A robust test for nonlinear mixture detection in hyperspectral images
Yoann Altmann, Nicolas Dobigeon, Jean-Yves Tourneret, José Carlos M. Bermudez

To cite this version:

HAL Id: hal-01150345
https://hal.archives-ouvertes.fr/hal-01150345
Submitted on 11 May 2015
A ROBUST TEST FOR NONLINEAR MIXTURE DETECTION IN HYPERSONTICAL IMAGES

Y. Altmann*, N. Dobigeon and J-Y. Tourneret
University of Toulouse
IRIT-ENSEEIHT
Toulouse, France

J.C.M. Bermudez†
Federal University of Santa Catarina
Florianópolis, SC, Brazil

ABSTRACT
This paper studies a pixel by pixel nonlinearity detector for hyper-spectral image analysis. The reflectances of linearly mixed pixels are assumed to be a linear combination of known pure spectral components (endmembers) contaminated by additive white Gaussian noise. Nonlinear mixing, however, is not restricted to any prescribed nonlinear mixing model. The mixing coefficients (abundances) satisfy the physically motivated sum-to-one and positivity constraints. The proposed detection strategy considers the distance between an observed pixel and the hyperplane spanned by the endmembers to decide whether that pixel satisfies the linear mixing model (null hypothesis) or results from a more general nonlinear mixture (alternative hypothesis). The distribution of this distance is derived under the two hypotheses. Closed-form expressions are then obtained for the probabilities of false alarm and detection as functions of the test threshold. The proposed detector is compared to another nonlinearity detector recently investigated in the literature through simulations using synthetic data. It is also applied to a real hyperspectral image.

Index Terms— Nonlinearity detection, Hyperspectral images, Linear mixing model.

1. INTRODUCTION
Identifying the macroscopic materials present in a scene as well as their proportions in each pixel of this scene is of prime interest when analyzing hyperspectral images. Most spectral unmixing (SU) strategies assume that the pixel reflectances are linear combinations of the endmembers [1–5]. However, as explained in [6], the linear mixing model (LMM) can be inappropriate for some hyperspectral images, such as those containing sand, trees or vegetation areas. Nonlinear mixing models provide an interesting alternative for overcoming the inherent limitations of the LMM. Some nonlinear models have been proposed in the literature to handle specific kinds of nonlinearity [7, 8]. For instance, the bidirectional reflectance-based model studied in [9] has been introduced for intimate mixtures. The bilinear models recently studied in [10–15] mainly focus on scattering effects observed in images with relief such as vegetation or urban areas.

Even if the consideration of nonlinear effects in hyperspectral images can provide more accurate results in terms of endmember and abundance identification, nonlinear models generally require a higher computational complexity than approaches based on the linear and nonlinear models used for hyperspectral image unmixing under the null and alternative hypotheses. Section 3 derives the distribution of the squared distance between the observed pixel and the low-dimensional hyperplane. Hence, we propose to design a statistical test based on the distance between the observed pixel and that hyperplane for deciding between the null and alternative hypotheses. As in [8], we assume that the endmembers contained in the image have been estimated by a geometric endmember extraction algorithm (EEA).

This paper is organized as follows. Section 2 introduces the linear and nonlinear models used for hyperspectral image unmixing under the null and alternative hypotheses. Section 3 derives the distribution of the squared distance between the observed pixel and the low-dimensional subspace containing linear mixtures of the endmembers. Section 4 derives two statistical tests for nonlinearity detection depending on the prior knowledge about the noise variance (that can be either known or unknown). Some simulation results conducted on synthetic and real data are shown in Section 5. Conclusions and future works are finally reported in Section 6.

2. MIXING MODELS
Let \( y \) be the \( L \times 1 \) pixel observed at \( L \) different spectral bands. The LMM assumes that \( y \) results from a mixture of \( R \) known endmembers \( m_1, \ldots, m_R \) as follows
\[
y = M\alpha + \epsilon
\] (1)
where \( M = [m_1, \ldots, m_R] \) is the \( L \times R \) endmember matrix, the elements \( \alpha_i \) of \( \alpha = [\alpha_1, \ldots, \alpha_R] \) are the proportions of each endmember in the mixture and \( \epsilon \) is an \( L \times 1 \) independent white Gaussian vector such that \( \epsilon \sim \mathcal{N}(0_L, \sigma^2 I_L) \). The elements of the abundance vector \( \epsilon \) follow a zero-mean Gaussian distribution with variance \( \sigma^2 \).

