Low-Rank Nonnegative Factorizations for Spectral Imaging Applications

Bob Plemmons
Wake Forest University

• Collaborators:
  Christos Boutsidis (U. Patras), Misha Kilmer (Tufts), Peter Zhang, Paul Pauca, (WFU)

• Interaction on spectral data with: Kira Abercromby (NASA-Houston)

• Related Papers at: http://www.wfu.edu/~plemmons

• Project Funded by AFOSR

Stanford Workshop on Algorithms for Modern Massive Data Sets, June 06
<table>
<thead>
<tr>
<th>Outline</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Object Identification from Spectral Data</strong></td>
</tr>
<tr>
<td><strong>Features-Based Clustering &amp; Classification</strong></td>
</tr>
<tr>
<td><strong>Nonnegativity Constrained Low-Rank Approximation for Blind Source Separation and Unsupervised Unmixing (ill-posed, nonlinear inverse problem)</strong></td>
</tr>
<tr>
<td><strong>Nonnegative Matrix Factorization (NMF)</strong></td>
</tr>
<tr>
<td><strong>Results using Air Force data from Maui and data from K. Abercromby at NASA JSC</strong></td>
</tr>
<tr>
<td><strong>Preliminary Results on using Perron-Frobenius Theory to Compress Hyperspectral Sensor Data</strong></td>
</tr>
<tr>
<td><strong>Comments on Nonnegative Tensor Factorization (NTF) for Image data (see poster by Christos Boutsidis)</strong></td>
</tr>
</tbody>
</table>
Simple Analog Illustration

Hidden Components in Light \(\downarrow\) Separated by a Prism

From Newton’s Notebook

Our purpose \(\downarrow\) finding hidden components by data analysis
Blind Source Separation for Finding Hidden Components (Endmembers)

Mixing of Sources

\[ \mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n] \]

\( \mathbf{X} \) is column vectors (1-D spectral scans)

Approximately factor

\[ \mathbf{X} \approx \mathbf{WH} = \sum_{1}^{k} \mathbf{w}^{(j)} \pm \mathbf{h}^{(j)} \]

\( \pm \) denotes outer product

\( \mathbf{w}^{(j)} \) is jth col of \( \mathbf{W} \), \( \mathbf{h}^{(j)} \) is jth col of \( \mathbf{H}^{T} \)

\( \mathbf{X} \) sensor readings (mixed components, observed data)

\( \mathbf{W} \) separated components (feature basis matrix, unknown, low rank)

\( \mathbf{H} \) hidden mixing coefficients (unknown), replaced later with abundances of materials that make up the object.
Typical Scan
NMF Allows only additive, not subtractive combinations of the original data, in comparison to orthogonal decomposition methods, e.g. PCA.


Historical perspective:

Problem 73-14, Rank Factorization of Nonnegative Matrices, by A. Berman and R.J. Ple., SIAM Review 15 (1973), p. 655: (Also in Berman/Ple. book)
Some General Applications of NMF Techniques

- Source separation in acoustics, speech, video
- EEG in Medicine, electric potentials
- Spectroscopy in chemistry
- Molecular pattern discovery - genomics
- Thermal nondestructive testing - aircraft and missile parts
- Email surveillance
- Document clustering in text data mining
- Atmospheric pollution source identification
- Hyperspectral sensor data compression
- **Spectroscopy for space applications** and spectral data mining
  - Identifying object surface materials and substances
Space Object Identification and Characterization from Spectral Reflectance Data

More than 15,000 known objects in orbit: various types of military and commercial satellites, rocket bodies, residual parts, and debris — need for space object database mining, object identification, clustering, classification, etc.
Maui Space Surveillance Site
Satellite Pass

Atmospheric Distortion

r0
Imaging Sciences for Space Situational Awareness by Monitoring Space Satellites (AFOSR)

- Listen (laser enabled vibrometry)
- Smell (chemical sensing with spectrometer)
- Touch (scatterometry/polarimetry for surface texture information)
- See (by sequential speckle <video> imaging)
- Characterize materials (spectral imaging)
The creation and observation of a reflectance spectrum
Spectral Imaging of Space Objects

- Current "operational" capability for spectral imaging of space objects
- Panchromatic images
- Non-imaging spectra
Why look at anything spectrally?
• Simple answer: Color vs. Black and white
• More involved answer: Spectral radiometry

For space objects we are looking at being able to:
• Differentiate between different material classes
• Material degradation
• Identify hidden payloads
• Anomaly resolution
Overview of the SOI Problem

- Space activities require accurate information about orbiting objects for space situational awareness

Many objects are either in

- Geosynchronous orbits (about 40,000 KM from earth), or
- Near-Earth orbits, but too small (e.g., space mines) to be resolved by optical imaging systems

Can approximately collect one pixel/object by optical telescope
Overview of the SOI Problem Continued

- Problem solution by learning the parts of objects (hidden components) by low rank nonnegative sparse representation.

