Multivariate Curve Resolution of Hyperspectral Images: Initialization and Functional Constraints

Neal B. Gallagher, Jeremy M. Shaver,
Barry M. Wise
Eigenvector Research, Inc.

FACSS, Portland, OR 2004

EIGENVECTOR RESEARCH INCORPORATED

Multivariate Curve Resolution

- MCR is most often used with spectra
  - aka end member extraction", self-modeling curve resolution
- Literature filled with examples from evolving data
  - LC-MS, GC-NIR, GC-GC …
- Newer examples include multivariate images
  - Image Mid-IR, NIR, UV-Vis ...

EIGENVECTOR RESEARCH INCORPORATED

MCR

- Based on the classical least squares (CLS) model, attempt to estimate **C** and **S** given **X**:

$$\mathbf{X} = \mathbf{C}\mathbf{S}^T + \mathbf{E}$$

where

X is a MxN matrix of measured responses,

C is a MxK matrix of pure analyte contributions,

S is a NxK matrix of pure analyte spectra, and

E is a MxN matrix of residuals.

EIGENVECTOR RESEARCH INCORPORATED

MCR Similar to PCA

- CLS and PCA look similar but with different mathematical properties

$\mathbf{X} = \mathbf{C}\mathbf{S}^T + \mathbf{E}$	$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E}$
$\mathbf{X}_{M\text{x}N}$	$\mathbf{X}_{M\text{x}N}$
$\mathbf{C}_{M\text{x}K}$, $c_{ij} \geq 0$	$\mathbf{T}_{M\text{x}K}$, $\mathbf{T}^T\mathbf{T}$ KxK diagonal
$\mathbf{S}_{N\text{x}K}$, $s_{ij} \geq 0$	$\mathbf{P}_{N\text{x}K}$, $\mathbf{P}^T\mathbf{P} = \mathbf{I}_{K\text{x}K}$
$\mathbf{E}_{M\text{x}N}$	$\mathbf{E}_{M\text{x}N}$

EIGENVECTOR RESEARCH INCORPORATED

ALS

- Alternating and constrained least squares is often used to get estimates of **C** and **S** e.g.
- 0) Start with a guess for $\mathbf{C} = \mathbf{C}_0$ (or $\mathbf{S}_0$)
- 1) Estimate $\mathbf{S}^T_r = \mathbf{C}^\dagger_{r-1}\mathbf{X}$ subject to $\mathbf{S}(i,j) \geq 0$
- 2) Estimate $\mathbf{C}_r = \mathbf{X}\mathbf{S}^{\dagger T}_r$ subject to $\mathbf{C}(i,j) \geq 0$
- 3) Continue Steps 1 and 2 until $\|\mathbf{E}\|$ within some tolerance or $r > r_{max}$

EIGENVECTOR RESEARCH INCORPORATED

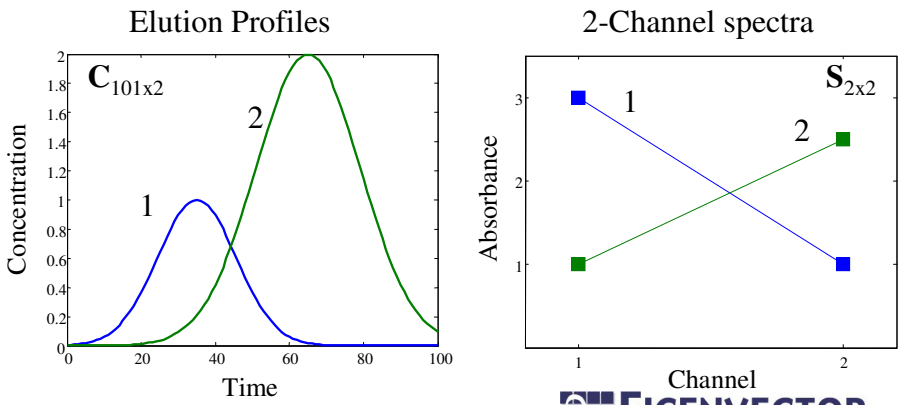
Initial Estimate

- With evolving data
  - can use Evolving Factor Analysis (EFA) and Evolving Window Factor Analysis (EWFA) to get a first estimate of **C**
  - more difficult to do with images since the pixels are not usually "ordered" in time
- Try to find "extreme" samples/pixels

EIGENVECTOR RESEARCH INCORPORATED

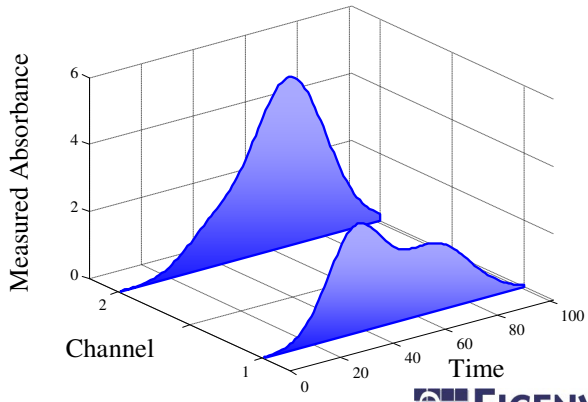
Example with Evolving Data

- Synthetic data from LC-NIR

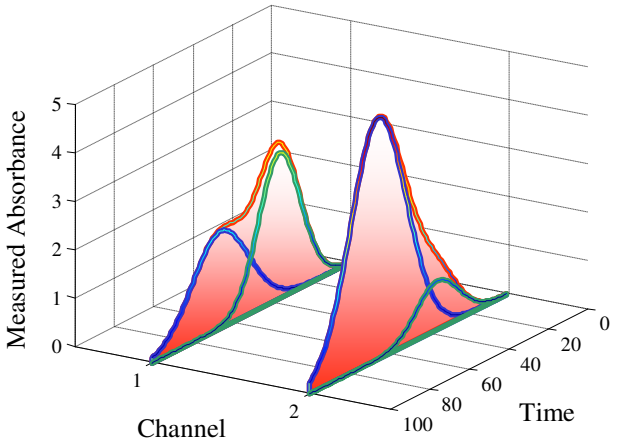


Elution Example No Noise

- Measured response (no noise)

