



Article A Probabilistic Weighted Archetypal Analysis Method with Earth Mover's Distance for Endmember Extraction from Hyperspectral Imagery

Weiwei Sun ^{1,2,*}, Dianfa Zhang ¹, Yan Xu ³, Long Tian ³, Gang Yang ¹ and Weiyue Li ^{4,*}

- ¹ Department of Geography and Spatial Information Techniques, Ningbo University, Ningbo 315211, China; zhangdianfa@nbu.edu.cn (D.Z.); yanggang@nbu.edu.cn (G.Y.)
- ² State Key Lab of Information Engineering on Survey, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
- ³ Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS 39762, USA; yx131@msstate.edu (Y.X.); lt766@msstate.edu (L.T.)
- ⁴ Institute of Urban Studies, Shanghai Normal University, Shanghai 200234, China
- * Correspondence: nbsww@outlook.com (W.S.); lwy_326@126.com (W.L.)

Received: 1 April 2017; Accepted: 9 August 2017; Published: 14 August 2017

A Probabilistic Weighted Archetypal Analysis method with Earth Mover's Abstract: Distance (PWAA-EMD) is proposed to extract endmembers from hyperspectral imagery (HSI). The PWAA-EMD first utilizes the EMD dissimilarity matrix to weight the coefficient matrix in the regular Archetypal Analysis (AA). The EMD metric considers manifold structures of spectral signatures in the HSI data and could better quantify the dissimilarity features among pairwise pixels. Second, the PWAA-EMD adopts the Bayesian framework and formulates the improved AA into a probabilistic inference problem by maximizing a joint posterior density. Third, the optimization problem is solved by the iterative multiplicative update scheme, with a careful initialization from the two-stage algorithm and the proper endmembers are finally obtained. The synthetic and real Cuprite Hyperspectral datasets are utilized to verify the performance of PWAA-EMD and five popular methods are implemented to make comparisons. The results show that PWAA-EMD surpasses all the five methods in the average results of spectral angle distance (SAD) and root-mean-square-error (RMSE). Especially, the PWAA-EMD obtains more accurate estimation than AA in almost all the classes of endmembers including two similar ones. Therefore, the PWAA-EMD could be an alternative choice for endmember extraction on the hyperspectral data.

Keywords: probabilistic weighted archetypal analysis; earth mover's distance; endmember extraction; hyperspectral imagery

1. Introduction

Hyperspectral imagery (HSI) processing is a hot topic in the field of remote sensing because the collected data has great potentials in differentiating distinct materials on the earth surface, using its hundreds of narrow spectral bands [1,2]. Particularly, spectral unmixing is a significant step of HSI data processing [3,4], and it correlates closely with realistic applications including land cover mapping [5], mine exploration [6], precision agriculture [7] and marine monitoring [8] and so on. The phenomenon of spectral mixture mainly results from limited spatial resolutions of imaging spectrometer (<30 m) and homogeneous mixture of distinct materials, and it refers to that the observed spectral reflectance at each pixel is physically a spectral mixture of several pure materials or called endmembers [9,10]. Spectral unmixing is to estimate spectral signatures of endmembers and meanwhile to quantify the fractions of each endmember (i.e., abundance) present in the mixed pixels [11,12].

Endmember extraction is a key but preliminary work for spectral unmixing, either for its linear or nonlinear spectral mixture model [3,10]. Proper endmembers bring accurate abundance estimation at each pixel and greatly benefit the spectral unmixing in realistic applications mentioned above and vice versa. Generally, spectral signatures of endmembers can be estimated from two divergent schemes [13,14]: (1) the reference-endmembers are manually measured on the ground or in the library using the field spectrometer, and (2) the image-endmembers are estimated from the HSI data using endmember extraction methods. Unfortunately, because of different collecting conditions (e.g., image sensors, atmospheric effects and scattering conditions) between hyperspectral imaging and the field spectrometer, the spectrum of reference-endmembers usually disagrees with those of image pixels [15]. Accordingly, complicated spectral calibration is mandatorily requiring for spectral matching between reference-endmembers and the image pixels. In contrast, the image-endmembers are directly estimated from the image scene and simpler procedures bring them more popularity in spectral unmixing [16].

Numerous image-endmember extraction methods have been presented in current literatures of remote sensing, and they can be grouped into two main aspects with opposite assumptions [10,17]: (1) the pure pixel scheme; and (2) the non-pure pixel scheme. The pure pixel scheme assumes that at least one pure endmember exists in the image scene and it finds pure pixels that contain only one material at the pixel [18]. It expects the volume of the inflating simplex composed by the HSI data to be as large as possible and then finds vertices of the convex simplex. The benchmark method of pixel purity index (PPI) iteratively projects each spectral vector onto skewers that are defined as a large set of random vectors, and it chooses the extreme pixels with highest accounting scores as the final endmembers [19]. N-FINDER estimates pure endmember signatures that correspond to a set of pixels defining the largest volume by inflating a simplex inside the HSI dataset [20]. The alternative volume maximization (AVMAX) maximizes the volume of the simplex defined by the endmembers with respect to only one endmember at one time [21]. The vertex component analysis (VCA) determines endmembers from the extreme of the projection that has the random direction orthogonal to the subspace spanned by the identified endmember signatures at each iteration [22]. More recent representative methods are the hierarchical clustering based on rank-two nonnegative matrix factorization (H2NMF) [23], recursive nonnegative matrix factorization (RNMF) [24], subspace vertex pursuit (SVP) [25] and self-dictionary multiple measurement vector (SDMMV) [18].

The non-pure pixel scheme presumes that no pure pixels exist in the image scene and it seeks "artificial" pure pixels or "virtual" endmembers for the HSI data. In this paper, we focus our topic in non-pure pixel scheme because the HSI scenarios of non-pure pixels are more realistic and the estimated "virtual" endmembers are more closely associated with physically meaningful spectral signatures of true materials in the spectral library. The non-pure pixel estimation methods can be summarized into three main classes: (1) the geometrical methods, (2) the matrix factorization methods and (3) the statistical methods. The geometrical methods find a "virtual" endmember set that minimizes the volume of the simplex it defines. Representative algorithms are convex analysis-based minimum volume enclosing simplex (MVES) [26] and minimum volume simplex analysis (MVSA) [27]. The matrix factorization methods formulate an optimization problem of blind source decomposition with many additive constraints (e.g., sparsity, low rank or positivity) to simultaneously estimate all the endmembers. Representative algorithms include sparse nonnegative matrix underapproximation (SNMU) [28], multilayer nonnegative matrix factorization (MLNMF) [29], robust nonnegative matrix factorization (RNMF) [30] and constrained nonnegative matrix factorization [12]. The statistical methods transform endmember extraction into a statistical inference problem and aim for the highly mixed hyperspectral image scenarios, with typical examples of independent component analysis (ICA) [31] and Bayesian algorithms such as normal endmember spectral unmixing [32] and the hierarchical Bayesian algorithm [33]. Unfortunately, all the above methods still have their own drawbacks or disadvantages. The MVSA considers vertices of the minimum-volume enclosing simplex of the HSI dataset as good endmembers, but noisy observations in the hyperspectral images would bring about bad proximity of endmembers to their true spectral signatures [34]. The convergence of

matrix factorization methods highly depends on the initialization of endmembers [35]. Moreover, it could not provide clear geometrical or physical explanations for the estimated endmembers [36]. The abundance fractions of ICA are supposed to be sampled from mutually independent sources and it contradicts with the case of the HSI data and could not guarantee good estimations of endmembers [31]. Therefore, significant improvement can be made to explore the endmember extraction problem from the non-pure pixel assumption.

