Anomaly Detection and Reconstruction From Random Projections

James E. Fowler, *Senior Member, IEEE*, and Qian Du, *Senior Member, IEEE*

Abstract—Compressed-sensing methodology typically employs random projections simultaneously with signal acquisition to accomplish dimensionality reduction within a sensor device. The effect of such random projections on the preservation of anomalous data is investigated. The popular RX anomaly detector is derived for the case in which global anomalies are to be identified directly in the random-projection domain, and it is determined via both random simulation, as well as empirical observation that strongly anomalous vectors are likely to be identifiable by the projection-domain RX detector even in low-dimensional projections. Finally, a reconstruction procedure for hyperspectral imagery is developed wherein projection-domain anomaly detection is employed to partition the data set, permitting anomaly and normal pixel classes to be separately reconstructed in order to improve the representation of the anomaly pixels.

Index Terms—Anomaly detection, compressed sensing (CS), hyperspectral data, principal component analysis (PCA).

I. INTRODUCTION

IN HYPERSPECTRAL image analysis, as well as in many other applications, anomalies are important occurrences since they often represent man-made targets or unusual situations of critical application interest. Anomalies can be considered to be outlier observations that appear inconsistent with the remainder of the data set [1]; their identification within a data set enables tasks such as automatic target detection and change detection [2]. For hyperspectral-imagery applications, in particular, anomaly detection is of paramount importance (e.g., [2]–[7]).

Most anomaly detectors identify targets that are distinct from their surroundings using no other *a priori* information beyond the fact that the targets are rare. Thus, for any lossy representation or compression, it is essential to preserve anomalies in order to enable target detection and discrimination in an unsupervised setting. However, anomalies are difficult to compress well due to the fact that their signal features are typically quite different from their surroundings. Indeed, anomalies often experience the largest distortion after traditional source coding due in large part to the fact that most lossy compression algorithms are designed—either explicitly or implicitly—to minimize some reconstruction-distortion criterion (such as a mean-square error) that is typically a global average across the entire data set and that gives little weight to rare anomalous features.

A straightforward approach to accommodate anomalies is to detect them in advance of compression and store or transmit them separately along with their location (e.g., [8] and [9]). However, anomaly detection is typically a fairly computationally expensive task that may not be amenable to encoder-side processing in severely resource-constrained sensors [6]. For example, the popular RX algorithm [10] uses a Mahalanobis distance to measure the degree of anomaly. However, a Mahalanobis distance requires an inverse covariance matrix; thus, one must acquire and store the data set in full resolution, calculate the sample covariance, and then invert the matrix. The computation and memory that this process entails may be too great for certain sensor platforms, e.g., satellite-borne devices.

However, there has been recent interest in shifting heavy computational loads from resource-constrained encoders to presumably more powerful "base-station" decoders. Much of this interest has centered on the use of sensors employing random projections to effectuate dimensionality reduction simultaneously with signal acquisition in an effort to obviate the need for explicit computation of compression within severely resource-constrained sensor devices. Much work along these lines is driven by the recent paradigm of compressed sensing (CS) (e.g., [11]–[14]), which theoretically permits a sparse signal to be sampled at sub-Nyquist rates and to be recovered with no loss. With the possibility of CS or other approaches (e.g., [15]) to recover from random projections, data can be sensed directly in a reduced dimensionality by the hardware of the sensor device. In such a setting, the ability of random projections to preserve anomalous information then becomes a critical issue.

We observe that the straightforward encoder-side approach of [8] and [9]—preserving anomalies by first detecting them and then separately compressing them—is infeasible for sensors in which the random projections occur simultaneously with signal acquisition. Consequently, we focus on anomaly detection located at the decoder side of the system along with signal reconstruction. We note too, that, although it would be possible to attempt traditional anomaly detection after a data set is reconstructed using an appropriate recovery, we are interested in the case in which anomaly detection is performed directly in the domain of the random projections. This permits recovery of the data set by first detecting anomalies and then providing separate recovery procedures for the anomaly and nonanomaly pixel classes with a goal for improved representation of the anomalies.

The authors are with the Department of Electrical and Computer Engineering and the Geosystems Research Institute, Mississippi State University, Starkville, MS 39762 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

In this paper, we investigate the interaction between anomaly detection and dimensionality reduction via random projections. The discussion that follows describes two main contributions. The first is an analysis that derives the RX anomaly detector of [10] for the situation in which anomalies are detected within random orthonormal projections. We find the solution to be a Mahalanobis distance in the projected domain similar to the solution given by [10] for the original signal domain. Additionally, we observe through random simulation that the probability that an anomaly is successfully identified by the projection-domain RX detector depends on the strength of the anomaly and the dimensionality of the projected space, with very strong anomalies likely to be correctly identified, even in low-dimensional projections. The second main contribution is a recovery procedure motivated by these analytic observations and designed to improve reconstruction of the anomalies in hyperspectral imagery. Specifically, projection-domain RX detection partitions the data set into anomaly and normal (nonanomaly) pixel classes, and separate reconstruction tailored to each class is performed. Experimental results demonstrate largely unchanged reconstruction performance for the normal pixels, whereas anomalous pixels exhibit significantly improved reconstruction.

The remainder of this paper is organized as follows: Section II presents our analytic derivation of the RX anomaly detector in the domain of random projections coupled with a random simulation of its probability of success. Next, Section III describes our proposed procedure for reconstructing anomaly and normal pixel classes separately as driven by projection-domain anomaly detection. Finally, Section IV presents experimental results, whereas Section V makes some concluding remarks. We note that, while our analysis is general and likely applicable in many settings in which anomaly detection is relevant, our experimental results are confined to hyperspectral imagery since anomaly detection is crucial in hyperspectral applications and since we have access to hyperspectral data with known ground truth, which permits objective performance assessment.

II. Anomaly Detection in Random Projections

Although it is straightforward to deploy processing after a signal has been reconstructed from random projections, it is possible to perform certain signal-processing tasks directly in the domain of the random projections—for example, [16] considers target-detection, classification, estimation, and filtering problems posed in a projected space. Here, we derive the RX anomaly detector [10] in the domain of random projections. Below, Section II-A first overviews traditional RX anomaly detection, Section II-B derives the projection-domain detector, Section II-C considers the probability of successful detection in the projected domain, and Section II-D examines the validity of the idealized noise model used in the derivation of the projected-domain detector.

A. RX Algorithm

Perhaps the most common form of anomaly detection is the RX algorithm, which originates in [10] as a generalized likelihood ratio test (GLRT) designed to detect a signal of known spatial distribution but unknown spectral distribution in a data

set of spectral pixel vectors. Specifically, consider target signal $\mathbf{b} \in \mathbb{R}^N$ (i.e., a single spectral signature of N bands), which is either present (hypothesis H_1) with strength s_m or not present (hypothesis H_0) in vector m of the data set, i.e.,

$$\mathbf{x}_m = \begin{cases} \mathbf{z}_m + \mathbf{b}s_m, & H_1, \\ \mathbf{z}_m, & H_0 \end{cases} \tag{1}$$

where $\mathbf{z}_m \in \mathbb{R}^N$ is a vector of random noise. Thus, for a data set of M vectors, $\mathbf{X} = [\mathbf{x}_1 \; \cdots \; \mathbf{x}_M]$, we have

$$\mathbf{X} = \begin{cases} \mathbf{Z} + \mathbf{b}\mathbf{s}^T, & H_1, \\ \mathbf{Z}, & H_0 \end{cases} \tag{2}$$

where $\mathbf{s} = [s_1 \; \cdots \; s_M]^T$ is the spatial distribution of $\mathbf{b}$ across the data set.

In general, there are two sources that can give rise to noise $\mathbf{z}_m$ in (1) that hinders the detection of target $\mathbf{b}$. The first is the "background" or "clutter" signal $\mathbf{c}_m$, which is simply the non-target ambient signal that exists in the absence of the target.[1] Another is sensor noise $\mathbf{n}_m$ due to the physics of the sensing device. We assume that these two noise sources are independent from one another and are additive such that $\mathbf{z}_m = \mathbf{c}_m + \mathbf{n}_m$, and

$$\mathbf{Z} = \mathbf{C} + \mathbf{N} \tag{3}$$

for the collection of M vectors. Furthermore, we assume each noise signal to be zero mean, Gaussian, and independent from pixel to pixel such that

$$\mathbf{C} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_M \otimes \boldsymbol{\Sigma}_{\mathbf{C}}) \tag{4}$$

$$\mathbf{N} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_M \otimes \boldsymbol{\Sigma}_{\mathbf{N}}) \tag{5}$$

$$\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_M \otimes \boldsymbol{\Sigma}) \tag{6}$$

where $\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}_{\mathbf{C}}$, and $\boldsymbol{\Sigma}_{\mathbf{N}}$ are the covariances of each $\mathbf{z}_m$, $\mathbf{c}_m$, and $\mathbf{n}_m$, respectively; $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{\mathbf{C}} + \boldsymbol{\Sigma}_{\mathbf{N}}$; $\mathbf{I}_M$ is the $M \times M$ identity matrix; and we follow the Kronecker notation of [17]. Clearly, this stationary noise model is highly idealized—we will return to consider the validity of this model in greater depth later in Section II-D.

Following known maximum likelihood results, [10] derives the GLRT to detect $\mathbf{b}$ in $\mathbf{X}$. Specifically, with $\mathbf{X}$ normally distributed as

$$\mathbf{X} \sim \begin{cases} \mathcal{N}(\mathbf{b}\mathbf{s}^T, \mathbf{I}_M \otimes \boldsymbol{\Sigma}), & H_1, \\ \mathcal{N}(\mathbf{0}, \mathbf{I}_M \otimes \boldsymbol{\Sigma}), & H_0 \end{cases} \tag{7}$$

[10] defines a likelihood function for hypothesis H_1 as

$$\mathcal{L}_{\mathbf{X}}(\mathbf{b}, \boldsymbol{\Sigma}) = (2\pi)^{-NM/2}|\boldsymbol{\Sigma}|^{-M/2}$$
$$\times \text{etr}\left[-\frac{1}{2}\boldsymbol{\Sigma}^{-1}(\mathbf{X} - \mathbf{b}\mathbf{s}^T)(\mathbf{X} - \mathbf{b}\mathbf{s}^T)^T\right] \tag{8}$$

and a similar function $\mathcal{L}_{\mathbf{X}}(\mathbf{0}, \boldsymbol{\Sigma})$ for hypothesis H_0; here, $\text{etr}(\mathbf{A}) = e^{\text{tr}(\mathbf{A})}$. As the centerpiece of the analysis in [10],

---

[1] We note that the original development of the RX detector [10] considered only $\mathbf{c}_m$ and not $\mathbf{n}_m$; however, the independent additive model of (3) renders the distinction between the two unnecessary for the derivation here. In the next section, however, the difference between these two noise sources becomes more relevant.

