SAND2017-3675C

Supervised non-negative tensor factorization for automatic hyperspectral feature extraction and target discrimination

Dylan Anderson[†], Aleksander Bapst[†], Joshua Coon[†], Aaron Pung[†], Michael Kudenov[‡]

† Sandia National Laboratories

‡ North Carolina State University

Sandia National Laboratories is a multi-mission laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. SAND NO. 2011-XXXXP

Overview

Outline

1. Problem definition and background
2. Tensor factorization
3. Supervised tensor factorization
4. Learning algorithm
5. Experimental results

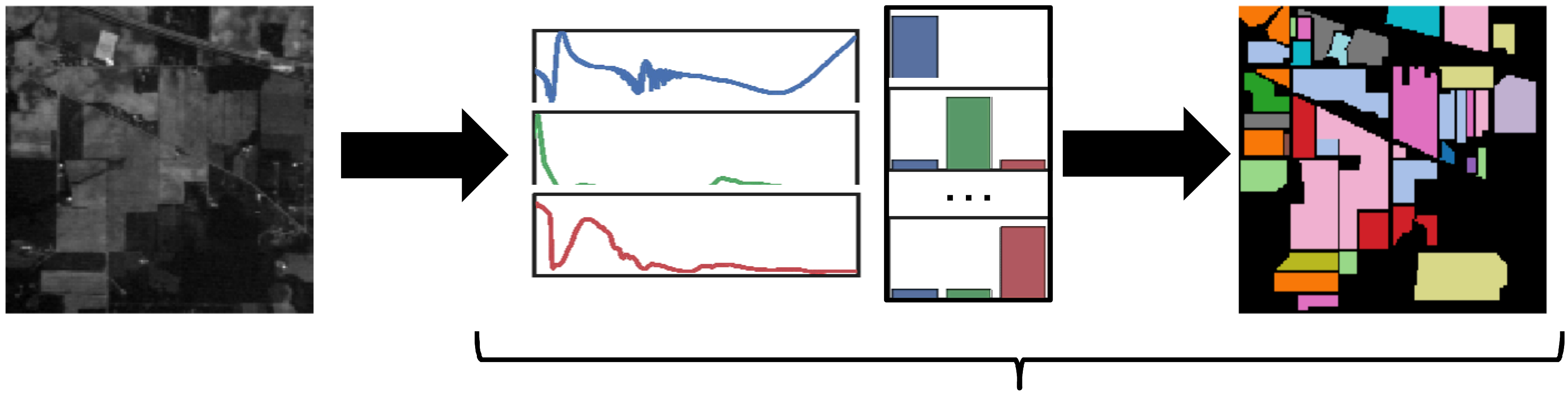
Key Attributes

- Jointly performs dimensionality reduction and classification
- Fully linear
- Non-negative constraints
- Extends to higher-order data (tensors)
- Decision function emulates multi-spectral sensor

Hyperspectral Classification

- Goal: Classification
- Classifying in hyperspectral space is prone to curse of dimensionality, overfitting
- Dimensionality reduction (DR) prior to classification

