Tensor Analysis of Massive Data Sets

Bob Plemmons
Wake Forest University

- Collaborators:
  - Peter Zhang, WFU Dept. Biostatistical Sciences
  - Han Wang, WFU Dept. Biostatistical Sciences
  - Paul Pauca, WFU Dept. Computer Science
- Interaction on spectral data with Kira Abercromby (NASA-JSC in Houston)
- Inter. Conf. on Interdisciplinary Stat. and Comb., UNCG, Oct., 07
- Related Papers at: http://www.wfu.edu/~plemmons

Outline

- Matrix Factorizations for Data Analysis, PCA, ICA…
- Nonnegative Matrix Factorization (NMF)
- Nonnegative Tensor Factorization (NTF), a Brief Overview in 3D
- An Application: Object Identification from Spectral Data
- Nonnegativity Constrained Low-Rank Approximation (Ill-posed nonlinear inverse problem)
- Computations using Air Force data from Maui and NASA, and Biomedical applications

Nonnegative Matrix Factorization (NMF)

[NMF Problem]

Given a nonnegative matrix $\mathrm{A} \in \mathrm{R}^{\mathrm{m}\times\mathrm{n}}$ and a positive integer $k < \min\{m, n\}$, find nonnegative matrices $\mathrm{W} \in \mathrm{R}^{\mathrm{m}\times\mathrm{k}}$ and $\mathrm{H} \in \mathrm{R}^{\mathrm{k}\times\mathrm{n}}$ to minimize the functional

$$f(\mathrm{W}, \mathrm{H}) = \frac{1}{2}\|\mathbf{A} - \mathbf{WH}\|_{\mathbf{F}}^{\mathbf{2}}.$$

Convex in **W** or **H**, but <u>not both</u>. A natural extension of nonnegative least squares.

NMF an SVD-Type Concept

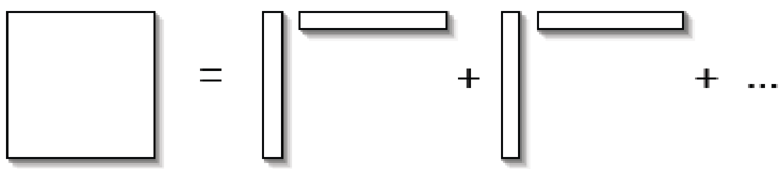
$\boldsymbol{X} = [X_1, X_2, \ldots, X_n]$ – column vectors (data)

Approximately factor

$$\boldsymbol{X} \approx \boldsymbol{W}\,\boldsymbol{H} = \sum_1^k \boldsymbol{w}^{(j)} \circ \boldsymbol{h}^{(j)}$$

$\circ$ denotes outer product

$\boldsymbol{w}^j$ is jth col of $\boldsymbol{W}$, $\boldsymbol{h}^j$ is jth col of $\boldsymbol{H}^T$



Approximate NMF in Data Analysis

- Utilize constraint that values in a data matrix $\boldsymbol{X}$ are <u>nonnegative</u>
- Apply non-negativity constrained low rank approximation for blind source separation, dimension reduction, and/or 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 initial <u>basis vectors</u> for database, may want smoothness and statistical independence in $\boldsymbol{W}$.
  - Columns of $\boldsymbol{H}$ represent <u>mixing coefficients or weights</u>, often desire statistical sparsity in $\boldsymbol{H}$ to force essential uniqueness in $\boldsymbol{W}$.

- Brief Overview -

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

PCA

- Based on eigen-decomposition of covariance matrix for $\boldsymbol{X} = [X_1, X_2, \ldots, X_m]$ – training set of column vectors, scaled and centered, $\boldsymbol{XX}^T$ (or SVD of $\boldsymbol{X}$ itself). Low-rank approx.
- 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.)
- Used for: orthogonal representation, dimension reduction, clustering into principal components, all computed by simple linear algebra.

Characteristics of PCA

- May not enforce nonnegativity in ***W*** and ***H***, where $\boldsymbol{X} \approx \boldsymbol{WH}$
- May not enforce statistical independence of basis vectors in ***W***
- May not enforce statistical sparsity in ***H*** (separation by parts).
- This led to work beginning in late 1990s to present on ICA, SCA, non-negative SCA, and NMF.

ICA

- Based on neural computation– unsupervised learning.
- Often 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 non-gausssian 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 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".

Some 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.
- Our papers on web page: http://www.wfu.edu/~plemmons

Nonnegative Matrix Factorization (NMF)

[NMF Problem]

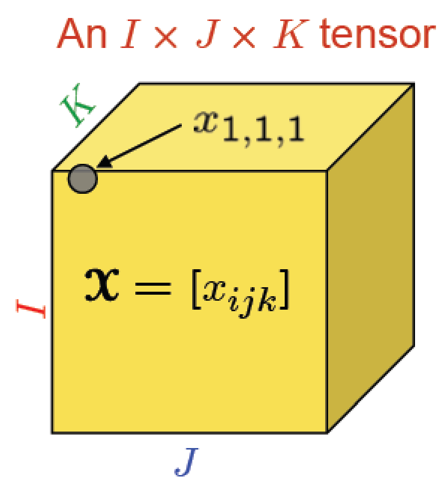
Given a nonnegative matrix $\mathrm{A} \in \mathrm{R}^{\mathbf{m}\times\mathbf{n}}$ **and a positive integer** $k < \min\{m, n\}$**, find non-negative matrices** $\mathrm{W} \in \mathrm{R}^{\mathbf{m}\times\mathbf{k}}$ **and** $\mathrm{H} \in \mathrm{R}^{\mathbf{k}\times\mathbf{n}}$ **to minimize the functional**

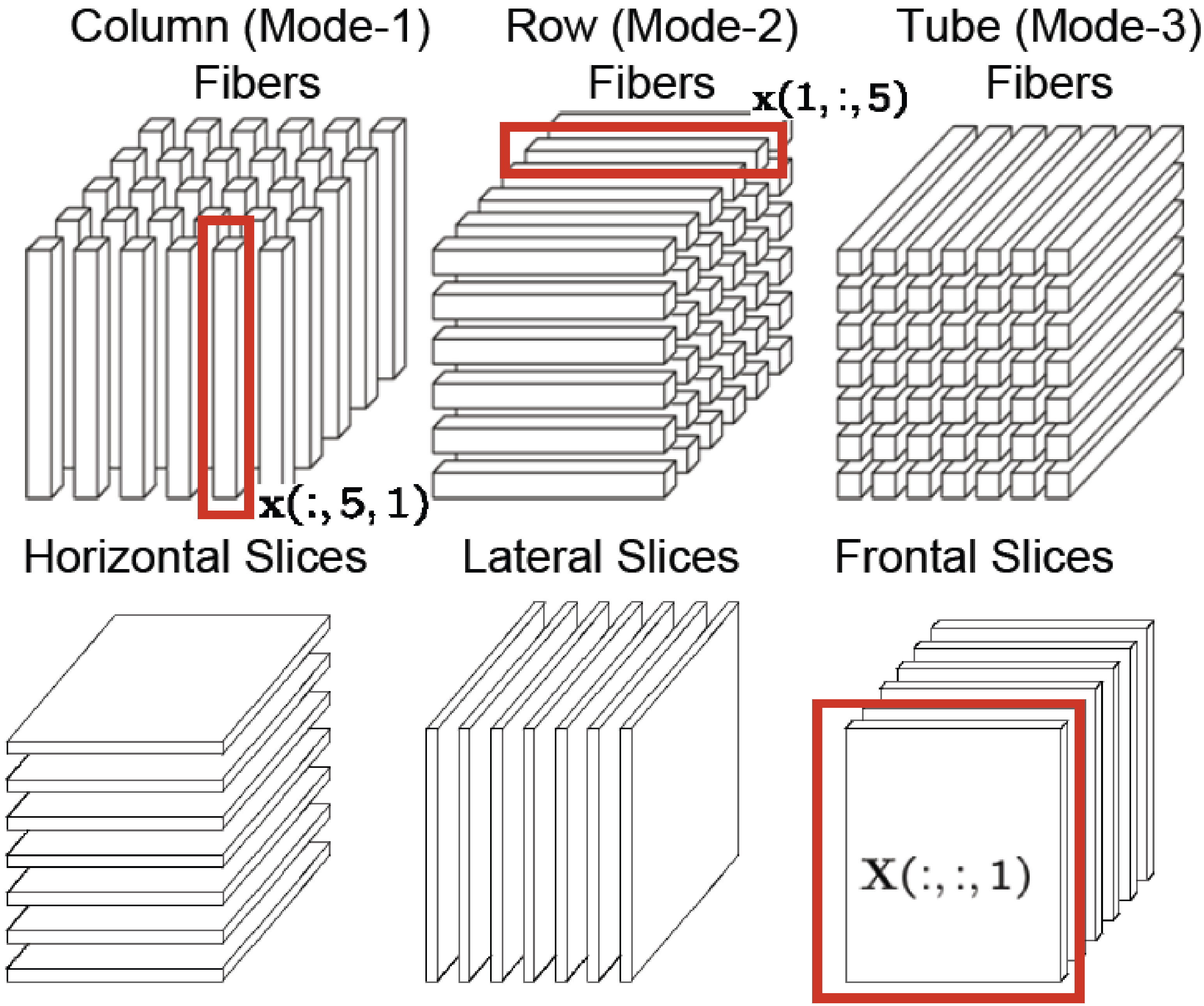
$$f(\mathrm{W}, \mathrm{H}) = \frac{1}{2}\|\mathbf{A} - \mathbf{WH}\|_{\mathbf{F}}^{\mathbf{2}}.$$

Will extend concept to tensors.

