On the Exclusion of Hyperspectral Sources

Zihan Zhang, *Student Member, IEEE,* and Thierry Blu, *Fellow, IEEE*

***Abstract*—This paper introduces a blind source separation approach based on a "source exclusion" principle for hyperspectral image unmixing (HSU). We define the exclusion mathematically as a metric quantifying the global purity of the pixels. We then develop an efficient algorithm to minimize this criterion (Weak Exclusion Principle—WEP), and devise a convex optimization strategy (WEP+) to enforce sum-to-one and non-negativity, which are common constraints for hyperspectral sources. Through comprehensive experimental validations against standard and state-of-the-art unmixing algorithms on synthetic and real-world datasets, we demonstrate the superior accuracy and computational efficiency of our WEP+ solution.**

***Index Terms*—Hyperspectral Image Unmixing, Blind Source Separation, Exclusion Principle, Hyperspectral Data Visualization.**

I. Introduction

A. Hyperspectral unmixing and blind source separation

HYPERSPECTRAL images (HSIs) capture hundreds of spectral bands across visible and infrared regions. While this provides high spectral resolution, the accompanying lower spatial resolution often results in mixed pixels containing multiple materials. Hyperspectral unmixing (HSU) thus becomes essential for material separation, as it is about decomposing each mixed pixel into its constituent pure materials (endmembers) and their fractional abundances. This process transforms three-dimensional hyperspectral data into spectral signatures and corresponding two-dimensional abundance maps. As a source separation technique, HSU facilitates various image processing tasks such as denoising, classification, and segmentation [2]. The performance of HSU algorithms, particularly in terms of accuracy and speed, thus plays a crucial role in remote sensing applications like cartographic mapping, agricultural planning, and mineral exploration.

Blind source separation (BSS) is a technique that identifies underlying sources and their mixing relationships from observed signals. This technique has applications across medical imaging, remote sensing, communications, etc. As an application of BSS, hyperspectral unmixing decomposes hyperspectral data from a $K \times L$ matrix $\mathbf{X}$ (where K represents pixels and L spectral channels) into a linear combination of M sources. In this context, abundances serve as the sources, while endmembers represent the mixing relationships. Specifically, the task involves determining the $K \times M$ source matrix $\mathbf{S}$ (abundance map) and the $M \times L$ mixing matrix $\mathbf{A}$ (endmembers). The noisy linear mixing can be written as

$$\mathbf{X} = \mathbf{S}\mathbf{A} + \mathbf{N}$$

$$\text{visually,} \quad \underset{K \times L}{\mathbf{X}} = \underset{K \times M}{\mathbf{S}} \times \underset{M \times L}{\mathbf{A}} + \mathbf{N} \tag{1}$$

The components of the matrices involved are denoted by lower-case letters

$$\mathbf{X} = [x_{k,l}]_{\substack{1 \le k \le K \\ 1 \le l \le L}}, \mathbf{S} = [s_{k,m}]_{\substack{1 \le k \le K \\ 1 \le m \le M}}, \text{ and } \mathbf{A} = [a_{m,l}]_{\substack{1 \le m \le M \\ 1 \le l \le L}},$$

where $\mathbf{N}$ represents noise, which is typically neglected in standard HSU algorithms and also not addressed in this work. Please note that, similar to [3], [4], [5], [6], [7], [8], we use a transposed formulation of the usual mixing equation in our setting. It is typical to assume that, M, the number of sources of the problem, is known. Hence, like most hyperspectral unmixing algorithms, we perform first a dimension reduction of the data matrix $\mathbf{X}$ by approximating it as a rank-M matrix (see Section II-B for details).

BSS is an ill-posed problem; yet, this ill-posedness can be narrowed down to the degrees of freedom of the mixing matrix, $\mathbf{A}$, because once it is identified, determining the source matrix $\mathbf{S}$ becomes unambiguous in general (e.g., $\mathbf{S} = \mathbf{X}\mathbf{A}^+$ using the pseudo-inverse of $\mathbf{A}$). Since these degrees of freedom are much fewer than the degrees of freedom of the source matrix ($ML \ll KM$: the number of bands is much smaller than the number of pixels), the key to effective HSU is finding an accurate mixing matrix.

B. Non-negativity and sum-to-one assumptions

Hyperspectral source images are often required to satisfy two constitutive constraints [9]: the abundance non-negativity constraint (ANC) which ensures nonnegative proportions, and the abundance sum-to-one constraint (ASC) which requires that proportions in each pixel sum to unity. In our mathematical setting:

$$\mathbf{S} \ge \mathbf{0}_K, \text{ and } \mathbf{S}\mathbf{1}_M = \mathbf{1}_K, \tag{2}$$

where $\mathbf{0}_D/\mathbf{1}_D$ are D-dimensional column vectors of zeros/ones (see the "Notations" subsection at the end of the introduction). An elegant geometric interpretation is that, seen as a collection of M-dimensional points, a noiseless source image is contained inside a simplex.

A very preliminary version of this work was presented at ICASSP 2022 [1].

Zihan Zhang is with the Department of Electronic Engineering, The Chinese University of Hong Kong, Hong Kong, Shatin, NT, Hong Kong SAR, China (e-mail: zhangzihan@link.cuhk.edu.hk).

Thierry Blu is with the Department of Electronic Engineering, The Chinese University of Hong Kong and the Department of Electrical Engineering, National Taiwan University, No.1, Sec. 4, Roosevelt Rd., Taipei 106319, Taiwan, R.O.C. (e-mail: thierry.blu@m4x.org).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org., provided by the author. The material includes experimental results. Contact thierry.blu@m4x.org for further questions about this work.

These constraints have drawn quite some interest since the 1970s with the work of Horwitz for multispectral data proportion estimation [10], [11]. The Constrained-Least-Squares (CLS) method, introduced by Shimabukuro and Smith in 1991 [9], marked the first classic method considering these two constraints within the multispectral images. Then, Heinz et al. in 1999 proposed the Fully Constrained Least Squares (FCLS) method [12], the first to solve two constraints for hyperspectral images. It is also called the Fully Constrained Linear Spectral Unmixing (FCLS) framework. Subsequently, the ANC and ASC became fundamental hypotheses in nearly all HSU algorithms. Building upon the FCLS framework, subsequent works were developed including the Fully Sets (FS) method [13], the Endmember Growing FCLS (EG-FCLS) method [14].

Unfortunately, the ANC and ASC constraints alone are not sufficient to make the blind HSU well-posed in general (e.g., adding a constant matrix to $\mathbf{S}$ and renormalizing the rows to $\mathbf{1}_K$ provides another valid solution).

C. Hyperspectral unmixing approaches

Based on [15], it is possible to identify four main types of methods for solving the blind HSU problem: geometrical, statistical, sparse regression and spatial-spectral contextual information based approaches. The geometrical-based approaches formulate the hyperspectral data as a simplex, which includes Pure Pixel (PP) based and Minimum Volume (MV) based. The statistical methods formulate the spectral unmixing as a statistical inference problem, such as the dependent component analysis (DECA) [16]. Sparse regression-based methods assume most coefficients of the source matrix are zero, a property that is exploited by the $\ell_{1/2}$ regularizer NMF ($\ell_{1/2}$-NMF) [17]. Spatial-spectral contextual information-based approaches focus on the inherent spatial variation of endmember spectra within the hyperspectral image. They extend the basic linear mixing model to account for endmember variations, such as the perturbed linear mixing model (PLMM) [18]. More recent approaches to the unmixing problem have used Autoencoder (AE) neural networks, such as the spectral variability-aware cascaded autoencoder (SVACA)[19], the joint multiscale graph attention and classify-driven autoencoder (MSGA-CD)[20], and autoencoders with temperature scaling [21], [22]. In the experiments of this paper, we have found that MiSiCNet, DFFN and SWC-Net [23], [24], [25] were performing best among the available autoencoder-based HSU algorithms. Unfortunately, the absence of sufficiently large hyperspectral image datasets makes a deep learning strategy currently unreliable.

Here, we are particularly interested in the pure-pixel approach because of its intuitiveness (hence, interpretability) and computational efficiency (no GPU needed), despite its relative lack of modernity. It is worth pointing out that the pure-pixel hypothesis is interwoven with the ANC/ASC constraints, which explains the apparent separation between HSU and BSS research. Thanks to the ANC and ASC constraints, the unknown endmembers can then be retrieved as the vertices of the simplex that bounds the hyperspectral data. Identifying pure pixels has led to several methods. In 1995, Boardman introduced the Pixel Purity Index (PPI) method [26], finding pure spectral signatures by searching vertices of a convex hull. In 1999, Winter proposed the N-FINDR method [27], which inflates a simplex inside the data, identifying the vertices of the largest simplex volume as pure pixels. Subsequently, in 2005, Nascimento and Dias developed the now standard Vertex Component Analysis (VCA) [28] by iteratively projecting the data onto a direction orthogonal to the subspace spanned by the endmembers already determined and the new endmember signature corresponds to the extreme of the projection. Further algorithmic contributions to the pure pixel assumption are fast iterative PPI (FIPPI) [29], p-norm-based pure pixel identification (TRI-P) [30], and random PPI (RPPI) [31]. However, it was soon realized that the pure pixel requirement is inaccurate for most data and it became reasonable to explore the no-pure-pixel situation, such as minimum volume enclosing simplex (MVES) [32], minimum volume constrained nonnegative matrix factorization (MVC-NMF) [33], simplex identification via split augmented Lagrangian (SISAL) [34], minimum volume simplex analysis (MVSA) [35], and hyperplane-based CSI (HyperCSI) [36]. Despite their ability to deal with non-pure pixels, minimum volume methods assume sources that, either have a few points that are exactly on the boundaries of the simplex (geometric approaches), or are statistically uniform inside the simplex (maximum likelihood approaches) to be optimal, relying on points near the edges and vertices for simplex identification [37]. If there are no vertices or edge pixels, their accuracy drops significantly: most of the other pixels do not contribute accurately to the estimation of the endmembers.

Pure-pixel methods face two key challenges: First, they rely on the existence of a few truly pure pixels in the data. Although minimum volume methods attempt to bypass this requirement, they perform significantly worse with heavily mixed data. Second, pixel purity itself lacks a precise definition: despite an attempt at quantifying pixel purity in [38], we are not aware of any use of it as an optimization criterion in later publications, as it seems to be limited to a theoretical analysis tool [39], [40], [41], [42]. Sparsity, a possible alternative purity metric, does not control the locations of the zeros, unfortunately, making it an inaccurate optimization criterion as soon as the hyperspectral pixels are not exactly pure [43]. What we propose in this paper is precisely a global quantification of purity and how to exploit it in an optimization setting.

D. Exclusion, a measure of purity

Intuitively, we expect that hyperspectral images can be segmented into homogeneous regions; i.e., into separate regions where a single material is dominant, or "excludes" the other materials. Following our preliminary work [1], we quantify this idea by defining a measure of this dominance, the "exclusion", and we posit that optimizing this criterion results in optimal source separation: this is what we call the Weak Exclusion Principle (WEP). The exclusion can be understood as a form of local control over sparsity. It regulates the positions of non-zero pixels and avoids the null solution

that arises from over-penalization in sparse methods [43]. In contrast with the pixel purity measures defined in [38] which consider only the pixels that are within the narrowest ring that contains the boundary of the data, the exclusion involves all the pixels in a way that is likely to be more robust to inaccuracies in real settings.