Hyperspectral Input | Dimensionality Reduction | Classification



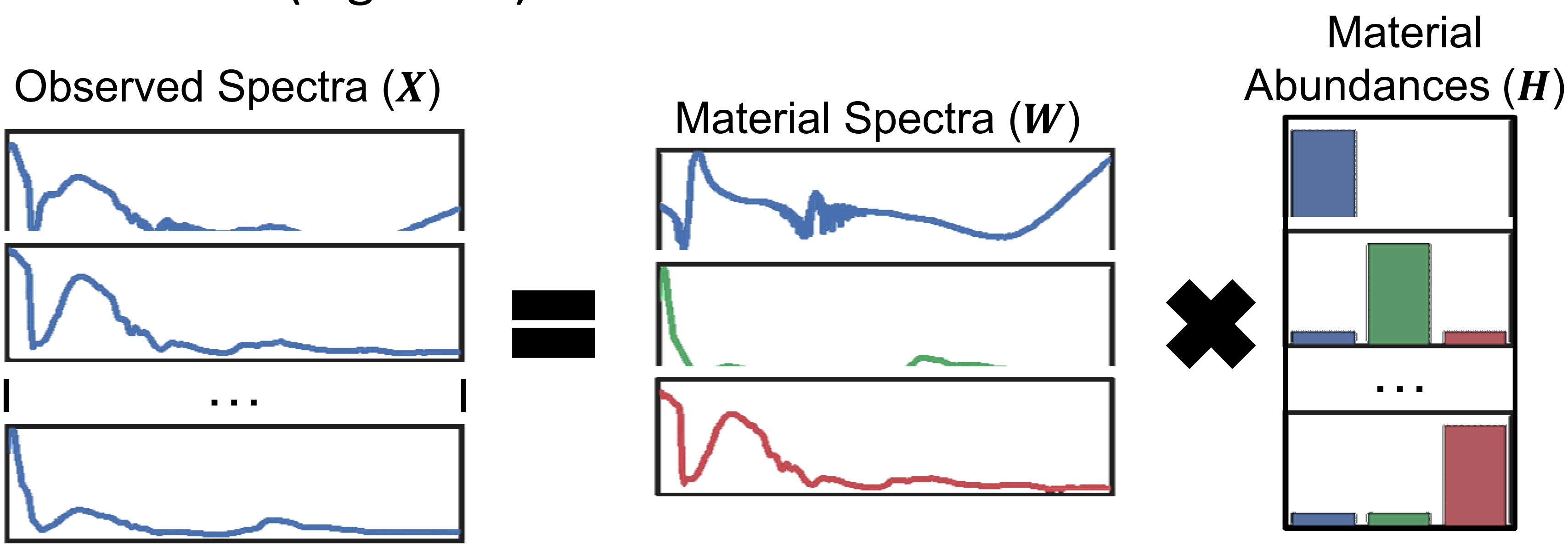
SNTF performs these two steps jointly

Matrix Factorization

- Linear DR can be cast as matrix factorization

$$\boldsymbol{X} \approx \boldsymbol{W}\boldsymbol{H}$$

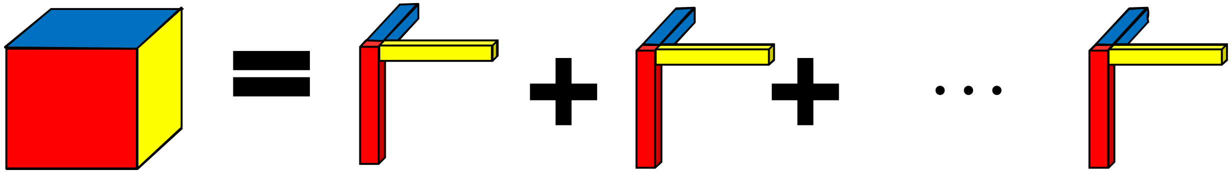
- Most common are Singular Value Decomposition (SVD) and Non-negative Matrix Factorization (NMF)
- Resulting abundances/mixing coefficients used to train a classifier (e.g. SVM)



Tensor Factorization

- CANDECOMP (also called PARAFAC)
  - Generalization of SVD to higher-order tensors
  - Approximates data-tensor as summation of rank-1 tensors:

$$\boldsymbol{\mathcal{X}} = \sum_{k=1}^{K} \boldsymbol{a}_k^{(1)} \circ \ldots \circ \boldsymbol{a}_k^{(N)} = [\![\boldsymbol{A}^{(1)}, \ldots, \boldsymbol{A}^{(N)}]\!]$$



- Imposing non-negativity (NTF) extends spectral unmixing framework to higher-order tensors
  - Spatial, temporal, polarimetric, etc.

$$D_{KL}\left(\boldsymbol{\mathcal{X}} \parallel [\![\boldsymbol{A}^{(1)}, \ldots, \boldsymbol{A}^{(N)}]\!]\right) = \sum_{\boldsymbol{i}} \left(x_{\boldsymbol{i}} \ln \frac{x_{\boldsymbol{i}}}{[\![\boldsymbol{A}^{(1)}, \ldots, \boldsymbol{A}^{(N)}]\!]_{\boldsymbol{i}}} - x_{\boldsymbol{i}} + [\![\boldsymbol{A}^{(1)}, \ldots, \boldsymbol{A}^{(N)}]\!]_{\boldsymbol{i}} \right)$$

NTF Learning

- Objective is convex per factor (set of N optimizations)
- Alternating minimization strategy

$$\left\{ \min_{\boldsymbol{A}^{(n)} > 0} D_{KL} \left(\boldsymbol{\mathcal{X}} \parallel [\![\boldsymbol{A}^{(1)}, \ldots, \boldsymbol{A}^{(N)}]\!] \right) \right\}_{n=1}^{N}$$

- Extend Lee-Seung NMF multiplicative update rule (MUR)
  - Alternate over each NTF factor matrix
  - Non-negative initialization guarantees non-negativity

$$\boldsymbol{A}^{(n)} \leftarrow \boldsymbol{A}^{(n)} \circledast \left[\left(\boldsymbol{X}_{(n)} \oslash \hat{\boldsymbol{X}}_{(n)} \right) \boldsymbol{A}^{\odot -n} \right]$$

