Nonparametric Detection of Nonlinearly Mixed Pixels and Endmember Estimation in Hyperspectral Images

José Carlos Moreira Bermudez[1]

Federal University of Santa Catarina, Florianópolis, SC, Brazil

This work was done in collaboration with

Tales Imbiriba (UFSC)
Cédric Richard (Université de Nice)
Jean-Yves Tourneret (INPT)

July 5, 2016

[1]This work was partly supported by CNPq grants 305377/2009-4, 400566/2013-3 and 141094/2012-5, by ANR grants ANR-12- BS03-003 (Hypanema), and ANR-11-LABX-0040-CIMI.

Hyperspectral Images

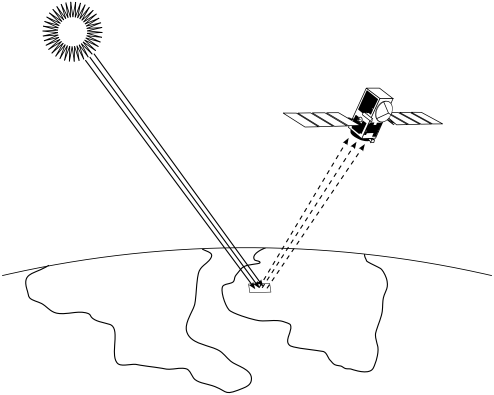


Figure: Remote Sensing: The Sun's radiation reflected on the Earth's surface is captured by an airborne or spaceborne hyperspectral sensor.

- High spectral resolution × poor spacial resolution

- One hyperspectral pixel $\boldsymbol{r}$ has hundreds of contiguous bands.

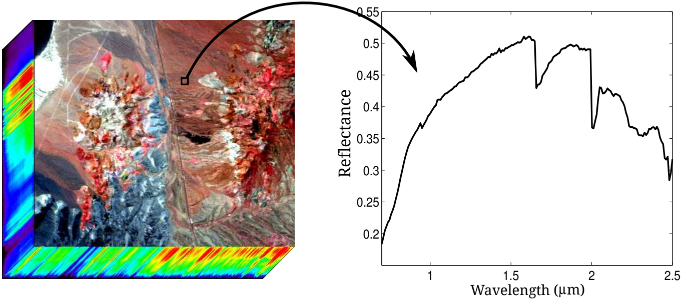


Figure: Illustration of the Hypercube captured by the AVIRIS instrument from the Cuprite field.

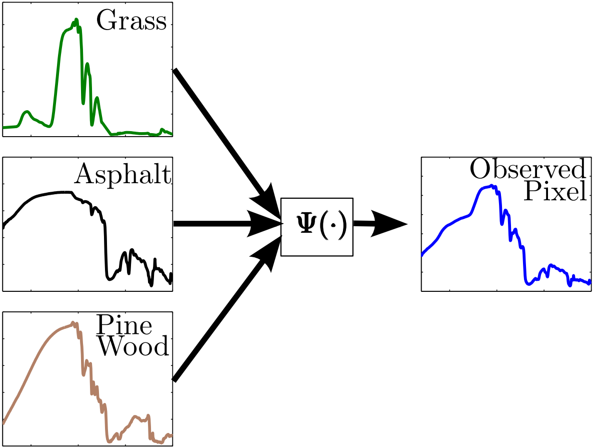


Figure: A observed pixel is in fact a mixture of spectral signatures.

Mixture Models

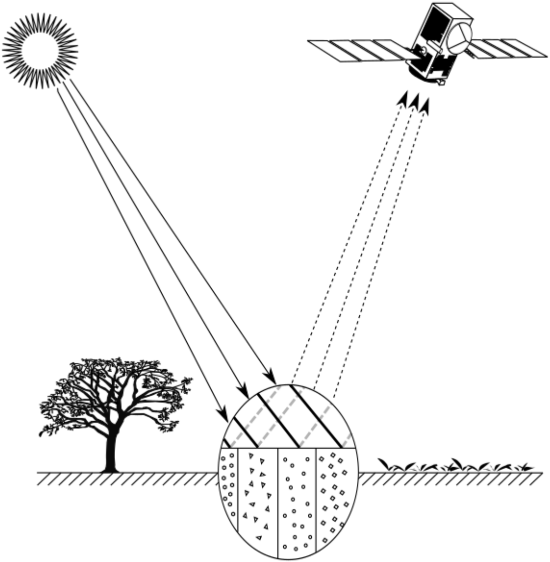


Figure: Linear Mixing.

The Linear Mixing Model - LMM

$$\boldsymbol{r} = \boldsymbol{M}\boldsymbol{\alpha} + \boldsymbol{n}$$

- $\boldsymbol{r}$: Observation vector ($L \times 1$)
- $\boldsymbol{M}$: Endmember matrix (R endmember spectra $\boldsymbol{m}_i$)
- $\boldsymbol{\alpha}$: Vector of abundances ($R \times 1$)
- $\boldsymbol{n}$: Noise vector (WGN, $\boldsymbol{n} \sim \mathcal{N}(\mathbf{0}_L, \sigma_n^2 \boldsymbol{I}_L)$)

Ref: [Keshava, 2002] [1]

Linear mixing model

$$\boldsymbol{r} = \boldsymbol{M}\boldsymbol{\alpha} + \boldsymbol{n}$$

Observations:

- This is a reasonable first order model
- Assumes that each ray of light interacts with only one material
- Simple to treat mathematically

A little complication ...

Abundance values should be constrained for physical meaning

- Positiveness

$$\alpha_k \geq 0, \quad \forall k \in \{1, \ldots, R\}$$

- Sum to one constraint (proportionality)

$$\sum_{k=1}^{R} \alpha_k = 1$$

- The constraints define simplexes

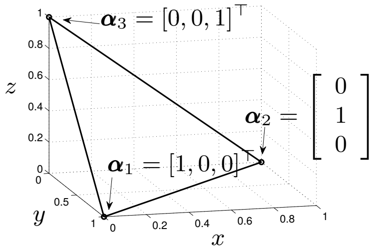
$$\mathcal{S}_{\boldsymbol{\alpha}} = \{\alpha \in \mathbb{R}^R | \boldsymbol{\alpha} \geq \mathbf{0}, \boldsymbol{\alpha}^\top \mathbf{1} = 1\}$$

For noiseless observations

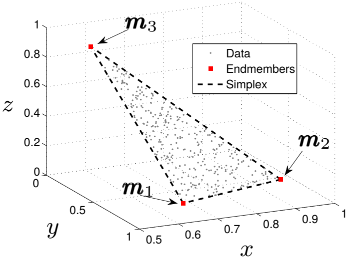
$$\mathcal{S}_{\boldsymbol{r}} = \{\boldsymbol{r} \in \mathbb{R}^L | \boldsymbol{r} = \boldsymbol{M}\boldsymbol{\alpha}, \boldsymbol{\alpha} \geq \mathbf{0}, \boldsymbol{\alpha}^\top \mathbf{1} = 1\}$$

$$\mathcal{S}_{\boldsymbol{\alpha}} = \{\alpha \in \mathbb{R}^R | \boldsymbol{\alpha} \geq \mathbf{0}, \boldsymbol{\alpha}^\top \mathbf{1} = 1\}$$

$$\mathcal{S}_{\boldsymbol{r}} = \{\boldsymbol{r} \in \mathbb{R}^L | \boldsymbol{r} = \boldsymbol{M}\boldsymbol{\alpha}, \boldsymbol{\alpha} \geq \mathbf{0}, \boldsymbol{\alpha}^\top \mathbf{1} = 1\}$$



(a) Unitary Simplex $\mathcal{S}_{\boldsymbol{\alpha}}$



(b) Data Simplex $\mathcal{S}_{\boldsymbol{r}}$.

