Solving the EEG inverse problem

Stefan Haufe

BBCI Winter School 2014, Berlin

Outline

1. Inverse source reconstruction

2. (Blind) source separation

Electroencephalography (EEG)

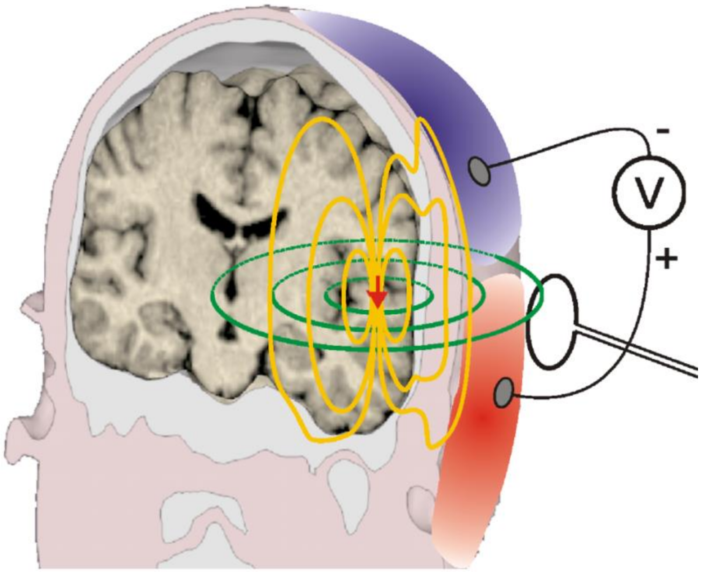
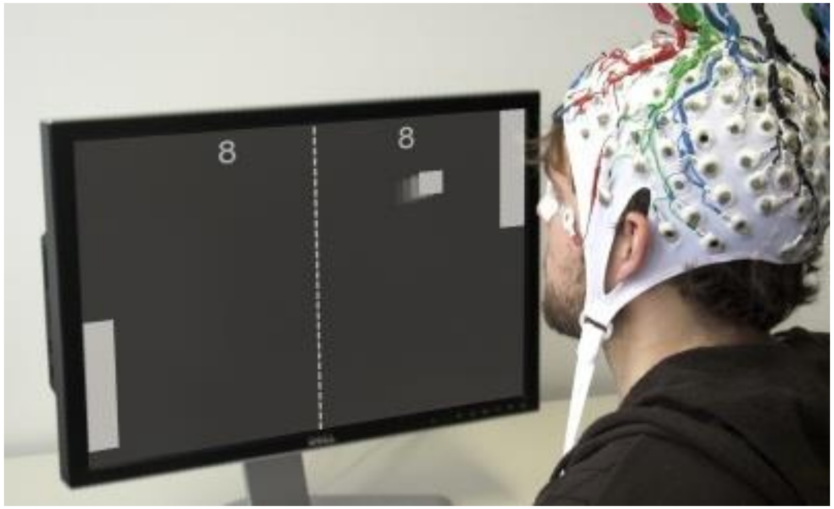