NTF Regularization

- Regularization penalties improves spectral factor interpretability, classification performance

Decorrelating	Smoothing
• Inner product between different spectral bands (correlation) • Factors specialize to different spectral regions $\alpha_{cr} Tr({\boldsymbol{A}^{(b)}}^T \mathbf{1}_{B\times B} \boldsymbol{A}^{(b)})$	• Finite central difference approximation • Smooth out thin spectral features, ringing $\frac{\alpha_{sm}}{2} \lVert \boldsymbol{L}\boldsymbol{A}^{(b)} \rVert_F^2$ $\boldsymbol{L} = \begin{bmatrix} -1 & 2 & -1 & & & 0 \\ & -1 & 2 & -1 & & \\ \vdots & & \ddots & \ddots & \ddots & \\ 0 & & & -1 & 2 & -1 \end{bmatrix}$

- Regularized NTF MUR:

$$\boldsymbol{A}^{(b)} \leftarrow \boldsymbol{A}^{(b)} \circledast \frac{\left(\boldsymbol{X}_{(n)} \oslash \hat{\boldsymbol{X}}_{(n)}\right) \boldsymbol{A}^{\odot -n}}{\mathbf{1}\mathbf{1}^T \boldsymbol{A}^{\odot -n} + \alpha_{sm} \boldsymbol{L}^T \boldsymbol{L} \boldsymbol{A}^{(b)} + \alpha_{cr} \boldsymbol{A}^{(b)} \mathbf{1}_{K\times K}}$$

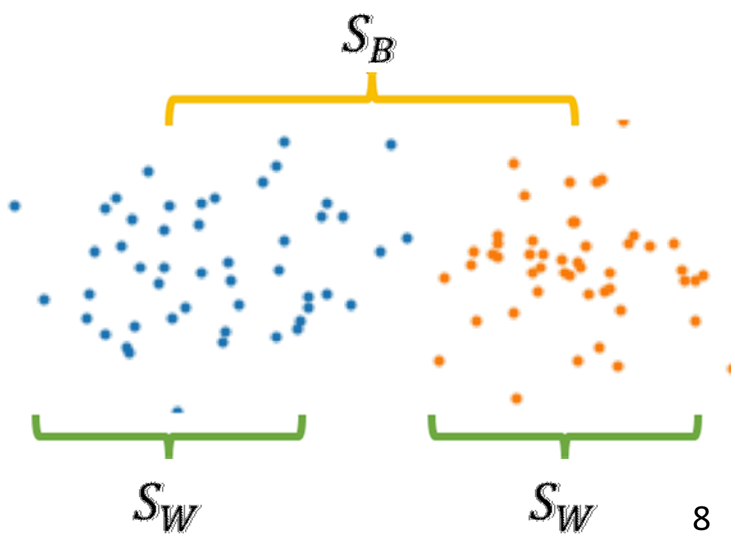
Supervised NTF

- Learn discriminative spectral factors applied as filters/bands

$$\left[\boldsymbol{\mathcal{X}} \times_b \boldsymbol{A}^{(b)} \right]_{(b)} = {\boldsymbol{A}^{(b)}}^T \boldsymbol{X}_{(b)}$$

- Fisher linear discriminant criterion:
  - Samples of the same class are close together (within-class scatter, $\boldsymbol{S}_w$)
  - Classes are far apart (between-class scatter, $\boldsymbol{S}_b$)
- Find spectral factors that minimize within-class scatter and maximize between class scatter

$$\min_{\boldsymbol{A}^{(b)} > 0} Tr\left({\boldsymbol{A}^{(b)}}^T (\lambda \boldsymbol{S}_w - \boldsymbol{S}_b) \boldsymbol{A}^{(b)} \right)$$



- Implies probabilistic classifier:
  - Gaussian mixture model with linked covariance (linear boundary)

Learning SNTF

- NTF objective function with decorrelating, smoothing penalties and Fisher discriminant criterion:

$$\min_{\boldsymbol{A}^{(n)}>0} \Big[D_{KL}\left(\boldsymbol{\mathcal{X}} \parallel [\![\boldsymbol{A}^{(1)},\ldots,\boldsymbol{A}^{(N)}]\!]\right) + \frac{\alpha_{sm}}{2}\|\boldsymbol{L}\boldsymbol{A}^{(b)}\|_F^2 + \alpha_{cr}Tr(\boldsymbol{A}^{(b)^T}\mathbf{1}_{K\times K}\boldsymbol{A}^{(b)}) + \frac{\alpha}{2}Tr\left(\boldsymbol{A}^{(b)^T}(\lambda\boldsymbol{S}_w - \boldsymbol{S}_b)\boldsymbol{A}^{(b)}\right)\Big] \; \forall n$$