Tensors

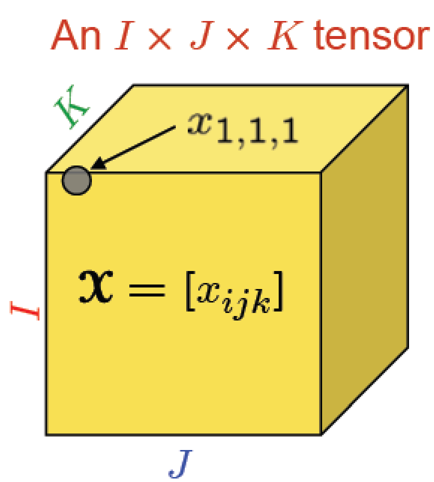
Multi-way Arrays
in Multilinear Algebra
3D Viewpoint – Tensor as a Box

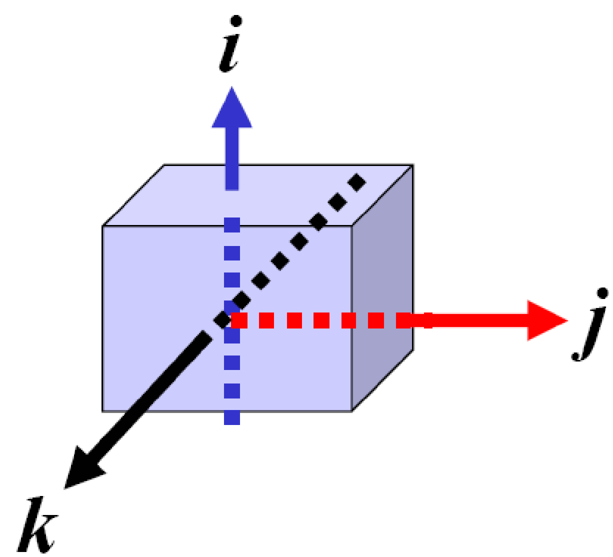



Column (Mode-1) Fibers
Row (Mode-2) Fibers
Tube (Mode-3) Fibers
x(:, 5, 1)
x(1, :, 5)
Horizontal Slices
Lateral Slices
Frontal Slices
X(:, :, 1)

Extend NMF: Nonegative Tensor Factorization (NTF)

- Our interest: 3-D data. 2-D images stacked into 3-D Array





Tensors in a Nutshell

- A matrix is an order-2 tensor, a 2-array of elements
- An *order-k tensor* is simply a k-array of elements
- Standard concepts for matrices, such as rank, eigenvalues, etc., are much more complicated for tensors
- Caution: What physicists and geometers call tensors are often tensor fields (i.e. tensor-valued functions on manifolds, e.g, stress tensor, Einstein tensor, etc.)
- Our concern is tensors in a multilinear algebra context:

$$X = [x_{j1\ldots jk}] \in R^{d_1 \times \ldots \times d_k}$$

Multilinear NMF = Nonnegative Tensor Factorization (NTF)

NTF (for 3-D Arrays)

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

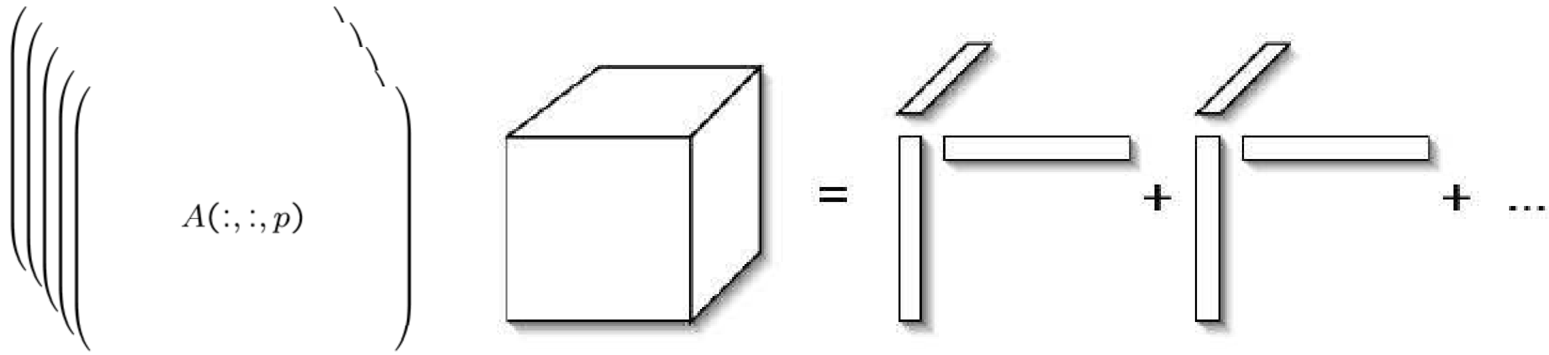
$$\frac{1}{2}\|\mathcal{A} - \sum_{i=1}^{k} u^{(i)} \circ v^{(i)} \circ w^{(i)}\|_F^2.$$

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

Datasets of images modeled as tensors

Goal: Extract features from a tensor dataset (naively, a dataset subscripted by multiple indices).

$m \times n \times p$ tensor A



Tensor (outer product) rank can be defined as the minimum number of rank 1 factors.

Properties of Matrix Rank

- Rank of $A \in R^{m \times n}$ easy to determine (Gauss elimination).
- Optimal rank-r approximation to A always exists, and easy to find (SVD).
- Pick $A \in R^{m \times n}$ at random, then A has full rank with probability 1, i.e., rank(A) = min{m,n}.
- Rank A is is the same whether we consider A as an element in $R^{m \times n}$ or in $C^{m \times n}$.
- Every statement above is **false** for order-k tensors, $k \geq 3$.

Some Open Problems

1. Extend theory of “nonnegative rank” for matrices to tensors (Cohen and Rothblum, 1993). Computing the regular rank of a tensor of order k $\geq$ 3 is np-hard.

2. Extend Perron-Frobenius theory to tensors in a meaningful way. Note, the spectra of tensors can be defined in various ways.

H.-K. Lim at Stanford MMDS workshop, June 2006 has provided a start. See:
http://www.stanford.edu/group/mmds/

NMF/NTF Applications to Data Analysis

- NMF Allows only additive, not subtractive combinations of the original data, in comparison to orthogonal decomposition methods, e.g. PCA.
- NMF used by Lee and Seung (MIT) in *Nature*, 1999, in biometrics, preceded and followed by numerous papers related to applications. Earlier work by Paatero, et al.
- Matlab Toolbox: NMFLAB, http://www.bsp.brain.riken.jp/
- Historical perspective:

  Problem 73-14, Rank Factorization of Nonnegative Matrices, by Berman and Ple., <u>SIAM Review</u> 15 (1973), p. 655: (Also in Berman/Ple. book)

Some General Applications of NMF/NTF Techniques

- Source separation in acoustics, speech, video
- EEG in Medicine, electric potentials
- Spectroscopy in chemistry
- Molecular pattern discovery - genomics
- Email surveillance
- Document clustering in text data mining
- Atmospheric pollution source identification
- Hyperspectral sensor data compression
- **Spectroscopy for space applications – spectral data mining**
  - **Identifying object surface materials and substances**

Outline

- Space Object Identification from spectral data (SOI)
- Nonnegative Matrix Factorization (NMF) Methods for Spectral Unmixing
- Nonnegative Tensor Factorization (NTF) for SOI
- 3-D Tensor Factorization: Applications to SOI, using Hyperspectral Data
  - Compression
  - Material identification
  - Spectral Abundances
  - Another Application (Medical)

Blind Source Separation for Finding Hidden Components (Endmembers)

Mixing of Sources

…basic physics often leads to linear mixing…

$\boldsymbol{X} = [\mathbf{X_1,X_2, \ldots,X_n}]$ –column vectors (1-D spectral scans)

Approximately factor

$$\boldsymbol{X} \approx \boldsymbol{W} \boldsymbol{H} = \sum_1^k \boldsymbol{w}^{(j)} \circ \boldsymbol{h}^{(j)}$$

$\circ$ denotes outer product

$\boldsymbol{w}^j$ is jth col of $\boldsymbol{W}$, $\boldsymbol{h}^j$ is jth col of $\boldsymbol{H}^T$

$\boldsymbol{X}$ sensor readings (mixed components – observed data)

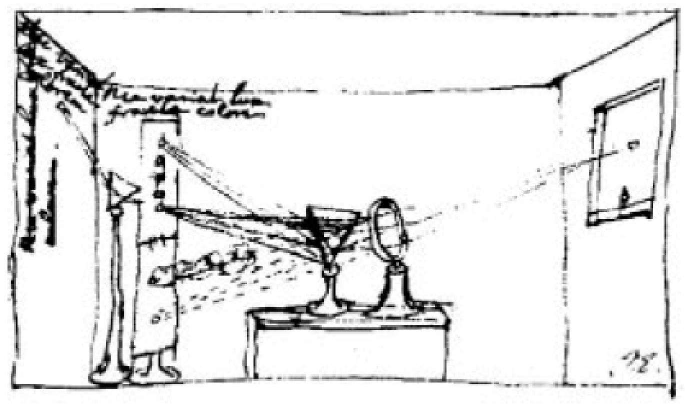
$\boldsymbol{W}$ separated components (feature basis matrix, unknown, low rank)

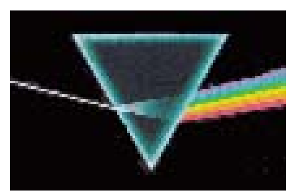
$\boldsymbol{H}$ hidden mixing coefficients (unknown), replaced later with abundances of materials that make up the object.

