Westfälische Wilhelms-Universität Münster

The Inverse Problem of EEG/MEG

Vorlesung “Spezielle Fragen der inversen Probleme in der Bildbearbeitung”

SS 2012

wissen.leben
WWU Münster

Felix Lucka

14.05.2012

Outline

Introduction & Motivation: EEG/MEG

Mathematical Forward Modeling & Simulation

The Inverse Problem: Concepts and Methods

Electroencephalography (EEG) and Magnetoencephalography (MEG)

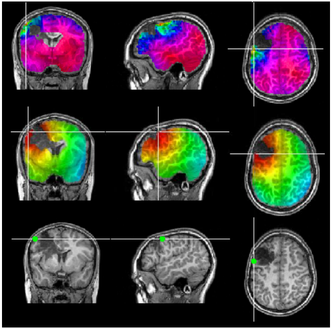
Aim: Reconstruction of brain activity by non-invasive measurement of induced electromagnetic fields (bioelectromagnetism) outside of the skull.

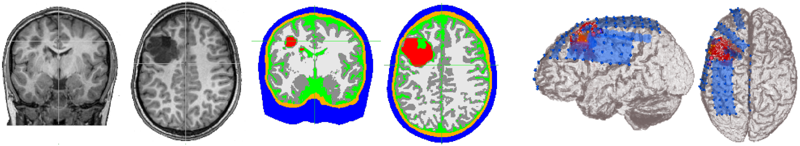
source: Wikimedia Commons

source: Wikimedia Commons

Applications of EEG/MEG

- Diagnostic tool in neurology, e.g., Epilepsy.
- Scientific applications:
  - Examination tool in several fields neuroscience.
  - Validation of therapeutic approaches in clinical neuroscience.
  - Examination tool for neurophysiology.



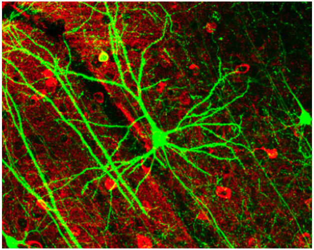


Neural Generators

Signals derive from the net effect of ionic currents flowing in the dendrites of neurons during correlated synaptic transmission.

EEG: Extracellular volume currents produced by postsynaptic potentials.
$\rightarrow$ strongly dependent on tissue's conductivity.

MEG: Intracellular currents associated with these postsynaptic potentials.
$\rightarrow$ less dependent on tissue's conductivity.



source: Wikimedia Commons

Outline

Introduction & Motivation: EEG/MEG

Mathematical Forward Modeling & Simulation

The Inverse Problem: Concepts and Methods

Basics of Mathematical Modeling

The interaction of electromagnetic fields with current and charge densities is described by Maxwell's equations.

Let $\vec{j}(\vec{r}) : \mathbb{R}^3 \to \mathbb{R}^3$ be a current density and $\rho(\vec{r}) : \mathbb{R}^3 \to \mathbb{R}$ be a charge density.
$\sigma(\vec{r}) : \mathbb{R}^3 \to \mathbb{R}$ is a conductivity distribution (isotropic).
The induced electric field $\mathbf{E}(\vec{r}) : \mathbb{R}^3 \to \mathbb{R}^3$ and the induced magnetic field $\mathbf{B}(\vec{r}) : \mathbb{R}^3 \to \mathbb{R}^3$ are given by:

Maxwell's equations (differential, microscopic form):

$$\mathrm{div}(\mathbf{E}) = \rho/\sigma \qquad \mathrm{rot}(\mathbf{E}) = -\partial_t \mathbf{B}$$

$$\mathrm{div}(\mathbf{B}) = 0 \qquad \mathrm{rot}(\mathbf{B}) = \mu_0 \cdot (\vec{j} - \sigma \partial_t \mathbf{E})$$

...4 coupled, (non-linear) time-dependent PDEs!

source: Wikimedia Commons

Basics of Mathematical Modeling

Maxwell's equations (differential, microscopic form):

$$\mathrm{div}(\mathbf{E}) = \rho/\sigma \qquad \mathrm{rot}(\mathbf{E}) = -\partial_t \mathbf{B}$$

$$\mathrm{div}(\mathbf{B}) = 0 \qquad \mathrm{rot}(\mathbf{B}) = \mu_0 \cdot (\vec{j} - \sigma \partial_t \mathbf{E})$$

source: Wikimedia Commons

Interpretation:

Under certain conditions, a vector field **A** is uniquely defined by its source-density div(**A**) and its circulation-density rot(**A**).

Maxwell's equations describe the origin of these two components for **E** and **B**.

Basics of Mathematical Modeling

Maxwell's equations (differential, microscopic form):

$$\mathrm{div}(\mathbf{E}) = \rho/\sigma \qquad \mathrm{rot}(\mathbf{E}) = -\partial_t \mathbf{B}$$

$$\mathrm{div}(\mathbf{B}) = 0 \qquad \mathrm{rot}(\mathbf{B}) = \mu_0 \cdot (\vec{j} - \sigma \partial_t \mathbf{E})$$

source: Wikimedia Commons

Interpretation:

$\mathrm{div}(\mathbf{E}) = \rho/\sigma$ is a differential formulation of Coulomb's law, which describes the force between charges.

In casual terms:

"The sources of the electric flux density ($D = \sigma \mathbf{E}$) are the free charges."

Basics of Mathematical Modeling

Maxwell's equations (differential, microscopic form):

$$\mathrm{div}(\mathbf{E}) = \rho/\sigma \qquad \mathrm{rot}(\mathbf{E}) = -\partial_t \mathbf{B}$$

$$\mathrm{div}(\mathbf{B}) = 0 \qquad \mathrm{rot}(\mathbf{B}) = \mu_0 \cdot (\vec{j} - \sigma \partial_t \mathbf{E})$$

source: Wikimedia Commons

Interpretation:

$\mathrm{rot}(\mathbf{E}) = -\partial_t \mathbf{B}$ is a differential formulation of Faraday's law of induction .

In casual terms:

"The temporal change of a magnetic field induces an electric curl field, oriented opposite to its cause. (Lenz's law)."

Basics of Mathematical Modeling

Maxwell's equations (differential, microscopic form):

$$\mathrm{div}(\mathbf{E}) = \rho/\sigma \qquad \mathrm{rot}(\mathbf{E}) = -\partial_t \mathbf{B}$$

$$\mathrm{div}(\mathbf{B}) = 0 \qquad \mathrm{rot}(\mathbf{B}) = \mu_0 \cdot (\vec{j} - \sigma \partial_t \mathbf{E})$$

source: Wikimedia Commons

Interpretation:

$\mathrm{div}(\mathbf{B}) = 0$ states, that the magnetic field is source free.

In casual terms:

"There are no magnetic monopoles"
or
"The magnetic field is a pure curl field"

Basics of Mathematical Modeling

Maxwell's equations (differential, microscopic form):

$$\mathrm{div}(\mathbf{E}) = \rho/\sigma \qquad \mathrm{rot}(\mathbf{E}) = -\partial_t \mathbf{B}$$

$$\mathrm{div}(\mathbf{B}) = 0 \qquad \mathrm{rot}(\mathbf{B}) = \mu_0 \cdot (\vec{j} - \sigma \partial_t \mathbf{E})$$

source: Wikimedia Commons

Interpretation:

$\mathrm{rot}(\mathbf{B}) = \mu_0 \cdot \vec{j}$ is a differential formulation of Ampère's circuital law.
In casual terms:

"The motion of electric charges (i.e., current) induces a magnetic curl field (orientation by right-hand rule)"

The addition of Maxwell's displacement current $(-\mu_0 \cdot \sigma \cdot \partial_t \mathbf{E})$ is less intuitive
.

Basics of Mathematical Modeling

Maxwell's equations (differential, microscopic form):

$$\text{div}(\mathbf{E}) = \rho/\sigma \qquad \text{rot}(\mathbf{E}) = -\partial_t \mathbf{B}$$
$$\text{div}(\mathbf{B}) = 0 \qquad \text{rot}(\mathbf{B}) = \mu_0 \cdot (\vec{j} - \sigma \partial_t \mathbf{E})$$

Still 4 coupled, (non-linear) time-dependent PDEs!

Simplifying assumptions:

- *Linearity:* Body ≈ passive conductor
- *Quasistatic approximation:* Temporal changes << spatial propagation velocity; Tissue is time-independent and has no inductance.
- *Charge-free:* No macroscopic charge aggregation.
- *Primary- and volume currents:* Separate current into a primary and resulting volume current.

