DETECTION OF NONLINEAR MIXTURES USING GAUSSIAN PROCESSES: APPLICATION TO HYPERSPECTRAL IMAGING

T. Imbiriba* J. C. M. Bermudez* J.-Y. Tourneret†
C. Richard‡[1]

* Federal University of Santa Catarina, Florianópolis, SC, Brazil
†University of Toulouse, IRIT-ENSEEIHT, CNRS, Toulouse, France
‡Université de Nice Sophia-Antipolis, CNRS, Nice, France

May 2014

[1]This work was partly supported by CNPq under grants Nos 305377/2009-4, 400566/2013-3 and 141094/2012-5, and by the Agence Nationale pour la Recherche, France, (Hypanema project, ANR-12- BS03-003), and by ANR-11-LABX-0040-CIMI within the program ANR-11-IDEX-0002-02.

Hyperspectral Images

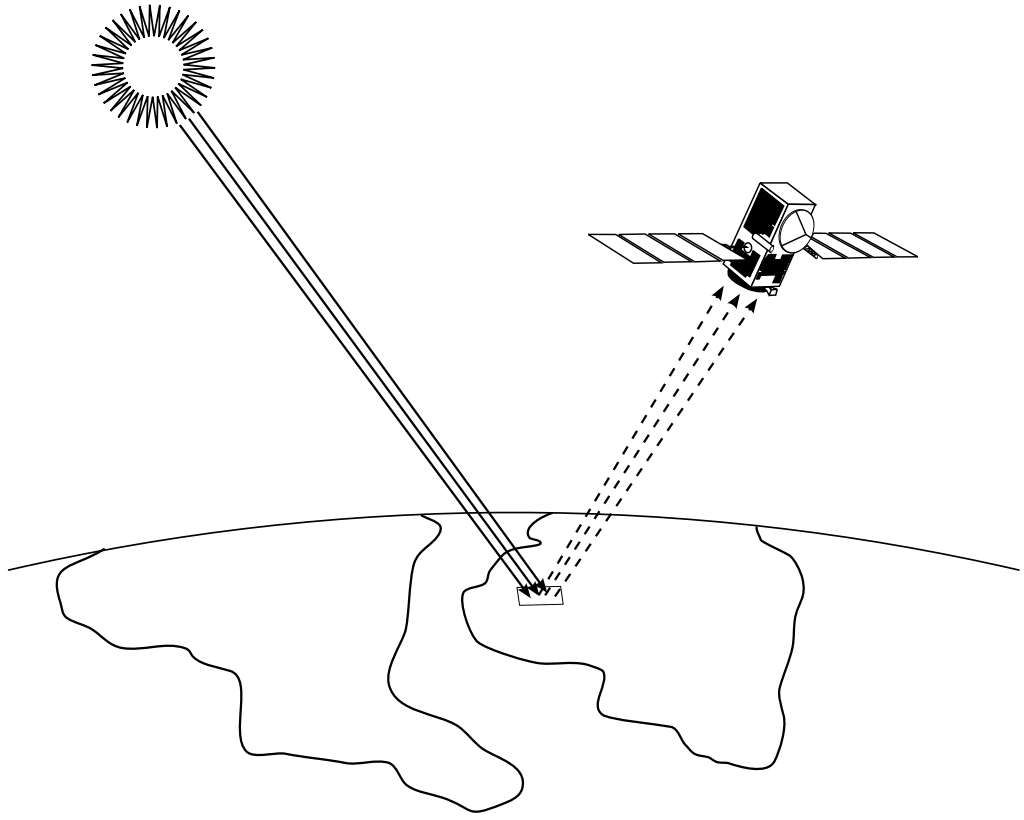


Figure: Remote Sensing: Sun's radiation reflects on the Earth's surface and is captured by an airborne or spaceborne hyperspectral sensor.

- High spectral resolution $\times$ poor spacial resolution
- One hyperspectral pixel $\boldsymbol{y}$ ➪ Hundreds of contiguous bands.