Theorem 10.1.1 of [17] is employed to derive that the GLRT, i.e.,

$$\frac{\max_{\mathbf{b},\boldsymbol{\Sigma}} \mathcal{L}_{\mathbf{X}}(\mathbf{b},\boldsymbol{\Sigma})}{\max_{\boldsymbol{\Sigma}} \mathcal{L}_{\mathbf{X}}(0,\boldsymbol{\Sigma})} \begin{array}{c} H_1 \\ \gtrless \\ H_0 \end{array} \tau_1 \tag{9}$$

reduces to

$$\frac{(\mathbf{X}\mathbf{s})^T (\mathbf{X}\mathbf{X}^T)^{-1} (\mathbf{X}\mathbf{s})}{\mathbf{s}^T \mathbf{s}} \begin{array}{c} H_1 \\ \gtrless \\ H_0 \end{array} \tau_2 \tag{10}$$

where τ_1 and τ_2 are test thresholds.

The term "RX anomaly detector" most commonly refers to a variant of (10) as used to determine whether a specific pixel vector is anomalous (e.g., [2]–[4]). Specifically, spatial pattern $\mathbf{s}$ is set to extract pixel m of interest, i.e., $\mathbf{s} = [0 \; \cdots \; 0 \, 1 \, 0 \; \cdots \; 0]$. Additionally, the mean is removed to match the zero-mean assumption on H_0 entailed by (2) and (6). Accordingly, (10) becomes

$$(\mathbf{x}_m - \boldsymbol{\mu}_{\mathbf{X}})^T \boldsymbol{\Sigma}_{\mathbf{X}}^{-1} (\mathbf{x}_m - \boldsymbol{\mu}_{\mathbf{X}}) \begin{array}{c} H_1 \\ \gtrless \\ H_0 \end{array} \tau_3 \tag{11}$$

where $\boldsymbol{\mu}_{\mathbf{X}}$ and $\boldsymbol{\Sigma}_{\mathbf{X}}$ are the sample mean and covariance, respectively, of $\mathbf{X}$. Consequently, the RX algorithm in the form of (11) is a threshold test on the well-known Mahalanobis distance.

B. RX Algorithm in Random Projections

We now derive that, when anomaly detection is deployed on orthogonally projected data, the RX algorithm is unchanged from (11); that is, the GLRT for the projected data becomes a threshold test on the Mahalanobis distance in the projected domain.

Suppose we want to distinguish between H_0 and H_1 as in (2), but, instead of observing $\mathbf{X}$, we have access only to projections of $\mathbf{X}$ in the form of

$$\widetilde{\mathbf{Y}} = \mathbf{P}^T \mathbf{X} + \widetilde{\mathbf{Z}} \tag{12}$$

where $\mathbf{P}$ is an $N \times K$ matrix that projects the N-dimensional vectors in $\mathbf{X}$ into K-dimensional vectors in $\widetilde{\mathbf{Y}}$. We assume that $\mathbf{P}$ is an orthonormal matrix in the sense that $\mathbf{P}^T \mathbf{P} = \mathbf{I}_K$. We again assume that the noise consists of both clutter and sensor noise, i.e.,

$$\widetilde{\mathbf{Z}} = \widetilde{\mathbf{C}} + \widetilde{\mathbf{N}} = \mathbf{P}^T \mathbf{C} + \widetilde{\mathbf{N}}. \tag{13}$$

In this model, the clutter noise arises in the original domain of $\mathbf{X}$ and is projected along with $\mathbf{X}$ into the domain of $\widehat{\mathbf{Y}}$. On the other hand, the sensor noise corrupts directly within the domain of the projections. This model represents the impact of noise when the random projections occur simultaneously with signal acquisition, as is assumed throughout our development. We note that [18] adopts a similar clutter/sensor noise model for a target-detection problem.

Under the noise model of (13), our hypothesis-testing problem is then

$$\widetilde{\mathbf{Y}} = \begin{cases} \widetilde{\mathbf{N}} + \mathbf{P}^T \mathbf{C} + \mathbf{P}^T \mathbf{b}\mathbf{s}^T, & H_1, \\ \widetilde{\mathbf{N}} + \mathbf{P}^T \mathbf{C}, & H_0 \end{cases} \tag{14}$$

which follows from (2). We make the same assumptions on the noise as in (4)–(6), specifically, that the cluster and sensor noises are mutually independent, zero mean, white, and Gaussian such that we have

$$\widetilde{\mathbf{N}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_M \otimes \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}) \tag{15}$$

$$\widetilde{\mathbf{C}} = \mathbf{P}^T \mathbf{C} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_M \otimes \mathbf{P}^T \boldsymbol{\Sigma}_{\mathbf{C}} \mathbf{P}) \tag{16}$$

where this last expression comes from known relations described in [[17, Sec. 2.2]. As a consequence, $\widehat{\mathbf{Y}}$ is normally distributed as

$$\widetilde{\mathbf{Y}} \sim \begin{cases} \mathcal{N}\left(\mathbf{P}^T \mathbf{b}\mathbf{s}^T, \mathbf{I}_M \otimes [\mathbf{P}^T \boldsymbol{\Sigma}_{\mathbf{C}} \mathbf{P} + \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}]\right), & H_1, \\ \mathcal{N}\left(\mathbf{0}, \mathbf{I}_M \otimes [\mathbf{P}^T \boldsymbol{\Sigma}_{\mathbf{C}} \mathbf{P} + \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}]\right), & H_0. \end{cases} \tag{17}$$

Considering the H_1 case, the likelihood function—the density of $\widetilde{\mathbf{Y}}$—depends on $\mathbf{b}$, $\boldsymbol{\Sigma}_{\mathbf{C}}$, and $\boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}$, i.e.,

$$\mathcal{L}_{\widetilde{\mathbf{Y}}}(\mathbf{b}, \boldsymbol{\Sigma}_{\mathbf{C}}, \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}) = (2\pi)^{-KM/2} |\mathbf{P}^T \boldsymbol{\Sigma}_{\mathbf{C}} \mathbf{P} + \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}|^{-M/2}$$
$$\times \mathrm{etr}\left[-\frac{1}{2} (\mathbf{P}^T \boldsymbol{\Sigma}_{\mathbf{C}} \mathbf{P} + \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}})^{-1} \right.$$
$$\left. \times (\widetilde{\mathbf{Y}} - \mathbf{P}^T \mathbf{b}\mathbf{s}^T)(\widetilde{\mathbf{Y}} - \mathbf{P}^T \mathbf{b}\mathbf{s}^T)^T \right]. \tag{18}$$

A similar expression gives $\mathcal{L}_{\widetilde{\mathbf{Y}}}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{C}}, \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}})$ for the H_0 hypothesis. The GLRT is then

$$\frac{\max_{\mathbf{b}, \boldsymbol{\Sigma}_{\mathbf{C}}, \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}} \mathcal{L}_{\widetilde{\mathbf{Y}}}(\mathbf{b}, \boldsymbol{\Sigma}_{\mathbf{C}}, \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}})}{\max_{\boldsymbol{\Sigma}_{\mathbf{C}}, \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}} \mathcal{L}_{\widetilde{\mathbf{Y}}}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{C}}, \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}})} \begin{array}{c} H_1 \\ \gtrless \\ H_0 \end{array} \tau_4. \tag{19}$$

Alternatively, we can define another likelihood function as

$$\widetilde{\mathcal{L}}_{\widetilde{\mathbf{Y}}}(\widetilde{\mathbf{b}}, \widetilde{\boldsymbol{\Sigma}}) = (2\pi)^{-KM/2} |\widetilde{\boldsymbol{\Sigma}}|^{-M/2}$$
$$\times \mathrm{etr}\left[-\frac{1}{2} \widetilde{\boldsymbol{\Sigma}}^{-1} (\widetilde{\mathbf{Y}} - \widetilde{\mathbf{b}}\mathbf{s}^T)(\widetilde{\mathbf{Y}} - \widetilde{\mathbf{b}}\mathbf{s}^T)^T \right] \tag{20}$$

such that, if $\widetilde{\mathbf{b}} = \mathbf{P}^T \mathbf{b}$ and $\widetilde{\boldsymbol{\Sigma}} = \mathbf{P}^T \boldsymbol{\Sigma}_{\mathbf{C}} \mathbf{P} + \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}$, then, clearly, $\widetilde{\mathcal{L}}_{\widetilde{\mathbf{Y}}}(\widetilde{\mathbf{b}}, \widetilde{\boldsymbol{\Sigma}}) = \mathcal{L}_{\widetilde{\mathbf{Y}}}(\mathbf{b}, \boldsymbol{\Sigma}_{\mathbf{C}}, \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}})$. For the H_0 hypothesis, we define $\widetilde{\mathcal{L}}_{\widetilde{\mathbf{Y}}}(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}})$ similarly.

