Retour sur... la ligne de base
BEADS: correction et filtrage conjoints de
mesures analytiques exploitant positivité et parcimonie

X. NING, I. W. SELESNICK
Polytechnic School of Engineering, New York University
L. DUVAL, A. PIRAYRE
IFP Energies nouvelles, Université Paris-Est

9 octobre 2017
Old peaks cast long shadows
Old peaks cast long shadows
Old peaks cast long shadows
Old peaks cast long shadows
Old peaks cast long shadows
The quick version

- **Issue**: how to accurately & repeatably quantize peaks?
  - avoiding separate baseline and noise removal
- **Question**: where is the string behind the bead?
  - without precise models for: peak, noise, baseline

- **Answer**: use main measurement properties + optimization
  - sparsity+symmetry, stationarity, smoothness
- **BEADS**: Baseline Estimation And Denoising w/ Sparsity
Outline

INTRODUCTION
   FOREWORD
   OUTLINE*
   BACKGROUND

BEADS model and algorithm
   NOTATIONS
   Compound sparse derivative modeling
   Majorize-Minimize type optimization

Evaluation and results
   GC: simulated baseline and Gaussian noise
   GC: simulated Poisson noise
   GC: real data
   GC×GC: real data
   Others

Conclusions
Background on background

Figure: Image processing: varying illumination

- Background affects quantitative evaluation/comparison
- In other domains: (instrumental) bias, (seasonal) trend
- In analytical chemistry: drift, continuum, wander, baseline
- Very rare cases of parametric modeling (piecewise linear, polynomial, spline)
Background on background

Figure: Econometrics: trends and seasonality

- Background affects quantitative evaluation/comparison
- In other domains: (instrumental) bias, (seasonal) trend
- In analytical chemistry: drift, continuum, wander, baseline
- Very rare cases of parametric modeling (piecewise linear, polynomial, spline)
Background on background

Figure: Biomedical: ECG isoelectric line or baseline wander

- Background affects quantitative evaluation/comparison
- In other domains: (instrumental) bias, (seasonal) trend
- In analytical chemistry: drift, continuum, wander, baseline
- Very rare cases of parametric modeling (piecewise linear, polynomial, spline)
Background on background

Figure: Gas chromatography: baseline

- Background affects quantitative evaluation/comparison
- In other domains: (instrumental) bias, (seasonal) trend
- In analytical chemistry: drift, continuum, wander, baseline
- Very rare cases of parametric modeling (piecewise linear, polynomial, spline)
Background on background

Analytical chemistry, biological data

- Signal separation into three main morphological components
Notations and assumptions

Morphological decomposition: $y = x + f + w$, signals in $\mathbb{R}^N$

- $y$: observation (spectrum, analytical data)
- $x$: clean series of peaks (no baseline, no noise)
- $f$: baseline
- $w$: noise

Assumption: without peaks, the baseline can be (approx.) recovered from noise-corrupted data by low-pass filtering

- $\hat{f} = L(y - \hat{x})$: $L$: low-pass filter; $H = I - L$: high-pass filter
- formulated as $\|y - \hat{x} - \hat{f}\|_2^2 = \|H(y - \hat{x})\|_2^2$
- Going further with $D_i$: differentiation operators
Compound sparse derivative modeling

An estimate $\hat{x}$ can be obtained via:

$$\hat{x} = \arg \min_x \left\{ F(x) = \frac{1}{2} \| H(y - x) \|_2^2 + \sum_{i=0}^{M} \lambda_i R_i(D_i x) \right\}.$$
Compound sparse derivative modeling

Examples of (smooth) sparsity promoting functions for $R_i$

- $\phi_i^A = |x|$
- $\phi_i^B = \sqrt{|x|^2 + \epsilon}$
- $\phi_i^C = |x| - \epsilon \log (|x| + \epsilon)$
Compound sparse derivative modeling

Take the positivity of chromatogram peaks into account:

\[
\hat{x} = \arg \min_x \left\{ F(x) = \frac{1}{2} \| H(y - x) \|_2^2 + \lambda_0 \sum_{n=0}^{N-1} \theta_\epsilon(x_n; r) + \sum_{i=1}^{M} \lambda_i \sum_{n=0}^{N_i-1} \phi ([D_i x]_n) \right\}.
\]