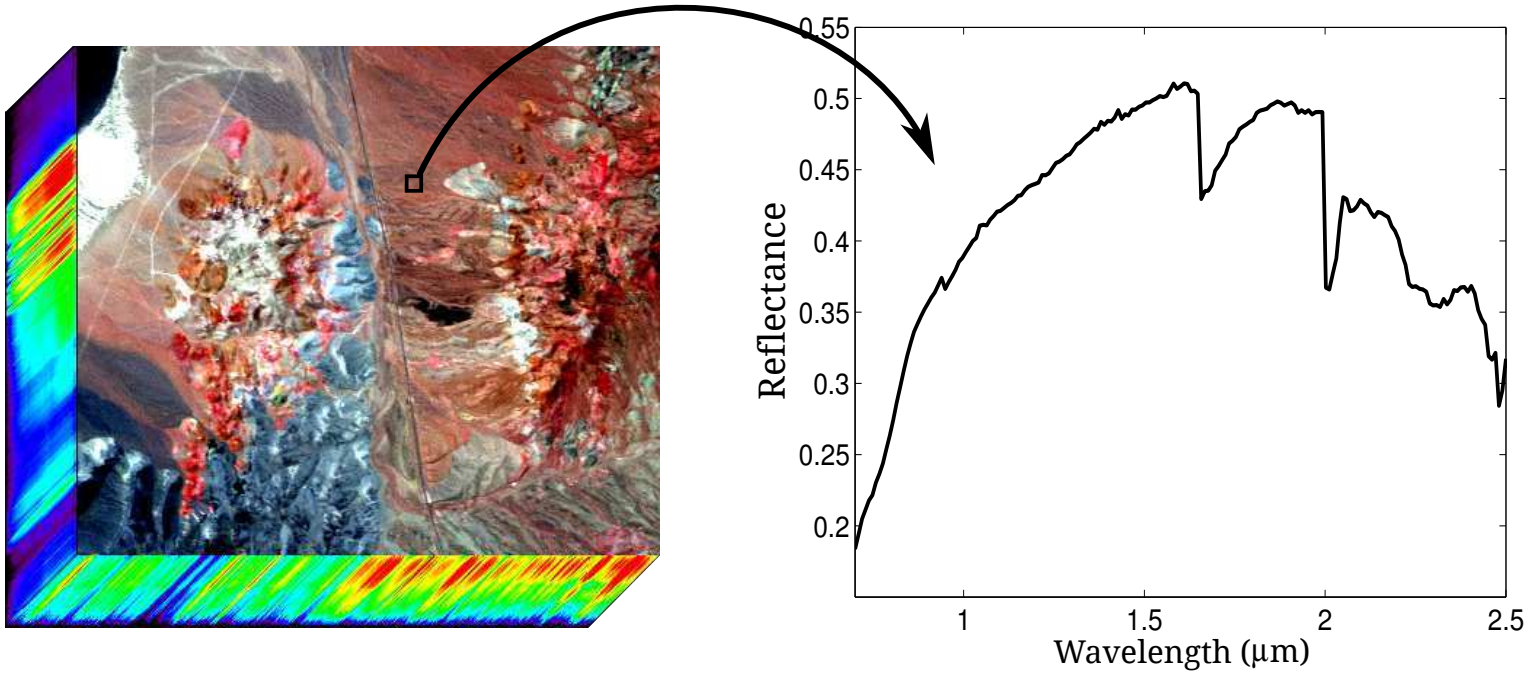


Figure: Hypercube captured by the AVIRIS from the Cuprite field.

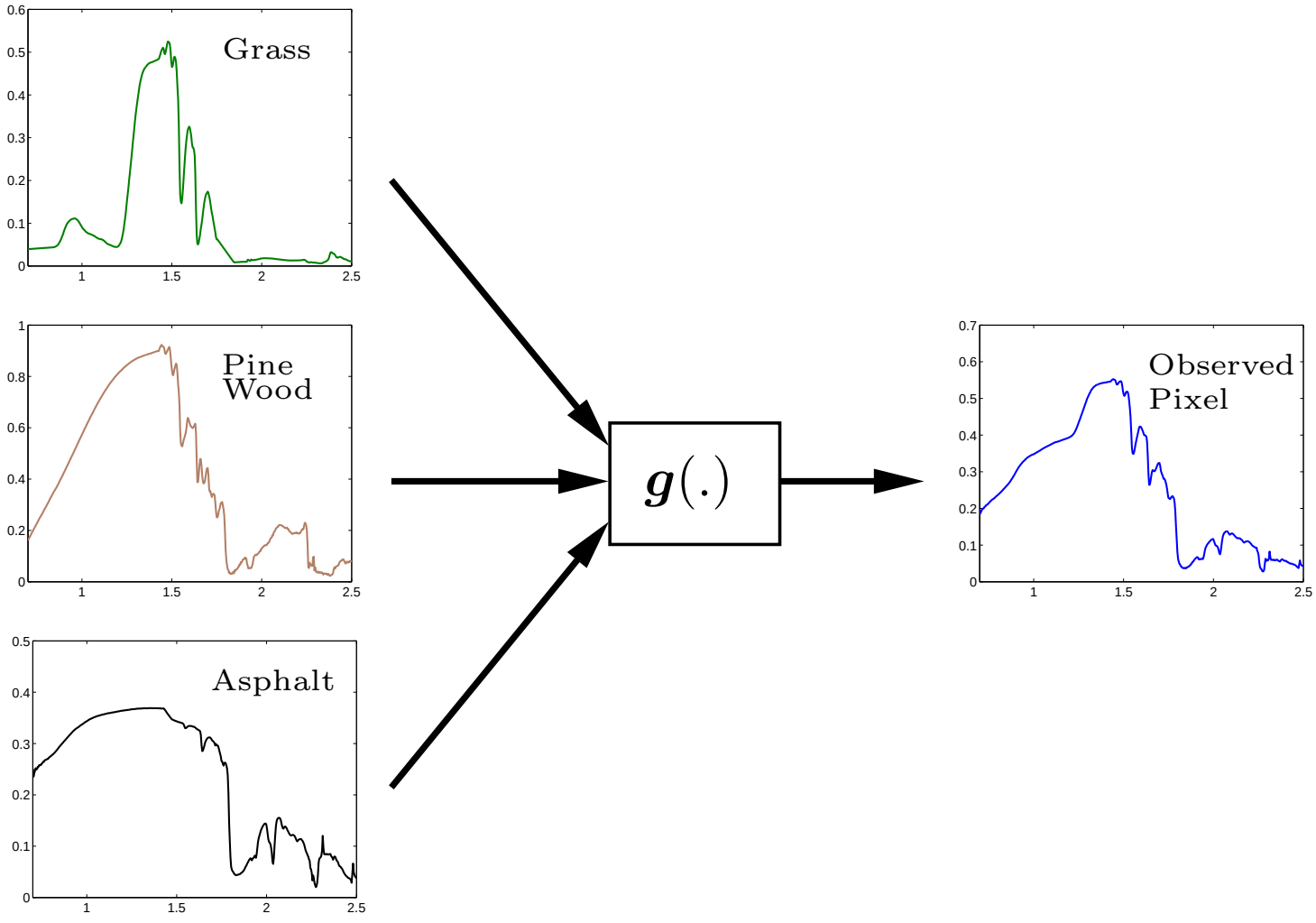
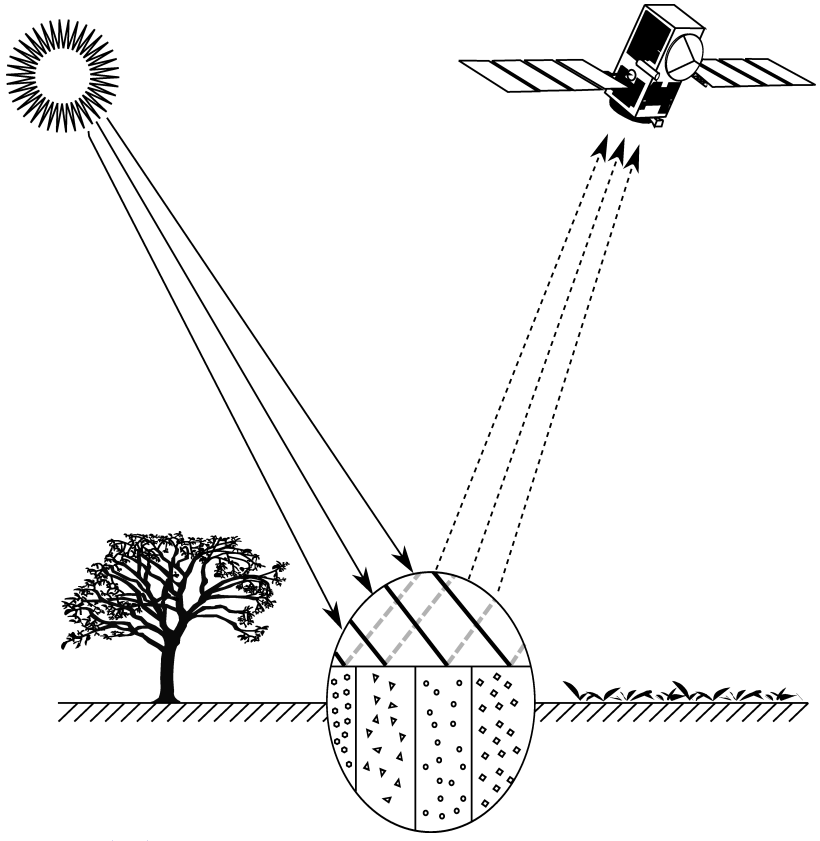
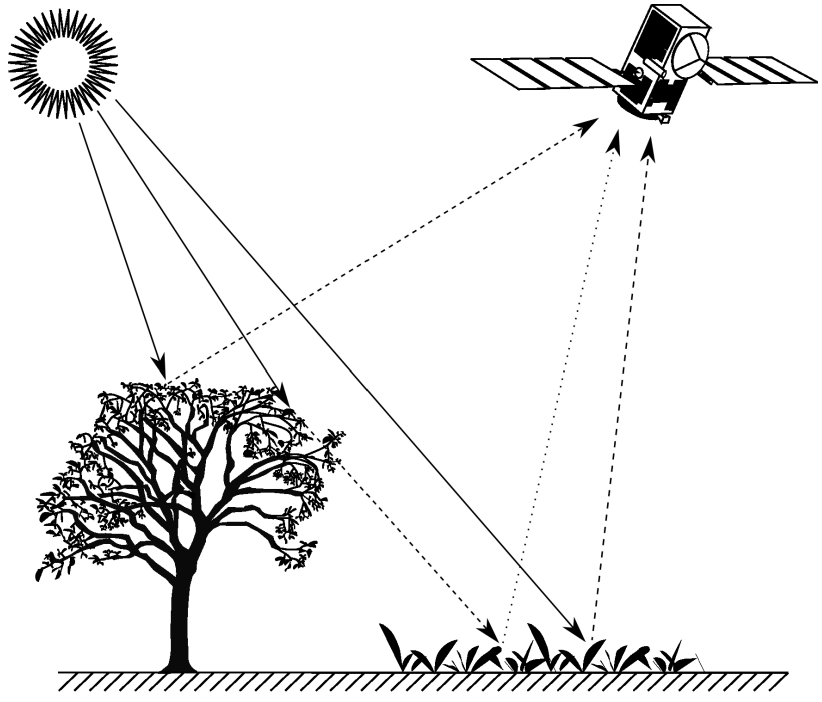