†Part of this work has been supported by Direction Generale de l’Armement, French Ministry of Defence, and by the Hypanema ANR Project ANR Project n° ANR-12-BS03-003.

*This work has been supported in part by Capes - Proc. 3367-11-8.
vector $a$ satisfy the physical positivity and sum-to-one constraints
\[ \sum_{r=1}^{R} a_r = 1, \quad a_r \geq 0, \forall r \in \{1, \ldots, R\}. \] (2)

Consider now the hyperplane $\mathcal{H}$ defined by
\[ \mathcal{H} : \{ z \mid z = Ma, \sum_{r=1}^{R} a_r = 1 \}. \] (3)

In the noise-free case, $\mathcal{H}$ lies in an $(R - 1)$-dimensional subspace embedding all observations distributed according to the LMM.

We also consider a general nonlinear mixing model as follows
\[ y = Ma + \mu + e \] (4)
where $\mu$ is an $L \times 1$ deterministic vector that does not belong to $\mathcal{H}$, i.e., $\mu \notin \mathcal{H}$ and $a$ satisfies the constraints (2). Note that $\mu$ can be a nonlinear function of the endmember matrix $M$ and/or the abundance vector $a$ and should be denoted as $\mu(M, a)$. However, the arguments $M$ and $a$ are omitted in this paper for brevity.

Given an observation vector $y$, we formulate the detection of nonlinear mixtures as the following binary hypothesis testing problem:
\[ \begin{align*}
H_0 &: \quad y \text{ is distributed according to } (1) \\
H_1 &: \quad y \text{ is distributed according to } (4).
\end{align*} \] (5)

Using the statistical properties of the noise $e$, we obtain $E[y|H_0] = Ma \in \mathcal{H}$ whereas $E[y|H_1] = Ma + \mu \notin \mathcal{H}$. As a consequence, it makes sense to consider the squared Euclidean distance
\[ \delta^2(y) = \min_{z \in \mathcal{H}} \| y - z \|^2 \] (6)
between the observed pixel $y$ and the hyperplane $\mathcal{H}$ to decide which hypothesis ($H_0$ or $H_1$) is true. The next section studies the distribution of $\delta^2(y)$ under the two hypotheses $H_0$ and $H_1$.

3. DISTRIBUTIONS OF THE SQUARE DISTANCE UNDER HYPOTHESES $H_0$ AND $H_1$

We now design a statistical test for hypotheses $H_0$ and $H_1$ by studying the distribution of $\delta^2(y)$ under each hypothesis.

Under $H_1(\mu \neq 0)$ or $H_0(\mu = 0)$, given the sum-to-one constraint on the abundance vector, the mixing model (4) can be rewritten as
\[ \hat{y} = y - m_R = K \epsilon + \mu + e \] (7)
where $\epsilon = [a_1, \ldots, a_{R-1}]^T \in (R - 1) \times 1$ and $K = [m_1 - m_R, \ldots, m_{R-1} - m_R] \in L \times (R - 1)$. Hence, $\delta^2(y)$ can be computed by solving the unconstrained least squares (LS) problem
\[ \hat{\delta}^2(y) = \min_{\epsilon \in \mathbb{R}^{R-1}} \| \hat{y} - Ke \|^2. \] (8)

It is well known that the solution to this problem is given by
\[ \hat{\epsilon}_{LS} = \left(K^T K \right)^{-1} K^T \hat{y} \] (9)
yielding the following $L \times 1$ residual vector $\hat{e}$
\[ \begin{align*}
\hat{e} &= \hat{y} - K \hat{\epsilon}_{LS} \\
&= \left[I_L - K \left(K^T K \right)^{-1} K^T \right] \hat{y} \\
&= H(\mu + e)
\end{align*} \] (10)
where $H = I_L - K (K^T K)^{-1} K^T$ is an $L \times L$ projection matrix of rank $K = L - R + 1$. Using $\hat{\delta}^2(y) = \hat{\epsilon}^T \hat{\epsilon}$ and $\hat{\epsilon} \sim \mathcal{N}(H \mu, \sigma^2 H)$, straightforward computations lead to [16]
\[ \frac{1}{\sigma^2} \delta^2(y) | H_1 \sim \chi^2_{K} \left( \frac{\mu^T H \mu}{\sigma^2} \right) \] (11)
where $\chi^2_{K}(\lambda)$ denotes the noncentral $\chi^2$ distribution with $K$ degrees of freedom and noncentrality parameter $\lambda$. The distribution of $\delta^2(y)$ under $H_0$ can be obtained by setting $\mu = 0$ in (11), yielding
\[ \frac{1}{\sigma^2} \delta^2(y) | H_0 \sim \chi^2_{K}(0) = \chi^2_{K} \] (12)
where $\chi^2_{K}$ is the $\chi^2$ distribution with $K$ degrees of freedom. Notice that the distributions of $\delta^2(y)$ under $H_0$ and $H_1$ depend on the known matrix $M$ but also on the usually unknown noise variance $\sigma^2$ and nonlinearity vector $\mu$. In the following we study nonlinearity detectors constructed for known and unknown noise variance $\sigma^2$.