The exclusion takes values in $[0, 1[$, expressed as percentages (exact exclusion $= 0\%$, good[1] exclusion values $\leq 20\%$). At 0%, every hyperspectral pixel is pure, which hints that our criterion is a global measure of purity—importantly, without nonnegativity and sum-to-one requirements. Of course, real data are not pure, and this is where quantifying purity is useful. Specifically, hyperspectral images are characterized by an exclusion of less than 12% (see Table I) which suggests that most of their pixels are close to pure. And, indeed, an accurate and robust recovery of a mixing matrix can be achieved by minimizing the exclusion of the resulting sources as we demonstrated in [1]. Notably, being unencumbered by the ANC and ASC constraints, the WEP optimization performs a factorization that is unlike nonnegative matrix factorization, and for this reason applies to broader blind source separation problems. In order to apply the exclusion principle to hyperspectral images, though, we will need to transform the WEP solution into an ANC+ASC-compliant solution (WEP+ algorithm).

As we will see in the experiment part (Section IV), the WEP+ algorithm is fast, accurate, and robust across many scenarios (exclusion/noise levels, numbers of sources, image sizes, real or synthetic data). It can also be used as initialization for other algorithms (e.g., segmentation) or as a preprocessing tool for neural network training.

E. Contributions of this paper

- We provide a quantitative definition of a pixel purity criterion—the "exclusion"—to address blind source separation problems, with additional insight compared to our earlier conference paper [1].
- We provide an efficient implementation of the minimization of exclusion (generic WEP solution). Although this algorithm was already described in [1], we provide new results and new comparisons with other algorithms that confirm the reliability of the WEP algorithm, despite not using the ANC/ASC constraints.
- We develop a novel algorithm to enforce the ANC and ASC constraints (WEP+ solution), from our WEP unmixing result—basically, from the segmentation map alone. Using convex optimization, we first recover the endmembers; then, we determine the optimal abundances using nonnegative least squares, in a faster implementation than that provided by FCLS.
- We demonstrate by comparisons with state-of-the-art algorithms (classical and recent neural network-based solutions) the superior accuracy and computational efficiency of the WEP+ solution over a comprehensive set of simulated (exclusion, noise level, number of sources, etc.) and real hyperspectral images.

[1]Empirical tests on random source matrices show that data with an exclusion smaller than 20% can be separated accurately (see Section 1 of the supplementary materials).

Notations

- K, number of pixels of the hyperspectral image;
- L, number of bands of the hyperspectral image;
- M, number of sources;
- $\mathbf{e}_m$, the M-dimensional unit vector whose m-th component is equal to 1 (hence all the others are zero);
- $\mathbf{0}_D$, the D-dimensional vector of zeros;
- $\mathbf{1}_D$, the D-dimensional vector of ones;
- $\mathbf{u}^{\mathrm{T}}$, $\mathbf{P}^{\mathrm{T}}$, the transpose of the vector $\mathbf{u}$, of the matrix $\mathbf{P}$;
- $\mathrm{Tr}(\mathbf{P})$, the trace of the (square) matrix $\mathbf{P}$ (i.e., the sum of its diagonal elements);
- $\|\mathbf{P}\|_{\mathrm{F}} = \sqrt{\mathrm{Tr}(\mathbf{P}\mathbf{P}^{\mathrm{T}})}$, the Frobenius (Euclidean) norm of the matrix $\mathbf{P}$;
- $\mathbf{P}^{+}$, the pseudo-inverse (Moore-Penrose) of the matrix $\mathbf{P}$;
- $\|\mathbf{v}\| = \sqrt{\mathbf{v}^{\mathrm{T}}\mathbf{v}}$, the Euclidean norm of the vector $\mathbf{v}$;
- $\widetilde{\mathbf{P}} = \left[\frac{\mathbf{P}\mathbf{e}_1}{\|\mathbf{P}\mathbf{e}_1\|} \quad \frac{\mathbf{P}\mathbf{e}_2}{\|\mathbf{P}\mathbf{e}_2\|} \quad \frac{\mathbf{P}\mathbf{e}_3}{\|\mathbf{P}\mathbf{e}_3\|} \quad \cdots\right]$, the column normalized form of the matrix $\mathbf{P}$.

II. Blind Source Separation Based on Source Exclusion

A. The Exclusion Principle

A source matrix $\mathbf{S}$ is exclusive whenever its rows contain only one non-zero element (see Fig. 1); i.e., for every row index $k \in \{1, \ldots, K\}$

$$s_{k,m} \neq 0 \Rightarrow s_{k,m'} = 0, \forall m' \neq m. \tag{3}$$

Equivalently, for any diagonal matrix $\mathbf{W}$, then $\mathbf{S}^{\mathrm{T}}\mathbf{W}\mathbf{S}$ is a diagonal matrix as well. To quantify how much a matrix is exclusive, it is useful to introduce the "exclusive" form $\mathring{\mathbf{S}}$ of the matrix $\mathbf{S}$ which is defined as follows

$$\mathring{s}_{k,m} \stackrel{\text{def}}{=} \begin{cases} \tilde{s}_{k,m}, & \left| \begin{array}{l} \text{if } \ |\tilde{s}_{k,m}| > |\tilde{s}_{k,m'}|, m' < m \\ \text{and } |\tilde{s}_{k,m}| \geq |\tilde{s}_{k,m'}|, m' > m \end{array} \right. \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where $\tilde{s}_{k,m}$ are the coefficients of the column-normalized form of $\mathbf{S}$. Note that this normalization just reflects the fact that, even in ideal cases, blind source separation provides a unique solution only up to a scaling of the sources (and a permutation). With this notation, we define the "exclusion" of a $K \times M$ matrix $\mathbf{S}$ by the non-negative scalar number

$$\mathscr{X}\{\mathbf{S}\} \stackrel{\text{def}}{=} 1 - \frac{\|\mathring{\mathbf{S}}\|_{\mathrm{F}}^2}{\mathrm{M}} = \frac{\|\mathring{\mathbf{S}} - \tilde{\mathbf{S}}\|_{\mathrm{F}}^2}{\|\tilde{\mathbf{S}}\|_{\mathrm{F}}^{\mathbf{2}}} = \frac{1}{\mathrm{SNR}(\mathring{\mathbf{S}}, \tilde{\mathbf{S}})}, \tag{5}$$

Obviously, an exactly exclusive matrix has zero exclusion (since, in that case, $\tilde{\mathbf{S}} = \mathring{\mathbf{S}}$); conversely, a matrix that has zero exclusion is exactly exclusive (since $\mathring{\mathbf{S}}$ is exclusive).

Typical exclusion values range between 0 and $1 - 1/M$, so we choose to express them as percentages. In practice, values that are $\lesssim 20\%$ indicate that the sources are sufficiently exclusive to warrant their robust recovery by minimizing the exclusion, as we will see later. Notably, the reference sources of standard real hyperspectral image datasets are well below this threshold, as shown in Table I.

TABLE I: Exclusion values of sources in some public datasets

Data Name	Samson	Jasper Ridge	Urban	Apex	WDC	Moffett
$\mathcal{E}\{\mathbf{S}\}$ (%)	6.53	9.58	11.57	7.33	11.34	7.87
Average	9.04					

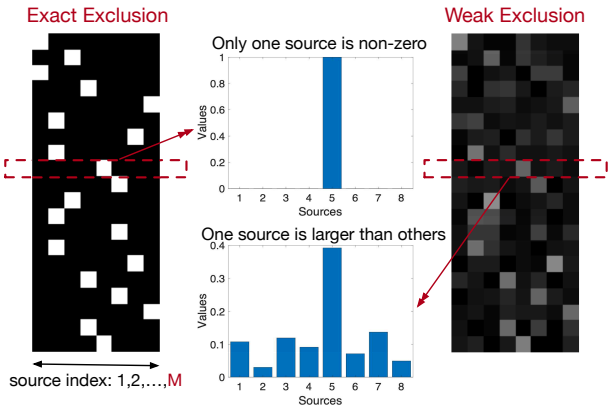


Fig. 1: Representation of exclusion and weak exclusion in the source matrix. In the source matrix, exact exclusion occurs when only one source has a non-zero value within a pixel. Weak exclusion occurs when this is not the case: how "weak" is quantified by the exclusion (5).

It is possible to express the exclusion in a form that is more optimization-friendly: consider the diagonal "source selection" matrices $\{\mathbf{W}_m\}_{m=1,2,\ldots,M}$ whose diagonal elements are obtained from the columns of the exclusive form of $\mathbf{S}$ according to

$$w_m[k,k] = \begin{cases} 1, & \text{if } \mathring{s}_{k,m} \neq 0; \\ 0, & \text{otherwise .} \end{cases} \tag{6}$$

for $k = 1, 2, \ldots, K$. Then, the exclusion takes the following quadratic form

$$\mathscr{X}\{\mathbf{S}\} = 1 - \frac{1}{M}\sum_{m=1}^{M} \|\mathbf{W}_m \tilde{\mathbf{S}} \mathbf{e}_m\|^2 \tag{7}$$

B. WEP: Optimizing the Exclusion

1) Pre-processing: We approximate the raw mixed source matrix $\mathbf{X}$ with a rank-M matrix $\mathbf{X}_{\text{pre}}\mathbf{A}_{\text{pre}}$, where $\mathbf{X}_{\text{pre}}$ is a full rank $K \times M$ orthogonal matrix, and $\mathbf{A}_{\text{pre}} = \mathbf{X}_{\text{pre}}^{\text{T}}\mathbf{X}$, an $M \times M$ matrix. There are two ways to find such an approximation, depending on whether we desire the unconstrained best rank-M matrix (Euclidean/Frobenius norm), or the best rank-M matrix that contains the vector of ones $\mathbf{1}_K$ in its span.

The first one consists in performing the SVD of $\mathbf{X}$ and keeping only its M most significant principal components in $\mathbf{X}_{\text{pre}}$. The second one consists in performing the SVD of the zero-mean matrix $\mathbf{X} - \mathbf{1}_K\mathbf{1}_K^{\text{T}}\mathbf{X}$ and stacking its $M-1$ most significant principal components with $\mathbf{1}_K/\sqrt{K}$ in $\mathbf{X}_{\text{pre}}$. Of course, the reason for considering this second option is that it ensures that the source matrices built from $\mathbf{X}_{\text{SVD}_1}$ are able to satisfy the ASC exactly.

Both rank-M approximations will be used as input to the algorithm that we are going to describe now.