Start from:

\[
\theta(x; r) = \begin{cases} 
  x, & x \geq 0 \\
  -rx, & x < 0 
\end{cases}
\]
Compound sparse derivative modeling

Take the positivity of chromatogram peaks into account:

\[
\hat{x} = \arg\min_x \left\{ F(x) = \frac{1}{2} \|H(y - x)\|_2^2 \right. \\
\left. + \lambda_0 \sum_{n=0}^{N-1} \theta_\epsilon(x_n; r) + \sum_{i=1}^M \sum_{n=0}^{N_i-1} \phi([D_i x]_n) \right\}.
\]

and majorize it

![Diagram showing the majorizer function](image-url)
Compound sparse derivative modeling

Take the positivity of chromatogram peaks into account:

\[ \hat{x} = \arg \min_x \left\{ F(x) = \frac{1}{2} \| H(y - x) \|_2^2 + \lambda_0 \sum_{n=0}^{N-1} \theta_{\epsilon}(x_n; r) + \sum_{i=1}^{M} \lambda_i \sum_{n=0}^{N_i-1} \phi([D_i x]_n) \right\}. \]

then smooth it:
Compound sparse derivative modeling

Take the positivity of chromatogram peaks into account:

\[ \hat{x} = \arg \min_x \left\{ F(x) = \frac{1}{2} \| H(y - x) \|_2^2 \right\} \]
\[ + \lambda_0 \sum_{n=0}^{N-1} \theta_{\varepsilon}(x_n; r) + \sum_{i=1}^{M} \lambda_i \sum_{n=0}^{N_i-1} \phi ([D_i x]_n) \}. \]

then majorize it:

\[ g_0(x, v) = \begin{cases} \frac{1+r}{4|v|} x^2 + \frac{1-r}{2} x + |v| \frac{1+r}{4}, & |v| > \varepsilon \\ \frac{1+r}{4\varepsilon} x^2 + \frac{1-r}{2} x + \varepsilon \frac{1+r}{4}, & |v| \leq \varepsilon. \end{cases} \]
Overall principle for
Majoration-Minimization-Maximization

Figure: Courtesy Peng Wang

HTTPS://COMMONS.WIKIMEDIA.ORG/W/INDEX.PHP?CURID=17689902
BEADS Algorithm (short)

Input: \( y, A, B, \lambda_i, i = 0, \ldots, M \)

1. \( b = B^TBA^{-1}y \)
2. \( x = y \) (Initialization)
   Repeat

3. \( [\Lambda_i]_{n,n} = \frac{\phi'(D_ix_{n})}{D_ix_{n}}, \quad i = 0, \ldots, M, \)

4. \( M = \sum_{i=0}^{M} \lambda_i D_i^T \Lambda_i D_i \)

5. \( Q = B^TB + A^TMA \)

6. \( x = AQ^{-1}b \)
   Until converged

8. \( f = y - x - BA^{-1}(y - x) \)

Output: \( x, f \)
Evaluation 1

Figure: Simulated chromatograms w/ polynomial+sine baseline
Evaluation 1 with Gaussian noise

<table>
<thead>
<tr>
<th></th>
<th>0 dB</th>
<th>10 dB</th>
<th>20 dB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
</tr>
<tr>
<td>BEADS</td>
<td>28.1</td>
<td>8.52</td>
<td>32.64</td>
</tr>
<tr>
<td>backcor</td>
<td>24.91</td>
<td>9.75</td>
<td>31.27</td>
</tr>
<tr>
<td>airPLS</td>
<td>20.26</td>
<td>9.65</td>
<td>22.54</td>
</tr>
</tbody>
</table>
Evaluation 2

Figure: Simulated chromatograms w/ limited power spectrum noise
Evaluation 2 with Gaussian noise

Experiment 2

<table>
<thead>
<tr>
<th></th>
<th>0 dB</th>
<th></th>
<th>10 dB</th>
<th></th>
<th>20 dB</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
<td>Std</td>
</tr>
<tr>
<td>BEADS</td>
<td>18.75</td>
<td>3.71</td>
<td>19.99</td>
<td>3.17</td>
<td>20.89</td>
<td>3.32</td>
</tr>
<tr>
<td>backcor</td>
<td>17.20</td>
<td>4.57</td>
<td>18.93</td>
<td>3.74</td>
<td>19.54</td>
<td>3.18</td>
</tr>
<tr>
<td>airPLS</td>
<td>16.71</td>
<td>4.80</td>
<td>17.52</td>
<td>5.54</td>
<td>17.98</td>
<td>4.82</td>
</tr>
</tbody>
</table>
Evaluation 3 with Poisson noise

