

Sensing Matrix Design for Compressive Spectral Imaging via Binary Principal Component Analysis

Jonathan Monsalve, Hoover Rueda-Chacon, *Member, IEEE*, and Henry Arguello, *Senior Member, IEEE*

Abstract—Compressive spectral imaging (CSI) is a framework that captures coded-and-multiplexed low-dimensional projections of spectral data-cubes. In general, the sensing process in many CSI architectures is described using binary matrices, so-called sensing/projection matrices, whose elements can be either random or designed. However, some characteristics of the spectral data, such as the ℓ_2 -norm or the second moment statistics, can be lost when this dimensionality reduction is performed. Similarly, principal component analysis (PCA) is a data dimensionality reduction technique that minimizes the least-squared error between the spectral data and its low-dimensional projection, but preserving its structure or variance. Thus, PCA can be used to guide the CSI acquisition process by designing the binary sensing matrix. Nonetheless, PCA requires to know the spectral image a-priori, and also, its associated projection matrix is not binary, as required by CSI optical architectures. Therefore, in this paper, an algorithm to design CSI sensing matrices by exploiting the structure-preserving property of the PCA projection is proposed. First, a set of compressive measurements obtained with random sensing matrices is used to rapidly estimate the covariance matrix associated with the spectral data. Then, a new sensing matrix is designed by solving a non-convex optimization problem that finds a set of binary vectors that approximate the principal components of the covariance matrix, thus maximizing the explanation of the data variance. Experimental results show an improvement of up to 3 dB in image reconstruction quality, in terms of the peak signal to noise ratio (PSNR), when the binary PCA-based sensing matrices are used and compared with conventional random sensing matrices and state-of-art designed matrices based on PCA.

Index Terms—Compressive spectral imaging, binary principal component analysis, sensing matrix design.

I. INTRODUCTION

COMPRESSIVE spectral imaging (CSI) is a framework to acquire and compress spectral images by means of coded bi-dimensional projections, such that, the number of required measurements for reconstruction are fewer than those needed by traditional techniques based on the Shannon-Nyquist sampling theorem [1], [2], [3], [4]. CSI exploits the fact that natural scenes can be accurately represented in a lower dimensional subspace. This concept is known as sparsity or low rank behavior [3], [5], [6], [7]. Further, the linear projector, so-called sensing matrix, used to capture the compressed

Jonathan Monsalve is with Department of Electrical Engineering, Hoover Rueda-Chacon and Henry Arguello are with the Department of Computer Science, Universidad Industrial de Santander, Bucaramanga, Colombia, 680002 e-mail: henarfu@uis.edu.co

This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org> provided by the author. The material includes proofs and additional figures. MATLAB code can be downloaded from <https://codeocean.com/capsule/8658864/>. Contact henarfu@uis.edu.co for further questions about this work.

version of the spectral image has to be incoherent with the representation basis, where the data becomes sparse. This, in turn, guarantees with high probability an accurate reconstruction, since an incoherent matrix has a dense representation in the basis domain, and so, no assumption on the behavior of the data is required.

Principal component analysis (PCA) is a technique used to reduce the dimensionality of a signal by projecting it into a lower dimension, such that, most of its variance is explained [7], [8], [9]. In particular for spectral images, PCA projects the spectral data using the eigenvectors associated with the m greatest eigenvalues of the covariance matrix Σ , resulting of the signal $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_n] \in \mathbb{R}^{l \times n}$, where l is the number of spectral bands, n the number of spatial pixels, and $\mathbf{f}_i \in \mathbb{R}^l$ is a pixel, for $i = 1, \dots, n$. Thus, a matrix $\mathbf{W}_m \in \mathbb{R}^{l \times m}$, with the m eigenvectors as columns, is used to project the data as $\tilde{\mathbf{F}} = \mathbf{W}_m^T \mathbf{F}$, with $\tilde{\mathbf{F}} \in \mathbb{R}^{m \times n}$ and $m < l$. This formulation has shown to achieve a small error in the Euclidean sense, described by $\|\mathbf{F} - \mathbf{W}_m \tilde{\mathbf{F}}\|$, while preserving the structure of the data in the low-dimensional space, and thus the direction of greatest variability [10].

In a similar way, the noiseless CSI sensing procedure can be expressed as $\mathbf{Y} = \mathbf{Q}^T \mathbf{F}$, where $\mathbf{Q} \in \mathbb{R}^{l \times m}$ is the sensing matrix and \mathbf{F} is the input image [7], [11]. CSI can be categorized as a dimensionality reduction technique, since it projects the spectral signal in a low-dimensional subspace spanned by the rows of the sensing matrix \mathbf{Q} . Note however that, \mathbf{Q} is either randomly generated or designed based on, the restricted isometry property (RIP) or its incoherence with a representation basis [12], [13], [14], [15]. In other words, \mathbf{Q} does not rely, conventionally, on the input signal. Therefore, much effort has been done in the signal processing community to design \mathbf{Q} such that, the structure of the data is preserved in the low-dimensional subspace. Figure 1 shows an example of how PCA better preserves the direction of greatest variability of the data compared to random matrices. In this figure a dataset in \mathbb{R}^3 is projected onto a \mathbb{R}^2 subspace using the eigenvectors associated with the covariance matrix of the data and compared against a conventional random matrix. Note that the PCA projection better preserves the data separability in the \mathbb{R}^2 subspace whereas the random projection mixes them all.

Remark that, if the CSI sensing matrix \mathbf{Q} is equal to the PCA matrix \mathbf{W}_m , the compression or low-dimensional projection, can be considered optimal in the least squares sense [16]. However, PCA is data-dependent, which requires to know the spectral image to be compressed beforehand, thus prohibiting its usage in CSI, where the target data are unknown a priori. Nevertheless, important information about the spectral data

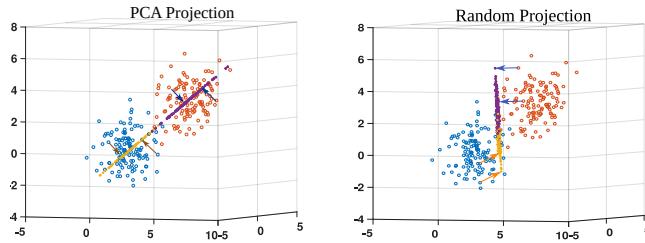


Fig. 1. Example of how PCA preserves the structure of the data. Blue data points are projected to the yellow ones, and orange data points are projected to purple ones. (Left) Data in \mathbb{R}^3 projected onto a subspace in \mathbb{R}^2 using 2 eigenvectors, where the separability and direction of greatest variability of data is preserved. (Right) The same data is projected using a random matrix.

can be extracted from some random compressed measurements \mathbf{Y} , so that, an approximation of the covariance matrix of \mathbf{F} can be attained, and thus exploited to design the subsequent sensing matrices via PCA. Another problem that appears in the design of \mathbf{Q} via PCA is that the entries of \mathbf{Q} are usually binary, as required by CSI optical architectures, while the entries of the principal component matrix \mathbf{W}_m are usually real. Therefore, the goal of this work is to use the PCA intuition to design binary matrices \mathbf{Q} that span a subspace where most of the variance of the signal \mathbf{F} is explained. More precisely, we aim to approximate the behavior of \mathbf{W}_m with a binary matrix \mathbf{Q} to be used as the CSI sensing matrix. For that reason, the traditional optimization problem, to find the eigenvectors, is modified by adding a binary constraint, and a computational algorithm is proposed to solve it efficiently.

A. Related Work

The use of PCA in compressive sensing has been previously studied to improve the signal reconstruction quality. For instance, Masiero *et al.*, in [17], used PCA to choose the best representation basis in the reconstruction process, whereas, Ke *et al.* used PCA to design sensing matrices for low-light-level imaging (L^3 -imaging) [18] and feature-specific imaging [19]. Although in [19] the design is adaptively obtained, the resulting vectors are chosen from a fixed basis like Hadamard. The ideas proposed in our paper are closely related to the work reported in [18], where the authors design binary matrices by solving an optimization problem that directly minimizes the error between the PCA matrix and its binary version using the Frobenius norm. However, [18] focused on data captured by L^3 -imaging and solve the Frobenius-based optimization problem by using the *sign* operator to force the binary restriction. In contrast, this work focuses on the correct acquisition of compressed spectral images through real implementable optical architectures and casts the Frobenius-based optimization problem as a non-convex optimization problem that maximizes the variance explained by the binary principal components. To solve it, a greedy-search-based algorithm is proposed.