- The convex geometry of facilitates the estimation of endmembers from data

- Most endmember extraction techniques are based on these properties

Linear Unmixing

- Step 1: Endmember extraction
  - Various techniques

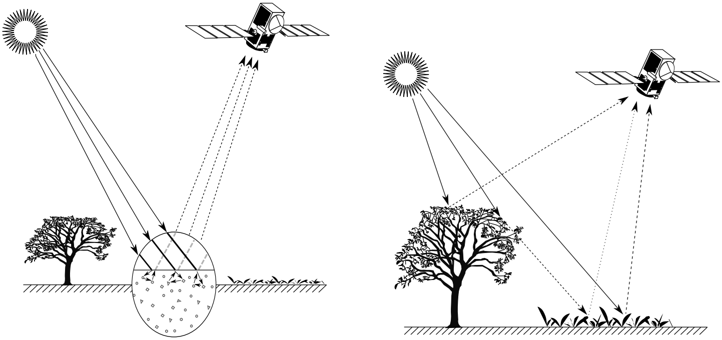
- Step 2: Abundance estimation

$$\boldsymbol{\alpha}^\star = \arg\min_{\boldsymbol{\alpha}} \|\boldsymbol{r} - \boldsymbol{M}\boldsymbol{\alpha}\|_2^2, \quad \text{s.t.} \begin{cases} \alpha_k \geq 0, & \forall k \in \{1, \dots, R\} \\ \sum_{k=1}^{R} \alpha_k = 1 \end{cases}$$

  - Fully constrained LS (FCLS) [Heinz et al., 2001] [2]
  - Geometrical approaches [Honeine et al., 2012] [3]
  - MVES (Minimum Volume-Enclosing Simplex) - joint $\boldsymbol{M}$ and $\boldsymbol{\alpha}$ estimation [Chan et al., 2009] [4]
  - Bayesian approaches [Dobigeon et al., 2009] [5]

Hum ... but life is not always that simple ...

Two other types of light interaction that complicate life ...



(a) Intimate mixing.   (b) Multiple scattering.

These types of light interaction lead to nonlinear mixing of endmember contributions!!!

Nonlinear Mixing Models

- How can we deal with nonlinearly mixed hyperspectral images?
- Several parametric nonlinear mixing model have been proposed
- Intimate Mixing Models
  - Radiative transfer model [Hapke, 1981] [6]
  - Models are physically motivated ☺
  - Manageable under simplifying approximations but still complex ☹

- Bilinear Mixing Models
  - Parametric models ☺
  - Tend to preserve definitions used in the LMM ☺
  - Have the LMM as a particular case ☺
  - Not really physically motivated ☹
- Post-Nonlinear Mixing Models
  - Allow for more general nonlinearities ☺
  - Usually require more work to estimate parameters ☹
  - May become quite simple in some cases ☺

Bilinear Mixing Models

- General expression

$$\boldsymbol{r} = \boldsymbol{f}(\boldsymbol{M}, \boldsymbol{\alpha}) + \boldsymbol{n}$$

where

$$\boldsymbol{f}(\boldsymbol{M}, \boldsymbol{\alpha}) = \sum_{k=1}^{R} \alpha_k \boldsymbol{m}_k + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} \beta_{i,j} \boldsymbol{m}_i \odot \boldsymbol{m}_j$$

$$\boldsymbol{f}(\boldsymbol{M}, \boldsymbol{\alpha}) = \sum_{k=1}^{R} \alpha_k \boldsymbol{m}_k + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} \beta_{i,j} \boldsymbol{m}_i \odot \boldsymbol{m}_j$$

- Fan's model: $\beta_{i,j} = \alpha_i \alpha_j$ [Fan et al., 2009] [7]

$$\boldsymbol{f}(\boldsymbol{M}, \boldsymbol{\alpha}) = \sum_{k=1}^{R} \alpha_k \boldsymbol{m}_k + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} \alpha_i \, \alpha_j \boldsymbol{m}_i \odot \boldsymbol{m}_j$$

subject to

$$\sum_{i=1}^{R} \alpha_i = 1, \quad \text{and} \quad \alpha_i \geq 0, \; i = 1, \ldots, R.$$

OBS:

- Nonlinearity must be presented if $\boldsymbol{m}_i$ is present
- The nonlinear term can place the vector anywhere outside the LMM simplex

$$\boldsymbol{f}(\boldsymbol{M}, \boldsymbol{\alpha}) = \sum_{k=1}^{R} \alpha_k \boldsymbol{m}_k + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} \beta_{i,j} \boldsymbol{m}_i \odot \boldsymbol{m}_j$$

- Nascimento's model: $\beta_{i,j}$ become abundances
  [Nascimento et al., 2009] [8]

$$\boldsymbol{f}(\boldsymbol{M}, \boldsymbol{\alpha}) = \sum_{k=1}^{R} \alpha_k \boldsymbol{m}_k + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} \beta_{i,j} \boldsymbol{m}_i \odot \boldsymbol{m}_j$$

  subject to
  - Positivity constraint: $\alpha_k \geq 0$, $\beta_{i,j} \geq 0$ for $\forall k$ and $\forall (i,j)$
  - Sum-to-one constraint: $\sum_{k=1}^{R} \alpha_k + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} \beta_{i,j} = 1$.

  OBS:
  - The nonlinear terms become new endmembers.

  ➪ Not practical for joint estimation of $\boldsymbol{M}$ and $\boldsymbol{\alpha}$

$$\boldsymbol{f}(\boldsymbol{M},\boldsymbol{\alpha}) = \sum_{k=1}^{R} \alpha_k \boldsymbol{m}_k + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} \beta_{i,j} \boldsymbol{m}_i \odot \boldsymbol{m}_j$$

- Generalized Bilinear Model (GBM): $\beta_{i,j} = \gamma_{i,j}\alpha_i\alpha_j$ [Halimi et al., 2011] [9]

$$\boldsymbol{f}(\boldsymbol{M},\boldsymbol{\alpha}) = \sum_{k=1}^{R} \alpha_k \boldsymbol{m}_k + \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} \beta_{i,j} \boldsymbol{m}_i \odot \boldsymbol{m}_j$$

subject to

$$\alpha_k \geq 0, \qquad \forall k \in \{1, \ldots, R\}, \qquad \sum_{k=1}^{R} \alpha_k = 1$$

$$0 \leq \gamma_{i,j} \leq 1, \qquad \forall i \in \{1, \ldots, R-1\}, \quad \forall j \in \{i+1, \ldots, R\}.$$

OBS:

- The endmember matrix is the same as in the LMM
- The constraints on α_k and $\gamma_{i,j}$ preclude the modeling of strong nonlinearities

Post Nonlinear Mixing Models

- General expression

$$\boldsymbol{r} = \boldsymbol{g}\left(\boldsymbol{M}\,\boldsymbol{\alpha}\right) + \boldsymbol{n}$$