We now establish that

$$\max_{\mathbf{b}, \boldsymbol{\Sigma}_{\mathbf{C}}, \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}} \mathcal{L}_{\widetilde{\mathbf{Y}}}(\mathbf{b}, \boldsymbol{\Sigma}_{\mathbf{C}}, \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}) = \max_{\widetilde{\mathbf{b}}, \widetilde{\boldsymbol{\Sigma}}} \widetilde{\mathcal{L}}_{\widetilde{\mathbf{Y}}}(\widetilde{\mathbf{b}}, \widetilde{\boldsymbol{\Sigma}}) \tag{21}$$

despite the fact that the left-hand maximization is over N-dimensional quantities $\mathbf{b}$ and $\boldsymbol{\Sigma}_{\mathbf{C}}$ (in addition to K-dimensional $\boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}$), whereas the right-hand maximization is over strictly K-dimensional quantities. First, observe that, for $\mathbf{b}^*$, $\boldsymbol{\Sigma}_{\mathbf{C}}^*$, and $\boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}^*$ that maximize $\mathcal{L}_{\widetilde{\mathbf{Y}}}$ on the left side of (21), we can define $\widetilde{\mathbf{b}}^* = \mathbf{P}^T \mathbf{b}^*$ and $\widetilde{\boldsymbol{\Sigma}}^* = \mathbf{P}^T \boldsymbol{\Sigma}_{\mathbf{C}}^* \mathbf{P} + \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}^*$ and have $\widetilde{\mathcal{L}}_{\widetilde{\mathbf{Y}}}(\widetilde{\mathbf{b}}^*, \widetilde{\boldsymbol{\Sigma}}^*) = \mathcal{L}_{\widetilde{\mathbf{Y}}}(\mathbf{b}^*, \boldsymbol{\Sigma}_{\mathbf{C}}^*, \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}^*)$. Since these $\widetilde{\mathbf{b}}^*$ and $\widetilde{\boldsymbol{\Sigma}}^*$ are necessarily in the set searched by the maximization on the right side of (21), we have

$$\max_{\widetilde{\mathbf{b}}, \widetilde{\boldsymbol{\Sigma}}} \widetilde{\mathcal{L}}_{\widetilde{\mathbf{Y}}}(\widetilde{\mathbf{b}}, \widetilde{\boldsymbol{\Sigma}}) \geq \max_{\mathbf{b}, \boldsymbol{\Sigma}_{\mathbf{C}}, \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}} \mathcal{L}_{\widetilde{\mathbf{Y}}}(\mathbf{b}, \boldsymbol{\Sigma}_{\mathbf{C}}, \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}). \tag{22}$$

On the other hand, for $\widetilde{\mathbf{b}}^*$ and $\widetilde{\boldsymbol{\Sigma}}^*$ that maximize $\widetilde{\mathcal{L}}_{\widetilde{\mathbf{Y}}}$ on the right side of (21), we can define $\mathbf{b}^* = \mathbf{P}\widetilde{\mathbf{b}}^*$, $\boldsymbol{\Sigma}_{\mathbf{C}}^* = (1/2)\mathbf{P}\widetilde{\boldsymbol{\Sigma}}^*\mathbf{P}^T$, and $\boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}^* = (1/2)\widetilde{\boldsymbol{\Sigma}}^*$. Thus, again, we have $\mathcal{L}_{\widetilde{\mathbf{Y}}}(\mathbf{b}^*, \boldsymbol{\Sigma}_{\mathbf{C}}^*, \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}^*) = \widetilde{\mathcal{L}}_{\widetilde{\mathbf{Y}}}(\widetilde{\mathbf{b}}^*, \widetilde{\boldsymbol{\Sigma}}^*)$. Since these $\mathbf{b}^*$, $\boldsymbol{\Sigma}_{\mathbf{C}}^*$, and $\boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}^*$ are necessarily in the set searched by the maximization on the left side of (21), we have

$$\max_{\mathbf{b}, \boldsymbol{\Sigma}_{\mathbf{C}}, \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}} \mathcal{L}_{\widetilde{\mathbf{Y}}}(\mathbf{b}, \boldsymbol{\Sigma}_{\mathbf{C}}, \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}) \geq \max_{\widetilde{\mathbf{b}}, \widetilde{\boldsymbol{\Sigma}}} \widetilde{\mathcal{L}}_{\widetilde{\mathbf{Y}}}(\widetilde{\mathbf{b}}, \widetilde{\boldsymbol{\Sigma}}). \tag{23}$$

The inequalities in (22) and (23) thus establish (21), and a similar argument for the H_0 hypothesis provides

$$\max_{\boldsymbol{\Sigma}_{\mathbf{C}}, \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}} \mathcal{L}_{\widetilde{\mathbf{Y}}}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{C}}, \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}) = \max_{\widetilde{\boldsymbol{\Sigma}}} \widetilde{\mathcal{L}}_{\widetilde{\mathbf{Y}}}(\widetilde{\mathbf{0}}, \widetilde{\boldsymbol{\Sigma}}). \tag{24}$$

As a consequence, the GLRT of (19) can be equivalently constructed in the projected domain as

$$\frac{\max_{\widetilde{\mathbf{b}}, \widetilde{\boldsymbol{\Sigma}}} \widetilde{\mathcal{L}}_{\widetilde{\mathbf{Y}}}(\widetilde{\mathbf{b}}, \widetilde{\boldsymbol{\Sigma}})}{\max_{\widetilde{\boldsymbol{\Sigma}}} \widetilde{\mathcal{L}}_{\widetilde{\mathbf{Y}}}(\mathbf{0}, \widetilde{\boldsymbol{\Sigma}})} \underset{H_0}{\overset{H_1}{\gtrless}} \tau_4. \tag{25}$$

At this point, we observe that (20) has the same form as (8), and (25) has the same form as (9), with $\widetilde{\mathbf{Y}}$, $\widetilde{\mathbf{b}}$, and $\widetilde{\boldsymbol{\Sigma}}$ replacing $\mathbf{X}$, $\mathbf{b}$, and $\boldsymbol{\Sigma}$, respectively. As a consequence, the analysis in [10] applies verbatim, meaning that the test in (25) can be reduced to an expression similar to (10), namely

$$\frac{(\widetilde{\mathbf{Y}}\mathbf{s})^T(\widetilde{\mathbf{Y}}\widetilde{\mathbf{Y}}^T)^{-1}(\widetilde{\mathbf{Y}}\mathbf{s})}{\mathbf{s}^T\mathbf{s}} \underset{H_0}{\overset{H_1}{\gtrless}} \tau_5. \tag{26}$$

Like (11), the resulting RX anomaly detector then becomes

$$(\widetilde{\mathbf{y}}_m - \widetilde{\boldsymbol{\mu}}_{\widetilde{\mathbf{Y}}})^T\widetilde{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{Y}}}^{-1}(\widetilde{\mathbf{y}}_m - \widetilde{\boldsymbol{\mu}}_{\widetilde{\mathbf{Y}}}) \underset{H_0}{\overset{H_1}{\gtrless}} \tau_6 \tag{27}$$

where $\widetilde{\boldsymbol{\mu}}_{\widetilde{\mathbf{Y}}}$ and $\widetilde{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{Y}}}$ are the sample mean and covariance, respectively, of $\widetilde{\mathbf{Y}}$.

C. Detection of Anomalies

The previous section derives that, given N-dimensional data with an unknown spectral-signature target, the GLRT for detecting the target from orthonormal K-dimensional projections of the data results in a test [i.e., (26)] with the same exact form that the test would take in the original N-dimensional space [i.e., (10)]. Consequently, the RX algorithm for anomaly detection similarly becomes a threshold test [i.e., (27)] on the Mahalanobis distance in the K-dimensional projected domain, identical to the form of the test in the original N-dimensional domain, [i.e., (11)]. This analysis, however, tells us nothing about whether a pixel identified as anomalous by (11) in the original signal domain would also be identified as anomalous by (27) in the projected domain. That is, for given target anomaly $\mathbf{b}$, it is possible for a particular projection $\mathbf{P}$ to result in significant foreshortening, such that $\mathbf{P}^T\mathbf{b}$ has very small length, even if $\mathbf{b}$ itself happens to be quite long. In such a case, (27) may choose H_0, even though (11) would choose H_1.

To gain some insight into this issue, we consider the probabilities of detection and false alarm of the RX detector. Specifically, for the RX detector in the original signal domain [i.e., (11)], [10] defines a "generalized" signal-to-noise ratio (SNR) as

$$\lambda = \mathbf{b}^T\boldsymbol{\Sigma}_{\mathbf{X}}^{-1}\mathbf{b} \tag{28}$$

where λ measures the strength of target anomaly $\mathbf{b}$ against that of the background clutter. Then, for $\boldsymbol{\mu}_{\mathbf{X}} = \mathbf{0}$, [10] derives the probability of false alarm to be

$$P_F(\tau) = 1 - I_\tau\left(\frac{N}{2}, \frac{M-N}{2}\right) \tag{29}$$

where τ is the test threshold, and $I_x(\alpha, \beta)$ is the regularized incomplete beta function (i.e., the cumulative distribution of a beta density); the probability of detection is

$$P_D(\tau, \lambda) = 1 - I_\tau\left(\frac{N}{2}, \frac{M-N}{2}; \lambda\right) \tag{30}$$

where $I_x(\alpha, \beta; \lambda)$ is the incomplete noncentral beta function (i.e., the cumulative distribution of a noncentral beta density for noncentrality parameter λ).

Cast into the projected domain, and assuming fixed $\boldsymbol{\Sigma}_{\mathbf{C}}$ and $\boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}$, the probabilities for the projected-domain detector (27) become

$$\widetilde{P}_F(\tau) = 1 - I_\tau\left(\frac{K}{2}, \frac{M-K}{2}\right) \tag{31}$$

$$\widetilde{P}_D\left(\tau, \widetilde{\lambda}(\mathbf{b}, \mathbf{P})\right) = 1 - I_\tau\left(\frac{K}{2}, \frac{M-K}{2}; \widetilde{\lambda}(\mathbf{b}, \mathbf{P})\right) \tag{32}$$

where the projected-domain generalized SNR, i.e.,

$$\begin{aligned}\widetilde{\lambda}(\mathbf{b}, \mathbf{P}) &= \widetilde{\mathbf{b}}^T\widetilde{\boldsymbol{\Sigma}}_{\widetilde{\mathbf{Y}}}^{-1}\widetilde{\mathbf{b}} \\ &= \mathbf{b}^T\mathbf{P}[\mathbf{P}^T\boldsymbol{\Sigma}_{\mathbf{C}}\mathbf{P} + \boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}]^{-1}\mathbf{P}^T\mathbf{b}\end{aligned} \tag{33}$$

depends on both $\mathbf{b}$ and $\mathbf{P}$.