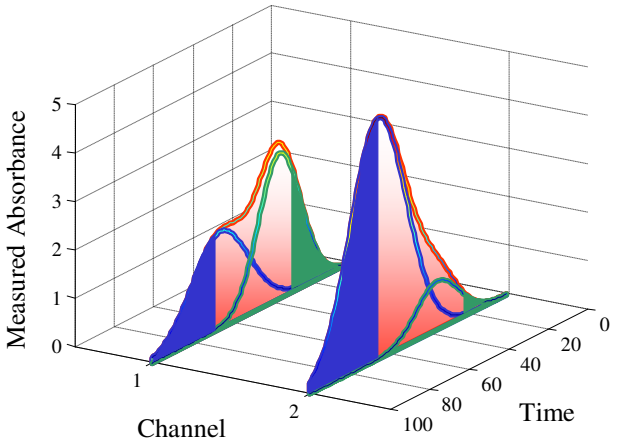


Measured Response (no noise)



EIGENVECTOR RESEARCH INCORPORATED

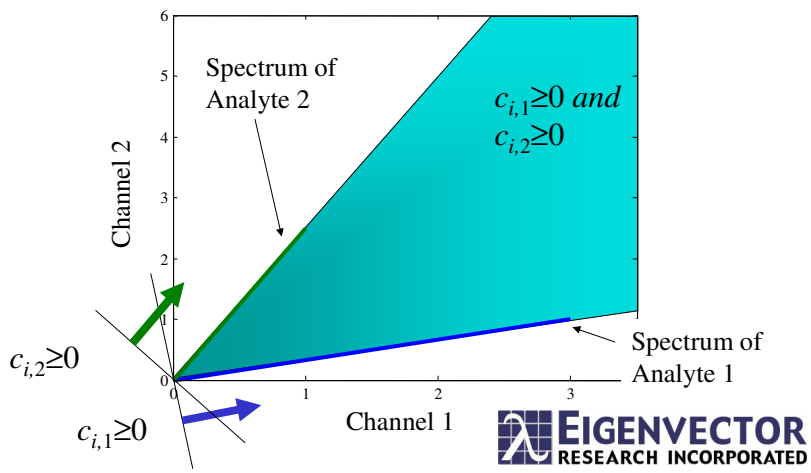
Measured Response (no noise)



EIGENVECTOR RESEARCH INCORPORATED

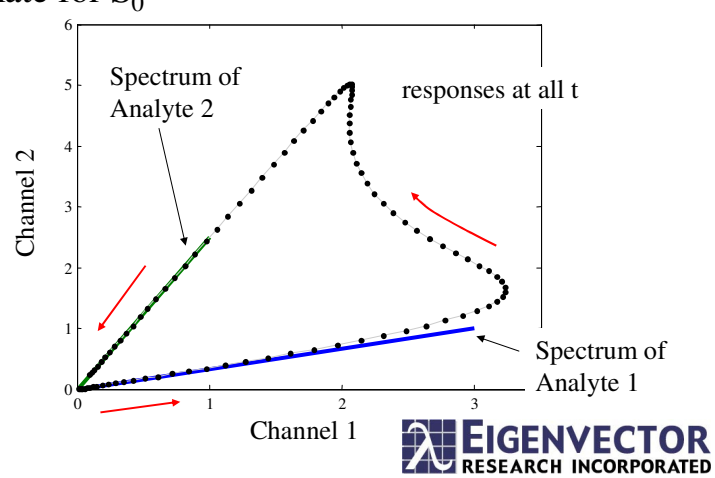
Non-negativity

- Plot **S** (Channel 2 versus 1)



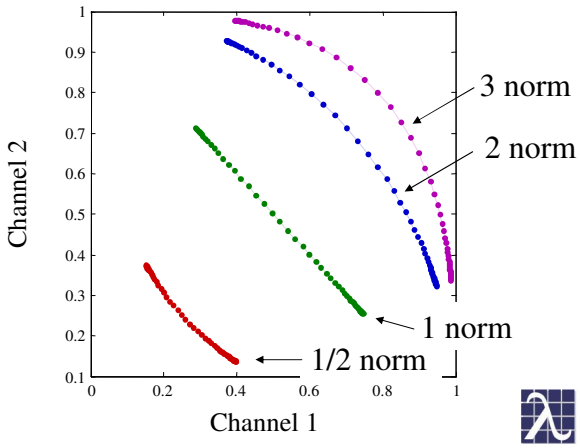
Elution Example No Noise

- Samples at the boudaries (extremes) are best estimate for $\mathbf{S}_0$



How to find the Extremes?

- Normalize each spectrum (which p?)



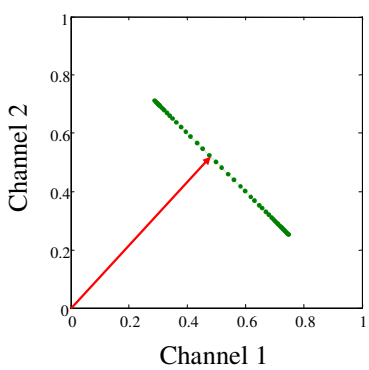
p norm

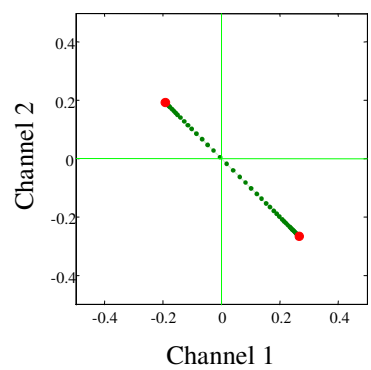
$$\mathbf{x} = \mathbf{x}\left(\sum_{j=1}^{N} x_j^p\right)^{1/p}$$

EIGENVECTOR RESEARCH INCORPORATED

Extreme Samples

- Mean centering the 1 norm spectra drops the rank





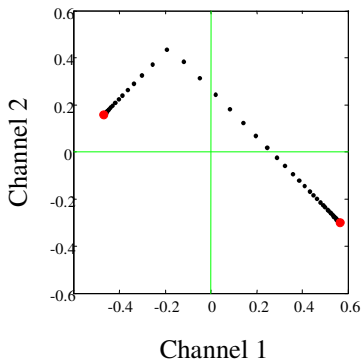
the extreme sample spectra are indistinguishable from the original analyte spectra

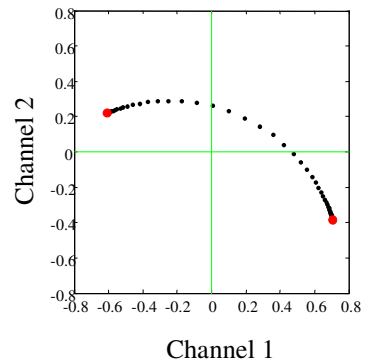
samples with 0 norm not used

EIGENVECTOR RESEARCH INCORPORATED

1 and 2 Norm

- Can use 2 norm to find extremes also
  - note that one of the pure spectra has a negative