Figure: A hyperspectral pixel is a **mixture** of the spectral signatures of the materials (or **endmembers**) in the area.

Mixture Models



(a) Linear Mixing Models.



(b) Bilinear Mixing Models.

- Linear mixing model - LMM [1]:

$$\boldsymbol{y} = \boldsymbol{M}\boldsymbol{a} + \boldsymbol{n}, \qquad (\boldsymbol{M} \text{ is } L \times R,\ L \gg R) \tag{1}$$

subjected to

$$\sum_{r=1}^{R} a_r = 1, \quad a_r \geq 0, \quad \forall r \in \{1, \ldots, R\}.$$

and $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}_L, \sigma_n^2 \boldsymbol{I})$.

- Many spectral unmixing techniques rely on a specific model to estimate $\boldsymbol{a}$ and $\boldsymbol{M}$

- Actual nature of mixture ➪ usually unknown

- Desirable: a model free nonlinearity detector

- General mixing model

$$\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{M}) + \boldsymbol{n}. \tag{2}$$

- Proposition:

  To model $\boldsymbol{g}(\boldsymbol{M})$ as a realization of a GP that describes a distribution over functions

Gaussian Process Regression

- We model (2) as

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{M}) + \boldsymbol{n}, \tag{3}$$

where $\boldsymbol{n} \sim \mathcal{N}(0, \sigma_n^2 \boldsymbol{I})$ and $\boldsymbol{f}(\cdot)$ is a smooth latent function

- Following [2] and considering the training set $\{\boldsymbol{y}, \boldsymbol{M}\}$, the prior distribution for $\boldsymbol{y}$ can be written as

$$\boldsymbol{y} \ \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{K} + \sigma_n^2 \boldsymbol{I}\right), \tag{4}$$

with $\boldsymbol{K}$ the Gram matrix whose entries $\boldsymbol{K}_{ij} = k(\boldsymbol{m}_i, \boldsymbol{m}_j)$ are the kernel (covariance) functions [3] of the inputs $\boldsymbol{m}_i$ and $\boldsymbol{m}_j$ (rows of $\boldsymbol{M}$), and $\boldsymbol{I}$ is the $L \times L$ identity matrix

