Low Rank Nonnegative Factorizations: Algorithms and Applications

Bob Plemmons

Wake Forest University

Collaborators:

Paul Pauca, Jon Piper (Wake Forest), Maile Giffin (Oceanit Labs – Maui)

plus

Michael Neumann (CT), Moody Chu (NCSU), Fasma Diele (Bari), Stafania Ragina (Bari), Michael Berry (UTK)

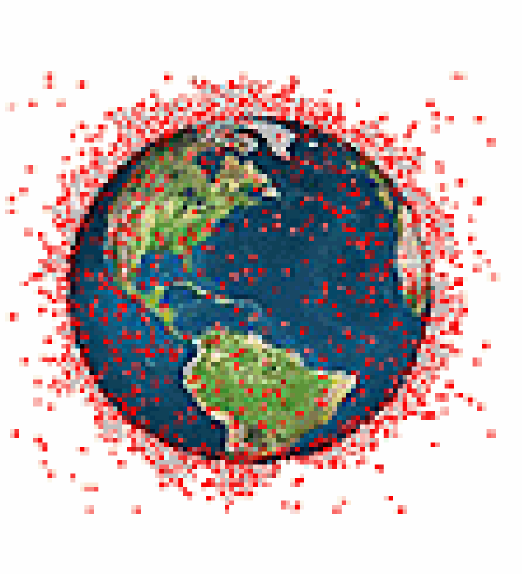
Papers: http://www.wfu.edu/~plemmons

Corton Italy Workshop, September 23, 2004

Alternate Title: Nonnegativity Constrained Low-Rank Matrix Approximation - Nonnegative Matrix Factorization (**NMF**), for Blind Source Separation and Unsupervised Unmixing

- Good Matrix Factorization Reference:
  Hubert, Meulmann, Heiser. “Two purposes of matrix factorization: A historical perspective”, Vol 42 SIREV, 2000.
- Good Matrix Approximation Reference:
  Nick Higham, “Nearest matrix approximations and applications”, Oxford Press, 1999.
- Various Constrained Low Rank Approximation References:
  M. Chu, R. Funderlic, Ple., B. Beckermann, B. De Moor, and numerous other authors.

One Application in this talk: Space Object Identification and Characterization from <u>Spectral Reflectance Data</u>



Perhaps 9,000 objects in orbit: various types of military and commercial satellites, rocket bodies, residual parts, and debris – space object database mining, object Identification, clustering, classification, etc.

General Applications of NMF Techniques

- Document clustering in text data mining (work with Mike Berry)
- Independent representation of image features - face recognition
- Source separation in acoustics, speech
- Hyperspectral imaging from satellites (our Maui project)
- EEG in Medicine, electric potentials
- MEG in medicine, magnetic fields
- Atmospheric pollution source identification (work with Moody Chu, Fasma Diele, Stafania Ragina)
- Sensorimotor processing in robots
- Spectroscopy in chemistry, etc.
- **Spectroscopy for space applications – spectral data mining**
  - **Identifying object surface materials and substances**

Computational Mathematics Space Investment

Partnership for Research Excellence and Transition (PRET) 2002 - 2007

- **PRET: A university based research program involving strong <u>industrial ties</u> to accelerate transition of research to industry**
- **PRET Objective: Explore and develop many of the basic sciences that form the basis for space situational awareness (SSA)**
- **Specific Research Areas:**
  - **<u>Spectral data mining</u>**
  - **Wave front sensor control**
  - **Image processing**
  - **<u>Enabling mathematics</u>**

The University of New Mexico

WAKE FOREST UNIVERSITY

Optical Sciences Center
The University of Arizona

Oceanit ...innovation through engineering and scientific excellence...

MAUI HIGH PERFORMANCE COMPUTING CENTER

Outline

- Background and Overview of the Problem
  - SOI (space object identification)
  - PCA, ICA, Sparse ICA, Non-Negative Sparse ICA
- Data Description
- Features-Based Identification & Classification
- Nonnegativity Constrained Low-Rank Approximation for Blind Source Separation and Unsupervised Unmixing
- Information-theoretic matching methods
- Preliminary Results using Spectrometer Data

Overview of the SOI Problem

- Space activities require accurate information about orbiting objects for space situational awareness and safety
- Many objects are either in
  - Geosynchronous orbits (about 40,000 KM from earth), or
  - Near-Earth orbits, but too small to be resolved by optical imaging systems
- Orbiting object identification and classification through reflectance spectroscopy sensor measurements
- Spectral measurements of reflected sunlight used to identify object surface materials and substances

Overview of the SOI Problem Continued

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

- Problem solution by learning the parts of objects (hidden components) by low rank non-negative sparse independent component analysis - a new approach for scientific data mining and unsupervised hyperspectral unmixing.

- Basis representation (dimension reduction) may enable near real-time object (target) recognition, object class clustering, and characterization.

- Space object identification from wavelength-resolved data is a promising and viable approach currently under investigation
  - Spectral measurements of reflected sunlight used as basis for object surface materials and substances identification
  - Enable satellite classification
  - Enable subsequent satellite shape determination
- Fundamental difficulty

  Determine from spectral measurements of an object:
  - *Endmembers*: the type of constituent material
  - *Fractional abundances*: the proportional amount of these materials that make up the object

Blind Source Separation for Finding Hidden Components

Mixing of Sources

…basic physics often leads to linear mixing…

$X = [X_1, X_2, \ldots, X_m]$ – training set of column vectors approximately factor

$$X \approx W H$$

X sensor readings (mixed components – observed data)

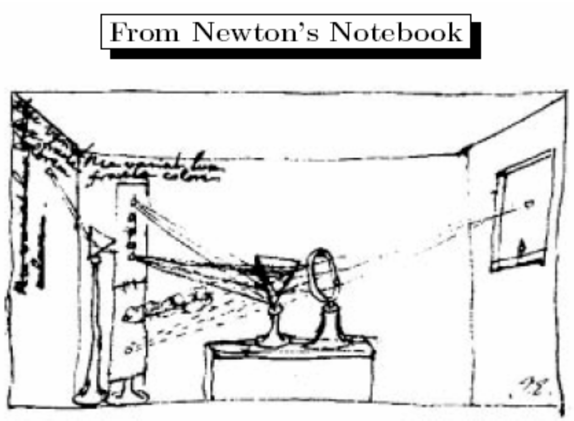
W separated components (feature basis matrix - unknown)

H hidden mixing coefficients (unknown)

Complete prior knowledge of basis matrix **W** would simplify problem, but **W** seldom known in practice.