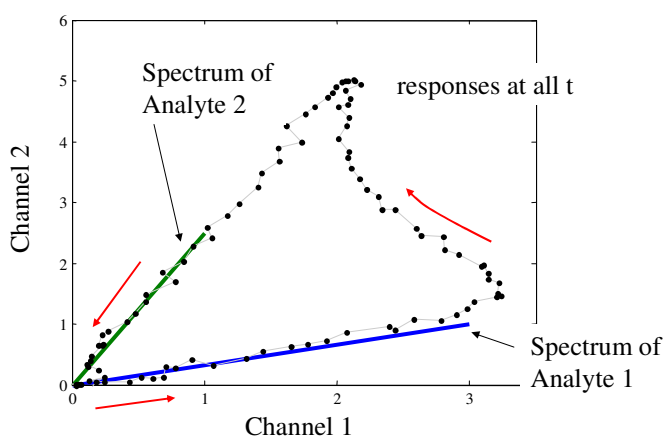




EIGENVECTOR RESEARCH INCORPORATED

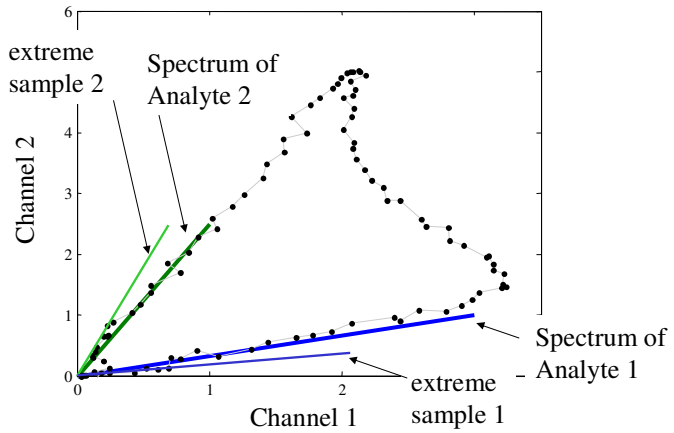
What Happens with Noise?

- Estimate extremes using samples with higher signal



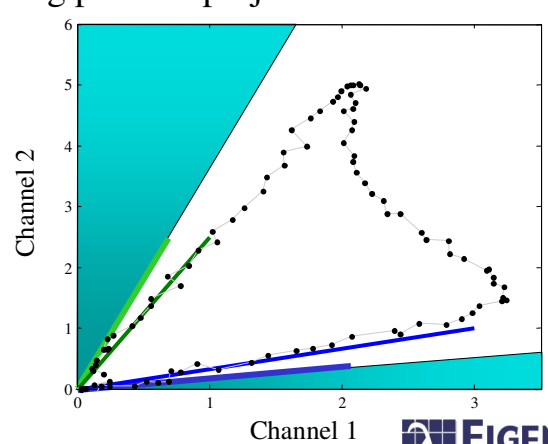
What Happens with Noise?

- Extremes from samples with norm >0.5



Feasible Solution

- Set of spectra (s_{ij}>0) that bounds the data with the c_{ij} having positive projections is a feasible solution



EIGENVECTOR RESEARCH INCORPORATED

Rotational Ambiguity

- For an invertible matrix **A**
  - such that constraints are satisfied
- Result is a rotational ambiguity
  - all solutions have the same fitness **E**

$$\mathbf{X} = \mathbf{C}\mathbf{S}^T + \mathbf{E} = (\mathbf{C}\mathbf{A})\left(\mathbf{A}^{-1}\mathbf{S}^T\right) + \mathbf{E}$$

Tauler, R., "Calculation of maximum and minimum band boundaries of feasible solutions for species profiles obtained by multivariate curve resolution", *J. Chemo.*, **15**, 627-646, 2001.

EIGENVECTOR RESEARCH INCORPORATED

Multiplicative Ambiguity

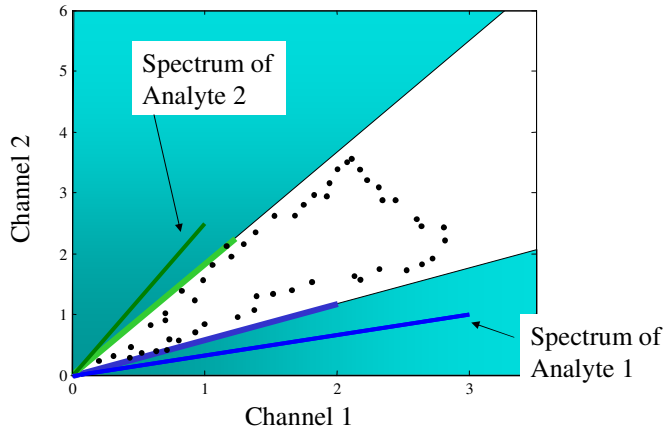
- For a diagonal matrix **A**
  - with non-zero elements
- Result is a multiplicative ambiguity
  - all solutions have the same fitness **E**

$$\mathbf{X} = \mathbf{C}\mathbf{S}^T + \mathbf{E} = (\mathbf{C}\mathbf{A})\left(\mathbf{A}^{-1}\mathbf{S}^T\right) + \mathbf{E}$$

EIGENVECTOR RESEARCH INCORPORATED

Lack of Selectivity

- Extremes are inside pure spectra



Manne, R., "On the Resolution Problem in Hyphenated Chromatography", *Chemo. Intll. Lab. Sys.*, **27**, 89–94, 1995.

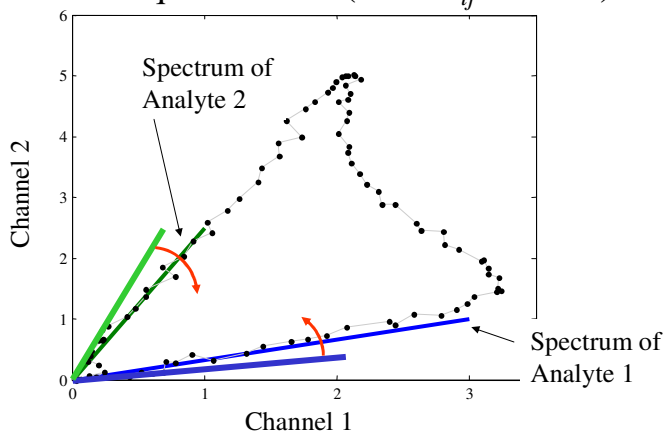
Use What You Know

- Multiplicative and rotational ambiguities, and lack of selectivity can some times be resolved using constraints
  - non-negativity, unimodality, smoothness, equality …
  - functional (chemistry/physics)
  - these can be easily incorporated into an ALS algorithm

EIGENVECTOR RESEARCH INCORPORATED

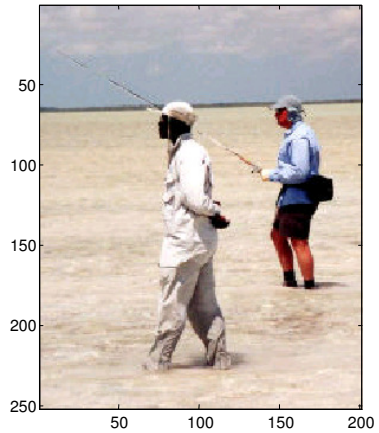
What Happens with Noise?