2) WEP algorithm: Looking for the most exclusive solution of the hyperspectral unmixing problem can be formulated through the minimization:

$$\min_{\mathbf{B}\in\mathbb{R}^M\times\mathbb{R}^M} \mathscr{X}\{\mathbf{X}_{\text{pre}}\mathbf{B}\}\,. \tag{8}$$

where $\mathbf{X}_{\text{pre}}$ is an orthogonal matrix. The optimization of (8) results in a square unmixing matrix $\mathbf{B}$ which in turn provides a source matrix $\mathbf{S}_{\text{WEP}} = \mathbf{X}_{\text{pre}}\mathbf{B}$ and a mixing matrix $\mathbf{A}_{\text{WEP}} = \mathbf{B}^{-1}\mathbf{A}_{\text{pre}}$. Note that the orthonormality of $\mathbf{X}_{\text{pre}}$ implies that $\mathbf{S}_{\text{WEP}}$ is column-normalized ($\mathbf{S}_{\text{WEP}} = \tilde{\mathbf{S}}_{\text{WEP}}$) iff $\mathbf{B}$ is column-normalized.

Solving the non-convex optimization problem (8) can be done by iteratively:

1. Optimizing (7) under the constraint that $\mathbf{X}_{\text{pre}}\mathbf{B}$ is column-normalized, assuming that the source selection matrices $\mathbf{W}_m$ are known. This provides $\mathbf{B}$, then an estimate of the source matrix $\mathbf{S} = \mathbf{X}_{\text{pre}}\mathbf{B} = \tilde{\mathbf{S}}$.
2. Updating $\mathbf{W}_m$ from $\mathbf{S}$ according to (6).

The advantage of assuming the matrices $\mathbf{W}_m$ known in the first step, is that the exclusion criterion becomes quadratic in $\mathbf{B}$ as seen in (7), so that we have to solve:

$$\begin{gathered} \min_{\mathbf{B}}\Big(1 - \frac{1}{M}\sum_{m=1}^{M} \|\mathbf{W}_m\mathbf{X}_{\text{pre}}\mathbf{B}\mathbf{e}_m\|^2\Big), \\ \text{constrained by } \operatorname{diag}\big(\mathbf{B}^{\text{T}}\mathbf{B}\big) = \mathbf{Id}. \end{gathered} \tag{9}$$

The optimal solution of this problem is obtained by selecting the largest eigenvalue λ_m of the SVD of $\mathbf{W}_m\mathbf{X}_{\text{pre}}$ for each m:

$$\mathbf{X}_{\text{pre}}^{\text{T}}\mathbf{W}_m\mathbf{X}_{\text{pre}}\mathbf{B}\mathbf{e}_m = \lambda_m\mathbf{B}\mathbf{e}_m, \quad m = 1, \ldots, M. \tag{10}$$

This provides $\mathbf{B}\mathbf{e}_m$ for each m (i.e., the mth column of the matrix $\mathbf{B}$), and eventually the matrix $\mathbf{B}$: see a summary of the sequence of instructions in the Algorithm box 1.

The algorithm starts from a random initialization of the source matrix and stops iterating when the matrices $\mathbf{W}_m$ do not change anymore. Due to the non-convexity of the optimization criterion (8), the final result of this iterative algorithm may depend on the initialization—the first estimate of the source matrix. Empirically, trying out no more than 10 different (random) initializations, and choosing the one that has the smallest exclusion is sufficient to achieve excellent results.

When the sources are exactly exclusive, the algorithm is provably exact within just one iteration. Even when the sources are weakly exclusive, this algorithm still provides an excellent classification of the hyperspectral image into dominant sources, accurately matching the reference segmentation/labeling in our tests (see Section IV); i.e., the source selection matrices $\mathbf{W}_m$, $m = 1, 2, \ldots M$ are accurately retrieved by the algorithm.

C. WEP+: enforcing ANC and ASC[2]

Applied to the two pre-processing options (see previous subsection), the WEP algorithm provides two results, of which we keep the one that has the smallest exclusion.

[2]Implementation available at https://www.ee.cuhk.edu.hk/~tblu/WEP.

Algorithm 1: WEP Optimization

Data: $K \times L$ mixed sources matrix $\mathbf{X}$
Result: $M \times L$ mixing matrix $\mathbf{A}_{\mathrm{WEP}}$
$K \times M$ source matrix $\mathbf{S}_{\mathrm{WEP}}$

$\mathbf{X} \xrightarrow[\text{approximation}]{\text{rank-}M\text{ SVD}} \mathbf{X}_{\mathrm{pre}}\mathbf{A}_{\mathrm{pre}}$ % Dimension % reduction

% Initialization
$\mathbf{S} = \mathrm{random}(K, M)$
$\mathbf{S} \to \tilde{\mathbf{S}}$ % Column normalization (4)
$\tilde{\mathbf{S}} \to \{\mathbf{W}_m\}_{m=1,2,\dots M}$ % Labeling (6)
% Minimization of the exclusion
repeat
  % Estimation of the unmixing matrix $\mathbf{B}$
  for $m = 1$ to M **do**
    $\mathbf{W}_m\mathbf{X}_{\mathrm{pre}} \xrightarrow{\text{pre}} \mathbf{U}\mathbf{D}\mathbf{V}^{\mathrm{T}}$
    $\mathbf{B}\mathbf{e}_m = \mathbf{V}(:, 1)$ % Largest singular value
  end
  % Update of the source matrix $\mathbf{S}$
  $\mathbf{S} = \mathbf{X}_{\mathrm{pre}}\mathbf{B}$
  % Update of the labeling (6)
  $\mathbf{S} \to \{\mathbf{W}_m\}_{m=1,2,\dots M}$
until $\mathbf{W}_m$ is unchanged between iterations
% Estimation of $\mathbf{A}_{\mathrm{WEP}}$ and $\mathbf{S}_{\mathrm{WEP}}$
$\mathbf{S}_{\mathrm{WEP}} = \mathbf{S}$
$\mathbf{A}_{\mathrm{WEP}} = \mathbf{B}^{-1}\mathbf{A}_{\mathrm{pre}}$

Although the resulting mixing matrix is already quite close to the reference (SAD metrics), the associated source matrix $\mathbf{S}_{\mathrm{WEP}}$ does not satisfy the hyperspectral unmixing constraints $\mathbf{S}_{\mathrm{WEP}} \geq \mathbf{0}_K$ (positivity/ANC) and $\mathbf{S}_{\mathrm{WEP}}\mathbf{1}_M = \mathbf{1}_K$ (sum-to-one/ASC). To mitigate this issue, we devise a two-step refinement that ensures that the final source matrix $\mathbf{S}_{\mathrm{WEP}+}$ satisfies these two constraints, while ensuring a low exclusion value. First, we leverage our accurate knowledge of the source selection matrices, $\mathbf{W}_m$, by reformulating exclusion using a convex (linear) term that regularizes a convex (quadratic) data term, under the convex sum-to-one positivity constraints: this provides our final estimate, $\mathbf{A}_{\mathrm{WEP}+}$, of the mixing matrix. Then, we retrieve our final estimate, $\mathbf{S}_{\mathrm{WEP}+}$, of the source matrix by solving another convex optimization problem.

1) Estimation of the mixing matrix $\mathbf{A}_{\mathrm{WEP}+}$: We solve for the $M \times M$ unmixing matrix $\mathbf{B}$ and for the $K \times M$ source matrix $\mathbf{S}$ that minimize the following convex problem

$$\min_{\substack{\mathbf{B},\mathbf{S}\geq\mathbf{0}_K \\ \mathbf{S}\mathbf{1}_M=\mathbf{1}_M}} \overbrace{\frac{1}{2}\underbrace{\|\mathbf{X}_{\mathrm{pre}}\mathbf{B}-\mathbf{S}\|_{\mathrm{F}}^2}_{\text{data term}} + \frac{\lambda}{\mathrm{M}}\underbrace{\mathrm{Tr}(\mathbf{W}^{\mathrm{T}}\mathbf{S})}_{\text{exclusion-like}}}^{\text{criterion}(\mathbf{B},\mathbf{S})}, \tag{11}$$

where the $K \times M$ matrix $\mathbf{W}$ is basically a binary mask that indicates whether, at a particular pixel, the source is not dominant: minimizing the non-dominant coefficients reduces the value of the exclusion. It is mathematically defined from the WEP source selection matrices $\mathbf{W}_m$, by

$$\mathbf{W} \stackrel{\mathrm{def}}{=} \mathbf{1}_K\mathbf{1}_M^{\mathrm{T}} - \sum_{m=1}^{M}\mathbf{W}_m\mathbf{1}_M\,\mathbf{e}_m^{\mathrm{T}}.$$

The regularization parameter λ provides a trade-off between the data term (how close is the source matrix to the unmixed raw data?) and the exclusion of the source matrix.

We know that there is a unique minimum (although possibly several equivalent solutions) to this optimization problem because the criterion is convex: any local minimum is a global minimum. To find this minimum, we alternate a minimization over $\mathbf{B}$ and over $\mathbf{S}$:

1. $\min_{\mathbf{B}}\{\mathrm{criterion}(\mathbf{B},\mathbf{S})\}$: solution of this unconstrained least-square fitting problem is
$$\mathbf{B} = \mathbf{X}_{\mathrm{pre}}^{\mathrm{T}}\mathbf{S},$$
taking into account the orthonormality of $\mathbf{X}_{\mathrm{pre}}$.
2. $\min_{\mathbf{S}}\{\mathrm{criterion}(\mathbf{B},\mathbf{S})\}$, subject to $\mathbf{S} \geq \mathbf{0}_K$ and $\mathbf{S}\mathbf{1}_M = \mathbf{1}_K$: solution of this constrained least-square fitting problem is
$$\mathbf{S} = (\mathbf{X}_{\mathrm{pre}}\mathbf{B} - \lambda\mathbf{W} - \mathbf{a}\mathbf{1}_M^{\mathrm{T}})_+$$
where $x_+ \stackrel{\mathrm{def}}{=} xH(x)$ is the one-sided power function if $H(x)$ is Heaviside's unit step function. Here $\mathbf{a}$ is a $K \times 1$ vector that is uniquely determined by the sum-to-one constraint $\mathbf{S}\mathbf{1}_M = \mathbf{1}_K$ (Lagrange multiplier).

This alternating minimization algorithm converges to a solution that minimizes the criterion: we achieve a satisfactory result after about 30 iterations.