- Some models
  - PNMM [Chen et al., 2013] [10]

$$\boldsymbol{r} = (\boldsymbol{M}\boldsymbol{\alpha})^{\xi} + \boldsymbol{n}$$

  - Post Polinomial Nonlinear Mixing Model [Altmann et al., 2011-2013] [11]

$$g(s_i) = s_i + b s_i^2 + ..., \qquad i = 1, \ldots, L$$

    where s_i is the i-th component of $\boldsymbol{M}\boldsymbol{\alpha}$

Nonlinear Unmixing

- Most techniques can be classified into two groups
  - Unmixing using parametric mixing models
  - Unmixing using model-free methods
- Both have pros and cons [Dobigeon et al., 2014] [12]

Nonlinear unmixing using parametric models

- Assuming $\boldsymbol{M}$ known (or estimated)

$$\boldsymbol{\theta}^\star = \arg\min_{\boldsymbol{\theta}} \|\boldsymbol{r} - \boldsymbol{\varphi}(\boldsymbol{M}, \boldsymbol{\theta})\|_2^2$$
$$\text{subject to} \quad \boldsymbol{\theta} \in \Omega,$$

  - $\boldsymbol{\theta}$: parameter vector (abundances + other parameters)
  - $\boldsymbol{\varphi}(\cdot)$: parametric mixing model
  - Ω: defines the feasible region
- This is a supervised parameter estimation problem
  - Optimization methods [Halimi et al., 2011a] [13]
  - Bayesian approaches [Halimi et al., 2011b] [9]

- For unknown $\boldsymbol{M}$ ➪ unsupervised estimation
  - Bayesian approach [Altmann et al., 2014] [14]
  - Manifold learning techniques [Heylen et al., 2012] [15]

Model-Free Nonlinear Unmixing

- If the type of nonlinearity is unknown

  ➪ More flexible approaches must be sought

- Methods based on reproducing kernels are attractive

  ➪ Function approximation without a parametric model

  - Supervised methods [Chen et al., 2013] [16]
  - Unsupervised Bayesian methods [Altmann et al., 2013] [17]
  - Unsupervised manifold learning [Nguyen et al., 2012] [18]

Facts About Hyperspectral Image Analysis

- Basically all hyperspectral images contain
  - Pixels that are "almost" linearly mixed
  - Pixels that are definitely nonlinearly mixed
- Nonlinear unmixing can lead to a better understand of the individual spectral contributions
- Nonlinear analysis is more challenging and more complex
- Unmixing linearly mixed pixels using nonlinear unmixing methods

  ➪ Usually poorer results than using linear unmixing

Natural Approach

- Detect the nonlinearly mixed pixels in the image
- Apply linear unmixing to linearly mixed pixels
- Apply nonlinear unmixing to nonlinearly mixed pixels

Detecting Nonlinearly Mixed Pixels

- Physically motivated models tend to be too complex [Borel et al., 1994] [19]
- Using surrogate data [Han et al., 2008] [20]

  ➪ Not very good results

- Using simplified parametric mixing models [Altmann et al., 2013a] [11]

  ➪ Not sure the model can capture the existing nonlinearity

- Using the distance between the pixel and the LMM simplex [Altmann et al., 2013b] [21]

  ➪ Conveys too little information about the nonlinearity

- Using a model-selection approach for detecting nonlinearities with different statistical properties [Altmann et al., 2014] [14]

  ➪ Bayesian supervised approach combining detection and unmixing (complexity and flexibility)

Proposed Nonlinearity Detection Strategy

- Objectives
  - To detect nonlinearly mixed pixels prior to unmixing
  - To obtain a model-free approach that generalizes well
  - The method should be unsupervised
- Strategy
  - Determine how well the observed pixel spectrum fits both a linear and a nonlinear recursion
  - Propose a hypothesis test assuming that:
    - Both estimators provide good results for linearly mixed pixels
    - The nonlinear estimator provides better results for nonlinearly mixed pixels

Estimators

- Linear estimator: Least-squares
- Nonlinear estimator: Gaussian process regression
  - Define a stochastic model for the unknown function
  - Perform inference in functional spaces
  - Modeling a nonlinear function as a Gaussian Process

    $f(\boldsymbol{x}_k)$ for inputs $\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_K$ modeled as jointly Gaussian random variables

Gaussian Process Regression

- Mathematical model for the ℓ-th band of the HI

$$r_\ell = \psi(\boldsymbol{m}_{\lambda_\ell}) + n_\ell, \qquad \ell = 1, \ldots, L$$

  - $\boldsymbol{m}_{\lambda_\ell}$: ℓ-th row of $\boldsymbol{M}$
  - $\boldsymbol{M} = [\boldsymbol{m}_{\lambda_1}, \ldots, \boldsymbol{m}_{\lambda_L}]^\top$
  - $\boldsymbol{r} = [r_1, \ldots, r_L]^\top$
  - n_ℓ is white Gaussian noise with power σ_n^2

- Gaussian process definition

$$\mathbb{E}\{\psi(\boldsymbol{m}_{\lambda_\ell})\} = 0$$
$$\mathbb{E}\{\psi(\boldsymbol{m}_{\lambda_\ell})\psi(\boldsymbol{m}_{\lambda_{\ell'}})\} = \kappa(\boldsymbol{m}_{\lambda_\ell}, \boldsymbol{m}_{\lambda_{\ell'}})$$

  - $\kappa(\cdot, \cdot)$ is a positive definite kernel [Rasmussen, 2006] [22]

- Prior of the noisy observation

$$\boldsymbol{r} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{K} + \sigma_n^2 \boldsymbol{I})$$

  - $\boldsymbol{K}$ is the Gram matrix, $\boldsymbol{K}_{ij} = \kappa(\boldsymbol{m}_{\lambda_i}, \boldsymbol{m}_{\lambda_j})$
  - σ_n^2 is the noise power

- Joint distribution of the observation $\boldsymbol{r}$ and $\psi_* \triangleq \psi(\boldsymbol{m}_{\lambda_*})$

$$\begin{bmatrix} \boldsymbol{r} \\ \psi_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{K} + \sigma_n^2 \boldsymbol{I} & \boldsymbol{\kappa}_* \\ \boldsymbol{\kappa}_*^\top & \kappa_{**} \end{bmatrix}\right)$$

$$\boldsymbol{\kappa}_* = [\kappa(\boldsymbol{m}_{\lambda_*}, \boldsymbol{m}_{\lambda_1}), \ldots, \kappa(\boldsymbol{m}_{\lambda_*}, \boldsymbol{m}_{\lambda_L})]^\top$$
$$\kappa_{**} = \kappa(\boldsymbol{m}_{\lambda_*}, \boldsymbol{m}_{\lambda_*})$$

- Predictive distribution (posterior) for $\psi_* \triangleq \psi(\boldsymbol{m}_{\lambda_*})$

$$\psi_* | \boldsymbol{r}, \boldsymbol{M}, \boldsymbol{m}_{\lambda_*} \sim \mathcal{N}\left(\boldsymbol{\kappa}_*^\top \left[\boldsymbol{K} + \sigma_n^2 \boldsymbol{I}\right]^{-1} \boldsymbol{r},\ \kappa_{**} - \boldsymbol{\kappa}_*^\top \left[\boldsymbol{K} + \sigma_n^2 \boldsymbol{I}\right]^{-1} \boldsymbol{\kappa}_*\right)$$

