MEG Source Localization Using an MLP With Distributed Output Representation

Sung Chan Jun, Barak A. Pearlmutter, Guido Nolte

CBLL Meeting – October 5, 2005
Source localization

- Definition: Identification of brain regions that emit detectable electromagnetic signals
- Assumption of dipole
- Application: detection of regions of the brain that cause epilepsy (for further neurosurgery)

Magnetoencephalography (MEG)

- Electrical activity produces magnetic fields
  Field $10^{-15}$ T (earth $0.5 \times 10^{-4}$ T)
  - Field recorded outside skull
    (Skull and tissue do not affect signal)
- SQUID = Superconductor Quantum Device, immersed in liquid He
  - Eliminates impedance
  - Allows for high sensitivity
- Magnetic shield room needed
- 4-D Neuroimag-122 as:
  122 pairs of sensors measuring signal over time (time resolution 1ms)

Source: www.elekta.com
How MEG works?

- Electromagnetic induction: magnetic field causes electric current in a moving coil
- Property used by magnetometers
- Gradiometers: Measure value of magnetic field between two different points, i.e. its gradient

Sources:
Example of MEG recordings
Why MEG?

- EEG = Electroencephalography
  - Time to place and calibrate EEG sensors
  - Skull has low conductivity to electric current but is transparent to magnetic fields
  - MEG not invasive, compared to placing EEG under skull
- fMRI = functional Magnetic Resonance Imaging
  - Time resolution 1s instead of 1ms
- Sources:
Inverse problem

- Localize source dipole \( \mathbf{x} \) given sensor activation \( \mathbf{B}_m \)
- General approach uses forward modeling:
  1) Compute \( \mathbf{B}_c(\mathbf{x}) \) from \( \mathbf{x} \)
  2) Compute cost \( c(\mathbf{x}) = |\mathbf{B}_c(\mathbf{x}) - \mathbf{B}_m|^2 \)
    Iterate on \( \mathbf{x} \) to minimize \( c(\mathbf{x}) \)
- Used algorithms: Simplex, LM, ANN

(Jun et al., 2002)
Forward model

- 122 pairs of sensors
- Source dipole location $\mathbf{x}$ and moment $\mathbf{Q}$ ($\mathbf{Q}$ can be deduced from $\mathbf{x}$ and $B_s$)
- Sensor $s$ has activation $B_s(\mathbf{x}, \mathbf{Q})$

$$B_s(\mathbf{x}, \mathbf{Q}) = \frac{[M(\mathbf{x}, \mathbf{Q}; \mathbf{t})|_{t=x_s^1} - M(\mathbf{x}, \mathbf{Q}; \mathbf{t})|_{t=x_s^2}] \cdot r_s}{|x_s^1 - x_s^2|}, \quad s = 1, \cdots, 122,$$

$$M(\mathbf{x}, \mathbf{Q}; x_s) = \frac{\mu_0}{4\pi} \frac{F \mathbf{Q} \times \mathbf{x} - (\mathbf{Q} \times \mathbf{x} \cdot x_s) \nabla F_x}{F^2},$$

$$F(\mathbf{x}, x_s) = d(x_s d + x_s^2 - (x_s \cdot x))$$

$$\nabla F_x(\mathbf{x}, x_s) = \left( \frac{d^2}{x_s} + \frac{(d \cdot x_s)}{d} + 2d + 2x_s \right) x_s - \left( d + 2x_s + \frac{(d \cdot x_s)}{d} \right) x$$

$$d = x_s - x, \quad d = |x_s - x|, \quad x_s = |x_s|,$$

(Jun et al., 2002)
About the dataset

- Geometry of sensors:
- Dataset consists in:
  - Pairs \((x, B_s)\)
  - Training dataset
  - Testing dataset
- Sensor activations = forward model + noise model
- Examples of noise model: abrupt visual stimulation followed by brief motor output and audio feedback, measured far from the stimulus or response

(Jun et al., 2002)
Distributed representation

- Usual architecture, “Cartesian-MLP” (Jun et al., 2002 and 2005) gives 3 outputs for dipole location

- Proposed approach in Jun et al., 2003: “distributed representation” Dipole location $\mathbf{x}$ deduced from $\mathbf{G(x)}$, vector of Gaussian receptive fields

- Receptive field $i$ located at $\mathbf{x}_i$ (regularly distributed) is:
  $$G_i(\mathbf{x}) = \exp\left(-\|\mathbf{x} - \mathbf{x}_i\|^2\right)/2\sigma^2$$
Decoding $G(x)$ to $x$