The mixing matrix $\mathbf{A}_{\mathrm{WEP}+}$ is then obtained as the minimum over $\mathbf{A}$ of the quadratic criterion $\|\mathbf{X} - \mathbf{S}\mathbf{A}\|^2$

$$\mathbf{A}_{\mathrm{WEP}+} = (\mathbf{S}^{\mathrm{T}}\mathbf{S})^{-1}\mathbf{S}^{\mathrm{T}}\mathbf{X}.$$

How should we set the regularization coefficient λ in (11)? We observed that λ is influenced by the exclusion of the source matrix, by noise, and by the conditioning of the data matrix $\mathbf{X}$. More quantitatively, by defining an estimated Signal-to-Noise Ratio and an estimated condition number according to

$$\mathrm{SNR}^{\mathrm{est}} = 10\log_{10}\left(\frac{\sum_{k\geq 1}\lambda_{\mathrm{k}}^2}{\sum_{k\geq M+1}\lambda_{\mathrm{k}}^2}\right), \tag{12}$$

$$\mathrm{cond}^{\mathrm{est}} = \left|\frac{\lambda_1}{\lambda_M}\right|, \tag{13}$$

where $\lambda_1, \lambda_2, \dots \lambda_L$ are the singular values of the matrix $\mathbf{X}$ (in descending order), the regularization parameter λ can be set to:

$$\lambda = 1.6 \times \begin{cases} 10^{-\frac{\mathrm{SNR}^{\mathrm{est}}}{20}}\mathrm{cond}^{\mathrm{est}}\mathrm{e}^{-\mathrm{SNR}^{\mathrm{est}}\mathscr{X}\{\mathbf{S}\}}, \\ \qquad \text{if } \mathrm{cond}^{\mathrm{est}} \leq \frac{9}{M}10^{\frac{\mathrm{SNR}^{\mathrm{est}}}{20}}; \\ 1, \qquad \text{if } \mathrm{cond}^{\mathrm{est}} > \frac{9}{M}10^{\frac{\mathrm{SNR}^{\mathrm{est}}}{20}}. \end{cases} \tag{14}$$

We determined this formula empirically. We do not claim that it is optimal: it is just sufficiently accurate so that it is not necessary to tune λ for every new mixed sources. To find this formula, we started with 4 sources and tested successively (a basic line search to minimize the SAD) the individual

influence of the SNR (in a 20–40 dB range), the exclusion (in a 5–20% range), and the condition number (in a 1–10 range), while keeping the other two fixed. This provided us with a rough expression in which we rounded the constants. Then, performing about 30000 tests varying randomly the three parameters above, we identified the "phase transition" behavior when the condition number becomes too small. Finally, we checked the limited sensitivity of this formula to the number sources (up to 10). Note that real datasets seem to have undergone some form of spectral denoising, which corrupts the expected relation between $\mathrm{SNR}^{\mathrm{est}}$ and $\mathrm{cond}^{\mathrm{est}}$ in (14): $\mathrm{SNR}^{\mathrm{est}}$ is artificially high (typically, about 40 dB), whereas $\mathrm{cond}^{\mathrm{est}}$ was left unchanged. As a consequence, we observed that $\lambda = 1.6$ is an appropriate choice for real data, consistent with a reduction of $\mathrm{SNR}^{\mathrm{est}}$ by ~ 20 dB. See further comments in Section IV-D.

2) Estimation of the source matrix $\mathbf{S}_{\mathrm{WEP}+}$: We calculate of the closest sum-to-one, nonnegative source matrix $\mathbf{S}_{\mathrm{WEP}+}$, i.e., the one that minimizes the RMSE, $\|\mathbf{X} - \mathbf{S}\mathbf{A}_{\mathrm{WEP}+}\|_{\mathrm{F}}$, over all $\mathbf{S}$ that satisfy the constraints $\mathbf{S} \geq \mathbf{0}_K$ and $\mathbf{S}\mathbf{1}_M = \mathbf{1}_M$. This convex problem was already addressed successfully in the literature, leading to the FCLS algorithm [12]. However, the relatively large computation time of this classical algorithm led us to revisit it.

This constrained optimization problem is decoupled row-wise, so we can simply minimize independently the contribution to the RMSE of each row; or, by transposition, minimize $\|\mathbf{X}^{\mathrm{T}}\mathbf{e}_k - \mathbf{A}_{\mathrm{WEP}+}^{\mathrm{T}}\mathbf{S}_{\mathrm{WEP}+}^{\mathrm{T}}\mathbf{e}_k\|$ for every row $k = 1, 2, \ldots K$. This is a much smaller quadratic programming problem: the unknown is $\mathbf{u}_k = \mathbf{S}_{\mathrm{WEP}+}^{\mathrm{T}}\mathbf{e}_k$, an M-dimensional vector which satisfies the positivity and sum-to-one constraints. To take advantage of Matlab's very efficient `lsqnonneg` function (nonnegative least squares algorithm due to Lawson and Hanson [44, p. 161]), we choose to solve it approximately by considering the close optimization problem

$$\min_{\mathbf{u}_k \geq \mathbf{0}_M} \|\mathbf{X}^{\mathrm{T}}\mathbf{e}_k - \mathbf{A}_{\mathrm{WEP}+}^{\mathrm{T}}\mathbf{u}_k\|^2 + \mu|\mathbf{1}_M^{\mathrm{T}}\mathbf{u}_k - 1|^2$$

where μ is large (typ. 10^6). We then just apply this function to every row of $\mathbf{X}$ to obtain all the rows of $\mathbf{S}_{\mathrm{WEP}+}$. This approach is significantly faster than FCLS by about 50% on average.

III. Metrics and Visualization

A. Metrics

To measure the quality of an unmixing algorithm, the accuracy of the source matrix (abundance maps) and of the mixing matrix (endmembers) have to be evaluated. Consistently with the literature on this topic, we choose the Root-Mean-Square Error (RMSE) and the Spectral Angle Distance (SAD) for this purpose:

$$\mathrm{RMSE\_S} = \frac{\|\mathbf{S}_{\mathrm{ref}} - \mathbf{S}_{\mathrm{est}}\|_{\mathrm{F}}}{\sqrt{KM}}, \tag{15}$$

$$\mathrm{SAD} = \frac{1}{M}\sum_{m=1}^{M} \arccos\left(\frac{|\mathbf{e}_m^{\mathrm{T}}\mathbf{A}_{\mathrm{ref}}\mathbf{A}_{\mathrm{est}}^{\mathrm{T}}\mathbf{e}_m|}{\|\mathbf{A}_{\mathrm{ref}}^{\mathrm{T}}\mathbf{e}_m\|\|\mathbf{A}_{\mathrm{est}}^{\mathrm{T}}\mathbf{e}_m\|}\right).$$

Additionally, we stress that the exclusion principle presupposes that hyperspectral images can be segmented into somewhat exclusive sources. Hence, the accuracy of such an exclusion-based labeling may be a good indicator of the overall quality of an unmixing algorithm. Given a reference source matrix $\mathbf{S}_{\mathrm{ref}}$, we first find the permutation of columns of the matrix $\mathbf{S}_{\mathrm{est}}$ that best agrees with $\mathbf{S}_{\mathrm{ref}}$ in terms of dominant sources. We then define the labeling error as the proportion of pixels whose labeling in $\mathbf{S}_{\mathrm{est}}$ disagrees with the labeling in $\mathbf{S}_{\mathrm{ref}}$.

B. Visualization

1) Exclusion map: Usually, the source matrices/abundance maps are shown individually, but comparing all these (often small) images between different algorithms is visually difficult. Moreover, outlining the purity of the source matrices (i.e., their exclusion) in such a representation would prove even harder. For this reason, we have resorted to showing the source maps on a unique image. The first step is to realize that when the sources are exactly exclusive, each of them can be represented with a unique color, as a labeling map. It is then natural to choose saturated colors for the labeling map, and to indicate the exclusion value by linking it to the saturation of that color: more saturation, more exclusive; less saturation (more whiteness), less exclusive, as illustrated in Figure 2.

2) Labeling error map: Considering a segmentation of hyperspectral sources based on the dominance of each source, we show labeling errors as bright pixels on a dimmed version of the exclusion map of the reference source matrix (see Fig. 3b). The percentage of such errors is shown in the title of this labeling error map.

IV. Experiments

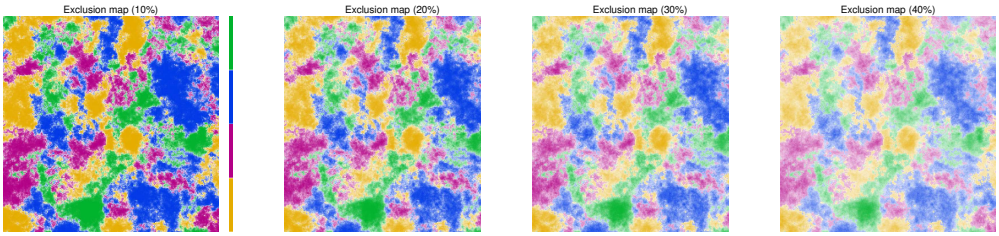
A. Experimental Data

Public hyperspectral datasets do not contain real ground-truth source images, but rather calculated reference images. In Table I, we showed that the exclusion values of these calculated references have an average exclusion value of 9.04%, which substantiates our WEP-based approach. We also choose this average exclusion as a target value when we generate synthetic datasets.

We first use synthetic data to perform a quantitative analysis of the parameters (typically, exclusion values, noise, image size, and number of sources) that affect the performance of HSU algorithms, expressed using the SAD and RMSE metrics. Detailed experimental protocols and their corresponding outcomes are listed in Section IV-C below. Then we test on real-world data, comparing algorithm results with the calculated references ("ground-truth"), and visualizing them using the tools described in Section III-B. The experimental data that we are showing here is from the Urban dataset (see Section IV-D below), but the results with other datasets are available in the supplementary materials.

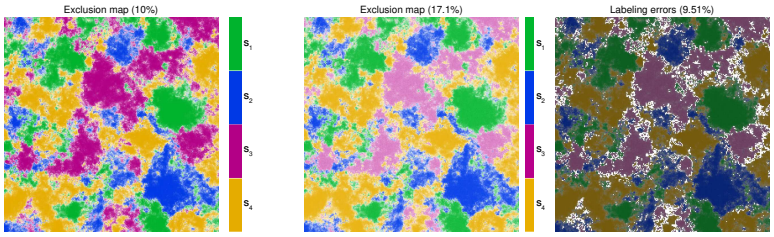
B. Algorithms Considered

Here is the list of algorithms, both classical and state-of-the-art, that we are comparing with:



Exclusion=10% Exclusion=20% Exclusion=30% Exclusion=40%

Fig. 2: Color saturation is used to represent exclusion values: higher exclusion, lower saturation.



(a) Exclusion map (ground-truth) (b) Exclusion map (MVES [38]) and labeling errors

Fig. 3: Visualization of unmixed sources (ground-truth or from unmixing algorithm).

- Fully Constrained Linear Spectral Unmixing with Vertex Component Analysis (VCA-FCLS)[3] [28], [45] uses the pure pixel hypothesis.
- Perturbed Linear Mixing Model (PLMM)[4] [18] operates based on the spectrum variability.
- Nonnegative Matrix Factorization Quadratic Minimum Volume (NMF-QMV)[5] [46] relies on the minimum volume hypotheses.
- Highly mixed/Ill-conditioned Spectrum UNmixing (HISUN)[6] [41] belongs to the geometry class.
- Minimum Simplex Convolutional Network (MiSiCNet)[7] [23] exploits the minimum volume and sparsity hypotheses.
- Dual-Feature Fusion Network (DFFN)[8] [24] builds an Autoencoder that focuses on spectral and spatial similarities.
- Stationary Wavelet Convolutional Network (SWC-Net)[9] [25] builds a network that focuses on spectral variation.

VCA-FCLS was used as the baseline methodology for comparative assessments. While PLMM represents a classical algorithmic approach, the remaining methods exemplify recent developments in the field. Notably, MiSiCNet, DFFN and SWC-Net incorporate autoencoder neural network architectures as their primary processing frameworks. Results of the raw WEP algorithm are given as reference only (to show its already good accuracy of the mixing matrix, and the low computation time), since the source matrix does not satisfy ANC+ASC. The results of our HSU algorithm are shown in the WEP+ column.