- Use positive minus negative terms to split into two positive matrices:

$$\lambda\boldsymbol{S}_w - \boldsymbol{S}_b = [\lambda\boldsymbol{S}_w - \boldsymbol{S}_b]_+ - [\boldsymbol{S}_b - \lambda\boldsymbol{S}_w]_+$$

- Updated MUR for spectral factors:

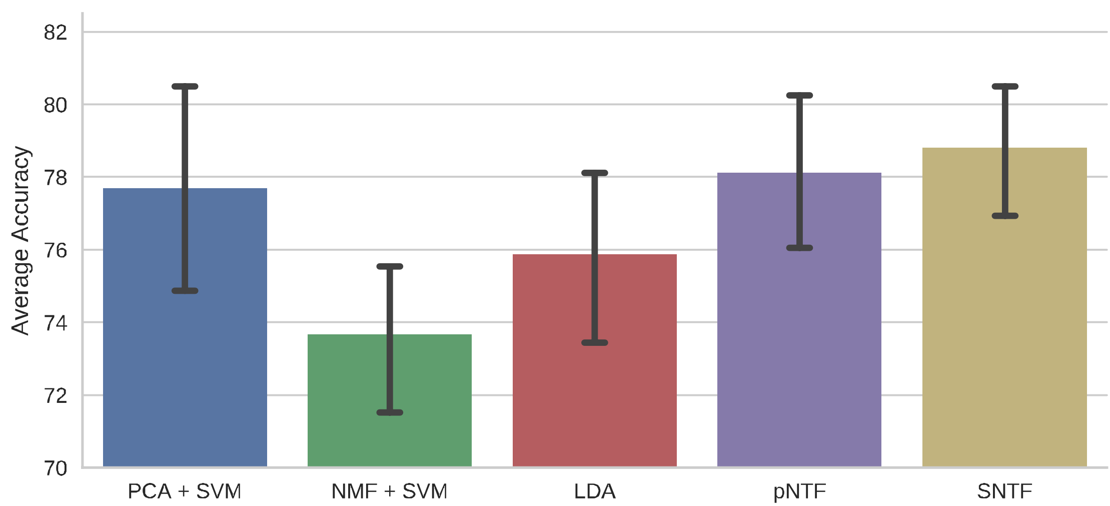
$$\boldsymbol{A}^{(b)} \leftarrow \boldsymbol{A}^{(b)} \circledast \frac{\left(\boldsymbol{X}_{(n)} \oslash \hat{\boldsymbol{X}}_{(n)}\right)\boldsymbol{A}^{\odot -n} + \alpha\,[\boldsymbol{S}_b - \lambda\boldsymbol{S}_w]_+}{\mathbf{1}\mathbf{1}^T\boldsymbol{A}^{\odot -n} + \alpha\,[\lambda\boldsymbol{S}_w - \boldsymbol{S}_b]_+ + \alpha_{sm}\boldsymbol{L}^T\boldsymbol{L}\boldsymbol{A}^{(b)} + \alpha_{cr}\boldsymbol{A}^{(b)}\mathbf{1}_{K\times K}}$$

Classification

- Indian Pines Agricultural HSI Dataset
  - 25% used as training, 10 repeated trials
  - Evaluated with average accuracy (imbalanced dataset)
- pNTF: NTF with regularization
- SNTF: NTF with Fisher criterion and regularization

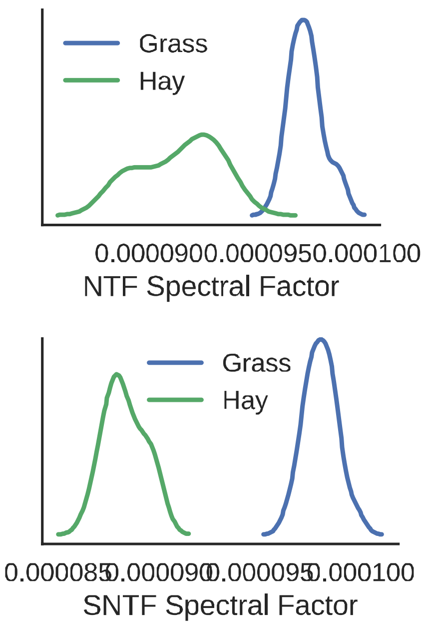
State-of-the-art performance with non-negativity and linear classifier