- Using the *marginalization property* [2], we can rewrite (4) for new inputs $\boldsymbol{M}_*$ as

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{f}_* \end{bmatrix} \sim \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{K} + \sigma_n^2 \boldsymbol{I} & \boldsymbol{K}_* \\ \boldsymbol{K}_*^\top & \boldsymbol{K}_{**} \end{bmatrix}\right) \tag{5}$$

where $[\boldsymbol{K}_*]_{ij} = k(\boldsymbol{m}_{*i}, \boldsymbol{m}_j)$ and $[\boldsymbol{K}_{**}]_{ij} = k(\boldsymbol{m}_{*i}, \boldsymbol{m}_{*j})$

- The multivariate predictive distribution of $\boldsymbol{f}_*$, or posterior of $\boldsymbol{f}_*$, can be obtained by conditioning (5) on the data as

$$\begin{aligned} \boldsymbol{f}_* | \boldsymbol{y}, \boldsymbol{M}, \boldsymbol{M}_* \sim \mathcal{N}\Big(& \boldsymbol{K}_*^\top \left[\boldsymbol{K} + \sigma_n^2 \boldsymbol{I}\right]^{-1} \boldsymbol{y}\,, \\ & \boldsymbol{K}_{**} - \boldsymbol{K}_*^\top \left[\boldsymbol{K} + \sigma_n^2 \boldsymbol{I}\right]^{-1} \boldsymbol{K}_*\Big) \end{aligned} \tag{6}$$

- Using the minimum mean squared error (MMSE) criterion, the predictor $\hat{\boldsymbol{y}}_g$ of $\boldsymbol{f}$ is defined as the mean of the predictive distribution in (6). Hence,

$$\hat{\boldsymbol{y}}_g = \widehat{\boldsymbol{f}}_*^{\text{MMSE}} = \boldsymbol{K}_*^\top \left[\boldsymbol{K} + \sigma_n^2 \boldsymbol{I}\right]^{-1} \boldsymbol{y}. \tag{7}$$

Parameter Estimation

- Using the Gaussian kernel for smoothness and non-informativeness

$$k(\boldsymbol{m}_p, \boldsymbol{m}_q) = \sigma_f^2 \exp\left\{-\frac{1}{2s^2}\|\boldsymbol{m}_p - \boldsymbol{m}_q\|^2\right\} \tag{8}$$

- We estimate the noise variance and the kernel hyperparameters in $\boldsymbol{\theta} = \{\sigma_f^2, s^2, \sigma_n^2\}$ by maximizing the marginal likelihood function $p(\boldsymbol{y}|\boldsymbol{M}, \boldsymbol{\theta})$

- Hence,

$$\hat{\boldsymbol{\theta}} = \arg_{\boldsymbol{\theta}} \max \log p(\boldsymbol{y}|\boldsymbol{M}, \boldsymbol{\theta}) \tag{9}$$

Nonlinearity Detector

- Hypothesis test

$$\begin{cases} \mathcal{H}_0 : \boldsymbol{y} = \boldsymbol{M}\boldsymbol{a} + \boldsymbol{n} \\ \mathcal{H}_1 : \boldsymbol{y} = \boldsymbol{g}(\boldsymbol{M}) + \boldsymbol{n} \end{cases} \tag{10}$$

- We assume that $\boldsymbol{M}$ is available or has been estimated from the image using an endmember extraction technique [5]

- We propose to compare the fitting errors resulting from estimating $\boldsymbol{y}$ using an LS estimator and the GP-based estimator (7)

- Under $\mathcal{H}_0$: both estimators should provide good estimates

- Under $\mathcal{H}_1$: the LS estimation error should be significantly larger

LS Fitting Error

- LS estimation error:

$$\boldsymbol{e}_\ell = \boldsymbol{y} - \hat{\boldsymbol{y}}_\ell \tag{11}$$

where $\hat{\boldsymbol{y}}_\ell = \boldsymbol{M}\hat{\boldsymbol{a}}$ is the LS estimator of $\boldsymbol{f}$

- Then, simple calculation yields

$$\boldsymbol{e}_\ell = \boldsymbol{P}\boldsymbol{y} \tag{12}$$

where $\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{M}(\boldsymbol{M}^\top \boldsymbol{M})^{-1}\boldsymbol{M}^\top$ is an $L \times L$ projection matrix of rank $\rho = L - R$.

GPM Fitting Error

- The GP-based estimation error is given by

$$\boldsymbol{e}_g = \boldsymbol{y} - \hat{\boldsymbol{y}}_g = \boldsymbol{y} - \widehat{\boldsymbol{f}}_*^{\mathrm{MMSE}}\Big|_{\boldsymbol{M}_*=\boldsymbol{M}} = \boldsymbol{H}\boldsymbol{y} \tag{13}$$

where $\boldsymbol{H} = \boldsymbol{I}_L - \boldsymbol{K}^\top \left[\boldsymbol{K} + \sigma_n^2 \boldsymbol{I}\right]^{-1}$ is a real-valued symmetric matrix of rank L

The Test Statistics

- We propose to compare the squared norms of the two fitting error vectors

  For that we need:
  - determine a test threshold from a given probability of false alarm (PFA);
  - a test statistics whose distribution is known or at least can be approximated

- We then propose the test

$$T = \frac{2\|\boldsymbol{e}_g\|^2}{\|\boldsymbol{e}_g\|^2 + \|\boldsymbol{e}_\ell\|^2} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\lessgtr}} \tau, \tag{14}$$

  where τ is the detection threshold. Equation (14) can be shown to follow a Beta distribution [6].