We utilized publicly available source codes with parameter configurations as specified in their respective publications. In cases where specific parameters were not explicitly documented, default values from the provided implementations were retained. We checked that we were using these implementations adequately by making sure to obtain the same results as the ones in the papers. The computational setup consists of Matlab R2021a and PyTorch, executed on a Windows 10 platform with 64GB RAM capacity. The computation times indicated in the tables include all the pre- and post-processing required by the different algorithms (including the multiple initializations of WEP).

[3]https://github.com/davidkun/HyperSpectralToolbox/tree/master/functions
[4]https://github.com/pthouvenin/unmixing-plmm
[5]https://github.com/LinaZhuang
[6]https://github.com/IHCLab/HiSun
[7]https://github.com/BehnoodRasti/MiSiCNet
[8]https://github.com/xuanwentao/DFFN.git
[9]https://github.com/UPCGIT/SWC-Net.git

C. Synthetic Data Experiments

Our simulations control for four parameters: exclusion values of source matrix, noise level, image size, and source number. Data were generated by the widely-recognized HYperspectral Data Retrieval and Analysis (HYDRA) toolbox [47], which has demonstrated extensive usefulness in simulation studies across the literature [48], [49], [50]. Considering the diversity of data, different abundance distributions were set using HYDRA, which may cause small discrepancies in numbers between different experiments. Throughout the data synthesis process, a standardized exclusion value of 9.04% (from Table I) was enforced in all datasets, except in the exclusion experiments. To ensure that the exclusion reaches a specific value (9.04%, but also 10%, 20%, 30% and 40%), we exponentiated the HYDRA source matrix element-wise, and adjusted the exponent by (dichotomous) line search.

Given that the spectral library contains a limited number of idealized waveforms, we realized that the mixing matrix it produces tends to be badly conditioned, leading to a value of $\mathrm{cond}^{\mathrm{est}}$ exceeding 35 (above 100 when seven or ten sources are considered), we supplemented our testing with randomly generated better-conditioned mixing matrices (resulting in $\mathrm{cond}^{\mathrm{est}} < 15$, typically): for every simulation, we

have used two groups of mixing matrices, providing a more comprehensive evaluation of the algorithms.

1) Simulation on the level of exclusion (SIM1): We first selected four distinct materials from the United States Geological Survey (USGS) spectral library[10] as endmembers: Jarosite GDS99K Sy 200C, Anorthite HS349.3B, Howlite GDS155, and Corrensite CorWa-1. We also generated randomly another mixing matrix with a better condition number (leading to $\mathrm{cond}^{\mathrm{est}} < 15$). A source matrix of size $256 \times 256 \times 4$ pixels was then synthesized, for four different exclusion values: 10%, 20%, 30%, and 40%. Following the linear mixing model, hyperspectral data were generated across 224 spectral channels, each channel being corrupted by $30\,\mathrm{dB}$-SNR Gaussian noise.

Table II quantifies the performance of WEP+ compared to other algorithms. In terms of mixing matrix accuracy, we can see that WEP is already good despite not enforcing ANC and ASC, and very fast (more than four times faster than the fastest algorithm tested). At a small increase in computational cost, WEP+ exhibits a significant reduction of SAD values to low single digits, even when the exclusion is large. As can be seen in this table, WEP+ outperforms most of the other algorithms for these metrics, whereas the second best performing algorithm, NMF-QMV, is 30–40 times slower.

Most algorithms exhibit progressive deterioration of the accuracy of the source matrix (RMSE_S) when exclusion increases. For instance, MiSiCNet's error rate escalates from 2.5% to 10% in Table II. Notably, WEP+ maintains remarkable stability with error rates consistently ranging between 1% and 3%, demonstrating robust performance, even when the exclusion reaches 40%.

2) Simulation on the level of noise (SIM2): Synthesized hyperspectral image with $256 \times 256 \times 224$ pixels, generated from a source matrix with $256 \times 256 \times 4$ pixels (exclusion $= 9.04\%$), and mixed with the same endmembers as SIM1 ($\mathrm{cond}^{\mathrm{est}} > 35$) or with randomly generated ones ($\mathrm{cond}^{\mathrm{est}} < 15$). To evaluate the robustness of HSU against noise perturbations, iid Gaussian noise was added to every spectral channel, with a signal-to-noise ratio (SNR) level of $20\,\mathrm{dB}$, $30\,\mathrm{dB}$, and $40\,\mathrm{dB}$.

The results, shown in Table III, demonstrate the consistently superior performance of WEP+ across all noise conditions in terms of accuracy of retrieval of the source matrix (RMSE_S). And we should point out that this performance is achieved without sacrificing computational efficiency. Notably, WEP+ achieved lower RMSE values without additional denoising, in contrast with some of the other algorithms (see, e.g., uDAS's integrated denoising component). This underscores the inherent robustness of WEP+ to noise perturbations.

3) Simulation on the size of HS images (SIM3): Contemporary applications of hyperspectral imaging, particularly in airborne remote sensing, frequently involve large-scale imagery. While subdividing large images into smaller sections for separate processing is feasible, this approach not only doubles computational overhead but also sacrifices inherent spatiotemporal correlation information. Therefore, the ability of the algorithm to process large-scale images directly should be tested.

The dataset was synthesized from sources of size $100 \times 100 \times 4$, $500 \times 500 \times 4$ and $1000 \times 1000 \times 4$ (9.04% exclusion each), mixed with the same endmembers as SIM1 ($\mathrm{cond}^{\mathrm{est}} > 35$) or with randomly generated ones ($\mathrm{cond}^{\mathrm{est}} < 15$), further corrupted by $30\,\mathrm{dB}$ noise.

The results shown in Table IV highlight the computational burden of most of the HSU algorithms tested here, to the exception of VCA-FCLS and WEP+. This is particularly obvious for the largest image size (1000×1000), where these algorithms need at least seven minutes to complete or simply run out of memory, whereas WEP+ and VCA-FCLS only need slightly more than one minute. Notably, the original WEP readily gets good SAD results within a few seconds. Again, this obvious computational advantage is not traded for a loss of accuracy since WEP+ is outperforming all the other algorithms with respect to all the metrics considered here (RMSE_S and SAD).

4) Simulation on the number of sources (SIM4): Using the HYDRA platform, we synthesized hyperspectral sources of size $256 \times 256 \times 4$, $256 \times 256 \times 7$, and $256 \times 256 \times 10$, mixed them with either 4, 7 or 10 endmembers from the USGS library spectrum, or from a better conditioned random set of endmembers ($\mathrm{cond}^{\mathrm{est}} < 15$), then added $30\,\mathrm{dB}$ noise.

As depicted in Table V, all the algorithms are able to deal with the increased complexity caused by a larger number of sources. Here again, WEP+ outperforms the other algorithms when the accuracy of the retrieved abundances is considered, although it is somewhat more challenged as far as SAD is concerned, at least when the mixing matrix is badly-conditioned. Not only is WEP+ consistently more accurate than the other algorithms in general, but also its computational cost is significantly lower: e.g., in the 10-source case in Table V, it is more than twice faster than VCA-FCLS, and at least seven times faster than the other algorithms.

D. Real Data Experiments

Datasets of real hyperspectral images do not seem to behave similarly to simulated datasets.

a) Inaccurate ground-truth: The primary issue with real datasets is the inability to measure reliable ground-truths. For most real datasets, the so-called ground-truths are actually calculated, through a process that involves manual endmember extraction from a spectral library (mixing matrix), followed by the use of a factorization algorithm (source matrix) [51], [52]. In some cases (e.g., WDC and Apex datasets), the calculated references and the actual hyperspectral data exhibit significant discrepancies, rendering the comparison with the "ground-truth" information questionable. This observation is supported by previous literature [53], which outlines that direct unmixing of these datasets rarely yields results conforming to their provided references.

Therefore, these so-called ground-truths are not fully reliable, and should instead be considered references—calculated references. This is a caution for the currently popular approach

[10]https://www.usgs.gov/labs/spectroscopy-lab/science/spectral-library

TABLE II: Simulations with different exclusion values

	Metric	Exclusion (%)	VCA-FCLS	PLMM	NMF-QMV	HISUN	MiSiCNet	DFFN	SWC-Net	WEP+	WEP
good conditioning $\mathrm{cond}^{\mathrm{est}} < 15$	SAD ($\times 10^{-2}$)	10	1.01	0.91	1.63	6.20	0.92	9.88	2.53	**0.12**	*5.21*
		20	1.57	1.36	1.20	3.30	2.32	7.03	2.91	**0.41**	*8.26*
		30	3.82	3.62	1.74	2.92	4.42	6.76	3.31	**1.07**	*10.58*
		40	8.27	8.46	3.63	4.43	7.03	13.25	3.60	**3.40**	*12.94*
	RMSE_S ($\times 10^{-2}$)	10	1.73	1.96	2.67	7.65	1.47	35.23	6.60	**1.09**	N/A
		20	2.57	3.00	1.87	3.75	2.90	56.39	3.63	**1.31**	N/A
		30	4.51	6.01	2.24	3.16	4.82	45.64	4.54	**1.76**	N/A
		40	10.25	10.35	3.54	5.08	6.92	42.46	7.43	**3.33**	N/A
	Time (s)	10	4.51	16.48	196.89	**2.77**	39.23	210.19	29.94	6.93	*0.62*
		20	**4.97**	16.46	195.66	18.38	31.36	296.73	25.91	7.26	*0.63*
		30	**4.72**	16.92	278.58	18.15	31.33	289.57	26.61	7.60	*0.77*
		40	**5.73**	22.64	297.43	18.40	31.39	258.92	24.75	7.98	*0.94*
bad conditioning $\mathrm{cond}^{\mathrm{est}} > 35$	SAD ($\times 10^{-2}$)	10	**0.92**	1.11	2.46	4.26	1.51	1.64	1.40	0.95	*5.22*
		20	0.99	1.10	1.85	3.95	3.52	6.43	2.30	**0.48**	*8.15*
		30	2.33	2.42	2.02	3.01	5.74	7.77	3.54	**0.53**	*10.19*
		40	5.30	5.43	3.05	1.46	8.07	15.05	4.35	**1.27**	*11.86*
	RMSE_S ($\times 10^{-2}$)	10	2.51	2.98	4.05	6.56	2.57	5.99	5.95	**1.78**	N/A
		20	2.42	2.93	2.78	4.92	4.94	17.54	5.45	**1.87**	N/A
		30	3.62	4.42	2.92	3.33	7.51	25.66	5.15	**2.30**	N/A
		40	7.33	9.15	3.37	3.03	10.16	40.46	5.00	**2.84**	N/A
	Time (s)	10	**3.99**	45.07	286.63	18.65	34.60	208.99	43.36	7.16	*0.64*
		20	**4.43**	56.68	183.38	18.79	30.48	207.97	47.11	7.26	*0.67*
		30	**4.16**	62.93	293.11	18.35	30.52	217.23	48.31	7.56	*0.79*
		40	**4.56**	63.31	285.08	18.42	30.61	221.94	46.42	7.89	*1.02*

Notes: $\mathrm{cond}^{\mathrm{est}}$ = condition number (13) of the mixed sources matrix (more details in Section IV-C). WEP results given as reference only, since the source matrix does not satisfy ANC+ASC.