B. Contributions

The main contribution of this paper is the development of a methodology to design binary sensing matrices, suitable for CSI architectures, using the structure-preserving PCA properties. The proposed methodology includes 3 steps: first, a set of

conventional random measurements/projections are captured in order to estimate the covariance matrix Σ of the spectral signal, if it is unknown; second, the subsequent sensing matrix is designed by solving a non-convex optimization problem that maximizes the variance explained by the approximated binary principal component matrix; finally, new compressive measurements are acquired with the designed matrix. Subsequently, the underlying data cube is reconstructed using the concatenation of both kind of compressive measurements i.e. those acquired with random and designed matrix, in a single linear system. Theoretical results show that the RIP constant can be considerably small when the sensing matrix is related to the eigenvectors. Additionally, computational results show an improvement in the reconstruction quality of up to 3 dB in terms of PSNR. This paper is organized as follows: Section II poses the non-convex optimization problem to find the binary principal components. Section III shows the theoretical results, specifically the RIP analysis for the proposed sensing matrix. Finally, Section IV shows the simulated experiments, and Section V discusses the findings, before concluding this work in Section VI.

II. SENSING MATRIX DESIGN VIA BINARY PCA

Let $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_n]$ be a matrix representation of the spectral image, where $\mathbf{f}_i \in \mathbb{R}^l$ is the i^{th} spectral pixel of the image with l bands. PCA projects the data into a subspace spanned by its own eigenvectors, such that a pixel \mathbf{f}_i is projected as $\tilde{\mathbf{f}}_i = \mathbf{W}^T \mathbf{f}_i$, where the columns of \mathbf{W} are the eigenvectors associated to the covariance matrix $\Sigma = \mathbf{F} \mathbf{F}^T / n$, assuming that \mathbf{F} is zero mean. The dimensionality reduction is achieved by constructing a matrix $\mathbf{W}_m \in \mathbb{R}^{l \times m}$ that contains only the m eigenvectors associated with the m largest eigenvalues of Σ . Therefore, the projection of the matrix \mathbf{F} can be obtained as

$$\tilde{\mathbf{F}} = \mathbf{W}_m^T \mathbf{F}. \quad (1)$$

PCA plays an important role in data dimensionality reduction and compression. The main advantage of PCA is that it preserves the structure of the data in a lower dimension by minimizing the error given by the signal \mathbf{F} and its orthonormal projection in the low-dimensional subspace $\mathbf{W}_m \mathbf{W}_m^T \mathbf{F}$. However, PCA is data dependent, thus, it requires the encoder to capture and calculate the projection matrix before the compression procedure can be applied. On the other hand, CSI is a framework used to capture and compress data directly in the detector, allowing to accurately recover them by assuming some specific behavior as sparsity, low-rank or eigenvalues eccentricity [1], [3], [20], [11]. Let $\mathbf{Q} \in \mathbb{R}^{l \times m}$ be a sampling matrix with $m < l$, thus, the noiseless sensing problem in CSI can be modeled as

$$\mathbf{Y} = \mathbf{Q}^T \mathbf{F}, \quad (2)$$

where $\mathbf{Y} \in \mathbb{R}^{m \times n}$ are the compressed measurements. Note that CSI does not require any prior knowledge of the data, and data compression is achieved without any calculation performed in the detector side. Note however that, the matrix \mathbf{Q} must be binary since implementable CSI architectures use optical devices that modulate the input source with binary

patterns. A traditional optimization problem to recover the signal \mathbf{F} from \mathbf{Y} is given by

$$\mathbf{F} = \underset{\mathbf{F}}{\operatorname{argmin}} \quad \|\mathbf{Y} - \mathbf{Q}^T \mathbf{F}\|_F^2 + \tau \|\Psi \operatorname{vec}(\mathbf{F})\|_1, \quad (3)$$

where $\|\cdot\|_F$ and $\|\cdot\|_1$ represent the Frobenius and ℓ_1 -norms respectively, Ψ is a sparsity-promoting representation basis, and $\operatorname{vec}(\cdot)$ represents the vectorization of a matrix.

Paraphrasing, the objective of this work is to design the matrix \mathbf{Q} in (2), so that, it approximately behaves as \mathbf{W}_m in (1). The motivation of making $\mathbf{Q} \approx \mathbf{W}_m$ is to preserve the ℓ_2 -norm of the data in the low dimensional space, which translates to a better RIP constant in the compressive sensing sense and thus to a better quality of the image reconstruction (as will be discussed in Section III-A). The matrix \mathbf{W}_m is conventionally found by solving the optimization problem [10]

$$\begin{aligned} \mathbf{W}_m = \underset{\mathbf{W}_m}{\operatorname{argmin}} \quad & \|\mathbf{F} - \mathbf{W}_m \mathbf{W}_m^T \mathbf{F}\|_F^2 \\ \text{subject to} \quad & \mathbf{W}_m^T \mathbf{W}_m = \mathbf{I}, \end{aligned} \quad (4)$$

where $\mathbf{I} \in \mathbb{R}^{m \times m}$ is an identity matrix. This problem can be easily solved with an alternate gradient algorithm, however, a binary restriction must be added to meet the requirements of the CSI architectures. To find the binary matrix, we first solve the intermediate problem, obtained by adding the restriction $Q_{k,j} \in \{0, 1/\sqrt{b_j}\}$ to (4), that limits the entries of the columns of the matrix. This leads to the problem

$$\begin{aligned} \mathbf{Q}_n = \underset{\mathbf{Q}_{n,b_j}}{\operatorname{argmin}} \quad & \|\mathbf{F} - \mathbf{Q}_n \mathbf{Q}_n^T \mathbf{F}\|_F^2 \\ \text{subject to} \quad & \mathbf{Q}_n^T \mathbf{Q}_n = \mathbf{I}, \quad Q_{k,j} \in \{0, 1/\sqrt{b_j}\}, \end{aligned} \quad (5)$$

where \mathbf{Q}_n is the designed matrix, b_j is the number of non-zero entries in its j^{th} column, for $k = 1, 2, \dots, l$ and $j = 1, 2, \dots, \tilde{m}$, with \tilde{m} the number of binary vectors. Note that, although the solution of (5) is not binary, in the sense that the entries of \mathbf{Q}_n can take more than two values, it is indeed a binary matrix with normalized columns. The binary matrix can be obtained as $\mathbf{Q} = [\sqrt{b_1} \mathbf{q}_1, \dots, \sqrt{b_{\tilde{m}}} \mathbf{q}_{\tilde{m}}]$, since \mathbf{Q} spans the same subspace of \mathbf{Q}_n . Solving (5) is hard due to the non-convexity entailed by the binary restriction. Therefore, we propose to solve a maximization-based problem that allows to directly design each binary vector, without relying on threshold operators. First, note that the problem in (4) is equivalent to [21], [22]

$$\begin{aligned} \mathbf{w}_j = \underset{\mathbf{w}_j}{\operatorname{argmax}} \quad & \mathbf{w}_j^T \Sigma \mathbf{w}_j \\ \text{subject to} \quad & \|\mathbf{w}_j\|_2^2 = 1, \end{aligned} \quad (6)$$

where $\mathbf{w}_j \in \mathbb{R}^l$ is the j^{th} column of the matrix \mathbf{W}_m . Problem (6) estimates a single eigenvector at a time, and then deflates the covariance matrix with the expression $\Sigma_j = \Sigma_{j-1} - \mathbf{w}_j \mathbf{w}_j^T \Sigma_{j-1} \mathbf{w}_j^T \mathbf{w}_j$, to remove the influence of the already estimated eigenvector. This is the procedure used in the power iteration method. Then, by adding the binary constraint, problem (6) becomes

$$\begin{aligned} \mathbf{q}_j = \underset{\mathbf{q}_{j,b_j}}{\operatorname{argmax}} \quad & \mathbf{q}_j^T \Sigma \mathbf{q}_j \\ \text{subject to} \quad & q_j^k \in \{0, 1/\sqrt{b_j}\}, \end{aligned} \quad (7)$$

where q_j^k is the k^{th} entry of the j^{th} column of \mathbf{Q}_n . The problem in (7) aims to estimate the subspace spanned by the binary vectors \mathbf{q}_j , that maximize the variance of the data concentrated in Σ . Therefore, \mathbf{q}_j is regarded as a binary principal component. Note that the covariance matrix Σ is unknown, since it depends on the data. Thus, a set of random projections must be acquired at first in order to estimate it. For this, there exist multiple algorithms in the literature that can be used [7], [8], [11]. This work employs the CPPCA approach, proposed in [7], due to its speed and reliability in estimating the covariance matrix. This approach will be addressed in detail in Section III-B.