Simulations

Synthetic Data

- linearly mixed pixels were generated using the LMM (1);
- nonlinearly mixed pixels were generated using the simplified generalized bilinear model (GBM) [7], with a new scaling that permits the control of the degree of nonlinearity for each nonlinear pixel generated
- The nonlinearly mixed pixels were generated using the model

$$\boldsymbol{y} = \kappa \boldsymbol{M}\boldsymbol{a} + \boldsymbol{\mu} + \boldsymbol{n} \tag{15}$$

where $0 \leq \kappa \leq 1$, $\boldsymbol{\mu} = \gamma \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} a_i a_j \boldsymbol{m}_i \odot \boldsymbol{m}_j$ is the nonlinear term, γ is the parameter that governs the amount of nonlinear contribution, and $\odot$ is the Hadamard product.

- Given the parameters $\boldsymbol{M}$, $\boldsymbol{a}$, γ and σ_n^2, this model generates samples with same energy and SNR as the LMM if

$$\begin{aligned} E_{n\ell} &= \|\kappa \boldsymbol{M}\boldsymbol{a} + \boldsymbol{\mu}\|^2 = E_\ell = \|\boldsymbol{M}\boldsymbol{a}\|^2 \\ k^2 E_\ell &+ 2kE_{\ell\mu} + E_\mu - E_\ell = 0 \end{aligned} \tag{16}$$

$$\kappa = \left[-2E_{\ell\mu} + \sqrt{4E_{\ell\mu}^2 - 4E_\ell(E_\mu - E_\ell)}\right] / 2E_\ell \tag{17}$$

where $E_\ell = \|\boldsymbol{y}_\ell\|^2$ is the energy of a noiseless linear pixel (i.e., $\boldsymbol{a}^\top \boldsymbol{M}^\top \boldsymbol{M}\boldsymbol{a}$), $E_{\ell\mu} = \boldsymbol{y}_\ell^\top \boldsymbol{\mu}$ is the "cross-energy" of the linear and nonlinear parts, and $E_\mu = \|\boldsymbol{\mu}\|^2$ is the energy of the nonlinear contribution.

Degree of Nonlinearity

- The degree of nonlinearity of a pixel is then defined as the ratio of the nonlinear portion to the total pixel energy

$$\eta_d = \frac{2\kappa E_{\ell\mu} + E_\mu}{\kappa^2 E_\ell + 2\kappa E_{\ell\mu} + E_\mu}. \tag{18}$$

- Synthetic data generation
  - $\boldsymbol{M}$ composed of $R = 3$ materials (green grass, olive green paint and galvanized steel metal) extracted from the spectral library of the software ENVI$^{\text{TM}}$ [8]
  - Each endmember $\boldsymbol{m}_r$ has $L = 826$ bands that were uniformly decimated to $L = 83$ bands for simplicity
  - Abundance vector $\boldsymbol{a} = [0.3, 0.6, 0.1]^\top$ arbitrarily fixed, and $\sigma_n^2 = 0.0011$ chosen to produce an SNR of 21dB for both linear and nonlinear samples
  - Simplified GBM with $\gamma = [1, 3, 5]$ ($\eta_d = [0.22, 0.55, 0.80]$) for nonlinearly mixed samples

M known

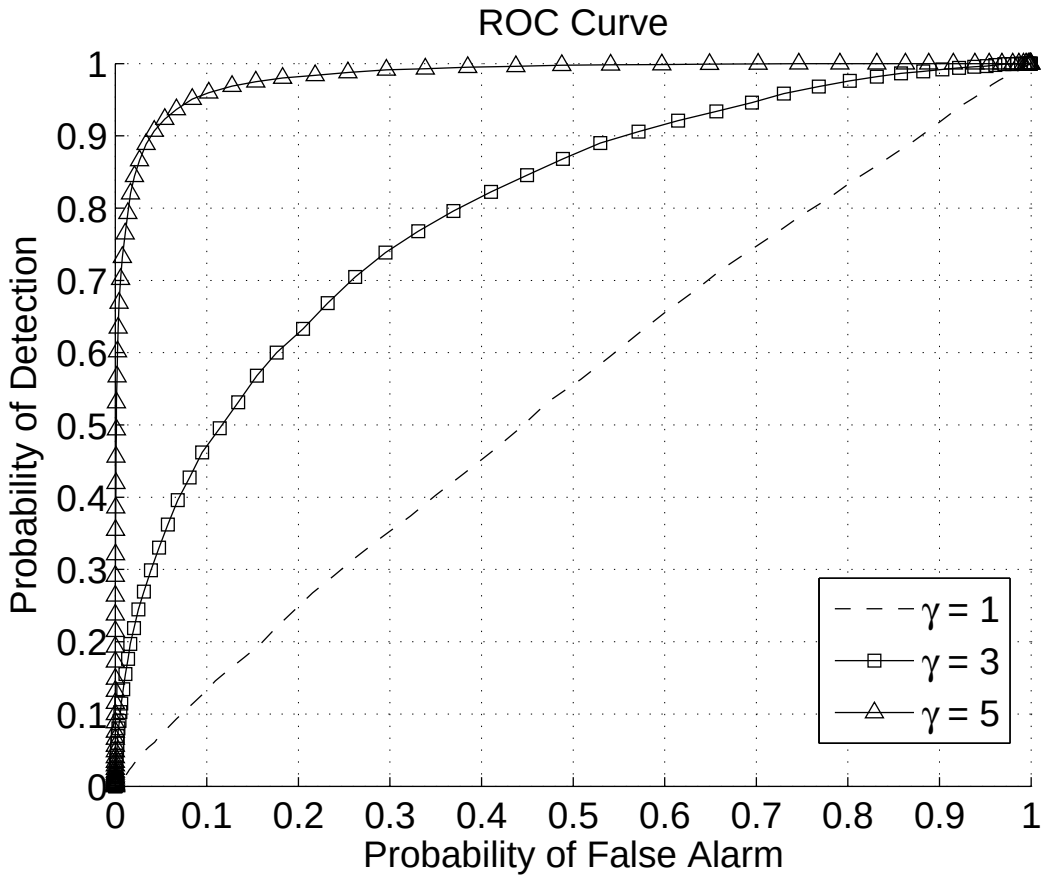


Figure: (a) Empirical ROCs for the LS-based detector [7] for 20,000 synthetic samples (10,000 for each hypothesis).