- Extension to multivariate test data $\boldsymbol{M}_* = [\boldsymbol{m}_{\lambda_{*1}}, \ldots, \boldsymbol{m}_{\lambda_{*L}}]^\top$

$$\boldsymbol{\psi}_* | \boldsymbol{r}, \boldsymbol{M}, \boldsymbol{M}_* \sim \mathcal{N}\left(\boldsymbol{K}_*^\top \left[\boldsymbol{K} + \sigma_n^2 \boldsymbol{I}\right]^{-1} \boldsymbol{r},\ \boldsymbol{K}_{**} - \boldsymbol{K}_*^\top \left[\boldsymbol{K} + \sigma_n^2 \boldsymbol{I}\right]^{-1} \boldsymbol{K}_*\right)$$

$$[\boldsymbol{K}_*]_{ij} = \kappa(\boldsymbol{m}_{\lambda_{\star i}}, \boldsymbol{m}_{\lambda_j}) \text{ and } [\boldsymbol{K}_{**}]_{ij} = \kappa(\boldsymbol{m}_{\lambda_{\star i}}, \boldsymbol{m}_{\lambda_{\star j}})$$

- Minimum mean square error (MMSE) estimator

$$\begin{aligned}\widehat{\boldsymbol{\psi}}_* &= \mathbb{E}\{\boldsymbol{\psi}_* | \boldsymbol{r}, \boldsymbol{M}, \boldsymbol{M}_*\} \\ &= \boldsymbol{K}_*^\top \left[\boldsymbol{K} + \sigma_n^2 \boldsymbol{I}\right]^{-1} \boldsymbol{r}.\end{aligned}$$

  - The estimator is a function of σ_n^2 and the parameter vector $\boldsymbol{\theta}$ (kernel parameters, for instance)
  - They must be estimated
  - We maxime the marginal likelihood $p(\boldsymbol{r}|\boldsymbol{M}, \sigma_n^2, \boldsymbol{\theta})$ with respect to $(\sigma_n^2, \boldsymbol{\theta})$

$$(\hat{\sigma}_n^2, \hat{\boldsymbol{\theta}}) = \arg\max_{\sigma_n^2, \boldsymbol{\theta}} \left(-\frac{1}{2}\boldsymbol{r}^\top \left[\boldsymbol{K} + \sigma_n^2 \boldsymbol{I}\right]^{-1} \boldsymbol{r} - \frac{1}{2} \log |\boldsymbol{K} + \sigma_n^2 \boldsymbol{I}|\right)$$

Parameter Estimation

- Gaussian kernel [Liu, 2010] [23]

$$\kappa(\boldsymbol{m}_{\lambda_p}, \boldsymbol{m}_{\lambda_q}) = \exp\left\{-\frac{1}{2s^2}\|\boldsymbol{m}_{\lambda_p} - \boldsymbol{m}_{\lambda_q}\|^2\right\}$$

- $\boldsymbol{\theta} = \{s^2, \sigma_n^2\}$

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

Nonlinearity Detector

- Binary hypothesis test problem

$$\begin{cases} \mathcal{H}_0 : \boldsymbol{r} = \boldsymbol{M}\boldsymbol{\alpha} + \boldsymbol{n} \\ \mathcal{H}_1 : \boldsymbol{r} = \boldsymbol{\psi}(\boldsymbol{M}) + \boldsymbol{n} \end{cases}$$

- At this point $\boldsymbol{M}$ is assumed known
- We compare the fitting errors
  - 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

- Linear fitting error

$$\boldsymbol{e}_{\text{lin}} = \boldsymbol{r} - \boldsymbol{M}\hat{\boldsymbol{\alpha}}$$

- LS estimator

$$\boldsymbol{e}_{\text{lin}} = \boldsymbol{P}\boldsymbol{r}$$

- $\boldsymbol{P} = \boldsymbol{I} - \boldsymbol{M}(\boldsymbol{M}^\top \boldsymbol{M})^{-1}\boldsymbol{M}^\top$

  $L \times L$ projection matrix of rank $\rho = L - R$

Gaussian process model (GPM) fitting error

- GP-based fitting error

$$\boldsymbol{e}_{\text{nlin}} = \boldsymbol{r} - \hat{\boldsymbol{r}}_g = \boldsymbol{r} - \widehat{\boldsymbol{\psi}}_*^{\text{MMSE}}\Big|_{\boldsymbol{M}_*=\boldsymbol{M}} = \boldsymbol{H}\boldsymbol{r}$$

- $\boldsymbol{H} = \boldsymbol{I}_L - \boldsymbol{K}^\top \left[\boldsymbol{K} + \sigma_n^2 \boldsymbol{I}\right]^{-1}$

  Real-valued symmetric matrix of rank L

The Test Statistics

- We compare the two error norms
- Desirable:
  - Be able to specify a probability of false alarm (PFA)

    ➪ Test statistics should at least approximate a known distribution under $\mathcal{H}_0$
- Proposed test statistics

$$T = \frac{2\|\boldsymbol{e}_{\text{nlin}}\|^2}{\|\boldsymbol{e}_{\text{nlin}}\|^2 + \|\boldsymbol{e}_{\text{lin}}\|^2} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \tau$$

- Reasoning
  - To be able to design PFA, we need a known pdf under $\mathcal{H}_0$
  - For the LS estimation error

$$\boldsymbol{e}_{\mathrm{lin}} | \mathcal{H}_0 \sim \mathcal{N}(0, \sigma_n^2 \boldsymbol{P})$$

  - Then,

$$\left. \frac{\|\boldsymbol{e}_{\mathrm{lin}}\|^2}{\sigma_n^2} \right| \mathcal{H}_0 \sim \chi_\rho^2(0), \quad \rho = L - R$$

  - The case of the nonlinear estimation is not that simple
  - We argue that under $\mathcal{H}_0$, both GP and linear estimations should achieve the same level of accuracy

➪ We assume that

$$\left. \frac{\|\boldsymbol{e}_{\mathrm{nlin}}\|^2}{\sigma_n^2} \right| \mathcal{H}_0 \sim \chi_\rho^2(0)$$

- Then, under $\mathcal{H}_0$, $\|\boldsymbol{e}_{\text{lin}}\|^2$ and $\|\boldsymbol{e}_{\text{nlin}}\|^2$ are correlated χ^2 variables

- Now, we can write

$$\boldsymbol{e}_{\text{lin}} = \boldsymbol{e}_{\text{nlin}} + \sqrt{2}\boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_\epsilon^2 \boldsymbol{I})$$

- Then, neglecting $\boldsymbol{e}_{\text{nlin}}^\top \boldsymbol{\epsilon}$ under $\mathcal{H}_0$

$$T \approx \frac{\|\boldsymbol{e}_{\text{nlin}}\|^2}{\|\boldsymbol{e}_{\text{nlin}}\|^2 + \|\boldsymbol{\epsilon}\|^2}$$

  where the two χ^2 variables are independent

- This ratio corresponds to a beta distribution [Johnson, 1995] [24]

- Fitted Beta vs Histogram for the test statistics T under $\mathcal{H}_0$;
- Synthetic data generated using the LMM with random abundances sampled from the unity simplex.

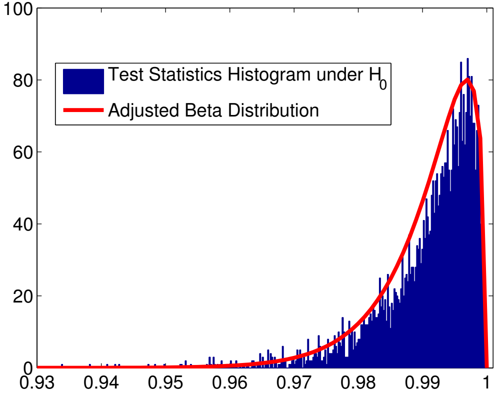


Figure: Histogram of the test statistics under $\mathcal{H}_0$ and the adjusted Beta distribution.