TABLE III: Simulations with different noise levels

	Metric	Noise (dB)	VCA-FCLS	PLMM	NMF-QMV	HISUN	MiSiCNet	DFFN	SWC-Net	WEP+	WEP
good conditioning $\mathrm{cond}^{\mathrm{est}} < 15$	SAD ($\times 10^{-2}$)	20	3.37	3.67	6.12	15.16	1.15	11.10	7.80	**0.89**	*5.35*
		30	1.25	1.14	1.71	7.07	0.65	13.06	2.74	**0.21**	*5.49*
		40	0.30	0.30	0.49	4.09	0.54	9.79	1.58	**0.09**	*5.51*
	RMSE_S ($\times 10^{-2}$)	20	5.50	6.83	7.78	13.97	3.59	46.09	10.47	**3.37**	N/A
		30	1.89	1.89	2.66	8.44	1.51	43.55	4.05	**1.07**	N/A
		40	0.55	0.57	0.82	5.44	1.27	40.45	4.99	**0.37**	N/A
	Time (s)	20	4.20	53.15	201.58	**2.78**	31.43	200.34	24.61	7.04	*0.56*
		30	**4.46**	16.24	235.24	14.68	31.23	222.77	24.98	7.05	*0.63*
		40	**3.88**	15.70	197.95	14.48	31.44	229.19	25.03	7.11	*0.55*
bad conditioning $\mathrm{cond}^{\mathrm{est}} > 35$	SAD ($\times 10^{-2}$)	20	2.84	2.00	8.69	10.66	1.53	3.95	4.07	**1.51**	*4.69*
		30	**0.87**	1.04	2.38	3.90	1.35	2.76	1.49	0.97	*4.86*
		40	0.29	**0.24**	0.67	1.30	1.38	2.42	1.00	0.34	*4.87*
	RMSE_S ($\times 10^{-2}$)	20	8.15	6.25	10.98	13.92	5.04	28.43	14.25	**4.04**	N/A
		30	2.80	2.84	4.14	6.51	2.38	25.83	17.00	**1.76**	N/A
		40	0.88	0.84	1.28	2.45	2.08	25.04	18.11	**0.64**	N/A
	Time (s)	20	**4.52**	79.67	186.01	18.64	34.97	266.95	38.54	6.78	*0.65*
		30	**4.20**	43.70	183.83	19.19	31.50	262.74	47.33	7.07	*0.59*
		40	**3.72**	27.52	285.78	17.92	31.25	247.67	44.02	7.23	*0.63*

Notes: $\mathrm{cond}^{\mathrm{est}}$ = condition number (13) of the mixed sources matrix (more details in Section IV-C). WEP results given as reference only, since the source matrix does not satisfy ANC+ASC.

of using these references as training sets for supervised learning networks, as this practice may compromise the authenticity of HSU results from the start.

b) Prior spectral denoising: Real datasets do not contain raw data, but already preprocessed data, although the detailed information of this processing may have been lost. In particular, it seems that it is standard for hyperspectral data to have undergone spectral denoising (e.g., using a Savitzky-Golay filter [54]). Such a preprocessing may conceal the actual amount of noise in the data, and may not be consistent with the actual number of sources that can be retrieved, or with the conditioning of the actual mixing matrix as pointed out in Section II-C. We deal with this issue by reducing the calculated noise level, $\mathrm{SNR}^{\mathrm{est}}$, by $\sim 20\,\mathrm{dB}$ in (14): this is consistent with the hidden spectral denoising mentioned above.

[11]https://lesun.weebly.com/hyperspectral-data-set.html

E. Real experiments on Urban

The Urban dataset consists of a $307 \times 307 \times 162$ hyperspectral image (2×2 m^2 resolution, wavelengths from 400 to 2500 nm). This image presents substantial HSU challenges due to its complex source composition and distribution. Captured as an aerial view of a community, it encompasses diverse urban elements including buildings (asphalt), roads, and vegetation (grass, trees). The spatial distribution of these elements is characterized by significant overlap and intricate mixing patterns. As shown in Table VI, WEP+ achieves a good SAD value (7.52%). Other methods yield SAD values between twenty and forty, indicating less satisfactory HSU performance. Figure 5 illustrates that these poor HSU results are mainly attributed to the difficulty in accurately extracting asphalt and road spectra, with most algorithms exhibiting substantial deviations from reference signatures for these two

TABLE IV: Simulations with different image sizes

	Metric	Size	VCA-FCLS	PLMM	NMF-QMV	HISUN	MiSiCNet	DFFN	SWC-Net	WEP+	WEP
good conditioning $\mathrm{cond}^{\mathrm{est}} < 15$	SAD ($\times 10^{-2}$)	100×100	0.95	0.84	1.61	6.65	3.34	5.22	3.07	**0.14**	*4.58*
		500×500	2.03	1.08	1.76	5.67	1.26	21.77	2.39	**0.11**	*5.20*
		1000×1000	1.37	OOM	1.05	7.18	2.18	27.18	2.09	**0.16**	*5.37*
	RMSE_S ($\times 10^{-2}$)	100×100	1.93	1.77	2.77	8.69	3.81	25.98	6.06	**1.03**	N/A
		500×500	3.15	2.16	2.75	7.18	2.38	38.68	5.11	**1.04**	N/A
		1000×1000	2.14	OOM	1.91	8.30	3.56	83.85	4.41	**1.08**	N/A
	Time (s)	100×100	**0.86**	4.03	13.36	1.73	13.23	45.53	5.86	0.88	*0.13*
		500×500	**20.73**	69.46	538.33	40.46	111.86	489.11	78.59	21.99	*2.06*
		1000×1000	**65.17**	OOM	1963.12	145.55	442.89	1864.91	525.86	85.54	*8.36*
bad conditioning $\mathrm{cond}^{\mathrm{est}} > 35$	SAD ($\times 10^{-2}$)	100×100	0.75	0.92	2.38	3.84	3.81	0.95	2.22	**0.93**	*4.59*
		500×500	1.09	1.03	2.51	3.98	3.64	9.28	2.94	**0.99**	*4.79*
		1000×1000	1.10	OOM	3.44	4.63	**0.86**	21.66	1.86	0.88	*4.69*
	RMSE_S ($\times 10^{-2}$)	100×100	2.38	2.02	4.18	6.33	5.68	11.37	9.76	**1.79**	N/A
		500×500	2.97	2.48	4.40	6.62	6.36	26.18	11.21	**1.92**	N/A
		1000×1000	3.28	OOM	4.15	7.53	6.02	52.82	8.87	**1.81**	N/A
	Time (s)	100×100	**0.92**	9.11	14.67	2.94	12.84	49.48	13.84	1.00	*0.13*
		500×500	**18.27**	228.89	529.97	48.13	110.00	996.95	135.35	27.04	*2.22*
		1000×1000	**77.37**	OOM	2787.42	174.30	443.88	2695.67	1879.60	91.35	*9.43*

Notes: $\mathrm{cond}^{\mathrm{est}}$ = condition number (13) of the mixed sources matrix (more details in Section IV-C).
WEP results given as reference only, since the source matrix does not satisfy ANC+ASC.

TABLE V: Simulations with different source numbers

	Metric	Number	VCA-FCLS	PLMM	NMF-QMV	HISUN	MiSiCNet	DFFN	SWC-Net	WEP+	WEP
good conditioning $\mathrm{cond}^{\mathrm{est}} < 15$	SAD ($\times 10^{-2}$)	4	1.00	1.04	1.71	7.07	0.60	13.06	2.82	**0.21**	*5.49*
		7	1.35	0.96	1.30	1.00	1.58	12.19	5.34	**0.87**	*6.64*
		10	1.63	**1.08**	1.80	OOM	3.00	16.48	6.50	1.78	*8.13*
	RMSE_S ($\times 10^{-2}$)	4	1.76	2.04	2.66	8.44	1.49	43.55	4.95	**1.07**	N/A
		7	1.38	1.28	1.60	1.33	1.33	41.17	5.55	**1.27**	N/A
		10	1.15	**1.08**	1.69	OOM	1.67	48.77	4.40	1.49	N/A
	Time (s)	4	**4.22**	15.71	200.86	14.56	31.36	222.39	25.09	6.31	*0.61*
		7	12.36	21.21	305.86	35.69	33.82	264.46	24.25	**9.23**	*0.80*
		10	25.65	24.78	413.39	OOM	36.37	277.11	24.73	**14.62**	*1.03*
bad conditioning $\mathrm{cond}^{\mathrm{est}} > 35$	SAD ($\times 10^{-2}$)	4	1.00	**0.75**	2.26	3.86	1.89	2.57	1.18	0.97	*4.88*
		7	1.34	**1.00**	7.06	1.67	2.72	11.58	22.40	3.16	*4.94*
		10	**1.55**	1.72	10.61	OOM	4.49	7.56	38.88	2.69	*5.54*
	RMSE_S ($\times 10^{-2}$)	4	3.25	2.77	3.88	6.54	4.22	13.50	6.64	**1.93**	N/A
		7	3.73	3.72	8.78	5.26	5.17	17.81	16.04	**3.68**	N/A
		10	3.07	4.43	8.88	OOM	5.89	15.14	13.17	**2.40**	N/A
	Time (s)	4	**4.57**	46.70	306.11	23.62	31.06	280.50	47.19	7.02	*0.57*
		7	13.75	74.34	307.44	37.35	34.27	270.68	46.81	**9.26**	*1.00*
		10	25.80	78.36	413.85	OOM	35.50	274.47	47.34	**13.17**	*1.27*

Notes: OOM means that the test ran out of memory.
$\mathrm{cond}^{\mathrm{est}}$ = condition number (13) of the mixed sources matrix (more details in Section IV-C).
WEP results given as reference only, since the source matrix does not satisfy ANC+ASC.

endmembers. Figure 4 presents the exclusion maps generated by WEP+ and other algorithms. From a holistic perspective, only WEP+, DFFN and MiSiCNet successfully unmix all four sources in their entirety. Examination of the labeling error maps reveals that, apart from WEP+ and DFFN, other algorithms exhibit large areas of source identification mistakes. These extensive labeling errors are consequences of significant inaccuracies in the mixing matrix estimation, which manifest as large-scale misidentifications in the corresponding source matrices. In conclusion, separating the Urban dataset proves to be a challenging task. However, WEP+ achieves largely satisfactory results, demonstrating its robustness in handling complex, highly mixed hyperspectral data. This performance underscores WEP+'s potential as an effective tool for urban remote sensing applications, where accurate material identification is crucial for urban planning and environmental monitoring.

We show more results on other standard hyperspectral datasets in the supplementary materials. The general observation is that WEP+ exhibits a performance that is on par with the best current algorithms and often, like for the Urban dataset, outperforms all the other algorithms, either when it is the SAD or the RMSE_S that is considered.