- Basis representation (dimension reduction) can enable near real-time object (target) recognition, object class clustering, and characterization. (ill-posed inverse problem)

- Match recovered hidden components with known spectral signatures from substances such as mylar, aluminum, white paint, kapton, and solar panel materials, etc. This is classification.

- Fundamental difficulty: Find from spectral measurements:
  - Endmembers: types of constituent materials
  - Fractional abundances: proportion of materials that comprise the object.
Approximate NMF

• Utilize constraint that sensor data values in $X$ are nonnegative.

• Apply non-negativity constrained low rank approximation for blind source separation, dimension reduction and unsupervised unmixing.

• Low rank approximation to data matrix $X$:

\[ X \approx WH, \quad W, H \geq 0 \]

- Columns of $W$ are initial basis vectors for spectral trace database, may want smoothness and statistical independence in $W$.

- Columns of $H$ represent mixing coefficients, desire statistical sparsity in $H$ to force essential uniqueness in $W$. May want sparsity for $H$. 
Some of our Data Obtained from a Spica (Space Infrared Telescope for Cosmology and Astrophysics) Spectrometer

- **Mission:** Support non-imaging SOI with spectroscopic observations
- **3 \(\text{Å}4\) angstrom resolution**
  - **Blue** mode: 3000 \(\text{Å}\)–6000 angstroms (.3 \(\text{Å}\)–.6 m)
  - **Red** mode: 6000 \(\text{Å}\)–9000 angstroms (.6 \(\text{Å}\)–.9 m)
- **Located on Maui**
Sample Raw Data Collected in Blue and Red Modes
Electromagnetic Spectrum: Spectral Signatures

- For any given material, the amount of solar (or other) radiation that it reflects, absorbs, or transmits varies with wavelength.

- This property of matter makes it possible to identify different substances out of 300+ and separate them by their spectral signatures (spectral curves) via spectral unmixing, finding fractional abundances.
An Approach to Finding Endmembers and Fractional Abundances

- **Vectorize** the spectral scans of space objects into columns of $X$ (works well for 1-D signals, not for 2-D images)
- **Cluster** the columns of $X$ using a NMF scheme
  \[ X \approx WH, \quad W \beta^2 \text{uo qqvj} + H \beta^2 \text{ur ctug} \]
  (We use a metric by Hoyer to enforce sparsity in $H$.)

\[
\text{sparseness}(x) = \frac{\sqrt{n} - \|x\|_1}{\|x\|_2} \div \sqrt{n - 1}
\]
Parts- Based Clustering & Classification

- Features from hidden components: parts-based learning algorithms from training set data

- Utilize constraint that spectral trace reflectance values are nonnegative

- Arrange the spectral traces into columns of a (nonnegative) database matrix denoted by $X$

- Non-negativity constrained low rank approximation for blind source separation and unsupervised unmixing

- Low rank approximation to data matrix $X$: $X \approx WH$, $W \beta^2, H \beta^2$
  - Columns of $W$ are basis vectors for spectral trace database (endmembers)
  - $H$ eventually discarded and new reduced $H$ computed
  - Alternating iterations used
NMF Problem Formulation

- Given initial database expressed as \( n \times m \) nonnegative matrix \( X \)

find two reduced-dimensional matrices \( W (n \times r) \) and \( H (r \times m) \) to:

\[
\min_{W,H} \| X - WH \|_F^2,
\]

plus constraints

where \( W_{ij} \beta^{2} \) and \( H_{ij} \beta 0 \) for each \( i \) and \( j \). Choice of \( r \ll m \) is often problem dependent. Can impose other (e.g., smoothness, sparsity) constraints on \( W \) and/or \( H \).
NMF - Continued

- Can use convex cone theoretic geometric concepts to determine conditions for uniqueness, up to permutation and scaling of the rows (Donoho and Stodden).

- Constraints on $H$ strongly affect uniqueness in $W$. 
Lee and Seung (1999) proposed a multiplicative alternating iteration scheme

1. Initialize $W$ and $H$ with nonnegative values and scale columns of $W$ to unit norm.
2. Iterate for each $c$, $j$ and $i$ until convergence or stop ($\epsilon$ is a machine dependent small positive pos. no.):

   (a) $H_{cj} \leftarrow H_{cj} \frac{(W^T X)_{cj}}{(W^T W H)_{cj} + \epsilon}$

   (b) $W_{ic} \leftarrow W_{ic} \frac{(X H^T)_{ic}}{(W H H^T)_{ic} + \epsilon}$

   (c) Scale the columns of $W$ to unit norm.

   - Process is essentially a diagonally-scaled gradient descent method of EM (R-L) type.

But, clustering is ill-posed. Regularization may be needed.
New Approach to Selecting Endmembers and Computing Fractional Abundances

- **Vectorize** the spectral scans of space objects into columns of a matrix $Y$

- **Cluster** the columns of $Y$ using a NMF scheme
  
  $$Y \approx WH, \quad W \beta^2, \quad H \beta^2$$

  (Enforce smoothness on $W$ and sparsity on $H$.)

- **Classify** the basis vectors in $W$ using lab data from Jorgersen and an information theoretic scoring method (Kullback-Leibler divergence, i.e., relative entropy). Represent these endmembers by a matrix $B$.