In this paper, inspired by Archetypal Analysis (AA), we present a Probabilistic Weighted Archetypal Analysis with Earth Mover's Distance (PWAA-EMD) to extract "virtual" endmembers for the HSI data. The kernel archetypal analysis (KAA) by Zhao [37] improves from AA and also aims for solving the nonpure endmember extraction problem in hyperspectral images. However, its motivation, object function and optimization scheme are different from our PWAA-EMD. The KAA utilizes nonlinear kernel function (e.g., Gaussian kernels) to replace the linear kernel in the optimization procedure of AA, and implements a projected gradient scheme to minimize its object function. Differently, our PWAA-EMD adds the dissimilarity information matrix from EMD into the AA model to weight the coefficient matrix, imposes the Bayesian framework to formulate the weighted version of AA into a Bayesian inference problem, and implements the iterative multiplicative update scheme to optimize the convex problem and estimate the desired "virtual" endmembers. Compared with other methods, the contributions of PWAA-EMD is in the following two aspects:

- (1) The PWAA-MED incorporates the dissimilarity information among pairwise pixels with the EMD metric to promote the behaviors of AA in selecting different endmembers. The EMD measure considers the manifold structure of the HSI data and it could fully describe spectral variations of all the pixels determined by low-dimensional manifolds of the hyperspectral data.
- (2) The PWAA-EMD adopts the Bayesian framework and formulates the endmember extraction of AA into a probabilistic inference problem. The Bayesian framework could represent spectral variability and accordingly the PWAA-EMD is more suitable for spectral unmixing in the realistic HSI data.

The rest of this paper is arranged as follows. Section 2 briefly reviews the AA method. Section 3 describes the methodology of PWAA-EMD for endmember extraction in the HSI data. Section 4 lists and discusses experimental results on simulated and real hyperspectral images for endmember extraction. Section 5 discusses the experimental results from the above Section 4. Section 5 states conclusions for our paper.

2. Brief Review of Archetypal Analysis

In this section, relevant knowledge of AA will be briefly reviewed. Archetypal Analysis, presented by Cutler and Breiman [38], is an unsupervised feature extraction or dimensionality reduction method designed for machine learning problems. It combines the virtues of clustering and the flexibility of matrix factorization, and aims to find the "extreme" point or archetypes from the principal convex hull of a dataset.

Consider the HSI data as a high-dimensional point set $\mathbf{Y} = {\{\mathbf{y}_i\}}_{i=1}^N \in \mathbf{R}^{D \times N}$, where the pixel \mathbf{y}_i corresponds to a real vector in the *D*-dimensional spectral space, *D* is equal to the number of bands in the image scene and *N* is the number of pixels with $D \ll N$. Consider the desired endmember set as $\mathbf{M} = {\{\mathbf{m}_j\}}_{j=1}^r \in \mathbf{R}^{D \times r}$, where *r* is the defined number of endmembers with 1 < r < N. AA assumes that each pixel in the HSI data is approximately a convex combination or linear mixture of all archetypes, while each endmember or archetypes \mathbf{m}_j is a convex combination or linear mixture of all pixels [36],

$$\mathbf{Y} = \mathbf{YBA} + \mathbf{E}, s.t., \begin{cases} \mathbf{b}_{j} \ge 0 \text{ and } \| \mathbf{b}_{j} \|_{1} = 1, \ j = 1, 2, \cdots, r \\ \mathbf{a}_{i} \ge 0 \text{ and } \| \mathbf{a}_{i} \|_{1} = 1, \ i = 1, 2, \cdots, N \end{cases}$$
(1)

where $\mathbf{B} = {\{\mathbf{b}_j\}}_{j=1}^r \in \mathbf{R}^{N \times r}$ and $\mathbf{A} = {\{\mathbf{a}_i\}}_{i=1}^N \in \mathbf{R}^{r \times N}$ are two unknown coefficient matrices respectively, and $\mathbf{E} \in \mathbf{R}^{D \times N}$ is the noise term resulting from modeling errors and measure noise. The constraints $\mathbf{a}_i \ge 0$ and $\|\mathbf{a}_i\|_1 = 1$ guarantee the convex combination property of each pixel by all endmembers. The constraints $\mathbf{b}_j \ge 0$ and $\|\mathbf{b}_j\|_1 = 1$ guarantee the convex combination property of each pixel by all endmember by all pixels. The solution of coefficient matrices \mathbf{B} and \mathbf{A} in AA can be transformed into the following optimization problem by minimizing a residual sum of squares,

$$\underset{\mathbf{B},\mathbf{A}}{\operatorname{argmin}} \parallel \mathbf{Y} - \mathbf{YBA} \parallel_{F}^{2} s.t., \begin{cases} \mathbf{b}_{j} \ge 0 \text{ and } \parallel \mathbf{b}_{j} \parallel_{1} = 1, \ j = 1, 2, \cdots, r \\ \mathbf{a}_{i} \ge 0 \text{ and } \parallel \mathbf{a}_{i} \parallel_{1} = 1, \ i = 1, 2, \cdots, N \end{cases}$$
(2)

where $\|\cdot\|_F$ denotes the Frobenius norm. Both two coefficient matrices **B** and **A** tend to have sparse entries in each column. The sparse coefficient matrices **A** correspond to abundance fractions of each pixel in its each line, and the abundances can be estimated from $\mathbf{M} = \mathbf{Y}\mathbf{\tilde{B}}$ via the optimized estimation of $\mathbf{\tilde{B}}$. Their geometrical explanations are archetypes of the minimal convex hull containing all the pixels in the HSI data.

3. The PWAA-EMD Model for Endmember Extraction

In this section, the methodology of PWAA-EMD is proposed. Section 3.1 describes the model of PWAA-EMD. Section 3.2 presents the solution for the PWAA-EMD model. Section 3.3 summarizes the procedure of PWAA-EMD for extracting endmembers.

3.1. The Formulation of PWAA-EMD Model

Endmember extraction is to select proper endmembers that represent all true materials in the image scene, and the endmembers should have big divergences from each other in the spectrum signatures [39]. In the AA, the estimated endmembers correspond to vertices of the minimal convex hull simplex enclosing all the pixels. Unfortunately, the simple nonnegativity and sum-to-one constraints in coefficient matrices **B** and **A** do not involve the dissimilarity information among all the pixels. It renders that minimizing the convex hull in the spectral observation space could not guarantee absolutely big divergences among selected endmembers, especially for a relatively large number of archetypes. Figure 1 shows the necessity of adding the dissimilarity information into the regular AA. For a high-dimensional HSI dataset in Figure 1a, Figure 1b plots the 10 archetypes of the HSI dataset from AA. The archetypes 4 and 5 are located close to each other and they might correspond to two pixels in the image scene with similar or even subtle spectrum divergences. Therefore, the dissimilarity information between pairwise pixels would be necessary to guarantee good behaviors of AA in endmember extraction.

Many measures have been proposed to quantify the dissimilarity between pairwise pixels. These measures can be computed very fast and often give good results of dissimilarity information, such as the Euclidean distance, Mahalanobis distance, Kullback-Leibler Divergence, Wasserstein distance [40] and so on. In the HSI data, spectral responses of each ground object change with its spatial geological or environmental conditions (e.g., soil composition, weather, terrain, hydrologic conditions). Unfortunately, the above measures could not take into account potential variations in the spectral signals at each pixel when comparing the spectral signals of pairwise pixels [41,42]. These unmodelled spectral variations may lead to large measure values for changes in the spectral responses of HSI pixels that are usually perceived to be small.



Figure 1. The necessity of adding dissimilarity information between two pixels into the AA.