Simple Analog Illustration

Hidden Components in Light – Separated by a Prism

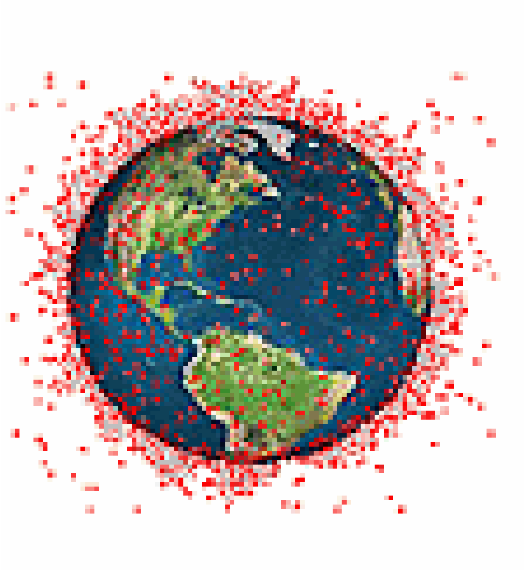
From Newton's Notebook

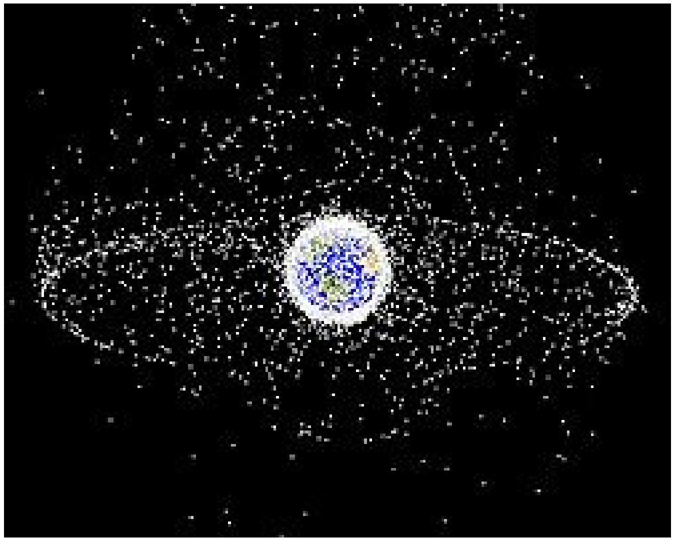




Our purpose – finding hidden components by <u>data analysis</u>

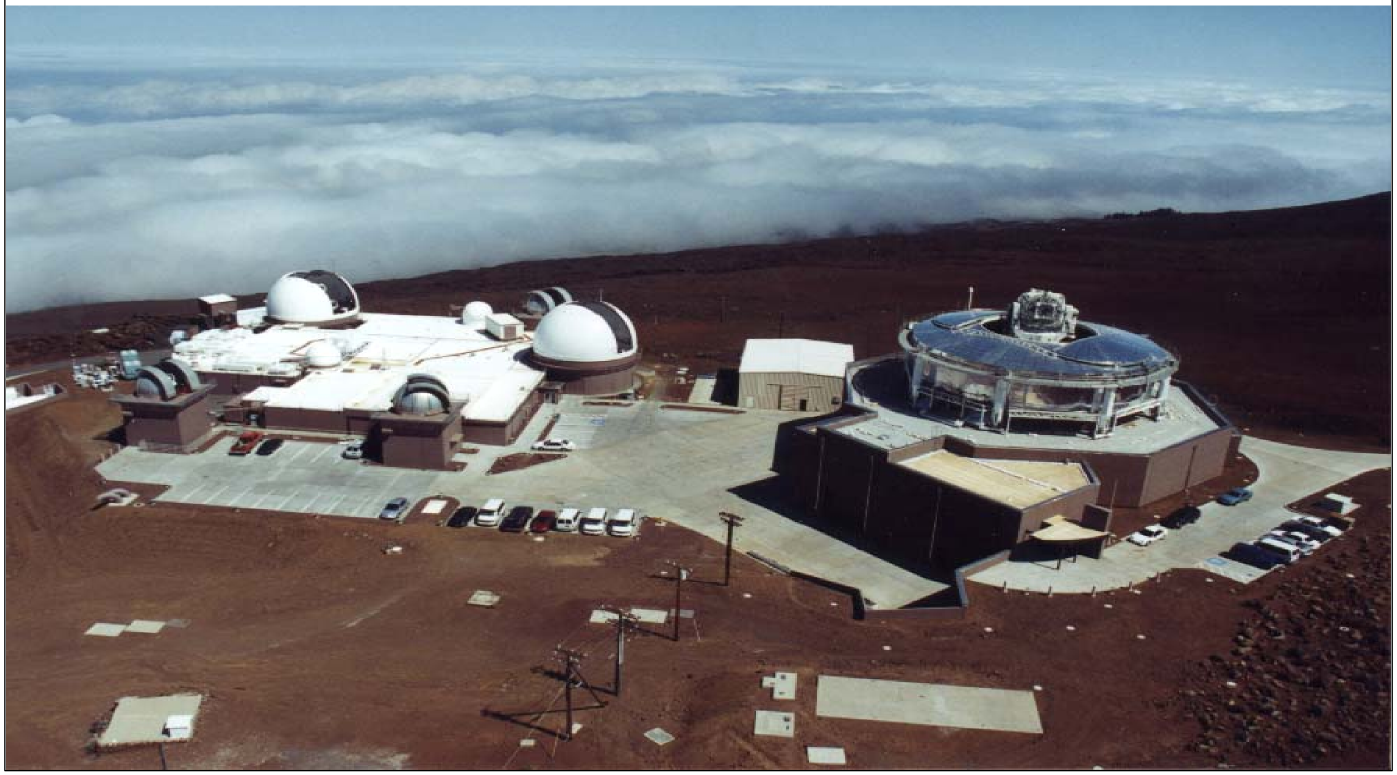
Space Object Identification and Characterization from <u>Spectral Reflectance Data</u>



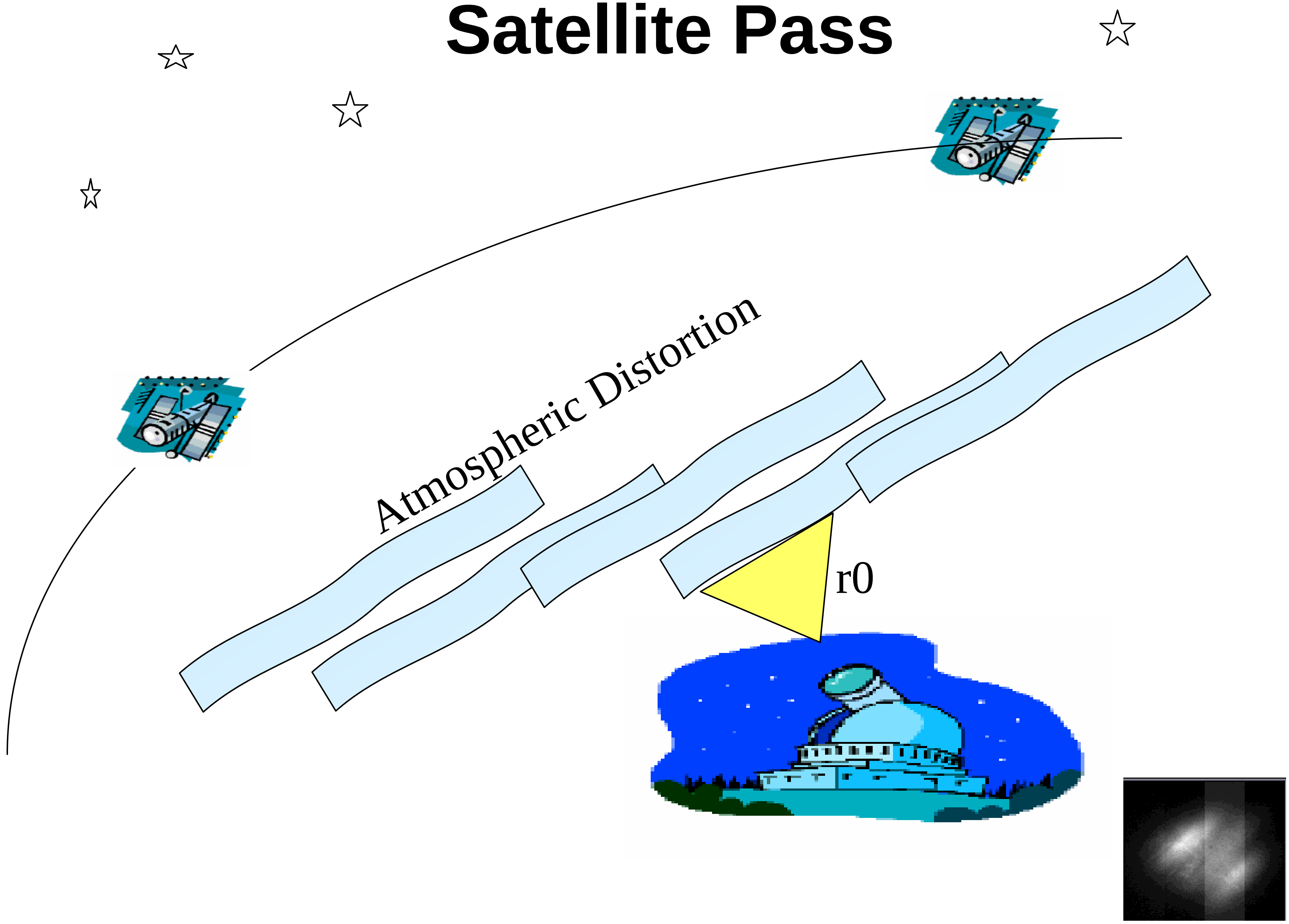


More than 20,000 known objects in orbit: various types of military and commercial <u>satellites</u>, rocket bodies, residual parts, and debris – need for space object database mining, object identification, clustering, classification, etc.