Simple Analog Illustration

Hidden Components in Light – Separated by a Prism



Our purpose – finding hidden components by <u>data analysis</u>

Some References:

Recent work involving co-authors of this presentation

- Pauca, Ple., Giffin, “Unmixing Spectral Data for Space Objects using Low-Rank Non-Negative Sparse Component Analysis”, to appear in Proc. Maui Amos Tech. Conf., 2004

- Pauca, Shahnaz, Berry and Ple., “Text Mining using Non-negative Matrix Factorization”, to appear in Proc. International Conf. on Data Mining, Orlando, 2004.

- Careal, Han, Neumann and Ple., “Reduced Rank Non-Negative Similarity Matrix Factorization”, to appear in LAA, 2004.

- Chu, Diele, Ple., Ragni, “Some Theory, Numerical Methods, and Applications of NMF”, draft 2004

Additional Related References

- Lee and Seung. “Learning the Parts of Objects by Non-Negative Matrix Factorization", Nature, 1999.
- Hoyer. “Non-Negative Sparse Coding", Neural Networks for Signal Proc., 2002.
- Hyvärinen and Hoyer. “Emergence of Phase and Shift Invariant Features by Decomposition of Natural Images into Independent Feature Subspaces", Neural Computation, 2000.
- David Donoho and Stodden. ``When does Nonnegative Matrix Factorization give a Correct Decomposition into Parts?", preprint, Dept. Stat., Stanford, 2003.
- Berman and Plemmons. Non-Negative Matrices in the Mathematical Sciences, SIAM Press, 1994.
- Sajda, Du, and Parra, “Recovery of Constituent Spectra using Non-negative Matrix Factorization”, Tech. Rept., Columbia U. & Sarnoff Corp. 2003.
- Cooper and Foote, “Summarizing Video using Non-Negative Similarity Matrix Factorization”, Tech. Rept. FX Palo Alto Lab, 2003.
- Szu and Kopriva, “Deterministic Blind Source Separation for Space Variant Imaging”, 4th Inter. Conf. Independent Component. Anal., Nara Japan, 2003.
- Umeyama, “Blind Deconvolution of Images using Gabor Filters and Independent Component Analysis”, 4th Inter. Conf. Independent Component. Anal., Nara Japan, 2003.

- Brief Review -

- Principal Component Analysis (PCA)
- Independent Component Analysis (ICA)
- Sparse Component Analysis (SCA)
- Non-Negative SCA

Various Approaches for BSS Can be Used PCA – Older Method

- Based on eigen-decomposition of covariance matrix for
  X = $[X_1, X_2, \ldots, X_m]$ – training set of column vectors, scaled and centered, XX^T (or SVD of X itself).
- In the PCA context each column of *W* represents an eigenvector (hidden component), and *H* represents eigenprojections.
- "Principal" components correspond to largest eigenvalues. Components called "eigenfaces" in face recognition applications.

- Advantages: orthogonal representation, dimension reduction, clustering into principal components, computed by simple linear algebra.
- Disadvantages: does not enforce nonnegativity in W and H.

ICA

- Based on neural computation studies – unsupervised learning.
- Identified with - blind source separation (BSS), feature extraction, finding hidden components.
- Most research based on equality, *X* = *WH*, not necessary.
- Statistical independence for components in W, a guiding principle, but seldom holds in practical situations.
- Data in X assumed to have nongausssian PDF, find hidden components as independent as possible – mutual information content in different components c_i, c_j, is (near) zero, or $p(c_i,c_j) \approx p(c_i)p(c_j)$.

- Next, sparse separation into parts, and use data non-negativity.

SCA

- Sparse (independent) component analysis – called sparse encoding in the neural information processing literature.
- Enforce sparsity for the hidden mixing components in *H*.
- PDF has sharp peak at zero and heavy tails
- Allows better separation of basis components by parts,
- Measures of sparsity: l^p functional, $p \leq 1$ (not a formal norm if $p < 1$). Other measures studied by Donoho, “beyond wavelets”.

Non-Negative SCA

- Utilize constraint that sensor data values in X are <u>nonnegative</u>

- Apply non-negativity constrained low rank approximation for blind source separation, dimension reduction (data compression) and unsupervised unmixing

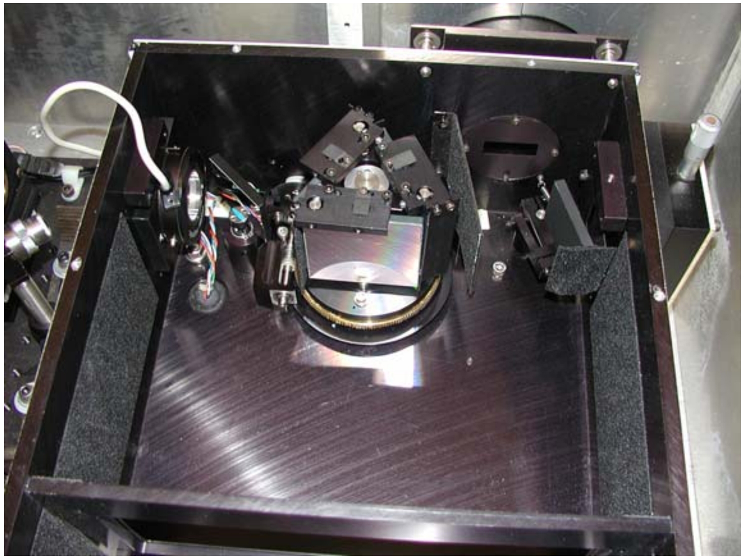
- Low rank approximation to data matrix $\boldsymbol{X}$:

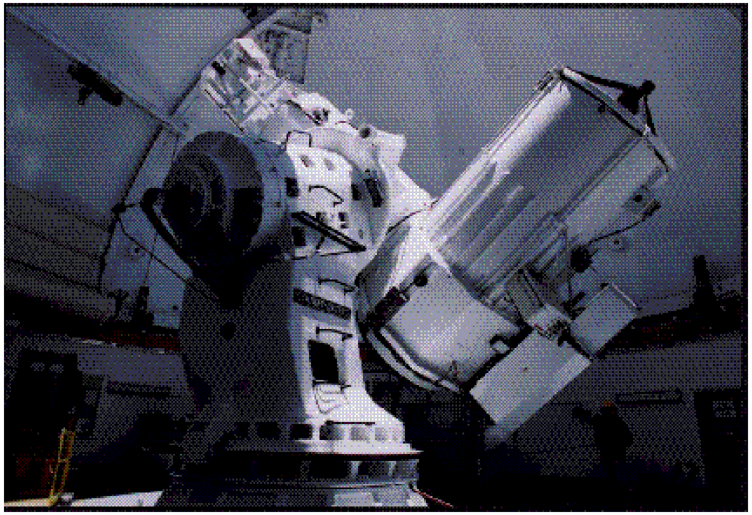
$$\boldsymbol{X} \approx \boldsymbol{WH}, \ \boldsymbol{W} \geq 0, \ \boldsymbol{H} \geq 0$$

  - Columns of $\boldsymbol{W}$ are basis vectors for spectral trace database, desire statistical independence in $\boldsymbol{W}$.
  - Columns of $\boldsymbol{H}$ represent mixing coefficients, desire statistical sparsity in $\boldsymbol{H}$.

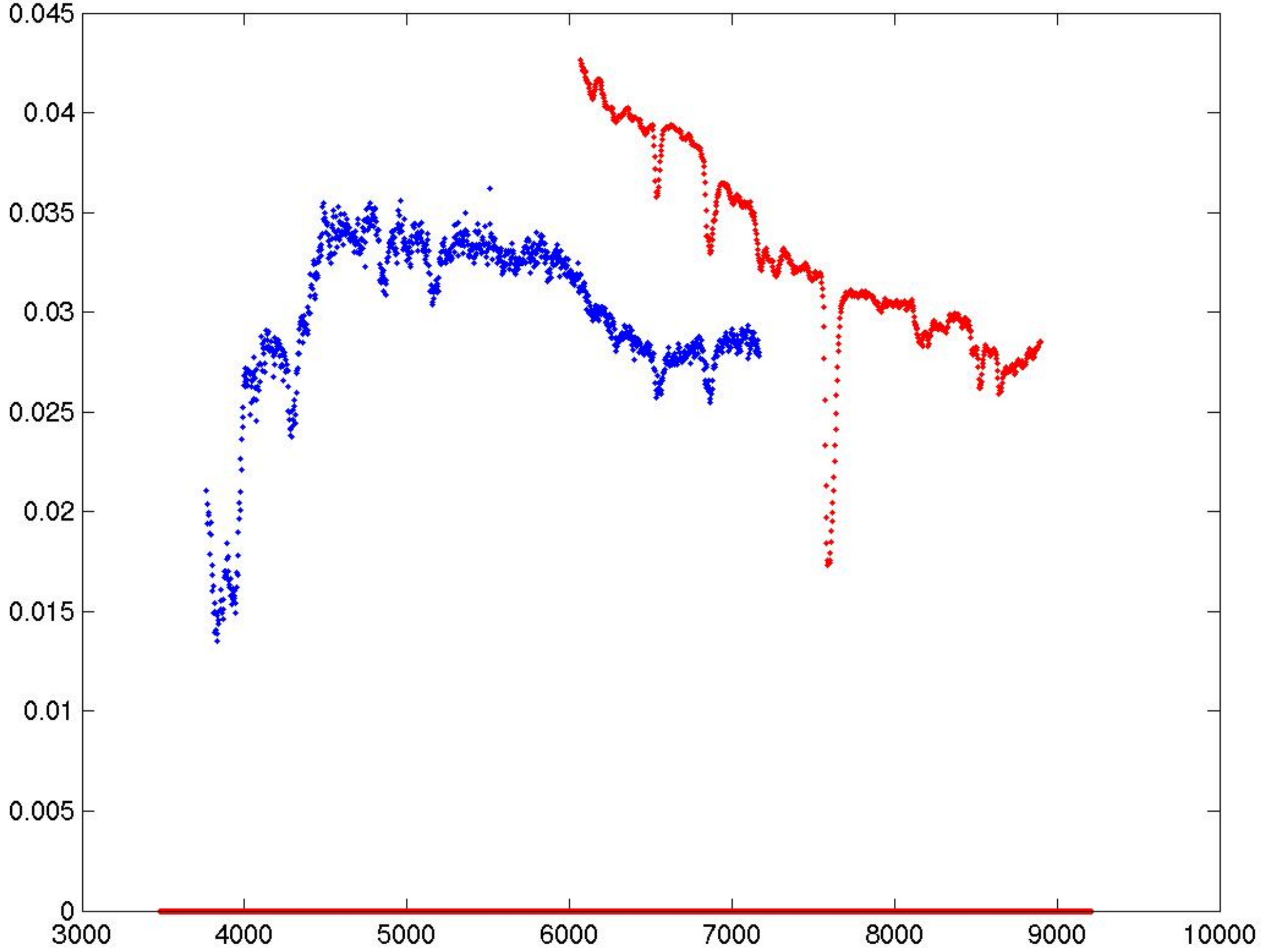
Data Obtained from a Spica (Space Infrared Telescope for Cosmology and Astrophysics) - type Spectrometer

- Mission: Support non-imaging SOI with spectroscopic observations
- 3 – 4 angstrom resolution
- Blue mode: 3000 – 6000 angstroms (.3 – .6 μm)
- Red mode: 6000 – 9000 angstroms (.6 – .9 μm)
- Located on the rear blanchard of a Maui 1.6m telescope
- Can acquire 15th magnitude objects (dim objects)





Sample Raw Data Collected in Blue and Red Modes



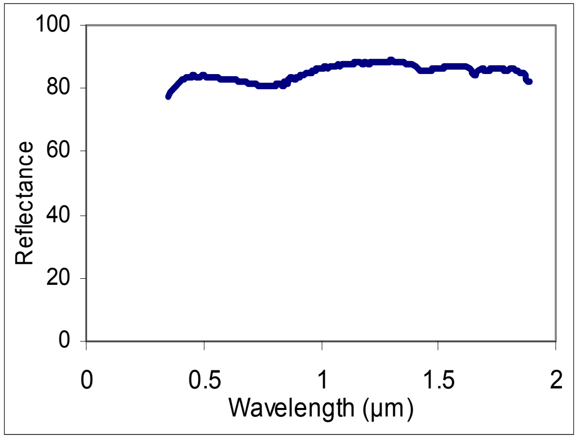
Electromagnetic Spectrum: Spectral Signatures