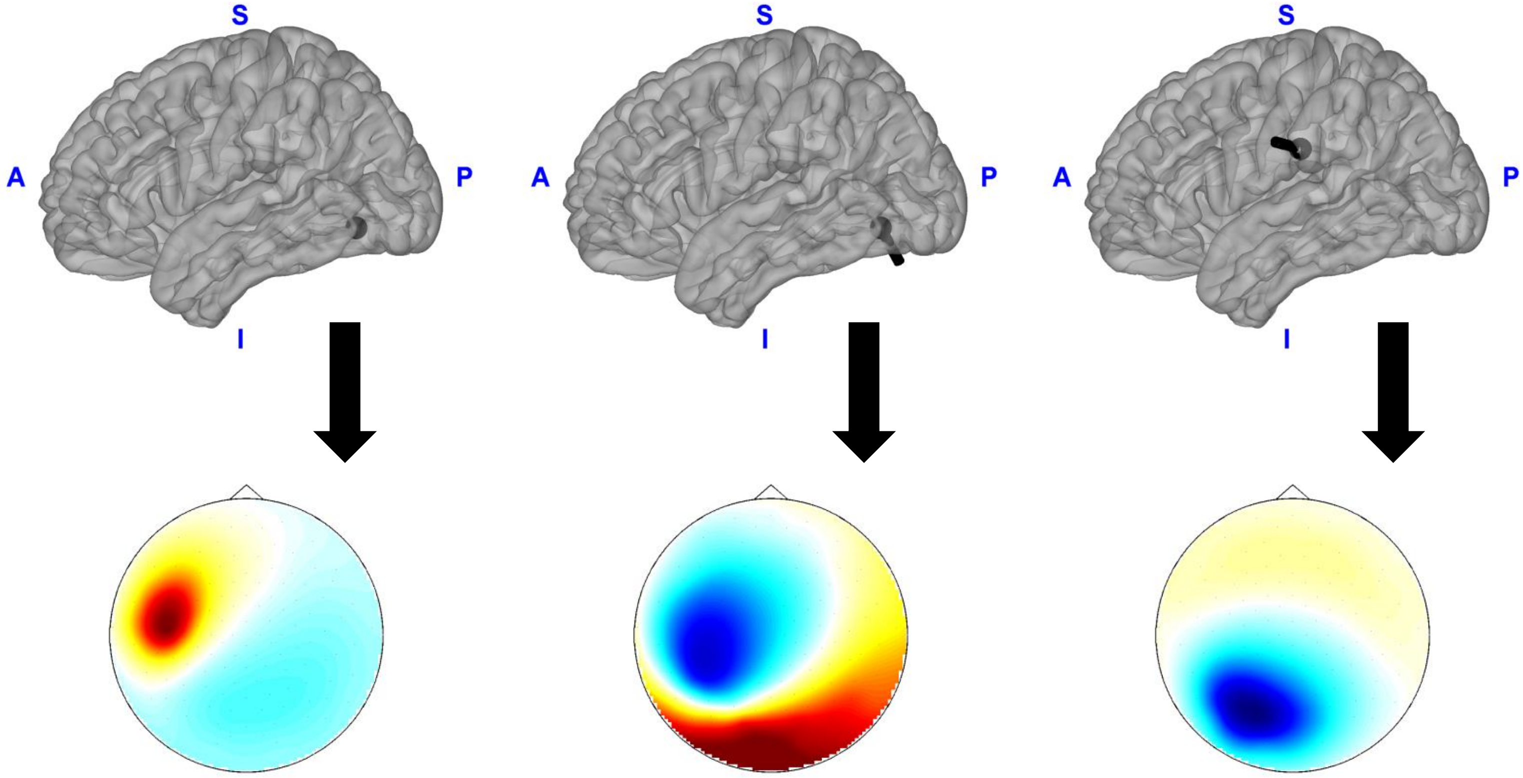
Figure by Lauri Parkkonen



Cellular (primary) currents due to synchronous firing of large populations of equally spatially coaligned neurons are accompagnied by extracellular return (secondary) currents measurable as extracranial electric potentials by EEG.

Volume conduction: attenuation and spatial smearing

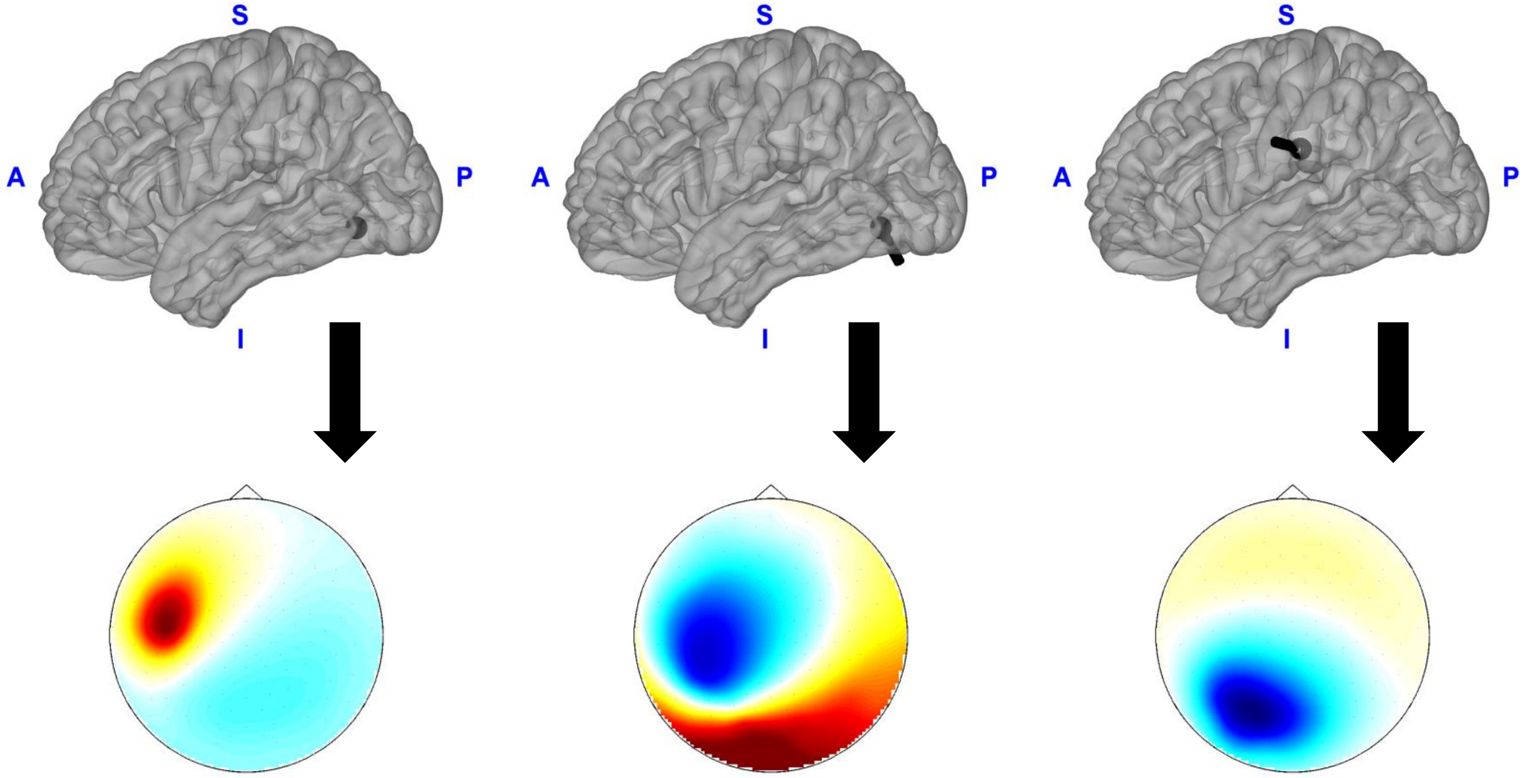
Primary current generator (dipole)



