{n_2}}^T)\notin\mathbf{U}_{\mathbf{\Phi}_i}}{\arg\max}\left\{\left\|\mathbf{P}_{\mathbf{\Phi}_k\mathbf{U}_i}^{\perp}\mathbf{vec}\left(\mathbf{\Phi}_{k_{n_1}}\mathbf{B}_l\mathbf{\Phi}_{k_{n_2}}^T\right)\right\|_2^2\right\}$
   c. Add the band to selected subset $\widetilde{\Omega}_{BS_{i+1}} = \widetilde{\Omega}_{BS_i} \cup (\mathbf{\Phi}_{k_{n_1}}\mathbf{B}_l\mathbf{\Phi}_{k_{n_2}}^T)^*$
   d. Increment the loop variable $i = i + 1$
   e. Test the loop condition, if $i = n_{VD}$ terminate, else go to step 5

Output: Selected band subset $\widetilde{\Omega}_{BS_p} = \left\{\widetilde{\mathbf{B}}_j\right\}_{j=1}^{p}$

The CS-based BS algorithm 1 can be extended to use BTs. Like algorithm 1, using tensors takes advantage of OSP and the SFS, but has some important differences. The first being that the algorithm must construct two sampling matrices $\mathbf{\Phi}_{k_{n_1}}$ and $\mathbf{\Phi}_{k_{n_2}}$. The second difference is the vectorization of the compressed BTs in steps 3) and 4.a) in Algorithm 2. Note that computing the compressed bands using CS applied to the BT as in (2.2) or (2.3) will typically be more efficient than BV due to the reduced size BVs as a result of dividing the sparsity into two orthogonal modes.

Compressive Sensing Band Subset Selection

Band subset selection is a simultaneous multiple band selection (SMMBS) selection algorithm which evaluates full band subsets in contrast to the method of SQMBS of the prior section [71], [88], [97]. By choosing a band subset cost function $J(\widetilde{\Omega})$ based on length or angle and then invoking RIP and RCP in (1.1) and (1.2), it is feasible to create algorithms which perform identically in both ODS and CSSD with the advantage of considerably reduced data in CSSD thus reducing the computational complexity.

Using the volume preservation in CSSD detailed in (3.22) of $J(\widetilde{\Omega}_{\mathbf{\Phi}}) \rightarrow J(\widetilde{\Omega})$ for both the BV and BT CS models, a BSS algorithm can be developed using the successive (SC) and sequential (SQ) search algorithms developed by Chang [65], [98], [99]. The algorithm begins by finding the size of the selected band subset using virtual dimensionality (VD) [86]. The HFC method developed by Chang has been used as an effective method to choosing the number of bands for unsupervised band selection [94]. The algorithm continues by choosing an initial band subsets using uniform band selection

(UBS). Then the SQ and SC methods are utilized to evaluate each band subset in turn to find the maximum subspace volume. The sequential CS-BSS procedure is in algorithm 3.

Algorithm 3. Band Subset Selection using Subspace Volume and Sequential Subset Search via Compressively Sensed Band Vectors

Inputs: Number of bands to select found using HFC = $p = n_{VD}$

Compressive sensing sampling rates m_{k_N}

1. Construct the random sampling matrices $\mathbf{\Phi}_{k_N}$ using m_{k_N} Gaussian realizations
2. Let $\widetilde{\Omega}^* = \widetilde{\Omega}^{(0)} = \left\{\tilde{\mathbf{b}}_1^{(0)}, \tilde{\mathbf{b}}_2^{(0)}, \ldots, \tilde{\mathbf{b}}_p^{(0)}\right\}$ be the initial set of band selected from. the band set Ω using uniform band subset selection
3. Simulate compressive sensor using (2.4) to transform all bands to the CSSD $\Omega = \left\{\mathbf{\Phi}_{k_N}\mathbf{b}_l\right\}_{l=1}^{L}$.
4. Outer Loop: Let $l \leftarrow 1$ until l equals L, $l \leftarrow l + 1$
5. Select the l^{th} band $\mathbf{b}_l$ to be the input to the inner loop
   a. Inner Loop: Let $j \leftarrow 1$ until j equals $p \leftarrow p + 1$
   b. Construct $\Omega_j^{(l)} = \left\{\mathbf{\Phi}_{k_N}\tilde{\mathbf{b}}_1^{(l)}, \ldots, \mathbf{\Phi}_{k_N}\tilde{\mathbf{b}}_{j-1}^{(l)}, \mathbf{\Phi}_{k_N}\mathbf{b}_l, \mathbf{\Phi}_{k_N}\tilde{\mathbf{b}}_{j+1}^{(l)}, \ldots, \mathbf{\Phi}_{k_N}\tilde{\mathbf{b}}_p^{(l)}\right\}$ by replacing the j^{th} band with the selected band
   c. Evaluate $J(\widetilde{\Omega}_j^{(l)})$ using volume cost function
   d. If $J\left(\Omega_j^{(l)}\right) > J(\Omega_j^*)$, then $\Omega_j^* \leftarrow \Omega_j^{(l)}$
6. Result BSS is Ω_j^* with maximal volume

This algorithm loops over the bands to find the position in the selected band subset which has the maximal cost value. This algorithm can also be developed using BTs in step 3. and 5b from (2.2) or (2.3).

The successive CS-BSS procedure is in algorithm 4 and is developed with BTs. This algorithm takes advantage of (3.31) of $J\left(\widetilde{\Omega}_{\mathbf{\Phi}}\right) \rightarrow J\left(\widetilde{\Omega}\right)$ preserves the cost function. The algorithm begins with UBS and the search strategy takes each band in the selected subset and swaps it out for a possibly better band by trying all other bands. Each position in the

subset is optimized separately. The result is the band subset which has maximal volume. The SC search strategy can also be developed using BVs.

Algorithm 4. Band Subset Selection using Subspace Volume and Successive Subset Search via Compressively Sensed Band Tensors

Inputs: Number of bands to select found using HFC = $p = n_{VD}$

Compressive sensing sampling rates m_{n_1} and m_{n_2}

1. Let $\widetilde{\Omega}^* = \widetilde{\Omega}^{(0)} = \left\{\mathbf{B}_1^{(0)}, \mathbf{B}_2^{(0)}, \ldots, \mathbf{B}_p^{(0)}\right\}$ be the initial set of band selected from the band set Ω using uniform band subset selection
2. Simulate the compressive sensing using (2.2) or (2.3) to transform all bands to CSSD so that $\Omega = \left\{\mathbf{\Phi}_{k_{n_1}} \mathbf{B}_l \mathbf{\Phi}_{k_{n_1}}^T\right\}_{l=1}^{L}$.
3. Outer Loop: Let $j \leftarrow 1$ until j equals $p \leftarrow p + 1$
   a. Inner Loop: Let $l \leftarrow 1$ until l equals L, $l \leftarrow l + 1$
   b. Construct the following BSS
      $\Omega_l^{(j)} = \left\{\mathbf{\Phi}_{k_{n_1}} \mathbf{B}_1^{(l)} \mathbf{\Phi}_{k_{n_1}}^T, \ldots, \mathbf{\Phi}_{k_{n_1}} \mathbf{B}_{j-1}^{(l)} \mathbf{\Phi}_{k_{n_1}}^T, \mathbf{\Phi}_{k_{n_1}} \mathbf{B}_l, \mathbf{\Phi}_{k_{n_1}} \mathbf{B}_{j+1}^{(l)} \mathbf{\Phi}_{k_{n_1}}^T, \ldots, \mathbf{\Phi}_{k_{n_1}} \mathbf{B}_p^{(l)} \mathbf{\Phi}_{k_{n_1}}^T\right\}$
      by replacing the j^{th} band with all the selected band in sequence
   c. Evaluate $J(\widetilde{\Omega}_l^{(j)})$ using volume cost function
   d. If $J\left(\Omega_l^{(j)}\right) > J(\Omega_l^*)$, then $\Omega_l^* \leftarrow \Omega_l^{(j)}$
4. Output is the BSS is Ω_l^* with maximal volumne

The SQ and SC based algorithms are related but traverse the search space in different ways. The SQ algorithm traverses the full bands space and then the selected band subset. In contrast the SC algorithm traverses the selected band subset followed by the full band space. The computational complexity of each is $O(pL)$ for L bands and p selected bands.

Band Selection Evaluation Metric

To evaluate the subsets of selected bands an unnormalized version of the Tanimoto index called the coincident band rate (CBR) (6.3) is used to compare the BS results from

ODS and CSSD [100]. This measures the cardinality of the intersection SBS using the full band data and the band subset selected from the compressed band data.

$$n_{CBR} = \left| \{\tilde{\mathbf{b}}_j\}_{j=1}^{p} \cap \{\mathbf{\Phi}_{k_N} \tilde{\mathbf{b}}_j\}_{j=1}^{p} \right| \tag{6.3}$$

This is computed for a range of $1 \leq m_N \leq N$ to validate the BV version of algorithm 1 and 3 and over $1 \leq m_{n_1} \leq n_1$ and $1 \leq m_{n_2} \leq n_2$ for the BT-based algorithms of 2 and 4. Due to the high correlation between HSI bands and the error inherent in the probabilistic approximation of CS, a BS algorithm may sometimes converge to a point that does not exactly correspond to the full SBS but approximately close. CBR measures the exact match of the selected band subsets.