- With appropriate constraints the ALS also fits the data in a least squares sense (some c_{ij}< ~ 0 ok)



Try an Example with an Image

- RGB image
  - 251x201x3
  - fly fishing for bone fish
- What are the 3 pure component spectra?
  - matricize to (251*201)x3 and look for the extremes



EIGENVECTOR RESEARCH INCORPORATED

Three Extreme Pixels

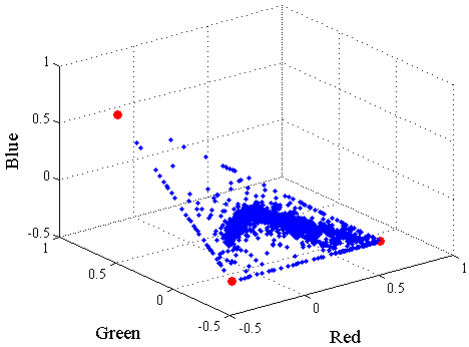
- 3 extreme samples are on the guide's face (red pixels)
  - in the shade where the signal is low
  - values are:

[12 0 0
0 0 5
0 1 0]

EIGENVECTOR RESEARCH INCORPORATED

Plot of Normalized Pixels

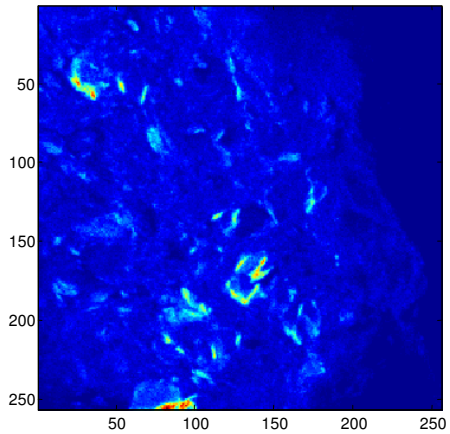
```
x = imread('c:\program files\qualcomm\eudora\attach\flat_harlan.jpg');
x = x(50:300,150:350,:);
y = double(reshape(x,251*201,3));
tol = 0.05;
[normy,norms] = normaliz(y,0,1);
i1  = find(norms>tol);
ymean = mncn(normy(i1,:));
isel = distslct(ymean,3);
y(isel,:)
  12    0    0
   0    0    5
   0    1    0
```



EIGENVECTOR RESEARCH INCORPORATED

Imaging Mass Spec

- Image is 256x256x90
- The mass spectrum was 41945 mass channels selected and binned into 90 channels
- Image of total ion count
  - false color



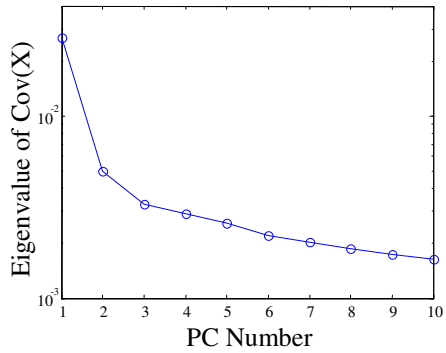
EIGENVECTOR RESEARCH INCORPORATED

PCA of the Image Data

1 norm, mean center for samples with norm>2

Percent Variance Captured by PCA Model

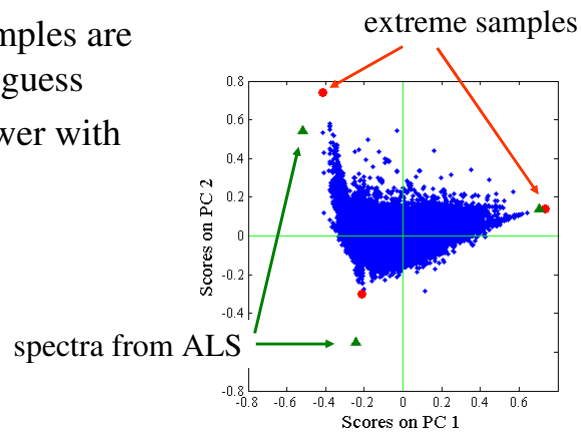
Principal Component Number	Eigenvalue of Cov(X)	% Variance Captured This PC	% Variance Captured Total
1	2.68e-002	35.40	35.40
2	4.97e-003	6.56	41.97
3	3.29e-003	4.34	46.31
4	2.90e-003	3.83	50.14
5	2.57e-003	3.40	53.54
6	2.22e-003	2.93	56.47
7	2.02e-003	2.67	59.15
8	1.88e-003	2.48	61.63
9	1.74e-003	2.31	63.94
10	1.63e-003	2.16	66.10



EIGENVECTOR RESEARCH INCORPORATED

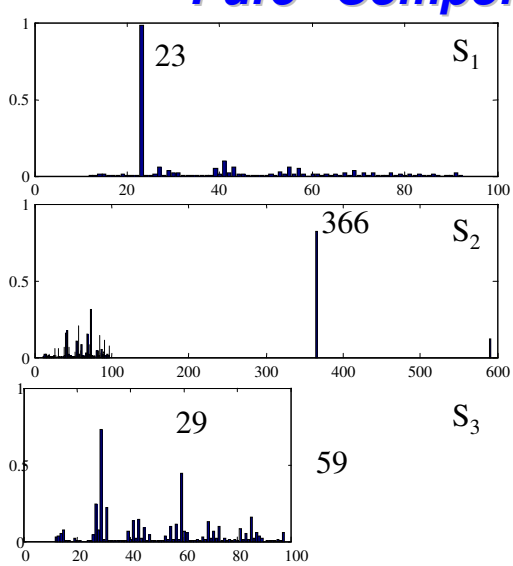
Scores on PC 2 vs. 1

- Extreme samples are a good first guess
- ALS has fewer with $c_{ij}<0$



EIGENVECTOR RESEARCH INCORPORATED

"Pure" Component Spectra

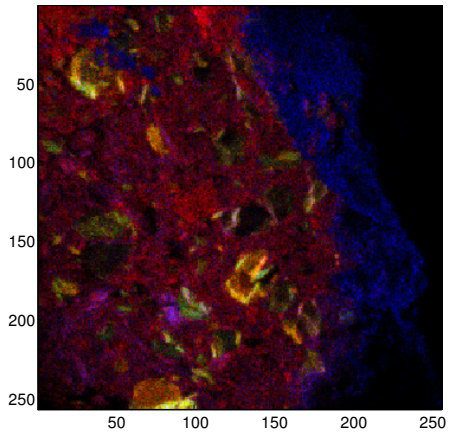


- Make a "true color" image where
  - Red - high at mass 23
  - Green - high at mass 366
  - Blue - high at masses 29 and 59