In contrast, the EMD presented by Monge [43] provides a powerful perspective for investigating the manifold structure of spectral signals [40]. The EMD metric has a great potential to describe spectral variations in different pixels determined by the manifold geometry of HSI data. Earth mover's distance, also called the Monge-Kantorovich mass transportation distance, is a cross-bin distance that mainly resolves the histogram or image matching problems by minimizing the amount of work to transform one distribution to another. Consider a HSI data set as $\mathbf{Y} = {\mathbf{y}_i}_{i=1}^N \in \mathbf{R}^{D \times N}$, where each column \mathbf{y}_i represents the source normalized histogram that corresponds to spectral responses of each pixel, and let spectral vector of another pixel $\mathbf{\hat{y}}_i$ be the target normalized histogram. The EMD distance between two histograms y_i and $\mathbf{\hat{y}}_i$ is formulated as a linear programming problem whose goal is to minimize the total cost in transforming the source spectral signal \mathbf{y}_i to the target one $\mathbf{\hat{y}}_i$, showing as follows:

$$EMD(\mathbf{y}_{i}, \hat{\mathbf{y}}_{i}) = \operatorname{argmin}_{j,k} \sum_{j,k} f_{i}(j,k)d(j,k),$$

s.t., $f_{i}(i,j) \ge 0, \sum_{j} f_{i}(j,k) \le y_{k,i}, \sum_{k} f_{i}(j,k) \le \hat{y}_{j,i}, \sum_{j,k} f_{i}(j,k) = 1$ (3)

where $f_i(j,k)$ and d(j,k) are the flow amount and flow cost between the *j*-th bin of the source histogram \mathbf{y}_i and the *k*-th bin of the target histogram $\mathbf{\hat{y}}_i$ respectively, and the flow cost measures with the L₁ ground distance. The constraint $\sum_{j,k} f_i(j,k) = 1$ guarantees that the total reconstruction weights on the same pixel from all the basis vectors sum to 1 and it results from the normalized histogram of each column vector. The computational complexity of EMD metric scales up to $O(D^3 log D)$, and therefore we adopt a fast and robust algorithm [44] to speed up the computation of EMD dissimilarity matrix. The algorithm improves the regular EMD with a thresholded ground distance $d(j,k) = \min(d(j,k),\epsilon)$, where the defined threshold $\epsilon > 0$. The algorithm reduces the computational complexity of EMD by an order of magnitude. Moreover, it considers noise distributions of spectral signals in the HSI data, and assigns different outliers with the same large ground distance to promote the robustness to noise and outliers.

In Equation (1), the coefficient matrix **B** determines the estimated abundances via $\mathbf{M} = \mathbf{YB}$, and we accordingly would like to impose the dissimilarity information $\mathbf{W} = \{EMD_{i,j}\} \in \mathbf{R}^{N \times N}$ into the matrix **B**. The formulated weighted AA with the EMD dissimilarity information is in the following:

$$\mathbf{Y} = \mathbf{YWBA} + \mathbf{E}, s.t., \begin{cases} \mathbf{b}_{j} \ge 0 \text{ and } \| \mathbf{b}_{j} \|_{1} = 1, j = 1, 2, \cdots, r \\ \mathbf{a}_{i} \ge 0 \text{ and } \| \mathbf{a}_{i} \|_{1} = 1, i = 1, 2, \cdots, N \end{cases}$$
(4)

where **B** and **A** is coefficient matrices respectively, and **E** is the noise term. The dissimilarity matrix **W** could enhance the sparsity of **B** and helps selecting proper endmembers with bigger divergences.

The coefficient matrices **B** and **A** are unknown, and the solution of Equation (4) is a typical blind source separation problem. The Bayesian framework is a powerful technique to solve the blind source separation problem and it builds a probabilistic model to formulate the solution problem in (4) as a Bayesian inference problem [45]. The Bayesian framework could represent spectral variability in HSI data and could better cope with the realistic HSI data unmixing [10]. Moreover, the Bayesian framework could fully incorporate the likelihood and prior distributions of spectrum signals and fraction abundances into the model. Besides, the Bayesian framework could automatically enforce other constraints such as sparsity, nonnegativity or sum-to-one constraint to ensure solutions within physical meaningful ranges and regularize solutions. Setting $\Theta = \mathbf{YW}$, the PWAA-EMD model is formulated to be a joint posterior density function in the Bayesian framework,

$$P(\boldsymbol{\Theta}, \mathbf{B}, \mathbf{A} | \mathbf{Y}) = \frac{P(\mathbf{Y} | \boldsymbol{\Theta}, \mathbf{B}, \mathbf{A}) \cdot P(\boldsymbol{\Theta}) \cdot P(\mathbf{B}) \cdot P(\mathbf{A})}{P(\mathbf{Y})} \propto P(\mathbf{Y} | \boldsymbol{\Theta}, \mathbf{B}, \mathbf{A}) \cdot P(\boldsymbol{\Theta}) \cdot P(\mathbf{B}) \cdot P(\mathbf{A})$$

$$s.t., \begin{cases} \mathbf{b}_{j} \ge 0 \text{ and } \| \mathbf{b}_{j} \|_{1} = 1, j = 1, 2, \cdots, r \\ \mathbf{a}_{i} \ge 0 \text{ and } \| \mathbf{a}_{i} \|_{1} = 1, i = 1, 2, \cdots, N \end{cases}$$
(5)

where $P(\Theta, \mathbf{B}, \mathbf{A} | \mathbf{Y})$ denotes the probability density function of $\Theta, \mathbf{B}, \mathbf{A}$ given \mathbf{Y} . $P(\mathbf{Y} | \Theta, \mathbf{B}, \mathbf{A})$ is the likelihood function depending on the observation model of the HSI data. $P(\mathbf{Y})$ and $P(\Theta)$ are normalized constants. The distributions $P(\mathbf{B})$ and $P(\mathbf{A})$ show prior knowledge about the two parameters and guarantee physical constraints inherent to the hyperspectral observation model.

3.2. The Solution of PWAA-EMD Model

The joint posterior density can be optimized via the famous maximum a posterior estimator, and accordingly the problem (5) can be transformed as

$$\underset{\mathbf{B},\mathbf{A}}{\operatorname{argmin}} - (\log P(\mathbf{Y}|\Theta, \mathbf{B}, \mathbf{A}) + \log P(\mathbf{B}) + \log P(\mathbf{A}))$$

$$\mathbf{B} \ge 0, \text{ and } \| \mathbf{b}_j \|_1 = 1, j = 1, 2, \cdots, r$$

$$\mathbf{A} \ge 0, \text{ and } \| \mathbf{a}_i \|_1 = 1, i = 1, 2, \cdots, N$$
(6)

where $\log(\bullet)$ represents the logarithmic operation. In the probabilistic theory, the nonnegativity and sum-to-one constraints on **B** and **A** are equal to the Dirichlet distributions in **b**_{*j*} and **a**_{*i*}, that is, **b**_{*j*} ~ Dir(1), and **a**_{*i*} ~ Dir(1), where Dir is the Dirichlet distribution. Both terms $\log P(\mathbf{B})$ and $\log P(\mathbf{A})$ are constant in Equation (5). On the other hand, considering the quantum nature of light in imaging spectrometer, the Poisson distribution [46] is adopted to describe the probability density distribution of (**Y**| Θ , **B**, **A**), showing as

$$P(\mathbf{Y}|\mathbf{\Theta}, \mathbf{B}, \mathbf{A}) = \prod_{k=1}^{D} \prod_{i=1}^{N} \frac{(\mathbf{\Theta}BA)_{ki}^{(\mathbf{Y})_{k,i}} \cdot \exp\left[-(\mathbf{\Theta}BA)_{k,i}\right]}{(\mathbf{Y})_{k,i}!}$$
(7)

The Equation (5) of joint posterior density can then be transformed into the following optimization problem

$$\operatorname{argmin}_{\mathbf{B},\mathbf{A}} \left(\sum_{k=1}^{D} \sum_{i=1}^{N} \left[-(\mathbf{Y})_{k,i} \log(\Theta B A)_{k,i} + (\Theta B A)_{ki} \right] \right)$$

$$\mathbf{B} \ge 0, \text{ and } \| \mathbf{b}_{j} \|_{1} = 1, j = 1, 2, \cdots, r$$

$$\mathbf{A} \ge 0, \text{ and } \| \mathbf{a}_{i} \|_{1} = 1, i = 1, 2, \cdots, N$$
(8)

The iterative multiplicative update scheme [47] is implemented to optimize the above equation. The scheme utilizes a suitably strong regularization parameter to relax the equality constraint. At the t