- **Strategy 1:**
  Find $i^* = \arg \max_i G_i(\tilde{x})$
  Interpolate between centers $x_j$ inside of a ball $B_{r_i}$ centered on $x_{i^*}$ and of radius 6cm
  \[ \sum_{x_i \in B_{i^*}} G_i(\tilde{x}) x_i / \sum_{x_i \in B_{i^*}} G_i(\tilde{x}) \]
  $6cm = \text{twice the inter-center distance}$

- **Strategy 2:**
  For each receptive field center $x_i$, compute within ball $B_i$
  \[ c_i = G_i(x) + \sum_{j \neq i, x_j \in B_i} G_i(x) / \|x_i - x_j\| \]
  Find $i^* = \arg \max_i c_i$
  Interpolate between centers $x_j$ inside of a ball $B_{r_i}$
  \[ \sum_{x_i \in B_{i^*}} G_i(\tilde{x}) x_i / \sum_{x_i \in B_{i^*}} G_i(\tilde{x}) \]
  (i.e. take into account neighborhood influence)
Soft-MLP architecture

Receptive field centers $\mathbf{x}_i$ evenly distributed every 3cm and cover training region

$$G_i(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{2\sigma^2}\right)$$

1.8cm

“Hyperbolic activation units to accelerate training”, c.f. LeCun 1991
Soft-MLP architecture

- Input data preprocessing: MEG sensor activations scaled to RMS=0.5
- Training algorithm: Backpropagation with online stochastic gradient descent
  Learning rate $\eta$ empirically chosen
Choice of MLP architecture

- Number of nodes in hidden layer
  \( N = 20, 40, 60, 80, 120, 160 \)
- 500 training epochs
- Training (noise free) dataset size:
  500, 1000, 2000, 4000, 8000, 16000, 32000
- Testing (noise free) dataset size:
  5000
- Generalization error averaged over 5 runs
  - Asymptote in generalization error:
    \( N = 80 \) hidden units, 8000 training samples
Dataset

- Training: 20000 samples (with brain noise)
- Testing: 4500 samples (with brain noise)
- 500 epochs
- 12 hours per training!?

- Signal to noise ratio in data:
  \[ \text{SNR} = 20 \log_{10} \frac{P_s}{P_n} \]
  (Noise has RMS:
  \( P_n = 50 \) to \( 100 \) fT/cm)
Results

- Soft-MLP faster and more accurate than Cartesian-MLP
- But hybrid methods (when MLP provides initial guess for LM, Levenberg-Marquard optimization) are still better
Comparison vs. LM alone

(Jun et al. 2002)

<table>
<thead>
<tr>
<th>algorithm</th>
<th>trained noise</th>
<th>computation time (ms)</th>
<th>localization error (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>fixed-4-start-LM</td>
<td>—</td>
<td>449</td>
<td>1.16</td>
</tr>
<tr>
<td>random-20-start-LM</td>
<td>—</td>
<td>2175</td>
<td>0.31</td>
</tr>
<tr>
<td>optimal-1-start-LM</td>
<td>—</td>
<td>22</td>
<td>0.23</td>
</tr>
<tr>
<td>N</td>
<td>0.3</td>
<td></td>
<td>2.70</td>
</tr>
<tr>
<td>W</td>
<td>0.3</td>
<td></td>
<td>1.64</td>
</tr>
<tr>
<td>C</td>
<td>0.3</td>
<td></td>
<td>2.06</td>
</tr>
<tr>
<td>B</td>
<td>0.3</td>
<td></td>
<td>1.15</td>
</tr>
<tr>
<td>N</td>
<td>53</td>
<td></td>
<td>0.84</td>
</tr>
<tr>
<td>W</td>
<td>41</td>
<td></td>
<td>0.44</td>
</tr>
<tr>
<td>C</td>
<td>49</td>
<td></td>
<td>0.67</td>
</tr>
<tr>
<td>B</td>
<td>36</td>
<td></td>
<td>0.28</td>
</tr>
</tbody>
</table>

"Real brain noise"
Comparison with commercial software in MEG Neuroimag-122

xfit = commercial software
Soft-MLP used with Strategy 2
Good initial guess
Possible extensions

- Cartesian-MLP can localize only 1 dipole
- Soft-MLP could localize more...
- Better than global search algorithm
Area of research since 1991

- Use of MLP with backpropagation on EEG signals
Questions…

- Why use only 1 sample of B's recording at a time? Does the shape of the MEG curves play a role?

- …