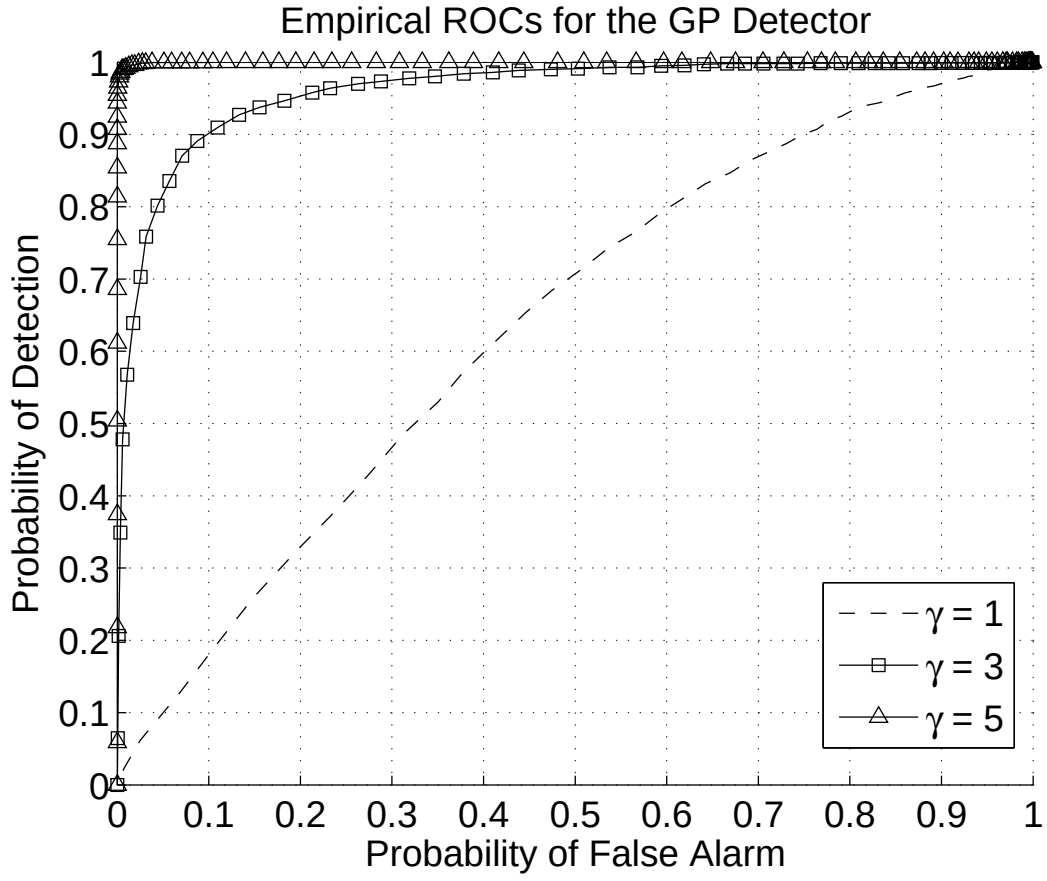


Figure: (b) Empirical ROCs for the GP detector for 4,000 samples (2,000 for each hypothesis), for nonlinear samples the simplified GBM was used with $\gamma = [1, 3, 5]$ ($\eta_d = [0.22, 0.55, 0.80]$).

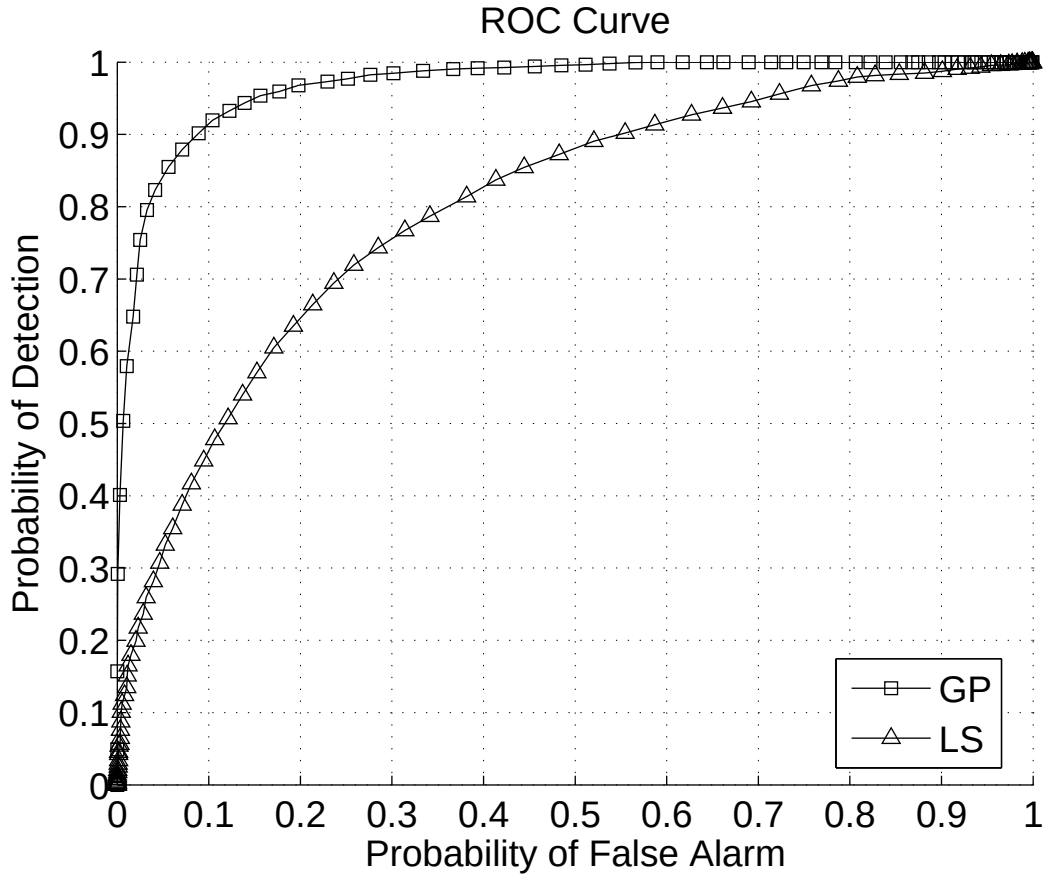


Figure: (c) Comparison of the empirical ROCs for GP and LS detectors for $\gamma = 3$ and 4,000 samples.