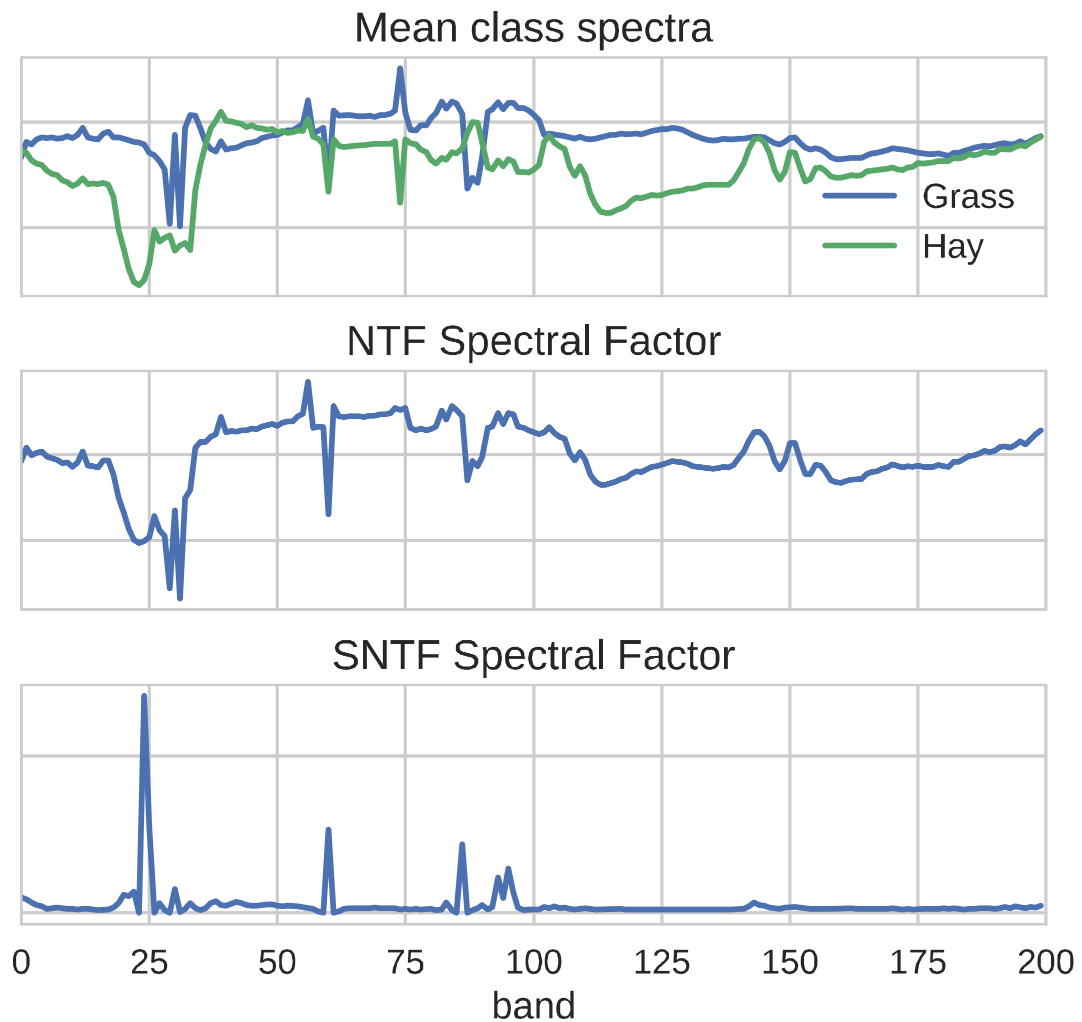
SNTF provides greatest benefit when number of spectral factors is limited (see manuscript).



Supervised Feature Extraction

- Two-class subset
  - “grass-pasture-mowed”
  - “hay-windrowed”
- Single spectral factor





SNTF emphasizes spectral features that separate classes.

Key Attributes of SNTF

- Jointly performs dimensionality reduction and classification
  - Spectral factors are better suited for classification, increased discriminative information in each factor
- Fully linear
  - Model results are easily understood and analyzed
- Non-negative constraints
  - Spectral factors are interpretable, physical
- Extends to higher-order data (tensors)
  - Improved performance and reduced data requirements in tensor settings
- Decision function emulates multi-spectral sensor
  - Task-specific sensor (multispectral bands) design tool

Thank you

Supervised non-negative tensor factorization for automatic hyperspectral feature extraction and target discrimination

Dylan Anderson
Sandia National Laboratories
dzander@sandia.gov