In order to empirically evaluate $\widetilde{P}_D$ for fixed $\widetilde{P}_F$, we consider a simplified special case in which the clutter covariance is $\boldsymbol{\Sigma}_{\mathbf{C}} = \sigma^2\mathbf{I}_N$, whereas the sensor noise is $\boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}} = \eta\sigma^2\mathbf{I}_K$ for $\eta \ll 1$. Given a constant probability of a false alarm, fixed target anomaly $\mathbf{b}$, and these $\boldsymbol{\Sigma}_{\mathbf{C}}$ and $\boldsymbol{\Sigma}_{\widetilde{\mathbf{N}}}$, we are interested in the probability of detection with projection matrix $\mathbf{P}$ randomly chosen from the set of all orthonormal $N \times K$ matrices. We can instead consider an equivalent "normalized" problem wherein we fix $\mathbf{b}$ to unit norm, fix projection $\mathbf{P}$, and have $\mathbf{b}$ range randomly over the surface of the unit hypersphere in N-dimensional space (see, e.g., [19] and [20]). For this equivalent problem, it is convenient to set $\mathbf{P}$ to be the first K columns of $\mathbf{I}_N$. Under these conditions, the generalized SNR (33) depends only on $\mathbf{b}$ as

$$\widetilde{\lambda}(\mathbf{b}) = \frac{\|\mathbf{P}^T\mathbf{b}\|^2}{(1+\eta)\sigma^2} \tag{34}$$

and the probability of detection of interest becomes

$$\begin{aligned}\widetilde{P}_D(\tau) &= \int_{\mathbf{b}} f(\mathbf{b})\widetilde{P}_D\left(\tau, \widetilde{\lambda}(\mathbf{b})\right) d\mathbf{b} \\ &= \int_{\mathbf{b}} f(\mathbf{b})\left[1 - I_t\left(\frac{K}{2}, \frac{M-K}{2}; \widetilde{\lambda}(\mathbf{b})\right)\right] d\mathbf{b}\end{aligned} \tag{35}$$
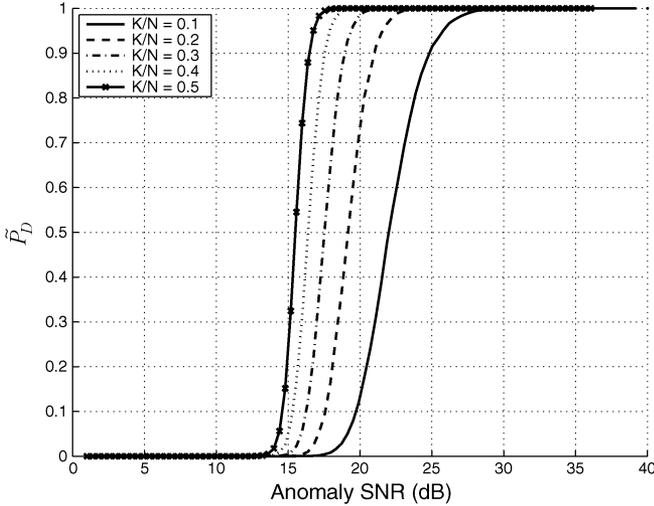
Fig. 1. Probability $\widetilde{P}_D$ of successful detection of a fixed anomaly from a random projection as a function of the anomaly SNR $= -10 \log_{10} \sigma^2$ in the original signal domain and subrate K/N of the projection. The curves are for a fixed probability of false alarm, i.e., $\widetilde{P}_F = 10^{-5}$, original signal dimension $N = 100$, a data set of $M = 1000$ vectors, sensor noise $\eta = 0.01$, and Monte Carlo simulation with $10\,000$ random trials.

where $f(\mathbf{b})$ is the probability density of $\mathbf{b}$. We resort to Monte Carlo integration to evaluate (35).

Fig. 1 plots $\widetilde{P}_D$ from (35) as a function of both the anomaly SNR in the original signal domain, i.e., $10 \log_{10}(\|\mathbf{b}\|^2/\sigma^2) = -10 \log_{10} \sigma^2$, and ratio K/N, which we call the subsampling rate, or subrate; a fixed $\widetilde{P}_F = 10^{-5}$ in (31) is assumed, and we arbitrarily set $N = 100$, $M = 1000$, and $\eta = 0.01$. In this plot, we can see that, for a strongly anomalous $\mathbf{b}$ (i.e., large SNR in Fig. 1), the probability that the RX detector in the projected domain will identify the projected vector $\mathbf{P}^T\mathbf{b}$ as anomalous is high, even for low subrate. On the other hand, for a weakly anomalous $\mathbf{b}$ (i.e., low SNR), it is unlikely that the projected vector will be identified as anomalous. Finally, for moderately anomalous $\mathbf{b}$, a higher subrate yields a higher chance of successful detection in the projected domain.

While it would be possible to carry out similar simulations for more general cases of specific nonidentity clutter and noise covariances, we instead consider a real hyperspectral data set later in Section IV-A to examine the success of RX detection in the projected domain. We see there that, empirically, strong anomalies are likely to be successfully detected in the projected domain, even at low subrate.

D. Noise Model and Local Mean Removal

The noise model represented by (3)–(6) is highly idealized. However, it is common in hyperspectral literature (e.g., [21] and [22]) to assume sensor noise that is white, zero mean, Gaussian, and independent from the other information-bearing signal. Thus, we briefly examine here the issue of the background clutter. It is argued in [23] and [24] that, although the clutter signal rarely possesses the white Gaussian distribution of (4), once a nonstationary local mean is subtracted, the resulting residual becomes, for many forms of imagery, not only approximately Gaussian, but white as well. Thus, the original RX

detector was derived [10] under the assumption of prewhitening via local mean removal.

Specifically, let vector m of the original (non-Gaussian) clutter signal be $\mathbf{c}'_m$ with local mean, i.e.,

$$\bar{\mathbf{c}}'_m = E\left[\mathbf{c}'_m\right]. \tag{36}$$

Typically, the local mean is adaptively estimated over a sliding window surrounding the pixel in question. For $\mathbf{C}' = [\mathbf{c}'_1 \; \cdots \; \mathbf{c}'_M]$ and $\bar{\mathbf{C}}' = [\bar{\mathbf{c}}'_1 \; \cdots \; \bar{\mathbf{c}}'_M]$, [23] and [24] argue that

$$\mathbf{C} = \mathbf{C}' - \bar{\mathbf{C}}' \tag{37}$$

is effectively distributed as a white Gaussian, as in (4). As a consequence, rather than performing (11) with sample mean $\boldsymbol{\mu}_\mathbf{X}$ globally calculated on the entire data set, many RX-detector implementations use a local mean in practice.[2]

Now, in our derivation of the RX detector as deployed on projected data, the hypothesis test of (14) has noise $\widetilde{\mathbf{Z}}'$ as

$$\widetilde{\mathbf{Z}}' = \mathbf{P}^T\mathbf{C}' + \widetilde{\mathbf{N}}. \tag{38}$$

If we apply the RX detector in the projected domain identically to as it is typically applied to the original signal, we perform local mean removal to yield the following:

$$\widetilde{\mathbf{Z}} = \widetilde{\mathbf{Z}}' - \bar{\widetilde{\mathbf{Z}}}' \tag{39}$$

where $\widetilde{\mathbf{Z}}' = [\widetilde{\mathbf{z}}'_1 \; \cdots \; \widetilde{\mathbf{z}}'_M]$, $\bar{\widetilde{\mathbf{Z}}}' = [\bar{\widetilde{\mathbf{z}}}'_1 \; \cdots \; \bar{\widetilde{\mathbf{z}}}'_M]$, and

$$\bar{\widetilde{\mathbf{z}}}'_m = E\left[\widetilde{\mathbf{z}}'_m\right] = E\left[\mathbf{P}^T\mathbf{c}'_m + \widetilde{\mathbf{n}}_m\right] = \mathbf{P}^T\bar{\mathbf{c}}'_m \tag{40}$$

under the reasonable assumption that sensor noise $\widetilde{\mathbf{N}}$ has zero mean. Thus, substituting (38) and (40) into (39) yields

$$\begin{aligned}\widetilde{\mathbf{Z}} &= \mathbf{P}^T\mathbf{C}' + \widetilde{\mathbf{N}} - \mathbf{P}^T\bar{\mathbf{C}}' \\ &= \mathbf{P}^T(\mathbf{C}' - \bar{\mathbf{C}}') + \widetilde{\mathbf{N}} \\ &= \mathbf{P}^T\mathbf{C} + \widetilde{\mathbf{N}} \tag{41}\end{aligned}$$

where the last expression uses (37). We note that (41) is identical to (13). Thus, we conclude that, if the ambient-domain clutter can be whitened via local mean removal, as in (37), to yield a Gaussian distribution, as in (4), then (41) implies that projection-domain noise removal, as in (39), produces projection-domain clutter that is also white Gaussian, as in (16).

III. ANOMALY-BASED RECONSTRUCTION FROM RANDOM PROJECTIONS

Given a data set of M vectors $\mathbf{X} = [\mathbf{x}_1 \; \cdots \; \mathbf{x}_M]$, where each $\mathbf{x}_m \in \mathbb{R}^N$, we assume the signal-acquisition device applies $N \times K$ orthonormal random projection $\mathbf{P}$ to obtain random projections $\widetilde{\mathbf{Y}} = [\widetilde{\mathbf{y}}_1 \; \cdots \; \widetilde{\mathbf{y}}_M]$, where each $\widetilde{\mathbf{y}}_m = \mathbf{P}^T\mathbf{x}_m$ has dimension K. To reconstruct an approximation $\widehat{\mathbf{X}}$ from projections $\widetilde{\mathbf{Y}}$, we could apply one of several existing algorithms; we overview two below, specifically, multitask Bayesian compressive sensing (MT-BCS) [25] in Section III-A and compressive-projection principal component analysis (CPPCA) [15] in Section III-B. Then, in Section III-C, we present a strategy that combines CPPCA and MT-BCS reconstructions with the spe-

---

[2]Covariance $\boldsymbol{\Sigma}_\mathbf{X}$ may be also locally estimated as well.

cific goal of producing significantly better recovery of anomalies in a hyperspectral image.