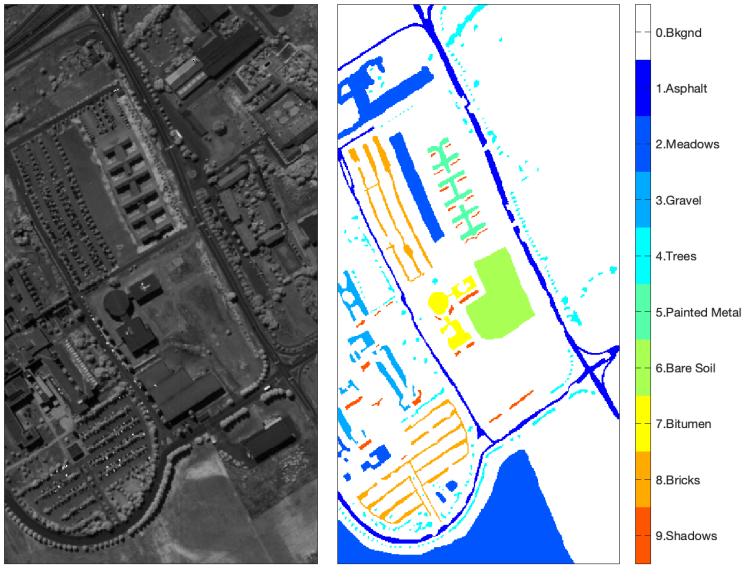
Real Hyperspectral Test Images

The algorithms developed in this chapter are verified through experiment with real HSI data sets. The HYDICE image from Fig. 12, Pavia University and Salina Valley dataset are used to verify the accuracy of the algorithms and the demonstrate the algorithm runtime savings provided by CS. These datasets are used because they exhibit high inter-band correlation which is an assumption made by the all BS algorithms. Sequential CS using recursive Kurtosis (SCS-RK) from Fig. 8 and Fig. 9 is used to estimate CSSR input the BS algorithms.

Pavia University Data Set

The Pavia Centre University scene was collected over Northern Italy using the ROSIS-03 sensor [9]. It has a spectral range of 0.43 – 0.86 μm and contains 610 x 340 pixels at 1.3 m GSD and 203 bands. The 12 bands with the largest noise were removed before processing. This scene is used for classification and has 9 classes and is a university

which is largely urban depicted in Fig. 55. The inter-band correlation is plotted in Fig. 56 showing a very high rate of data redundancy which is a result of the narrow spectral range of the sensor. The kurtosis is plotted across CSSR ranges for each mode of PV, BT and BV in Fig. 57 and Fig. 58. These plots are used to estimate CSSR required to construct the sampling matrices.



(a) (b)

Fig. 55. Pavia University hyperspectral test image and classification ground truth

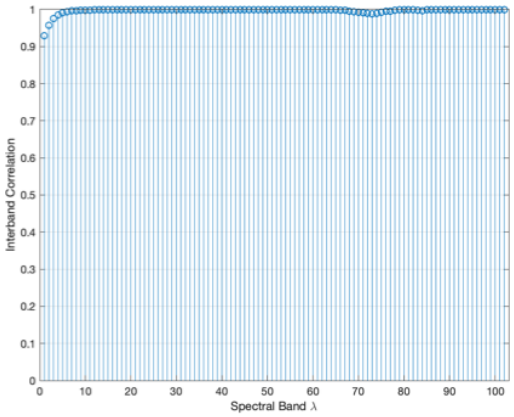


Fig. 56. Inter-band correlation of the Pavia University Scene

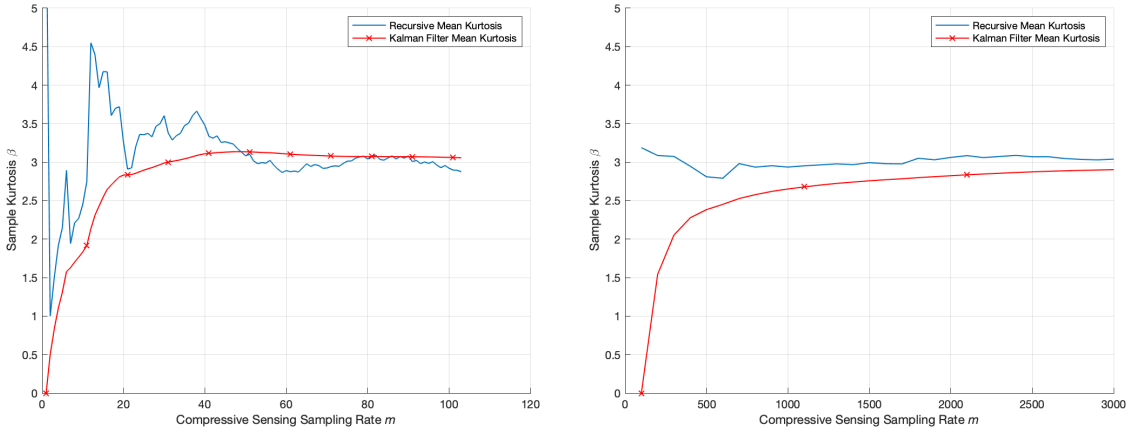


Fig. 57. Pavia University Scene CSSR convergence (a) PV plot (b) BV plot

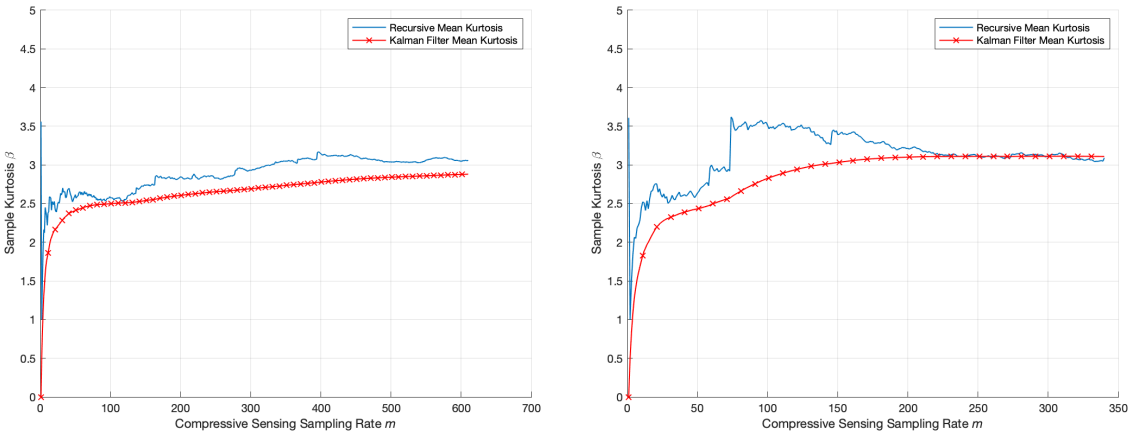
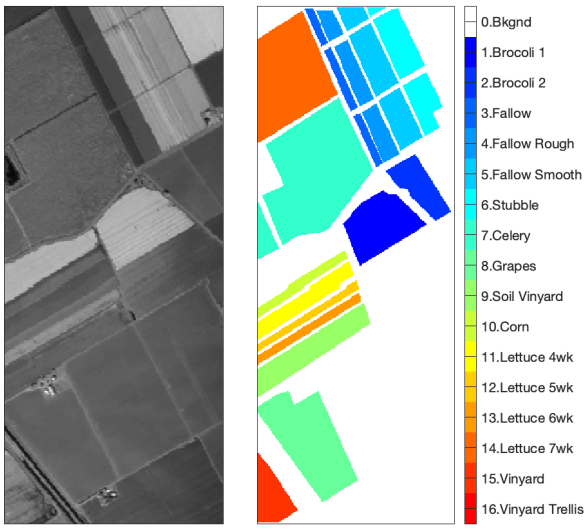


Fig. 58. Pavia University Scene CSSR convergence of BT (a) columns (b) rows

Salinas Data Set

The Salinas image was acquired using the AVIRIS sensor over Salinas Valley, California on October 9, 1998. The sensor has spectral range of 0.4 – 2.5 μm and consists of 512 x 217 pixels at 3.7m GSD and 224 bands. As with the Indian Pines data, the 20 water absorption bands were discarded 108-112, 154-167 and 224 leaving 104 spectral bands in the image. The scene is used to validate classification algorithms and has 16 classes and is largely agricultural depicted in Fig. 59. The inter-band correlation of the Salinas scene is plotted in Fig. 60 and shows considerable data redundancy among the spectral bands. Again, the kurtosis is plotted in Fig. 61 and Fig. 62 for the PV, BV and BT.



(a) (b)