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

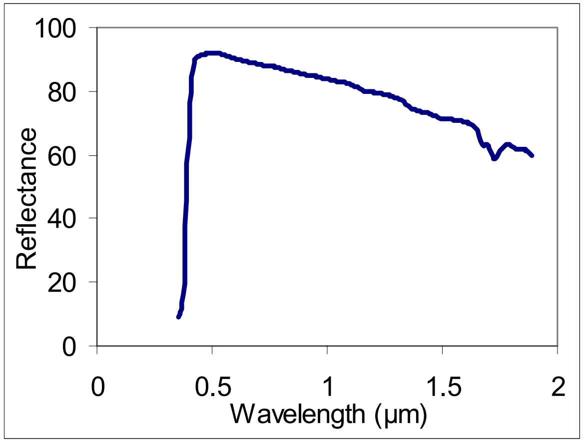
- **This property of matter makes it possible to identify different substances and separate them by their spectral signatures (spectral curves) – hyperspectral unmixing.**

  - **Complexity arises since objects can be composed of many materials, each with their own spectral signature.**

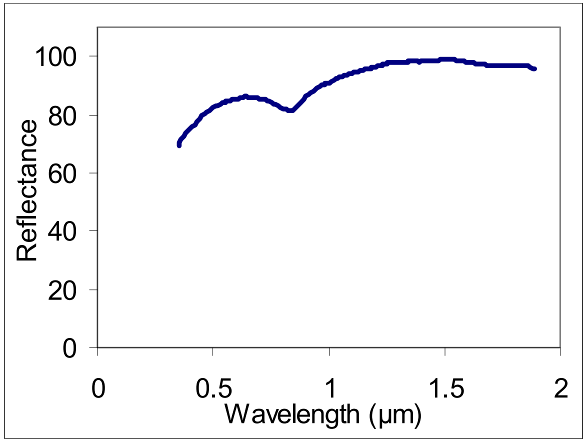
Some Laboratory Electromagnetic Spectral Signatures



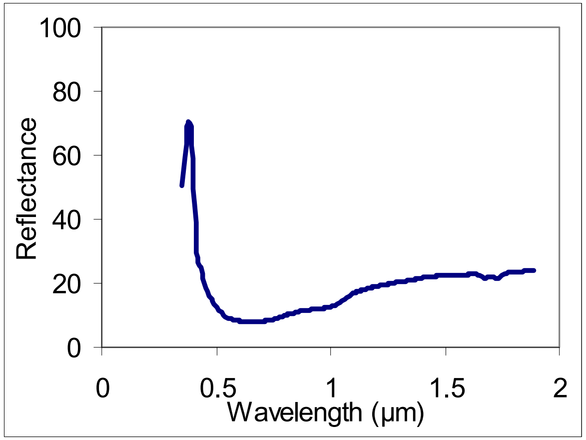
Mylar



White Paint



Aluminum



Solar Cell

Database Description

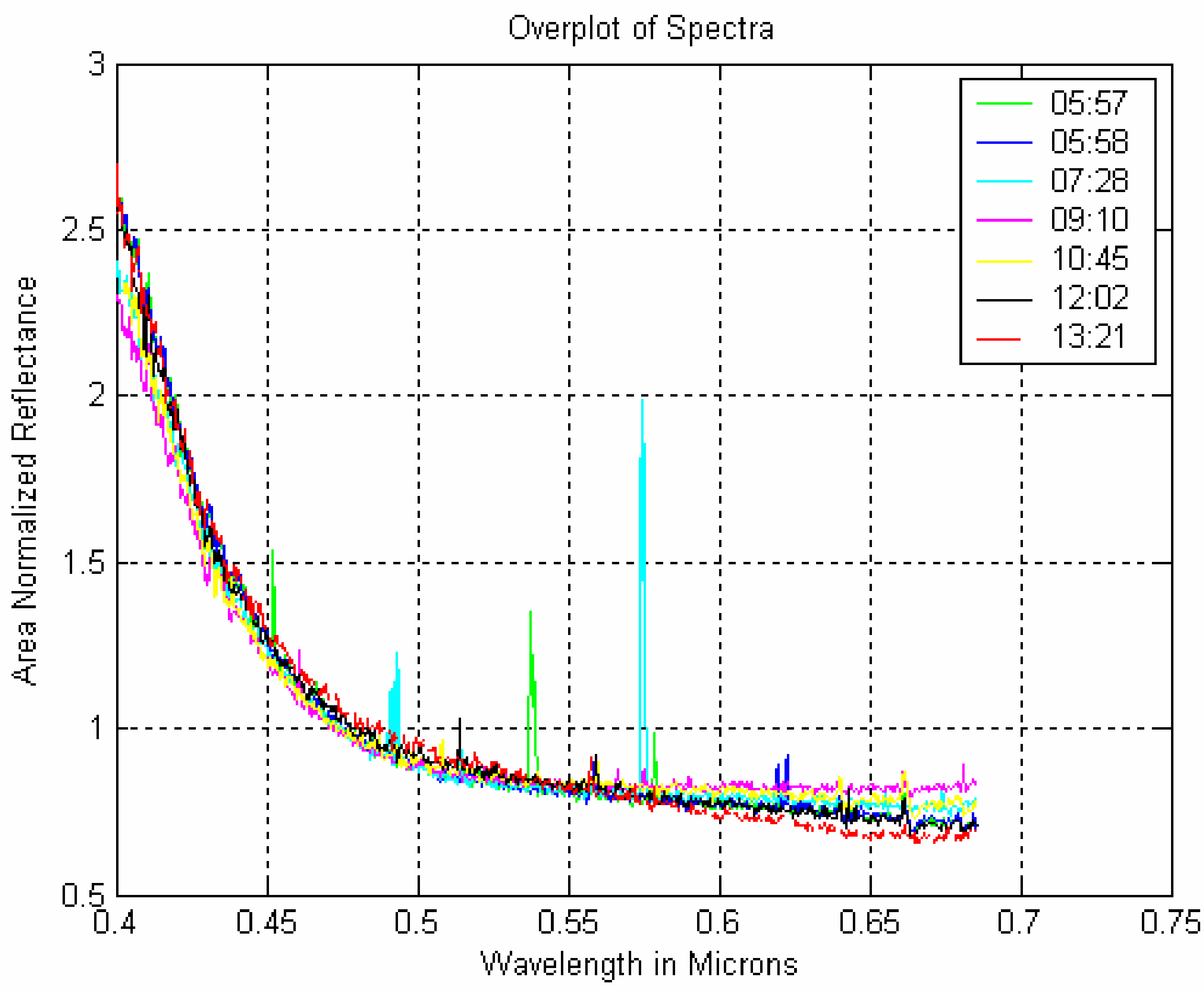
- Spica dataset used for test purposes consists of 2,392 spectral traces of various space objects. Training data matrix ***X*** is 5,732 X 2,392.