- $B$ represents a compressed database for $Y$ and has a variety of uses, e.g., ....

- Determine the **spectral abundances** of the space object spectral scans in columns of $Y$ by iteratively solving nonlinear least squares problems with matrix $B$ containing the classified endmembers.

  (We use a nonlinear least squares scheme to compute material abundancies.)
Minimize a functional $F(W, H)$ by solving the following constrained optimization problem. (Here $\chi$ and $\delta$ are regularization parameters).

$$\min_{W,H} \left\{ \|Y - WH\|_F^2 + \alpha J_1(W) + \beta J_2(H) \right\}, \text{ for } W \geq 0 \text{ and } H \geq 0$$

where $\alpha J_1(W)$ and $\beta J_2(H)$ are used to enforce certain application-dependent characteristics on the solution.

Determine gradients for $W$ and $H$ and set each to zero (alternating iterations).
Sparse CNMF

We define sparseness of a vector $x$ of length $n$ as

$$\text{sparseness}(x) = \frac{\sqrt{n} - \|x\|_1}{\sqrt{n} - 1}$$

Given $A$, a matrix of arbitrary size, let

$$\tilde{A} = \text{vec}(A),$$

denote the vector formed by stacking the columns of $A$.

Now consider the following objective function:

$$F(W, H) = \frac{1}{2}\|Y - WH\|_F^2 + \frac{\beta}{2}(\omega\|H\|_2 - \|H\|_1)^2,$$

where $Y$ is $m \times n$, $W$ is $m \times k$, and $H$ is $k \times n$, and

$$\omega = \sqrt{kn} - (\sqrt{kn} - 1) \text{sparseness}(H).$$

Compute gradient, insert in basic optimization expression, and apply alternating iterations. Results in basis matrix $W$ with a sparse mixing matrix $H$. 
Sample Results - Finding only Endmembers
We Form Simulated Satellites from NASA Data
A Few Combined Traces
(time varying mixtures)
Blind Source Separation Using NMF
Extend NMF: Nonegative Tensor Factorization (NTF) 
Joint project with Christos Boutsidis and Peter Zhang

Our interest: 3-D data. 2-D images stacked into 3-D Array, forming a "box".

\[
\begin{align*}
&\text{ Our interest: 3-D data. 2-D images stacked into } \\
&\text{3-F "Cube" into 3-D "Box".}
\end{align*}
\]
Datasets of images modeled as tensors

**Goal:** Extract features from a tensor dataset (naively, a dataset subscripted by multiple indices). Image samples with diversities, e.g., eigenviews.

\[ m \leq n \leq p \text{ tensor } A \]
What is NTF (for 3-D Arrays)?

Given a nonnegative tensor $A \in \mathbb{R}^{m \times n \times p}$ and a positive integer $k$, find nonnegative vectors $u^{(i)} \in \mathbb{R}^{m \times 1}$, $v^{(i)} \in \mathbb{R}^{n \times 1}$ and $w^{(i)} \in \mathbb{R}^{p \times 1}$ to minimize the functional

$$\frac{1}{2} \| A - \sum_{i=1}^{k} u^{(i)} \circ v^{(i)} \circ w^{(i)} \|_F^2.$$

Here $\circ$ denotes “outer product”. The rank-one matrices $u^{(i)} \circ v^{(i)}$ are the desired basis components, and $w^{(i)}$ the weights.

- See poster by Christos Boutsidis
- **Issues**: Uniqueness, Initialization, Efficient optimization algorithms
Iris Recognition Images
Recovered Images using PARAFAC ~1 hr
Recovered Images using Boutsidis/Zhang ~ 5 min
## Compression

<table>
<thead>
<tr>
<th>Layer</th>
<th>Dimension</th>
<th>Original Size</th>
<th>Compressed Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original array</td>
<td>120x160x30</td>
<td>4,608,000 bytes</td>
<td>48,000 bytes</td>
</tr>
<tr>
<td>X</td>
<td>120x50</td>
<td></td>
<td>64,000 bytes</td>
</tr>
<tr>
<td>Y</td>
<td>160x50</td>
<td></td>
<td>12,000 bytes</td>
</tr>
<tr>
<td>Z</td>
<td>30x50</td>
<td></td>
<td>12,000 bytes</td>
</tr>
</tbody>
</table>

Compression ratio 37 to 1
Columbia in Final Orbit over Maui Space Center
K-means clustering
# Summary and NMF Applications for Spectral Data

- Classification of objects in terms of material features and fractional abundances
- Database compression, including hyperspectral data
- Fast determination of whether a new object spectral trace is in the database, using basis matrix approximation
- Multiple observations with object in different orientations can provide object shape information
- Low-rank representation can enable fast object (target) recognition and tracking (Kullback-Leibler matching)
- Enabled in part by modified nonnegative matrix factorization and information theoretic techniques (relative entropy)
- Compression and reconstruction of image arrays data

*Some related papers at:* [http://www.wfu.edu/~plemmons](http://www.wfu.edu/~plemmons)