Simulations

Degree of nonlinearity for synthetic data

$$\boldsymbol{r} = \boldsymbol{r}_{\mathrm{lin}} + \boldsymbol{r}_{\mathrm{nlin}}$$

- The energy of $\boldsymbol{r}$ is given by

$$E = \|\boldsymbol{r}\|^2 = \|\boldsymbol{r}_{\mathrm{lin}}\|^2 + 2\,\boldsymbol{r}_{\mathrm{lin}}^\top \boldsymbol{r}_{\mathrm{nlin}} + \|\boldsymbol{r}_{\mathrm{nlin}}\|^2,$$

- $E_{\mathrm{lin}} = \|\boldsymbol{r}_{\mathrm{lin}}\|^2$ is the energy of the linear contribution

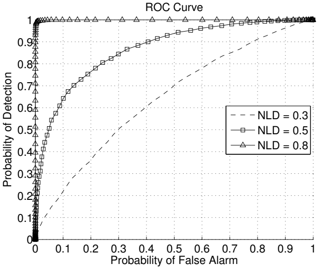
- $E_{\mathrm{nlin}} = 2\,\boldsymbol{r}_{\mathrm{lin}}^\top \boldsymbol{r}_{\mathrm{nlin}} + \|\boldsymbol{r}_{\mathrm{nlin}}\|^2$ is the part of the pixel energy affected by the nonlinear mixing.

- Define η_d (degree of nonlinearity) such that $0 \leq \eta_d \leq 1$

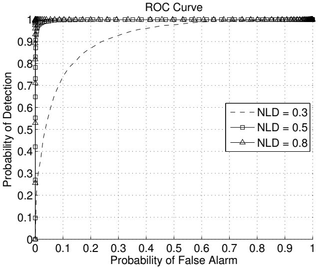
$$\eta_d = \frac{E_{\mathrm{nlin}}}{E} = \frac{1}{1+A}, \quad A = \|\boldsymbol{r}_{\mathrm{lin}}\|^2 / (2\,\boldsymbol{r}_{\mathrm{lin}}^\top \boldsymbol{r}_{\mathrm{nlin}} + \|\boldsymbol{r}_{\mathrm{nlin}}\|^2)$$

Known $\boldsymbol{M}$

- 4000 samples (2000 LMM, 2000 modified GBM) ;
- $R = 3$ endmembers extracted from ENVI$^{\text{TM}}$(green grass, olive green paint and galvanized steel metal);
- fixed abundances $\boldsymbol{\alpha} = [0.3,\ 0.6,\ 0.1]^\top$;
- different η_d (0.3, 0.5, 0.8);
- SNR = 21dB.



(a) Robust LS detector.



(b) Proposed GP detector.

- GP vs LS ($\eta_d = 0.5$)

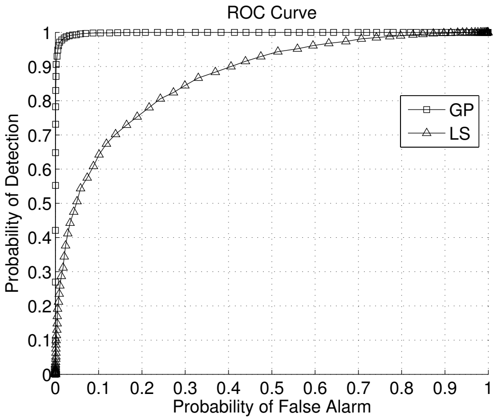


Table: Abundance estimation RMSE for $\boldsymbol{M}$ known and using the GBM mixing model (SNR = 21dB, $\eta_d = 0.5$).

	Image I: LMM + GBM			
Model	FCLS	SK-Hype	D.+U. GP (C.E.%)	D.+U. LS (C.E.%)
LMM	**0.0095**	0.0205	0.0097 (0.6)	0.0096 (0.2)
NLM	0.0624	**0.0312**	0.0324 (5.6)	0.0509 (51.4)
F.Img	0.0446	0.0264	**0.0239** (3.1)	0.0366 (25.8)

Table: Abundance estimation RMSE for $\boldsymbol{M}$ known and using the PNMM mixing model (SNR = 21dB, $\eta_d = 0.5$).

	Image II: LMM + PNMM			
Model	FCLS	SK-Hype	D.+U. GP (C.E.%)	D.+U. LS (C.E.%)
LMM	**0.0095**	0.0205	0.0099 (1.2)	**0.0095** (0)
NLM	0.0958	**0.0440**	0.0443 (0.8)	0.0483 (17)
F.Img	0.0681	0.0344	**0.0321** (1)	0.0348 (8.5)

C.E. - classification error (%)

One-tailed Wilcoxon signed rank test (Sig. level 0.05)

- Null hypothesis

$$\text{median}(\text{RMSE}_{\text{proposed}}) = \text{median}(\text{RMSE}_{\text{other}})$$

- Assign $\mathcal{A}$ if the null hypothesis is rejected.

- Assign "-" if the null hypothesis cannot be rejected.

Table: Image I.

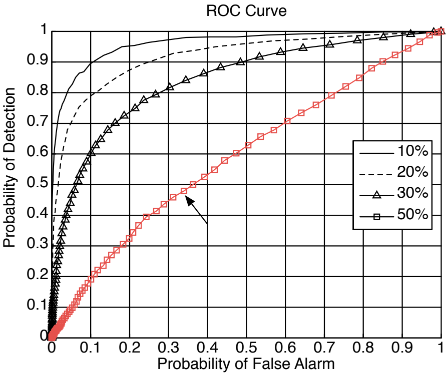
	FCLS	SK-Hype	D.+U. LS
LMM	-	$\mathcal{A}$	-
NLM	$\mathcal{A}$	-	$\mathcal{A}$
F.Img.	$\mathcal{A}$	$\mathcal{A}$	$\mathcal{A}$

Table: Image II.

	FCLS	SK-Hype	D.+U. LS
LMM	-	$\mathcal{A}$	-
NLM	$\mathcal{A}$	-	$\mathcal{A}$
F.Img.	$\mathcal{A}$	$\mathcal{A}$	$\mathcal{A}$

Unknown $\boldsymbol{M}$

- Different proportion of nonlinearly mixed pixels (10-50 %)
- $\eta_d = 0.5$
- $\boldsymbol{M}$ estimated using VCA
- keep in mind the red line (we'll return to it later!)