Basics of Mathematical Modeling

Forward/Direct Problem of EEG/MEG

Let $\sigma(\vec{r})$ be the conductivity and $\vec{j}^{pri}(\vec{r})$ a primary current density in $\Omega \subset \mathbb{R}^3$.
The electric potential u on $\partial\Omega$ is given by::

$$\begin{aligned} \nabla \cdot (\sigma \nabla u) &= \nabla \cdot \vec{j}^{pri} && \text{in } \Omega \\ n \cdot (\sigma \nabla u) &= 0 && \text{on } \partial\Omega \ (\textit{no-penetration condition}) \\ \int_{\partial\Omega} u \cdot dS &= 0 && (\textit{fix ground potential}) \end{aligned}$$

The magnetic field $\mathbf{B}$ can be conducted by (Biot-Savart):

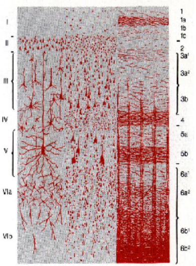
$$\mathbf{B}(\vec{r}) = \frac{\mu_0}{4\pi} \int_\Omega \left\{ \vec{j}^{pri}(\vec{r}\,') - \sigma(\vec{r}\,') \cdot \nabla u(\vec{r}\,') \right\} \times \frac{\vec{r} - \vec{r}\,'}{\|\vec{r} - \vec{r}\,'\|^3} d\vec{r}\,' \qquad \text{for } \vec{r} \in \mathbb{R}^3 \backslash \bar{\Omega}$$

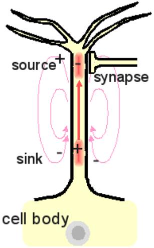
Solving the forward problem necessitates concerning 3 things:

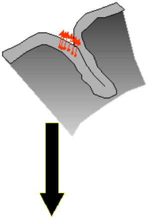
- A source-model for $\vec{j}^{pri}$: How can we model the macroscopic current-flows?
- A volume-conductor-model of $\sigma(\vec{r})$: How can we model the dielectric properties of the different tissues?
- A numerical method for solving the PDE w.r.t. to source and volume conductor model; mostly FEM or BEM approaches.

Basics of Mathematical Modeling

Common source model: Equivalent current dipoles, $\vec{j}^{pri}(x) = \sum_i M_i \delta(x - x_i)$





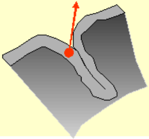


Equivalent Current Dipole (Primary current) (~50 nAm)

parameters:

position : x_0

moment : M

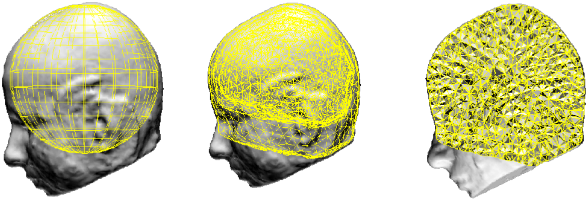


Size of Macroscopic Neural Activity

~30 mm² = **5.5×5.5 mm²**

Problem for certain numerical methods:
$\delta(r) \in H^{-3/2-\varepsilon}(\Omega)\ \forall \varepsilon > 0$ and $D^\alpha \delta(r) \in H^{-3/2-|\alpha|-\varepsilon}(\Omega)\ \forall \varepsilon > 0.$

Forward Computation Methods



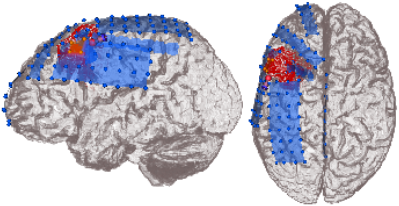
Sphere: Under the assumption of modeling the head by a multi-layer sphere model, a (quasi-)analytic solution exists.

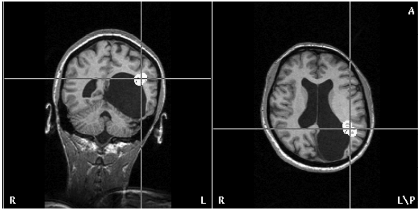
BEM: Assuming a nested shell topography boundary element methods can be used, demanding the discretization of the compartment boundaries.

FEM: Finite element methods are based upon a discretization of the whole volume conductor.

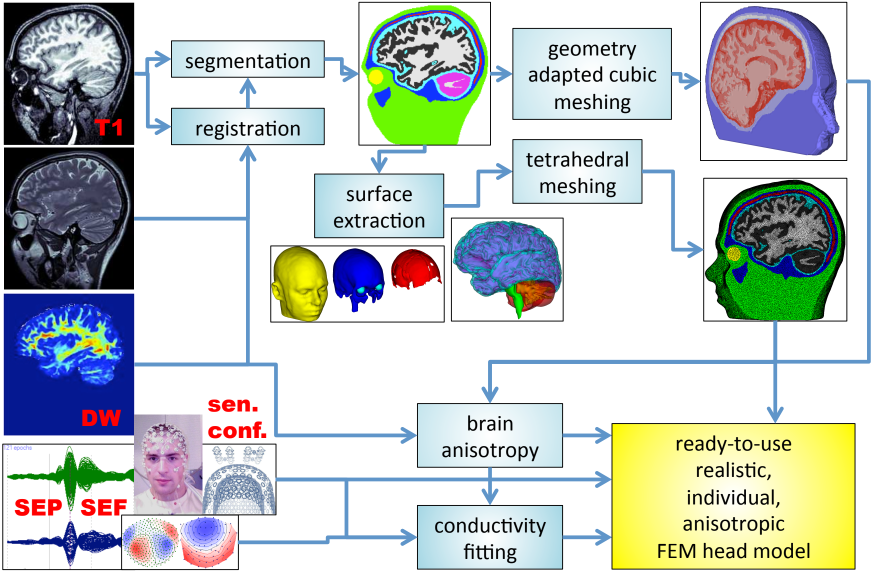
Why FEM?

- Possibility to incorporate nearly arbitrary complex geometries and arbitrary number of compartments:
  - CSF, gray/white matter, cerebellum, brain steam, muscles, dura mater, blood vessels;
  - Realistic skull modeling: Skull holes, three-layeredness.
  - Anatomical anomalies from surgeries / brain damages.
- Modeling of invasive recording devices (ECoG, depth-electrodes)
- Inclusion of anisotropic conductivities, e.g., white matter anisotropy

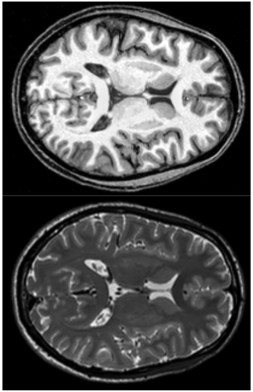




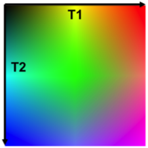
Realistic, individual head modeling for bioelectromagnetic applications