A. MT-BCS

In order to reconstruct a vector from random projections, the first strategy that probably comes to mind is CS. In brief, CS (e.g., [11]–[14]) produces a sparse signal representation directly from a small number of projections onto another basis, recovering the sparse transform coefficients via nonlinear reconstruction. The main tenet of CS theory holds that, if signal $\mathbf{x} \in \mathbb{R}^N$ can be sparsely represented (i.e., using only L nonzero coefficients) with some basis, then we can recover $\mathbf{x}$ from K-dimensional projection $\widetilde{\mathbf{y}} = \mathbf{P}^T\mathbf{x}$ under certain conditions, even when $K \ll N$. For recovery of a set of multiple, possibly correlated vectors $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_M]$, there have been proposals for multivector extensions of CS under the name of "multitask" [25] or "distributed" [26] CS; these, in turn, closely link to a larger body of literature on "simultaneous sparse approximation" (e.g., [27]–[31]). Below, we focus on MT-BCS [25], which introduces a hierarchical Bayesian framework into the multivector CS-recovery problem to share prior information across the multiple vectors. MT-BCS is particularly convenient for our work here in that it can function on arbitrarily small numbers of vectors, even down to $M = 1$ (in which case, MT-BCS becomes the special case of "single-task" CS recovery).

B. CPPCA

Principal component analysis (PCA) has been extensively used for dimensionality reduction and compression in many signal-processing applications. However, PCA is a data-dependent transform arising from the eigendecomposition of the covariance matrix of the signal in question. Thus, in traditional compression and communication applications using PCA, the encoder must calculate the PCA transform before it can be applied to the data; however, the computational burden that this process entails may well exceed the limited capabilities of many encoding platforms. In [15], we presented CPPCA, which is a technique that effectively shifts the computational burden of PCA from a resource-constrained encoder to the decoder. On the encoder side, CPPCA is driven by projections at the sensor onto lower-dimensional subspaces chosen at random. The CPPCA decoder, given only these random projections, recovers not only the coefficients associated with the PCA transform but also an approximation to the PCA transform basis itself.

CPPCA reconstruction of a set of randomly projected vectors first entails an eigenvector-reconstruction process based on an approximation that uses Ritz vectors as close representations of orthonormal projections of eigenvectors. Specifically, the PCA transform of $\mathbf{x}_m$ in data set $\mathbf{X} = [\mathbf{x}_1 \cdots \mathbf{x}_M]$ is $\breve{\mathbf{x}}_m = \mathbf{W}^T\mathbf{x}_m$, where $N \times N$ transform matrix $\mathbf{W}$ emanates from the eigendecomposition of $\mathbf{\Sigma}_{\mathbf{X}}$, i.e.,

$$\mathbf{\Sigma}_{\mathbf{X}} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^T \tag{42}$$

where $\mathbf{W}$ contains the N unit eigenvectors of $\mathbf{\Sigma}_{\mathbf{X}}$ columnwise. CPPCA reconstruction uses the first L Ritz vectors (essentially the eigenvectors of $\widetilde{\mathbf{\Sigma}}_{\widetilde{\mathbf{Y}}} = \mathbf{P}^T\mathbf{\Sigma}_{\mathbf{X}}\mathbf{P}$) to obtain approximations

of the first L principal eigenvectors in $\mathbf{W}$ corresponding to the L largest eigenvalues in $\mathbf{\Lambda}$. These approximate eigenvectors are then assembled into $N \times L$ matrix $\mathbf{\Psi}$. The reconstruction of the data set $\mathbf{X}$ is then produced in a pseudoinverse-based recovery of the PCA coefficients, i.e.,

$$\widehat{\mathbf{X}} = \mathbf{\Psi}\left(\mathbf{P}^T\mathbf{\Psi}\right)^+ \widetilde{\mathbf{Y}}. \tag{43}$$

The reader is referred to [15] for greater detail.

C. Reconstruction via Partitioning and Residue

It is well known that PCA does not tend to provide efficient representation of anomalies within a data set. That is, a basis composed of the principal eigenvectors tends to represent well only the major signal features that are present in abundance throughout the data set. On the other hand, anomalies, which by definition rarely occur, do not appear in any one component but rather tend to be spread out across all the minor components of the PCA representation. That is, to represent an anomaly, typically nonnegligible transform coefficients are needed for all the minor eigenvectors. Thus, an efficient compact representation of anomalies is not usually possible with a PCA basis, and CPPCA inherits from PCA this difficulty. However, if anomaly detection can be successfully performed on the random projections $\widetilde{\mathbf{Y}}$ directly, then the reconstruction can process the normal and anomaly pixels differently. Consequently, we propose the scheme illustrated in Fig. 2 wherein the data set is partitioned into normal and anomaly pixel classes based on a projection-domain anomaly detection in order to apply a reconstruction tailored to each class. Thus, the first step in the reconstruction in Fig. 2 is to partition $\widetilde{\mathbf{Y}}$ into normal pixels $\widetilde{\mathbf{Y}}_n$ and anomaly pixels $\widetilde{\mathbf{Y}}_a$ based on projection-domain RX detection, i.e., (27), in Section II-B.

For normal pixels $\widetilde{\mathbf{Y}}_n$, CPPCA reconstruction, as described in Section III-B, is applied; this is due to the fact that the experimental results in [15] indicate that CPPCA outperforms MT-BCS by a wide margin for the reconstruction of a hyperspectral data set as a whole. In this CPPCA reconstruction, L eigenvectors of the PCA transform for $\mathbf{X}$ are approximated and assembled in $N \times L$ matrix $\mathbf{\Psi}$. The approximation to the normal coefficients $\mathbf{X}_n$ is then produced via (43), i.e.,

$$\widehat{\mathbf{X}}_n = \mathbf{\Psi}(\mathbf{P}^T\mathbf{\Psi})^+ \widetilde{\mathbf{Y}}_n. \tag{44}$$

To recover the anomaly pixels, we note that, in many cases, even though anomaly pixels can be quite different from normal pixels, it is unlikely that the anomalies are completely uncorrelated with the normal pixels since each pixel identified as an anomaly is likely to be in actuality some mixture of both anomaly and normal spectral signatures. As a consequence, our recovery of anomaly pixels consists of first removing a prediction of the normal-pixel content from the anomalies and then performing a CS reconstruction on the resulting residual. Specifically, in reference to Fig. 2, we predict the anomaly pixels using only L' eigenvectors out of the L eigenvectors approximated via CPPCA over the normal pixels, i.e.,

$$\widehat{\mathbf{X}}_p = \mathbf{\Psi}'(\mathbf{P}^T\mathbf{\Psi}')^+ \widetilde{\mathbf{Y}}_a \tag{45}$$
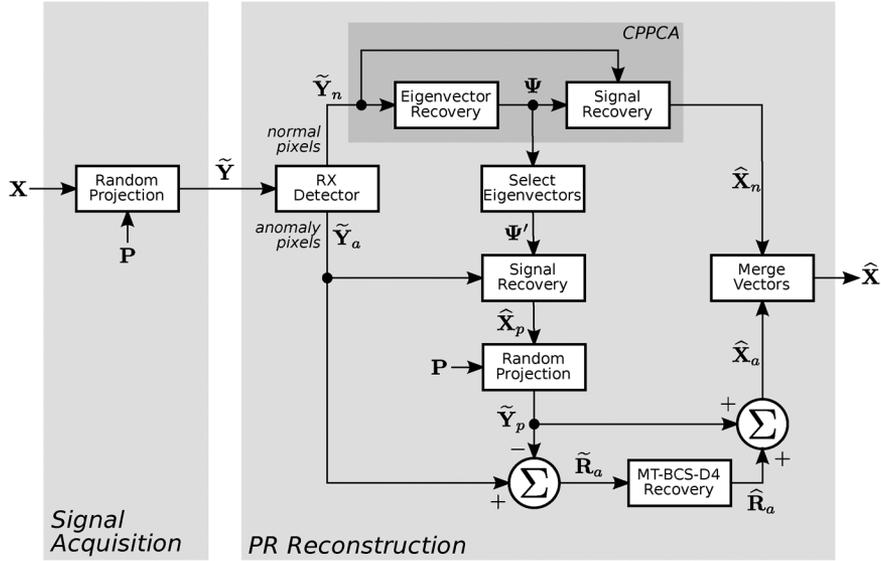
Fig. 2. PR algorithm.

where $\mathbf{\Psi}'$ consists of the first L' columns of $\mathbf{\Psi}$. We then apply CS reconstruction to the projection-domain residual as follows:

$$\widetilde{\mathbf{R}}_a = \widetilde{\mathbf{Y}}_a - \widetilde{\mathbf{Y}}_p \qquad (46)$$

where $\widetilde{\mathbf{Y}}_p = \mathbf{P}^T \widehat{\mathbf{X}}_p$. We employ MT-BCS with a Daubechies length-8 wavelet as the sparsity transform to reconstruct the residual $\widehat{\mathbf{R}}_a$ from the projected residual $\widetilde{\mathbf{R}}_a$, resulting in the reconstructed anomalies being formed as

$$\widehat{\mathbf{X}}_a = \widetilde{\mathbf{Y}}_p + \widehat{\mathbf{R}}_a. \qquad (47)$$

We note that MT-BCS is used for this residual reconstruction since CPPCA requires a relatively large number of vectors to function well, whereas the number of anomalies is necessarily small. Finally, since our technique involves both partitioning the data set into normal and anomaly pixel classes, as well as a residual reconstruction, we refer to it as partitioned-residual (PR) reconstruction.