Iterative Endmember Estimation/Detection in Nonlinearly Mixed HIs

Basic steps:

1. assume a relaxing factor r_f between 0 and 1

2. estimate $\boldsymbol{M}$ using the MVES

3. compute the detection threshold τ and its relaxed version $\tau_r = r_f \times \tau$

4. detect and discard nonlinearly mixed pixels using τ_r

5. make r_f less "relaxing" :-) (i.e., make r_f closer to 1)

6. return to 1 until some stopping criterion on $(T_{\max} - T_{\min})$ is satisfied

Algorithm 1: Iterative endmember estimation

Input : The hyperspectral image $\boldsymbol{R}$, and the number of endmembers R

Output: Estimated endmember matrix $\widehat{\boldsymbol{M}}$

Initialization: $T_{\max} = 1$, $T_{\min} = 0$, $\varepsilon = 0.05$, $\boldsymbol{R}_{\text{tmp}} = \boldsymbol{R}$, $N_{\max} = 10$, $cc = 0$, $r_f = 0.9$, $r_{inc} = (1 - r_f)/N_{\max}$, $\text{PFA} = 0.05$;

$\widehat{\boldsymbol{M}} = \text{MVES}(\boldsymbol{R}_{\text{tmp}}, R)$;

Compute τ using a Beta distribution;

$\tau_r = r_f \times \tau$; %% (relaxed threshold)

while $T_{\max} - T_{\min} > \varepsilon$ *&* $cc < N_{\max}$ **do**

    for $i = 1$ **to** N_{pixels} **do**

        Compute $\boldsymbol{T}(i)$;

    end

    Remove all pixels with $\boldsymbol{T}(i) \leq \tau_r$ from $\boldsymbol{R}_{\text{tmp}}$;

    $r_f = r_f + r_{inc}$; %% (relaxing factor)

    $\tau_r = r_f \times \tau$;

    $T_{\max} = \max(\boldsymbol{T})$; $T_{\min} = \min(\boldsymbol{T})$;

    $cc = cc + 1$;

    $\widehat{\boldsymbol{M}} = \text{MVES}(\boldsymbol{R}_{\text{tmp}}, R)$;

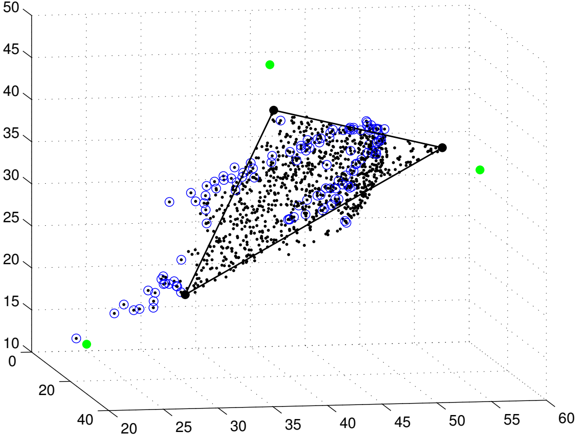
end

Example using synthetic data

- 2000 pixels
- 3 endmembers
- 50% of nonlinearly mixed pixels (randomly selected)
- nonlinearity: GBM with $\eta_d = 0.5$

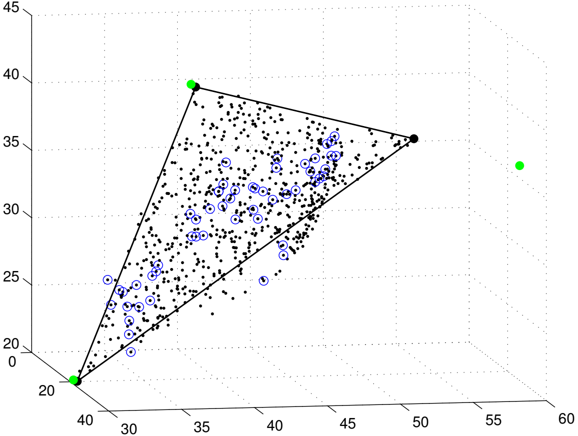
Example using synthetic data (iteration 1)

- Black dots are the real endmembers projection onto $\boldsymbol{M}$
- green dots are the estimated endmembers projection
- blue circles indicate pixels that are being discarded



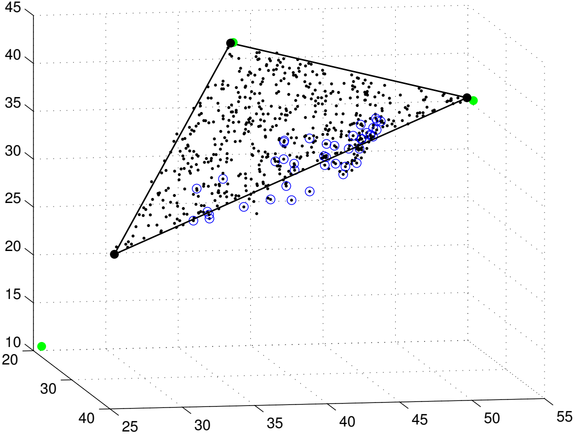
Example using synthetic data (iteration 4)

- Black dots are the real endmembers projection onto $\boldsymbol{M}$
- green dots are the estimated endmembers projection
- blue circles indicate pixels that are being discarded



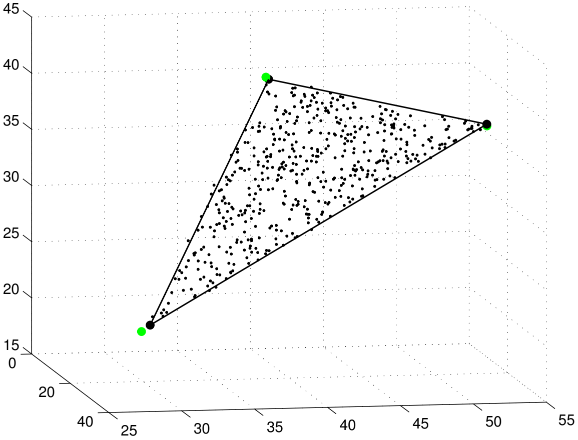
Example using synthetic data (iteration 7)

- Black dots are the real endmembers projection onto $\boldsymbol{M}$
- green dots are the estimated endmembers projection
- blue circles indicate pixels that are being discarded



Example using synthetic data (iteration 10)

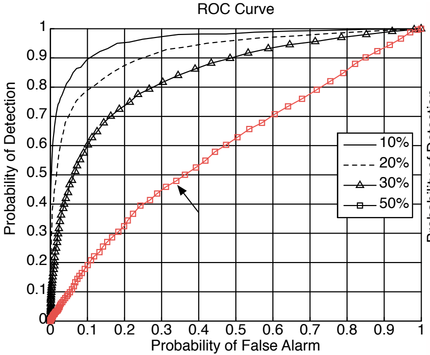
- Black dots are the real endmembers projection onto $\boldsymbol{M}$
- green dots are the estimated endmembers projection
- blue circles indicate pixels that are being discarded



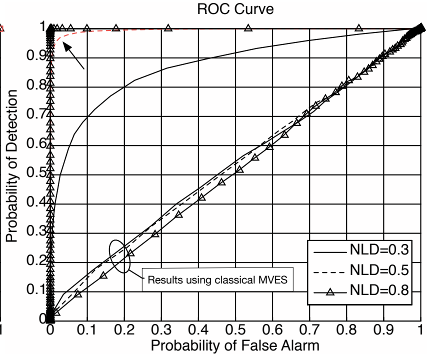
Performance of the proposed estimation method

- Different η_d (NLD in the figure)
- Results using VCA or MVES $\times$ using proposed method

Before



After



Synthetic data extracted from a real scene (Cuprite Mining Field - Nevada - CA)

- Controlled environment with labeled data;
- Alunite Hill is known to have 3 endmembers (alunite, muscovite, and kaolinite) [5];

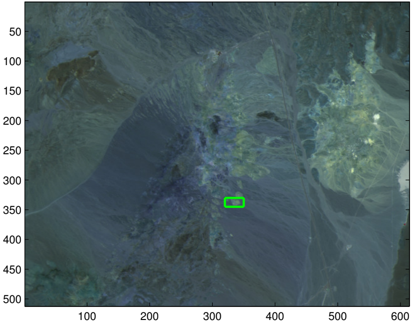
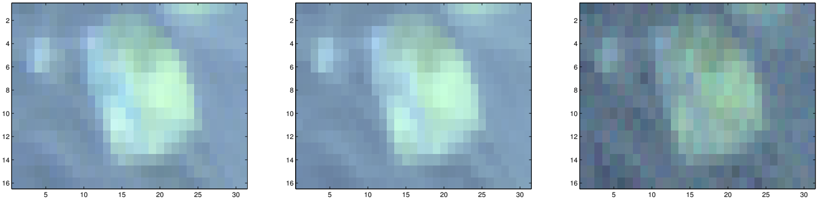


Figure: Cuprite mining site. The green box corresponds to the alunite hill scene.