4. NONLINEARITY DETECTION

As mentioned above, the distributions of $\delta^2(y)$ under $H_0$ and $H_1$ depend on $\sigma^2$ and on $\mu$, which are usually unknown. We address first the scenario in which the noise variance is known but the nonlinearity under $H_1$ is unknown.

4.1. Known $\sigma^2$, unknown $\mu$

For known $\sigma^2$, the distribution of $\delta^2(y)$ is perfectly known under $H_1$ and partially known under $H_0$. Thus, we use a statistical test that does not depend on the unknown nonlinearity $\mu$ to decide between $H_0$ and $H_1$. Here we propose to use the following statistical test
\[ T = \frac{\delta^2(y) - H_1}{\sigma^2} \geq \eta \] (13)
where $\eta$ is a threshold related to the probability of false alarm (PFA) of the test
\[ P_{FA} = P \left[T > \eta | H_0\right] \] (14)
or equivalently,
\[ \eta = F^{-1}_{\chi^2_{K}}(1 - P_{FA}) \] (15)
where $F^{-1}_{\chi^2_{K}}$ is the inverse cumulative distribution function of the $\chi^2_{K}$-distribution. For a given $\mu$, the power $P_{Det}(\mu)$ of the test is
\[ P_{Det}(\mu) = 1 - F^{-1}_{\chi^2_{K}(\lambda)}(\eta) \] (16)
where $F^{-1}_{\chi^2_{K}(\lambda)}(\eta)$ is the cumulative distribution function of the $\chi^2_{K}(\lambda)$-distribution and $\lambda = \sigma^{-2} \mu^T H \mu$. Notice that the probability of detection (PD) $P_{Det}(\mu)$ is an increasing function of $\lambda$ for a fixed threshold $\eta$. This makes sense, as the higher the power $\mu^T H \mu$ of the nonlinearity orthogonal to $\mathcal{H}$, the better the detection performance. Moreover, the lower the noise variance, the better the nonlinearity detection. Unfortunately, the noise variance is unknown in most practical applications. In these cases, the test (13) cannot be used. We study a nonlinearity detector for unknown $\sigma^2$ in the next section.
4.2. Unknown $\sigma^2$, unknown $\mu$

A simple solution if $\sigma^2$ is unknown is to replace the actual noise variance in (13) by its estimate $\hat{\sigma}^2$. This yields the following test

$$T' = \frac{\delta^2(y)}{\hat{\sigma}^2} \geq \eta,$$  \hspace{1cm} (17)

where $\eta$ is the threshold defined in (15). The PFA and PD of the test (17) are then given by

$$P_{FA} = \mathbb{P} \left[ T' > \eta \mid H_0 \right] = \mathbb{P} \left[ T > \frac{\delta^2}{\sigma^2} \eta \mid H_0 \right],$$

$$P_{PD}^\ast = \mathbb{P} \left[ T' > \eta \mid H_1 \right] = \mathbb{P} \left[ T > \frac{\delta^2}{\sigma^2} \eta \mid H_1 \right].$$  \hspace{1cm} (18)

The better the estimation of $\sigma^2$, the closer the distributions of $T$ and $T'$ and thus the closer the performances of the tests (13) and (17). Here we propose to estimate $\sigma^2$ through an eigenanalysis of the sample covariance matrix of a set of $N$ pixels assumed to share the same noise variance. The estimate $\hat{\sigma}^2$ is then determined as the average of the $p \leq L$ smallest eigenvalues of the sample covariance matrix. The accuracy of the estimator will depend on the choice of $p$, which will be discussed in the simulation section. For a given $P_{FA}$ and associated threshold $\eta$ given by (14), the distribution of $T'$ is shifted to the left if $\sigma^2$ is overestimated, i.e., $\hat{\sigma}^2 > \sigma^2$. This will lead to a $P_{FA}^\ast$ of the test (17) that is lower than the $P_{FA}$ of the test (13). Conversely, if $\hat{\sigma}^2 < \sigma^2$, the distribution of $T'$ is shifted to the right leading to $P_{FA}^\ast > P_{FA}$. Thus, it seems reasonable to overestimate $\sigma^2$ to ensure $P_{FA}^\ast$ is upper bounded by a $P_{FA}$ fixed by the user. This observation will be used in Section 5 to adjust the value of $p$.