Figure: Simulated chromatograms w/ Poisson noise
Results: mono-dimensional chromatography (data 1)

Figure: Original, superimposed, clean, noise
Results: two-dimensional chromatography (data 2)

Figure: Original data
Results: two-dimensional chromatography (data 2)

Figure: 2D background (estimated)
Results: two-dimensional chromatography (data 2)

Figure: Noise (estimated)
Results: two-dimensional chromatography (data 2)

Figure: BEADS corrected data
Results: two-dimensional chromatography (data 2)

Figure: Original data (again!)
Results: two-dimensional chromatography (data 3)

Figure: Original data
Results: two-dimensional chromatography (data 3)

Figure: 2D background (estimated)
Results: two-dimensional chromatography (data 3)

Figure: Noise (estimated)
Results: two-dimensional chromatography (data 3)

Figure: BEADS corrected data
Results: two-dimensional chromatography (data 3)

Figure: Original data (again!)
Results: performance

Figure: Linear cost per sample (almost)
Other known uses

- A fairly generic model (sparsity, positivity/negativity)
  - gas chromatography: mono-dimensional and comprehensive/two-dimensional
  - Raman spectra: biological and biomedical
  - MUSE (Multi Unit Spectroscopic Explorer): astronomical hyperspectral galaxy spectrum
  - X-ray absorption spectroscopy (XAS), X-ray diffraction (XRD), and combined XAS/XRD
  - high-resolution mass spectrometry
  - postprandial Plasma Glucose (PPG), multichannel electroencephalogram (EEG) and single-channel electrocardiogram (ECG)
  - arabic characters
Other known uses

- A fairly generic model (sparsity, positivity/negativity)
  - gas chromatography: mono-dimensional and comprehensive/two-dimensional
  - Raman spectra: biological and biomedical
  - MUSE (Multi Unit Spectroscopic Explorer): astronomical hyperspectral galaxy spectrum
  - X-ray absorption spectroscopy (XAS), X-ray diffraction (XRD), and combined XAS/XRD
  - high-resolution mass spectrometry
  - postprandial Plasma Glucose (PPG), multichannel electroencephalogram (EEG) and single-channel electrocardiogram (ECG)
  - arabic characters
Conclusions

- Joint Baseline Estimation and Denoising
  - Little "hard" modeling
  - Codes available in Matlab\(^2\) and R\(^3\)

![MathWorks File Exchange](http://lc.cx/beads)

- Interaction between "separative science" and "source separation"

\(^2\) http://lc.cx/beads
\(^3\) http://www.laurent-duval.eu/lcd-publications.html#beads-r-code
Work in progress

- Ongoing tests on analytical chemistry data: NIR, NMR, MS
- Better documentation and usability
- Estimated baseline and noise use?
- Novel metrics: errors related to peak quantities
- Novel filtering: an update on Savitzky-Golay filters
- Novel deconvolution: sparse & positive with norm ratios
Characterisation of middle-distillates by comprehensive two-dimensional gas chromatography (GC × GC):
A powerful alternative for performing various standard analysis of middle-distillates.

C. Vendeuvre, R. Ruiz-Guerrero, F. Bertoncini, L. Duval, and D. Thiébaut.
Comprehensive two-dimensional gas chromatography for detailed characterisation of petroleum products.

X. Ning, I. W. Selesnick, and L. Duval.
Chromatogram baseline estimation and denoising using sparsity (BEADS).

A. Repetti, M. Q. Pham, L. Duval, E. Chouzenoux, and J.-C. Pesquet.
Euclid in a taxicab: Sparse blind deconvolution with smoothed $\ell_1/\ell_2$ regularization.

BARCHAN: Blob Alignment for Robust CHromatographic ANalysis.

