Bayesian Blind Deconvolution

Kristian Zarb Adami

"the only relevant thing is uncertainty—the extent of our knowledge and ignorance. The actual fact of whether the events considered are in some sense determined, or known by other people is of no consequence"
Measurement

- Measurement is made with the purpose to increase the knowledge of the person performing it.
- Measurement constitutes an interaction between a detector and the environment.
- Measurement involves uncertainty in many shapes and sizes.
- General form given by \( D = R(\theta) + \nu \).
- Inference is the extraction of the desired parameters from the data.
The Beam

- The **beam function** is often complicated

- This makes **direct inversion** impossible

- We require an **approximate** method to form a suitable **pseudo-inverse**

- This process is **ill-determined** hence we require a systematic way of choosing the best ‘**pseudo-inverse**’
Deconvolution

- Requires the inverse of the beam matrix
- Is an example of an ill-posed problem
- Does not have a unique solution
- Requires a probabilistic solution

Source

* 

Beam

* =

RED
ERING
HIER
INSIVE

* =

RED
ERING
HIER
INSIVE
Inference for Inversion

- Need to select the best of all the ‘pseudo-inverses’

- Requires an amount of ‘Model Selection’

- ‘Best’ is always a trade-off between a ‘Cost’ function and a ‘Regularising’ function

- ‘Cost’ function ($\chi^2$) and ‘Regularising’ function ($-p_i \log p_i$)

- Bayes’ theorem provides a consistent framework for model selection
  \[
  Posterior = \frac{Likelihood \times Prior}{Evidence}
  \]
Bayesian Inference

- ‘Likelihood’ function determines how the ‘effect’ arises from the ‘cause’ - this contains all the parameters we are seeking

- ‘Prior’ reflects the range in which your experiment is designed to work and any assumptions made

- ‘Posterior’ quantifies belief in the parameters extracted from the measurement
Evidence

- Is the **likelihood** of the specified hypothesis
- Provides a **consistent framework** to compare hypotheses
- Involves **integrals** over hypothesis/parameter space
- Requires ‘clever’ **approximations** in order to evaluate it
- The **three** most widely used techniques are Laplacian, Markov Chain Monte Carlo and Variational Methods

“As far as the laws of maths refer to reality they are not certain, and as far as they are certain they do not refer to reality”
Approximate Inference

- Asymptotic Laplace approximation useful for large amounts of data
- Monte Carlo methods suitable for low dimensional models for $D < 10$
- Markov Chain Monte Carlo methods suitable for high dimensional models
- Variational methods are also suitable for high dimensional problems
Variational approximation

- Find a suitable analytical approximation to the posterior
- Impose a tight bound on the evidence value
- Optimise the bound
- This approximation is more compact than the true distribution.
Variational Approach

- Minimise the divergence between the true and approximate posterior

\[
D_{KL}(Q\|P) = \int_{\Theta} Q(\Theta) \log \left[ \frac{Q(\Theta)}{P(\Theta|D,H)} \right] d\Theta
\]

- This is equivalent to maximising a bound on the evidence value

\[
C_{KL}(Q\|P) = D_{KL}(Q\|P) - \log P(D|H)
= \int_{\Theta} Q(\Theta) \log \frac{Q(\Theta)}{P(D|\Theta,H)P(\Theta|H)} d\Theta
\geq -\log P(D|H)
\]

- This approach provides flexibility in specifying the prior distributions and a deterministic way to obtain a bound on the evidence value

- It is sensitive to probability mass rather than density
Interpolation as an example

- Consider **interpolating a curve** through a set of points. Our **model** will be
  \[ D_i = \sum_{n=1}^{N} w_n f_{ni} + \nu_i \]
- The **likelihood** is given by
  \[ p(D|w, \gamma, H) = \prod_{i=1}^{I} Gauss(D_i | \sum_{n=1}^{N} w_n f_{ni}, \gamma) \]
- Choosing **conjugate priors** for the **parameters** such that
  \[ p(w|H) = Gauss(w|0, a^w \mathbf{I}) \]
  \[ p(\gamma|H) = Gamma(\gamma|a^{(\gamma)}, b^{(\gamma)}) \]
- where the **hyperparameters** are chosen to provide sufficiently broad priors
- **Assume** posterior distributions are **separable**
Posterior distributions