5. SIMULATIONS

5.1. Synthetic data: known $\sigma^2$, unknown $\mu$

We first investigate the performance of the test (13), which assumes $\sigma^2$ known. We consider a mixture of $R = 3$ materials (green grass, olive green paint and galvanized steel metal) whose spectral signatures $m_i$ composed of $L = 826$ bands have been extracted from the spectral libraries provided with the ENVI software [17]. The abundance vector is fixed to $\alpha = [0.3, 0.6, 0.1]^T$ and the noise variance to $\sigma^2 = 10^{-3}$. The nonlinearity $\mu$ is set as follows

$$\mu = \nu \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} a_{ij} m_i \odot m_j,$$  \hspace{1cm} (19)

where $\nu$ is a scaling factor selected from the set {0.4, 0.5, 0.6, 0.7}. This nonlinearity corresponds to the generalized bilinear model (GBM) studied in [13] with $\gamma_{ij} = \nu$ for all $(i, j)$. The specific form in (19) was chosen so that the impact of the nonlinearity is governed by a single parameter $\nu$. Fig. 1 shows the theoretical and empirical (N = 20000 noise realizations) operating characteristics (ROCs) [18, p. 74-75] for the test (13). Each value of $\nu$ corresponds to a different noncentrality parameter $\lambda$ for the noncentral $F$ distribution, ranging from $\lambda = 49$ to $\lambda = 150$. These results confirm that the performance of test (13) improves for larger values of $\lambda$ (or $\nu$).

5.2. Synthetic data: unknown $\sigma^2$, unknown $\mu$

We now study the performance of the nonlinearity detector when $\sigma^2$ is replaced by its estimate $\hat{\sigma}^2$. Fig. 2 shows the ROCs of test (17) for $\lambda = 70$ and for three values of $\hat{\sigma}^2$: 0.95$\sigma^2$, $\sigma^2$ and 1.05$\sigma^2$. Note that all ROCs coincide. However, different estimates $\hat{\sigma}^2$ correspond to different points on the curve for a fixed PFA. For instance, if the PFA is fixed to $P_{FA} = 0.1$ and the noise variance is correctly estimated, the corresponding PFA is around $P_{FA} \approx 0.65$ (see Fig. 2 (middle)). If $\hat{\sigma}^2 = 0.95\sigma^2$, the PFA of test (17) rises to $P_{FA} \approx 0.41$, leading to $P_{FA} \approx 0.92$. Conversely, if $\hat{\sigma}^2 = 1.05\sigma^2$, the PFA of test (17) falls to $P_{FA} \approx 0.01$, leading to $P_{FA} \approx 0.27$.

![Fig. 1. Actual (solid lines) and empirical (diamonds) receiver operating characteristics (ROCs) of the first test (known noise variance) for $\nu = 0.4$ (blue), $\nu = 0.5$ (red), $\nu = 0.6$ (green) and $\nu = 0.7$ (black).](image1)

![Fig. 2. Theoretical ROCs of the test (17) (unknown noise variance) for $\sigma^2 = 0.95\sigma^2$ (left), $\sigma^2 = \sigma^2$ (middle) and $\hat{\sigma}^2 = 1.05\sigma^2$ (right). The point corresponding with $P_{FA} = 0.1$ is the intersection of the black lines.](image2)
using the standard linear mixing model (LMM) whereas the sub-images \( S_2, S_3 \) and \( S_4 \) were generated according to the nonlinear mixing model (20) with \( b = 0.1 \) and \( \theta \in \{ \pi/4, 3\pi/8, \pi/2 \} \). For each sub-image, the abundance vectors \( a_n, n = 1, \ldots, 2500 \), were generated uniformly in the admissible set defined by the positivity and sum-to-one constraints. All sub-images were corrupted by an additive white Gaussian noise of variance \( \sigma^2 = 10^{-3} \) corresponding to an average SNR \( \approx 21\text{dB} \). Table 1 shows the means and standard deviations of the noise variance estimates obtained for different values of \( p \) (for 50 Monte Carlo runs). This table shows that fixing \( p = L - R + 1 \) provides accurate estimates of \( \sigma^2 \) for these examples. Fig. 3 compares the empirical ROCs constructed from the number of pixels detected as linear and nonlinear for the different tests (known and unknown noise variance). The empirical ROCs for the test studied in [8] are also displayed in these figures. Fig. 3 (top left) shows that the three detectors are able to respect the PFA constraint. The three other subfigures of Fig. 3 display the ROCs for the three different values of \( \theta \). For small values of \( \theta \), the norm of the nonlinearity projection onto the vector \( \mu \) is large. Hence, the nonlinearity detector based on the PPNMM studied in [8] outperforms the tests studied in this paper (top right subfigure). However, the performance of test in [8] degrades as the portion of the nonlinearity that is orthogonal to \( \mu \) becomes predominant (bottom subfigures). The two proposed tests (known and unknown noise variance) perform similarly. Moreover, these two tests seem to be more robust to the type of nonlinearity. Finally, the proposed tests (13) and (17) only require one projection (10) of each pixel (and eventually the noise variance estimation procedure) while the test studied in [8] requires the minimization of a more complex cost function and the derivation of Cramér-Rao bounds, leading to higher computation costs when compared to the proposed method.

