Empirical Bayes for data fusion at the National Ignition Facility (NIF)

CASIS 2017
May 17, 2017

Florin Munteanu
National Ignition Facility

World’s most energetic laser (192 laser beams, a total of 1.8 MJ Energy)

A large amount of radiation is produced and recorded.

Data analysis offers insight into the state of the matter.

Data is corrupted by noise and instrumentation mis-calibration.

Inferring the true original signal is key to correct physical interpretation.
Neutron Time of Flight (NTOF) signal

The signal is recorded by multiple scopes at different attenuations.

- Goal: **Retrieve the original signal!**
- The fiducial ensures good registration.
- Attenuator values measured with errors.
NTOF simple stitch

For each portion select the trace with best signal to noise ratio (SNR).

Main issues:
1. Many available points from lower SNR traces not used -> noise.
2. The Attenuator mis-calibrations are evident in the gaps/overlaps between different traces.
Alternative: NTOF weighted stitch

\[ \hat{s} = \sum_{i=1}^{M} w_i y_i = \sum_{i=1}^{M} w_i (A_i s + n_i) \]

\[ E[\hat{s}] = s \rightarrow \sum_{i=1}^{M} w_i A_i = 1 \]

\[ Var[\hat{s}] \rightarrow \text{min} \]

\[ w_i = \frac{A_i}{\sigma_i^2} \sum_{k=1}^{p} \left( \frac{A_k}{\sigma_k} \right)^2 \]

Makes use of all available data
Still does not solve the attenuator calibration problem.
Proposed: Data fusion

Signal at time “k” from the scope “i” (N data, M signals)

\[ y_i(k) = A_i s(k) + n_i(k) \]

\[ n_i(k) = N(0, \sigma_i^2) \]

\[ y_i \equiv [y_i(1), y_i(2), \ldots, y_i(N)]^T \]

\[ n_i \equiv [n_i(1), n_i(2), \ldots, n_i(N)]^T \]

\[
\begin{bmatrix}
  y_1 \\
  y_2 \\
  \vdots \\
  y_M \\
\end{bmatrix}_{NM \times 1} =
\begin{bmatrix}
  A_1 I_N \\
  A_2 I_N \\
  \ldots \\
  A_M I_N \\
\end{bmatrix}_{NM \times N} s_N +
\begin{bmatrix}
  n_1 \\
  n_2 \\
  \vdots \\
  n_M \\
\end{bmatrix}_{NM \times 1}
\]

Known values:
- measured traces - \( y_i \)
- measurement noise - \( \sigma_i \)
- initial \( A_i^0 \) estimates

Variables to determine:
- true signal - \( s(k) \)
- true attenuator values - \( A_i \)
Signal distribution

Bayes theorem

\[ P(s|y_1, \ldots, y_M, \sigma_1, \ldots, \sigma_M, A_1, \ldots, A_M) \propto P(y_1, \ldots, y_M|s, \sigma_1, \ldots, \sigma_M, A_1, \ldots, A_M)P(s) \prod_{i=1}^{M} P(A_i) \prod_{i=1}^{M} P(\sigma_i) \]

\[
\begin{align*}
P(y_1, \ldots, y_M|s, \sigma_1, \ldots, \sigma_M, A_1, \ldots, A_M) &= \frac{1}{(2\pi)^{NM/2} \prod_{i=1}^{M} \sigma_i^{N_i/2}} e^{-\frac{1}{2} \sum_{i=1}^{M} \frac{(y_i - A_i s)^T (y_i - A_i s)}{2\sigma_i^2}} \\
A_i^0 \text{ measured attens} \\
\sigma_A \text{ unknown!} \\
\text{Separate estimation} \\
\text{compounds the errors in } A_i \\

P(s) &\propto e^{-\frac{1}{2\lambda^2} \sum_{k=1}^{N} [s(k) - s^T h_k]^2} \\
\hat{s}, \hat{A}_i, \hat{\sigma}_A &= \arg\min_{s, A_i, \sigma_A} \left[ -\log(P(s|y_i, \sigma_i, A_i, \sigma_A)) \right] \\
\end{align*}
\]

\textbf{‘}h\textbf{’} suitable filter (Savitzky–Golay, for example)
Empirical Bayes for $\sigma_A$ determination