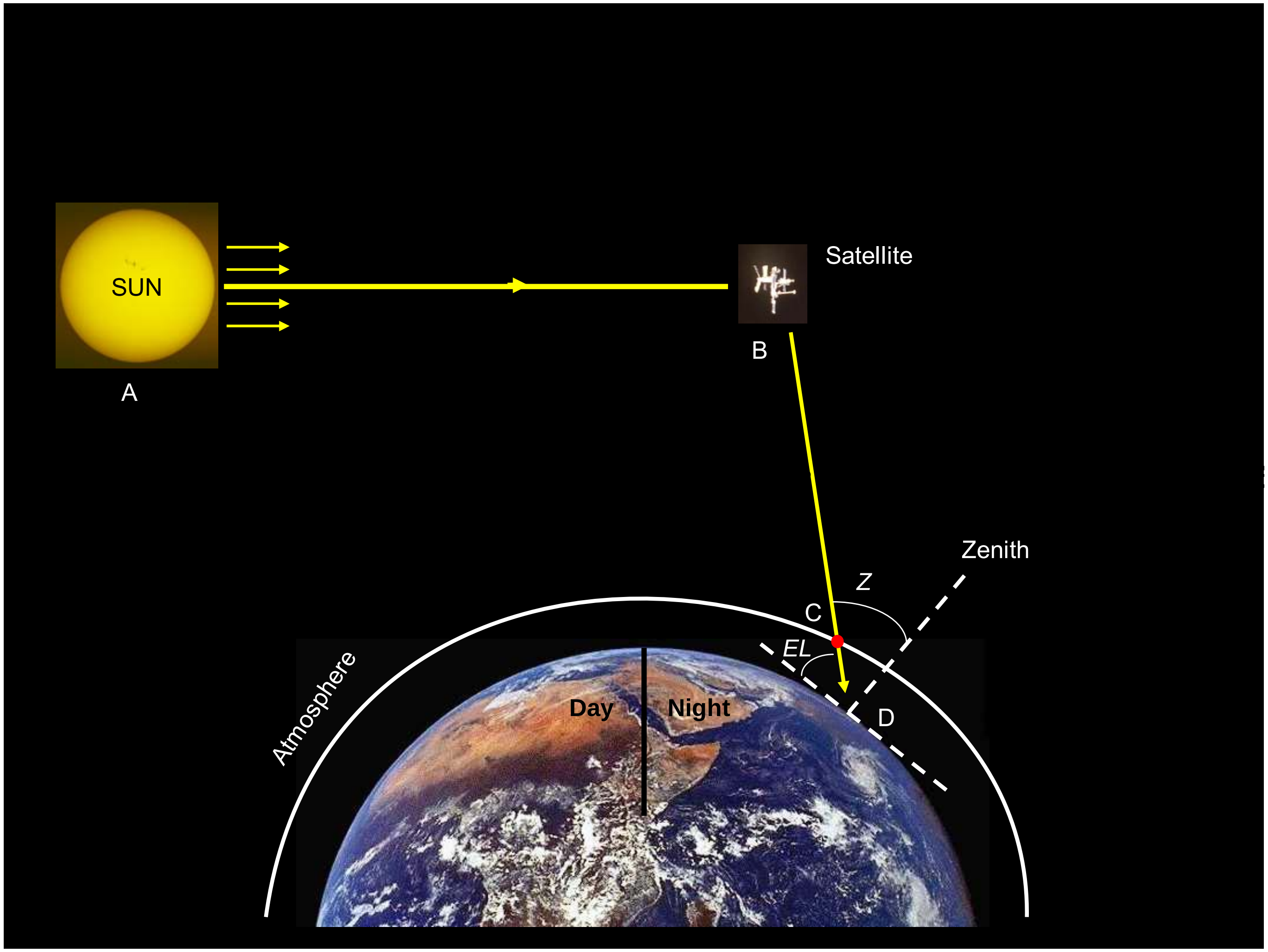
Maui Space Surveillance Site



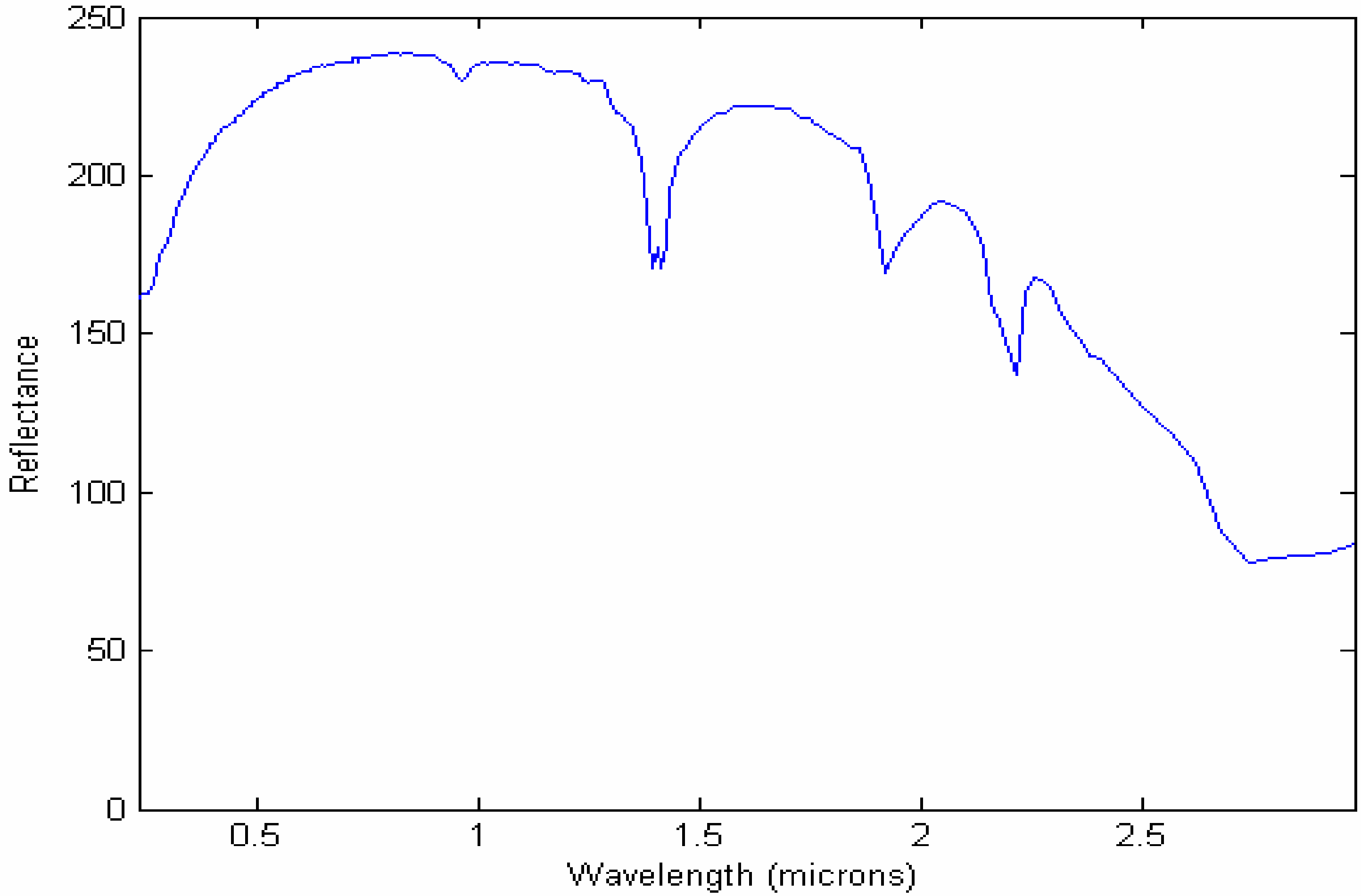
Satellite Pass



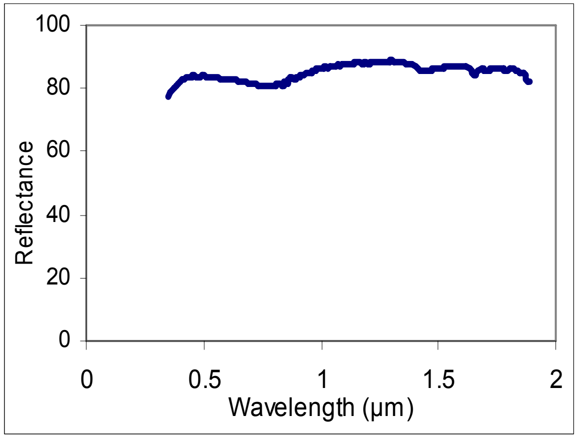
The creation and observation of a reflectance spectrum



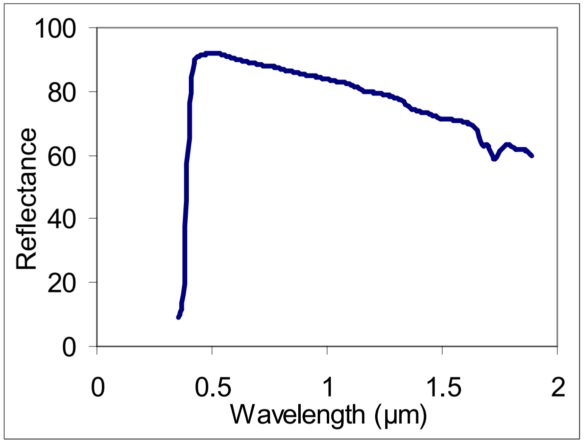
Typical Scan



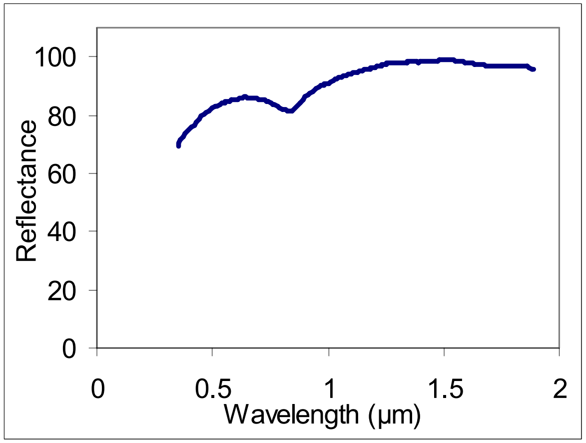
Some Laboratory Electromagnetic Spectral Signatures