IV. EXPERIMENTAL RESULTS

In order to evaluate the efficacy of the PR scheme previously proposed, we consider the two hyperspectral images used in [4]. The first is the hyperspectral digital image collection experiment (HYDICE) image depicted in Fig. 3, whereas the second is the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) image of the Lunar Crater Volcanic Field (LCVF) scene shown in Fig. 4.

The HYDICE image consists of 64×64 pixel vectors acquired in 210 spectral bands over the range of 0.4 to 2.5 μm. High-noise and water-absorption bands have been removed, leaving a total of 169 spectral bands. The spatial resolution is 1.56 m. The scene depicts 15 panels in a 5×3 arrangement; the panels have varying sizes and are painted with various substances. A precise ground-truth location of these panels yields a map of 19 anomalous pixels, as depicted in Fig. 3(b). This ground-truth map thus partitions the hyperspectral data
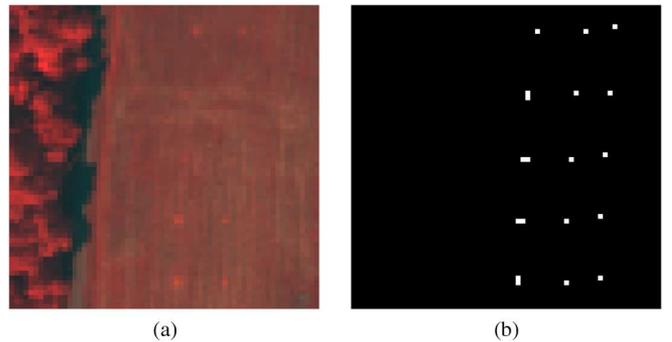

Fig. 3. Fifteen-panel HYDICE scene. (a) A color infrared representation. (b) The ground-truth map of the anomalies consisting of 19 anomalous pixels. white = anomaly pixels. black = normal pixels.
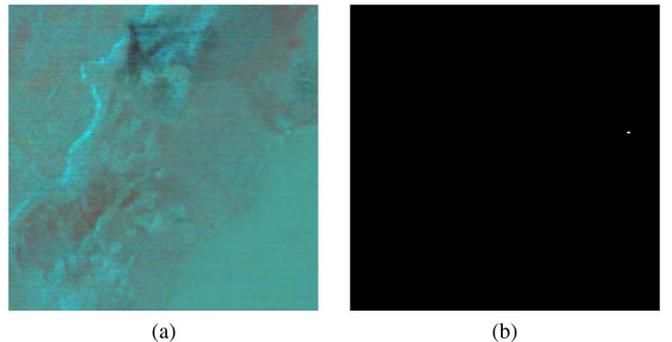

Fig. 4. LCVF scene. (a) A color infrared representation. (b) The ground-truth map of the anomalies consisting of two anomalous pixels. white = anomaly pixels. black = normal pixels.

set into two classes of pixels—"normal" nonanomalous pixels [black in Fig. 3(b)] and "anomaly" pixels [white in Fig. 3(b)].

The LCVF image consists of 200×200 pixel vectors acquired in 224 spectral bands over the range of 0.4 to 2.5 μm. High-noise and water-absorption bands have been removed, leaving a total of 158 spectral bands. The spatial resolution

is 30 m. As discussed in [4] and [32], the LCVF image has a two-pixel anomaly, as depicted in the ground-truth map in Fig. 4(b).

Below, we will measure performance over the normal and anomaly pixel classes separately based on the ground-truth maps in Figs. 3(b) and 4(b). For a given vector $\mathbf{x}_m$, we can measure the quality of its reconstruction in several ways, for instance, by using an SNR or a spectral-angle distortion measure. In the following results, we use a vector-based SNR measured in decibels, i.e.,

$$\text{SNR}(\mathbf{x}_m, \widehat{\mathbf{x}}_m) = 10 \log_{10} \frac{\text{var}(\mathbf{x}_m)}{\text{MSE}(\mathbf{x}_m, \widehat{\mathbf{x}}_m)} \quad (48)$$

where the $\text{var}(\mathbf{x}_m)$ is the variance of the components of vector $\mathbf{x}_m$, and the mean-squared error (MSE) is

$$\text{MSE}(\mathbf{x}_m, \widehat{\mathbf{x}}_m) = \frac{1}{N} \|\mathbf{x}_m - \widehat{\mathbf{x}}_m\|^2. \quad (49)$$

We then average the vector-based SNR over all vectors of the specific pixel class (normal or anomaly) of interest. Alternatively, we can measure the quality of a reconstructed hyperspectral data set using an average spectral angle, where the spectral angle in degrees between the reconstructed hyperspectral pixel vector and its corresponding original vector is averaged over the pixel class, i.e., $\bar{\xi} = \text{mean}(\xi_m)$ where

$$\xi_m = \angle(\mathbf{x}_m - \widehat{\mathbf{x}}_m). \quad (50)$$

A. Anomaly Detection in the Domain of Random Projections

As discussed in Section II-A, a popular approach to automatically detecting anomalies in a hyperspectral image is to apply the RX anomaly detector (11) to the spectral pixel vectors. The result obtained when the RX detector is applied to the HYDICE image in Fig. 3(a) and the threshold set such that exactly 19 pixels yield detector output above the threshold is shown in Fig. 5(a).

The PR algorithm previously proposed in Section III-C is driven by partitioning the data set into normal/anomaly classes by using (27) directly in the random projections produced by the CS-based sensor. The RX detector, again instructed to identify 19 anomalies, applied in the random-projection domain for a subrate of 10%–30% yields the results shown in Fig. 5(b)–(d) for the HYDICE image. We note that, although the maps in Fig. 5(b)–(d) differ somewhat from the true ground-truth map in Fig. 3(b), over 99.5% of the pixels are correctly classified in all cases. Additionally, even though the entire rightmost column of panels is not detected, these panels are not detected by the RX detector applied to the original pixel vectors in Fig. 5(a); this is likely due to the fact that the panels in the rightmost column are 1 m^2 in size, below the 1.56-m spatial resolution of the pixels. Overall, the RX detector applied to the random projections produces anomaly-map results similar to those of the RX detector applied to the original data. This is as expected from the simulations in Section II-C.

Additionally, the receiver operator characteristic (ROC) that plots the probability of detection against the probability of false alarm as the threshold of the RX detector is varied is shown in Fig. 6 for the HYDICE image. Fig. 6 gives the ROC curves
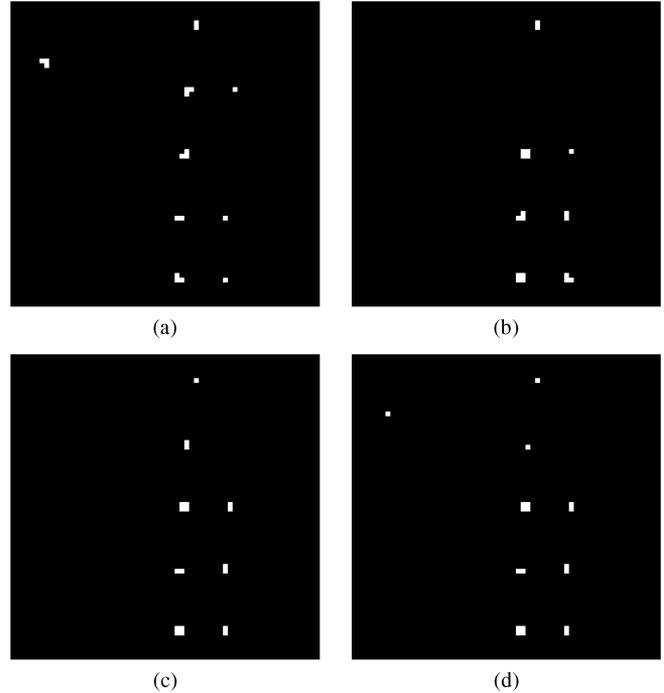

Fig. 5. Anomalies as identified by the RX detector for the 15-panel HYDICE scene. white = anomaly pixels. black = normal pixels. (a) RX detection on the original image [99.658% of pixels correctly identified with respect to ground-truth map in Fig. 3(b)]. (b) RX detection on random projections with subrate $K/N = 0.1$ (99.512% of pixels correctly identified). (c) RX detection with subrate $K/N = 0.2$ (99.658% of pixels correctly identified). (d) RX detection with subrate $K/N = 0.3$ (99.609% of pixels correctly identified).
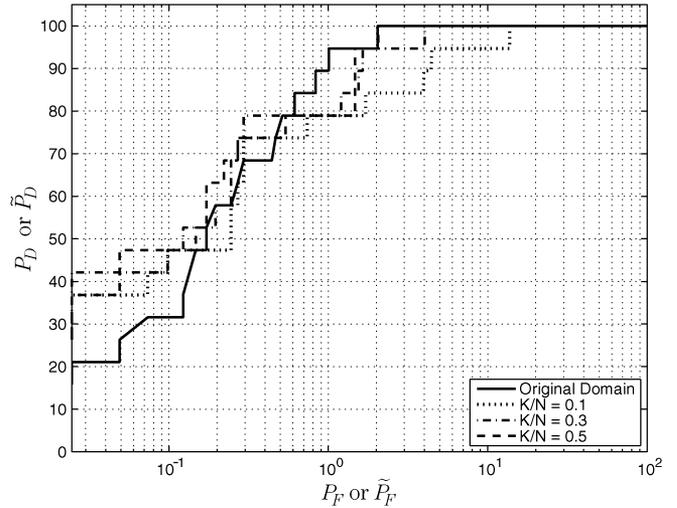

Fig. 6. ROC curves for the HYDICE scene for the RX detector in the original signal domain, as well as in the projected domain for subrates of 10%, 30%, and 50%.

for both the original signal domain and the projected domain for various subrates. Interestingly, Fig. 6 reveals that the projected-domain RX detector tends to achieve a somewhat higher probability of detection at low false-alarm rates, whereas the original-signal-domain detector appears to do slightly better at higher false-alarm rates.