- Individual trace wavelengths ranged between about .3 to .9 microns, collected in a blue mode (wavelength .3 to .6 microns) and red mode (.6 to .9 microns).

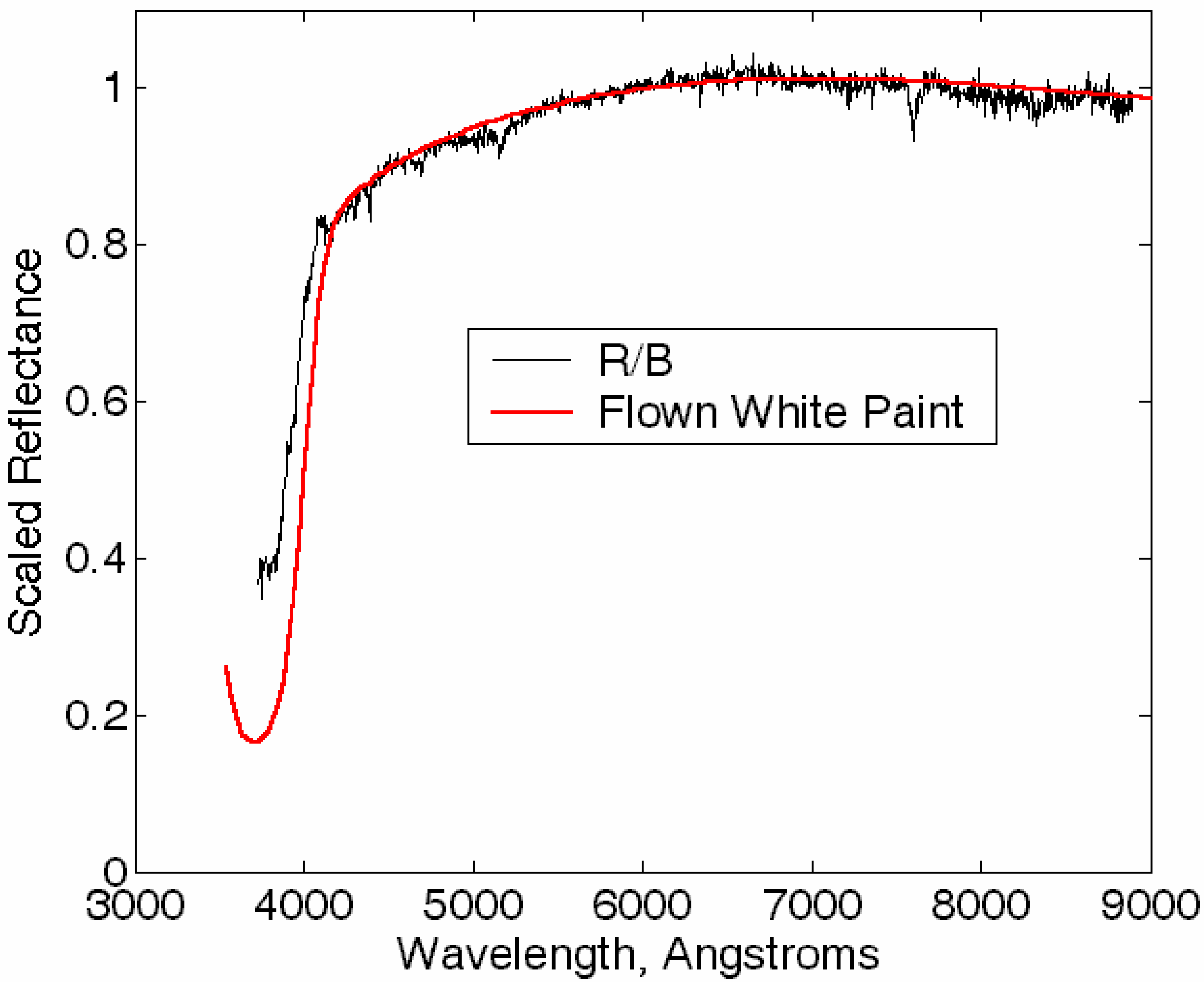
- Spectral traces are pre-processed to correct for cosmic rays, etc., and have background and atmospheric absorption effects removed. CCD read noise and thermal noise are also present, but at small levels.

Spica Observations Of Galaxy V Provide Test Cases

Some Raw Data Spectral Observations of Galaxy V Satellite



NASA data showing spectra of a white painted rocket body matched with a laboratory spectra of white paint (angstrom = 10^4 microns)



Parts- Based Feature Identification & 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 $\boldsymbol{X}$: $\boldsymbol{X} \approx \boldsymbol{WH}$, $\boldsymbol{W} \geq 0$, $\boldsymbol{H} \geq 0$
  - Columns of ***W*** are basis vectors for spectral trace database
  - Columns of ***H*** represent mixing coefficients
- Low rank representation may allow near real-time object (target) recognition and classification using reduced dimension basis matrix W

Learning the Hidden Components of Objects by Nonnegative Matrix Factorization (NMF) – a Recent Approach for Mining Nonnegative Scientific Data

- First proposed by Lee and Seung (MIT) in *Nature*, 1999.
- Idea - use NMF to find a set of nonnegative basis functions to represent image-related data where the basis functions enable the identification of "intrinsic parts or features" of objects and spectral abundances.
- Allows only additive, not subtractive combinations of the original data, in comparison to other decomposition methods such as principal component analysis (PCA) and independent component analysis (ICA).
- Problem solution by unsupervised hyperspectral unmixing
- NMF has also been used successfully for "unmixing" data consisting of spectral traces of ARVIS (Airborne Visible/IR Spectrometer) observations
- Other spectral unmixing a applications include Raman spectroscopy and chemical shift imaging in biochemistry classification of nuclear magnetic resonance spectral data in medicine, by Sajda, et al, Columbia U. and Sarnoff Corp., 2003.