• The corresponding optimal approximate distributions are given by

\[ Q(w) = Gauss(w|\mu, \sigma) \]
\[ Q(\gamma) = Gamma(\gamma|a, b) \]

where the parameters are given by

\[ \sigma = a(w)I + \langle \gamma \rangle_Q ff^T \]
\[ \mu = \sigma^{-1} \langle \gamma \rangle_Q fD \]
\[ a = a(\gamma) + \frac{1}{2} \sum_{i=1}^{I} \left\langle \left( D_i - \sum_{n=1}^{N} w_n f_{ni} \right)^2 \right\rangle_Q \]
\[ b = b(\gamma) + \frac{I}{2} \]

Since parameters are coupled the above equations form a Markov blanket and several iterations are required for them to stabilise
Interpolation

- Real Model
- Data
- Maximum Likelihood
- Variational Ensemble Learning
MCMC methods

- Sample from a probability distribution

- Shift the probability distribution such that it tends asymptotically to the posterior

- Sample from the posterior distribution and then calculate the evidence using thermodynamic integration
The evidence is given as
\[ E = P(D|H) = \int P(D|\Theta, H)P(\Theta|H)d\Theta = \int Ld\pi \]
and can be generalised to
\[ E = \int L^\lambda d\pi \]
which can be differentiated to give
\[ \frac{d(\log E)}{d\lambda} = \frac{\int L^\lambda \log Ld\pi}{\int L^\lambda d\pi} = \langle \log L \rangle_\lambda \]
Because \( E(0) = 1 \), the evidence we are trying to calculate becomes
\[ E = E(1) = \exp \int_0^1 \frac{d(\log E)}{d\lambda} d\lambda = \exp \int_0^1 \langle \log L \rangle_\lambda d\lambda \]
Mixture Models

- Allow us to obtain **distributions** with a **richer** probability density
- **Mixture models** are used as source priors in **Blind Deconvolution**
- For **distributions** in the **exponential** family the **exact expectation** of mixture models is often **intractable**
- **Jensen’s inequality** has to be invoked which affects the **evidence bound**
- This allows us write the **cost function** as a **sum of logarithms**
Cluster Model Selection

- Ensure both **MCMC methods** and **Variational methods** perform inference correctly and select the correct number of clusters.

**MCMC samples**

**Variational Method**

<table>
<thead>
<tr>
<th>No. of Clusters</th>
<th>Evidence bound &lt; Evidence bound</th>
<th>Evidence bound &lt; Evidence bound</th>
<th>Evidence bound &lt; Evidence bound</th>
<th>Evidence bound &lt; Evidence bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-581</td>
<td>-462</td>
<td>-491</td>
<td>-470</td>
</tr>
<tr>
<td>3</td>
<td>436</td>
<td>800</td>
<td>796</td>
<td>787</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Deconvolution

- Assume a **gaussian** beam function

- **Convolve** it with a **source** and **add** some **noise** for old times’ sake

- Assume each **pixel** is **independent** and **positive** hence choose a **prior** which reflects this

- **Deconvolve** it using a **variational** approach
Blind Deconvolution

• What if we’re not sure what the beam function is?

• Now we need to infer both the beam function and estimate the source function

• Assumption #1 – the beam function is smaller than the source function

• Assumption #2 – the beam function is INDEPENDENT of the source function
Independence

• **Statistical independence** implies a **factorisation** of the joint probability densities

• **Independence** is measured by the **mutual information** between sources which is **defined** in terms of their **entropies**

• **Minimising** the mutual information between sources ensures independence

• Various other **measures** of independence exist – the most noted being **non-gaussianity**
Frequentist Approach

- Use kurtosis as an estimator of non-gaussianity
- This is defined as $E\{x^4\} - 3(E\{x^2\})^2$ which is zero for gaussian distributions
- The main advantage of this approach is its computational and theoretical simplicity
- The main disadvantage is that kurtosis is very sensitive to outliers which does not result in great separation in noisy environments
Independent Sources
Bayesian Approach

• Maximise the **likelihood** to find a **single mixing matrix** and **source model parameters** which most likely gave rise to the **observed data**

• It is **difficult** to incorporate **prior knowledge** into a **maximum likelihood** framework

• **Maximum likelihood** models have a tendency to **over-fit**

• Require the use of **priors** to form the **posterior density**

• **Spurious** structure in the **posterior** may also lead to **over-fitting**
Results

- Implementation of **blind deconvolution** algorithm for **sparse sources**

[Images of data, recovered source, inferred beam, and input map]
The Struggle

- Diamond Peak

[Images of recovered source, recovered filter, reconstructed data, and data plots]
Current Work

- Using independent component analysis to reduce test time at ST Microelectronics
- Implementing Laplacian priors in fMRI time series to ensure positivity
- Mixture of Variational/MCMC methods for efficient exploration of posterior distributions
- Studying the advantages of using the Variational approximation for proposal densities