For the LCVF image, the two anomaly pixels are quite strongly anomalous. Thus, the RX detector applied in the
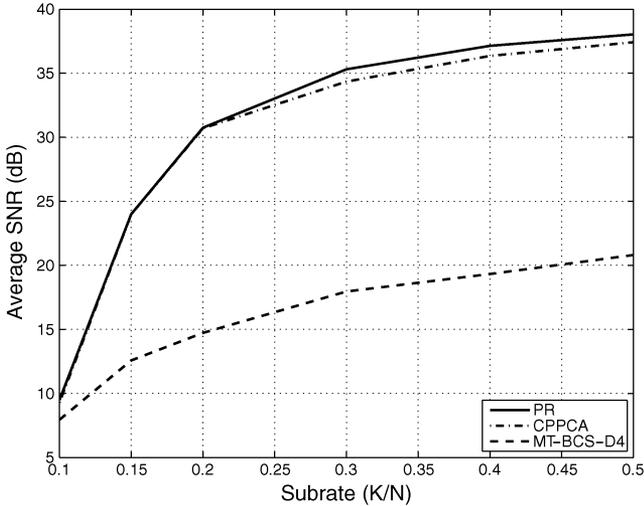
Fig. 7. HYDICE scene—average SNR performance over pixels identified as normal (nonanomalous) in the ground-truth map in Fig. 3(b).
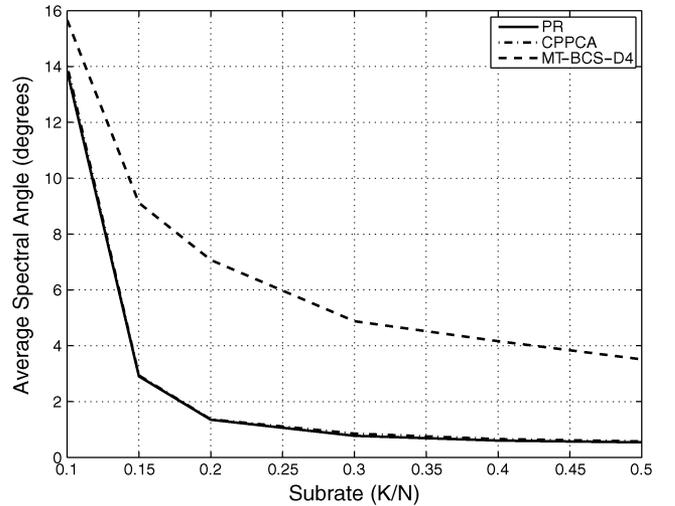

Fig. 9. HYDICE scene—average spectral-angle performance over pixels identified as normal (nonanomalous) in the ground-truth map in Fig. 3(b).
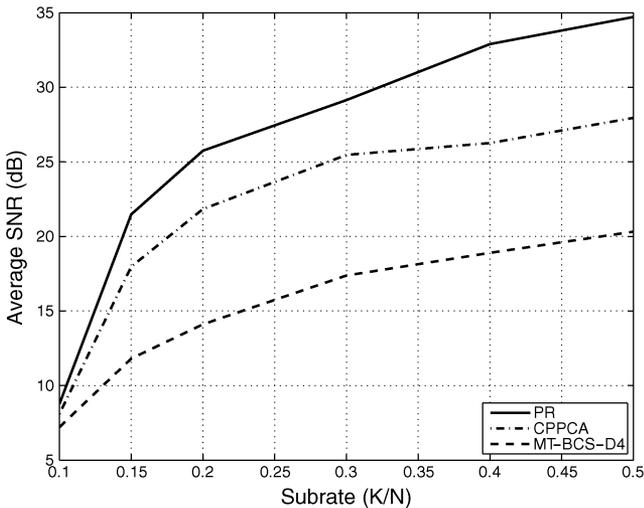

Fig. 8. HYDICE scene—average SNR performance over pixels identified as anomalous in the ground-truth map in Fig. 3(b).
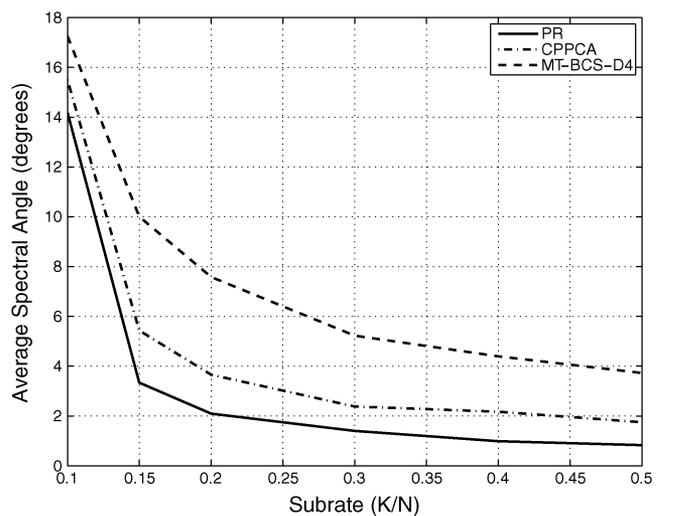

Fig. 10. HYDICE scene—average spectral-angle performance over pixels identified as anomalous in the ground-truth map in Fig. 3(b).

random-projection domain with threshold set such that two pixels yield detector output above the threshold achieves 100% anomaly-normal pixel classification for every subrate tested (10%–50%). The resulting projection-domain anomaly maps are consequently identical to Fig. 4(b) in every case.

B. Performance of the PR Algorithm

The SNR performance for the HYDICE image in terms of various reconstruction strategies is depicted for normal and anomaly pixels in Figs. 7 and 8, respectively. These graphs show the performance of the PR algorithm previously proposed in Section III-C, wherein the RX detector is applied to the random projections to partition the data set into normal/anomaly pixel classes, with CPPCA reconstructing the normal class and MT-BCS applied for a residual reconstruction of the anomalies. In these plots, we compare the PR approach to two other, more straightforward, reconstructions—"CPPCA" refers to a reconstruction using CPPCA to reconstruct every pixel of the data set, whereas "MT-BCS-D4" refers to a reconstruction using

MT-BCS for every pixel. We note that all three techniques use exactly the same random projections. For CPPCA, we recover $L = 10$ eigenvectors; we use the implementation available from the CPPCA website.[3] For MT-BCS, we use a Daubechies length-8 wavelet as the sparsity basis and the implementation available from its authors.[4] Prediction $\widehat{\mathbf{X}}_p$ is formed using $\mathbf{\Psi}'$ with the $L' = 3$ principal eigenvectors. We note that, in all cases, performance is measured by partitioning the HYDICE data set into normal and anomaly pixels according to the ground-truth map in Fig. 3(b); this means that, in the case of PR, the normal/anomaly partition for performance measurement slightly differs from that used in the actual reconstruction of the data set [i.e., as represented by Fig. 5(b)–(d)]. In Figs. 7 and 8, we see that, while the proposed PR reconstruction achieves performance very similar to CPPCA over the normal pixels of the data set, for the anomaly pixels, the story is quite

[3]http://www.ece.msstate.edu/fowler/CPPCA/

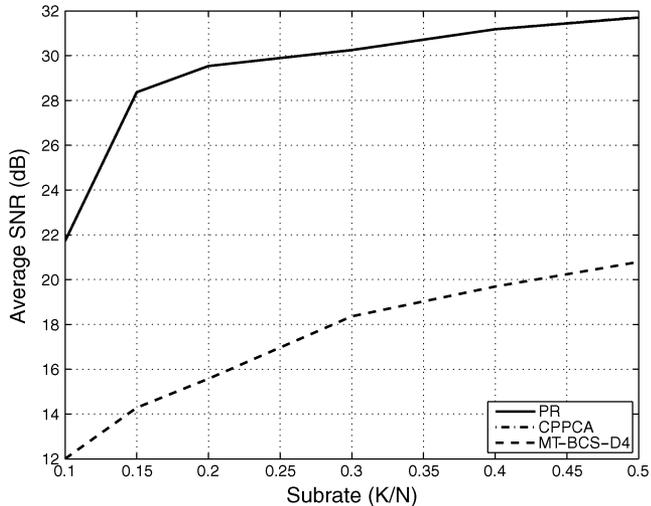[4]http://people.ee.duke.edu/lihan/cs/

Fig. 11. LCVF scene—average SNR performance over pixels identified as normal (nonanomalous) in the ground-truth map in Fig. 4(b) (the PR and CPPCA curves are coincident).
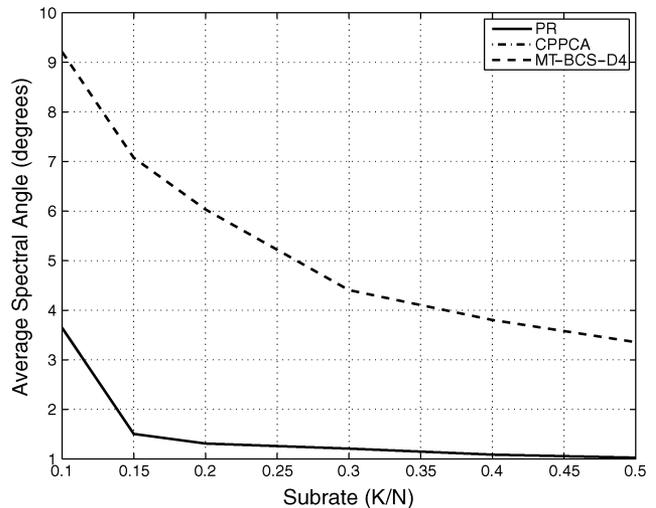

Fig. 13. LCVF scene—average spectral-angle performance over pixels identified as normal (nonanomalous) in the ground-truth map in Fig. 4(b) (the PR and CPPCA curves are coincident).
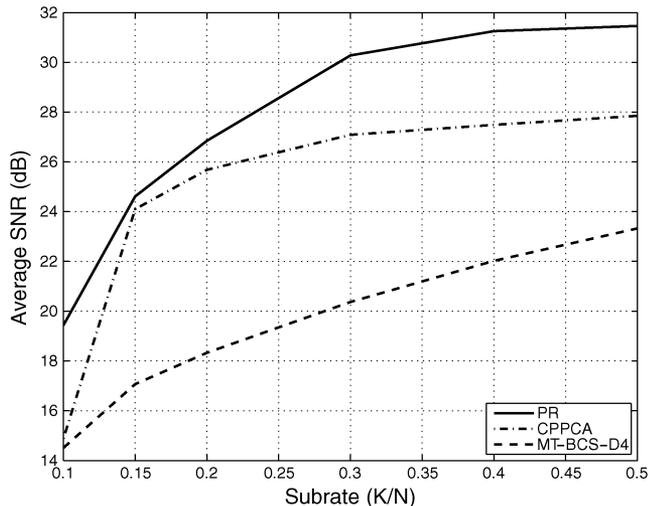