Fig. 59. Salinas hyperspectral test image and classification ground truth

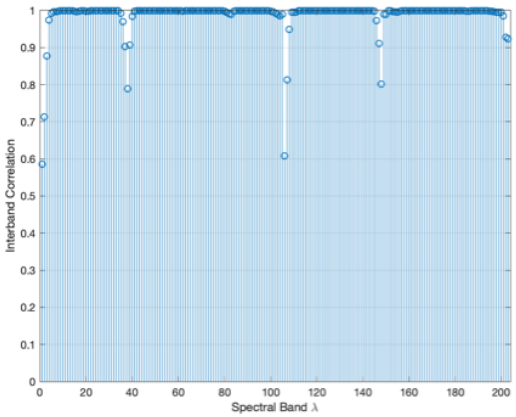
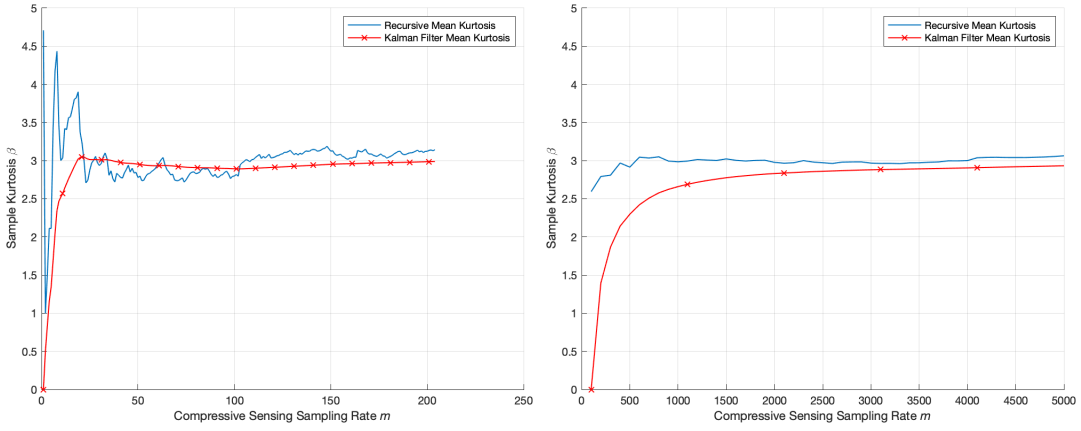
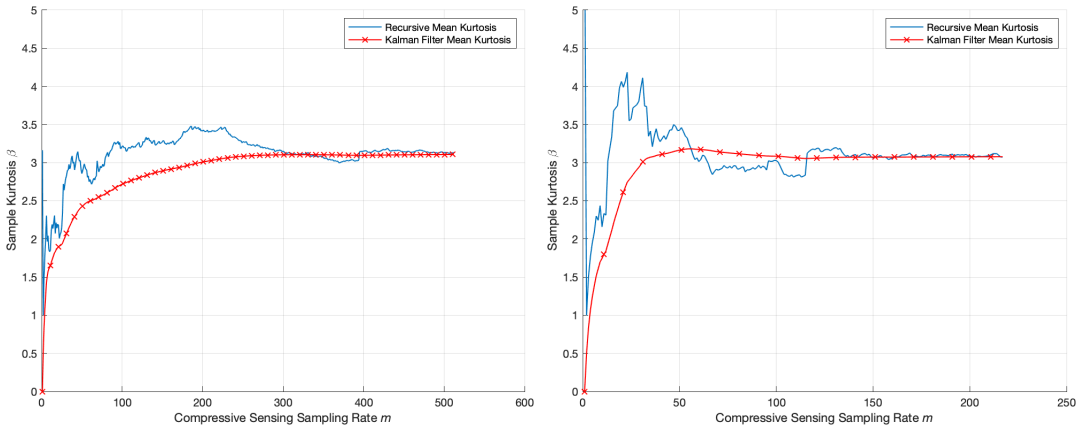


Fig. 60. Inter-band correlation of the Salinas Scene



(a) (b)

Fig. 61. Salinas Scene CSSR convergence (a) PV plot and (b) BV plot



(a) (b)

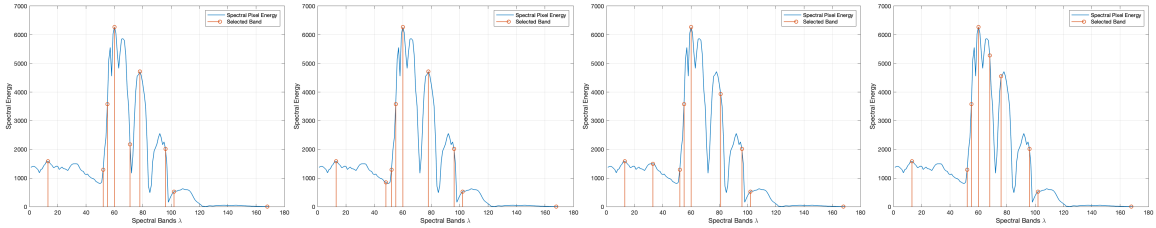
Fig. 62. Salinas Scene CSSR convergence of BT (a) columns and (b) rows

Band Selection Empirical Results

Algorithm 1 and 2 were verified through experiment using the HYDICE dataset depicted in Fig. 12. CSSR for the BV CS model in (2.4) was estimated from the convergence of the SCS-RK plot in Fig. 14(b) to be approximately $m_N = 1350$ which is a CSSR of $\frac{1350}{4096} = 0.3296$. CSSR for BT-CS model in (2.2) or (2.3) was estimated from the convergence of SCS-RK plot in Fig. 15 to be $m_{n_1} = m_{n_2} = 37$ for a total of 1369 CS observations which is a CSSR of $\frac{1369}{4096} = 0.3342$. In addition, the HFC estimate of $p = n_{VD} = 9$ bands as the SBS size. The SBS was computed using algorithms 1 and 2 is in Table 6. The SBS results from Table 6 show the coincidence of the algorithm 1 and 2 output when using the data in ODS and CSSD. CBR shows the closeness of the result with variations due to the limited size of the random sampling matrix which

Table 6. HYDICE CS Band Selection Results

Algorithm	CSSR	CBR	Selected Bands (in order of selection)
Full Band	No CS	N/A	168 60 52 13 102 55 78 96 71
CSBS-BV	1350 / 4096	8 / 9	168 60 52 13 102 55 78 48 96
CSBS-T	1369 / 4096	7 / 9	168 60 52 13 102 55 81 96 33
CSBS-TSVD	1369 / 4096	7 / 9	168 60 52 13 102 55 76 68 96

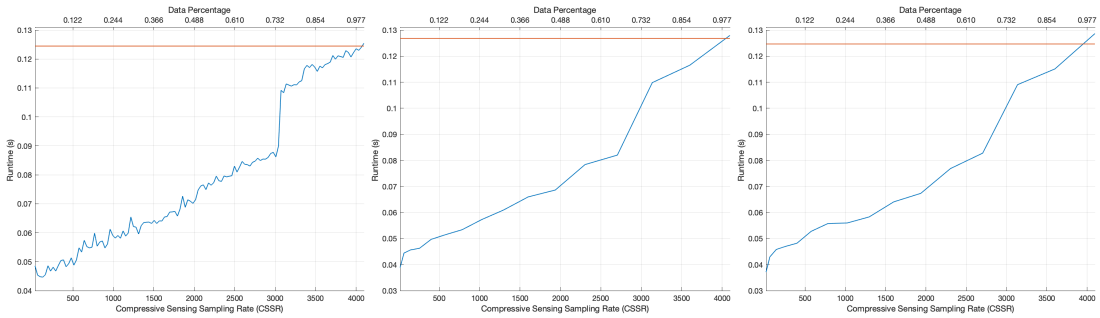


(a) (b) (c) (d)

Fig. 63. HYDICE selection result (a) ODS, (b) BV, (c) BT, (d) BT-SVD

converges to identity when CSSR goes to infinity. Note that the bands which are differing are among the bands chosen last. The bands from Table 6 are plotted in Fig. 63 which shows the quality of the band selection result and the coincidence between the results in ODS to CSSD. Finally, the execution time was recorded as the CSSR was increased in Fig. 64 from $m_N = [1,4096]$ and $m_{n_1} = m_{n_2} = [1,64]$. The linear trend below the runtime of ODS is evident which underpins the advantages of performing the BS algorithm 1 in CSSD.

Algorithms 1 and 2 were also verified through experiment using the Pavia University dataset depicted in Fig. 55. CSSR for the BV CS model in (2.4) was estimated from the convergence of SCS-RK plot in Fig. 57(b) to be approximately $m_N = 30{,}000$ which is a CSSR of $\frac{30{,}000}{207{,}400} = 0.1446$. CSSR for the BT-CS model in (2.2) or (2.3) was estimated from the convergence of the SCS-RK plot in Fig. 58 to be $m_{n_1} = 350, m_{n_2} = 200$ for a total of 66,980 CS observations which is a CSSR of



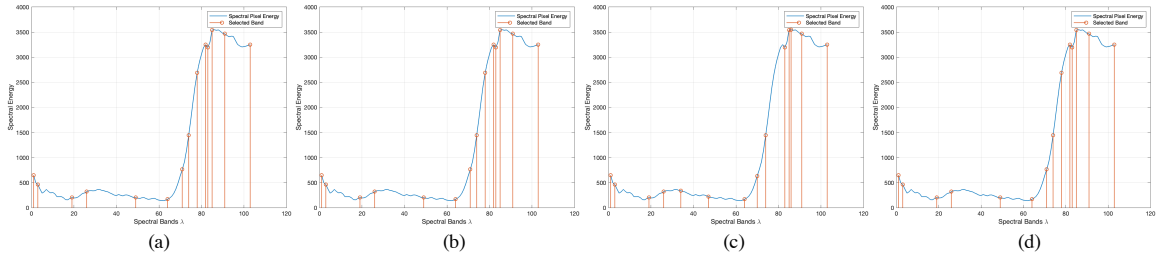
(a) (b) (c)

Fig. 64. HYDICE selection timing (a) BV, (b) BT and (c) BT-SVD

Table 7. Pavia University CS Band Selection Results

Algorithm	CSSR	CBR	Selected Bands (in order of selection)													
Full Band	No CS	NA	3	91	19	26	49	64	71	74	103	82	85	83	1	78
CSBS-BV	30000 / 207400	14 / 14	3	91	19	26	49	64	71	74	103	82	85	83	1	78
CSBS-T	66980 / 207400	10 / 14	3	91	19	26	70	64	34	74	103	83	86	47	85	1
CSBS-TSVD	66980 / 207400	14 / 14	3	91	19	26	49	64	71	74	103	82	85	83	1	78

$\frac{66,980}{207,400} = 0.3230$. In addition, the HFC estimated the $p = n_{VD} = 14$ bands is the estimated SBS size. The SBS was computed using Algorithm 1 and 2 is in Table 7. The SBS results from Table 7 show the coincidence of the algorithm 1and 2 output when using the data in ODS and CSSD. CBR shows the closeness of the result with variations due to the limited size of the random sampling matrix which converges to identity as CSSR goes to infinity. Note that the bands which are differing are among the bands which are chosen last. The bands from Table 7 are plotted in Fig. 65 which shows the quality of the band selection result and the coincidence between the results obtained in ODS to CSSD. Finally, the execution time was recorded as CSSR was increased in Fig. 66 from $m_N = [1,30000]$ and $m_{n_1} = [1,610], m_{n_2} = [1,340]$. The linear trend below the runtime of using data in ODS is evident which underpins the advantages of performing the BS algorithms 1 and 2 in CSSD. The runtime graph in Fig. 66(a) using the BV model was computed for $m_N = [1,30000]$ due to computational limitations.