Additionally, note that (7) is non-convex because the convex combination of two binary vectors does not necessarily result in a binary vector, i.e. $\lambda \mathbf{q}_j + (1 - \lambda) \mathbf{q}_{j'}$ is not necessarily binary. Thus, Algorithm 1 has been developed to approximate the solution of (7). This algorithm is greedy-search-based since it iteratively looks for the best position within the vector that maximizes the objective function. In spite of its greedy characteristics, the proposed algorithm exhibits good performance, as it will be shown in the simulations, and furthermore, it is parameter-free.

In detail, Algorithm 1 requires an initial estimation of the covariance matrix Σ_1 and the number of binary vectors \tilde{m} to be designed. Then, it initializes the coding pattern \mathbf{q}_j with a vector of zeros (Line 3). In each iteration, it looks for the best position where a binary value of $1/\sqrt{b}$ maximizes the objective function, but keeping fixed the positions previously found. That is, in each iteration, a previously selected position is never evaluated or tested again. The for-loop in Line 4 only iterates over (l/\tilde{m}) to promote the resulting vectors to be orthonormal to each other, which implies that the transmittance of the resulting binary vector is forced to be less or equal than $1/\tilde{m}$. Note that, the transmittance is defined as the ratio between the number of nonzero elements and the total number of elements.

Algorithm 1 B-PCA: Binary PCA estimation

```

1: input:  $\Sigma_1, \tilde{m}$ 
2: for  $j = 1$  to  $\tilde{m}$  do
3:    $\mathbf{q}_j \leftarrow \mathbf{0}; max \leftarrow 0; list = 1, 2, \dots, l$ 
4:   for  $k = 1$  to  $\operatorname{round}(l/\tilde{m})$  do impose transmit.  $1/\tilde{m}$ 
5:     for  $i = \text{each element in } list$  do
6:        $\mathbf{q}_j^i \leftarrow 1$  place a one in the  $i^{th}$  position
7:        $c^k \leftarrow \frac{\mathbf{q}_j^T}{\|\mathbf{q}_j\|} \Sigma_j \frac{\mathbf{q}_j}{\|\mathbf{q}_j\|}$  objective function (7)
8:       if  $c^k > max$  then
9:          $max = c^k; index = i$ 
10:      end if
11:       $\mathbf{q}_j^i \leftarrow 0$  remove the one value
12:    end for
13:     $\mathbf{q}_j^{index} \leftarrow 1$  place the one in the best position
14:     $list.remove(index); index \leftarrow 0$ 
15:  end for
16:   $\mathbf{P} \leftarrow \frac{\mathbf{q}_j \mathbf{q}_j^T}{\|\mathbf{q}_j\|^2}$ 
17:   $\Sigma_{j+1} \leftarrow \Sigma_j - \mathbf{P} - \mathbf{P} \Sigma_j + \mathbf{P} \Sigma_j \mathbf{P}$ 
18: end for
19: output:  $\mathbf{Q} = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_{\tilde{m}}]$ 

```

In Line 6, a value of one is placed in a certain position and the objective function is calculated (Line 7) and compared with its previous value. If the current position maximizes the function, it is considered as a candidate (Line 8 to 10). To test other positions, the one is removed in Line 11. At the end of the loop, the best position is stored in the variable *index*, such that a value of one is fixed there (Line 13). After that, the covariance matrix Σ is deflated by subtracting the influence of the vector \mathbf{q}_j (Line 16 and 17) and the for-loop continues. The proof that expression in Line 17 subtracts the influence of the subspace spanned by \mathbf{q} is shown in Appendix A, in the supplementary material. Algorithm 1 has computational complexity $O(l^3)$ and its proof is deferred to Appendix B, in the supplementary material. Note that, although the complexity of the algorithm is dominated by the cubic term, spectral images usually span along few hundreds of bands, thus it does not considerably increase the computational complexity of the problem. Additionally, an analysis of the convergence of the algorithm is presented in Appendix C, in the supplementary material, explaining why the restriction $q_j^k \in \{0, 1/\sqrt{b_j}\}$ is needed. Furthermore, a flowchart that describes the Algorithm 1 is shown in Fig. S1, in the supplementary material, and a Matlab code implementation of this algorithm can be found and tested at <https://codeocean.com/capsule/8658864/>.

III. THEORETICAL RESULTS

A. Restricted Isometry Property Analysis

This section shows that the use of PCA satisfies the RIP in specific cases. It is important to show that the RIP holds when the sensing matrix $\mathbf{Q} = \mathbf{W}_m$, and the signal can be accurately represented in the subspace given by $\text{span}(\mathbf{W}_m)$, linking the use of PCA to the CSI sensing procedure. Additionally, it sets a relationship between the variance and the expected value of the pixel norm linking PCA and the RIP.

Theorem 1: Let $\mathbf{f} = \mathbf{W}\boldsymbol{\theta} \in \mathbb{R}^l$ be a pixel of a spectral image and $\mathbf{W} \in \mathbb{R}^{l \times l}$ be the matrix whose columns are the eigenvectors of the covariance matrix $\Sigma = \mathbf{W}\Lambda\mathbf{W}^T$. It holds that

$$(1 - \delta_m)\|\boldsymbol{\theta}_P\|_2^2 \leq \|\mathbf{A}_P\boldsymbol{\theta}_P\|_2^2 \leq (1 + \delta_m)\|\boldsymbol{\theta}_P\|_2^2, \quad (8)$$

with

$$\delta_m = |P_2|\lambda_{m+1}/(\sum_{i \in P_1} \lambda_i + \sum_{i \in P_2} \lambda_i), \quad (9)$$

and $P_1 \subset \{1, 2, \dots, m\}$, $P_2 \subset \{m+1, \dots, l\}$, \mathbf{A}_P is a sub-matrix whose columns are a subset of the columns of $\mathbf{A} = \mathbf{W}_m^T \mathbf{W}$, $P = P_1 \cup P_2$, and λ_i is the i^{th} eigenvalue of the covariance matrix of the data. Note that, in most natural scenes, pixels can be accurately represented in a low dimensional space, hence, most of the information is kept in the first eigenvalues which implies that $0 < \delta_m \ll 1$.

The proof of this theorem is deferred to Appendix D, in the supplementary material.

Theorem 1 shows that, when the eigenvectors are used as linear projectors, the norm of a vector changes in proportion to the least important eigenvalues. Further, note that binary versions of the eigenvectors are used, given by \mathbf{Q} and thus,

(8) does not hold since $\mathbf{Q}^T \mathbf{W} \neq \mathbf{I}$. The RIP for \mathbf{Q} is stated in the Corollary 1.1

Corollary 1.1: For an arbitrary sensing matrix \mathbf{A} , the RIP constant is bounded by $\delta_{\tilde{m}} = (\sum_{k \in P_1} \lambda_k |\beta_k| + \sum_{k' \in P_2} \lambda_{k'} |\beta_{k'}|)/(\sum_{k \in P_1} \lambda_k + \sum_{k' \in P_2} \lambda_{k'})$ with $|\beta_k| = |1 - \sum_{i=1}^{\tilde{m}} (a_k^i)^2| = |1 - \sum_{i=1}^{\tilde{m}} (\mathbf{q}_k^T \mathbf{w}_i)^2|$ and $|\beta_{k'}| = |1 - \sum_{i=1}^{\tilde{m}} (a_{k'}^i)^2| = |1 - \sum_{i=1}^{\tilde{m}} (\mathbf{q}_{k'}^T \mathbf{w}_i)^2|$ for $k = 1, \dots, \tilde{m}$ and $k = \tilde{m} + 1, \dots, l$

The proof of the corollary is detailed in Appendix E, in the supplementary material.