<table>
<thead>
<tr>
<th>( L - p )</th>
<th>( S_1 )</th>
<th>( S_2 )</th>
<th>( S_3 )</th>
<th>( S_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19.94 (±0.27)</td>
<td>21.32 (±0.35)</td>
<td>20.69 (±0.30)</td>
<td>19.96 (±0.27)</td>
</tr>
<tr>
<td>2</td>
<td>9.99 (±0.01)</td>
<td>9.98 (±0.01)</td>
<td>9.99 (±0.01)</td>
<td>9.99 (±0.01)</td>
</tr>
<tr>
<td>3</td>
<td>9.97 (±0.01)</td>
<td>9.97 (±0.01)</td>
<td>9.97 (±0.01)</td>
<td>9.97 (±0.01)</td>
</tr>
<tr>
<td>4</td>
<td>9.95 (±0.01)</td>
<td>9.95 (±0.01)</td>
<td>9.96 (±0.01)</td>
<td>9.96 (±0.01)</td>
</tr>
<tr>
<td>5</td>
<td>9.94 (±0.01)</td>
<td>9.94 (±0.01)</td>
<td>9.94 (±0.01)</td>
<td>9.94 (±0.01)</td>
</tr>
</tbody>
</table>

Table 1. Means and standard deviations (in brackets) of the estimated noise variance \( (\times10^{-3}) \) for different values of \( p \) \( (\sigma^2 = 10^{-3}) \).

5.3. Real data

The real image considered in this section is composed of spectral bands and was acquired in 1997 by the satellite AVIRIS over the Moffett Field, CA. A subimage of size \( 50 \times 50 \) pixels (which we considered in previous work) was chosen to evaluate the proposed nonlinearity detector. The scene is mainly composed of water, vegetation, and soil. The corresponding endmembers were estimated by the vertex component analysis (VCA) [19] with \( R = 3 \). Fig. 4 shows an example of detection map (for \( P_{PFA} = 10^{-5} \)) obtained with the proposed detector. It can be seen that nonlinear mixtures are mainly detected in the coastal area and in regions where mixtures of vegetation and soil occur (as also observed in [13]).

6. CONCLUSIONS

This paper presented a new pixel by pixel nonlinearity detector for hyperspectral image analysis when the noise variance is known. The proposed detector is based on the distance between the observed pixel and the low dimensional subspace spanned by the endmembers. For unknown noise variance, a similar detector was proposed by replacing the actual noise variance by an accurate estimator based on eigenvalues resulting from the eigenanalysis of the sample covariance matrix of a set of image pixels. The main advantages of the proposed method are the absence of prior knowledge about the type of the nonlinearity and its low computational cost. Simulations on synthetic data illustrated the robustness of the method to detect various nonlinearities. The results obtained with a real AVIRIS hyperspectral image are also encouraging. Future works include the detection of nonlinearity in the case of unknown (or partially unknown) endmembers.

7. RELATION TO PRIOR WORK

As mentioned above, one of the main properties of the proposed nonlinearity detection method is that it does not rely on prior knowledge about the type of the nonlinearity used for spectral unmixing. Conversely, the pixel by pixel nonlinearity detector presented in [8] assumed a polynomial form for the nonlinearity with respect to the endmembers. We showed in this paper that the proposed detector has clear advantages when the nonlinear mixture of the endmembers cannot be approximated by a polynomial.
8. REFERENCES