+ 1 iteration, coefficient matrices $\mathbf{A}^{(t+1)}$ and $\mathbf{B}^{(t+1)}$ can be iteratively updated via the following two closed forms:

$$\mathbf{A}^{(t+1)} = \mathbf{A}^{(t)} \bigodot \frac{\nabla_{\mathbf{A}^{(t)}}^{-}}{\nabla_{\mathbf{A}^{(t)}}^{+}}, \nabla_{\mathbf{A}_{j,i}}^{+} = \sum_{k,i} \Theta_{k,i} \mathbf{B}_{i,j}^{(t)} + \lambda, \nabla_{\mathbf{A}_{j,i}}^{-} = \sum_{k} \frac{\mathbf{Y}_{k,i} \sum_{i} (\Theta_{k,i} \mathbf{B}_{i,j}^{(t)})}{\sum_{i,j} (\Theta_{k,i} \mathbf{B}_{i,j}^{(t)} \mathbf{A}_{j,i}^{(t)})} + \frac{\lambda}{\sum_{j} \mathbf{A}_{j,i}^{(t)}}$$
(9)

$$\mathbf{B}^{(t+1)} = \mathbf{B}^{(t)} \bigodot \frac{\nabla_{\mathbf{B}^{(t)}}^{-}}{\nabla_{\mathbf{B}^{(t)}}^{+}}, \nabla_{\mathbf{B}_{i,j}}^{+} = \sum_{k,i} \Theta_{k,i} \mathbf{A}_{j,i}^{(t+1)} + \lambda, \nabla_{\mathbf{B}_{i,j}}^{-} = \sum_{k,i} \frac{\mathbf{Y}_{k,i} \Theta_{k,i} \mathbf{A}_{j,i}^{(t+1)}}{\sum_{i,j} (\Theta_{k,i} \mathbf{B}_{i,j}^{(t)} \mathbf{A}_{j,i}^{(t+1)})} + \frac{\lambda}{\sum_{i} \mathbf{B}_{ij}^{(t)}}$$
(10)

where the operation \odot denotes the Hadamard product, λ is a defined regularization parameter, and the indices $i = 1, 2, \dots, N, j = 1, 2, \dots, r$ and $k = 1, 2, \dots, D$. The above update procedures are repeated until satisfying the convergence condition or the number of iterations exceeds the predefined maximal iteration number *T*. The convergence condition is set as $\|\mathbf{B}^{(t+1)} - \mathbf{B}^{(t)}\|_{\infty} \le \varepsilon$ and $\|\mathbf{A}^{(t+1)} - \mathbf{A}^{(t)}\|_{\infty} \le \varepsilon$ and $\|\mathbf{Y} - \mathbf{\Theta}\mathbf{B}^{(t+1)}\mathbf{A}^{(t+1)}\|_{\infty} \le \varepsilon$, where ε is the defined error tolerance. The estimated endmembers are achieved via $\mathbf{\hat{M}} = \mathbf{YWB}^{(t+1)}$.

Careful initialization of variable **B** is proven to promote the convergence speeds of Equation (8). In this paper, we implement the two-stage algorithm [48] to initialize the endmember set $\mathbf{M}^{(0)}$ and estimate the initial coefficient matrices $\mathbf{A}^{(0)}$ and $\mathbf{B}^{(0)}$. The algorithm has a lower computational complexity of $O(D^2N)$ and it selects *r* columns (i.e., corresponds to *r* pixels) that maximize the volume of the parallelepiped spanned by the selected columns. The algorithm implements with two stages the stochastic stage and the deterministic stage, showing as follows:

- (1) *The stochastic stage:* the algorithm computes the $\mathbf{V}_r^{\mathrm{T}}$ that consists the top *r* right singular vectors of \mathbf{Y} , and selects $O(r \log r)$ random columns of $\mathbf{V}_r^{\mathrm{T}}$. The columns are carefully chosen according to the nonuniform probability distribution that depends on the information in the top-*r* right singular subspace of \mathbf{Y} .
- (2) *The deterministic stage:* the algorithm applies a deterministic column-selection procedure [49] to select exactly *r* columns from the set of $O(r \log r)$ columns of $\mathbf{V}_r^{\mathrm{T}}$ selected from the first stage. The algorithm finally outputs exactly *r* columns of the HSI data and we set it to be the initial endmembers $\mathbf{M}^{(0)}$.

3.3. The Summary of PWAA-EMD Model for Endmember Extraction

The PWAA-EMD improves from AA and adds the EMD dissimilarity information to promote the behavior of AA in selecting endmembers with bigger spectrum divergences. The EMD metric considers manifold structure of the HSI data and it could describe spectral variations of all the pixels determined by the manifold geometry of hyperspectral data. Meanwhile, the PWAA-EMD adopts the Bayesian framework and formulates the endmember extraction into a probabilistic inference problem. The Bayesian framework could represent spectral variability and it is more suitable for spectral unmixing in the realistic HSI data. The solution of PWAA-EMD model is formulated into optimizing a joint posterior density via the maximum a posterior estimator. The iterative multiplicative update scheme is utilized to optimize the object function and the two-stage algorithm is employed to initialize the two coefficient matrices. Figure 2 illustrates the procedure of PWAA-EMD in endmember extraction and it includes the following steps:

- (1) Hyperspectral images are transformed from a data cube into a matrix $\mathbf{Y} \in \mathbf{R}^{D \times N}$, where *D* is the number of bands and *N* is the number of pixels, respectively.
- (2) The dissimilarity information among all pixels are computed with the EMD measure in (3) and the dissimilarity information matrix **W** are obtained.
- (3) The coefficient matrix **B** is weighted by the EMD dissimilarity matrix **W** and the Bayesian framework is then utilized to formulate the model of PWAA-EMD in (5).

(1)

- (4) The solution of PWAA-EMD is transformed into an optimization problem of a joint posterior density via the maximum a posterior estimator in (8). The Poisson distribution is utilized to quantify the prior knowledge of the HSI data from the consideration of quantum nature in hyperspectral imaging.
- (5) The two-stage algorithm is adopted to initialize the two coefficient matrices and the iterative multiplicative update rules in (9) and (10) are implemented to optimize the problem.
- (6) The coefficient matrix $\mathbf{B}^{(t+1)}$ at the stopping iteration is set as the final coefficient matrix and the proper endmembers are finally estimated from $\hat{\mathbf{M}} = \mathbf{YWB}^{(t+1)}$.



Figure 2. The procedure of PWAA-EMD for endmember extraction.

4. Experimental Results

In this section, we will design several groups of experiments on both synthetic and real Cuprite hyperspectral data to investigate the performance of PWAA-EMD in extracting endmembers. To make a holistic comparison, five state-of-the-art endmember extraction methods including ICA [50], MLNMF [29], CONMF [51], RNMF [30] and AA [36] have been implemented and their endmember results are compared against that of PWAA-EMD. The extracted endmembers are evaluated with two popular measures, spectral angle distance (SAD) and root-mean-square-error (RMSE). For the *j*-th pure endmember, the formulation of SAD is defined as follows:

$$SAD_{j} = \arccos\left(\frac{\mathbf{m}_{j}^{\mathrm{T}} \cdot \hat{\mathbf{m}}_{j}}{\|\mathbf{m}_{j}\|_{2} \|\hat{\mathbf{m}}_{j}\|_{2}}\right)$$
(11)

where \mathbf{m}_j and $\hat{\mathbf{m}}_j$ are an estimated endmember and the reference endmember respectively, and $\|\cdot\|_2$ is the norm operation of one endmember vector. For the *j*-th endmember, the formulation of RMSE is defined in the following:

$$\text{RMSE}_j = \sqrt[2]{\frac{1}{D} \|\mathbf{m}_j - \hat{\mathbf{m}}_j\|_2^2}$$
(12)

where *D* is the number of bands.

4.1. The Experiments on the Synthetic Data

In this section, synthetic hyperspectral images are utilized to test the performance of PWAA-EMD. Six endmembers are picked from the spectral library of USGS [52], including asphalt-gds367, brick-gds350, cedar-gds360, particleboard-gds364, plastic-gds394 and woodbeam-gds363. The initial spectrum of six endmembers has 2151 bands, ranging the spectrum wavelength from 0.1 to 150 μ m. After spectral resampling, final 108 bands are used in the experiments. Figure 3a plots spectrum signatures of all the six endmembers, and Figure 3b illustrates the simulated abundance images of the six spectrums, having the image size of 100 \times 100 pixels. The abundance vectors were randomly generated following the uniform Dirichlet distribution and no pure pixels exist in the image. The synthetic images follow the linear mixture model with nonnegative abundances summing to one at each pixel. The particleboard-gds364 and woodbeam-gds363 are similar to each other in spectral signals, and it is difficult to accurately estimate the two different endmembers.