Corollary 1.1 shows that if the columns of the sensing matrix are unit-norm, the RIP constant is small. However, for the case of the eigenvectors, it is enough to ensure that this holds for the columns related to the largest eigenvalues. Thus, in order to guarantee that $\delta_{\tilde{m}}$ is small, $|\beta| = |1 - \sum_{i=1}^{\tilde{m}} (a_k^i)^2| = 0$ should hold, i.e. $\sum_{i=1}^{\tilde{m}} (a_k^i)^2 = \sum_{i=1}^{\tilde{m}} (\mathbf{q}_k^T \mathbf{w}_i)^2 = 1$. Intuitively, it can be seen that if $\mathbf{q}_k \approx \mathbf{w}_k$, $a_k^i \approx 1$ and

$$\sum_{i=1}^{\tilde{m}} (a_k^i)^2 \approx \sum_{i=1}^{k-1} (a_k^i)^2 + 1 + \sum_{i=k+1}^{\tilde{m}} (a_k^i)^2. \quad (10)$$

Therefore, if \mathbf{q}_i is orthogonal to \mathbf{q}_k , for $i \neq k$, we get that

$$\sum_{i=1}^{\tilde{m}} (a_k^i)^2 \rightarrow 1, \text{ and thus, } \delta_{\tilde{m}} \rightarrow 0. \quad (11)$$

B. Covariance Matrix Estimation

For the covariance matrix estimation, the CPPCA approach was adopted [7]. For that, define

$$\mathbf{F}_i = [\mathbf{f}_{\Omega_i^1}, \dots, \mathbf{f}_{\Omega_i^{n/p}}], \Omega_i^j \neq \Omega_k^l, i \neq k, \forall i, k, \quad (12)$$

as a subset of the hyperspectral image \mathbf{F} introduced in (1), where $\Omega_i \subset \Omega = \{1, \dots, n\}$, and Ω_i^j refers to the j^{th} element in the subset Ω_i with $i = 1, \dots, p$ such that $p \ll n$. Additionally, not only one projection matrix $\mathbf{Q} \in \mathbb{R}^{l \times v}$ is used, but p different projection matrices $\mathbf{Q}_i \in \mathbb{R}^{l \times v}$, which allow to rewrite the problem in (2) as

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{H}}\tilde{\mathbf{F}}, \quad (13)$$

where $\tilde{\mathbf{Y}} = [\mathbf{Y}_1^T, \dots, \mathbf{Y}_p^T]^T$, $\mathbf{Y}_i \in \mathbb{R}^{v \times n/p}$, $\tilde{\mathbf{H}} = \text{diag}(\mathbf{Q}_1^T, \dots, \mathbf{Q}_p^T)$, $\tilde{\mathbf{F}} = [\mathbf{F}_1^T, \dots, \mathbf{F}_p^T]$ and v is the number of random acquisitions or randomly generated rows. This p -partition sensing approach is borrowed from [23] following the sensing strategy of HYCA (hyperspectral coded aperture).

This sensing strategy can be implemented in CSI optical architectures, such as, the 3D-CASSI [2] and the DD-CASSI [24]. Using this approach, the covariance matrix can be rapidly estimated from a set of random projections using a projection-onto-convex-sets based algorithm (POCS). The reader is encouraged to check [7] for more details on POCS. Additionally, the data must be centered, but since the spectral images take values between 0 and $2^{\text{bits}} - 1$, the zero mean assumption is unrealistic. Reconstructing the image in order to estimate its mean will result in high computational costs, therefore, the mean should be calculated directly from the compressive

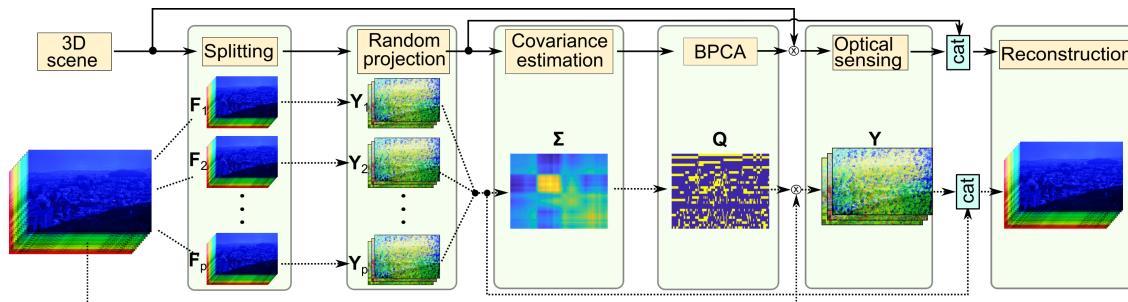


Fig. 2. Flowchart for the sensing and reconstruction algorithm. Solid lines show a block diagram of the procedure and dotted lines show graphically the same procedure. First, the image is divided into p subsets and projected using random matrices. The covariance matrix is estimated using these random projections and it is used to design the binary sensing matrix.

measurements using the randomly generated sensing matrices as [8]

$$\hat{\mathbf{f}} = \alpha \sum_{i=1}^p \sum_{j=1}^k \mathbf{Q}_i^T (\mathbf{Q}_i \mathbf{Q}_i^T)^{-1} \mathbf{y}_i^j, \quad (14)$$

where $\alpha = m/n$, \mathbf{y}_i^j is the j -th pixel in the i -th partition or subset, and k is the number of pixels in each partition p , such that $pk = n$. It was proved that Eq. (14) converges to the true mean when $n \rightarrow \infty$ [8]. Once the mean is estimated, the measurements are centered by subtracting it as $\tilde{\mathbf{Y}}_i = \mathbf{Y}_i - \mathbf{Q}_i^T (\hat{\mathbf{f}} \otimes \mathbf{1}^T)$, where \otimes represents the Kronecker product, and $\mathbf{1} \in \mathbb{R}^k$ is a k -long one-valued vector. This operation replicates the mean and subtracts it from the compressive measurements. Performing this operation only once usually does not produce very accurate results, however, estimating the mean and subtracting it multiple times in a for-loop usually produces better results [8].

C. Sensing and Reconstruction Methodology

In summary, the proposed sensing and reconstruction methodology is detailed in Algorithm 2 and sketched in Fig. 2. First, p randomly generated matrices $\{\mathbf{Q}_i\}_{i=1}^p \in \mathbb{R}^{l \times v}$, following a Bernoulli distribution, are used as the initial sensing matrices. Using the random compressed measurements $\{\mathbf{Y}_i\}_{i=1}^p$ of each disjoint subset $\{\mathbf{F}_i\}_{i=1}^p$, the covariance matrix is estimated (Lines 6-7) and the subsequent matrix is designed following Algorithm 1 (Line 8). Afterwards, the designed matrix $\mathbf{Q} \in \mathbb{R}^{l \times \tilde{m}}$ is concatenated with the initial random matrices \mathbf{Q}_i , as stated in Lines 9 to 12, and the sensing process is repeated using the resulting matrix in Line 13. The concatenation of random and designed matrices is performed in order to improve the condition of the problem. This is, the rank of the designed matrix is at most \tilde{m} and the rank of the concatenation is at most $v + \tilde{m}$. However, if the covariance matrix is known a-priori, the random measurements are not required. Finally, the optimization problem to recover the image is solved in Line 14. The latter can be done using algorithms like GPSR [25], SALSA [26], or SpaRSA [27].

IV. SIMULATIONS AND RESULTS

In this section, two hyperspectral images are used as input to demonstrate the effectiveness of the designed binary PCA matrices. The hyperspectral scenes are the Urban dataset [28]

Algorithm 2 Proposed sensing and reconstruction protocol

```

1: input:  $\{\mathbf{F}_i\}_{i=1}^p, \tilde{m}$ 
2: for  $i = 1$  to  $p$  do
3:    $\mathbf{Q}_i \leftarrow \text{randbinary}(l, v)$ 
4:    $\mathbf{Y}_i \leftarrow \mathbf{Q}_i^T \mathbf{F}_i$ 
5: end for
6:  $\{\tilde{\mathbf{F}}_i\}_1^p = \text{CPPCA}(\{\mathbf{Q}_i\}_1^p, \{\mathbf{Y}_i\}_1^p)$   $\triangleright$  Image estimation
7:  $\tilde{\Sigma} = \frac{1}{n} \sum_{i=1}^p \tilde{\mathbf{F}}_i \tilde{\mathbf{F}}_i^T$   $\triangleright$  Estimate covariance matrix
8:  $\mathbf{Q} \leftarrow \text{B-PCA}(\tilde{\Sigma}, \tilde{m})$   $\triangleright$  Find binary principal components
9: for  $i = 1$  to  $p$  do
10:    $\bar{\mathbf{Q}}_i \leftarrow [\mathbf{Q}_i^T, \mathbf{Q}^T]^T$   $\triangleright$  Concatenate sensing matrices
11:    $\mathbf{Y}_i \leftarrow [\mathbf{Y}_i^T, (\mathbf{Q}^T \mathbf{F}_i)^T]^T$   $\triangleright$  Update measurements
12: end for
13:  $\tilde{\mathbf{H}} \leftarrow \text{diag}(\bar{\mathbf{Q}}_1^T, \dots, \bar{\mathbf{Q}}_p^T)$ 
14:  $\tilde{\mathbf{F}} \leftarrow \text{argmin}_{\tilde{\mathbf{F}}} \|\tilde{\mathbf{Y}} - \tilde{\mathbf{H}} \tilde{\mathbf{F}}\|_F^2 + \tau \|\Psi \text{vec}(\tilde{\mathbf{F}})\|_1$ 
15: Output:  $\tilde{\mathbf{F}}$ 

```

with 256×256 pixels of spatial resolution and $l = 128$ spectral bands, and a section of the Pavia centre recorded by the ROSIS sensor [29], with 512×512 pixels of spatial resolution and $l = 102$ spectral bands. A single spectral band and three pixels of each dataset are shown in Figs. 3 and 4, respectively.