EIGENVECTOR RESEARCH INCORPORATED

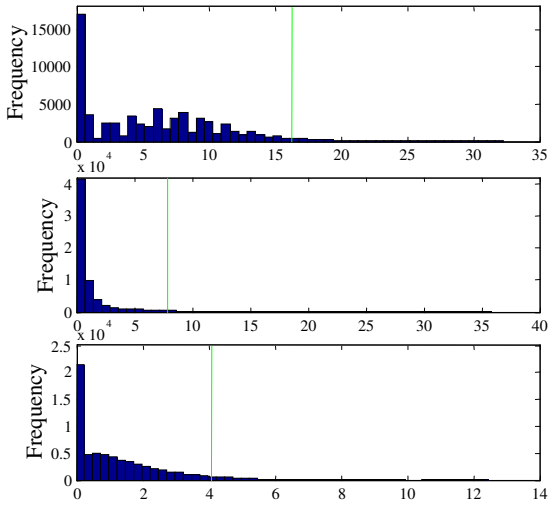
Concentration Image

- Convert **C** to uint8
  - 0 to 255



EIGENVECTOR RESEARCH INCORPORATED

Contrast Enhancement

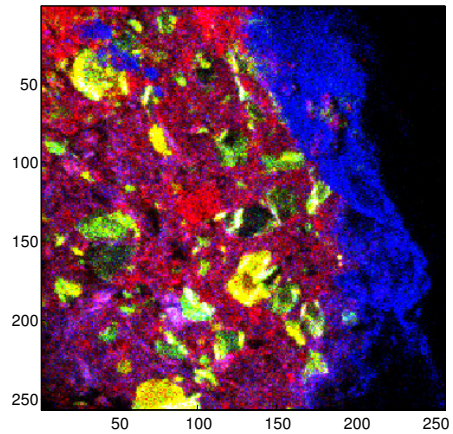


- Only a few channels use up the dynamic range
  - identify high limit and saturate values above
  - can also use a low limit

EIGENVECTOR RESEARCH INCORPORATED

Contrasted Concentration Image

- Convert **C** to uint8
  - "true color" image



EIGENVECTOR RESEARCH INCORPORATED

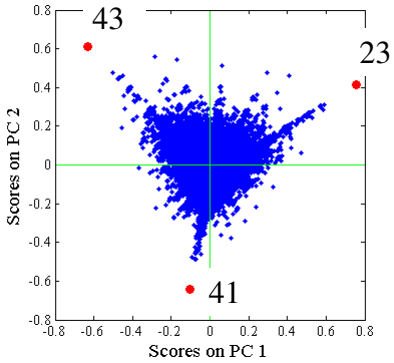
Summary of Procedure

- Normalize rows of $\mathbf{X} \Rightarrow \mathbf{X}_{\text{norm}}$
  - 1 norm easy to use if **C**>0 and **S**>0 (in a single quadrant)
  - use only rows with norm > tolerance (~noise level)
- Mean center the columns of $\mathbf{X}_{\text{norm}} \Rightarrow \mathbf{X}_{\text{norm,mean}}$
- Identify the extreme samples of $\mathbf{X}_{\text{norm,mean}} \Rightarrow i_{sel}$
- Initial guess $\mathbf{S}_0 = \mathbf{X}_{\text{norm}}(i_{sel},:)$ [stop?]
- Use ALS on $\mathbf{X}_{\text{norm}}$ with $\mathbf{S}_0$ as an initial guess
  - include known constraints

EIGENVECTOR RESEARCH INCORPORATED

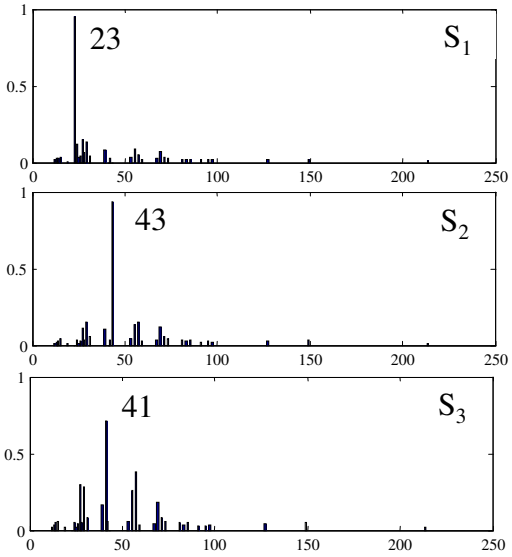
Cloth Example

- Cloth 256x256x34 mass spec
- PCA of 1 norm (for norm>2), mean centered
- The extreme samples represent "single channel spectra"



EIGENVECTOR RESEARCH INCORPORATED

MCR via ALS



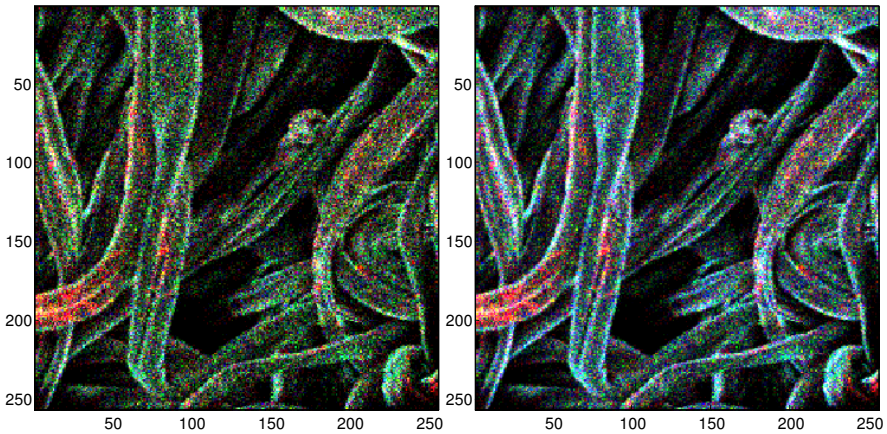
- For both the extreme spectra and ALS results
  - Red - high in 23
  - Green - high in 43
  - Blue - high in 41