(a) (b) (c) (d)

Fig. 65. Pavia University selection result (a) ODS, (b) BV, (c) BT, (d) BT-SVD

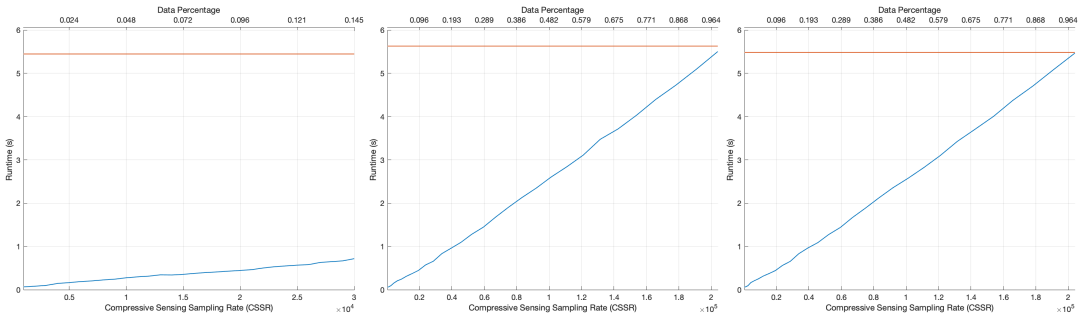


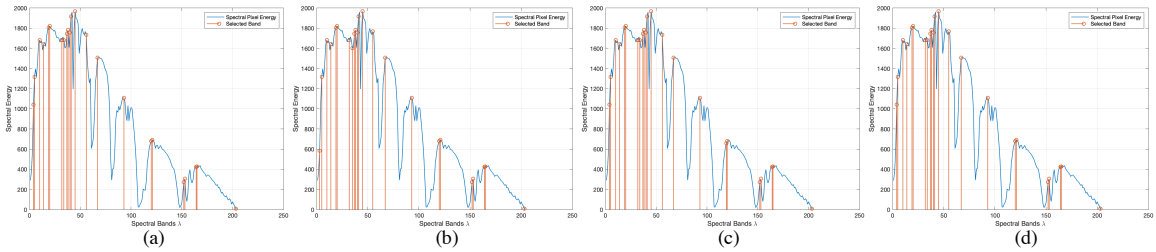
Fig. 66. Pavia University selection timing (a) BV, (b) BT and (c) BT-SVD

Generating a full sampling matrix for the BV model would require a matrix 207,400 x 207,400 which when realized is over 1.6 GB.

The final dataset used to validate Algorithms 1 and 2 through experiment is the Salinas Valley scene in Fig. 59. CSSR for the BV-CS model in (2.4) was estimated from the convergence of SCS-RK plots in Fig. 61 to be approximately $m_N = 5000$ which is a CSSR of $\frac{5{,}000}{111{,}104} = 0.0450$. CSSR for the BT-CS model in (2.2) or (2.3) was estimated from the convergence of SCS-RK plots in Fig. 62 to be $m_{n_1} = 300, m_{n_2} = 100$ for a total of 30,000 CS observations which is a CSSR of $\frac{30{,}000}{111{,}104} = 0.2700$. In addition, the HFC estimated the $p = n_{VD} = 23$ bands is the estimated SBS size. SBS was computed using Algorithm 1 and 2 is in Table 8. The SBS results from Table 8 show the coincidence of the algorithm 1and 2 output when using the data in ODS and

Table 8. Salinas Valley CS Band Selection Results

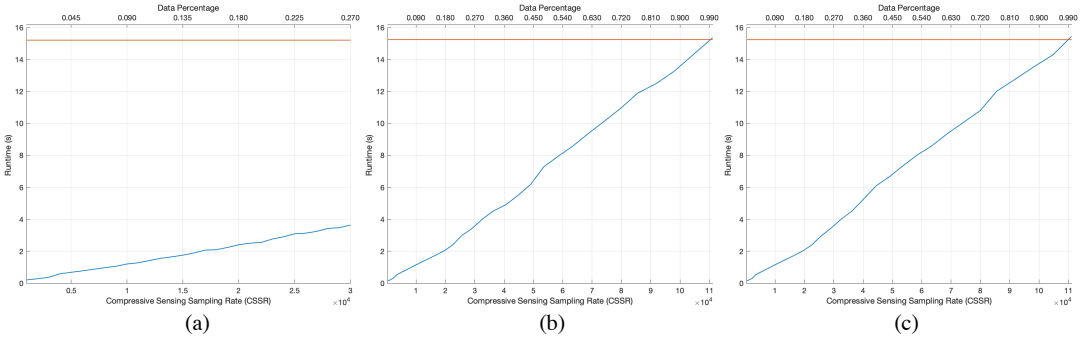
Algorithm	CSSR	CBR	Selected Bands (in order of selection)
Full Band	No CS	NA	203 45 32 38 165 5 93 67 152 20 121 55 41 10 34 14 4 40 37 120 153 19 164
CSBS-BV	5000 / 111104	21 / 23	203 45 32 38 165 5 93 67 152 20 120 56 41 10 14 37 40 4 34 153 19 119 164
CSBS-T	30000 / 111104	22 / 23	203 45 32 38 165 5 93 67 152 20 14 121 56 41 10 4 40 37 34 120 19 153 164
CSBS-TSVD	30000 / 111104	21 / 23	203 45 32 38 165 5 93 67 152 20 121 55 41 10 35 3 14 40 37 120 19 153 164



(a) (b) (c) (d)

Fig. 67. Salinas Valley selection results (a) ODS, (b) BV, (c) BT, (d) BT-SVD

CSSD. CBR shows the closeness of the result with variations due to the limited size of the random sampling matrix which converges to identity as CSSR goes to infinity. Note that the bands which are differing are among the bands which are chosen last. The bands from Table 8 are plotted in Fig. 67 which shows the quality of the band selection result and the coincidence between the results using data in ODS to and data in CSSD. Finally, the execution time was recorded as CSSR was increased in Fig. 68 from $m_N = [1{,}3000]$ and $m_{n_1} = [1{,}300], m_{n_2} = [1{,}100]$. The linear trend below the runtime of using data in ODS is evident which underpins the advantages of performing the BS algorithms 1 and 2 in CSSD. The runtime graph in Fig. 68(a) using the BV model was computed for $m_N = [1{,}3000]$ due to computational limitations. The full matrix would be 111,104 x 111,104 Gaussian realizations which would require a matrix over 800MB in size.



(a) (b) (c)

Fig. 68. Salinas Valley selection timing (a) BV, (b) BT and (c) BT-SVD

Band Subset Selection Empirical Results

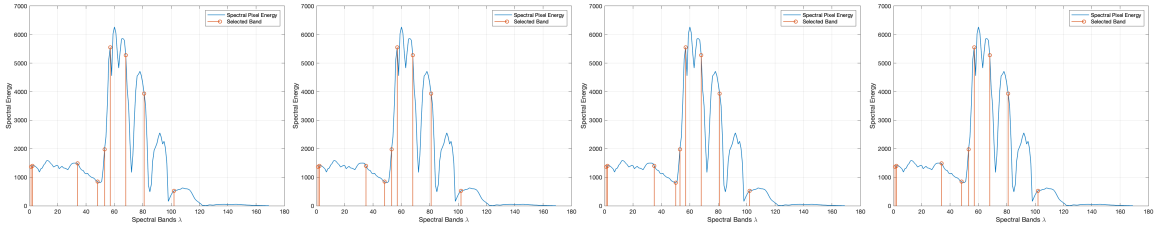
The sequential and successive band subset of algorithms 3 and 4 were validated using the HYDICE image from Fig. 12. The BV, BT and BT-SVD models in (2.2), (2.3) and (2.4) where used to compressively sense the data. Using SCS-RK in Fig. 9 the BV-CSSR was estimated using Fig. 14(b) to be $m_N = 1350$ giving a CSSR of $\frac{1350}{4096} = 0.3296$. In addition, CSSR of the BT models was estimated using Fig. 15 to be $m_{n_1} = 37$ and $m_{n_2} = 37$ for a total of 1369 compressive samples giving a CSSR of $\frac{1369}{4096} = 0.3342$. The results of algorithms 3 and 4 are in Table 9 and Table 10. The SBS results from Table 9 and Table 10 exhibit the concurrence of the algorithms 3 and 4 when using data in ODS and CSSD. CBR shows the closeness of the result with variations due to the limited size of the random sampling matrix which converges to identity as CSSR goes to infinity. The bands from in Table 9 and Table 10 are plotted in Fig. 69 and Fig. 70 showing the quality of the band selection result and the similarity between the results in ODS and in CSSD. Finally, the execution time was recorded as

Table 9. HYDICE CS Sequential Band Subset Selection Results

Algorithm	CSSR	CBR	Selected Bands (in order of selection)
Full Band	No CS	NA	1 57 53 81 34 102 48 68 2
CSBSS-BV	1350 / 4096	8 / 9	1 57 48 81 68 102 35 53 2
CSBSS-T	1369 / 4096	7 / 9	1 68 50 81 57 102 35 53 2
CSBSS-TSVD	1369 / 4096	9 / 9	1 57 48 81 34 102 53 68 2