Figure 3. The synthetic hyperspectral data. (**a**) Spectrum plots of six endmembers in the synthetic data (**b**) The six abundance images of the synthetic HSI data.

4.1.1. The Experiment on the Synthetic Data without Gaussian Noise

The experiment is to testify the performance of PWAA-EMD in estimating endmembers from the synthetic data without Gaussian noise. For the PWAA-EMD method, using cross-validation, the threshold ϵ is set as 1; the error tolerance ϵ and the maximum iteration time *T* are manually set as 10^{-5} and 500 respectively. For the MLNMF method, using cross-validation, the sparsity regularization parameters a_0 and τ are set as 0.1 and 25 respectively; the regularization parameter δ_1 for abundance sum-to-one constraint is set as 25; and the maximum number of layers *L* and maximum iteration times of each layers L_{max} are set as 10 and 400 respectively. For the CONMF method, using cross-validation, the maximal iteration T_{max} and the relative construction error δ_2 are set as 200 and 10^{-4} respectively. Using cross-validation, the penalization weight parameter λ_1 and the divergence shape parameter β_1 in the RNMF are set as 0.1 and 1 respectively; and the maximum iteration time T_m and the error tolerance ϵ_1 are set as 500 and 10^{-5} respectively. For the AA method, using cross-validation, the error tolerance threshold ϵ_2 and the maximum iteration time T_t are set as 10^{-7} and 1000 respectively.

Table 1 lists the comparison in the SAD and RMSE between PWAA-EMD and other five methods on the synthetic data without adding Gaussian noise. The CONMF performs the best, and the penultimate are ICA and RNMF. The PWAA-EMD is inferior to CONMF but behaves better than other four methods including ICA, RNMF, MLNMF and AA. The MLNMF and AA have similar performances in SAD and RMSE and they surpass ICA and RNMF.

Criteria	Endmember Extraction Methods						
	ICA	MLNMF	CONMF	RNMF	AA	PWAA-EMD	
SAD RMSE	0.1275 0.1042	0.0374 0.0404	<u>0.0074</u> 0.0063	0.2629 0.1016	0.0349 0.0410	0.0188 0.0137	

Table 1. The contrast in SAD and RMSE of all the six methods on the synthetic data without Gaussian noise.

4.1.2. The Experiment on the Synthetic Data with Gaussian Noise

We further testify the robustness of PWAA-EMD and other five methods in resisting Gaussian noise by changing the levels of signal-noise-ratio (SNR) of the synthetic data. The Gaussian white noise is added to the synthetic data with zero-mean and the SNR is defined as follows:

$$SNR = 10 \log_{10} \frac{E[\mathbf{y}_i^{\mathrm{T}} \mathbf{y}_i]}{E[\mathbf{e}_i^{\mathrm{T}} \mathbf{e}_i]}$$
(13)

where \mathbf{y}_i and \mathbf{e}_i are spectrum observation and noise of pixel *i*, and $E[\cdot]$ denotes the expectation operator. In the experiment, the SNR of the synthetic data changes between 10 dB and 60 dB with a step interval of 10 dB. Parameter configurations of all the involved methods are the same with their counterparts in the previous experiment on the noiseless synthetic data.

Figure 4 shows the plots of SAD and RMSE from all six methods when varying the noise levels of the synthetic data. In the figure, the RNMF and CONMF improves much more greatly in the SAD and RMSE with the rising levels of SNR respectively. Especially, the CONMF promotes the SAD from 0.7765 to 0.0092 and its RMSE from 0.1613 to 0.0065, and it behaves best among all the methods when having the SNR level over 60 dB. In contrast, the other four methods ICA, MLNMF, AA and PWAA-EMD have more robust behaviors in terms of Gaussian noise, although their curves of SAD and RMSE are falling down with the changing SNR levels. Both ICA and RNMF have similar performance with a larger SNR over 20 dB but they both are inferior to MLNMF, AA and PWAA-EMD in all cases of SNR levels. The PWAA-EMD surpasses AA in all SNR levels and it performs better than MLNMF with a SNR over 30 dB. The MLNMF obtains slightly better or comparable performance to AA in all SNR levels and that coincides with observations in the previous experiments.



Figure 4. Comparison of all six methods with different levels of noise in terms of (a) SAD and (b) RMSE.

Moreover, Tables 2 and 3 list the SAD and RMSE results of all the six methods from the synthetic data with the SNR equal to 30 dB. Figure 5 illustrates the comparison between the PWAA-EMD endmembers and their references from the USGS library at the SNR equal to 30 dB. We did not list the comparison results of all the other methods from the consideration of overlarge space. From the tables, the ICA, CONMF and RNMF have worse performance than other three methods. The MLNMF

estimates the best endmembers in particleboard-gds364 and plastic-gds394, but it performs worse than PWAA-EMD in the majority classes of endmember extraction. The PWAA-EMD surpasses AA in all six endmembers, especially the two similar endmembers including particleboard-gds364 and woodbeam-gds363 have smaller divergences from their reference endmembers. Moreover, the PWAA-EMD has smaller average SAD and RMSE among all the six methods.

Table 2. The comparison in SAD results from all six methods on the synthetic data with SNR = 30 dB.

Fndmembers	SAD						
	ICA	MLNMF	CONMF	RNMF	AA	PWAA-EMD	
asphalt-gds367	0.1822	0.0694	0.1770	0.1481	0.0340	<u>0.0174</u>	
brick-gds350	0.2584	0.0391	0.1921	0.1717	0.0423	0.0257	
cedar-gds360	0.3074	0.0200	0.1245	0.1773	0.0412	0.0367	
particleboard-gds364	0.0895	<u>0.0129</u>	0.0355	0.0938	0.0479	0.0310	
plastic-gds394	0.2043	0.0092	0.3111	0.2545	0.0436	0.0287	
woodbeam-gds363	0.0979	0.0547	0.6074	0.1206	0.0376	<u>0.0356</u>	
Average	0.1899	0.0342	0.2413	0.1610	0.0411	0.0292	

Table 3. The comparison in RMSE results from all six methods on the synthetic data with SNR = 30 dB.

Fndmembers	RMSE						
Litumentoers —	ICA	MLNMF	CONMF	RNMF	AA	PWAA-EMD	
asphalt-gds367	0.1786	0.0500	0.1695	0.0502	0.0187	<u>0.0098</u>	
brick-gds350	0.1100	0.0481	0.0741	0.0666	0.0404	0.0207	
cedar-gds360	0.1207	0.0342	0.0626	0.0680	0.0421	<u>0.0313</u>	
particleboard-gds364	0.0919	<u>0.0319</u>	0.0704	0.0666	0.0596	0.0346	
plastic-gds394	0.0898	<u>0.0304</u>	0.1282	0.0988	0.0445	0.0370	
woodbeam-gds363	0.0988	0.0627	0.2066	0.1142	0.0544	<u>0.0335</u>	
Average	0.1150	0.0487	0.1186	0.0774	0.0433	0.0278	



Figure 5. The PWAA-EMD endmembers on the synthetic data with SNR = 30 dB. (a) asphalt-gds367; (b) brick-gds350; (c) cedar-gds360; (d) particleboard-gds364; (e) plastic-gds394; and (f) woodbeam-gds363.

4.2. The Experiments on the Cuprite Data

In this section, we further investigate the performance of endmember extraction on real Cuprite hyperspectral data. The Cuprite HSI dataset is a popular benchmarking dataset for testing endmember extraction methods [11–14,17]. The dataset was collected by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) in 1997, covering the area of Cuprite, Nevada, USA. The original imagery has 224 bands ranging the wavelength from 370 to 2480 nm and 20 m spatial resolutions. The bands (1–3, 33, 97, 104–115, 148–170, and 221–224) were removed from the original band set due to water absorptions and too low SNR, with final 180 bands left for our experiment. Figure 6a shows the realistic spatial distributions of different materials in the Cuprite data. A smaller dataset in Figure 6b is manually selected for our experiment, with the image size of 250×191 pixels.