EIGENVECTOR RESEARCH INCORPORATED

Contrasted Concentration Images

using extreme spectra

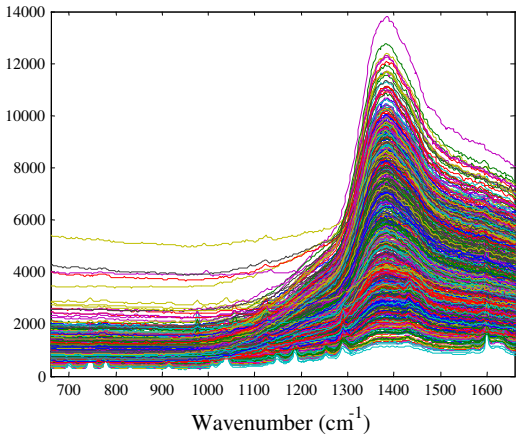
using ALS spectra



EIGENVECTOR RESEARCH INCORPORATED

Aspirin in Polymer

- Raman 22x33x501 (635-1660 cm^{-1})



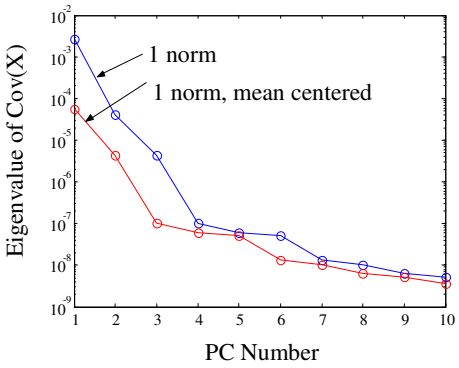
EIGENVECTOR RESEARCH INCORPORATED

Aspirin in Polymer

- PCA, 1 norm, mean centered

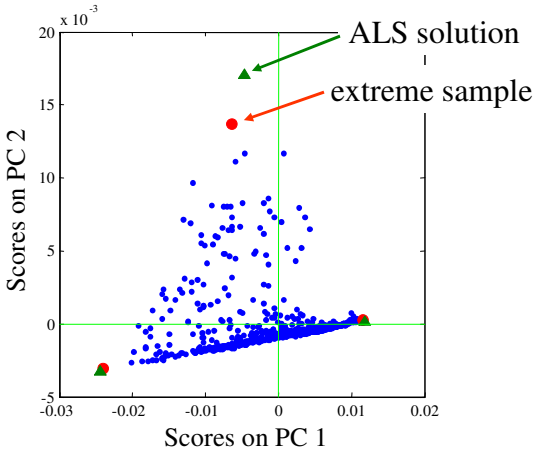
Percent Variance Captured by PCA Model

Principal Component Number	Eigenvalue of Cov(X)	% Variance Captured This PC	% Variance Captured Total
1	5.60e-005	92.20	92.20
2	4.37e-006	7.20	99.40
3	1.04e-007	0.17	99.57
4	5.92e-008	0.10	99.67
5	5.03e-008	0.08	99.75
6	1.35e-008	0.02	99.77
7	1.00e-008	0.02	99.79
8	6.25e-009	0.01	99.80