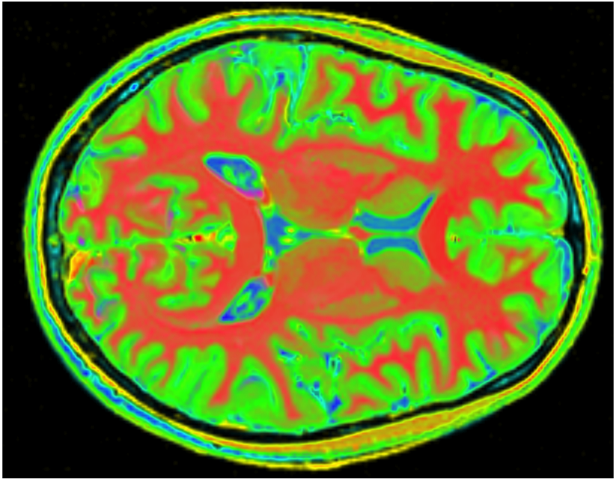


Part 1: MRI Processing, Structural Scans

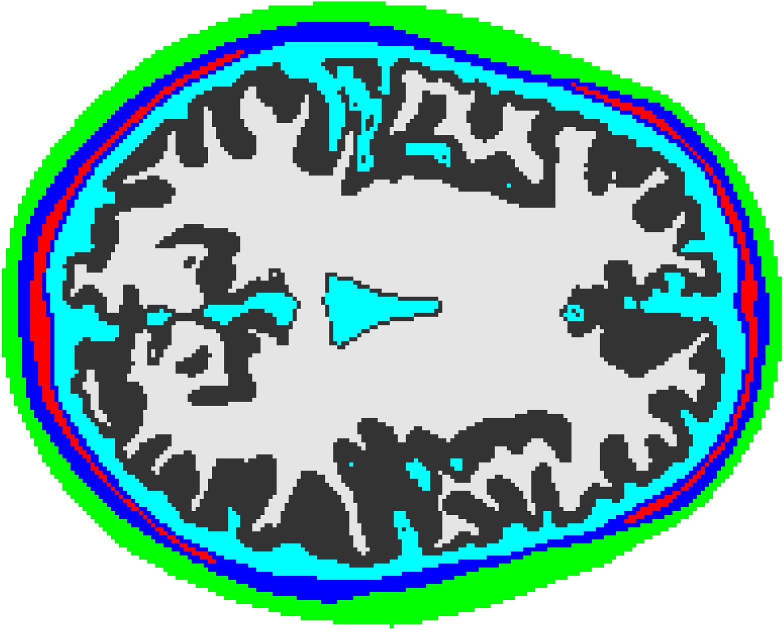


RGB map



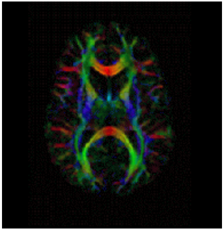


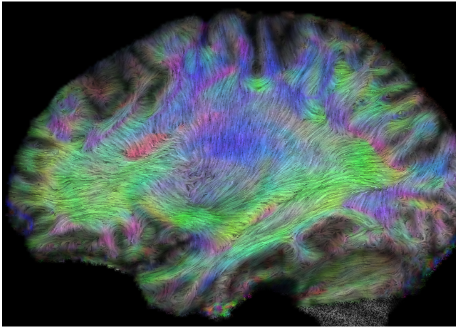
RGB composite of T1 and T2 MRI scan

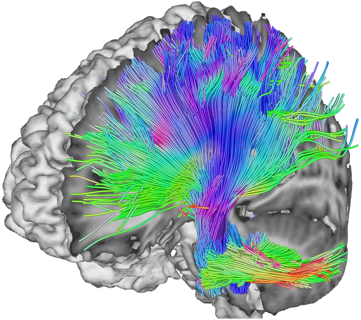


Part 1: MRI Processing, Diffusion Weighted MRI

- DW-MRI allows the mapping of diffusion processes of molecules in biological tissues, in vivo.
- Clinical application: Localization of white matter lesions in stroke patients, surgical planning.
- Key imaging modality to assess connectivity via tractography .
- We use it to compute conductivity tensors.

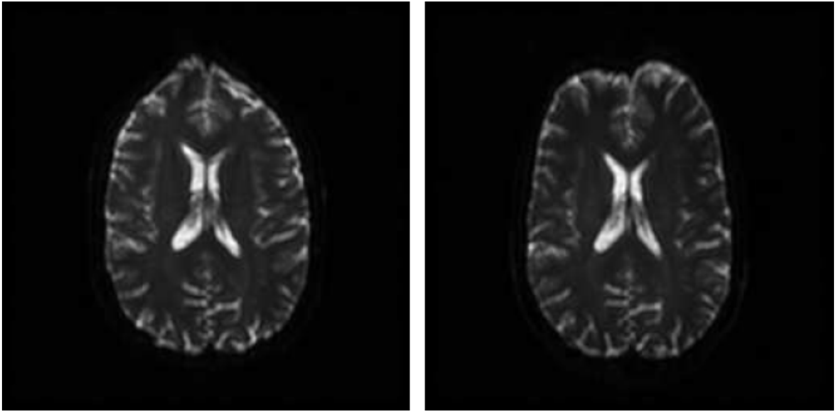






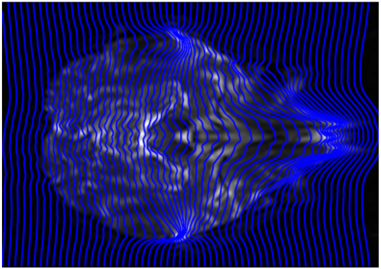
Part 1: MRI Processing, Diffusion Weighted MRI

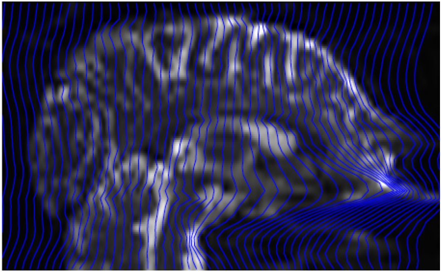
Fast Echo-Planar Imaging (EPI)



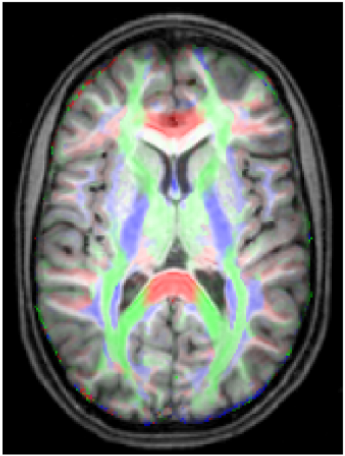
Part 1: MRI Processing, Diffusion Weighted MRI

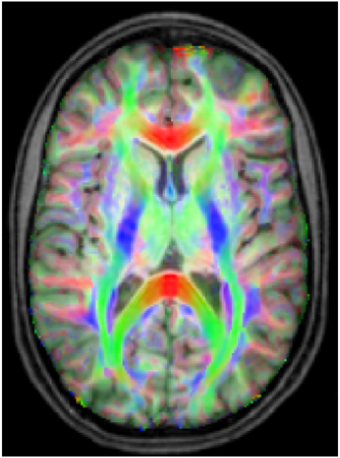
New, non-linear, variational registration approach (DA and PhD by Lars Ruthotto):