NMF Problem Formulation

- Given initial database expressed as n x m nonnegative matrix **X**

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

$$\min_{W,H} \|X - WH\|_F^2,$$ plus constraints

where $W_{ij} \geq 0$ and $H_{ij} \geq 0$ for each i and j. Choice of r << m is often problem dependent. Can impose other (e.g., smoothness) constraints on W and/or H.

NMF - Continued

- Of course ***W*** and ***H*** are not unique without further constraints. ***W*** (DP)(DP)$^{-1}$***H***, etc.

- Donoho, et al, 2003, used <u>convex cone theoretic geometric concepts</u> to determine conditions for uniqueness, up to permutation and scaling of the rows.

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 (eps is a machine dependent small positive pos. no.):

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

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

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

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

A Non-Negative Sparse Coding Approach Suggested Hoyer and Donaho in the Blind Source Separation Literature

Initialization of ***W*** and ***H*** as with GD-CLS Algorithm

Iterate until convergence or stop:

(a) $W \leftarrow W - \mu(WH - X)H^T$

(b) Set any negative values of W to zero.

(c) Rescale the columns of W to unit norm.

(d) $H \leftarrow H.*(W^TX./(W^TWH + \lambda./(H+\gamma).\wedge 2)$

We replaced (a) by (a) $W_{ic} \leftarrow W_{ic}\frac{(XH^T)_{ic}}{(WHH^T)_{ic} + eps}$, for each c and i

(b) Is then not needed.

We replaced step (d) above by: $H \leftarrow H.*(W^TX)./(W^TWH + \lambda)$

Equivalent to using a sparsity constraint for non-negative H.

A Non-Negative Sparse ICA Scheme

Modified Hoyer Algorithm for Non-Negative Sparse Coding

1. Initialize W and H with non-negative values, and scale the columns of W to unit norm.
2. Iterate until convergence or stop:

   (a) $W_{ic} \leftarrow W_{ic} \dfrac{(XH^T)_{ic}}{(WHH^T)_{ic} + eps}$, for each c and i

   (b) Rescale the columns of W to unit norm.

   (c) $H \leftarrow H.*(W^T X./(W^T WH + \lambda)$

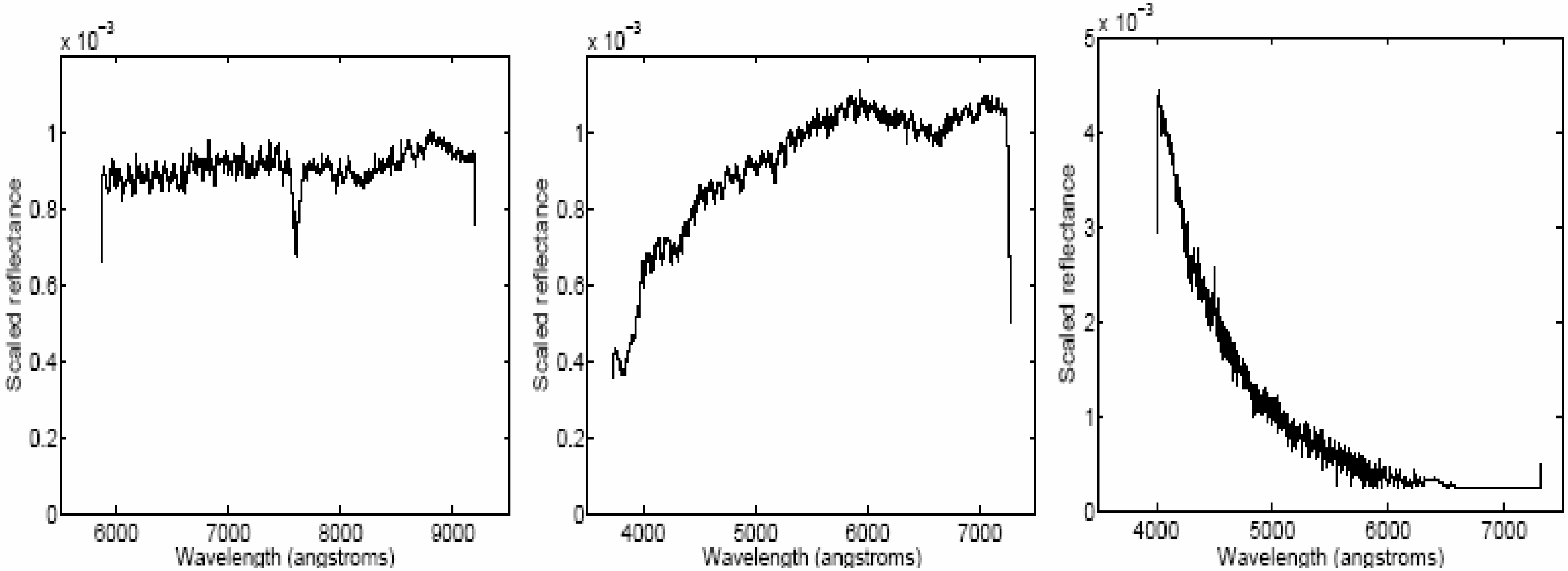
Experimental Results using the Spica Database

- ***X*** contains m = 2,392 spectral traces, each represented by a vector of dimension n = 5,732, corresponding to the wavelength range used

- Number of columns used for the basis matrix ***W*** and rows of the spectral abundance matrix ***H*** was arbitrarily set at ***k = 30*** , for test purposes, as an estimate for an upper bound on the number of distinct material traces present in the space objects

- The non-negative sparse coding ICA algorithm were applied to Spica database

Some Columns of the Computed Basis Matrix *W* Showing Intrinsic Hidden Component Material Spectra (endmembers)



mylar (left), white paint (center), and solar cell (right)

How well does the factorization *WV* approximate *Y* for satellite identification?