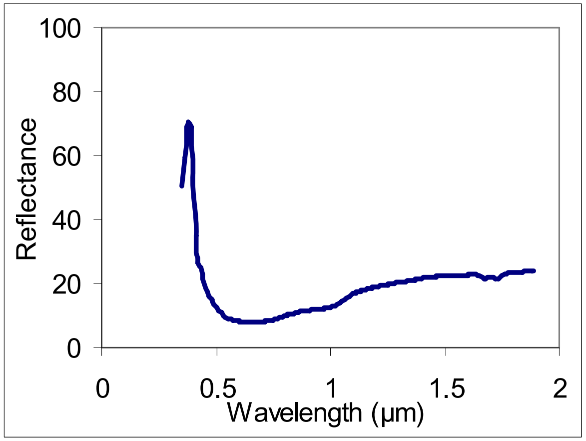
Mylar



White Paint



Aluminum



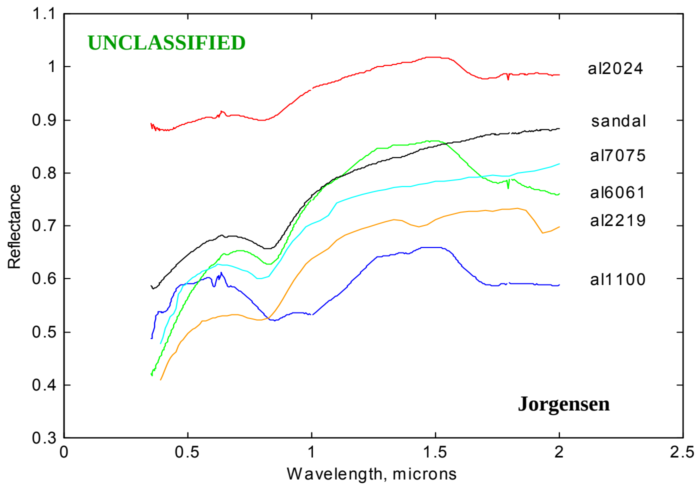
Solar Cell

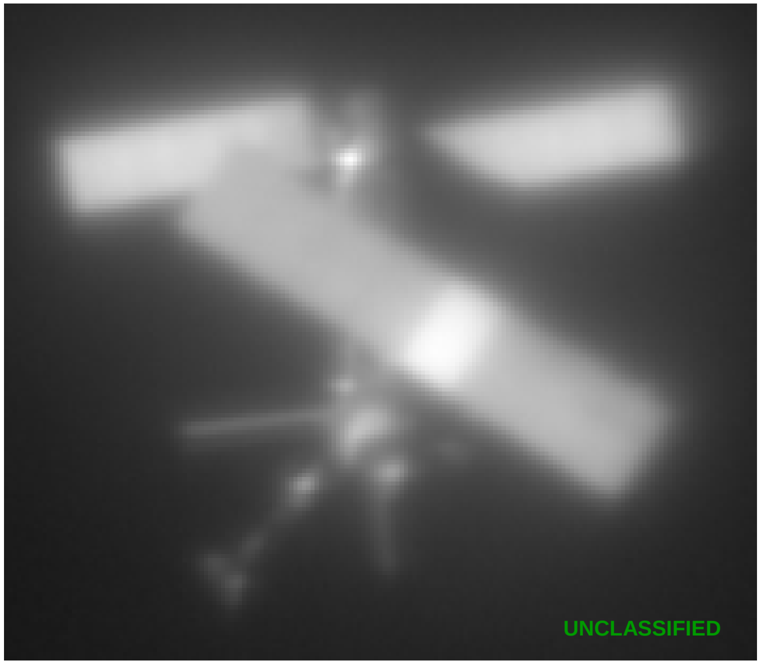
Spectral Imaging of Space Objects

- Current "operational" capability for spectral imaging of space objects

- Panchromatic images

- Non-imaging spectra





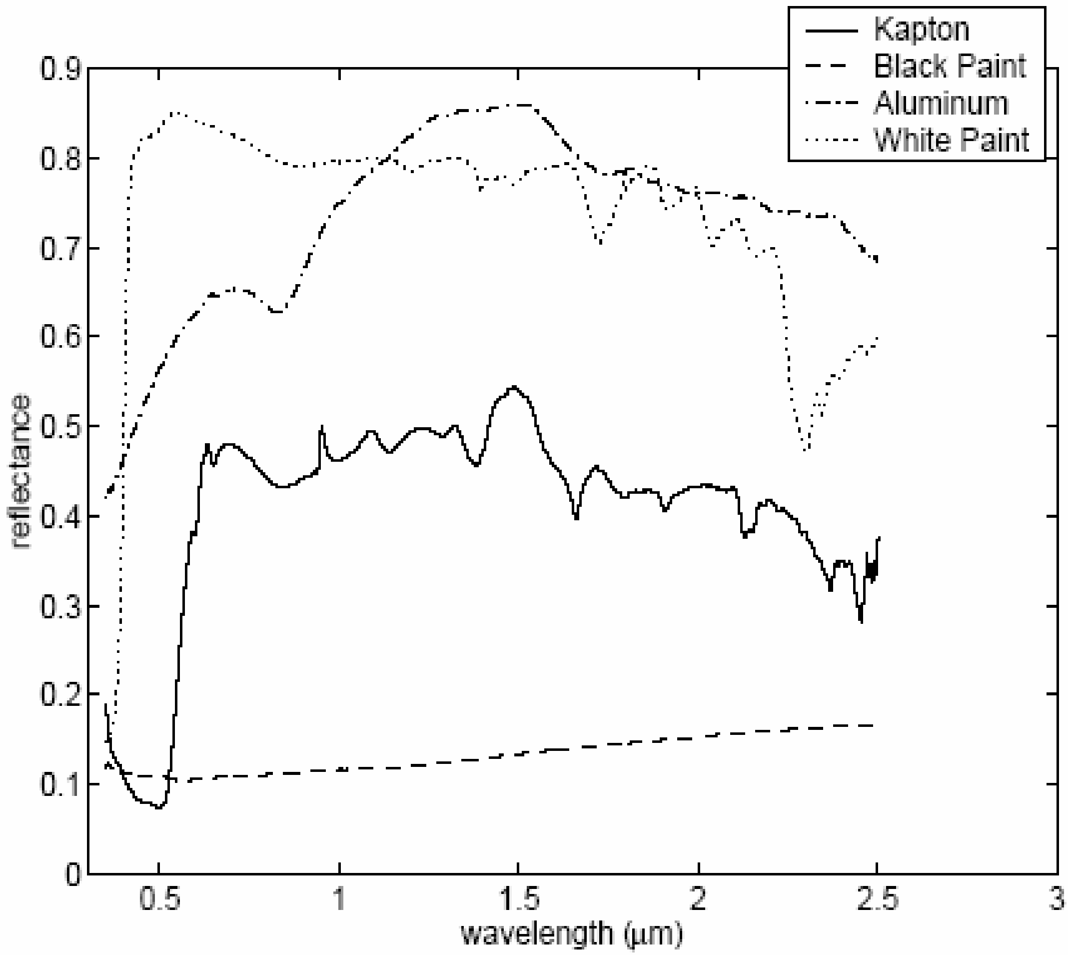
Solve the following constrained optimization problem. (Here α and β are regularization parameters).

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

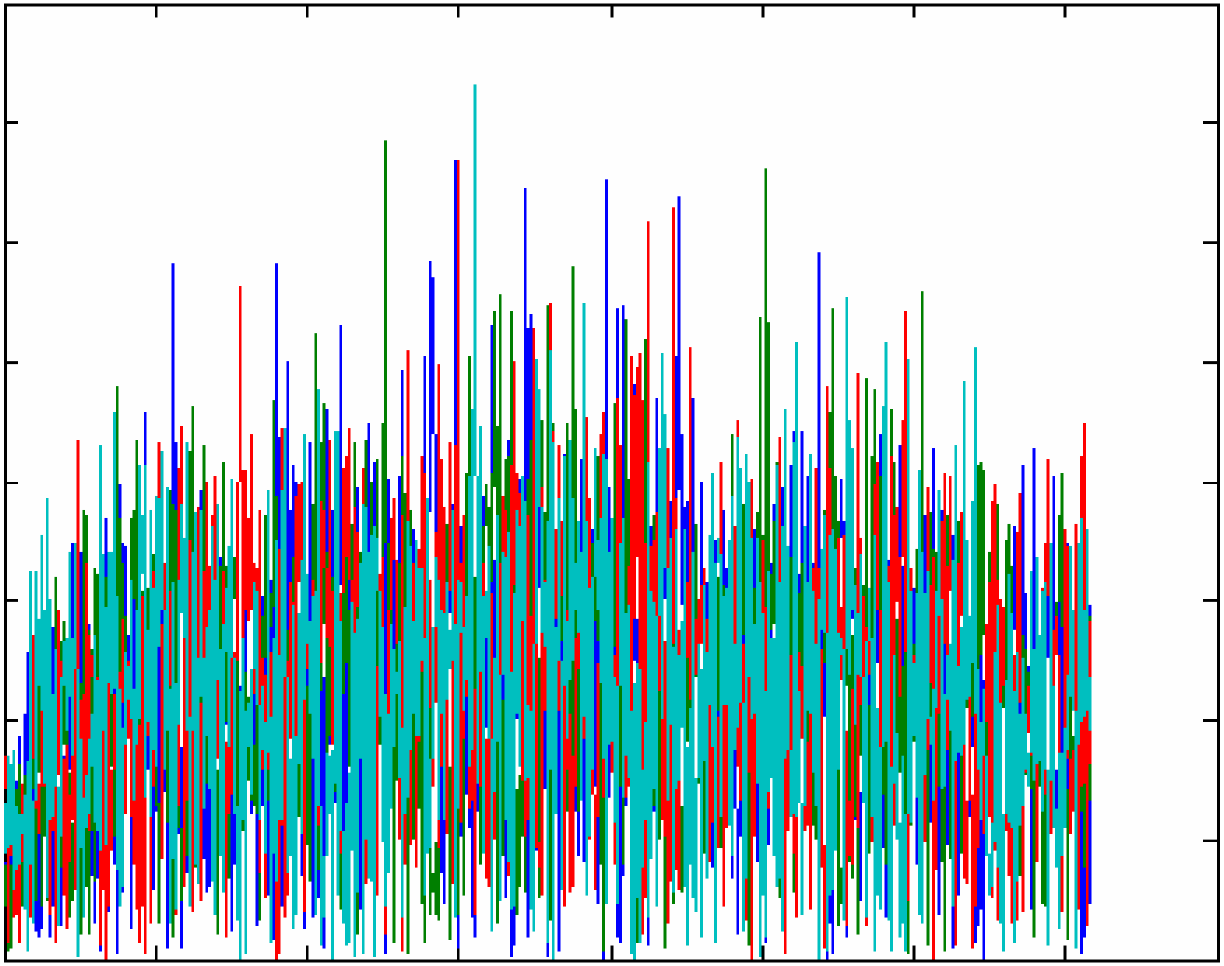
where $\alpha J_1(W)$ and $\beta J_2(H)$ are used to enforce certain application-dependent characteristics on the solution