Referring from the ground truth of different mineral materials in Figure 6a, twelve different classes of minerals exist in the image scene of Figure 6b, including Alunite1, Alunite2, Pyrophyllite, Buddingtonite, Chalcedony, Jarosite, Kaolinite1, Kaolinite2, Montmorillonite, Muscovite1, Muscovite2 and Nontronite. The reference spectral signatures of twelve materials were obtained from the USGS spectral library [52]. After careful spectrum matching with the wavelength of AVIRIS data, Figure 6c plots the reference spectrum curves of twelve different materials in the dataset. Similar spectrum responses exist between Alunite1 and Alunite2, and that increases the difficulty in estimating in different endmembers for the Cuprite data.



Figure 6. The Cuprite HSI data. (a) The Ground truth of different mineral materials in the Cuprite data; (b) the image scene of our experiment dataset; and (c) Spectrum plots of twelve materials in our experiment dataset.

In the experiment, for the PWAA-EMD method, the threshold ϵ is set as 1.05 via cross-validation; the error tolerance ϵ and the maximum iteration time T are manually set as 10^{-7} and 1000 respectively. For the MLNMF method, using cross-validation, the sparsity regularization parameters a_0 and τ are set as 0.15 and 30 respectively, and the regularization parameter δ_1 for abundance sum-to-one constraint is set as 30. Using cross-validation, the maximal iteration T_{max} and the relative construction error δ_2 in the CONMF are set as 300 and 10^{-5} respectively. For the RNMF method, using cross-validation, the penalization weight parameter λ_1 and the divergence shape parameter β_1 are set as 0.15 and 1.5 respectively; and the maximum iteration time T_m and the error tolerance ϵ_1 are set as 1000 and 10^{-6} respectively. The error tolerance threshold ε_2 and the maximum iteration time T_t in AA are set as 10^{-6} and 800 respectively, using cross-validation. Other parameters unmentioned are the same as their counterparts in the previous experiments on the synthetic data.

Tables 4 and 5 list the results of SAD and RMSE from all the six methods on the Cuprite data. Figure 7 plots the comparison in spectrum curves between the PWAA-EMD endmembers and the reference endmembers on the Cuprite data. We did not list those of other five methods with the consideration of overlarge space. The CONMF, ICA and RNMF perform worse than other three methods MLNMF, AA and PWAA-EMD and they have larger SAD and RMSE in almost all classes of mineral materials. The PWAA-EMD endmembers surpasses those of all the other five methods in the majority classes of mineral materials, and its averaged SAD and RMSE are smallest among all the six methods involved in the comparison. The MLNMF behaves better than PWAA-EMD in a few classes such as Chaledony, but its overall performance is inferior to PWAA-EMD and it is slightly worse than AA. When comparing PWAA-EMD against AA, the former estimates more accurate endmembers than the later in almost all classes of mineral materials. Moreover, the PWAA-EMD endmembers has smaller differences from the reference endmembers on the two similar classes Alunite1 and Alunite2 because of its smaller SAD and RMSE.



Figure 7. The PWAA-EMD endmembers on the Cuprite data. (a) Alunite1; (b) Alunite2; (c) Pyrophyllite;
(d) Buddingtonite; (e) Chalcedony; (f) Jarosite; (g) Kaolinite1; (h) Kaolinite2; (i) Montmorillonite;
(j) Muscovite1; (k) Muscovite2; and (l) Nontronite.

Table 4. The comparison in SAD results from all six methods on the Cuprite data.

Endmembers —	Endmember Extraction Methods							
	ICA	MLNMF	CONMF	RNMF	AA	PWAA-EMD		
Alunite1	0.2530	0.1112	0.1370	0.2548	0.1119	<u>0.0651</u>		
Alunite2	0.2357	0.1894	<u>0.1338</u>	0.2994	0.1904	0.1724		
Pyrophyllite	0.0922	0.1007	0.1325	0.1976	0.0802	<u>0.0776</u>		
Buddingtonite	0.1411	0.0845	0.1377	0.2257	0.1112	0.1037		
Chaledony	0.1415	0.0820	0.1401	0.2600	0.1186	0.0937		
Jarosite	0.2249	<u>0.1846</u>	0.2681	0.2596	0.2242	0.2081		
Kaolinite1	0.1858	0.1673	0.2386	0.2912	0.2122	<u>0.1373</u>		
Kaolinite2	0.2397	0.2432	0.3855	0.3286	0.2678	<u>0.2105</u>		
Montmorillonite	0.2180	0.2846	0.4359	0.3411	0.2497	<u>0.1706</u>		
Muscovite1	<u>0.0630</u>	0.1150	0.2923	0.2448	0.0952	0.1231		
Muscovite2	0.1359	0.1893	0.4252	0.3846	0.1668	<u>0.1276</u>		
Nontronite	0.1122	0.2826	0.6233	0.3851	0.1691	<u>0.1024</u>		
Average	0.1702	0.1695	0.2792	0.2894	0.1665	0.1327		

Fndmembers	Endmember Extraction Methods						
	ICA	MLNMF	CONMF	RNMF	AA	PWAA-EMD	
Alunite1	0.1584	0.1282	0.1608	0.1710	0.1383	<u>0.0741</u>	
Alunite2	0.1655	0.1711	0.2657	0.1687	0.1056	<u>0.1021</u>	
Pyrophyllite	0.1056	0.0791	0.1016	0.1255	0.0746	0.0778	
Buddingtonite	0.0697	0.0911	0.1393	0.1551	0.0686	<u>0.0647</u>	
Chaledony	0.1326	<u>0.0776</u>	0.1343	0.1296	0.1282	0.0854	
Jarosite	<u>0.1290</u>	0.1481	0.1311	0.1439	0.1336	0.1346	
Kaolinite1	0.1939	0.1600	0.2320	0.1864	0.1435	<u>0.0946</u>	
Kaolinite2	0.1118	0.1553	0.1540	0.1379	0.1151	<u>0.1031</u>	
Montmorillonite	0.1545	0.1386	0.1798	0.1677	0.1671	0.1057	
Muscovite1	0.0896	<u>0.0723</u>	0.1427	0.1223	0.1420	0.0874	
Muscovite2	0.0949	0.1257	0.0933	0.1009	0.1330	0.0876	
Nontronite	0.0731	0.0895	0.1222	0.1031	0.0726	<u>0.0673</u>	
Average	0.1232	0.1197	0.1547	0.1427	0.1185	0.0904	

Table 5. The comparison in RMSE results from all six methods on the Cuprite data.

5. Discussion

The above experiments testify the performance of PWAA-EMD in extracting endmembers on both the synthetic and Cuprite HSI datasets, and the extracted endmembers are compared against five state-of-the-art methods including ICA, MLNMF, CONMF, RNMF and AA.

Experimental results on the synthetic data show that all the six methods own improving performance with the increasing SNR levels from 10 dB to 60 dB. That explains all the comparison methods are negatively affected from the Gaussian noise. The ICA, MLNMF, AA and PWAA-EMD have relatively smaller improvements in SAD and RMSE with the rising SNR levels, and therefore they are more robust to Gaussian noise than CONMF and RNMF. The CONMF performs worse in the case of low SNR smaller than 40 dB whereas it behaves best of all in the noiseless synthetic data or with a SNR over 60 dB. For a certain SNR level, the ICA and RNMF always perform worse than MLNMF, AA and PWAA-EMD. The PWAA outperforms MLNMF when having a larger SNR level over 30 dB. The PWAA-EMD surpasses AA in all six endmembers including the two similar endmembers of particleboard-gds364 and woodbeam-gds363, regardless of the noiseless or noisy synthetic data.