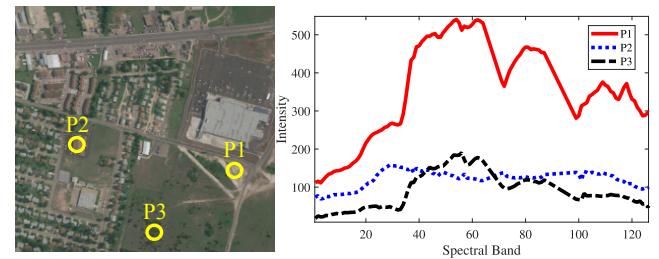


Fig. 3. Urban dataset. (Left) RGB composite of Urban dataset. (Right) Three spectral signatures at different pixels of the image.

The random sensing matrices \mathbf{Q}_i are generated following a Bernoulli distribution. The reconstruction of the full datacube from the set of random+designed measurements, Line 14 in Algorithm 2, is performed using the Split Augmented Lagrangian Shrinkage Algorithm (SALSA) [26], using a signal sparsity prior over the 3D Kronecker basis formed by the Wavelet 2D Symlet 8 and the Discrete Cosine transform (DCT). The relative variation of the objective function is used as the stopping criterion, and it is set

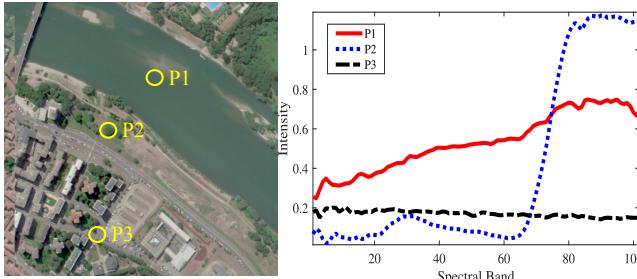


Fig. 4. Pavia centre dataset. (Left) RGB composite of Pavia dataset. (Right) Three spectral signatures at different pixels of the image.

to $1e-5$. The number of partitions is set to $p = l$, for simplicity. The performance is measured in terms of the spectral peak signal to noise ratio (PSNR), defined as $\text{PSNR} = 1/n \sum_{i=1}^n (10 \log A^2 / ||\mathbf{f}_i - \tilde{\mathbf{f}}_i||_2^2)$, where A is the maximum amplitude value of a pixel, \mathbf{f}_i is the i^{th} pixel of the reference image and $\tilde{\mathbf{f}}_i$ is the reconstructed pixel, the mean squared error (MSE), defined as $1/n \sum_{i=1}^n ||\mathbf{f}_i - \tilde{\mathbf{f}}_i||_2^2$, and the spectral angle mapper (SAM) defined as $\arccos(\mathbf{f}^T \tilde{\mathbf{f}} / (||\mathbf{f}|| ||\tilde{\mathbf{f}}||)) (180/\pi)$.

A. Mean estimation

As hyperspectral images do not have zero mean, it is first estimated, following Eq. (14), and then subtracted from the measurements. For Urban and Pavia, we test the quality of this estimation by setting $p = l$, and varying the number of random projections from 6 to 18. The normalized mean squared error, $NMSE = MSE / ||\mathbf{f}||$, is used to measure the quality of the estimated means. Table I shows the overall results in terms of the NMSE and SAM, where it can be seen that the estimated mean does not differ much from the true mean as the number of projections increases.

TABLE I
NMSE OF THE ESTIMATED MEAN VARYING THE NUMBER OF SHOTS

Image	Metric	Shots			
		6	10	14	18
Urban	NMSE	0.182	0.116	0.074	0.046
	SAM	10.402	6.470	4.172	2.574
Pavia	NMSE	0.171	0.095	0.053	0.031
	SAM	9.3690	5.1879	3.0096	1.6407

B. Quality of the PCA-based Designed Matrices

First, the quality of the designed sensing matrices is evaluated by testing only the ℓ_2 approximation in (3), setting $\tau = 0$. Note that the reconstruction using only the ℓ_2 -term can be done in closed-form via the Moore-Penrose pseudo-inverse. The results attained with the designed matrices are compared against the ones with randomly generated matrices, the ones generated with the algorithm proposed in [18](Section 3, which will be termed QPCA) and the ground-truth data. In this test, the results attained with the designed matrix but calculated from the ground-truth covariance matrix are also included, in order to check the induced error when it is designed directly from the random projections. Table II shows the overall results in terms of MSE, PSNR and SAM, and Fig.

TABLE II
OVERALL PERFORMANCE OF THE SENSING MATRICES. THE COLUMN “RANDOM” REPRESENTS THE RANDOMLY GENERATED MATRICES, “ESTIMATED Σ ” REPRESENTS THE DESIGNED MATRICES USING THE ESTIMATED COVARIANCE MATRIX AND “A-PRIORI Σ ” THE DESIGNED MATRICES BUT USING THE TRUE COVARIANCE MATRIX.

Image	Metric	Random		Estimated Σ		A-priori Σ	
		Prop.	QPCA	Prop.	QPCA	Prop.	QPCA
Urban	MSE	292.03	106.1	240.6	105.3	230.7	
	PSNR	10.31	19.27	11.51	20.31	12.82	
	SAM	35.86	4.692	27.50	4.59	21.64	
Pavia	MSE	324.7	147.9	212.1	135.2	186.3	
	PSNR	10.12	16.30	11.74	16.47	13.38	
	SAM	37.34	6.23	29.99	5.82	14.44	

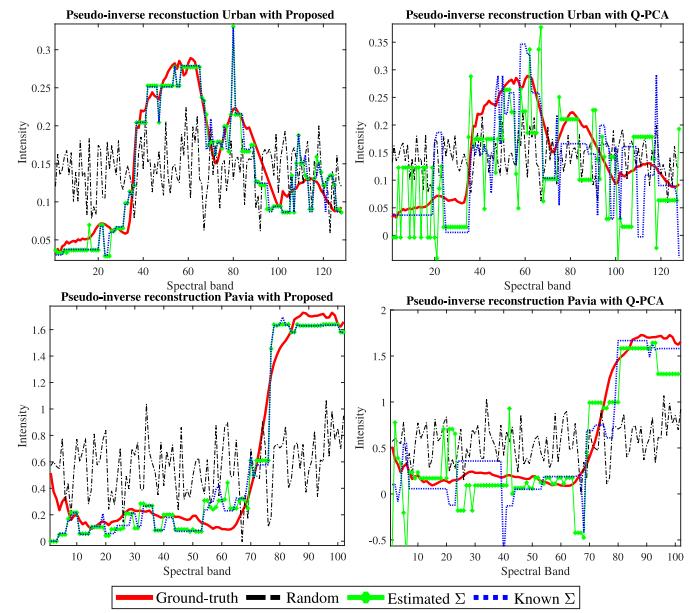


Fig. 5. Comparison of a reconstructed pixel solving only the ℓ_2 term in (3) via the Moore-Penrose pseudo-inverse, using 8 random measurements (black line), 8 designed measurements with prior Σ (blue lines), and designed matrix with estimated Σ (green lines). (Top-Left) Results for the Urban dataset using the proposed matrix. (Top-Right) Results for the Urban dataset using QPCA. (Bottom-left) Results for Pavia dataset using the proposed matrix. (Bottom-right) Results for Pavia dataset using QPCA.

5 depicts the behavior of the reconstruction at a specific pixel of the datasets. It can be noticed that the proposed sensing matrices achieve a better performance in terms of the three metrics, disregarding how Σ is estimated. In terms of MSE, the proposed matrices achieve a 3-fold improvement compared with random measurements, and around 2-fold in terms of PSNR. Remark that the results with the estimated Σ closely approximate the ones from the true Σ , confirming the good quality of the statistics estimated from the random projections.