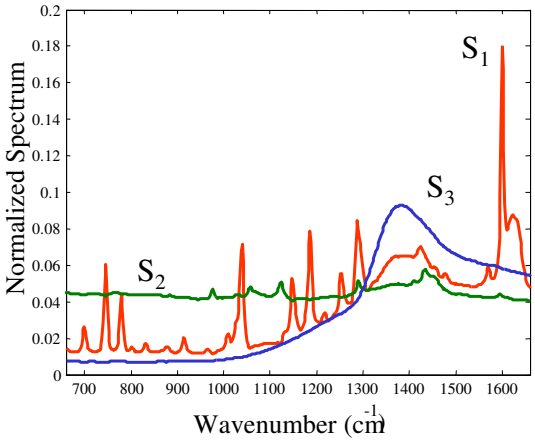
EIGENVECTOR RESEARCH INCORPORATED

Extreme Samples and ALS



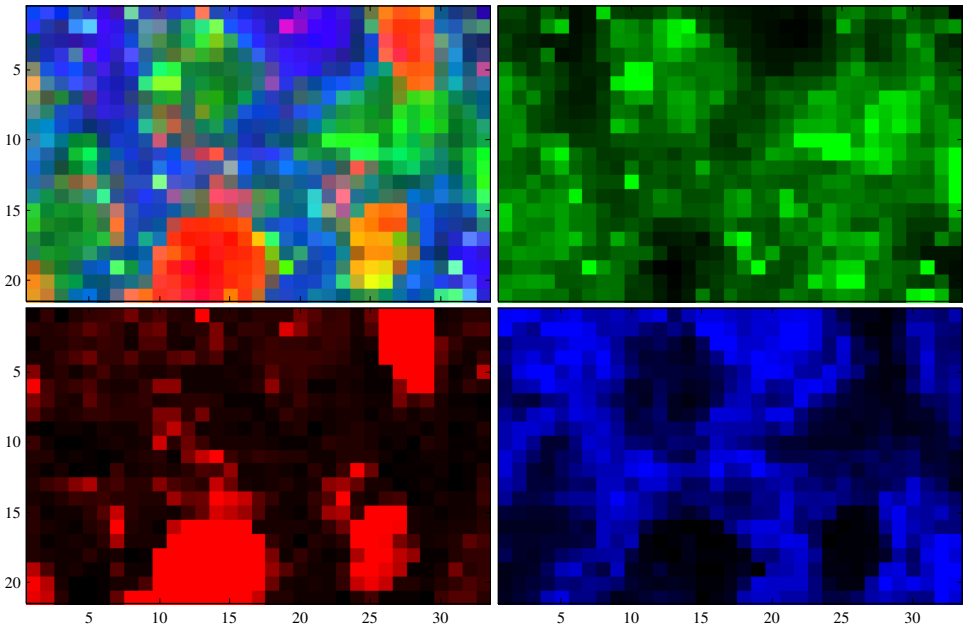
EIGENVECTOR RESEARCH INCORPORATED

Extracted Spectra

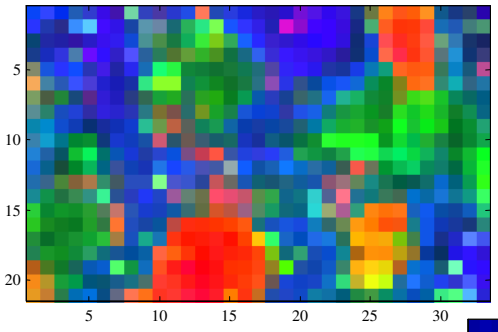


EIGENVECTOR RESEARCH INCORPORATED

Concentration Images



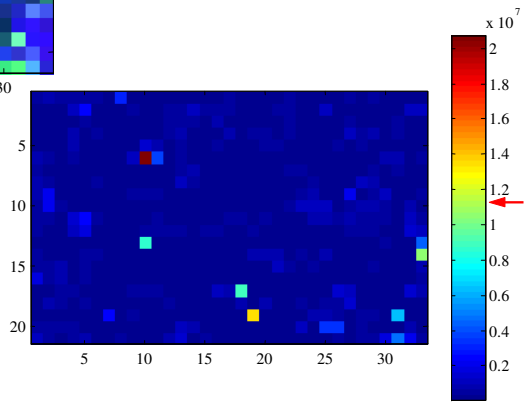
Concentration Images

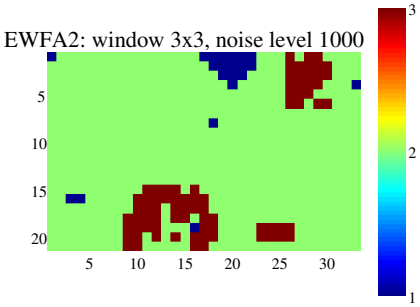


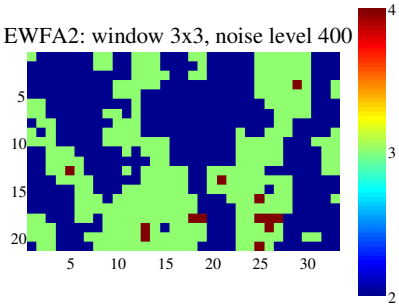
RGB Image of the 3 factors

Sum of squared residuals

arrow indicates ~95% confidence limit line



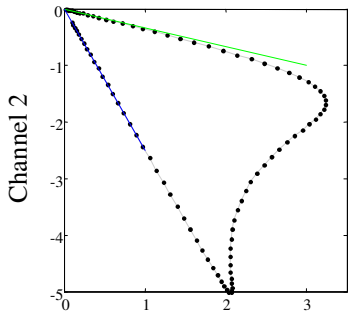




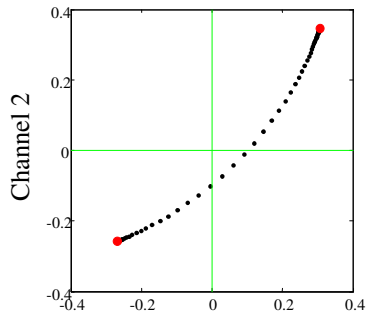
EIGENVECTOR RESEARCH INCORPORATED

2nd Derivative

- Find the extreme samples of 2 norm of 2nd derivative spectra (positive and negative channels)



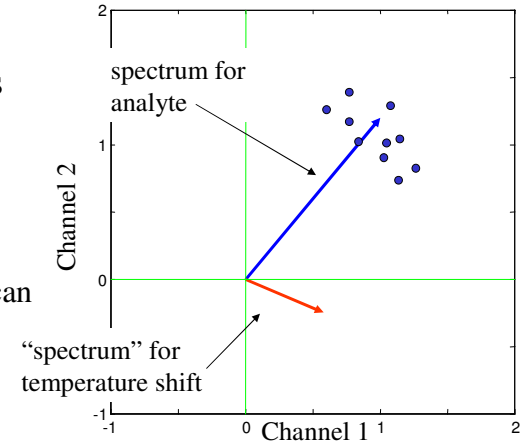
non-negativity on **C** means the data lie between the spectra



EIGENVECTOR RESEARCH INCORPORATED

Temperature Shifts

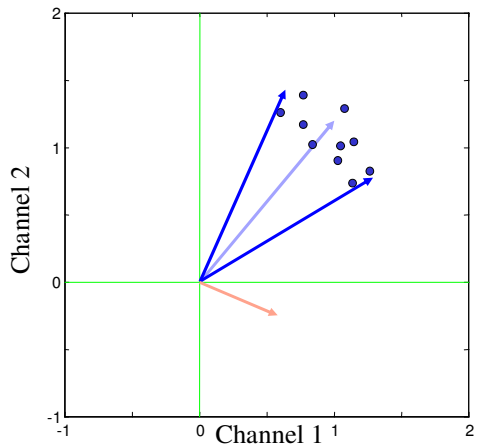
- Temp changes often result in spectral shifts
  - can often be modeled using a low number of "extra factors" or "temperature spectra"
  - but temp contribution can be >0 or <0 ⇒lot's of feasible solutions



EIGENVECTOR RESEARCH INCORPORATED

Temperature Shifts

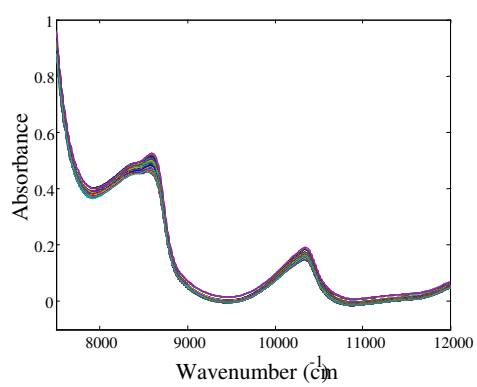
- Non-negativity on **C**
  - rotates the temperature spectrum so a soln can be obtained
  - but also rotates the other spectrum away
    - it is the spectrum at an extreme temperature
    - can constrain if **C** is known



EIGENVECTOR RESEARCH INCORPORATED

Example: 2nd Deriv and T Shifts

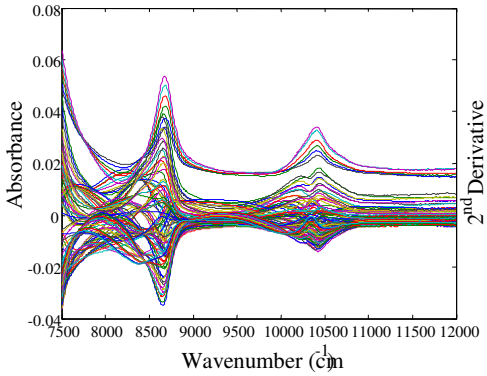
- Caustic data
  - SWIR (7500-12000 cm^{-1})
  - NaCl, NaOH, (wt%) in hot brine with variable T
  - 70 calibration samples
  - 25 test samples