- Test by <u>scoring</u> with a fixed spectral scan *q* of Sat #21906 (Galaxy V), using spectral data from blue mode, observations taken on various days – significant!
- Matching done by using the information-theoretic Kullback-Liebler Divergence Measure
- Thirty basis vectors are used.
- We derive a method imposing useful constraints. Notation change: Y for X.

- New Approach
  - Solve a constrained optimization problem,

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

    where $\alpha J_1(W)$ and $\beta J_2(H)$ are used to enforce certain application-dependent characteristics on the solution
  - For spectral unmixing, we set $J_1(W) = \|W\|_F^2$ to enforce smoothness on the basis vectors in W and $\beta = 0$
  - Constrained version of NMF is derived and convergence property proven (CNMF, like NMF, is an alternating modified steepest descent type algorithm)
  - Found that resulting basis vectors better resemble material spectral signatures and give better approximation to Y

Constrained NMF Algorithm (CNMF)

1. Choose the number of basis vectors k.

2. Initialize $W^{(0)}$ and $H^{(0)}$ with non-negative values.

3. For $t = 1, 2, \ldots,$ until $F(W^{(t)}, H^{(t)}) \leq tol$, for a chosen tolerance value $tol \in \mathbb{R}$, compute

   (a) $$H_{cj}^{(t)} \leftarrow H_{cj}^{(t-1)} \frac{\left((W^{(t-1)})^T Y\right)_{cj} - \beta H_{cj}^{(t-1)}}{\left((W^{(t-1)})^T W^{(t-1)} H^{(t-1)}\right)_{cj} + \epsilon}, \quad 1 \leq c \leq k, \quad 1 \leq j \leq n,$$

   (b) $$W_{ic}^{(t)} \leftarrow W_{ic}^{(t-1)} \frac{\left(Y (H^{(t)})^T\right)_{ic} - \alpha W_{ic}^{(t-1)}}{\left(W^{(t-1)} H^{(t)} (H^{(t)})^T\right)_{ic} + \epsilon}, \quad 1 \leq i \leq m, \quad 1 \leq c \leq k.$$

where $\epsilon \approx 10^{-9}$ is used to avoid possible division by zero.

Endmember Selection

- For two random variables $\mathbf{w}$ and $\mathbf{d}$, with probability vectors $\mathbf{p} = (1/\sum w_i)\mathbf{w}$ and $\mathbf{q} = (1/\sum d_i)\mathbf{d}$, the Kullback-Leibler divergence is defined as

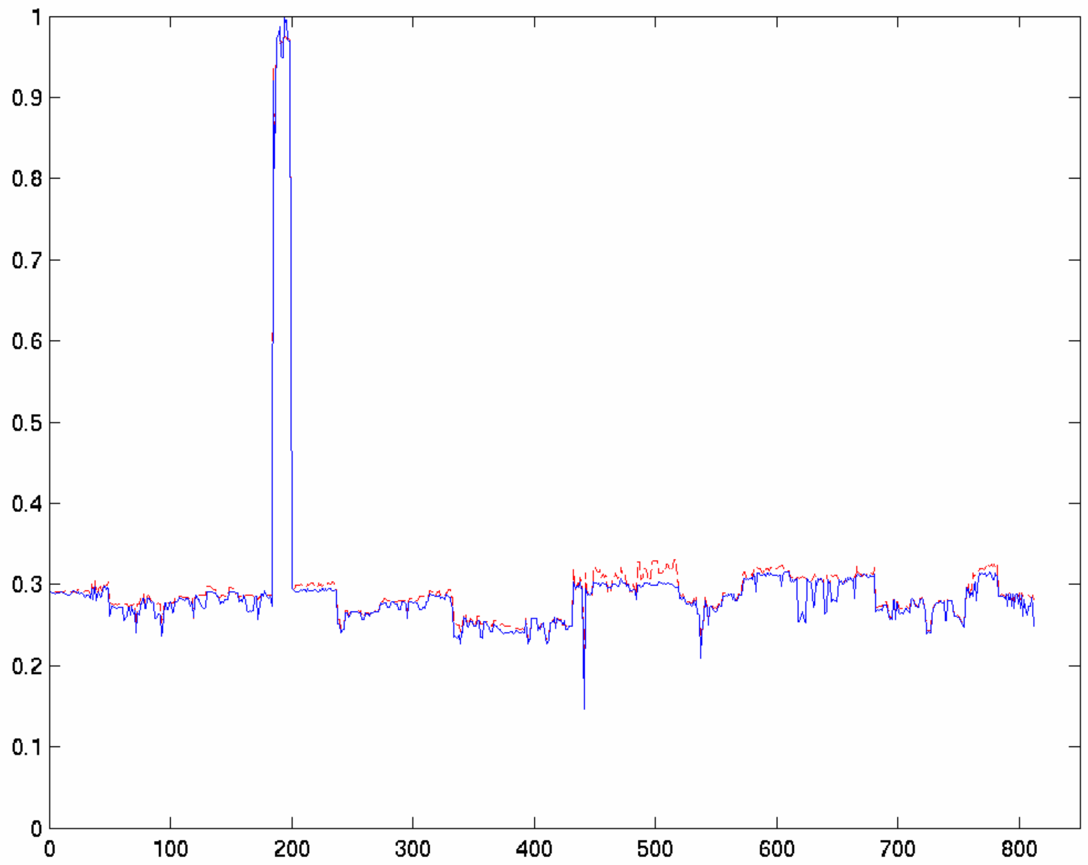
$$D(\mathbf{w}||\mathbf{d}) = \sum p_i(\log(p_i) - \log(q_i)) = \sum p_i \log(\frac{p_i}{q_i}),$$

- Here, we use a symmetrized version

$$D_s(\mathbf{w}, \mathbf{d}) = D(\mathbf{w}||\mathbf{d}) + D(\mathbf{d}||\mathbf{w})$$

- For each $\mathbf{d}_j$, form pair $(\mathbf{d}_j, \arg\min_{\mathbf{w}_i} D_s(W, \mathbf{d}_j))$, the basis vector in W to which $\mathbf{d}_j$ is closest to in terms of D_s.
- A tolerance τ is used to reject pairs for which D_s is larger than τ
- The basis vectors in remaining pairs are used to form matrix B

Columns B used as final endmember set



Scoring truth for database is BLUE. Scoring approximation using 30 basis in endmember matrix B vectors is RED.

