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

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 …
**MCR**

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

\[
X = CS^T + E
\]

where

- \( X \) is a \( M \times N \) matrix of measured responses,
- \( C \) is a \( M \times K \) matrix of pure analyte contributions,
- \( S \) is a \( N \times K \) matrix of pure analyte spectra, and
- \( E \) is a \( M \times N \) matrix of residuals.

**MCR Similar to PCA**

- CLS and PCA look similar but with different mathematical properties

\[
X = CS^T + E \quad \text{and} \quad X = TP^T + E
\]

where

- \( X_{M \times N} \), \( C_{M \times K} \), \( c_{ij} \geq 0 \)
- \( S_{N \times K} \), \( s_{ij} \geq 0 \)
- \( E_{M \times N} \)
- \( T_{M \times K} \), \( T^T T_{K \times K} \) diagonal
- \( P_{N \times K} \), \( P^T P = I_{K \times K} \)
- \( E_{M \times N} \)
**ALS**

- Alternating and constrained least squares is often used to get estimates of \( C \) and \( S \) e.g.
  - 0) Start with a guess for \( C = C_0 \) (or \( S_0 \))
  - 1) Estimate \( S_{r}^T = C_{r}^\dagger \cdot X \) subject to \( S(i,j) \geq 0 \)
  - 2) Estimate \( C_r = X S_{r}^{\dagger T} \) subject to \( C(i,j) \geq 0 \)
  - 3) Continue Steps 1 and 2 until \( \|E\| \) within some tolerance or \( r > r_{\text{max}} \)

**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
**Example with Evolving Data**

- Synthetic data from LC-NIR

**Elution Profiles**

- Measured response (no noise)
Measured Response (no noise)
Non-negativity

- Plot $S$ (Channel 2 versus 1)

Elution Example No Noise

- Samples at the boundaries (extremes) are best estimate for $S_0$
How to find the Extremes?

- Normalize each spectrum (which $p$?)

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

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
### 1 and 2 Norm

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

![Graph of 1 and 2 Norm](image)

### What Happens with Noise?

- Estimate extremes using samples with higher signal

![Graph of What Happens with Noise?](image)
**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
Rotational Ambiguity

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

$$\mathbf{X} = \mathbf{CS}^T + \mathbf{E} = (\mathbf{CA})(\mathbf{A}^{-1}\mathbf{S}^T) + \mathbf{E}$$


Multiplicative Ambiguity

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

$$\mathbf{X} = \mathbf{CS}^T + \mathbf{E} = (\mathbf{CA})(\mathbf{A}^{-1}\mathbf{S}^T) + \mathbf{E}$$
**Lack of Selectivity**

- Extremes are inside pure spectra

![Graph showing lack of selectivity](image)


---

**Use What You Know**

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

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

\[
\begin{bmatrix}
12 & 0 & 0 \\
0 & 0 & 5 \\
0 & 1 & 0
\end{bmatrix}
\]

Plot of Normalized Pixels

```matlab
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 = distslec(ymean,3);
y(isel,:)
```

\[
\begin{bmatrix}
12 & 0 & 0 \\
0 & 0 & 5 \\
0 & 1 & 0
\end{bmatrix}
\]
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

PCA of the Image Data

1 norm, mean center for samples with norm>2

<table>
<thead>
<tr>
<th>Principal Component Number</th>
<th>Cov(X)</th>
<th>% Variance This PC</th>
<th>Total % Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.68e-002</td>
<td>35.40</td>
<td>35.40</td>
</tr>
<tr>
<td>2</td>
<td>4.97e-003</td>
<td>6.56</td>
<td>41.97</td>
</tr>
<tr>
<td>3</td>
<td>3.29e-003</td>
<td>4.34</td>
<td>46.31</td>
</tr>
<tr>
<td>4</td>
<td>2.90e-003</td>
<td>3.83</td>
<td>50.14</td>
</tr>
<tr>
<td>5</td>
<td>2.57e-003</td>
<td>3.40</td>
<td>53.54</td>
</tr>
<tr>
<td>6</td>
<td>2.22e-003</td>
<td>2.93</td>
<td>56.47</td>
</tr>
<tr>
<td>7</td>
<td>2.02e-003</td>
<td>2.67</td>
<td>59.15</td>
</tr>
<tr>
<td>8</td>
<td>1.88e-003</td>
<td>2.48</td>
<td>61.63</td>
</tr>
<tr>
<td>9</td>
<td>1.74e-003</td>
<td>2.31</td>
<td>63.94</td>
</tr>
<tr>
<td>10</td>
<td>1.63e-003</td>
<td>2.16</td>
<td>66.10</td>
</tr>
</tbody>
</table>
Scores on PC 2 vs. 1

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

“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
Concentration Image

- Convert C to uint8
  - 0 to 255

Contrast Enhancement

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

- Convert C to uint8
  - “true color” image

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
**Cloth Example**

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

**MCR via ALS**

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

- Using extreme spectra
- Using ALS spectra

**Aspirin in Polymer**

- Raman 22x33x501 (635-1660 cm\(^{-1}\))
Aspirin in Polymer

- PCA, 1 norm, mean centered

### Percent Variance Captured by PCA Model

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

### Extreme Samples and ALS

ALS solution

extreme sample

Eigenvector Research Incorporated
Extracted Spectra

Concentration Images
**Concentration Images**

RGB Image of the 3 factors

**Sum of squared residuals**
arrow indicates ~95% confidence limit line

EWFA2: window 3x3, noise level 1000

EWFA2: window 3x3, noise level 400

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

**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 $<0 \Rightarrow$ lot’s of feasible solutions
Temperature Shifts

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

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
Second Derivative

mean centering shows offsets

2nd derivative removes offsets (mean centered)

savgol(x,15,2,2)

PCA of Normalized 2nd Deriv

2nd derivative (not normalized)

1 norm

1 norm, mean centered
**Extreme Samples**

**Recovered Spectra**

S2: only C constrained >0

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

S1: used to predict NaCl and NaOH for test samples

![Graphs showing predicted vs. known concentrations of NaCl and NaOH with RMSEP values of 0.06 wt% and 0.04 wt% respectively.]

**Curve Resolution for Multivariate Images**

- Summary ...