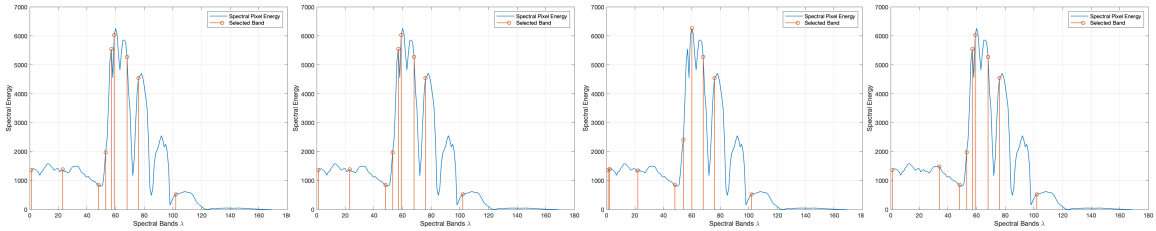
Table 10. HYDICE CS Successive Band Subset Selection Results

Algorithm	CSSR	CBR	Selected Bands (in order of selection)
Full Band	No CS	NA	53 1 48 57 68 76 23 59 102
CSBSS-BV	1350 / 4096	9 / 9	53 1 48 57 68 76 23 59 102
CSBSS-T	1369 / 4096	5 / 9	54 1 48 60 68 76 22 2 102
CSBSS-TSVD	1369 / 4096	8 / 9	53 1 48 57 68 76 59 34 102



(a) (b) (c) (d)

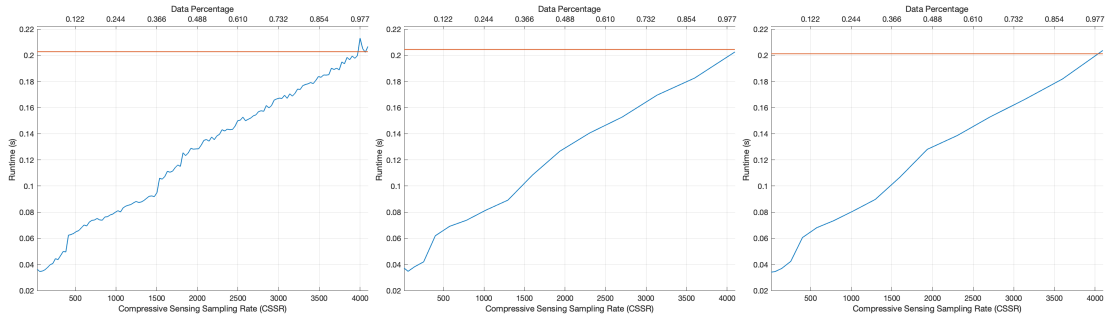
Fig. 69. HYDICE sequential subset selection (a) ODS, (b) BV, (c) BT, (d) BT-SVD



(a) (b) (c) (d)

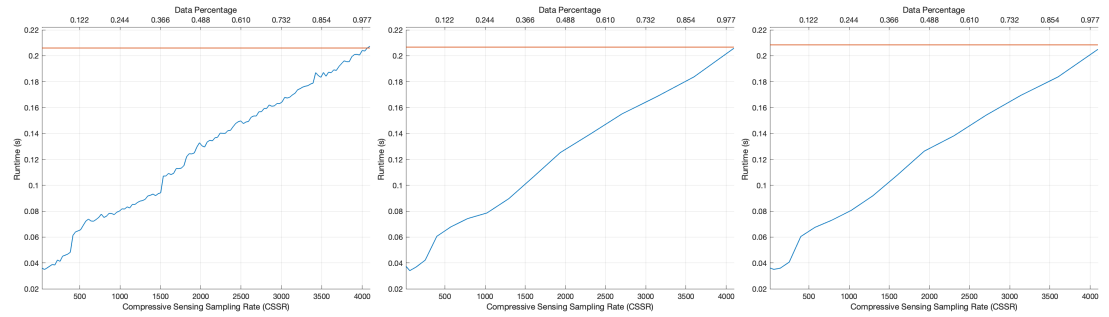
Fig. 70. HYDICE successive subset selection (a) ODS, (b) BV, (c) BT, (d) BT-SVD

CSSR was increased in Fig. 71 and Fig. 72 from $m_N = [1,4096]$ and $m_{n_1} = [1,64], m_{n_2} = [1,64]$. The linear trend below the runtime of using data in ODS is evident which underpins the advantages of performing the BS algorithms 3 and 4 in CSSD.



(a) (b) (c)

Fig. 71. HYDICE sequential subset selection timing (a) BV, (b) BT and (c) BT-SVD



(a) (b) (c)

Fig. 72. HYDICE successive subset selection timing (a) BV, (b) BT and (c) BT-SVD

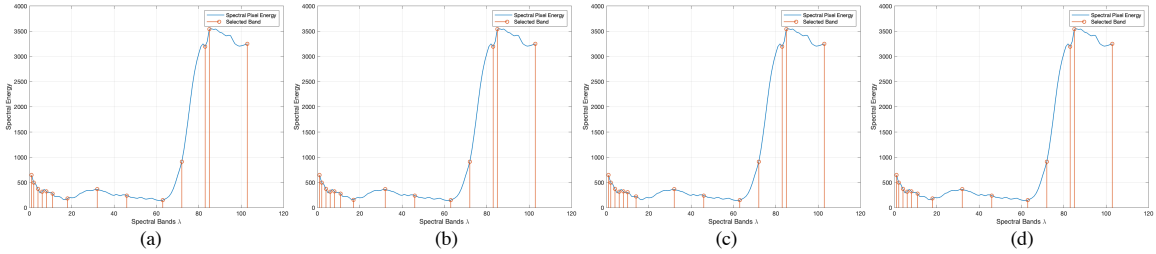
Another dataset of Pavia University from Fig. 55 is used to validate the sequential and successive band subset of algorithms 3 and 4. Again, the BV, BT and BT-SVD models in (2.2), (2.3) and (2.4) where used to compressively sense the data. Using SCS-RK in Fig. 9 the BV-CSSR was estimated using Fig. 57(b) to be $m_N = 30{,}000$ giving a CSSR of $\frac{30{,}000}{207{,}400}$. In addition, CSSR of the BT models was estimated using Fig. 58 to be $m_{n_1} = 350$ and $m_{n_2} = 200$ for a total of 70,000 compressive samples giving a CSSR of $\frac{70{,}000}{207{,}400} = 0.3375$. The results of algorithms 3 and 4 are in Table 11 and Table 12. The SBS results from in Table 11 and Table 12 display the correspondence of the algorithms 3 and 4 when using data from ODS and CSSD. CBR shows the closeness of the result with variations due to the limited size of the random sampling matrix which converges to identity as CSSR goes to infinity. The bands from in in Table 11 and Table 12 are plotted in Fig. 73 and Fig. 74 showing the quality of the band selection result and the similarity between the results obtained from using data in ODS and in CSSD. Finally, the execution time was

Table 11. Pavia University CS Sequential Band Subset Selection Results

Algorithm	CSSR	CBR	Selected Bands (in order of selection)													
Full Band	No CS	NA	1	8	18	4	32	2	46	6	63	72	83	85	11	103
CSBSS-BV	30000 / 207400	13 / 14	1	8	17	4	32	2	46	6	63	72	83	85	11	103
CSBSS-T	70000 / 207400	12 / 14	1	8	14	4	32	2	46	6	63	72	83	85	10	103
CSBSS-TSVD	70000 / 207400	14 / 14	1	8	18	4	32	2	46	6	63	72	83	85	11	103

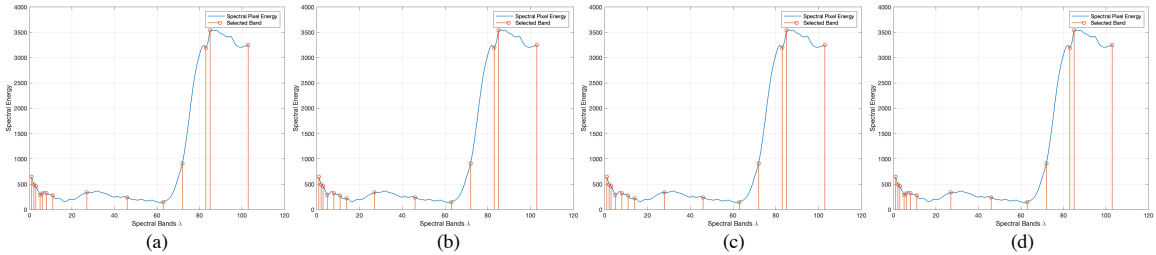
Table 12. Pavia University CS Successive Band Subset Selection Results

Algorithm	CSSR	CBR	Selected Bands (in order of selection)													
Full Band	No CS	NA	1	3	5	8	2	27	83	46	63	72	11	85	6	103
CSBSS-BV	30000 / 207400	13 / 14	1	3	5	8	2	27	83	46	63	72	11	85	14	103
CSBSS-T	70000 / 207400	12 / 14	1	3	5	8	2	28	83	46	63	72	11	85	14	103
CSBSS-TSVD	70000 / 207400	14 / 14	1	3	5	8	2	27	83	46	63	72	11	85	6	103



(a) (b) (c) (d)

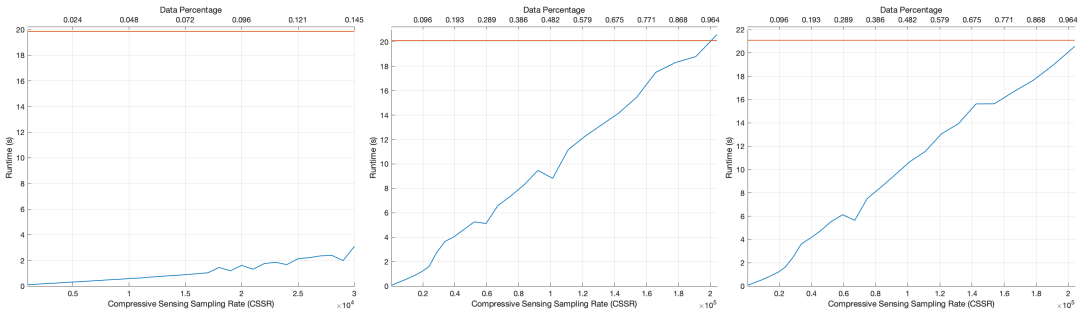
Fig. 73. Pavia U. sequential subset selection (a) ODS, (b) BV, (c) BT, (d) BT-SVD