Additionally, the optimization problem proposed in (7) was compared with the standard PCA problem given in (6). For that, the variance explained by varying the number of vectors is analyzed. The variance explained or retained is defined as

$$d = \text{trace}(\mathbf{Q}^T \Sigma (\mathbf{Q}^\dagger)^T) / \text{trace}(\Sigma). \quad (15)$$

Figure 6 was generated by designing 35 vectors using the proposed optimization problem and the method proposed in [18], and then calculating the explained variance using (15) when a subset of vectors is used. Figure 6 shows that the

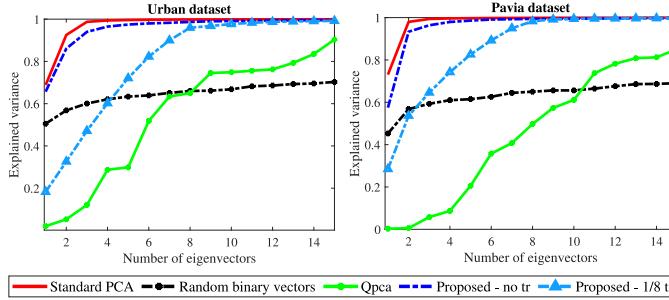


Fig. 6. Explained variance by varying the number of eigenvectors with different methods. Red - solid line represents the variance for traditional PCA (theoretical limit). Blue dot-dashed line is used for the proposed binary eigenvectors when no transmittance restriction is imposed, dashed line with triangles represents the proposed binary eigenvectors when the transmittance is set to 12.5% (1/8), green-solid line with marks represents Qpca, whereas Black-dashed line is for a random matrix.

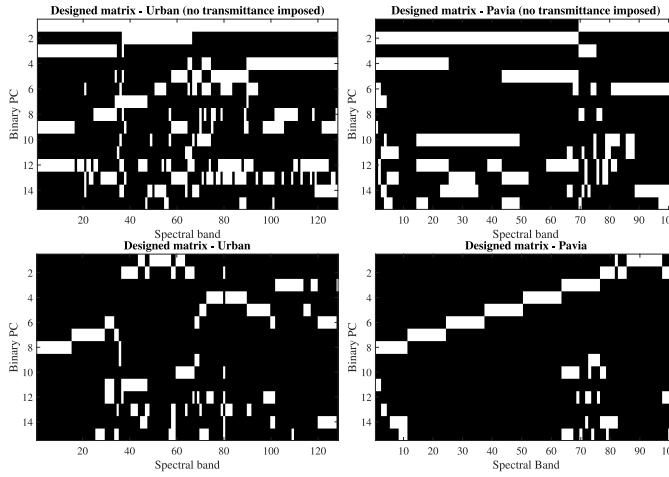


Fig. 7. Realization of the designed matrices for Urban and Pavia datasets with 15 vectors. (Top-row) Designed matrices when no transmittance is imposed. (Bottom-row) Designed matrices with 1/8 of transmittance.

proposed binary eigenvectors explain better the variance of the data in comparison with Qpca and random vectors when no restriction in the transmittance is imposed. For instance, for 98% of variance, standard PCA requires only 3 eigenvectors, the proposed binary PCA requires 5 to 7 depending on the image, and Qpca is not able to reach this amount of explained variance. However, as the theoretical results suggest, restricting the transmittance improves the RIP, thus the sensing matrices for the reconstruction experiments are calculated with the transmittance restriction imposed. Although the explained variance reduces, it is still larger than Qpca and random. As an example, Fig. 7 shows a realization of the binary matrix obtained by our method, when no transmittance is imposed and when it is imposed, for both datasets.

C. Reconstruction Performance under Noisy Scenarios

In this experiment, the performance of the designed matrices is evaluated by reconstructing the datasets using the $\ell_2 - \ell_1$ optimization problem in (3). For this, the number of measurements is varied from 18 to 36, which roughly represents a 14% to 28% compression ratio for the Urban dataset, and a 17% to

TABLE III
RELATIONSHIP BETWEEN THE NUMBER OF RANDOM AND DESIGNED VECTORS USED IN THE SENSING PROCEDURE.

Image	Kind	Shots					
		18	21	24	27	30	33
Urban	Random	10	13	15	18	21	24
	Designed	8	8	9	9	9	9
Pavia	Random	13	15	18	20	23	26
	Designed	5	6	6	7	7	8

35% compression ratio for the Pavia dataset. Intuitively, it will be ideal to use as many designed measurements as possible, however, since the covariance matrix is estimated from the compressed measurements, some of these vectors should be generated at random. For instance, when 18 measurements are being captured, one could use 9 random and 9 designed, or 10 random and 8 designed, or any other combination. However, the number of designed binary vectors \tilde{m} in this work is calculated by computing the variance explained by the binary eigenvectors as in (15). Thus, by fixing this percentage to be 98%, different numbers of random and designed vectors were used for the reconstructions, as shown in Table III.

Remark that the different proportions are due to the fact that, at some point, using more designed measurements do not contribute significantly in terms of data variance, but, as shown in Table I, since the covariance matrix is estimated from the random measurements, increasing its number improves the estimation of Σ .

Reconstruction results using the PSNR as the performance metric, for the two datasets, are summarized in Fig. 8. In this figure, two noisy scenarios are tested, with signal-to-noise-ratios (SNR) of 15 and 25 dB. Overall, the designed matrices (solid lines) outperform the random (dotted lines) by up to 3 dB in the two noisy scenarios, and up to 2 dB in comparison with Qpca (dot-dashed lines) with high compression ratios. Additionally, note that, the entries of the matrices produced by [18] are $\{-1, 1\}$, which entail problems when being implemented, such as the requirement of an additional all-pass shot in order to produce the -1 code word. Furthermore, since this type of coding cannot be implemented directly, its emulation increases the noise by a factor of 5 [30]. Under this scenario the reconstructions obtained by the Qpca in Figs. 8 to 10 would be dramatically affected.

Additionally, to visually compare the attained reconstructions, Figs. 9 and 10 show the reconstructions at two randomly selected spatial pixels, denoted as P1 and P2, and a specific spectral band when 21 measurements are captured (16% and 20% of the data for Urban and Pavia datasets, respectively). There, it can be noticed that the spectral signatures attained with the designed matrices closely resemble the ground truth, and the difference of the spectral bands show finer details in the reconstructions attained with the designed matrices.

V. DISCUSSION

We present a brief analysis of the differences in the explained variance of the proposed algorithm in comparison to Qpca. Additionally, it is discussed why the proposed binary

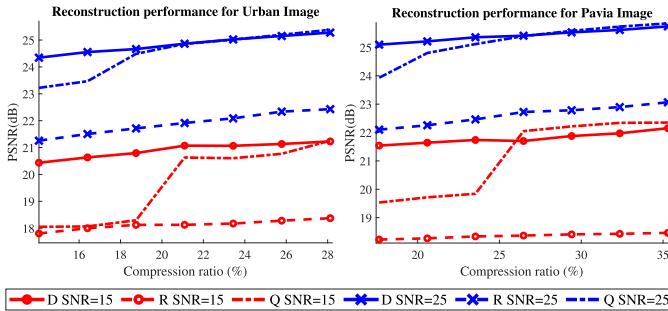


Fig. 8. Average PSNR of the reconstructed hyperspectral datasets for various compression ratios, at 2 noise scenarios with SNR = 15 and 25 dB. “D” stands for designed (solid lines), “Q” for those proposed in [18] (QPCA) (dot-dashed lines) and “R” for random (dotted lines) matrices. Different colors represent different noise levels. (Left) Urban dataset. (Right) Pavia dataset

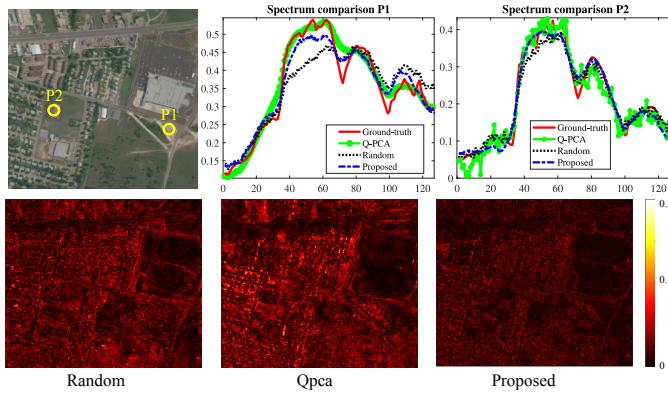


Fig. 9. Comparison between some pixels (P1, P2) of the Urban dataset using 21 measurements (20.58% of the data, 13 random and 8 designed). (Top-left) RGB composite of the ground-truth. (Top-middle) Comparison between reconstructed and ground-truth for P1. (Top-right) Comparison between reconstructed and ground-truth for P2. (Bottom-left) Normalized residual of the reconstruction using random matrices. (Bottom-middle) Normalized residual using QPCA. (Bottom-right) Normalized residual using the proposed matrices.

eigenvectors exhibit a close behavior to those obtained with standard PCA in terms of explained variance.