Quantification of Fractional Abundancies

We use PMRNSD from RestoreTools and endmember matrix B to iteratively solve for fractional abundances vector x, given a spectral trace vector y.

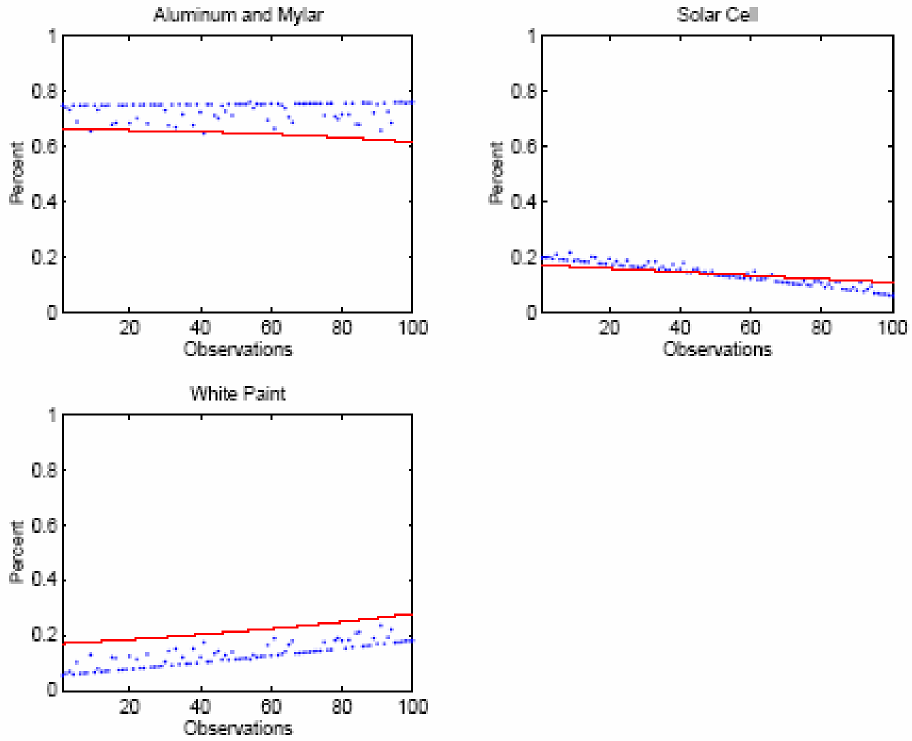
- For each $\mathbf{y}$ parameterize $\mathbf{x}$ to solve an equivalent unconstrained minimization problem

$$J(\mathbf{x}) = \min_{\mathbf{x}} \|\mathbf{y} - B\mathbf{x}\|_2^2, \quad \text{for } \mathbf{x} = e^{\mathbf{z}}.$$

x provides percentages of aluminum, mylar, solar cell, paint, etc.

Sample Simulation Results

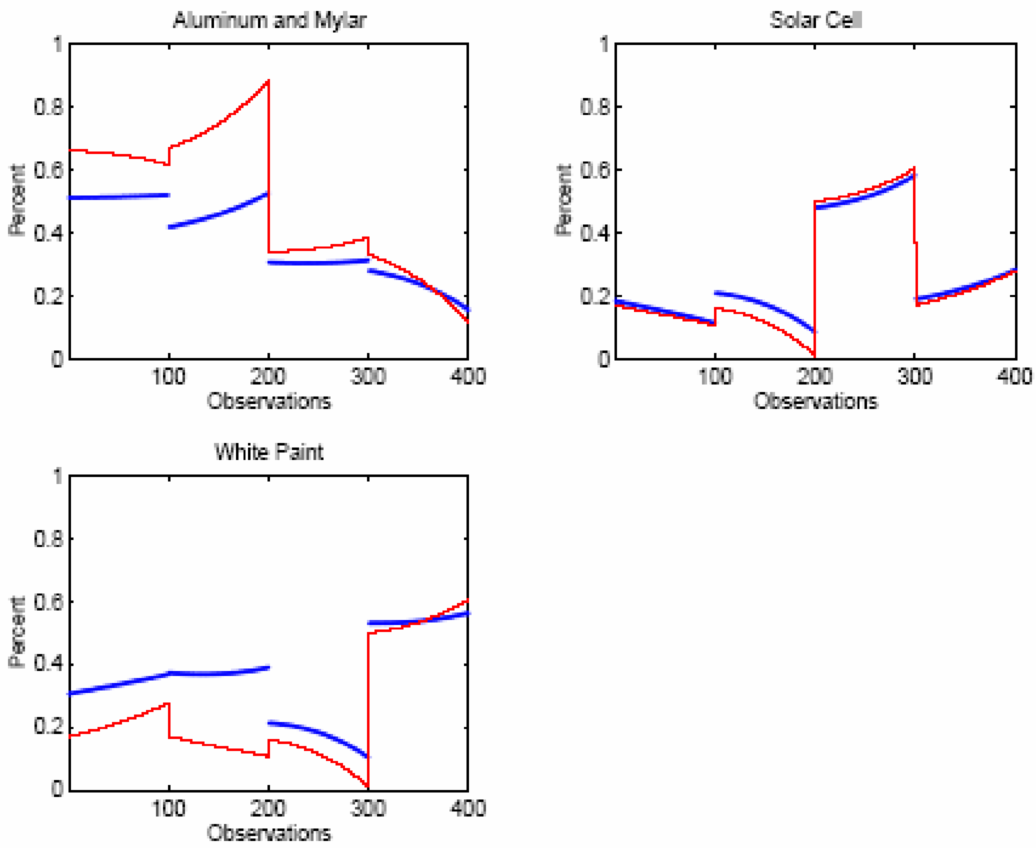
PMRNSD computed fractional abundances for Satellite 1



- Red line represents true fractional abundances and blue line the computed fractional abundances
- Best results for solar cell and white paint

Satellite 5 simulates (real) Galaxy 5 at different observation times and in different orientations

PMRNSD computed fractional abundances for Satellite 5



– Computed fractional abundances follow well trend of change in material composition

Applications to Object (Target) Feature Identification

- Classification of objects in terms of material features and fractional abundances
- Database compression
- Fast determination of whether a new object spectral trace is in the database, using basis matrix *B*
- Multiple observations with object in different orientations can provide object shape information
- Low-rank representation to enable near real-time object (target) recognition and tracking

- The End -