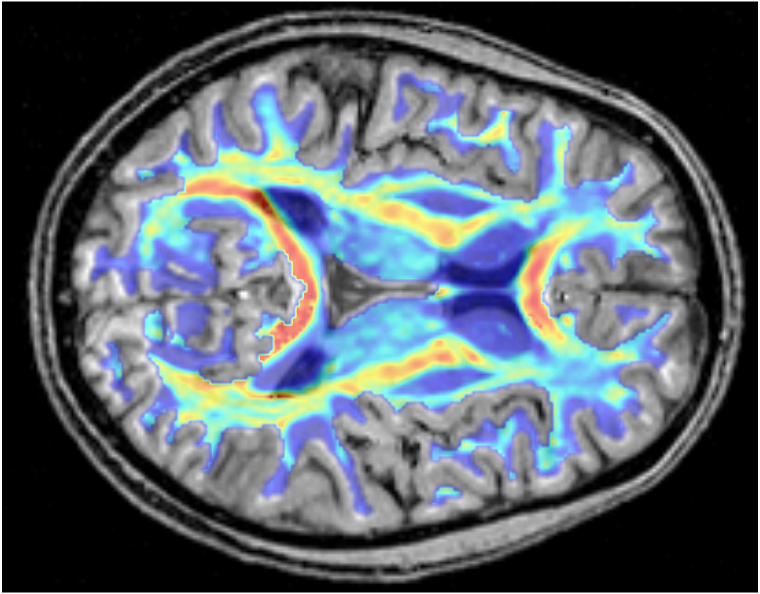


Part 1: MRI Processing, Diffusion Weighted MRI

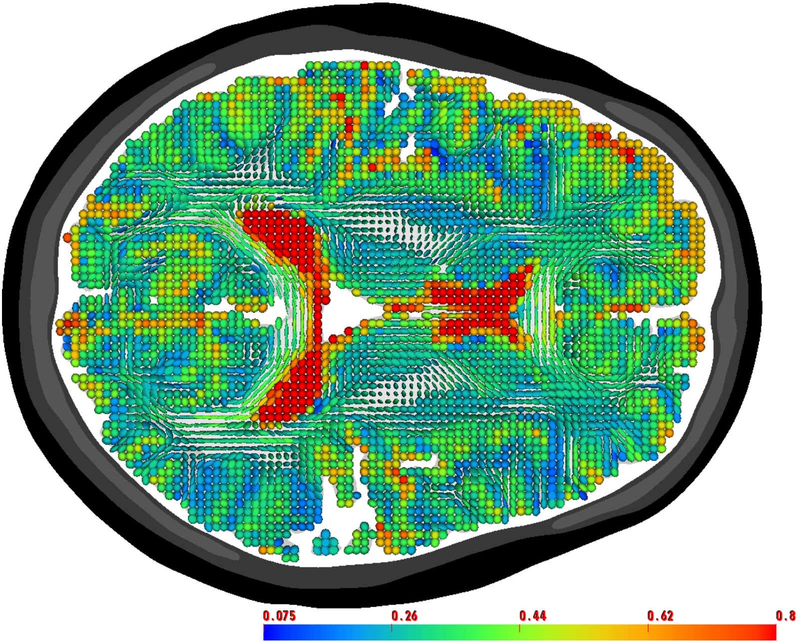




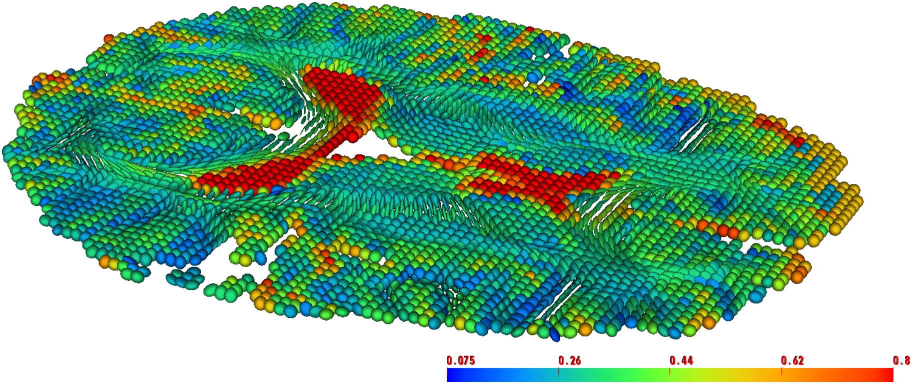
Part 1: MRI Processing, Diffusion Weighted MRI



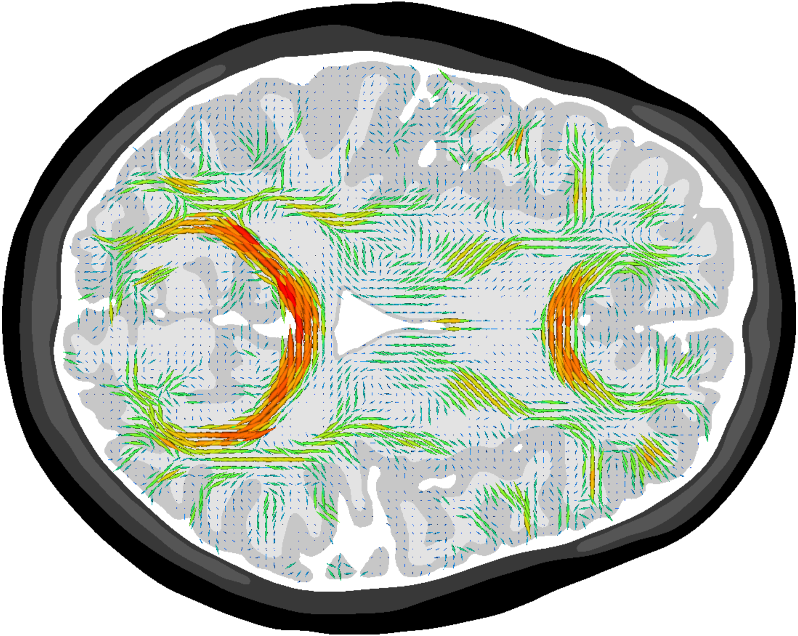
Part 1: MRI Processing, Diffusion Weighted MRI



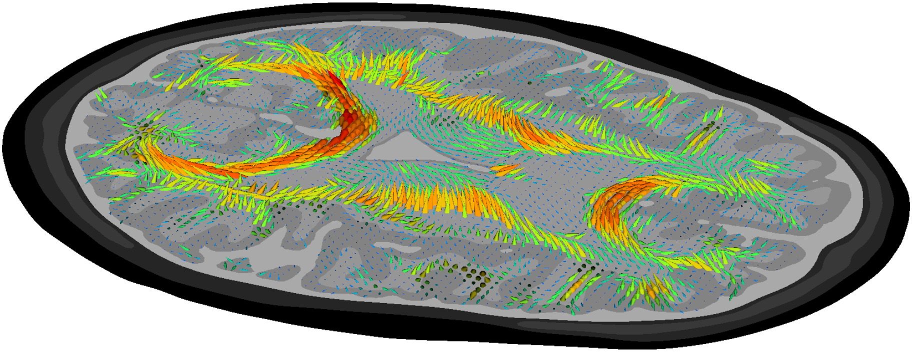
Part 1: MRI Processing, Diffusion Weighted MRI



Part 1: MRI Processing, Diffusion Weighted MRI

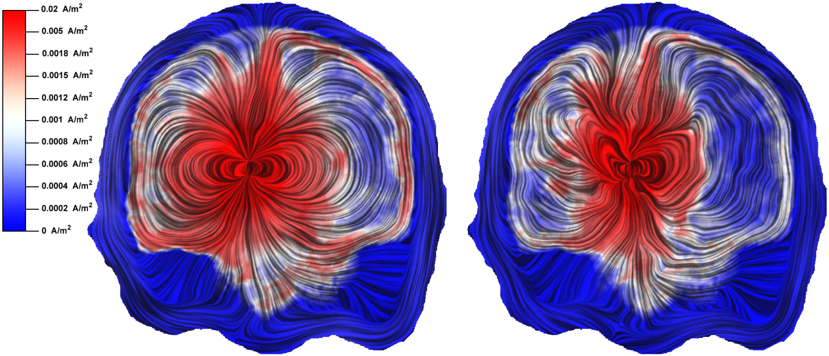


Part 1: MRI Processing, Diffusion Weighted MRI

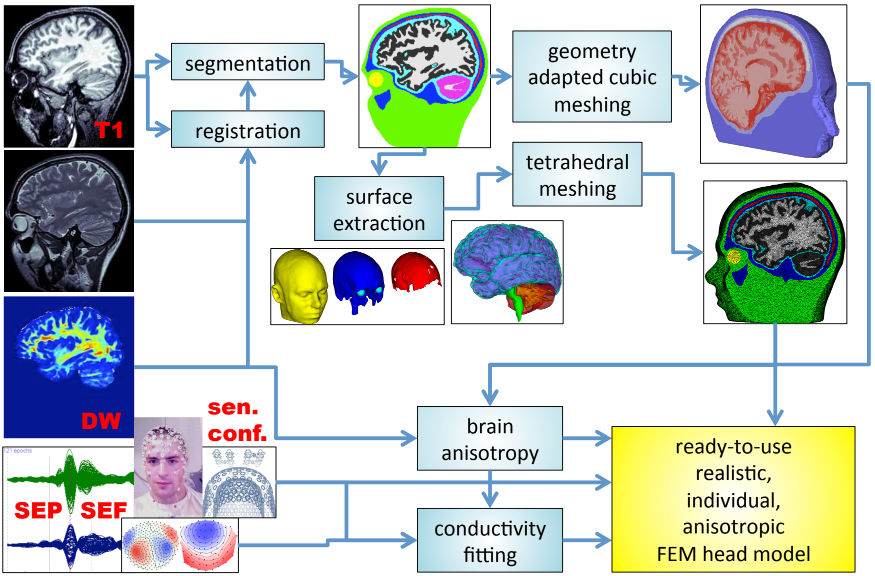


Part 1: MRI Processing, Diffusion Weighted MRI

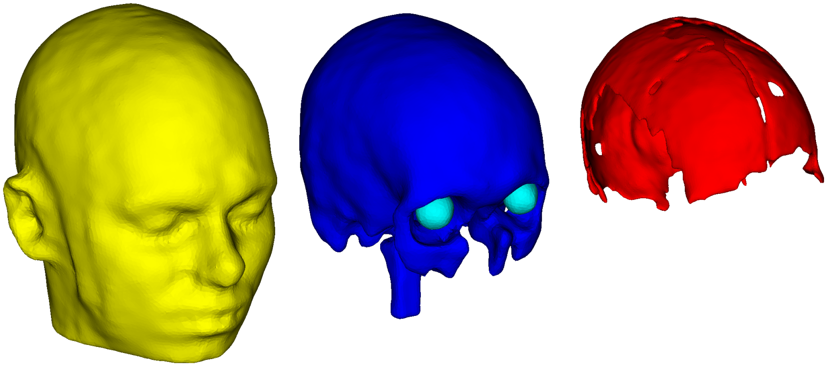
Effects of white matter anisotropy on thalamic source:



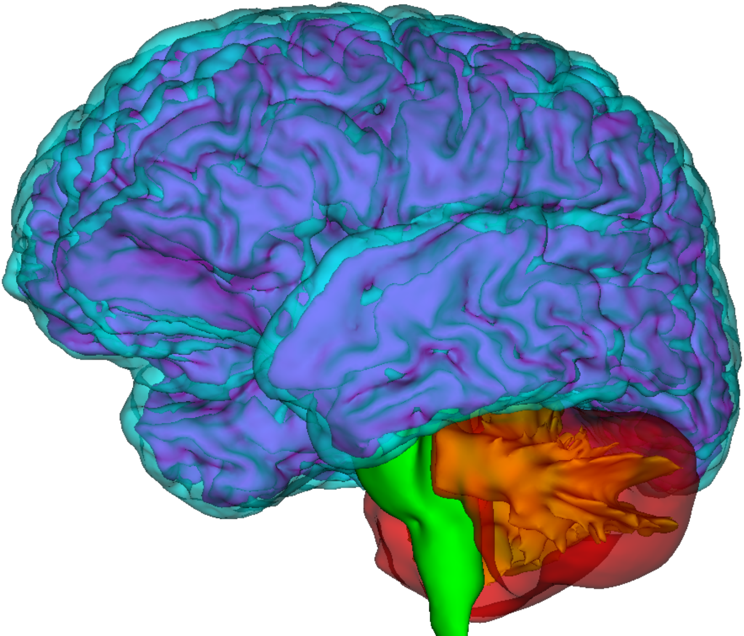
Realistic, individual head modeling for bioelectromagnetic applications



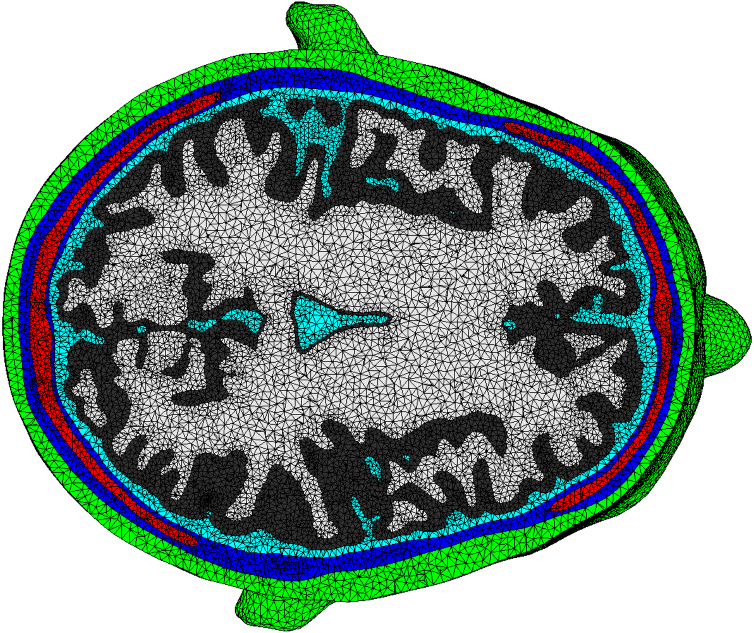
Part 2: FEM Meshing



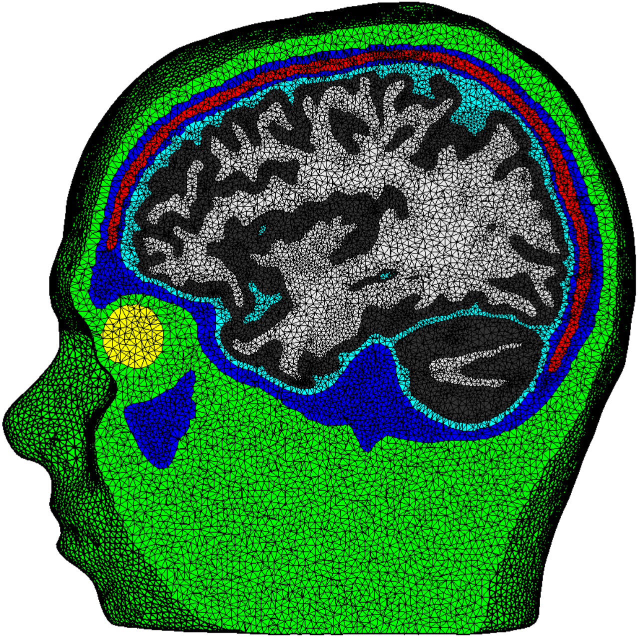
Part 2: FEM Meshing



Part 2: FEM Meshing



Part 2: FEM Meshing



Outline

Basics of the Inverse Problem

Inverse Problem of EEG/MEG Source Reconstruction

Given

- measurements b of the electric potential u and/or of the normal-component of the magnetic field $\langle n, \mathbf{B} \rangle$ on the surface $\partial\Omega$;
- a volume-conductor-model of $\sigma(\vec{r})$;
- a source model $\mathcal{J} \subset \mathcal{D}'(\Omega, \mathbb{R}^3)$;

estimate the primary current $\vec{j}^{pri} \in \mathcal{J}$ (source) that is consistent with b and the neurophysiological constrains of brain activity.

Solving the inverse problem (source reconstruction) necessitates concerning 3 things:

- Data preprocessing: How can we clean/filter the data from external sources (artifacts), noise, unwanted brain activity components?
- A-prior modeling: How much and which assumptions on brain activity do we need to incorporate and how do we model them to stabilize the inverse problem?
- Implementation: How do we solve the inverse problem practically?

Characteristic Features of Inverse Problems

Hadamard's definition of *well-posed* problems:

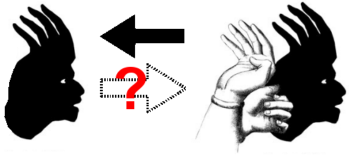
1. A solution exists.
2. The solution is unique.
3. The solution depends continuously on the data.

If one of the conditions does not hold, the problem is called ill-posed.

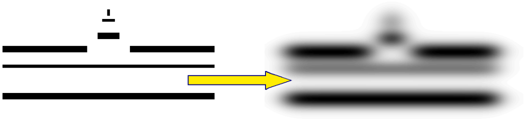
Inverse problems are typically ill-posed.

Jacques Salomon Hadamard
(1865-1963)

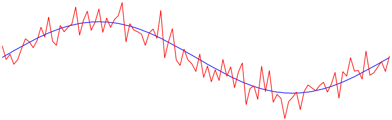
What About the Inverse Problem of EEG/MEG?



- (Presumably) under-determined



- Severely ill-conditioned



- Low SNRs

What About the Inverse Problem of EEG/MEG?

Summary: The problem is severely ill-posed.

Measurements alone are insufficient and unsuitable to determine solution.

$\Longrightarrow$ Incorporation of a-priori information about the solution in an explicit or implicit way:

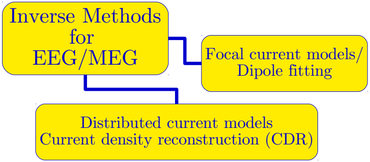
- Knowledge about general/specific brain activity?
- Integration of multimodal information (fMRI, DW-MRI, PET)?
- Mathematical formulation?
- Computational implementation?

$\Longrightarrow$ Variety of inverse methods for EEG/MEG (“*curse of interdisciplinary*“)

Different Approaches to the Inverse Problem

Inverse Methods
for
EEG/MEG

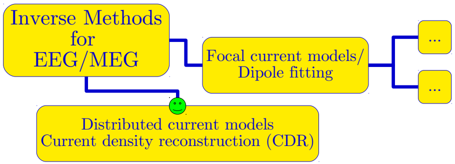
Different Approaches to the Inverse Problem



Focal current modeling: Reconstruction using a small number of *dipoles* with arbitrary locations and orientations.
Unknown number of sources/spatial extent? $\Longrightarrow$ not suitable

Distributed current modeling: Discretization of underlying current field using a large number of focal elementary sources with fixed locations and orientations.

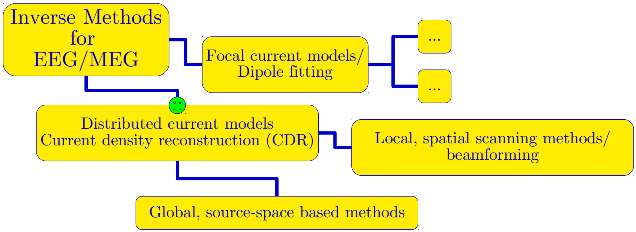
Different Approaches to the Inverse Problem



Focal current modeling: Reconstruction using a small number of *dipoles* with arbitrary locations and orientations.
Unknown number of sources/spatial extent? $\Longrightarrow$ not suitable

Distributed current modeling: Discretization of underlying current field using a large number of focal elementary sources with fixed locations and orientations.