Peaks, baseline and noise separation.
Chapter in *Source Separation in Physical-Chemical Sensing*, 2018.
BEADS Algorithm

We now have a majorizer for $F$

$$G(x, v) = \frac{1}{2} \|H(y - x)\|_2^2 + \lambda_0 x^T [\Gamma(v)] x$$

$$+ \lambda_0 b^T x + \sum_{i=1}^{M} \left[ \frac{\lambda_i}{2} (D_i x)^T [\Lambda(D_i v)] (D_i x) \right] + c(v).$$

Minimizing $G(x, v)$ with respect to $x$ yields

$$x = \left[ H^T H + 2\lambda_0 \Gamma(v) + \sum_{i=1}^{M} \lambda_i D_i^T [\Lambda(D_i v)] D_i \right]^{-1} \left( H^T H y - \lambda_0 b \right).$$

with notations

$$c(v) = \sum_n \left[ \phi(v_n) - \frac{v_n}{2} \phi'(v_n) \right].$$
BEADS Algorithm

We now have a majorizer for $F$

$$G(x, v) = \frac{1}{2} \|H(y - x)\|^2_2 + \lambda_0 x^T [\Gamma(v)] x$$

$$+ \lambda_0 b^T x + \sum_{i=1}^{M} \left[ \frac{\lambda_i}{2} (D_i x)^T [\Lambda(D_i v)] (D_i x) \right] + c(v).$$

Minimizing $G(x, v)$ with respect to $x$ yields

$$x = \left[ H^T H + 2\lambda_0 \Gamma(v) + \sum_{i=1}^{M} \lambda_i D_i^T [\Lambda(D_i v)] D_i \right]^{-1} \left( H^T H y - \lambda_0 b \right).$$

with notations

$$[\Gamma(v)]_{n,n} = \begin{cases} \frac{1+r}{4|v_n|}, & |v_n| \geq \epsilon \\ \frac{1+r}{4\epsilon}, & |v_n| \leq \epsilon \end{cases}$$
**BEADS Algorithm**

We now have a majorizer for \( F \)

\[
G(x, v) = \frac{1}{2} \|H(y - x)\|^2_2 + \lambda_0 x^T [\Gamma(v)] x + \lambda_0 b^T x + \sum_{i=1}^{M} \frac{\lambda_i}{2} (D_i x)^T [\Lambda(D_i v)] (D_i x) + c(v).
\]

Minimizing \( G(x, v) \) with respect to \( x \) yields

\[
x = \left[ H^T H + 2\lambda_0 \Gamma(v) + \sum_{i=1}^{M} \lambda_i D_i^T [\Lambda(D_i v)] D_i \right]^{-1} \left( H^T H y - \lambda_0 b \right).
\]

with notations

\[
[\Lambda(v)]_{n,n} = \frac{\phi'(v_n)}{v_n}
\]
**BEADS Algorithm**

We now have a majorizer for $F$

\[
G(x, v) = \frac{1}{2} \| H(y - x) \|^2 + \lambda_0 x^T [\Gamma(v)] x \\
+ \lambda_0 b^T x + \sum_{i=1}^{M} \left[ \frac{\lambda_i}{2} (D_i x)^T [\Lambda(D_i v)] D_i x \right] + c(v).
\]

Minimizing $G(x, v)$ with respect to $x$ yields

\[
x = \left[ H^T H + 2\lambda_0 \Gamma(v) + \sum_{i=1}^{M} \lambda_i D_i^T [\Lambda(D_i v)] D_i \right]^{-1} \left( H^T H y - \lambda_0 b \right).
\]

with notations

\[
[b]_n = \frac{1 - r}{2}
\]
BEADS Algorithm

Writing filter $H = A^{-1}B \approx BA^{-1}$ (banded matrices) we have

$$x = AQ^{-1} \left( B^TBA^{-1}y - \lambda_0 A^Tb \right)$$

where $Q$ is the banded matrix,

$$Q = B^TB + A^TMA,$$

and $M$ is the banded matrix,

$$M = 2\lambda_0 \Gamma(v) + \sum_{i=1}^{M} \lambda_i D_i^T [\Lambda(D_i v)] D_i.$$
BEADS Algorithm

Using previous equations, the MM iteration takes the form:

\[
M^{(k)} = 2\lambda_0 \Gamma(x^{(k)}) + \sum_{i=1}^{M} \lambda_i D_i^T \Lambda(D_i x^{(k)}) D_i.
\]

\[
Q^{(k)} = B^T B + A^T M^{(k)} A
\]

\[
x^{(k+1)} = A[Q^{(k)}]^{-1} \left( B^T B A^{-1} y - \lambda_0 A^T b \right)
\]