$\boldsymbol{M}$ unknown

- 4 experiments using 5,000 synthetic samples;
- the proportion of nonlinearly mixed pixels in the image varying from 10% to 50%;
- random abundance vectors;
- $\boldsymbol{M}$ was estimated using the well known *vertex component analysis* (VCA) [9].

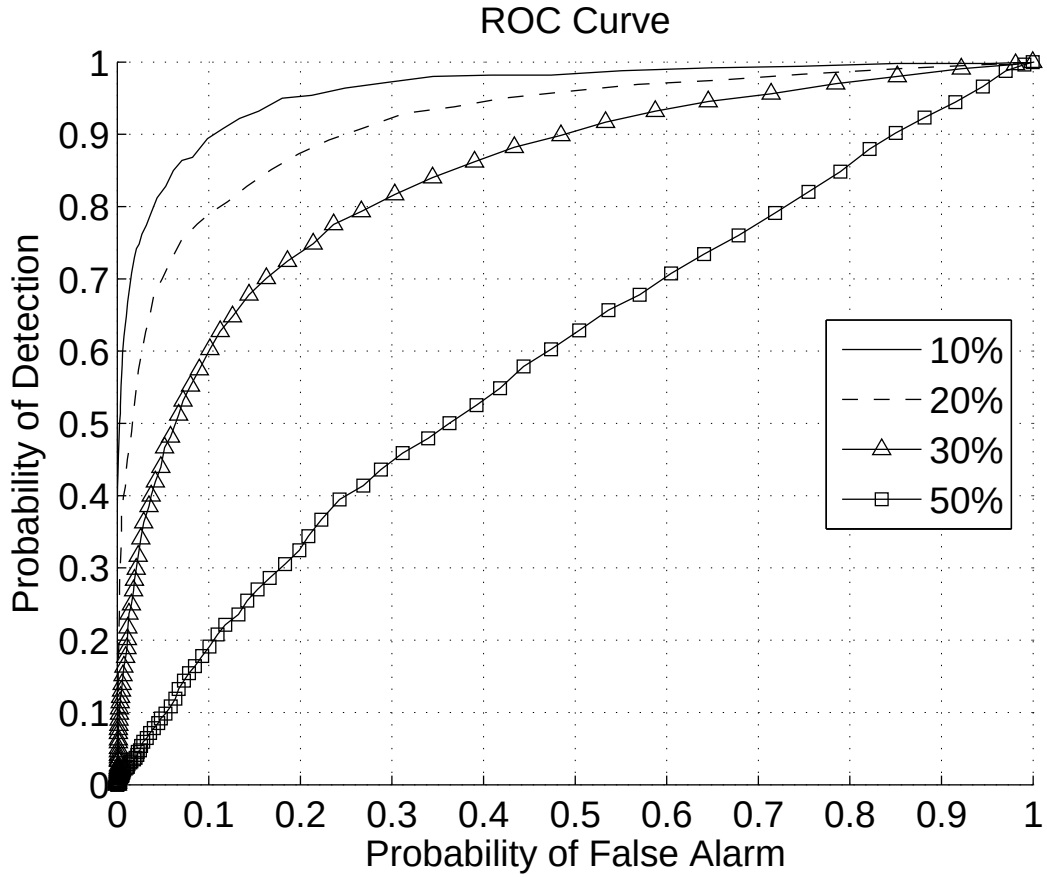
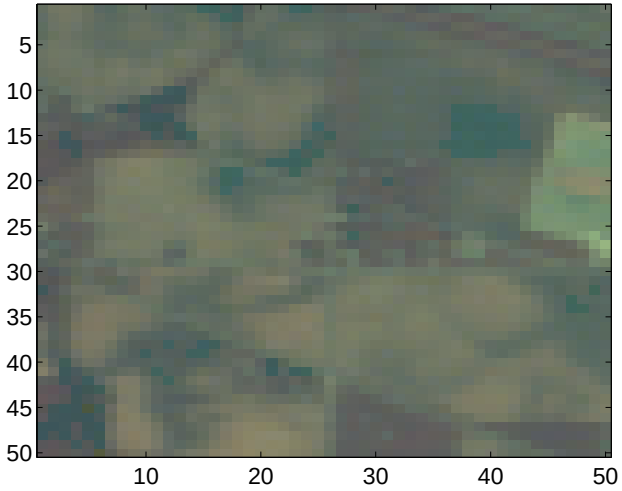


Figure: ROCs for different proportions of nonlinearly mixed pixels.

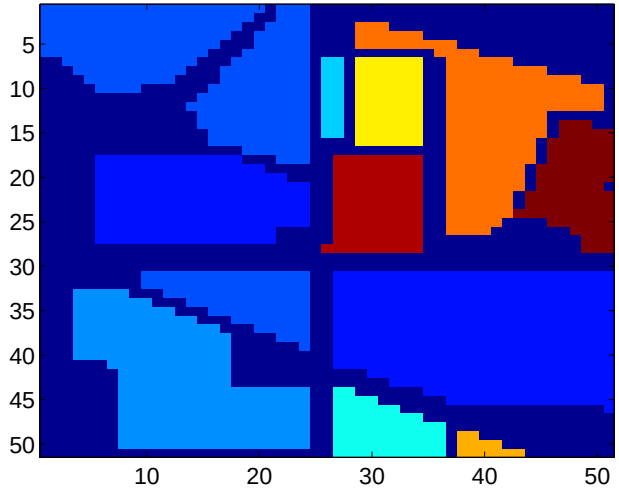
Detection performance degrades as the number of nonlinear pixels increases and VCA looses accuracy in extracting the endmembers. Thus, alternatives to VCA must be sought for nonlinearly-mixed pixels.