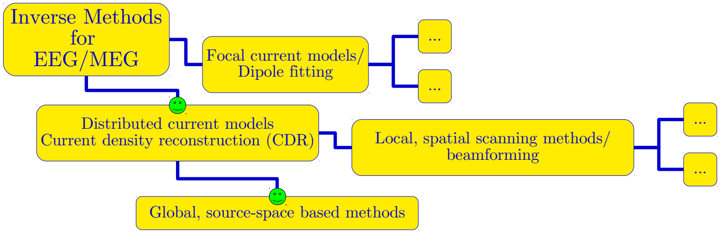
Different Approaches to the Inverse Problem



Local, spatial scanning methods/beamforming: Repeatedly optimize the estimate at a single location or a small region while suppressing crosstalk from other areas.

Global, source-space based methods: Incorporate a-priori information on the global properties of the solution.

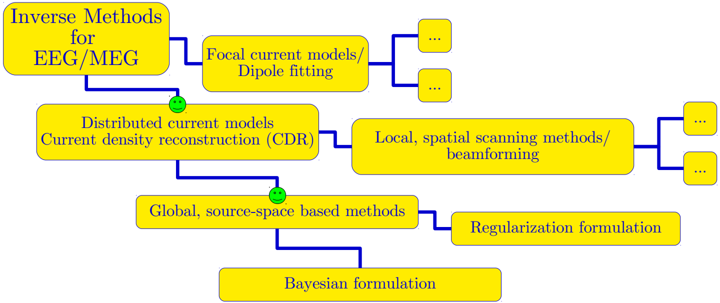
Different Approaches to the Inverse Problem



Local, spatial scanning methods/beamforming: Repeatedly optimize the estimate at a single location or a small region while suppressing crosstalk from other areas.

Global, source-space based methods: Incorporate a-priori information on the global properties of the solution.

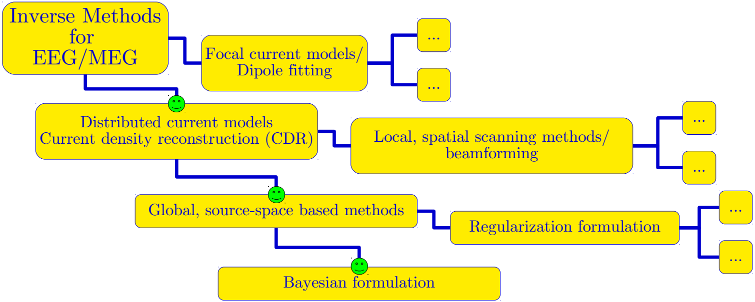
Different Approaches to the Inverse Problem



(Variational) regularization: Incorporate a-priori information through a variational framework (Tikhonov regularization).

Bayesian inference: Incorporate a-priori information through a statistical framework.

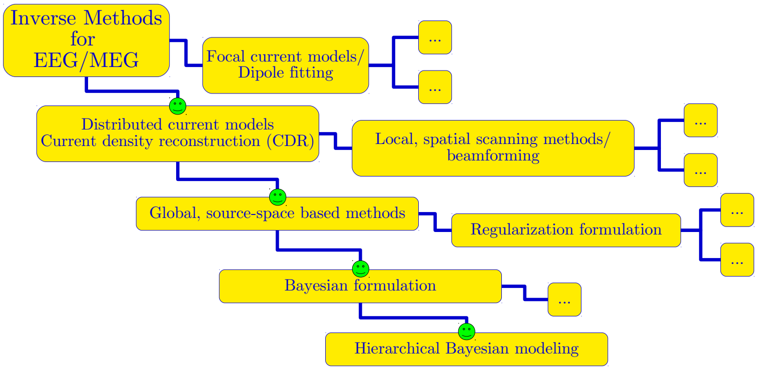
Different Approaches to the Inverse Problem



(Variational) regularization: Incorporate a-priori information through a variational framework (Tikhonov regularization).

Bayesian inference: Incorporate a-priori information through a statistical framework.

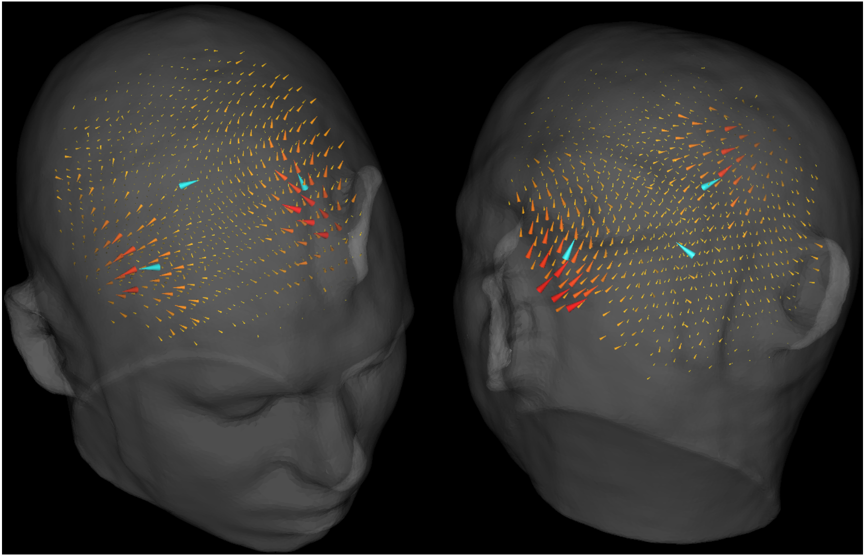
Different Approaches to the Inverse Problem



Hierarchical/Empirical Bayesian modeling: More later...

Current Density Reconstruction

Discretization of an underlying continuous current distribution by large number of current dipoles with fixed location and orientation.



Current Density Reconstruction

Lead-field matrix concept:

- $L \in \mathbb{R}^{m \times n}$; columns represent measurements at m sensors caused by the n single current dipoles.
- Linear combination of the dipoles is represented by source vector $s \in \mathbb{R}^n$.
- Measurements $b \in \mathbb{R}^m$ caused by s can then be calculated via:

$$b = L\,s$$

Current Density Reconstruction

Lead-field matrix concept:

- $L \in \mathbb{R}^{m \times n}$; columns represent measurements at m sensors caused by the n single current dipoles.
- Linear combination of the dipoles is represented by source vector $s \in \mathbb{R}^n$.
- Measurements $b \in \mathbb{R}^m$ caused by s can then be calculated via:

$$b = L\,s$$

Infer s from b? Apparently ill-posed problem:

- $n \gg m. \implies b = L\,s$ is under-determined.
- L inherits the bad condition of the continuous problem.
- Noise $\mathcal{E} \sim \mathcal{N}(0, \sigma^2 \mathrm{Id})$ is added to signal.

Variational/Tikhonov Regularization

- Original ill-posed problem is approximated by a well-posed problem.
- Degree of approximation is controlled by means of a regularization parameter λ, such that $\lambda \rightarrow 0$ corresponds to the solution given by a generalized inverse of L.
- Can be formulated as a minimization problem

$$s_\lambda = \underset{s \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \mathcal{D}(s) + \lambda \cdot \mathcal{P}(s) \right\}; \qquad \mathcal{D}(s) = \|b - Ls\|_2^2,$$

$\mathcal{D}(s)$: Data functional, controls the deviation from measurements.
$\mathcal{P}(s)$: Penalty functional, renders the problem well-posed, promotes solutions with certain qualities by penalizing all others.

Variational/Tikhonov Regularization

Theoretical perception of the inverse problem:

Find a generalized inverse to the forward operator with certain additional properties.

- Appropriate, and most often used point of view for the analysis of methods based on these approaches is functional analysis.
- Noise $\Longrightarrow$ choice of the data functional; seen as implicit obstacle; stochastic nature of the process (and thus of the solution given by the method) is more or less neglected.
- The penalization of unwanted features is a form of an implicit modeling; no reference to an underlying explicit model.
- High dimension of the source space $\Longrightarrow$ minimization can involve sophisticated optimization techniques.

Variational/Tikhonov Regularization, Examples

Classical minimum norm estimate:
$\mathcal{P}(s) = \|s\|_2^2 \quad \longrightarrow \quad s_\lambda = \operatorname{argmin} \; \|b - Ls\|_2^2 + \lambda \|s\|_2^2$

Weighted minimum norm estimate:
$\mathcal{P}(s) = \|Ws\|_2^2 \quad \longrightarrow \quad s_\lambda = \operatorname{argmin} \; \|b - Ls\|_2^2 + \lambda \|Ws\|_2^2$
Possible weightings: Depth weighting, spatial smoothness, fMRI/SPEC/PET.