(a) (b) (c) (d)

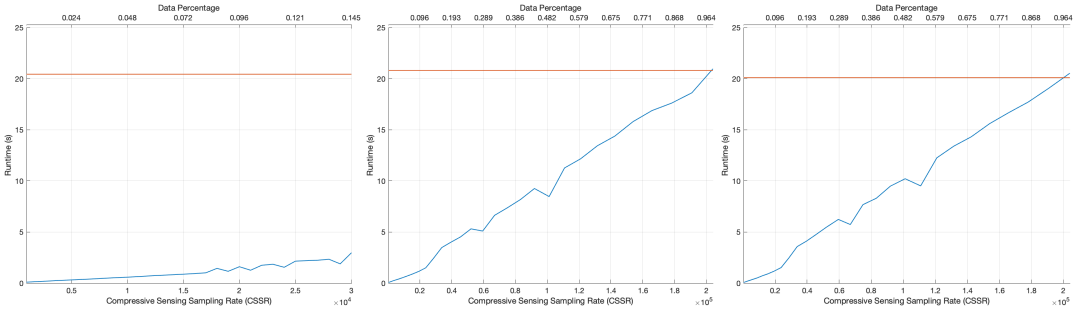
Fig. 74. Pavia U. successive subset selection (a) ODS, (b) BV, (c) BT, (d) BT-SVD

recorded for a varying CSSR of $m_N = [1{,}30000]$ and $m_{n_1} = [1{,}610]$ and $m_{n_2} = [1{,}340]$ with results plotted in Fig. 75 and Fig. 76. The plots display the linear trend evident in prior experiments. Also, note that in Fig. 75(a) and Fig. 76(a) are limited due to computational constraints mentioned prior.



(a) (b) (c)

Fig. 75. Pavia U. sequential subset selection timing (a) BV, (b) BT and (c) BT-SVD



(a) (b) (c)

Fig. 76. Pavia U. successive subset selection timing (a) BV, (b) BT and (c) BT-SVD

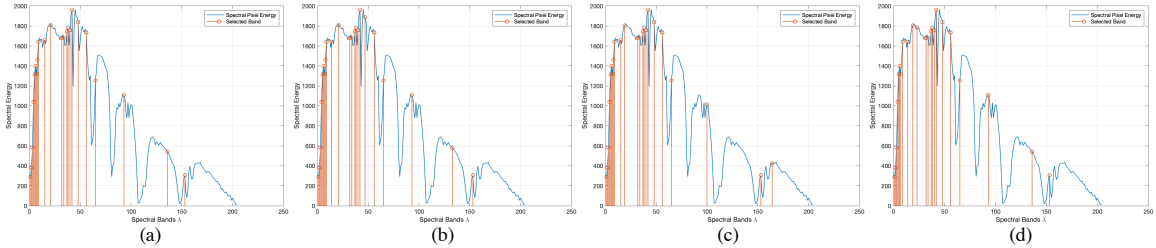
The final dataset of Salinas Valley from Fig. 59 is used to validate the sequential and successive band subset of algorithms 3 and 4. Again, the BV, BT and BT-SVD models in (2.2), (2.3) and (2.4) where used to compressively sense the data. Using SCS-RK in Fig. 9 the BV CSSR was estimated using Fig. 61 to be $m_N = 5{,}000$ giving a CSSR of $\frac{5{,}000}{111{,}104} = 0.0450$. In addition, CSSR of BT models was estimated using Fig. 62 to be $m_{n_1} = 300$ and $m_{n_2} = 100$ for a total of 30,000 compressive samples giving a CSSR of $\frac{30{,}000}{111{,}104} = 0.2700$. The results of algorithms 3 and 4 are in Table 13 and Table 14. The SBS results display the correspondence of the algorithms 3 and 4 when using data in ODS and CSSD. CBR shows the closeness of the result with variations due to the limited size of the random sampling matrix which converges to identity as CSSR goes to infinity. The selected bands are plotted in Fig. 77 and Fig. 78 showing the quality of the band selection result and the similarity between the results obtained using data in ODS and in CSSD.

Table 13. Salinas Valley CS Sequential Band Subset Selection Results

Algorithm	CSSR	CBR	Selected Bands (in order of selection)
Full Band	No CS	NA	1 38 21 32 37 48 56 65 34 8 93 42 9 15 4 136 2 153 40 6 7 3 5
CSBSS-BV	5000 / 111104	20 / 23	1 40 21 32 37 47 56 65 34 9 93 38 8 42 4 133 2 153 14 6 7 3 5
CSBSS-T	30000 / 111104	20 / 23	1 40 19 32 37 48 56 65 34 8 100 164 9 15 4 42 2 153 38 6 7 3 5
CSBSS-TSVD	30000 / 111104	20 / 23	1 34 19 32 37 48 56 65 38 9 93 40 42 23 4 136 2 153 14 6 7 3 5

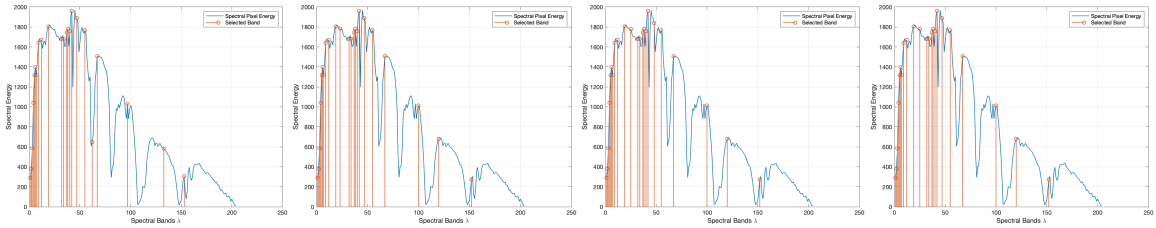
Table 14. Salinas Valley CS Successive Subset Selection Results

Algorithm	CSSR	CBR	Selected Bands (in order of selection)
Full Band	No CS	NA	2 5 19 34 38 42 3 1 55 4 12 67 32 40 6 97 47 37 7 133 9 153 62
CSBSS-BV	5000 / 111104	19 / 23	2 5 19 34 38 42 3 1 55 4 12 67 32 40 6 100 47 37 7 9 120 152 23
CSBSS-T	30000 / 111104	18 / 23	2 5 19 34 38 42 3 1 55 4 12 67 32 40 6 100 48 37 7 9 120 152 25
CSBSS-TSVD	30000 / 111104	19 / 23	2 5 19 34 38 42 3 1 55 4 12 67 32 40 6 100 47 37 7 9 120 152 25



(a) (b) (c) (d)

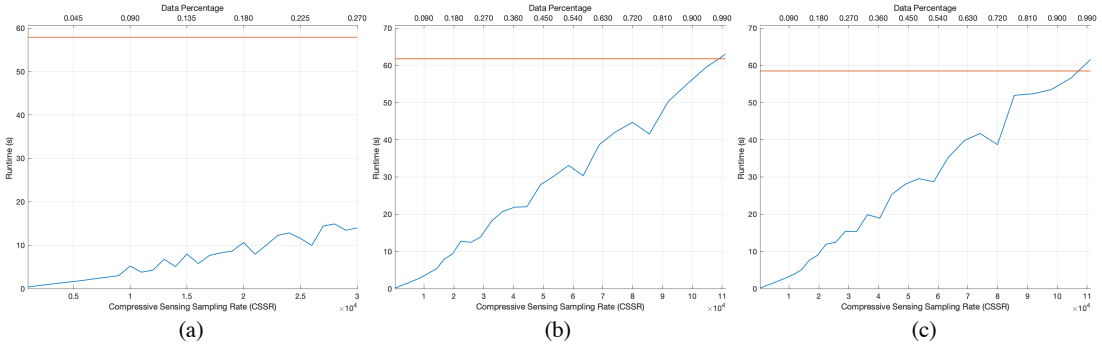
Fig. 77. Salinas sequential subset selection (a) ODS, (b) BV, (c) BT, (d) BT-SVD



(a) (b) (c) (d)

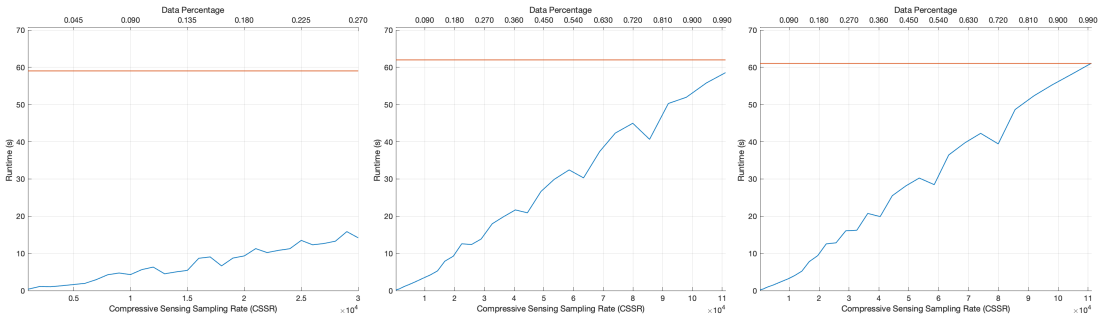
Fig. 78. Salinas successive subset selection (a) ODS, (b) BV, (c) BT, (d) BT-SVD

Finally, the execution time was recorded for a varying CSSR of $m_N = [1{,}5000]$ and $m_{n_1} = [1{,}512]$ and $m_{n_2} = [1{,}217]$ with results plotted in Fig. 79 and Fig. 80. The plots display the linear trend underpinning the time savings of using CS and BS.