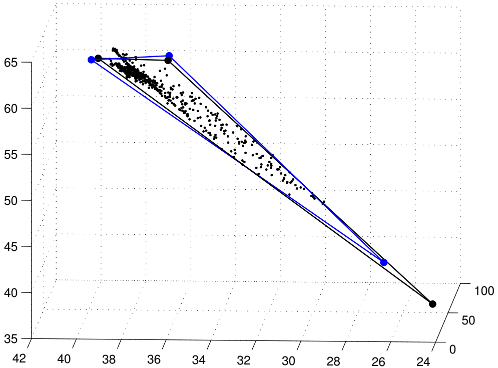


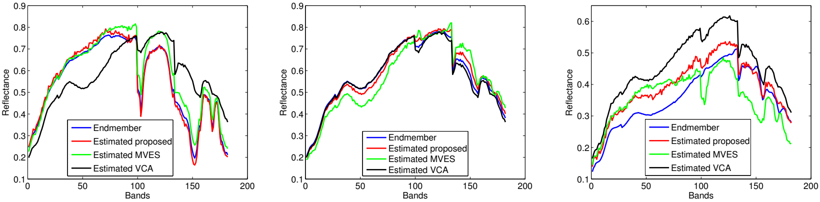
(a) Alunite hill. (b) LMM. (c) GBM + WGN.

Figure: (a) Plot of the alunite hill with bands 30, 70 and 100. (b) Reconstruction of the scene using the LMM. (c) Adding 30 % of nonlinearly mixed pixels ($\eta_d = 0.3$) and WGN to give a 30dB SNR.

Estimated endembers

- black circles (true endembers);
- blue circles (estimated using the prop. method) after 10 iterations.





(a) Alunite. (b) Kaolinite. (c) Muscovite.

Figure: Endmember estimations for the nonlinearly mixed image with different extraction techniques.

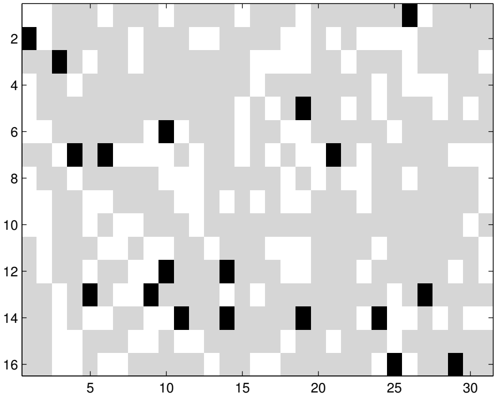


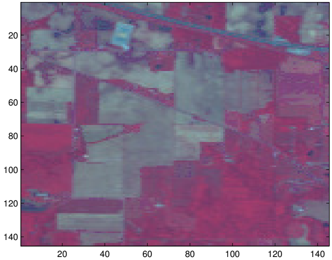
Figure: Detection map and true nonlinear map. Linearly mixed pixels in gray, nonlinearly mixed pixels in white, and misclassified pixels in black.

Table: RMSE for the abundances in the alunite hill scene.

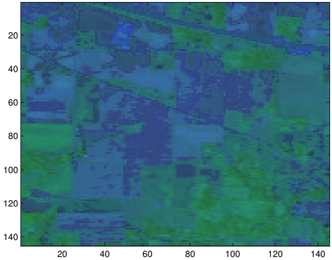
Algorithm	RMSE $\pm$ STD (Class. Err. %)
FCLS	0.0797 ± 0.0123 (-)
SK-Hype	0.0824 ± 0.0059 (-)
detect-then-unmix	**0.0671 ± 0.0049** (3.83)

Real Data 1: Indian Pines

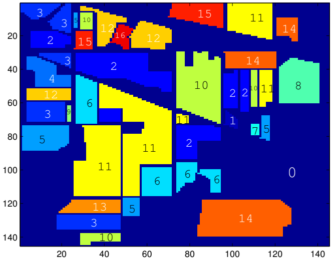
- 16 non-mutually exclusive classes;
- divided in 8 subimages by grouping classes with similar number of pixels;



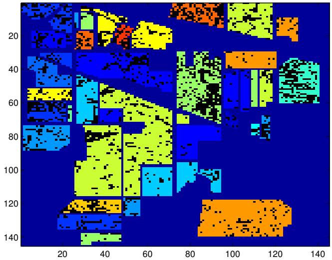
(a) 3-band combination #1



(b) 3-band combination #2



(a) Ground truth.

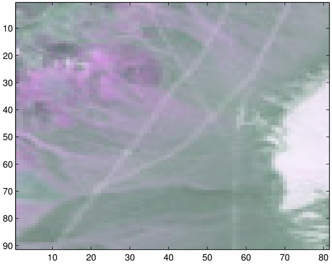


(b) Detection map.

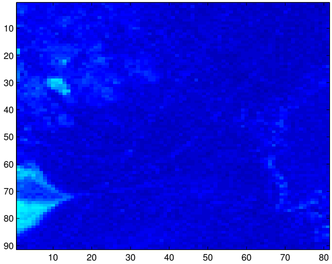
Table: Indian Pines recontruction error (RMSE) by subimage.

Subimg.	RMSE $\pm$ STD		
	FCLS	SK-Hype	detect-then-unmix
1	**0.0028627**	0.0030332	0.0029083
2	0.0038963	0.003881	**0.0038391**
3	0.0044259	0.0035981	**0.0035537**
4	0.0040145	0.0039097	**0.0038895**
5	0.0030848	0.0032353	**0.0030527**
6	0.0039905	0.004055	**0.0039644**
7	0.0034804	0.0035049	**0.0034552**
8	0.0037665	0.0039314	**0.0037531**

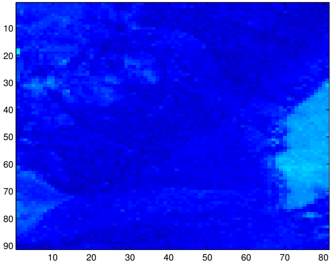
Real Data 2: Cuprite



(a) Cuprite scene.