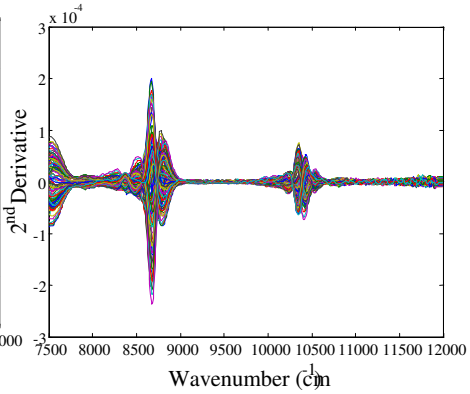
EIGENVECTOR RESEARCH INCORPORATED

Second Derivative

mean centering shows offsets



2nd derivative removes offsets (mean centered)

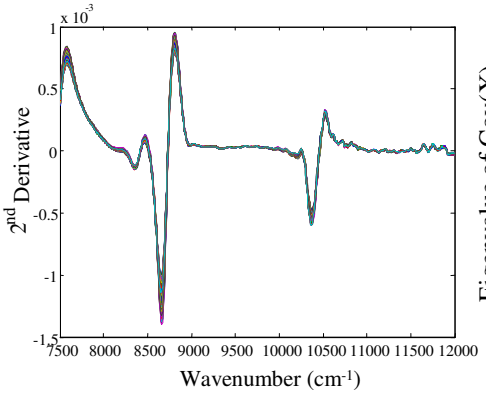


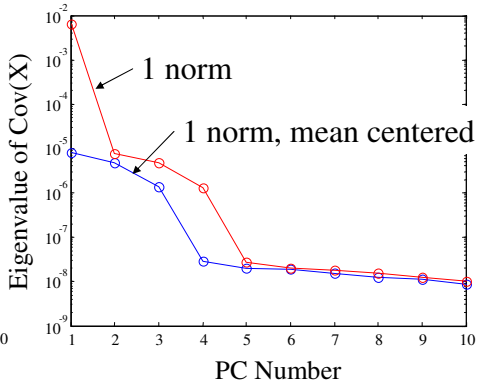
savgol(x,15,2,2)

EIGENVECTOR RESEARCH INCORPORATED

PCA of Normalized 2nd Deriv

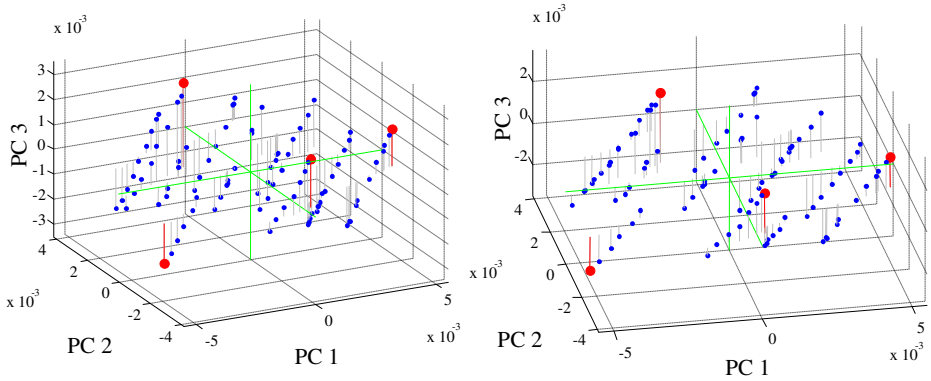
2nd derivative (not normalized)





EIGENVECTOR RESEARCH INCORPORATED

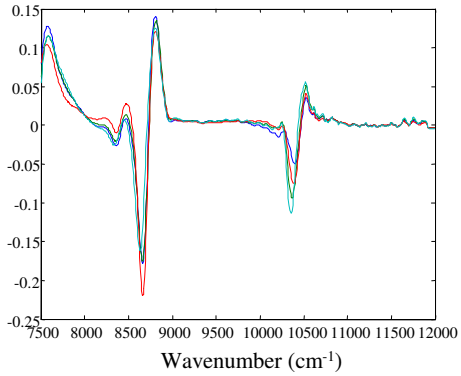
Extreme Samples



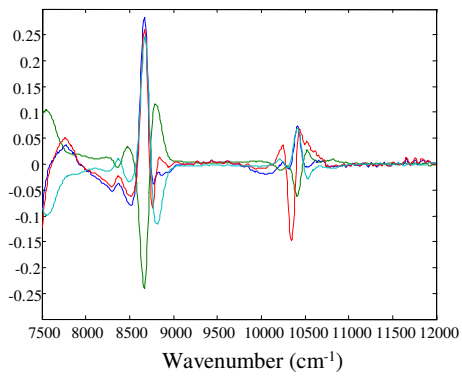
EIGENVECTOR RESEARCH INCORPORATED

Recovered Spectra

S2: only C constrained >0



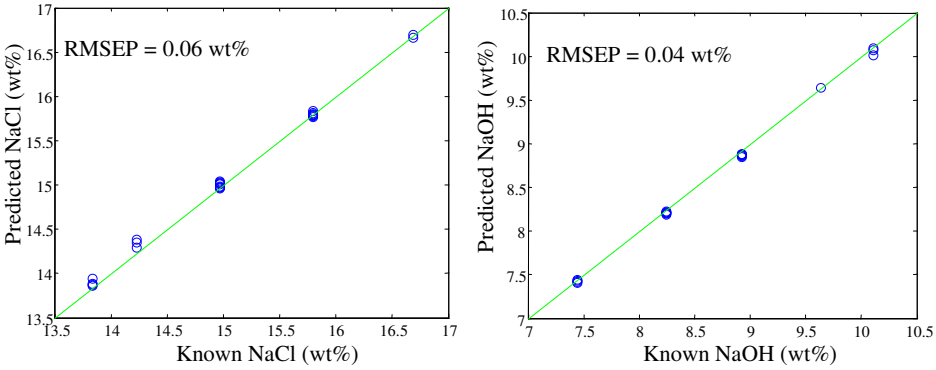
S1: C constrained >0, and equality constraints for known NaCl and NaOH



EIGENVECTOR RESEARCH INCORPORATED

CLS Prediction

S1: used to predict NaCl and NaOH for test samples



EIGENVECTOR RESEARCH INCORPORATED

Curve Resolution for Multivariate Images

- Summary ...

EIGENVECTOR RESEARCH INCORPORATED