Minimum current estimate. Let s_i^* denote the amplitude of the current vector at position i:
$\mathcal{P}(s) = \|s^*\|_1 \quad \longrightarrow \quad s_\lambda = \operatorname{argmin} \; \|b - Ls\|_2^2 + \lambda \|s^*\|_1$

Weighted minimum current estimates:
$\mathcal{P}(s) = \|Ws^*\|_1 \quad \longrightarrow \quad s_\lambda = \operatorname{argmin} \; \|b - Ls\|_2^2 + \lambda \|Ws^*\|_1$
Possible weightings: Total Variation (TV, $W = \nabla$), depth weighting, fMRI/SPEC/PET

Bayesian Inference

Back to

$$b = L\,s + \mathcal{E}$$

- $n \gg m. \implies b = L\,s$ is under-determined.
- L inherits the bad condition of the continuous problem.
- Noise $\mathcal{E}$ is added to signal.

Bayesian Inference

Back to

$$b = L\,s + \mathcal{E}$$

- $n \gg m. \implies b = L\,s$ is under-determined.
- L inherits the bad condition of the continuous problem.
- Noise $\mathcal{E}$ is added to signal.

High uncertainty and under-determinateness of a problem?

$\implies$ Account for them explicitly by formulating the problem in a statistical framework

Strategy of Bayesian Inference

1. Make stochastic model for the relation between parameters, data and noise.

   - $B = L\, s + \mathcal{E}, \quad \mathcal{E} \sim \mathcal{N}(0, \sigma^2 \mathrm{Id})$ (b is now random variable B)

   - Compute probability density of B given s (likelihood):

$$p_{like}(b|s) \propto \exp\left(-\frac{1}{2\,\sigma^2}\|b - L\,s\|_2^2\right)$$

Strategy of Bayesian Inference

1. Make stochastic model for the relation between parameters, data and noise: $p_{like}(b|s)$.

2. Supplement information given by the data by a-priori information about the parameters of interest. $\longrightarrow$ Bayesian modeling:

- s is considered to be a random variable itself ($s \rightarrow S$).
- Its distribution $p_{prior}(s)$ reflects a-priori assumptions/knowledge.
- Task of the prior: Render the estimation problem well-posed.

Strategy of Bayesian Inference

1. Make stochastic model for the relation between parameters, data and noise: $p_{like}(b|s)$.

2. Supplement information given by the data by a-priori information about the parameters of interest: $p_{prior}(s)$

3. Merge information before the measurement (prior) with the information gained after performing the measurement (likelihood) by Bayes rule:

$$p_{post}(s|b) = \frac{p_{like}(b|s)p_{prior}(s)}{p(b)}$$

- Conditional distribution of S given B is called posterior distribution.
- Represents all information on S given the realization of $B = b$.
- Complete solution to the inverse problem in Bayesian Inference

Strategy of Bayesian Inference

1. Make stochastic model for the relation between parameters, data and noise: $p_{like}(b|s)$.

2. Supplement information given by the data by a-priori information about the parameters of interest: $p_{prior}(s)$

3. Merge information before the measurement (prior) with the information gained after performing the measurement (likelihood) by Bayes rule: $p_{post}(s|b)$

4. Exploit a-posteriori information by inferring point estimates:
    1. *Maximum a-posteriori-estimate (MAP)*: $\hat{s}_{\text{MAP}} := \operatorname{argmax}_{s \in \mathbb{R}^n} \; p_{post}(s|b)$.
       Practically: High-dimensional optimization problem.
    2. *Conditional mean-estimate (CM)*: $\hat{s}_{\text{CM}} := \mathbb{E}\,[s|b] = \int_{\mathbb{R}^n} s \; p_{post}(s|b) ds$.
       Practically: High-dimensional integration problem.

Strategy of Bayesian Inference, Connections to Variational Regularization

Consider

$$p_{prior}(s) \propto \exp\left(-\frac{\lambda}{2\,\sigma^2}\mathcal{P}(s)\right),$$

then

$$p_{post}(s|b) \propto \exp\left(-\frac{1}{2\,\sigma^2}\|b - L\,s\|_2^2 - \frac{\lambda}{2\,\sigma^2}\mathcal{P}(s)\right),$$

and

$$\begin{aligned}\hat{s}_{\mathrm{MAP}} &= \underset{s\in\mathbb{R}^n}{\operatorname{argmax}}\left\{\exp\left(-\frac{1}{2\,\sigma^2}\|b - L\,s\|_2^2 - \frac{\lambda}{2\,\sigma^2}\mathcal{P}(s)\right)\right\}\\ &= \underset{s\in\mathbb{R}^n}{\operatorname{argmin}}\left\{\|b - L\,s\|_2^2 + \lambda\mathcal{P}(s)\right\}\end{aligned}$$

The same applies for CM estimates in a less explicit way...

Bayesian Inference strategies

But there is more to Bayesian inference:

- Confidence intervals estimates
- Conditional covariance estimates
- Histogram estimates
- Marginalization
- Model selection or averaging
- Experiment design

And more complex models, i.e., empirical/hierarchical Bayesian models.

Empirical Bayesian Inference

Problem: Brain activity is too complex (or our knowledge is too limited) to be captured in a fixed but sufficiently informative prior.

Empirical Bayesian Inference

Problem: Brain activity is too complex (or our knowledge is too limited) to be captured in a fixed but sufficiently informative prior.

Solution: Let the same data determine the prior used for the inference based on this data!

Empirical Bayesian Inference

Problem: Brain activity is too complex (or our knowledge is too limited) to be captured in a fixed but sufficiently informative prior.

Solution: Let the same data determine the prior used for the inference based on this data!

Sounds like...

Empirical Bayesian Inference

Problem: Brain activity is too complex (or our knowledge is too limited) to be captured in a fixed but sufficiently informative prior.

Solution: Let the same data determine the prior used for the inference based on this data!

Sounds like...

...but can be formulated into a consistent, statistical reasoning by adding a new dimension of inference: Hyperparameters and hyperpriors.

Top-down construction scheme → Hierarchical Bayesian modeling (HBM).

Hierarchical Bayesian Modeling (HBM)

Overview:

- Current trend in all areas of Bayesian inference.
- Flexible framework for the construction of complex models.
- Adds an adaptive, data-driven element into the estimation.
- Automated reduction of complex models.
- Different levels for the embedding of qualitative or quantitative a-priori information.
- Embeds several heuristic approaches into sound mathematical framework.
- Comprises many former EEG/MEG methods.
- Offers various new ways of inference: Full-MAP, Full-CM, γ-MAP, S-MAP, VB

Example: Hierarchical Bayesian Modeling of Focal Activity

Wanted: A prior promoting focal source activity.

First try:

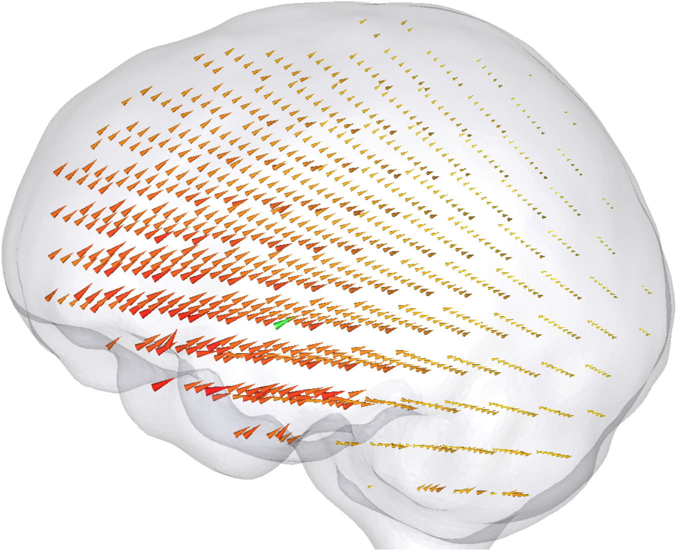
- Take Gaussian prior with zero mean and covariance $\Sigma_s = \gamma \cdot \mathrm{Id}, \quad \gamma > 0$ (*Minimum norm estimation*).
- Compute MAP or CM estimate (equal)!

$$\hat{s}_{\text{MAP}} := \underset{s \in \mathbb{R}^n}{\operatorname{argmax}} \left\{ \exp\left(-\frac{1}{2\,\sigma^2} \|b - \mathrm{L}\, s\|_2^2 - \frac{1}{2\gamma} \|s\|_2^2 \right) \right\}$$

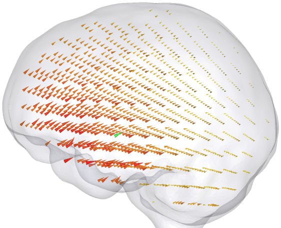
$$= \underset{s \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ \|b - \mathrm{L}\, s\|_2^2 + \frac{\sigma^2}{\gamma} \|s\|_2^2 \right\}$$

Example: Hierarchical Bayesian Modeling of Focal Activity

First try: NOT a focal reconstruction.