(a) (b) (c)

Fig. 79. Salinas sequential subset selection timing (a) BV, (b) BT and (c) BT-SVD



(a) (b) (c)

Fig. 80. Salinas successive subset selection (a) BV, (b) BT and (c) BT-SVD

Conclusion

This chapter presented the notion of using CS to reduce the size of the bands in hyperspectral images using a band vector, band tensor or band tensor SVD-CS models minimally affects the orthogonal subspace projection lengths as well as the subspace volumes. The experiments conducted showed the convergence of the band selection outputs for each data set at an appropriate CSSR. The errors in the band selection output converge to the truth exponentially while the runtime increases linearly. This shows that a considerably smaller compressively sampled set of data can be used to perform band selection using orthogonal subspace projection or subspace volume.

Chapter 7: Conclusions and Future Work

This dissertation investigates various approaches to hyperspectral image analysis via compressive sensing. By exploiting the length and angle preserving principles of the related restricted isometric property and restricted conformal property in CS, this dissertation demonstrates that the applications of target detection and band selection can be derived in the compressively sensed sample domain. The principal advantage of formulating these applications in compressively sensed sample or band domain is to reproduce nearly the same results using compressively sensed data compared to using the data in original data space so that significantly reduced computational complexity and tremendous memory/storage savings can be achieved. Ideally, a compressive sensor would acquire these samples directly and collection systems could immediately perform data exploitation without the need for reconstruction.

The restricted isometric property is also shown to be extensible to include the probabilistic measure of information entropy. As a result, not only can length and angle be preserved but also the Kullback Leibler distance is preserved as well. This has implications for further hyperspectral processing algorithms based on the spectral information divergence. Future developments are the formulating of data processing algorithms in terms of the information theoretic restricted entropy and restricted spectrum properties proposed in this dissertation.

As a final contribution, this dissertation develops a practical study and presents a novel estimation method to find the compressive sensing sampling rates. The proposed method is recursive, computationally light-weight and can be built in-line with data

collection. Future work includes the refinement of this method casting it as a detection problem so that the rates can be ascertained automatically.

References

[1] E. J. Candes and M. B. Wakin, "An Introduction To Compressive Sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, 2008.

[2] C.-I Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification*. Springer US, 2003.

[3] C.-I Chang, *Hyperspectral Data Processing: Algorithm Design and Analysis*. Hoboken, NJ: Wiley-Interscience, 2013.

[4] Chein-I Chang and Shao-Shan Chiang, "Anomaly detection and classification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 6, pp. 1314–1325, 2002.

[5] C.-I Chang, *Real-Time Progressive Hyperspectral Image Processing*. New York, NY: Springer Berlin Heidelberg, 2016.

[6] C.-I Chang, *Real-Time Recursive Hyperspectral Sample and Band Processing*. New York, NY: Springer Science+Business Media, 2017.

[7] P. A. Mitchell, "Hyperspectral Digital Imagery Collection Experiment (HYDICE)," in *Geographic Information Systems, Photogrammetry, and Geological/Geophysical Remote Sensing*, 2004, vol. 2587, pp. 70–95.

[8] D. Snyder, J. Kerekes, I. Fairweather, R. Crabtree, J. Shive, and S. Hager, "Development of a web-based application to evaluate target finding algorithms," *Int. Geosci. Remote Sens. Symp.*, vol. 2, no. 1, pp. 915–918, 2008.

[9] G. Romay, Manuel, "Grupo De Intelligencia Computacional (GIC)," 2019. [Online]. Available: http://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.

[10] D. A. Kusinsky and M. E. Leeser, "FPGA-based hyperspectral covariance coprocessor for size, weight, and power constrained platforms," *2013 IEEE High Perform. Extrem. Comput. Conf. HPEC 2013*, pp. 1–6, 2013.

[11] E. Puckrin *et al.*, "Airborne infrared hyperspectral imager for intelligence, surveillance and reconnaissance applications," *Electro-Optical Remote Sensing, Photonic Technol. Appl. VI*, vol. 8542, no. November 2012, p. 854226, 2012.

[12] R. M. Willett, M. F. Duarte, M. A. Davenport, and R. G. Baraniuk, "Sparsity and Structure in Hyperspectral Imaging," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 116–126, 2014.

[13] E. Candes, "The Restricted Isometry Property and Its Implications for Compressed Sensing," *Comptes Rendus Math.*, vol. 346, no. 9–10, pp. 589–592, 2008.

[14] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral Image Processing for Automatic Target Detection Applications," *Lincoln Lab. J.*, vol. 14, no. 1, pp. 79–116, 2003.

[15] I. S. Read and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust.*, vol. 38, no. 19, pp. 1760–1770, 1990.

[16] C.-I Chang and Hsuan Ren, "Linearly constrained minimum variance beamforming approach to target detection and classification for hyperspectral imagery," pp. 1241–1243, 2003.

[17] C.-I Chang, "Orthogonal subspace projection (OSP) revisited: a comprehensive study and analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 502–518, 2005.

[18] W. Zhang, X. Li, Y. Dou, and L. Zhao, "A geometry-based band selection approach for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4318–4333, 2018.

[19] Liguo Wang, Ye Zhang, and Yanfeng Gu, "Unsupervised Band Selection Method Based on Improved N-FINDR Algorithm for Spectral Unmixing," *Int. Symp. Syst. Control Aerosp. Astronaut.*, no. 1, pp. 1018–1021, 2006.

[20] C. E. Shannon, "Communication in the Presence of Noise," *Proc. IRE*, vol. 37, no. 1, pp. 10–21, 1949.

[21] C. E. Shannon, "A Mathematical Theory of Communication*," *Mob. Comput. Commun. Rev.*, vol. 5, no. 1, pp. 3–55, 1948.

[22] E. Candes and J. Romberg, "Sparsity and Incoherence in Compressive Sampling," *Inverse Prob.*, vol. 23, no. 3, pp. 969–985, 2007.

[23] E.J. Candes and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[24] D. L. Donoho, “Compressed Sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
[25] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. 2013.
[26] M. Naureen, S. Lee, S. Qaisar, R. M. Bilal, and W. Iqbal, “Compressive sensing: From theory to applications, a survey,” *J. Commun. Networks*, vol. 15, no. 5, pp. 443–456, 2014.
[27] N. Hurley and S. Rickard, “Comparing measures of sparsity,” *IEEE Trans. Inf. Theory*, vol. 55, no. 10, pp. 4723–4741, 2009.
[28] W. Meiniel, Y. Le Montagner, E. Angelini, and J. C. Olivo-Marin, “Image denoising by multiple compressed sensing reconstructions,” *Proc. - Int. Symp. Biomed. Imaging*, vol. 2015–July, pp. 1232–1235, 2015.
[29] H. Abdi and L. J. Williams, “Principal Component Analysis,” *WIREs Comput. Stat.*, vol. 2, no. 4, pp. 433–459, Jul. 2010.
[30] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, “A Transformation for Ordering Multispectral Data in Terms of Image Quality with Implications for Noise Removal,” *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 65–74, 1988.
[31] A. Hyvärinen and E. Oja, “Independent Component Analysis: Algorithms and Applications,” *Neural Netw.*, vol. 13, no. 4–5, pp. 411–430, May 2000.
[32] P. O. Hoyer, “Non-negative Matrix Factorization with Sparseness Constraints,” *J. ofMachine Learn. Res.*, vol. 5, pp. 1457–1469, 2004.
[33] M. E. Lopes, “Estimating Unknown Sparsity in Compressed Sensing,” in *Estimating Unknown Sparsity in Compressed Sensing Miles*, 2013, vol. 28, no. 1.
[34] M. E. Lopes, “Unknown sparsity in compressed sensing: Denoising and inference,” *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 5145–5166, 2016.
[35] R. Baraniuk, Richard, Romberg, Justin, Nowak, “Compressive Sensing: A New Framework for Imaging.” 2006.
[36] S. Dasgupta and A. Gupta, “An Elementary Proof of a Theorem of Johnson and Lindenstrauss,” *Random Struct. Algorithms*, vol. 22, no. 1, pp. 60–65, 2003.
[37] D. Achlioptas, “Database-friendly random projections: Johnson-Lindenstrauss with binary coins,” *J. Comput. Syst. Sci.*, vol. 66, no. 4, pp. 671–687, 2003.
[38] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, “A simple proof of the restricted isometry property for random matrices,” *Constr. Approx.*, vol. 28, no. 3, pp. 253–263, 2008.
[39] T. Cheng, “Restricted conformal property of compressive sensing,” *2014 11th Int. Comput. Conf. Wavelet Act. Media Technol. Inf. Process. ICCWAMTIP 2014*, pp. 152–161, 2014.
[40] L. H. Chang and J. Y. Wu, “Achievable angles between two compressed sparse vectors under norm/distance constraints imposed by the restricted isometry property: A plane geometry approach,” *IEEE Trans. Inf. Theory*, vol. 59, no. 4, pp. 2059–2081, 2013.
[41] D. M. Malioutov, S. R. Sanghavi, and A. S. Willsky, “Sequential compressed sensing,” *IEEE J. Sel. Top. Signal Process.*, vol. 4, no. 2, pp. 435–444, 2010.
[42] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, “Uniform uncertainty principle for Bernoulli and subgaussian ensembles,” *Constr. Approx.*, vol. 28, no. 3, pp. 277–289, 2008.
[43] R. Ward, “Compressed sensing with cross validation,” *IEEE Trans. Inf. Theory*, vol. 55, no. 12, pp. 5773–5782, 2009.
[44] D. M. Malioutov, S. Sanghavi, and a S. Willsky, “Compressed Sensing with Sequential Observations,” *Bernoulli*, pp. 3357–3360, 2008.
[45] D. Bertsekas, *Convex optimization theory*, vol. 25, no. 3. Cambridge, UK ; New York: Cambridge University Press, 2009.
[46] S. S. Chen, D. L. Donoho, and M. A. Saunders, “Atomic Decomposition by Basis Pursuit,” *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 2003.
[47] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via OMP,” *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
[48] T. Blumensath and M. E. Davies, “Iterative hard thresholding for compressed sensing,” *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, 2009.
[49] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
[50] D. Sarvaiya and J. Amin, “Compression of Hyperspectral Image using Compressive Sensing,” *Int. J. Adv. Res. Innov. Ideas Educ.*, vol. 2, no. 3, pp. 1904–1911, 2016.