Experimental results on the Cuprite data also prove the advantage of PWAA-EMD over the AA. The PWAA-EMD could better identify two similar endmembers Alunite1 and Alunite2 than AA. The reason for that is the EMD metric considers the dissimilarity information from the manifold structure of the HSI data and it enhances the sparsity of coefficient matrix **B** with the dissimilarity weights to guarantee selecting more different endmembers. Moreover, the PWAA-EMD surpasses the AA in almost all other endmembers, with a smaller SAD and RMSE. The explanation is that the Bayesian framework considers the spectral variability of the HSI data and improves the performance of AA in selecting more accurate endmembers. The RNMF, ICA and CONMF behave worse than other three methods including AA, MLNMF and PWAA-EMD. The explanation for worse behaviors of ICA is that its assumption of independent and robust statistics of endmembers could not be satisfied in most hyperspectral Images including ours [3]. The reason for worse performance of RNMF we guessed is that the nonlinear mixing assumption in RNMF [30] is not suitable for the synthetic and Cuprite HSI data we implemented. The shifting behaviors of CONMF endmembers from noiseless synthetic data to noisy synthetic and Cuprite datasets again demonstrate its severe sensitivity to larger noise levels in the hyperspectral data [53]. The PWAA-EMD outperforms MLNMF in the majority classes of mineral materials and it has the smallest averaged SAD and RMSE among all the six methods. That coincides with the observations of PWAA-EMD on the synthetic data.

Unfortunately, the PWAA-EMD still has some drawbacks and needs careful investigations in the future work. First, we did not carefully explore the parameter setting of threshold error and

iteration times in the PWAA-EMD. The automatic setting schemes on the two parameters will be carefully studied in our following work. Second, the computational complexity of fast EMD is still too high relative to its peer methods like Euclidean distances and kernel distances. The graphical processing unit (GPU) computing and parallel computing schemes will be adopted to improve the computational efficiency of PWAA-EMD in the future. Third, the size problem of endmember set was not carefully considered in the paper. Some recently proposed schemes such as the statistics of the indegree distribution (IDD) [54] and the collaborative sparsity [55] will be adopted into the method to ameliorate the results of endmember extraction. Finally, more real hyperspectral datasets will be utilized to further verify the performance of the PWAA-EMD and improve it for realistic applications.

6. Conclusions

In the paper, we presented a PWAA-EMD method to investigate the endmember extraction problem on hyperspectral images. The PWAA-EMD improves the AA by imposing a EMD weighted matrix into the coefficient matrix to help selecting different endmembers. Compared with other measures, the EMD metric considers manifold structures of the HSI data and it could better represent spectral variations of all the pixels in the HSI data. Meanwhile, the PWAA-EMD adopts the Bayesian framework and formulates endmember extraction into a probabilistic inference problem via optimizing a joint posterior density function. The iterative multiplicative update scheme is utilized to solve the optimization problem and the two-stage algorithm is utilized to make careful initialization. The synthetic and real Cuprite HSI datasets were utilized to testify the performance of PWAA-EMD, and the results of SAD and RMSE were compared with five state-of-the-art methods including ICA, MLNMF, CONMF, RNMF and AA. The results show that PWAA-EMD outperforms AA in almost all the classes of endmembers including similar ones. Moreover, the PWAA-EMD behaves better than MLNMF, ICA, RNMF and CONMF, having smaller averaged SAD and RMSE.

Acknowledgments: This work was funded by National Natural Science Foundation (41671342, U1609203, 41401389), by the Chinese Postdoctoral Science Foundation (2015M570668, 2016T90732), by Public Projects of Zhejiang Province (2016C33021), and by the K. C. Wong Magna Fund in Ningbo University. The authors would like to thank the editor and referees for their suggestions that improve the paper.

Author Contributions: All coauthors made significant contributions to the paper. Weiwei Sun presented the key idea of the PWAA-EMD method and carried on the majority of comparison experiments. Dianfa Zhang designed the comparison experiments between the PWAA-EMD and the other five methods. Yan Xu and Long Tian helped to revise the paper and made major tune works about the parameters involved in the five comparison methods. Gang Yang helped to design the procedures of experiments in the paper. Weiyue Li provided the background knowledge of hyperspectral processing and helped to revise the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Bioucas-Dias, J.M.; Plaza, A.; Camps-Valls, G.; Scheunders, P.; Nasrabadi, N.; Chanussot, J. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geosci. Remote Sens. Mag.* 2013, 1, 6–36. [CrossRef]
- Chen, C.; Li, W.; Su, H.; Liu, K. Spectral-spatial classification of hyperspectral image based on kernel extreme learning machine. *Remote Sens.* 2014, 6, 5795–5814. [CrossRef]
- 3. Ma, W.-K.; Bioucas-Dias, J.M.; Chan, T.-H.; Gillis, N.; Gader, P.; Plaza, A.J.; Ambikapathi, A.; Chi, C.-Y. A signal processing perspective on hyperspectral unmixing: Insights from remote sensing. *IEEE Signal Process. Mag.* **2014**, *31*, 67–81. [CrossRef]
- 4. Li, C.; Ma, Y.; Mei, X.; Liu, C.; Ma, J. Hyperspectral unmixing with robust collaborative sparse regression. *Remote Sens.* **2016**, *8*, 588. [CrossRef]
- 5. Priem, F.; Canters, F. Synergistic use of lidar and apex hyperspectral data for high-resolution urban land cover mapping. *Remote Sens.* **2016**, *8*, 787. [CrossRef]
- 6. Makki, I.; Younes, R.; Francis, C.; Bianchi, T.; Zucchetti, M. A survey of landmine detection using hyperspectral imaging. *ISPRS J. Photogramm. Remote Sens.* **2017**, 124, 40–53. [CrossRef]

- Zarco-Tejada, P.; González-Dugo, M.; Fereres, E. Seasonal stability of chlorophyll fluorescence quantified from airborne hyperspectral imagery as an indicator of net photosynthesis in the context of precision agriculture. *Remote Sens. Environ.* 2016, 179, 89–103. [CrossRef]
- 8. Simon, A.; Shanmugam, P. Estimation of the spectral diffuse attenuation coefficient of downwelling irradiance in inland and coastal waters from hyperspectral remote sensing data: Validation with experimental data. *Int. J. Appl. Earth Obs. Geoinf.* **2016**, *49*, 117–125. [CrossRef]
- 9. Cerra, D.; Bieniarz, J.; Müller, R.; Storch, T.; Reinartz, P. Restoration of simulated enmap data through sparse spectral unmixing. *Remote Sens.* **2015**, *7*, 13190–13207. [CrossRef]
- 10. Bioucas-Dias, J.M.; Plaza, A.; Dobigeon, N.; Parente, M.; Du, Q.; Gader, P.; Chanussot, J. Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 354–379. [CrossRef]
- 11. Liu, R.; Du, B.; Zhang, L. Hyperspectral unmixing via double abundance characteristics constraints based nmf. *Remote Sens.* **2016**, *8*, 464. [CrossRef]
- 12. Du, B.; Wang, S.; Wang, N.; Zhang, L.; Tao, D.; Zhang, L. Hyperspectral signal unmixing based on constrained non-negative matrix factorization approach. *Neurocomputing* **2016**, *204*, 153–161. [CrossRef]
- Xu, M.; Zhang, L.; Du, B.; Zhang, L.; Fan, Y.; Song, D. A mutation operator accelerated quantum-behaved particle swarm optimization algorithm for hyperspectral endmember extraction. *Remote Sens.* 2017, *9*, 197. [CrossRef]
- Liu, R.; Zhang, L.; Du, B. A novel endmember extraction method for hyperspectral imagery based on quantum-behaved particle swarm optimization. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2017, 99, 1–22. [CrossRef]
- 15. Stagakis, S.; Vanikiotis, T.; Sykioti, O. Estimating forest species abundance through linear unmixing of chris/proba imagery. *ISPRS J. Photogramm. Remote Sens.* **2016**, *119*, 79–89. [CrossRef]
- Sun, W.; Ma, J.; Yang, G.; Du, B.; Zhang, L. A poisson nonnegative matrix factorization method with parameter subspace clustering constraint for endmember extraction in hyperspectral imagery. *ISPRS J. Photogramm. Remote Sens.* 2017, 128, 27–39. [CrossRef]
- 17. Sun, W.; Liu, C.; Sun, Y.; Li, W.; Li, J. Extracting pure endmembers using symmetric sparse representation for hyperspectral imagery. *J. Appl. Remote Sens.* **2016**, *10*, 045023. [CrossRef]
- Fu, X.; Ma, W.-K.; Chan, T.-H.; Bioucas-Dias, J.M. Self-dictionary sparse regression for hyperspectral unmixing: Greedy pursuit and pure pixel search are related. *IEEE J. Sel. Top. Signal Process.* 2015, 9, 1128–1141. [CrossRef]
- 19. Chang, C.-I.; Plaza, A. A fast iterative algorithm for implementation of pixel purity index. *IEEE Geosci. Remote Sens. Lett.* **2006**, *3*, 63–67. [CrossRef]
- Winter, M.E. N-findr: An algorithm for fast autonomous spectral end-member determination in hyperspectral data. In Proceedings of the 1999 SPIE's International Symposium on Optical Science, Engineering, and Instrumentation, International Society for Optics and Photonics, Denver, CO, USA, 18–23 July 1999; pp. 266–275.
- 21. Chan, T.-H.; Ambikapathi, A.; Ma, W.-K.; Chi, C.-Y. Robust affine set fitting and fast simplex volume max-min for hyperspectral endmember extraction. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 3982–3997. [CrossRef]
- 22. Nascimento, J.M.; Dias, J.M. Vertex component analysis: A fast algorithm to unmix hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 898–910. [CrossRef]
- 23. Gillis, N.; Kuang, D.; Park, H. Hierarchical clustering of hyperspectral images using rank-two nonnegative matrix factorization. *IEEE Trans. Geosci. Remote Sens.* 2015, 53, 2066–2078. [CrossRef]
- 24. Gillis, N.; Vavasis, S.A. Fast and robust recursive algorithmsfor separable nonnegative matrix factorization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 698–714. [CrossRef] [PubMed]
- 25. Qu, Q.; Nasrabadi, N.M.; Tran, T.D. Subspace vertex pursuit: A fast and robust near-separable nonnegative matrix factorization method for hyperspectral unmixing. *IEEE J. Sel. Top. Signal Process.* **2015**, *9*, 1142–1155. [CrossRef]
- 26. Chan, T.-H.; Chi, C.-Y.; Huang, Y.-M.; Ma, W.-K. A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing. *IEEE Trans. Signal Process.* **2009**, *57*, 4418–4432. [CrossRef]
- 27. Li, J.; Agathos, A.; Zaharie, D.; Bioucas-Dias, J.M.; Plaza, A.; Li, X. Minimum volume simplex analysis: A fast algorithm for linear hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 5067–5082.