- **Determine <u>gradients</u> for W and H and set each to zero (alternating iterations).**
- **<u>Initialize</u> W and H using PCA, based on Perron-Frobenius theory.**

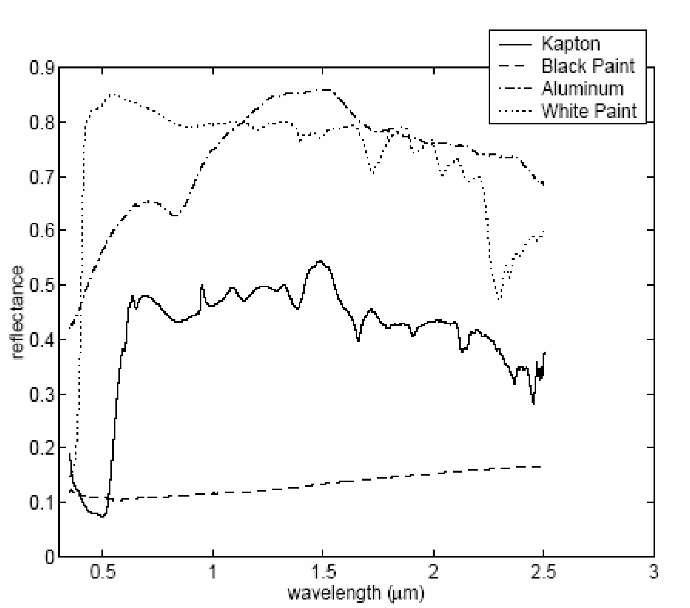
Sample Results – Finding only Endmembers We Form Simulated Satellites from NASA Data Imaging a Single Pixel



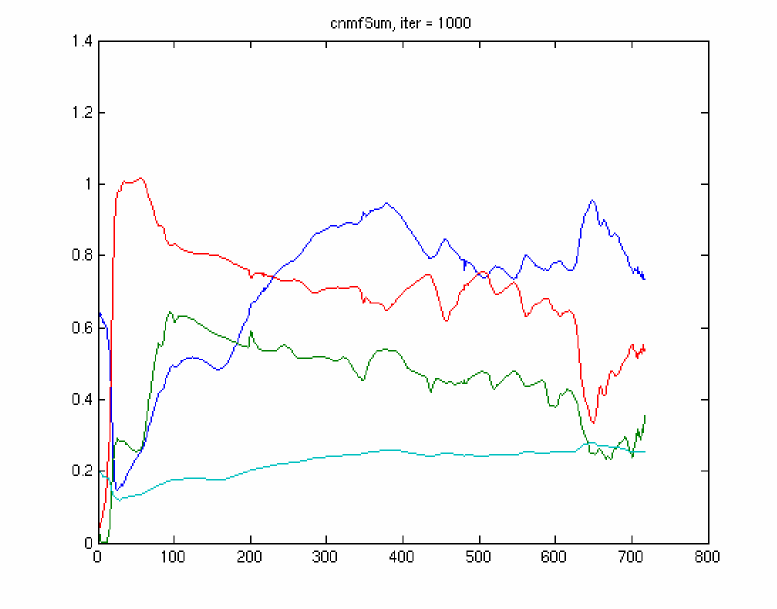
Blind Source Separation of Time Varying Mixture Below Using NMF



Original



Recovered



Tensors

Higher Order Arrays

Recall NTF (for 3-D Arrays)

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

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

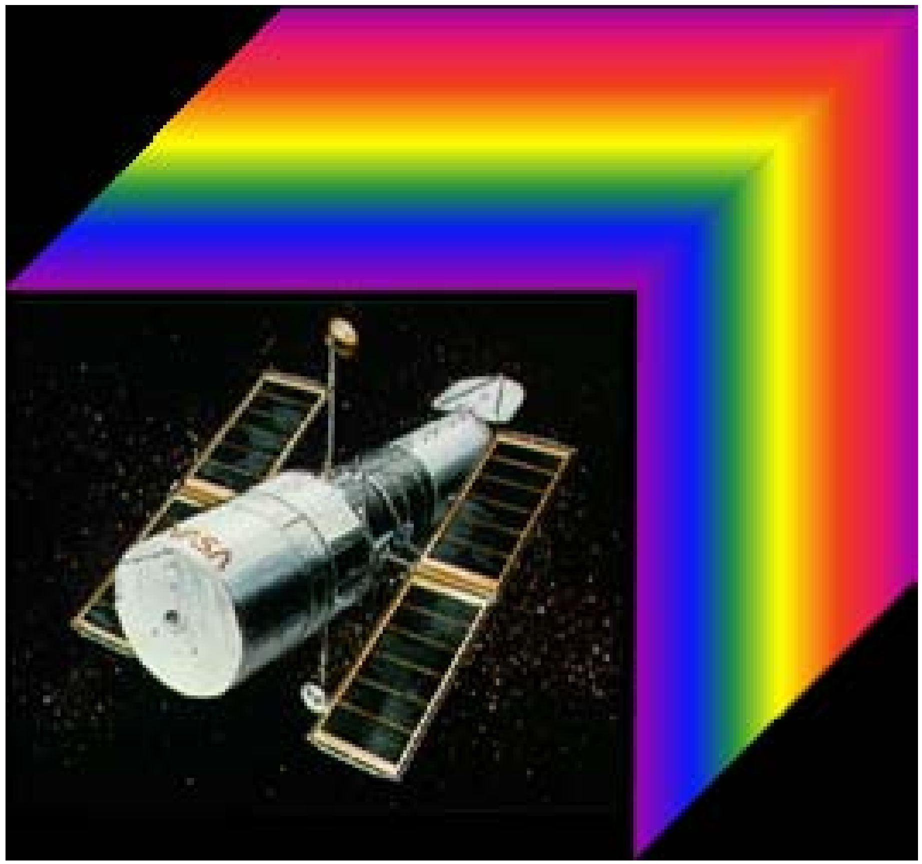
Here $\circ$ *denotes "outer product". The rank-one matrices* $u^{(i)} \circ v^{(i)}$ *are the desired basis components, and* $w^{(i)}$ *the weights.*

- Optimal solution exists (Lim 2006, Stanford MMDS workshop)
- Issues: initialization, efficient optimization algorithms, avoid local minima

Hyperspectral Imaging

- Hyperspectral remote sensing technology allows one to capture images objects (multiple pixels), using spectra from ultraviolet and visible to infrared.
- Multiple images of a scene or object are created using light from different parts of the spectrum.
- Hyperspectral images can be used to:
  - Identify surface minerals, objects, buildings, etc. from space.
  - Enable space object identification from the ground.

Hyperspectral Data Cube for Hubble Telescope

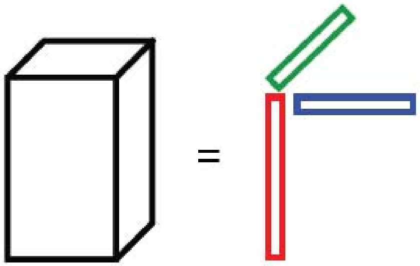


Outer, Kronecker, & Khatri-Rao Products

3-Way Outer Product

$$\mathcal{X} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c}$$

$$x_{ijk} = a_i b_j c_k$$



Rank-1 Tensor

Review: Matrix Kronecker Product

$$\underset{M \times N}{\mathbf{A}} \otimes \underset{P \times Q}{\mathbf{B}} = \underbrace{\begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1N}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2N}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{M1}\mathbf{B} & a_{M2}\mathbf{B} & \cdots & a_{MN}\mathbf{B} \end{bmatrix}}_{MP \times NQ}$$

$$= \begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{b}_1 & \mathbf{a}_1 \otimes \mathbf{b}_2 & \cdots & \mathbf{a}_N \otimes \mathbf{b}_Q \end{bmatrix}$$

Matrix Khatri-Rao Product

$$\underset{M \times R}{\mathbf{A}} \odot \underset{N \times R}{\mathbf{B}} = \underbrace{\begin{bmatrix} \mathbf{a}_1 \otimes \mathbf{b}_1 & \mathbf{a}_2 \otimes \mathbf{b}_2 & \cdots & \mathbf{a}_R \otimes \mathbf{b}_R \end{bmatrix}}_{MN \times R}$$

C.G. KHATRI AND C.R. RAO, Solutions to some functional equations and their applications to characterization of probability distributions, *Sankhya*, **30** (1968), 51–69.

C.R. RAO AND M.B. RAO, *Matrix Algebra and its Applications to Statistics and Econometrics*, World Scientific, Singapore, (1998).