Real Data

- Indian Pines test site in North-western Indiana [10]
- Captured by the AVIRIS (Airborne Visible/Infrared Imaging Spectrometer), the image has 145×145 samples, and has 220 contiguous bands (366 to 2497 nm)
- Noisy and water absorption bands were removed resulting in a total of 200 bands that were decimated to 50 to speed up simulations
- Ground truth map that divides the samples in 17 classes



(a) Indian Pines.



(b) Ground Truth.

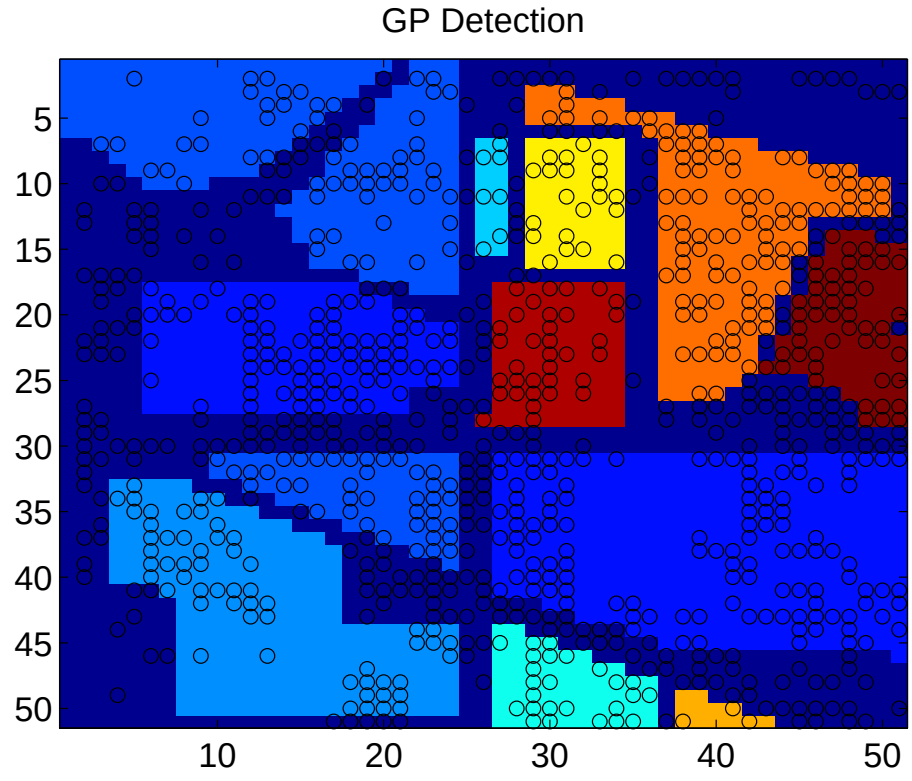


Figure: Detection map: black circles indicate the pixels detected as non-linearly mixed.

Conclusions

- A GP-based nonlinearity detection strategy was introduced to detect nonlinearly mixed pixels in hyperspectral images
- The proposed detector does not require the use of a parametric model for the underlying nonlinear mixing functions
- Simulations using synthetic data indicate that the proposed detector outperforms a robust method previously presented in the literature
- The detector was also tested on the Indian Pines image showing that pixels close to the class boundaries and in the background seem to be nonlinearly mixed
- Future work includes joint detection of nonlinear mixtures and unmixing

Thanks

This work was partly supported by CNPq under grants Nos 305377/2009-4, 400566/2013-3 and 141094/2012-5, and by the Agence Nationale pour la Recherche, France, (Hypanema project, ANR-12- BS03-003), and by ANR-11-LABX-0040-CIMI within the program ANR-11-IDEX-0002-02.

References

[1] N. Keshava and J. Mustard,
"Spectral unmixing,"
IEEE Signal Processing Magazine, vol. 19, no. 1, pp. 44–57, 2002.

[2] C. E. Rasmussen and C. K. I. Williams,
Gaussian Processes for Machine Learning,
The MIT Press, 2006.

[3] B. Schölkopf and A. J. Smola,
Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond,
The MIT Press, 2001.

[4] W. Liu, J. C. Prncipe, and S. Haykin,
Kernel Adaptive Filtering: A Comprehensive Introduction,
Wiley, 2010.

[5] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. Nasrabadi, and J. Chanussot,
"Hyperspectral remote sensing data analysis and future challenges,"
IEEE Geoscience and Remote Sensing Magazine, vol. 1, no. 2, pp. 6–36, 2013.

[6] N. J. Johnson, S. Kotz, and N. Balakrishnan,
Continuous Univariate Distributions, vol. 1,
Wiley-Interscience, 1994.

[7] Y. Altmann, N. Dobigeon, J.-Y. Tourneret, and J. C. M. Bermudez,
"A robust test for nonlinear mixture detection in hyperspectral images,"
in *Proc. IEEE ICASSP*, 2013.

[8] RSI (Research Systems Inc.,
"Envi user's guide version 4.0," Sept. 2013.

[9] J. M. P. Nascimento and J. M. Bioucas 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.

[10] I. Dopido, M. Zortea, A. Villa, A. Plaza, and P. Gamba,
"Unmixing prior to supervised classification of remotely sensed hyperspectral images,"
IEEE Geoscience and Remote Sensing Letters, vol. 8, no. 4, pp. 760–764, 2011.