[51] Z. Wang, Y. Feng, and Y. Jia, "Spatial-spectral compressive sensing of hyperspectral image," *2013 IEEE 3rd Int. Conf. Inf. Sci. Technol. ICIST 2013*, pp. 1254–1259, 2013.
[52] P. Li, S. Yang, L. Jiao, L. Jin, M. Wang, and B. Wu, "Compressive Hyperspectral Imaging via Sparse Tensor and Nonlinear Compressed Sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 5943–5957, 2015.
[53] S. Friedland, Q. Li, and D. Schonfeld, "Compressive sensing of sparse tensors," *IEEE Trans. Image Process.*, vol. 23, no. 10, pp. 4438–4447, 2014.
[54] N. D. Sidiropoulos and A. Kyrillidis, "Multi-way compressed sensing for sparse low-rank tensors," *IEEE Signal Process. Lett.*, vol. 19, no. 11, pp. 757–760, 2012.
[55] Q. Du and H. Yang, "Similarity-Based Unsupervised Band Selection for Hyperspectral Image Analysis," *EEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 564–568, 2008.
[56] S. Sarvotham, D. Baron, and R. G. Baraniuk, "Measurements vs . Bits : Compressed Sensing meets Information Theory," pp. 1–5, 2006.
[57] M. A. A. Davenport, M. B. Wakin, and R. G. Baraniuk, "Detection and Estimation with Compressive Measurements," *Dept. ECE, Rice Univ. Tech. Rep*, pp. 1–16, 2006.
[58] Y. Chen, "Effects of linear projections on the performance of target detection and classification in hyperspectral imagery," *J. Appl. Remote Sens.*, vol. 5, no. 1, p. 053563, 2011.
[59] H. Cheng, "The Fundamentals of Compressed Sensing," in *Sparse Representation, Modeling and Learning in Visual Recognition*, London: Springer London, 2015, pp. 21–53.
[60] E. C. Caby, "An Introduction to Statistical Signal Processing," *Technometrics*, vol. 48, no. 4, pp. 572–573, 2009.
[61] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. New York, NY, USA: Wiley-Interscience, 2000.
[62] S. Report, "Formulas for Robust , One-Pass Parallel Computation of Covariances and Arbitrary-Order Statistical Moments," no. September. Sandia National Laboratories, 2008.
[63] H. Musoff and P. Zarchan, *Fundamentals of Kalman Filtering: A Practical Approach, Third Edition*, no. v. 190. Reston, Va: American Institute of Aeronautics and Astronautics, 2012.
[64] S. Liu, M. Yamada, N. Collier, and M. Sugiyama, "Change-point detection in time-series data by relative density-ratio estimation," *Neural Networks*, vol. 43, pp. 72–83, 2013.
[65] W. Xiong, C.-I Chang, C. C. Wu, K. Kalpakis, and H. M. Chen, "Fast Algorithms to Implement N-FINDR for Hyperspectral Endmember Extraction," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 4, no. 3, pp. 545–564, 2011.
[66] A. Martínez-Usó, F. Pla, J. M. Sotoca, and P. García-Sevilla, "Clustering-based multispectral band selection using mutual information," *Proc. - Int. Conf. Pattern Recognit.*, vol. 2, pp. 760–763, 2006.
[67] C.-I Chang, "Constrained subpixel target detection for remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1144–1159, 2000.
[68] D. Manolakis, C. Siracusa, and G. Shaw, "Hyperspectral subpixel target detection using the linear mixing model," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1392–1409, 2001.
[69] C. D. Meyer, Ed., *Matrix Analysis and Applied Linear Algebra*. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2000.
[70] J. Haupt and R. Nowak, "A generalized restricted isometry property," *Univ. Wisconsin-Madison, Tech. Rep*, pp. 1–16, 2007.
[71] L. Wang *et al.*, "Band Subset Selection for Anomaly Detection in Hyperspectral Imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 4887–4898, 2017.
[72] Q. Du and H. Yang, "Unsupervised band selection for hyperspectral image analysis," *Int. Geosci. Remote Sens. Symp.*, vol. 5, no. 4, pp. 282–285, 2007.
[73] L. Wang, D. Liu, and Q. Wang, "Geometric method of fully constrained least squares linear spectral mixture analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 6, pp. 3558–3566, 2013.
[74] X. Kang, S. Li, and J. A. Benediktsson, "Spectral – Spatial Hyperspectral Image Classification With Edge-Preserving Filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, 2014.
[75] D. C. Heinz and C.-I Chang, "Mixture Analysis Method for Material Quantification in Hyperspectral Imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 3, pp. 529–545, 2001.
[76] C.-I Chang, "Spectral information divergence for hyperspectral image analysis," no. 1, pp. 509–

511, 2003.
[77] S. Author, S. Kullback, and R. A. Leibler, “On Information and Sufficiency,” 1951.
[78] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. New York, NY, USA: Wiley-Interscience, 2006.
[79] D. G. Manolakis, G. A. Shaw, and N. Keshava, “Comparative Analysis of Hyperspectral Adaptive Matched Filter Detectors,” in *Proceedings of SPIE*, 2000, no. August 2000.
[80] H. V. Poor, *An Introduction to Signal Detection and Estimation (2Nd Ed.)*. Berlin, Heidelberg: Springer-Verlag, 1994.
[81] N. Keshava and P. W. Boettcher, “Relationships between physical phenomena, distance metrics, and best-bands selection in hyperspectral processing,” in *Algorithms for Multispectral, Hyperspectral, and Ultraspectral Imagery VII*, 2003, vol. 4381, no. August 2001, pp. 55–67.
[82] N. Keshava, “Best bands selection for detection in hyperspectral processing,” in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2001.
[83] C.-I Chang, W. Xiong, W. Liu, M. L. Chang, C. C. Wu, and C. C. C. Chen, “Linear spectral mixture analysis based approaches to estimation of virtual dimensionality in hyperspectral imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 3960–3979, 2010.
[84] C.-I Chang, Q. Du, T. Sun, M. L. G. Althouse, “A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification,” *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2631–2641, 1999.
[85] S. De Backer, P. Kempeneers, W. Debruyn, and P. Scheunders, “Band Selection for Hyperspectral Remote Sensing,” *Pattern Recognit.*, vol. 2, no. 3, pp. 319–323, 2005.
[86] C.-I Chang, “A Review of Virtual Dimensionality for Hyperspectral Imagery,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 11, no. 4, pp. 1285–1305, 2018.
[87] C.-I Chang, W. Xiong, and C. H. Wen, “A theory of high-order statistics-based virtual dimensionality for hyperspectral imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 188–208, 2014.
[88] C.-I Chang, L. C. Lee, B. Xue, M. Song, and J. Chen, “Channel Capacity Approach to Hyperspectral Band Subset Selection,” *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 10, no. 10, pp. 4630–4644, 2017.
[89] P. Pudil, J. Novovičová, and J. Kittler, “Floating search methods in feature selection,” *Pattern Recognit. Lett.*, vol. 15, no. 11, pp. 1119–1125, 1994.
[90] Y. Xu, Q. Du, and N. Younan, “Particle swarm optimization-based band selection for hyperspectral target detection,” *2016 IEEE Int. Geosci. Remote Sens. Symp.*, vol. 1, no. 4, pp. 5872–5875, 2016.
[91] H. Su, B. Yong, and Q. Du, “Hyperspectral band selection using improved firefly algorithm,” *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 1, pp. 68–72, 2016.
[92] A. Martínez-Usó, F. Pla, J. M. Sotoca, and P. García-Sevilla, “Clustering-based hyperspectral band selection using,” *IEEE Trans. Geosci. Remote Sens.*, vol. 12, pp. 4158–4171, 2007.
[93] Yuan Yuan, Guokang Zhu, and Qi Wang, “Hyperspectral Band Selection by Multitask Sparsity Pursuit,” *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 631–644, 2014.
[94] C.-I Chang, “Spectral Inter-Band Discrimination Capacity of Hyperspectral Imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1749–1766, 2018.
[95] H. Yang, Q. Du, H. Su, and Y. Sheng, “An efficient method for supervised hyperspectral band selection,” *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 138–142, 2011.
[96] C.-I Chang and Q. Du, “Estimation of number of spectrally distinct signal sources in hyperspectral imagery,” *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 608–619, 2004.
[97] C. Yu, M. Song, and C.-I Chang, “Band subset selection for hyperspectral image classification,” *Remote Sens.*, vol. 10, no. 1, pp. 1–25, 2018.
[98] L. C. Lee, Y. C. Ouyang, S. Y. Chen, and C.-I Chang, “An information theoretical approach to multiple-band selection for hyperspectral imagery,” in *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016, vol. 2016–Novem, no. 1, pp. 2773–2776.
[99] L. Wang and C.-I Chang, “Multiple band selection for anomaly detection in hyperspectral imagery,” in *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2016, no. 1, pp. 7022–7025.
[100] T. T. Tanimoto, *An elementary mathematical theory of classification and prediction by T.T. Tanimoto*. International Business Machines Corporation New York, 1958.