Calculate the marginal distribution over $A_i$

$$m(s|y_i, \sigma_i, \sigma_A) = \prod_{i=1}^{M} \int P(s|y_i, \sigma_i, A_i, \sigma_A) dA_i$$

$$P(s|y_i, \sigma_i, A_i, \sigma_A) \propto \frac{1}{(2\pi \sigma_A^2)^{M/2}} \prod_{i=1}^{M} e^{-\left[ \sum_{k=1}^{N} \frac{[y_i(k) - A_i s(k)]^2}{2\sigma_i^2} + \frac{(A_i - A_i^0)^2}{2\sigma_A^2} \right]}$$

Minimize with respect $\sigma_A$

$$-\log [m(s|y_i, \sigma_i, \sigma_A)] = M \log (\sigma_A) + \sum_{i=1}^{M} \frac{A_i^0}{2\sigma_A^2} - \sum_{i=1}^{M} \left[ \sum_{k=1}^{N} \frac{s(k) y_i(k)}{2\sigma_i^2} + \frac{A_i^0}{2\sigma_A^2} \right]^2$$
Iterative algorithm

1. Start with the best weighted stitched signal

\[ \hat{s} = \arg\min_s \left[ -\log(P(s|y_i, \sigma_i, A_i)) \right] \]

2. Estimate the unknown signal ‘s’

3. Estimate the value of \( \sigma_A \) by minimizing the marginal

\[ \hat{\sigma}_A = \arg\min_{\sigma_A} \left[ -\log(m(s|y_i, \sigma_i, A_i, \sigma_A)) \right] \]

4. Estimate the attenuations

\[ \hat{A}_i = \arg\min_{A_i} \left[ -\log(P(A_i|s, y_i, \sigma_i, \sigma_A)) \right] \]

\[ A_i = \frac{\sum_{k=1}^{N} \frac{s(k)y_i(k)}{\sigma_i^2} + \frac{A^0_i}{\sigma_A^2}}{\sum_{k=1}^{N} \frac{s(k)^2}{\sigma_i^2} + \frac{1}{\sigma_A^2}} \quad \rightarrow \quad A^0_i \quad \sigma_A \rightarrow 0 \]

5. Repeat 2-4
Results – experimental data

Converged after 12 iterations

Finding the correct attenuator values removed the gaps in data

<table>
<thead>
<tr>
<th>Attens measured</th>
<th>Attens estimated</th>
<th>Change (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2529</td>
<td>0.2589</td>
<td>2.32</td>
</tr>
<tr>
<td>0.1262</td>
<td>0.1233</td>
<td>-2.35</td>
</tr>
<tr>
<td>0.0632</td>
<td>0.0643</td>
<td>1.71</td>
</tr>
<tr>
<td>0.0316</td>
<td>0.0322</td>
<td>1.86</td>
</tr>
</tbody>
</table>

$\sigma_A = 0.003$
Simple stitch is badly corrupted when attenuator mis-calibrations are present (red trace lower than its true value!).

Weighted stitch reduces the random noise but is still affected by attenuator mis-calibration.
Result – synthetic data

Converged after 5 iterations

<table>
<thead>
<tr>
<th>Attens true</th>
<th>Attens measured</th>
<th>Attens determined</th>
<th>Error (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9958</td>
<td>1</td>
<td>1.0124</td>
<td>-1.67</td>
</tr>
<tr>
<td>0.7034</td>
<td>0.7</td>
<td>0.7151</td>
<td>-1.66</td>
</tr>
<tr>
<td>0.5021</td>
<td>0.5</td>
<td>0.5103</td>
<td>-1.63</td>
</tr>
<tr>
<td>0.1738</td>
<td>0.2</td>
<td>0.1767</td>
<td>-1.67</td>
</tr>
</tbody>
</table>

$\sigma_A$ true  \hspace{2cm} $\sigma_A$ determined

0.01 \hspace{2cm} 0.012

Finding the correct attenuator values removed the gaps/overlap in data
Summary

- Hardware mis-calibration and instrumentation noise affects negatively the quality of the original signal estimation.

- Simultaneous estimation of the signal and hardware calibration values is necessary.

- Empirical Bayes approach offers a valid alternative for solving a difficult optimization problem.