NTF Algorithm

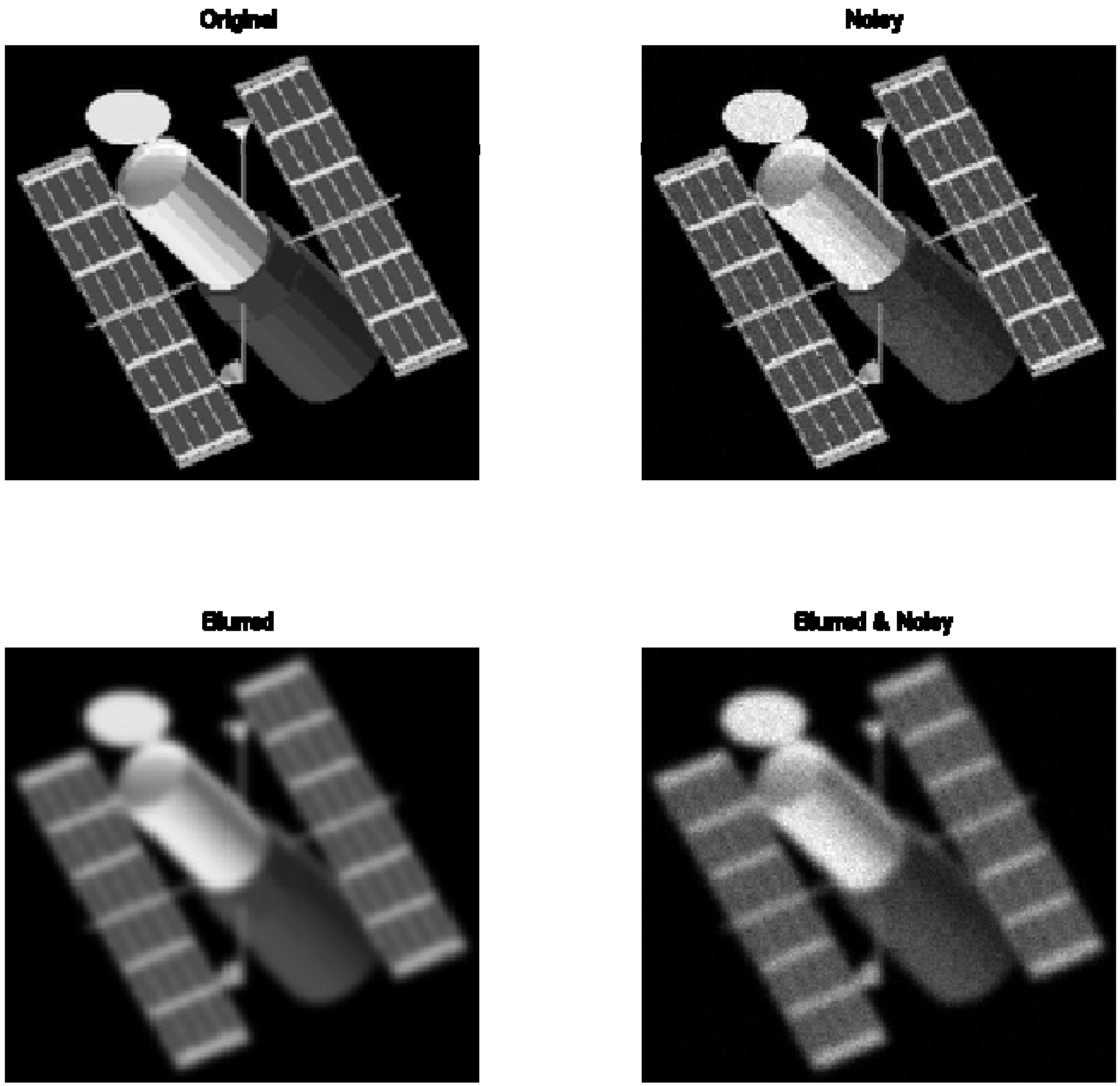
- Group $\mathbf{x}_i$'s, $\mathbf{y}_i$'s and $\mathbf{z}_i$'s as columns in $\boldsymbol{X} \in \mathbb{R}_+^{D_1 \times k}$, $\boldsymbol{Y} \in \mathbb{R}_+^{D_2 \times k}$ and $\boldsymbol{Z} \in \mathbb{R}_+^{D_3 \times k}$ respectively.
- Initialize $\boldsymbol{X}, \boldsymbol{Y}$.
  - Nonnegative Matrix Factorization of the mean slice,
$$\min ||\boldsymbol{A} - \boldsymbol{X}\boldsymbol{Y}||_F^2. \tag{1}$$
    where $\boldsymbol{A}$ is the mean of $\boldsymbol{\mathcal{T}}$ across the 3^{rd} dimension.
- Iterative Tri-Alternating Minimization

1. Fix $\boldsymbol{\mathcal{T}}, \boldsymbol{X}, \boldsymbol{Y}$ and fit $\boldsymbol{Z}$ by solving a NMF problem using a projected gradient descent algorithm [12].
$$\min_{\boldsymbol{Z}} ||\boldsymbol{T}_z - \boldsymbol{C}_z \boldsymbol{Z}||_F^2, \tag{2}$$
   where $C_z = \boldsymbol{X} \bigodot \boldsymbol{Y} \in \mathbb{R}^{D_1 D_2 \times k}$, T_z is the unfolding tensor across the 3^{rd} dimension and $\bigodot$ stands for the Khatri-Rao product.
2. Fix $\boldsymbol{\mathcal{T}}, \boldsymbol{X}, \boldsymbol{Z}$, fit for $\boldsymbol{Y}$, but add in the sparsity constraint.
$$\min_{\boldsymbol{Y}} ||\boldsymbol{T}_y - \boldsymbol{C}_y \boldsymbol{Y}||_F^2 + \alpha \frac{||\boldsymbol{Y}||_1^2}{||\boldsymbol{Y}||_2^2}, \tag{3}$$
3. Fix $\boldsymbol{\mathcal{T}}, \boldsymbol{Y}, \boldsymbol{Z}$ and fit for $\boldsymbol{X}$.
$$\min_{\boldsymbol{X}} ||\boldsymbol{T}_x - \boldsymbol{C}_x \boldsymbol{X}||_F^2 + \alpha \frac{||\boldsymbol{X}||_1^2}{||\boldsymbol{X}||_2^2}, \tag{4}$$

Simulated Data

- 177 x 193 x 100   3-D model of Hubble satellite.
- Assign each pixel a certain spectral signature from lab data supplied by Kira Abercromby (NASA). 8 materials used.
- Bands of spectra ranging from .4 μm to 2.5 μm, with 100 evenly distributed spectral values.
- Spatial blurring followed by Gaussian and Poisson noise and applied over the spectral bands.

Images from Data Cube at 1.4 μm



Materials Assigned to Pixels

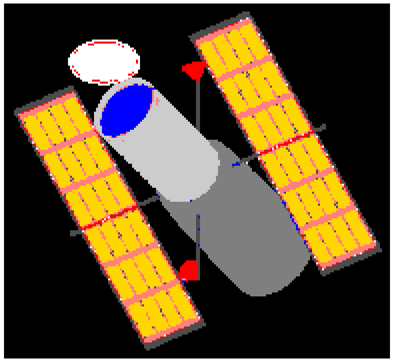
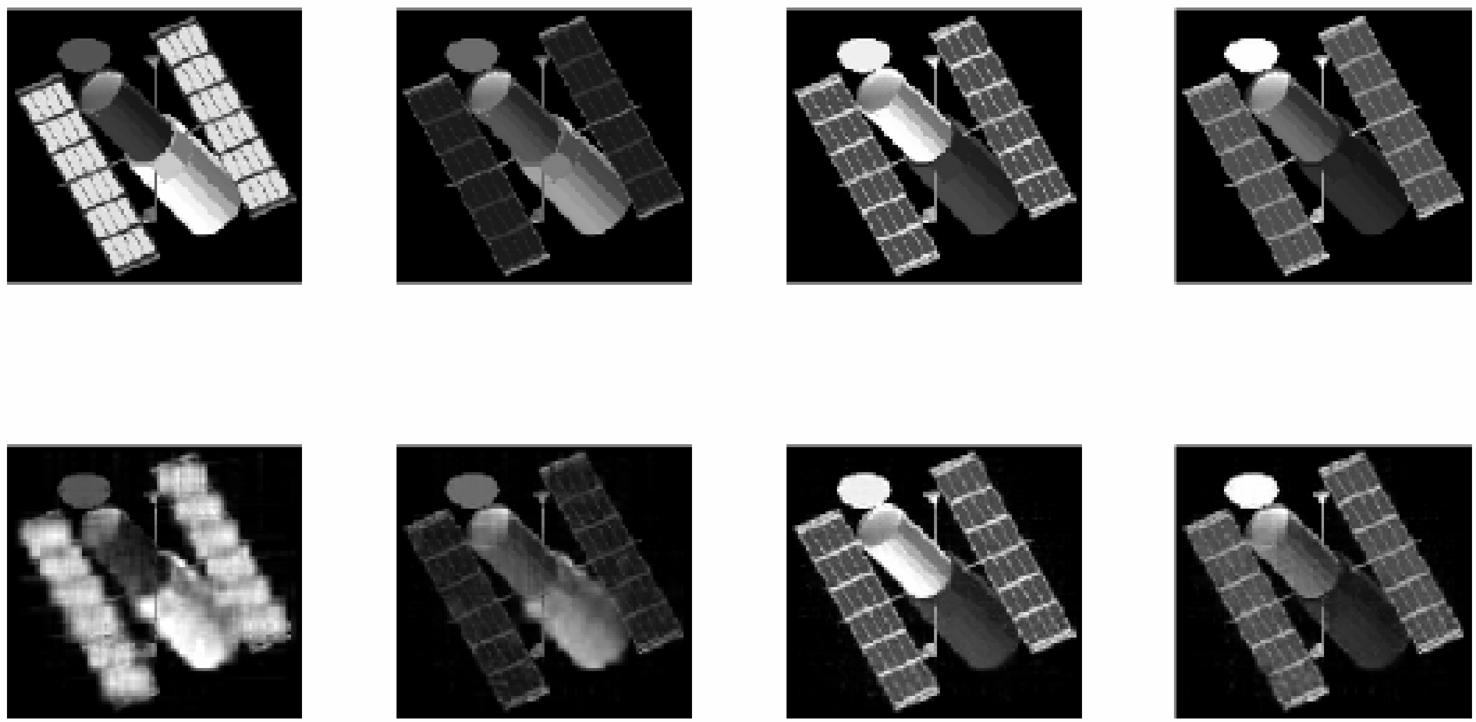