V. Discussion

Based on unmixing experiments of synthetic and real data, the WEP+ algorithm that we are presenting in this paper outperforms in general all the other algorithms in terms of quality (SAD, RMSE_S) and even in terms of computational cost. We believe that this is a validation of the main criterion—"exclusion"—that we are optimizing here. There are a few points on which we would like to elaborate:

- **Sum-to-one and positivity**—As a general blind source separation algorithm, the WEP does not enforce the ANC and ASC. For this reason, we had to "patch" it by post-processing its output in such a way that these two constraints are eventually satisfied (see Section II-C), resulting in the somewhat slower WEP+ algorithm—albeit much faster than most of the other algorithms tested. Yet, it is obvious that minimizing exclusion under ANC and ASC directly is

TABLE VI: Comparison of unmixing algorithms on the Urban dataset[11]

Metric	Source	VCA-FCLS	PLMM	NMF-QMV	HISUN	MiSiCNet	DFFN	SWC-Net	WEP+	WEP
SAD ($\times 10^{-2}$)	1	20.95	**8.04**	30.78	16.79	13.04	19.83	11.20	10.46	*11.34*
	2	33.39	34.18	14.54	7.16	70.59	14.10	21.11	**4.10**	*4.06*
	3	17.84	24.49	20.16	30.60	10.53	10.60	9.62	**4.92**	*8.00*
	4	72.21	73.16	**8.88**	49.85	18.15	36.74	29.83	10.59	*5.07*
	Overall	36.10	34.97	18.59	26.10	28.08	20.32	17.94	**7.52**	*7.12*
RMSE_S ($\times 10^{-2}$)	1	28.84	31.21	26.59	34.60	27.64	**16.75**	18.69	20.59	
	2	50.11	46.33	36.92	54.88	40.31	14.27	25.35	**13.42**	
	3	39.03	32.06	32.23	34.65	25.06	16.57	13.39	**7.55**	N/A
	4	17.90	25.77	19.82	19.09	12.89	33.82	**12.13**	14.98	
	Overall	36.01	34.69	29.59	38.00	28.21	21.81	18.16	**14.88**	
Time (s)	-	10.55	93.53	278.12	24.64	44.58	559.03	42.15	**8.15**	*0.93*

Note: WEP results given as reference only, since the source matrix does not satisfy ANC+ASC.

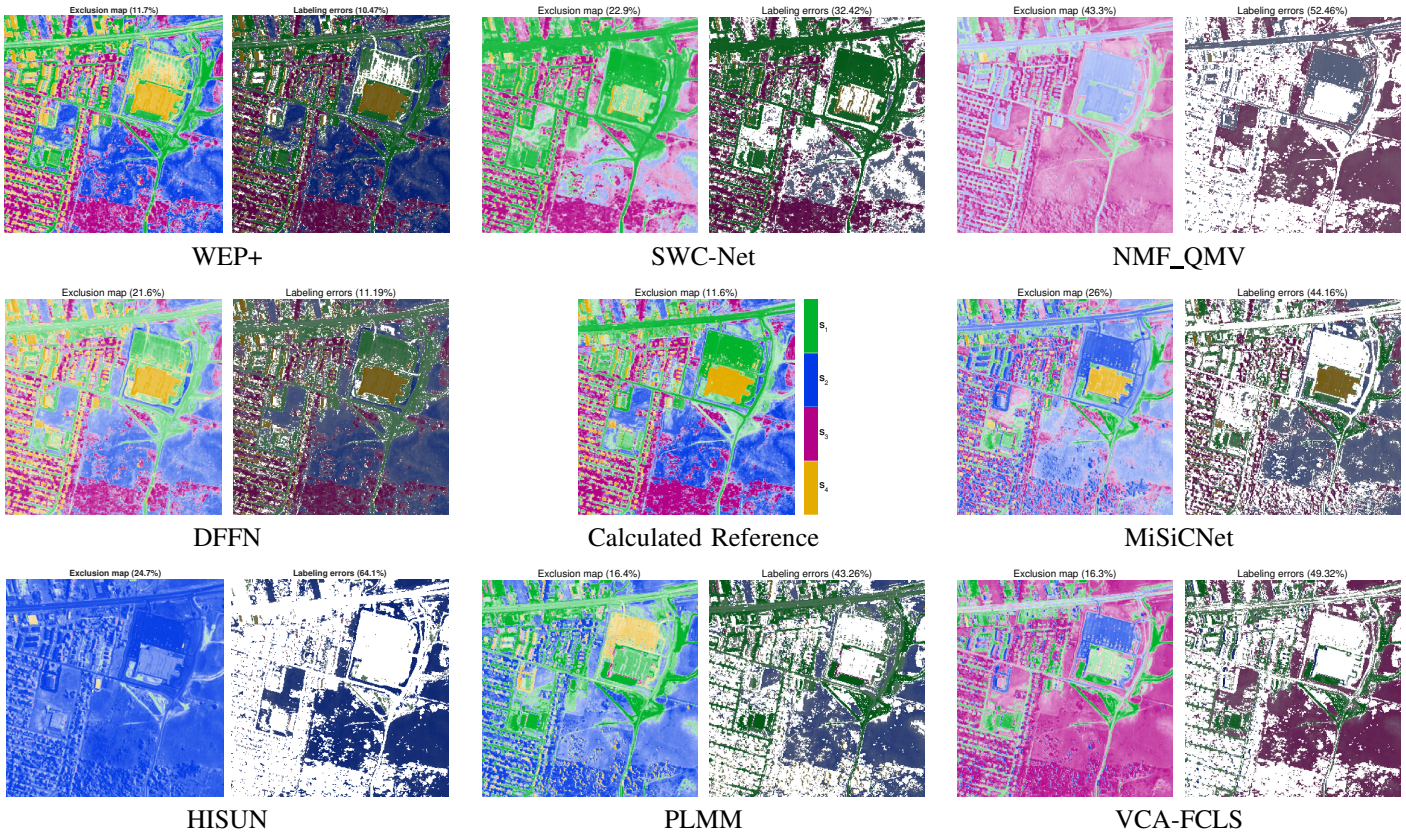


Fig. 4: Exclusion maps obtained by WEP+ and other algorithms on the Urban data (calculated reference shown for comparison): see Table VI for more details.

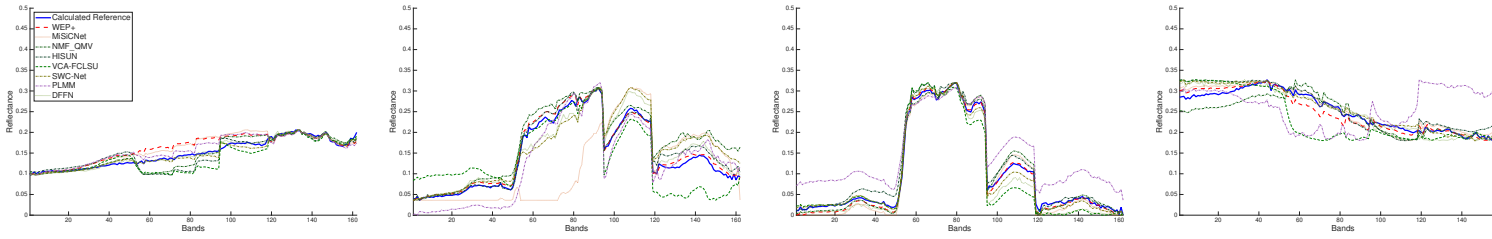


Fig. 5: Spectral endmembers calculated by WEP+ and other algorithms on the Urban data.

possible. We plan to develop WEP in this direction in the future.

- **Pre-processing**—The WEP algorithm takes as input a dimension-reduced version of the hyperspectral image, $\mathbf{X}_{\text{pre}}$ (see Section II-B). One of the goals of this pre-processing is noise reduction. But we have two options for reducing the dimension: while the first one is a standard PCA, the ANC justification for the second one may not seem compelling. In fact, the main reason for this second pre-processing is to ensure that sources with similar endmembers, but different

amplitudes, are identified as different by the WEP algorithm: one column of $\mathbf{X}_{\rm pre}$ is the same constant $(1/\sqrt{K})$ for every row (i.e., every pixel) while the other columns have row-dependent amplitudes proportional to the intensity of the sources. This is important in the real hyperspectral images that we had to deal with because "water" usually has a significantly lower intensity than other sources, and could easily be misclassified with other sources if intensity is not taken into account.

- **Real hyperspectral images**—We observe that real data behave very differently from synthetic data. For instance, they require the saturated version of this formula, as if the noise was large, in apparent contradiction with the cleanliness of the mixed sources. We surmised that this contradiction results from some form of spectral denoising. But it is also clear that the "noise" in hyperspectral images is highly correlated across bands, in contrast with synthetic data. This can be a way to tell apart real from synthetic data, but we believe that the current linear mixing hypothesis is not sufficiently accurate to deal with real data, probably because it does not take into some well-known artifacts like the Keystone and smile properties [55].
- **Parameter tuning**—The only parameter that we had to set in the WEP+ algorithm is the trade-off λ between the fidelity to the raw data and the level of exclusion. Indeed, it would be possible to tune this parameter for each new image, or to leave it as a button for user interaction. But our choice in this paper has been to "learn" its value in function of several characteristics of the mixed hyperspectral image, based on tens of thousands of synthetic tests. Our formula (14) is very reliable, and accounts for real data as well.

VI. Conclusion

We introduced a criterion, the "exclusion", which quantifies the global purity of a hyperspectral image. This led us to the principle that sources are optimally separated when they minimize the exclusion. Specifically, we developed an efficient algorithm for minimizing this criterion, while ensuring that the solution satisfies the ANC and ASC constraints that are intrinsic to hyperspectral sources. The high accuracy of the exclusion principle and the algorithmic efficiency of our implementation have been validated by comprehensive experiments using both simulated and real data, as compared with a host of standard unmixing methods. In a future contribution, we plan to provide theoretical performance guarantees to validate the WEP as an optimization principle for blind source separation.

Acknowledgments

We thank the anonymous reviewers for their constructive comments and Professor Jizhou Li for his insightful suggestions. This work was supported by the Yushan Fellowship awarded to Professor Blu.

References

[1] Z. Zihan and T. Blu, "Blind source separation via a weak exclusion principle," in ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2022, pp. 2699–2703.

[2] F. Perez, S. Rojas, C. Hinojosa, H. Rueda-Chacón, and B. Ghanem, "Unmix-nerf: Spectral unmixing meets neural radiance fields," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2025, pp. 26 284–26 293.

[3] S. Henriet, U. Şimşekli, S. Dos Santos, B. Fuentes, and G. Richard, "Independent-variation matrix factorization with application to energy disaggregation," IEEE Signal Processing Letters, vol. 26, no. 11, pp. 1643–1647, 2019.

[4] C. H. Ding, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorizations," IEEE transactions on pattern analysis and machine intelligence, vol. 32, no. 1, pp. 45–55, 2008.

[5] N. Seichepine, S. Essid, C. Févotte, and O. Cappé, "Piecewise constant nonnegative matrix factorization," in 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2014, pp. 6721–6725.