A. Explained Variance Comparison

The proposed algorithm performs very well in terms of explained variance in contrast to QPCA. This behavior is expected since the QPCA algorithm minimizes the objective function given by

$$\begin{aligned} \arg \min_{\mathbf{D}} \quad & \|\sqrt{N} \mathbf{D} \mathbf{W}_m^T - \text{sign}(\mathbf{D} \mathbf{W}_m^T)\|_F^2 \\ \text{subject to} \quad & \mathbf{D} \mathbf{D}^T = \mathbf{I}, \end{aligned} \quad (16)$$

which does not take into account the eigenvalues. Considering the eigenvalues is critical in hyperspectral imaging, since usually the first eigenvectors are much more important than the last ones. In fact, the proposed algorithm uses the covariance matrix rather than just the eigenvectors as in [18], which implicitly takes into account the eigenvalues. To show that, we present additional results in Fig. 11, when the matrix \mathbf{W}_m is scaled by a diagonal matrix \mathbf{B} such that, the first eigenvector is scaled by a larger number than the last eigenvector, i.e. $\mathbf{W}_m = \mathbf{W}_m \mathbf{B}$. Note however that by doing this, the QPCA algorithm

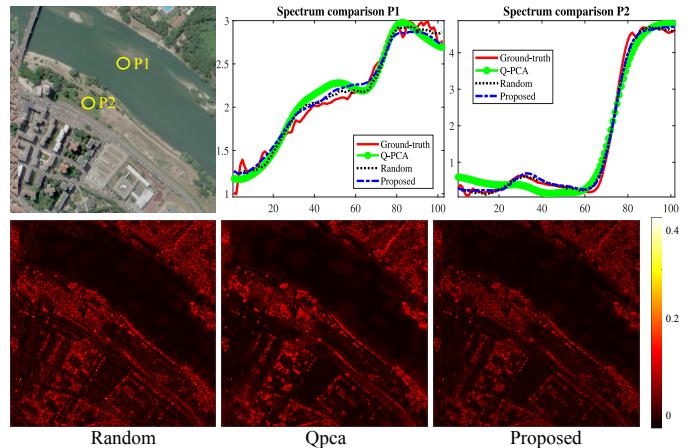


Fig. 10. Comparison between some pixels (p1, P2) of the Pavia dataset using 21 measurements (16.41% of the data, 15 random and 6 designed). (Top-left) RGB composite of the ground-truth. (Top-middle) Comparison between reconstructed and ground-truth for P1. (Top-right) Comparison between reconstructed and ground-truth for P2. (Bottom-left) Normalized residual of the reconstruction using random matrices. (Bottom-middle) Normalized residual using QPCA. (Bottom-right) Normalized residual using the proposed matrices.

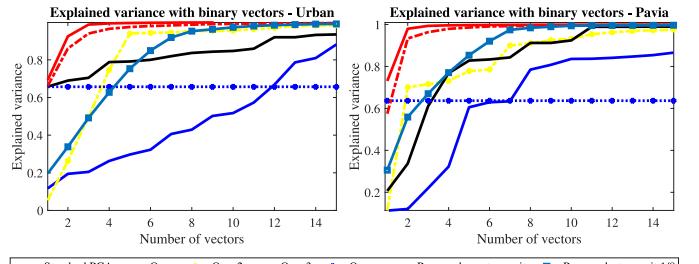


Fig. 11. Explained variance with different realizations of QPCA. QPCA represents the explained variance when the pure eigenvectors are used; QPCA2 is obtained by scaling the eigenvectors with a diagonal matrix whose values are shown in Fig. 12(b); QPCA3 is obtained by scaling the eigenvectors with a diagonal matrix whose values are shown in Fig. 12(c); QPCAC is obtained by scaling the eigenvectors with the eigenvalues, which are shown in Fig. 12(d)

should be adjusted since it assumes that $\mathbf{W}_m^T \mathbf{W}_m = \mathbf{I}$, and it does not hold for $\mathbf{W}_m = \mathbf{W}_m \mathbf{B}$. Nevertheless, it can be seen that by scaling the eigenvectors according to their eigenvalues, the explained variance of the resulting QPCA binary matrix improves, at the cost of rank deficiency as shown in Fig. 12.

In Figs. 11 and 12, it can be seen that by taking into account the importance of each eigenvector (determined by its associated eigenvalue) the QPCA algorithm converges to a solution where the variance is better explained. However, its convergence is not guaranteed since the rank of the resulting matrix degenerates as the matrix \mathbf{B} resembles the eigenvalues. Note that, when the actual eigenvalues are used, the resulting binary matrix is rank 1, which negatively affects the conditioning of the problem. In contrast, in Figs. 11 and 7 it can be seen that the proposed method better explains the variance while the rank of the designed matrix does not degenerate.

B. Binary Matrix vs. Discrete Matrix

On the other hand, note that during the designing procedure, the sensing matrix is not binary at all, but discrete. However,

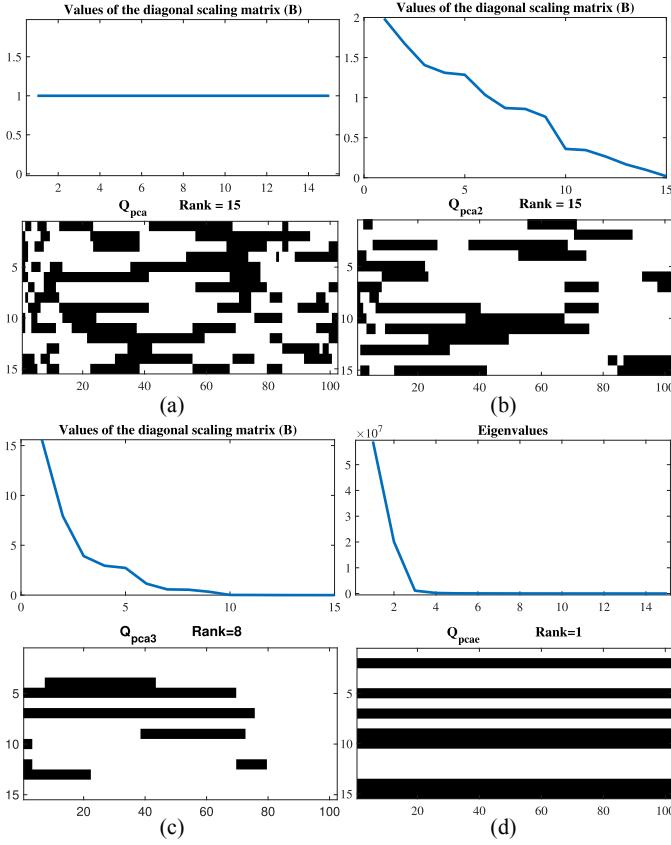


Fig. 12. Values used in the diagonal matrix \mathbf{B} to scale the eigenvectors of the Qpca algorithm, along with the resulting binary matrix. (a) Not scaling. (b) Smooth scaling. (c) Steep scaling. (d) Eigenvalue scaling.

once the algorithm finishes, the matrix is scaled to become binary. Hence, it should be highlighted that this scaling does not affect the resulting matrix, since an eigenvector defines a direction of projection, and so, any scaled version of an eigenvector is also an eigenvector. To see that, let's use the eigendecomposition of the covariance matrix

$$\Sigma = \mathbf{W} \Lambda \mathbf{W}^{-1}. \quad (17)$$

Hence, if \mathbf{W} is scaled by a constant b , it holds that,

$$\Sigma = b\mathbf{W}\Lambda b^{-1}\mathbf{W}^{-1}, \quad (18)$$

is still an eigenvector. We expect the latter holds also for binary eigenvectors. To further support this affirmation, Fig. S2 shows that the explained variance is exactly the same when the binary or their discrete (normalized-column) version are used.

Furthermore, it can be seen in Fig. S2 of the supplementary material, that the explained variance obtained with the proposed binary eigenvectors is quite close to the standard PCA. This behavior is mainly due to hyperspectral images exhibit a low-rank behavior [7]. For a signal with not such behavior, the explained variance using binary eigenvectors is not that close to the standard, as shown and discussed in Fig. S3 of the supplementary material, where a random signal is compared against the Urban dataset.