Table 1: Materials, colors and fractional abundances

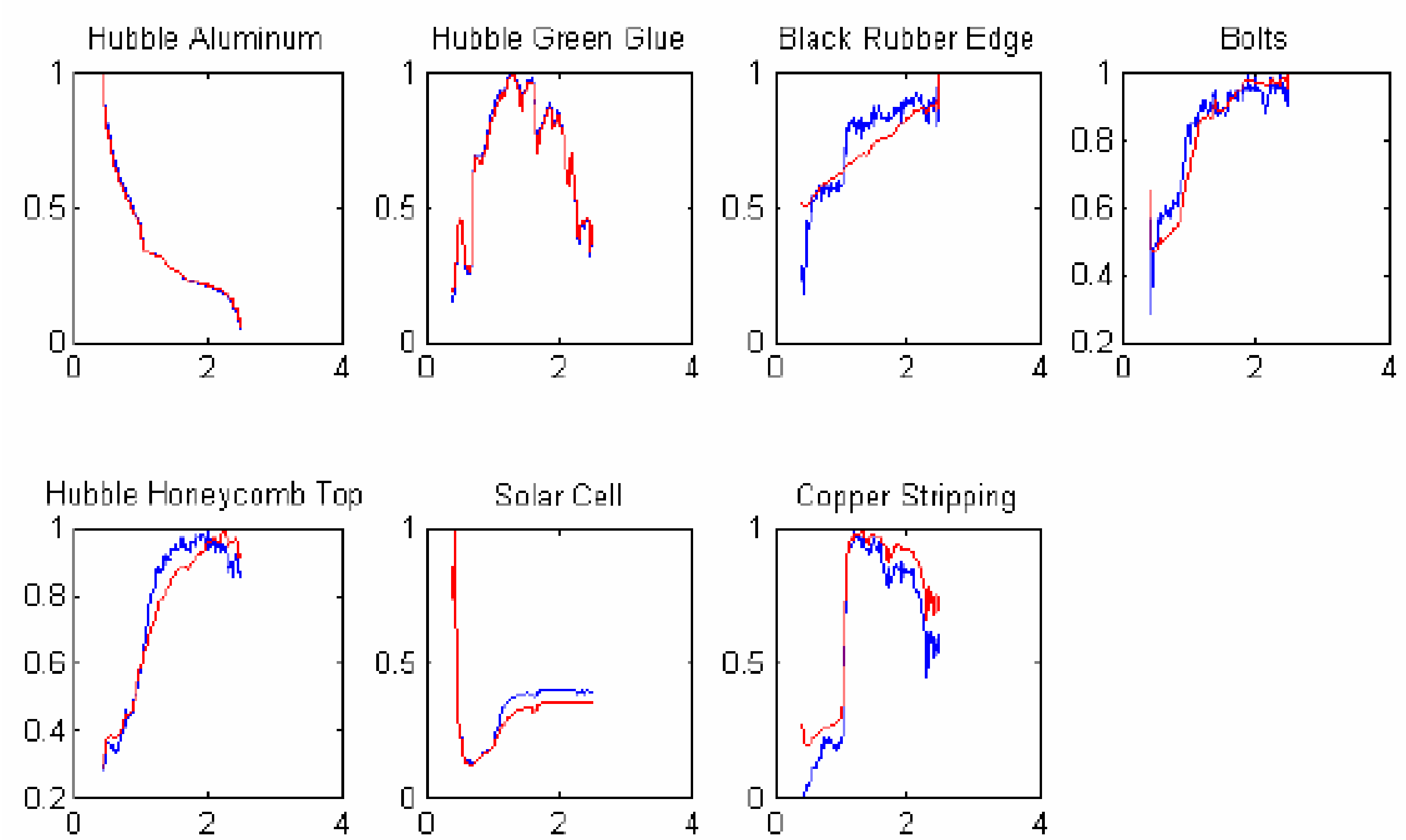
Material	Color	Fractional Abundance (%)
Hubble Aluminum	light gray	19
Hubble Green Glue	dark gray	12
Hubble Honeycomb Top	white	4
Hubble Honeycomb Side	blue	3
Solar Cell	gold	37
Bolts	red	3
Black Rubber Edge	dark gray	8
Copper Stripping	cyan	13

Reconstruction of Blurred & Noisy Data at wavelengths: .42, .68, 1.55, 2.29 μm



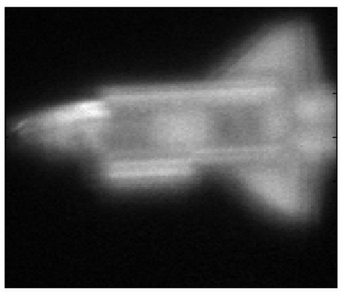
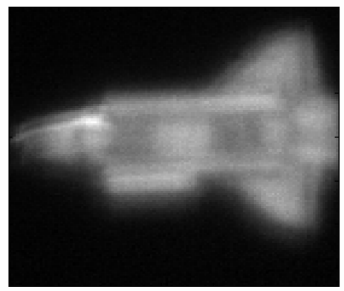
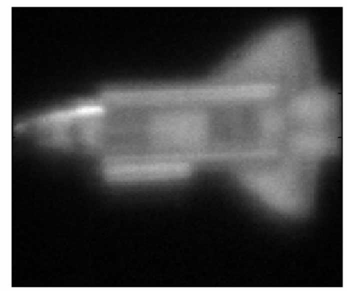
Top row – Original Images
Bottom row – Reconstructions
Compression factor - 73

Some Spectral Scan Matching Graphs

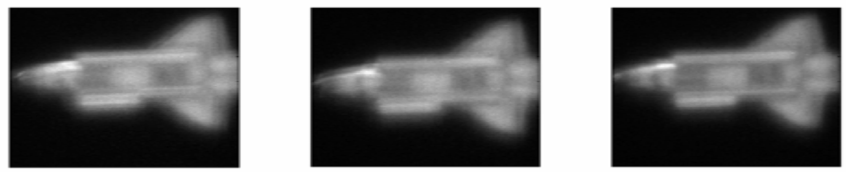


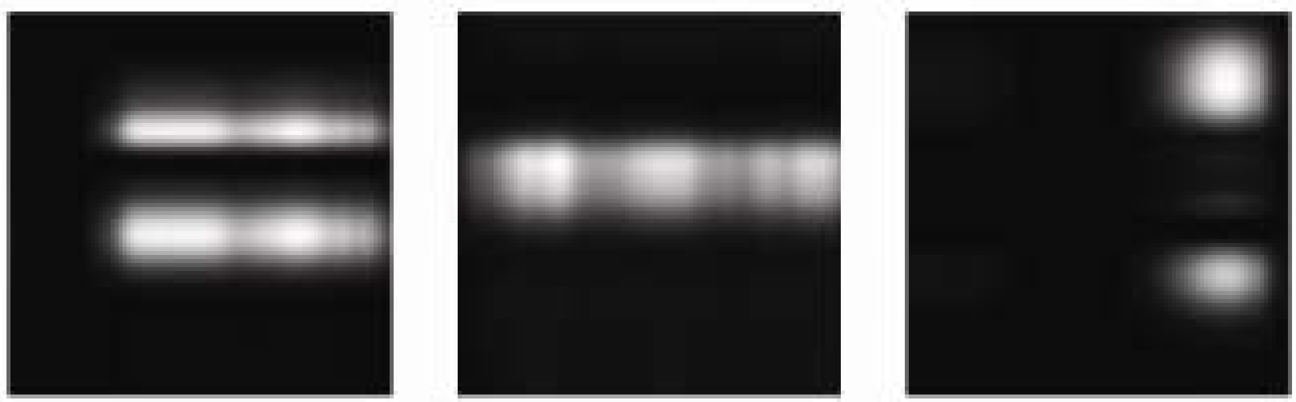
Red is original, Blue is matched endmember scan matched from Z-factors. Black rubber edge (8%) not as well matched. Blurred and noisy data cube.

Telescope images taken at the Air Force Maui Space Center of the shuttle Columbia on it's final orbit before disintegration on re-entry, Feb. 2003. Below are 3 of 64 video images.

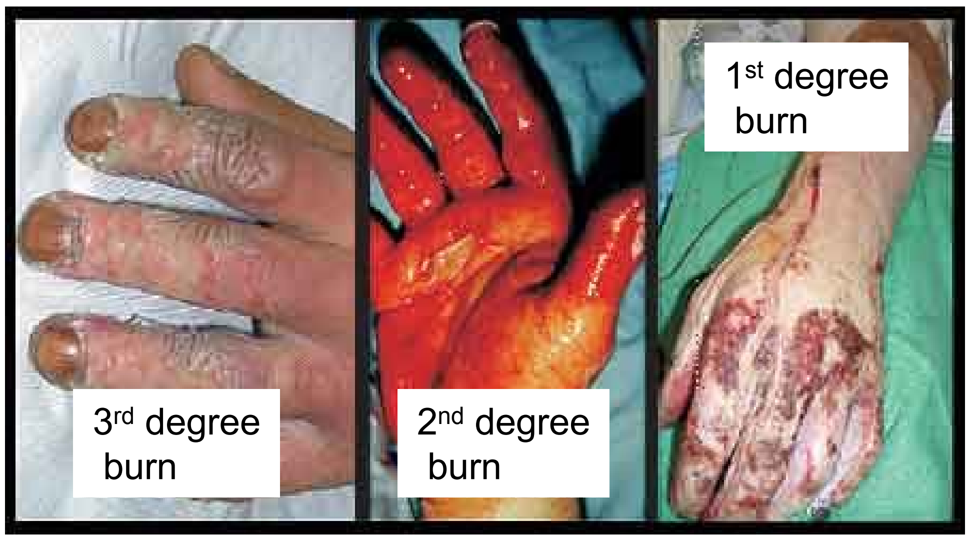
Separation by Parts using NTF





Potential Applications using Spectral Imaging and Tensor Analysis with NTF

- IARPA: Burn and tissue wound assessment (with the WFU Hospital Burn Center)



Summary

- Low-rank matrix factorizations for data analysis, PCA, IGA…
- Nonnegative matrix factorization (NMF)
- Nonnegative tensor factorization (NTF), brief overview and some open problems
- An Application: Object Identification from Spectral Data
- Computations using AF data from Maui and NASA
- Recent applications to SOI and biomedical data analysis