Example: Hierarchical Bayesian Modeling of Focal Activity



What went wrong?

- Gaussian variables = characteristic scale given by variance. (not *scale invariant*)
- All sources have variance $\gamma \Longrightarrow$ Similar amplitudes are likely.
- $\Longrightarrow$ Focal activity is very unlikely.

Example: Hierarchical Bayesian Modeling of Focal Activity

Idea:

- Let sources at single locations i have different variances γ_i.
- Let the data determine $\gamma_i \Longrightarrow$ New level of inference!
  - $\boldsymbol{\gamma} = (\gamma_i)_i$ are called hyperparameters.
  - Bayesian inference: $\boldsymbol{\gamma}$ are random variables as well.
  - Their prior distribution $p_{hyper}(\boldsymbol{\gamma})$ is called hyperprior.

Example: Hierarchical Bayesian Modeling of Focal Activity

Idea:

- Let sources at single locations i have different variances γ_i.
- Let the data determine $\gamma_i \Longrightarrow$ New level of inference!
  - $\boldsymbol{\gamma} = (\gamma_i)_i$ are called hyperparameters.
  - Bayesian inference: $\boldsymbol{\gamma}$ are random variables as well.
  - Their prior distribution $p_{hyper}(\boldsymbol{\gamma})$ is called hyperprior.
- Encode focality assumption into hyperprior:
  - Focality: Nearby sources should a-priori not be mutually dependent.
  - Focality: Most sources silent, few with large amplitude;
  - No location preference for activity should be given a priori.

Example: Hierarchical Bayesian Modeling of Focal Activity

Idea:

- Let sources at single locations i have different variances γ_i.
- Let the data determine $\gamma_i \Longrightarrow$ New level of inference!
  - $\boldsymbol{\gamma} = (\gamma_i)_i$ are called hyperparameters.
  - Bayesian inference: $\boldsymbol{\gamma}$ are random variables as well.
  - Their prior distribution $p_{hyper}(\boldsymbol{\gamma})$ is called hyperprior.
- Encode focality assumption into hyperprior:
  - γ_i should be stochastically independent.
  - Focality: Most sources silent, few with large amplitude;
  - No location preference for activity should be given a priori.

Example: Hierarchical Bayesian Modeling of Focal Activity

Idea:

- Let sources at single locations i have different variances γ_i.
- Let the data determine $\gamma_i \Longrightarrow$ New level of inference!
  - $\boldsymbol{\gamma} = (\gamma_i)_i$ are called hyperparameters.
  - Bayesian inference: $\boldsymbol{\gamma}$ are random variables as well.
  - Their prior distribution $p_{hyper}(\boldsymbol{\gamma})$ is called hyperprior.
- Encode focality assumption into hyperprior:
  - γ_i should be stochastically independent.
  - Sparsity inducing hyperprior, e.g., inverse gamma distribution.
  - No location preference for activity should be given a priori.

Example: Hierarchical Bayesian Modeling of Focal Activity

Idea:

- Let sources at single locations i have different variances γ_i.
- Let the data determine $\gamma_i \Longrightarrow$ New level of inference!
  - $\boldsymbol{\gamma} = (\gamma_i)_i$ are called hyperparameters.
  - Bayesian inference: $\boldsymbol{\gamma}$ are random variables as well.
  - Their prior distribution $p_{hyper}(\boldsymbol{\gamma})$ is called hyperprior.
- Encode focality assumption into hyperprior:
  - γ_i should be stochastically independent.
  - Sparsity inducing hyperprior, e.g., inverse gamma distribution.
  - γ_i should be equally distributed.

Example: Hierarchical Bayesian Modeling of Focal Activity

In formulas:

$$p_{prior}(s|\gamma) \sim \mathcal{N}(0, \Sigma_s(\gamma)), \quad \text{where} \quad \Sigma_s(\gamma) = \mathrm{diag}\left(\gamma_i \cdot \mathrm{Id}_3, i = 1, \ldots, k\right)$$

$$p_{hyper}(\gamma) = \prod_{i=1}^{k} p^i_{hyper}(\gamma_i) = \prod_{i=1}^{k} p_{hyper}(\gamma_i) = \prod_{i=1}^{k} \frac{\beta^\alpha}{\Gamma(\alpha)} \gamma_i^{-\alpha-1} \exp\left(-\frac{\beta}{\gamma_i}\right)$$

$\alpha > 0$ and $\beta > 0$ determine *shape* and *scale*, $\Gamma(x)$ denotes the Gamma function.

Joint prior: $p_{pr}(s, \gamma) = p_{prior}(s|\gamma)\, p_{hyper}(\gamma)$

Implicit prior:
$$p_{pr}(s) = \int p_{prior}(s|\gamma)\, p_{hyper}(\gamma) d\gamma$$
$$= \int \mathcal{N}(0, \Sigma_s(\gamma))\, p_{hyper}(\gamma) d\gamma \quad \rightsquigarrow \text{“Gaussian scale mixture”}$$

Example: Hierarchical Bayesian Modeling of Focal Activity

Posterior, general:

$$p_{post}(s, \gamma | b) \propto p_{like}(b|s) \; p_{prior}(s|\gamma) \; p_{hyper}(\gamma)$$

$$\text{Comparison:} \quad p_{post}(s|b) \propto p_{like}(b|s) \; p_{prior}(s)$$

Posterior, concrete:

$$p_{post}(s, \gamma | b) \propto$$
$$\exp\left(-\frac{1}{2\sigma^2} \|b - L\,s\|_2^2 - \sum_{i=1}^{k} \left(\frac{\frac{1}{2}\|s_{i*}\|^2 + \beta}{\gamma_i} + (\alpha + \tfrac{5}{2}) \ln \gamma_i \right) \right)$$

Analytical advantages...

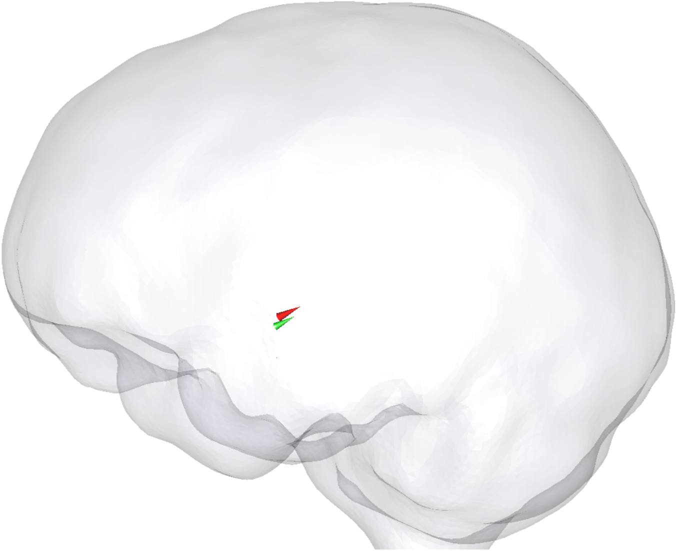
- Energy is quadratic with respect to s
- Factorizes over γ_i's.

and disadvantages...

- Energy is non-convex w.r.t. (s, γ) (posterior is multimodal)

Example: Hierarchical Bayesian Modeling of Focal Activity

Full-CM estimate computed via blocked Gibbs MCMC integration, see Calvetti et al., 2009.



Static vs. Dynamic Inverse Methods

Up to now we only looked at the static inverse problem...but:

- EEG and MEG offer an excellent temporal resolution.
- The temporal characteristics of brain dynamics attract growing attention.
- Incorporating the complete temporal data could stabilize the ill-posed inverse problem.

Approaches:

- Single-pass strategies: Extract information from one domain to enhance the reconstruction in the other domain.
- State-space approaches: Iterate between space and time to balance both sources of information.
- Multi-pass strategies: Incorporate information from both domains simultaneously (spatio-temporal inversion). Computationally expensive.


That's all Folks!