[6] T. Heittola, A. Mesaros, T. Virtanen, and A. Eronen, "Sound event detection in multisource environments using source separation," in Machine Listening in Multisource Environments, 2011.

[7] D. R. Hundley, M. J. Kirby, and M. Anderle, "Blind source separation using the maximum signal fraction approach," Signal processing, vol. 82, no. 10, pp. 1505–1508, 2002.

[8] T. Virtanen, A. Mesaros, and M. Ryynänen, "Combining pitch-based inference and non-negative spectrogram factorization in separating vocals from polyphonic music." in SAPA@ INTERSPEECH. Citeseer, 2008, pp. 17–22.

[9] Y. E. Shimabukuro and J. A. Smith, "The least-squares mixing models to generate fraction images derived from remote sensing multispectral data," IEEE Transactions on Geoscience and Remote sensing, vol. 29, no. 1, pp. 16–20, 1991.

[10] H. M. Horwitz, R. F. Nalepka, P. D. Hyde, and J. P. Morgenstern, "Estimating the proportions of objects within a single resolution element of a multispectral scanner." in International Symposium on Remote Sensing of Environment, 7th, University of Michigan, Ann Arbor, May 1971, Report No. 10259-1-X.

[11] H. Horwitz, J. Lewis, and A. Pentland, "Estimating proportions of objects from multispectral scanner data," Tech. Rep., 1975.

[12] D. Heinz, C.-I. Chang, and M. L. Althouse, "Fully constrained least-squares based linear unmixing [hyperspectral image classification]," in IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No. 99CH36293), vol. 2. IEEE, 1999, pp. 1401–1403.

[13] J. L. Silván-Cárdenas and L. Wang, "Fully constrained linear spectral unmixing: Analytic solution using fuzzy sets," IEEE Transactions on Geoscience and Remote Sensing, vol. 48, no. 11, pp. 3992–4002, 2010.

[14] S.-Y. Chen, Y.-C. Ouyang, C. Lin, H.-M. Chen, C. Gao, and C.-I. Chang, "Progressive endmember finding by fully constrained least squares method," in 2015 7th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS). IEEE, 2015, pp. 1–4.

[15] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," IEEE journal of selected topics in applied earth observations and remote sensing, vol. 5, no. 2, pp. 354–379, 2012.

[16] J. M. Nascimento and J. M. Bioucas-Dias, "Hyperspectral unmixing based on mixtures of Dirichlet components," IEEE Transactions on Geoscience and Remote Sensing, vol. 50, no. 3, pp. 863–878, 2011.

[17] Y. Qian, S. Jia, J. Zhou, and A. Robles-Kelly, "Hyperspectral unmixing via $l_{1/2}$ sparsity-constrained nonnegative matrix factorization," IEEE Transactions on Geoscience and Remote Sensing, vol. 49, no. 11, pp. 4282–4297, 2011.

[18] P.-A. Thouvenin, N. Dobigeon, and J.-Y. Tourneret, "Hyperspectral unmixing with spectral variability using a perturbed linear mixing model," IEEE Transactions on Signal Processing, vol. 64, no. 2, pp. 525–538, 2015.

[19] G. Zhang, S. Mei, Y. Wang, H. Han, Y. Feng, and Q. Du, "Spectral variability-aware cascaded autoencoder for hyperspectral unmixing," IEEE Transactions on Geoscience and Remote Sensing, 2025.

[20] F. Cao, Y. Situ, and H. Ye, "A joint multi-scale graph attention and classify-driven autoencoder framework for hyperspectral unmixing," IEEE Transactions on Geoscience and Remote Sensing, 2025.

[21] J. Xu, M. Xu, S. Liu, H. Sheng, and Z. Yang, "Temperature scaling unmixing framework based on convolutional autoencoder," International Journal of Applied Earth Observation and Geoinformation, vol. 129, p. 103864, 2024.

[22] Z. Wang, J. Xu, G. Wei, J. Wang, and Y. Yan, "Multi-scale spectral-spatial unmixing network with boltzmann-inspired adaptive tempera-

ture," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2025.

[23] B. Rasti, B. Koirala, P. Scheunders, and J. Chanussot, "MiSiCNet: Minimum simplex convolutional network for deep hyperspectral unmixing," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–15, 2022.

[24] X. Tao, B. Koirala, A. Plaza, and P. Scheunders, "A new dual-feature fusion network for enhanced hyperspectral unmixing," IEEE Transactions on Geoscience and Remote Sensing, 2024.

[25] M. Xu, J. Xu, S. Liu, H. Sheng, B. Shen, and K. Hou, "Stationary wavelet convolutional network with generative feature learning for hyperspectral unmixing," IEEE Transactions on Geoscience and Remote Sensing, 2024.

[26] J. W. Boardman, F. A. Kruse, and R. O. Green, "Mapping target signatures via partial unmixing of AVIRIS data," in Summaries of the fifth annual JPL airborne earth science workshop. Volume 1: AVIRIS workshop, 1995.

[27] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," in Imaging Spectrometry V, vol. 3753. SPIE, 1999, pp. 266–275.

[28] J. M. Nascimento and J. M. Dias, "Vertex component analysis: A fast algorithm to unmix hyperspectral data," IEEE transactions on Geoscience and Remote Sensing, vol. 43, no. 4, pp. 898–910, 2005.

[29] C.-I. Chang and A. Plaza, "A fast iterative algorithm for implementation of pixel purity index," IEEE Geoscience and Remote Sensing Letters, vol. 3, no. 1, pp. 63–67, 2006.

[30] A. Ambikapathi, T.-H. Chan, C.-Y. Chi, and K. Keizer, "Hyperspectral data geometry-based estimation of number of endmembers using p-norm-based pure pixel identification algorithm," IEEE Transactions on Geoscience and Remote Sensing, vol. 51, no. 5, pp. 2753–2769, 2012.

[31] C.-I. Chang, C.-C. Wu, and H.-M. Chen, "Random pixel purity index," IEEE Geoscience and Remote Sensing Letters, vol. 7, no. 2, pp. 324–328, 2009.

[32] T.-H. Chan, C.-Y. Chi, Y.-M. Huang, and W.-K. Ma, "A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," IEEE Transactions on Signal Processing, vol. 57, no. 11, pp. 4418–4432, 2009.

[33] L. Miao and H. Qi, "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," IEEE Transactions on Geoscience and Remote Sensing, vol. 45, no. 3, pp. 765–777, 2007.

[34] J. M. Bioucas-Dias, "A variable splitting augmented Lagrangian approach to linear spectral unmixing," in 2009 First workshop on hyperspectral image and signal processing: Evolution in remote sensing. IEEE, 2009, pp. 1–4.

[35] J. Li and J. M. Bioucas-Dias, "Minimum volume simplex analysis: A fast algorithm to unmix hyperspectral data," in IGARSS 2008-2008 IEEE International Geoscience and Remote Sensing Symposium, vol. 3. IEEE, 2008, pp. III–250.

[36] C.-H. Lin, C.-Y. Chi, Y.-H. Wang, and T.-H. Chan, "A fast hyperplane-based minimum-volume enclosing simplex algorithm for blind hyperspectral unmixing," IEEE Transactions on Signal Processing, vol. 64, no. 8, pp. 1946–1961, 2015.

[37] B. Jang and A. Hero, "Minimum volume topic modeling," in The 22nd International Conference on Artificial Intelligence and Statistics. PMLR, 2019, pp. 3013–3021.

[38] C.-H. Lin, W.-K. Ma, W.-C. Li, C.-Y. Chi, and A. Ambikapathi, "Identifiability of the simplex volume minimization criterion for blind hyperspectral unmixing: The no-pure-pixel case," IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 10, pp. 5530–5546, 2015.

[39] C.-H. Lin and J. M. B. Dias, "Linear spectral unmixing via matrix factorization: Identifiability criteria for sparse abundances," in IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2018, pp. 6155–6158.

[40] ——, "New theory for unmixing ill-conditioned hyperspectral mixtures," in 2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM). IEEE, 2018, pp. 430–434.

[41] C.-H. Lin and J. M. Bioucas-Dias, "Nonnegative blind source separation for ill-conditioned mixtures via john ellipsoid," IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 5, pp. 2209–2223, 2020.

[42] L. Drumetz, J. Chanussot, C. Jutten, W.-K. Ma, and A. Iwasaki, "Spectral variability aware blind hyperspectral image unmixing based on convex geometry," IEEE Transactions on Image Processing, vol. 29, pp. 4568–4582, 2020.

[43] H. Zou, T. Hastie, and R. Tibshirani, "On the "degrees of freedom" of the lasso," Annals of Statistics, vol. 35, no. 5, pp. 2173–2192, October 2007.

[44] C. L. Lawson and R. J. Hanson, Solving least squares problems. SIAM, 1995.

[45] D. C. Heinz et al., "Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery," IEEE transactions on geoscience and remote sensing, vol. 39, no. 3, pp. 529–545, 2001.

[46] L. Zhuang, C.-H. Lin, M. A. Figueiredo, and J. M. Bioucas-Dias, "Regularization parameter selection in minimum volume hyperspectral unmixing," IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 12, pp. 9858–9877, 2019.

[47] Grupo de Inteligencia Computacional, Universidad del País Vasco / Euskal Herriko Unibertsitatea (UPV/EHU), Spain, "Hyperspectral imagery synthesis (EIAs) toolbox," http://www.ehu.es/ccwintco/index.php/Hyperspectral_Imagery_Synthesis_tools_for_MA 2023.

[48] H. K. Aggarwal and A. Majumdar, "Hyperspectral unmixing in the presence of mixed noise using joint-sparsity and total variation," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 9, no. 9, pp. 4257–4266, 2016.

[49] M. Zhao, L. Yan, and J. Chen, "LSTM-DNN based autoencoder network for nonlinear hyperspectral image unmixing," IEEE Journal of Selected Topics in Signal Processing, vol. 15, no. 2, pp. 295–309, 2021.

[50] M. Zhao, S. Shi, J. Chen, and N. Dobigeon, "A 3-D-CNN framework for hyperspectral unmixing with spectral variability," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–14, 2022.

[51] F. Zhu, "Hyperspectral unmixing: ground truth labeling, datasets, benchmark performances and survey," arXiv preprint arXiv:1708.05125, 2017.

[52] S. Jia and Y. Qian, "Spectral and spatial complexity-based hyperspectral unmixing," IEEE Transactions on Geoscience and Remote Sensing, vol. 45, no. 12, pp. 3867–3879, 2007.

[53] P. Ghosh, S. K. Roy, B. Koirala, B. Rasti, and P. Scheunders, "Hyperspectral unmixing using transformer network," IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1–16, 2022.

[54] J. M. Amigo and C. Santos, "Preprocessing of hyperspectral and multispectral images," in Data handling in science and technology. Elsevier, 2019, vol. 32, pp. 37–53.

[55] L. Guanter, K. Segl, and H. Kaufmann, "Simulation of optical remote-sensing scenes with application to the enmap hyperspectral mission," IEEE Transactions on Geoscience and Remote Sensing, vol. 47, no. 7, pp. 2340–2351, 2009.