Fig. 12. LCVF scene—average SNR performance over pixels identified as anomalous in the ground-truth map in Fig. 4(b).
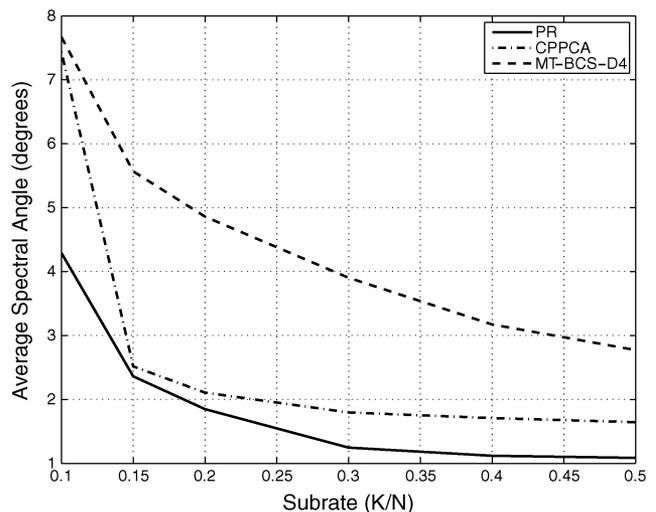

Fig. 14. LCVF scene—average spectral-angle performance over pixels identified as anomalous in the ground-truth map in Fig. 4(b).

different—except at the lowest subrates, PR provides a significantly superior representation of the anomaly pixels, achieving some 3–7 dB higher SNR for the subrates considered.

Similar conclusions are shown in Figs. 9 and 10, which depict performance in terms of average spectral angle $\bar{\xi}$ for the HYDICE image. We see that, for anomaly pixels, PR achieves $1°–2°$ smaller $\bar{\xi}$ than CPPCA alone and $3°–6°$ smaller $\bar{\xi}$ than MT-BCS alone. Finally, the corresponding SNR and spectral-angle performance for the LCVF image is presented in Figs. 11–14. Again, the PR algorithm yields significant performance improvement over both CPPCA and MT-BCS used alone.

V. CONCLUSION

The emerging CS paradigm has increased efforts to shift computational burdens from resource-constrained sensor platforms to presumably more powerful base-station decoders. These efforts assume that dimensionality reduction occurs simultaneously with signal acquisition directly within the sensor device through the use of random projections, thereby obviating the need for explicit computation of compression onboard the sensor. Since anomalies are critical to many end applications, the ability of random projections to preserve anomalous information is of paramount importance.

In this paper, we have analytically derived that the popular RX anomaly detector, when deployed in the domain of random orthonormal projections, remains unchanged in form—specifically, it is a threshold test on a Mahalanobis distance, as calculated in the projected domain. We have also observed, both in random simulations and in empirical results on real hyperspectral data with known ground truth, a high likelihood that strong anomalies are successfully detected by the projection-domain RX detector even at low subrates. Finally, we have proposed a process for reconstruction of a hyperspectral image from random projections that partitions the data set into anomaly and normal pixel classes and tailors separate

recoveries for the two classes. Specifically, CPPCA is used to recover the normal-pixel class, while, in the anomaly-pixel recovery, a prediction is made from the normal-pixel class, and CS reconstruction is applied to the resulting residual. The proposed system demonstrated that not only can anomalies be successfully detected with high likelihood in random projections but also projection-domain anomaly detection can drive reconstruction from random projections with significantly improved representation of the anomalous pixels.

An ability to imbue anomalies with increased fidelity is of significant importance in applications such as hyperspectral imagery wherein analysis tasks such as automated target detection and change detection are likely to follow data set reconstruction. Reconstructions that emphasize the preservation of anomalies, such as the PR approach proposed here, are likely to be critical to the widespread adoption of CS and random-projection methodologies in such applications.

REFERENCES

[1] V. J. Hodge and J. Austin, "A survey of outlier detection methodologies," *Artif. Intell. Rev.*, vol. 22, no. 2, pp. 85–126, Oct. 2004.

[2] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.

[3] E. A. Ashton and A. Schaum, "Algorithms for the detection of sub-pixel targets in multispectral imagery," *Photogramm. Eng. Remote Sens.*, vol. 64, no. 7, pp. 723–731, Jul. 1998.

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

[5] A. Bannerjee, P. Burlina, and C. Diehl, "A support vector method for anomaly detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2282–2291, Aug. 2006.

[6] O. Duran and M. Petrou, "A time-efficient method for anomaly detection in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3894–3904, Dec. 2007.

[7] Y. Gu, Y. Liu, and Y. Zhang, "A selective KPCA algorithm based on high-order statistics for anomaly detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 1, pp. 43–47, Jan. 2008.

[8] B. Penna, T. Tillo, E. Magli, and G. Olmo, "Hyperspectral image compression employing a model of anomalous pixels," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 664–668, Oct. 2007.

[9] Q. Du, W. Zhu, and J. E. Fowler, "Anomaly-based JPEG2000 compression of hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 696–700, Oct. 2008.

[10] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.

[11] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

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

[13] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[14] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

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

[16] M. A. Davenport, P. T. Boufounos, M. B. Wakin, and R. G. Baraniuk, "Signal processing with compressive measurements," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 445–460, Apr. 2010.

[17] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. Hoboken, NJ: Wiley, 2005.

[18] K. Krishnamurthy, M. Raginsky, and R. Willett, "Hyperspectral target detection from incoherent projections: Nonequiprobable targets and inhomogeneous SNR," in *Proc. Int. Conf. Image Process.*, Hong Kong, Sep. 2010, pp. 1357–1360.

[19] S. Dasgupta and A. Gupta, An elementary proof of a theorem of Johnson and Lindenstrauss UC Berkeley, Berkeley, CA, Tech. Rep. TR-99-006, Mar. 1999.

[20] D. Achlioptas, " Database-friendly random projections: Johnson-Lindenstrauss with binary coins," *J. Comput. Syst. Sci.*, vol. 66, no. 4, pp. 671–687, Jun. 2003.

[21] J. C. Harsanyi and C.-I Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, Jul. 1994.

[22] N. Keshava and J. F. Mustard, "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44–57, Jan. 2002.

[23] A. Margalit, I. S. Reed, and R. M. Gagliardi, "Adaptive optical target detection using correlated images," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-21, no. 3, pp. 394–405, May 1985.

[24] J. Y. Chen and I. S. Reed, "A detection algorithm for optical targets in clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-23, no. 1, pp. 46–59, Jan. 1987.

[25] S. Ji, D. Dunson, and L. Carin, "Multitask compressive sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 92–106, Jan. 2009.

[26] M. F. Duarte, S. Savrvotham, D. Baron, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing of jointly sparse signals," in *Proc. 39th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Oct. 2005, pp. 1537–1541.

[27] M. Fornasier and H. Rauhut, "Recovery algorithms for vector-valued data with joint sparsity constraints," *SIAM J. Numer. Anal.*, vol. 46, no. 2, pp. 577–613, Mar. 2008.

[28] M. Mishali and Y. C. Eldar, "Reduce and boost: Recovering arbitrary sets of jointly sparse vectors," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 4692–4702, Oct. 2008.

[29] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, Jul. 2007.

[30] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.

[31] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Process.*, vol. 86, no. 3, pp. 589–602, Mar. 2006.

[32] C.-I Chang and D. C. Heinz, "Constrained subpixel target detection for remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 3, pp. 1144–1159, May 2000.

James E. Fowler (S'91–M'96–SM'02) received the B.S. degree in computer and information science engineering and the M.S. and Ph.D. degrees in electrical engineering from Ohio State University, Columbus, in 1990, 1992, and 1996, respectively.

In 1995, he was an Intern Researcher with AT&T Laboratories in Holmdel, NJ, and, in 1997, he held a National Science Foundation-sponsored postdoctoral assignment at the Université de Nice-Sophia Antipolis, Nice, France. In 2004, he was a Visiting Professor in the Département TSI, École Nationale Supérieure des Télécommunications, Paris, France. He is currently a Billie J. Ball Professor and the Graduate Program Director of the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, where he is also a Researcher with the Geosystems Research Institute.

Dr. Fowler is an Associate Editor for IEEE TRANSACTIONS ON IMAGE PROCESSING and the European Association for Signal Processing *Journal of Image and Video Processing*; he formerly served as an Associate Editor for IEEE TRANSACTIONS ON MULTIMEDIA and IEEE Signal Processing Letters. He is the Vice-Chair of the Image, Video, and Multidimensional Signal Processing Technical Committee of the IEEE Signal Processing Society and the Publicity Chair of the program committee for the Data Compression Conference.

Qian Du (S'98–M'00–SM'05) received the Ph.D. degree in electrical engineering from the University of Maryland Baltimore County, Cotonsville, in 2000.

She was with the Department of Electrical Engineering and Computer Science, Texas A&M University, Kingsville, from 2000 to 2004. In fall of 2004, she joined the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, where she is currently an Associate Professor. Her research interests include remote-sensing image analysis, pattern classification, data compression, and neural networks.

Dr. Du is a member of The International Society for Optical Engineers, the American Society for Photogrammetry and Remote Sensing, and the American Society for Engineering Education. She currently serves as the Cochair for the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society. She also served as a Guest Editor for the special issue on Spectral Unmixing of Remotely Sensed Data in the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and a Guest Editor for the special issue on High-Performance Computing in Earth Observation and Remote Sensing in the *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (J-STARS).