- 28. Gillis, N.; Plemmons, R.J. Sparse nonnegative matrix underapproximation and its application to hyperspectral image analysis. *Linear Algebra Appl.* **2013**, *438*, 3991–4007. [CrossRef]
- 29. Rajabi, R.; Ghassemian, H. Spectral unmixing of hyperspectral imagery using multilayer nmf. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 38–42. [CrossRef]
- 30. Févotte, C.; Dobigeon, N. Nonlinear hyperspectral unmixing with robust nonnegative matrix factorization. *IEEE Trans. Image Process.* **2015**, *24*, 4810–4819. [CrossRef] [PubMed]
- 31. Wang, N.; Du, B.; Zhang, L.; Zhang, L. An abundance characteristic-based independent component analysis for hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 416–428. [CrossRef]
- 32. Zhuang, L.; Zhang, B.; Gao, L.; Li, J.; Plaza, A. Normal endmember spectral unmixing method for hyperspectral imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2015, *8*, 2598–2606. [CrossRef]
- 33. Altmann, Y.; Pereyra, M.; McLaughlin, S. Bayesian nonlinear hyperspectral unmixing with spatial residual component analysis. *IEEE Trans. Comput. Imaging* **2015**, *1*, 174–185. [CrossRef]
- Lin, C.-H.; Chi, C.-Y.; Wang, Y.-H.; Chan, T.-H. A fast hyperplane-based mves algorithm for hyperspectral unmixing. In Proceedings of the 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, Australia, 19–24 April 2015; pp. 1384–1388.
- 35. Sun, L.; Zhao, G.; Du, X. Cur based initialization strategy for non-negative matrix factorization in application to hyperspectral unmixing. *J. Appl. Math. Phys.* **2016**, *4*, 614–617. [CrossRef]
- 36. Mørup, M.; Hansen, L.K. Archetypal analysis for machine learning and data mining. *Neurocomputing* **2012**, *80*, 54–63. [CrossRef]
- 37. Zhao, C.; Zhao, G.; Jia, X. Hyperspectral image unmixing based on fast kernel archetypal analysis. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 331–346. [CrossRef]
- 38. Cutler, A.; Breiman, L. Archetypal analysis. Technometrics 1994, 36, 338–347. [CrossRef]
- Wang, N.; Du, B.; Zhang, L. An endmember dissimilarity constrained non-negative matrix factorization method for hyperspectral unmixing. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 2013, *6*, 554–569. [CrossRef]
- Nakhostin, S.; Courty, N.; Flamary, R.; Corpetti, T. Supervised planetary unmixing with optimal transport. In Proceedings of the WHISPERS 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing, Los Angeles, CA, USA, 21–24 August 2016; pp. 1–6.
- 41. Hegde, C.; Sankaranarayanan, A.C.; Baraniuk, R.G. Learning manifolds in the wild. *J. Mach. Learn. Res.* **2012**, 1, 4.
- 42. Venkatasubramanian, S. Moving heaven and earth: Distances between distributions. *ACM SIGACT News* **2013**, *44*, 56–68. [CrossRef]
- 43. Monge, G. M'emoire sur la theorie des deblais et des remblais. In *Histoire de L'academie Royale des Sciences de Paris, avec les Memoire de Mathematique et de Physique Pour la Meme Annee;* De l'Imprimerie Royale: Paris, France, 1781; pp. 666–704.
- 44. Pele, O.; Werman, M. Fast and robust earth mover's distances. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 27 September–4 October 2009; pp. 460–467.
- 45. Bernardo, J.M.; Smith, A.F. Bayesian Theory; IOP Publishing: Bristol, UK, 2001.
- Ma, L.; Moisan, L.; Yu, J.; Zeng, T. A dictionary learning approach for poisson image deblurring. *IEEE Trans. Med. Imaging* 2013, 32, 1277–1289. [PubMed]
- 47. Seth, S.; Eugster, M.J. Probabilistic archetypal analysis. Mach. Learn. 2016, 102, 85–113. [CrossRef]
- Boutsidis, C.; Mahoney, M.W.; Drineas, P. An improved approximation algorithm for the column subset selection problem. In Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms, New York, NY, USA, 4–6 January 2009; Society for Industrial and Applied Mathematics: Philadelphia, PA, USA, 2009; pp. 968–977.
- 49. Pan, C.-T. On the existence and computation of rank-revealing Lu factorizations. *Linear Algebra Appl.* 2000, 316, 199–222. [CrossRef]
- 50. Wang, J.; Chang, C.-I. Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 1586–1600. [CrossRef]
- 51. Li, J.; Bioucas-Dias, J.M.; Plaza, A. Collaborative nonnegative matrix factorization for remotely sensed hyperspectral unmixing. In Proceedings of the 2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Munich, Germany, 22–27 July 2012; pp. 3078–3081.

- 52. Clark, R.N.; Swayze, G.A.; Wise, R.; Livo, K.E.; Hoefen, T.; Kokaly, R.F.; Sutley, S.J. USGS Digital Spectral Library splib06a; US Geological Survey: Denver, CO, USA, 2007; p. 231.
- 53. Li, J.; Bioucas-Dias, J.M.; Plaza, A.; Liu, L. Robust collaborative nonnegative matrix factorization for hyperspectral unmixing. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6076–6090. [CrossRef]
- 54. Heylen, R.; Parente, M.; Scheunders, P. Estimation of the number of endmembers in a hyperspectral image via the hubness phenomenon. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2191–2200. [CrossRef]
- 55. Jutten, C. Estimating the number of endmembers to use in spectral unmixing of hyperspectral data with collaborative sparsity. In Proceedings of the 13th International Conference Latent Variable Analysis and Signal Separation (LVA/ICA 2017), Grenoble, France, 21–23 February 2017; Springer: Berlin, Germany; pp. 1–10.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).