VI. CONCLUSIONS

This paper introduced the design of binary sensing matrices, commonly used in real CSI architectures, via binary PCA. The designed matrices exploit the structure-preserving properties of PCA, where most of the data variance is explained by a set of binary vectors, estimated directly from some compressed random measurements. The performance of the designed matrices was evaluated over two real hyperspectral datasets, Urban and Pavia center, achieving an overall 3 dB improvement in the reconstruction quality compared with conventional random sensing matrices and up to 2 dB compared with state-of-art sensing matrices based on PCA. Additionally, an analysis of the RIP for the proposed method was introduced, which provided a bound for the RIP when the standard PCA technique is performed and when the binary PCA is used. The proposed algorithm to design the matrices is greedy-search-based with low computational complexity, and the results show that it is able to retain variance of the data in a better fashion than state-of-art methods.

ACKNOWLEDGMENTS

The authors would like to thank the Vicerrectoría de Investigación y Extensión of the Universidad Industrial de Santander (VIE-UIS) for supporting this work registered under the project “Sistema óptico-computacional para mejorar la resolución espacial de imágenes hiperespectrales usando técnicas de fusión a partir de datos obtenidos con esquemas de adquisición basados en muestreo compresivo” under VIE code 2436. Jonathan Monsalve is supported by a Colciencias/Department of Santander scholarship. Hoover Rueda-Chacon was supported by the postdoctoral internship program of the VIE-UIS.

REFERENCES

- [1] G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle, “Compressive Coded Aperture Spectral Imaging: An Introduction,” *IEEE Signal Processing Magazine*, vol. 31, pp. 105–115, jan 2014.
- [2] X. Cao, T. Yue, X. Lin, S. Lin, X. Yuan, Q. Dai, L. Carin, and D. J. Brady, “Computational Snapshot Multispectral Cameras: Toward dynamic capture of the spectral world,” *IEEE Signal Processing Magazine*, vol. 33, pp. 95–108, sep 2016.
- [3] E. Candes and M. Wakin, “An Introduction To Compressive Sampling,” *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.
- [4] E. J. Candes and T. Tao, “Decoding by linear programming,” *Information Theory, IEEE Transactions on*, vol. 51, no. 12, pp. 4203–4215, 2005.
- [5] H. Arguello and G. R. Arce, “Restricted Isometry Property in coded aperture compressive spectral imaging,” in *2012 IEEE Statistical Signal Processing Workshop (SSP)*, pp. 716–719, IEEE, IEEE, aug 2012.
- [6] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [7] J. Fowler, “Compressive-Projection Principal Component Analysis,” *IEEE Transactions on Image Process*, vol. 18, pp. 2230–2242, oct 2009.
- [8] H. Qi and S. M. Hughes, “Invariance of principal components under low-dimensional random projection of the data,” *Proceedings - International Conference on Image Processing, ICIP*, pp. 937–940, 2012.
- [9] Y. Zhang, A. D’Aspremont, and L. El Ghaoui, “Sparse PCA: Convex relaxations, algorithms and applications,” *International Series in Operations Research and Management Science*, vol. 166, pp. 915–940, 2012.
- [10] N. Kwak, “Principal Component Analysis Based on L1-Norm Maximization,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 1672–1680, sep 2008.
- [11] J. M. Bioucas-Dias, D. Cohen, and Y. C. Eldar, “Covalsa: Covariance estimation from compressive measurements using alternating minimization,” in *2014 22nd European Signal Processing Conference (EUSIPCO)*, pp. 999–1003, Sept 2014.

- [12] C. V. Correa, H. Arguello, and G. R. Arce, "Spatiotemporal blue noise coded aperture design for multi-shot compressive spectral imaging," *J. Opt. Soc. Am. A*, vol. 33, pp. 2312–2322, Dec 2016.
- [13] H. Rueda, H. Arguello, and G. R. Arce, "Compressive spectral testbed imaging system based on thin-film color-patterned filter arrays," *Appl. Opt.*, vol. 55, pp. 9584–9593, Nov 2016.
- [14] C. V. Correa, C. A. A. Hinojosa, G. R. Arce, and H. Arguello, "Multiple snapshot colored compressive spectral imager," *Optical Engineering*, vol. 56, no. 4, p. 10, 2016.
- [15] X. Lin, Y. Liu, J. Wu, and Q. Dai, "Spatial-spectral encoded compressive hyperspectral imaging," *ACM Trans. Graph.*, vol. 33, no. 6, pp. 233:1–233:11, 2014.
- [16] J. M. Bioucas-Dias and J. M. Nascimento, "Hyperspectral subspace identification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2435–2445, 2008.
- [17] G. Quer, R. Masiero, G. Pillonetto, M. Rossi, and M. Zorzi, "Sensing, compression, and recovery for WSNs: Sparse signal modeling and monitoring framework," *IEEE Transactions on Wireless Communications*, vol. 11, no. 10, pp. 3447–3461, 2012.
- [18] J. Ke and E. Y. Lam, "Fast compressive measurements acquisition using optimized binary sensing matrices for low-light-level imaging," *Optics Express*, vol. 24, no. 9, p. 9869, 2016.
- [19] J. Ke, A. Ashok, and M. A. Neifeld, "Object reconstruction from adaptive compressive measurements in feature-specific imaging," *Applied optics*, vol. 49, no. 34, pp. H27–H39, 2010.
- [20] W. Li and J. E. Fowler, "Decoder-side dimensionality determination for compressive-projection principal component analysis of hyperspectral data," in *2011 18th IEEE International Conference on Image Processing*, no. 2, pp. 321–324, IEEE, sep 2011.
- [21] B. Jiang, S. Ma, and S. Zhang, "Tensor principal component analysis via convex optimization," *Mathematical Programming*, vol. 150, pp. 423–457, may 2015.
- [22] M. Journée, Y. Nesterov, P. Richtárik, and R. Sepulchre, "Generalized power method for sparse principal component analysis," *Neural Computation*, vol. 21, pp. 3179–3213, nov 2008.
- [23] G. Martín, J. M. Bioucas-Dias, and A. Plaza, "HYCA: A New Technique for Hyperspectral Compressive Sensing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, pp. 2819–2831, may 2015.
- [24] M. E. Gehm, R. John, D. J. Brady, R. M. Willett, and T. J. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Optics Express*, vol. 15, no. 21, pp. 14013–14027, 2007.
- [25] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics Signal Process*, vol. 1, no. 4, pp. 586–597, 2007.
- [26] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Fast Image Recovery Using Variable Splitting and Constrained Optimization," *IEEE Transactions on Image Processing*, vol. 19, pp. 2345–2356, sep 2010.
- [27] S. Wright, R. Nowak, and M. Figueiredo, "Sparse Reconstruction by Separable Approximation," *IEEE Transactions on Signal Processing*, vol. 57, pp. 2479–2493, jul 2009.
- [28] F. Zhu, Y. Wang, S. Xiang, B. Fan, and C. Pan, "Structured sparse method for hyperspectral unmixing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 88, pp. 101–118, 2014.
- [29] A. A. Mueller, A. Hausold, and P. Strobl, "Hysens-dais/rosis imaging spectrometers at dlr," in *SPIE 4545, Remote Sensing for Environmental Monitoring, GIS Applications, and Geology*, vol. 4545, p. 11, Jan 2002.
- [30] S. E. Pinilla, H. M. Vargas García, and H. Arguello Fuentes, "Probability of correct reconstruction in compressive spectral imaging," *Ingeniería e Investigación*, vol. 36, no. 2, pp. 68–77, 2016.



Hoover Rueda-Chacon (M'12) received the B. Sc. and M. Sc. degrees in Computer Science from Universidad Industrial de Santander, Colombia, in 2009 and 2012, respectively, and the M. Sc. and Ph. D. degrees in Electrical and Computer Engineering from the University of Delaware, Newark, DE, USA, in 2015 and 2017, respectively, sponsored by a Fulbright-Colciencias scholarship. His main research areas are computational imaging, compressive sensing, high dimensional signal processing and optimization algorithms.



Henry Arguello (SM'17) received the M.Sc. in Electrical Engineering from the Universidad Industrial de Santander, Colombia, and the Ph.D. degree in Electrical and Computer Engineering from University of Delaware, Newark, DE, USA. He is currently a Titular Professor with the School of Systems Engineering at Universidad Industrial de Santander. He was named a Senior Member of IEEE in 2017. Dr. Arguello's research interests include statistical signal processing, high dimensional signal coding and processing, optical and computational imaging, optical code design, compressed sensing, hyperspectral imaging and numerical optimization using stochastic algorithms.



Jonathan Monsalve Received the B. Sc. and M. Sc. degrees in Computer Science from Universidad Industrial de Santander, Colombia, in 2015 and 2018, respectively. He is currently pursuing the Ph.D. in Electrical Engineering at Universidad Industrial de Santander, sponsored by a Colciencias scholarship. His main research areas are computational imaging, compressive sensing, optical code design and subspace learning.