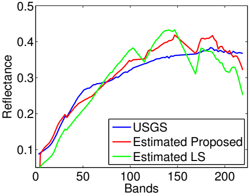


(b) Prop. EEA + SK-Hype

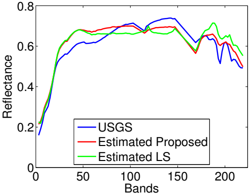


(c) VCA + SK-Hype

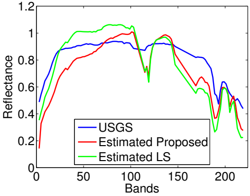
Estimated endmembers and USGS spectra



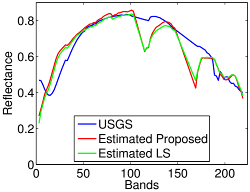
(a) Sphene



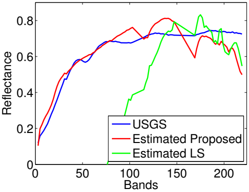
(b) Montmorillonite



(c) Kaolinite

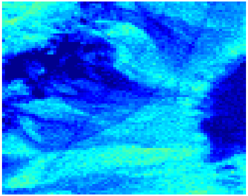


(d) Dumortierite

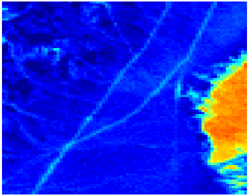


(e) Pyrope

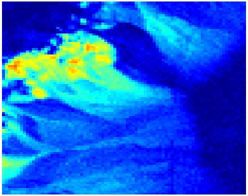
Estimated Abundance Maps



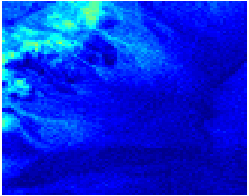
(a) Sphene



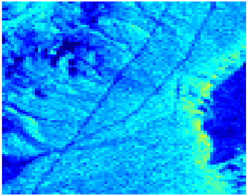
(b) Montmorillonite



(c) Kaolinite



(d) Dumortierite



(e) Pyrope

Table: Spectral angles (in rad) between estimated and USGS spectra.

Endmemeber	IEE	LS	VCA	MVES
Sphene	**0.0799**	0.1498	0.3634	0.2457
Montmorillonite	**0.0615**	0.0852	0.0888	0.0762
Kaolinite	**0.1471**	0.1689	0.2022	0.2559
Dumortierite	0.1054	0.1008	**0.0942**	0.1422
Pyrope	**0.1035**	0.9792	0.1760	0.1588

Final remarks

- The proposed nonparametric method for detecting nonlinear mixtures in HIs outperforms other nonparametric detection methods in the literature
- The unmixing performance shows improvement when compared to state-of-the-art methods
- The unmixing results are statistically consistent
- The degree of mixture nonlinearity was defined allowing one to compare results using different models
- The iterative EEA algorithm proposed presented good results when compared to traditional VCA and MVES
- Simulations using different scenarios corroborate the conclusions

- Main results published in [25] and [26]

- source code available at `https://github.com/talesim/NP_NL_Det_EE_HI/archive/master.zip`

References

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

[2] D. C. Heinz and C.-I. Chang,
"Fully constrained least squares linear mixture analysis for material quantification in hyperspectral imagery,"
IEEE Transactions on Geoscience and Remote Sensing, vol. 39, no. 3, pp. 529–545, 2001.

[3] P. Honeine and C. Richard,
"Geometric unmixing of large hyperspectral images: A barycentric coordinate approach,"
IEEE Transactions on Geoscience and Remote Sensing, vol. 50, no. 6, pp. 2185–2195, 2012.

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

[5] N. Dobigeon, S. Moussaoui, M. Coulon, J.-Y. Tourneret, and A. O. Hero,
"Joint bayesian endmember extraction and linear unmixing for hyperspectral imagery,"
IEEE Transactions on Signal Processing, vol. 57, no. 11, pp. 4355–4368, 2009.

[6] B. Hapke,
"Bidirectional reflectance spectroscopy, 1, Theory,"
Journal of Geophysical Research, vol. 86, no. B4, pp. 3039–3054, 1981.

[7] W. Fan, B. Hu, J. Miller, and M. Li,
"Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data,"
International Journal of Remote Sensing, vol. 30, no. 11, pp. 2951–2962, 2009.

[8] J. M. P. Nascimento and J. M. Bioucas-Dias,
"Nonlinear mixture model for hyperspectral unmixing,"
in *SPIE Europe Remote Sensing*. International Society for Optics and Photonics, 2009, pp. 74770I–74770I.

[9] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret,
"Nonlinear Unmixing of Hyperspectral Images Using a Generalized Bilinear Model,"
IEEE Transactions on Geoscience and Remote Sensing, vol. 49, no. 11, pp. 4153–4162, Nov. 2011.

[10] J. Chen, C. Richard, and P. Honeine,
"Estimating abundance fractions of materials in hyperspectral images by fitting a post-nonlinear mixing model,"
in *Proc. IEEE WHISPERS*, 2013, pp. 1–4.

[11] Y. Altmann, N. Dobigeon, and J.-Y. Tourneret,
"Nonlinearity detection in hyperspectral images using a polynomial post-nonlinear mixing model.,"
IEEE Transactions on Image Processing, vol. 22, no. 4, pp. 1267–1276, 2013.

[12] N. Dobigeon, J.-Y. Tourneret, C. Richard, J. C. M. Bermudez, S. McLaughlin, and A. O. Hero,
"Nonlinear unmixing of hyperspectral images: Models and algorithms,"
IEEE Signal Processing Magazine, vol. 31, no. 1, pp. 82–94, Jan 2014.

[13] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret,
"Unmixing hyperspectral images using the generalized bilinear model,"
in *2011 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. 2011, pp. 1886–1889, IEEE.

[14] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret,
"Residual component analysis of hyperspectral images–application to joint nonlinear unmixing and nonlinearity detection.,"
IEEE Transactions on Image Processing, vol. 23, pp. 2148–2158, May 2014.

[15] R. Heylen and P. Scheunders,
"Calculation of geodesic distances in nonlinear mixing models: Application to the generalized bilinear model,"
Geoscience and Remote Sensing Letters, IEEE, vol. 9, no. 4, pp. 644–648, 2012.

[16] J. Chen, C. Richard, and P. Honeine,
"Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model,"
IEEE Transactions on Signal Processing, vol. 61, pp. 480–492, Jan 2013.

[17] Y. Altmann, N. Dobigeon, S. McLaughlin, and J.-Y. Tourneret,
"Nonlinear spectral unmixing of hyperspectral images using gaussian processes,"
IEEE Transactions on Signal Processing, vol. 61, pp. 2442–2453, May 2013.

[18] N. H. Nguyen, C. Richard, P. Honeine, and C. Theys,
"Hyperspectral image unmixing using manifold learning methods: derivations and comparative tests,"
in *2012 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 2012.

[19] C. C. Borel and S. A. W. Gerstl,
"Nonlinear spectral mixing models for vegetative and soil surfaces,"
Remote Sensing of Environment, vol. 47, no. 3, pp. 403–416, Jan 1994.

[20] T. Han and D. G. Goodenough,
"Investigation of nonlinearity in hyperspectral imagery using surrogate data methods,"
IEEE Transactions on Geoscience and Remote Sensing, vol. 46, pp. 2840–2847, Oct 2008.

[21] Y. Altmann, N. Dobigeon, J.-Y. Tourneret, and J. C. M. Bermudez,
"A robust test for nonlinear mixture detection in hyperspectral images,"
in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013.

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

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

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

[25] T. Imbiriba, J. C. M. Bermudez, J.-Y. Tourneret, and C. Richard,
"Detection of nonlinear mixtures using gaussian processes: Application to hyperspectral imaging,"
in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Jan 2014, pp. 7949–7953.

[26] T. Imbiriba, J. C. M. Bermudez, C. Richard, and J.-Y. Tourneret,
"Nonparametric detection of nonlinearly mixed pixels and endmember estimation in hyperspectral images,"
IEEE Transactions on Image Processing, vol. 25, no. 3, pp. 1136–1151, March 2016.