Resulting EEG scalp potential: ~5-10 μV

Volume conduction: attenuation and spatial smearing

Primary current generator (dipole)

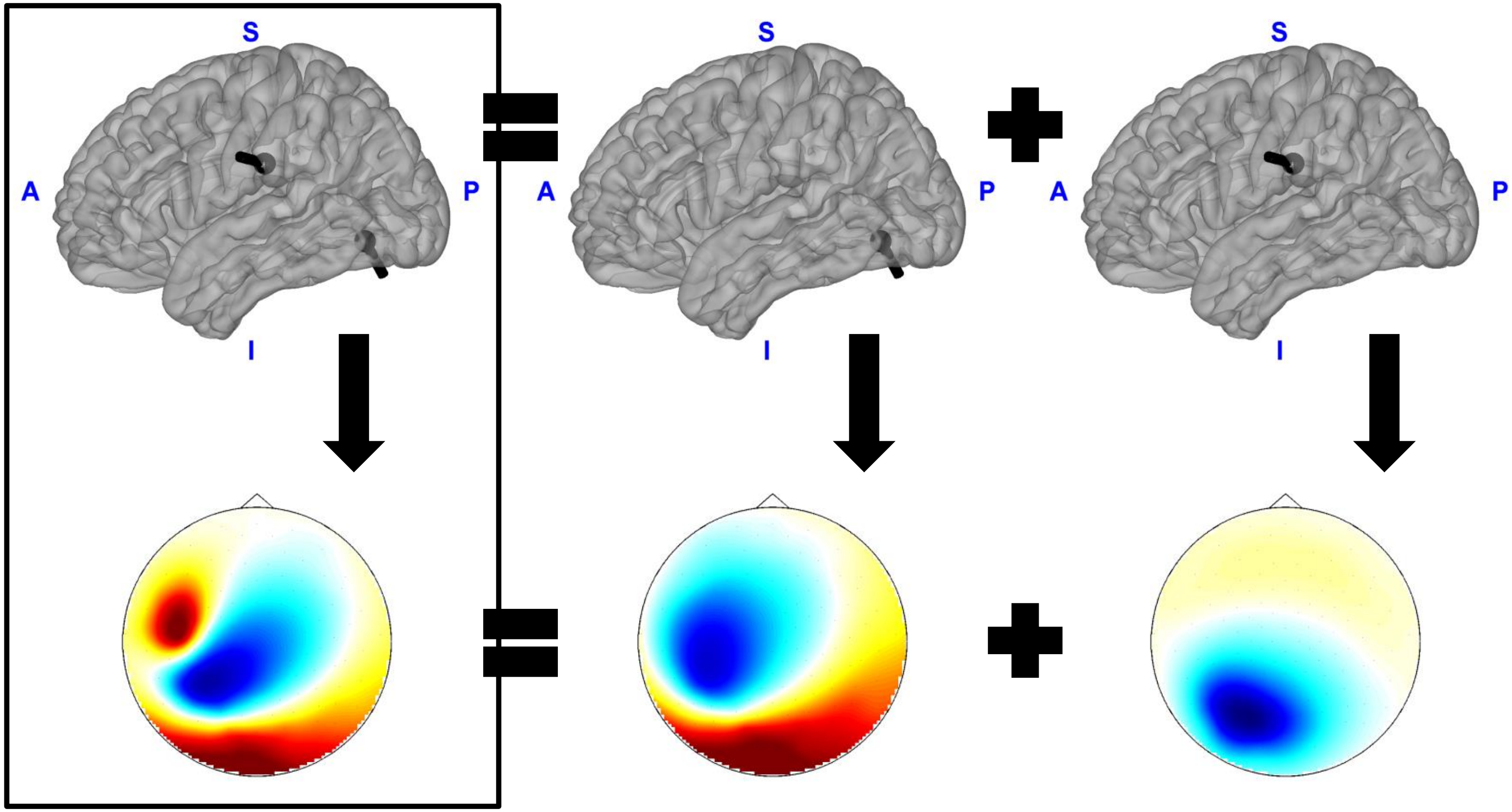


Resulting EEG scalp potential: ~5-10 µV

Artifacts: ~100 µV

Volume conduction: superposition of activity

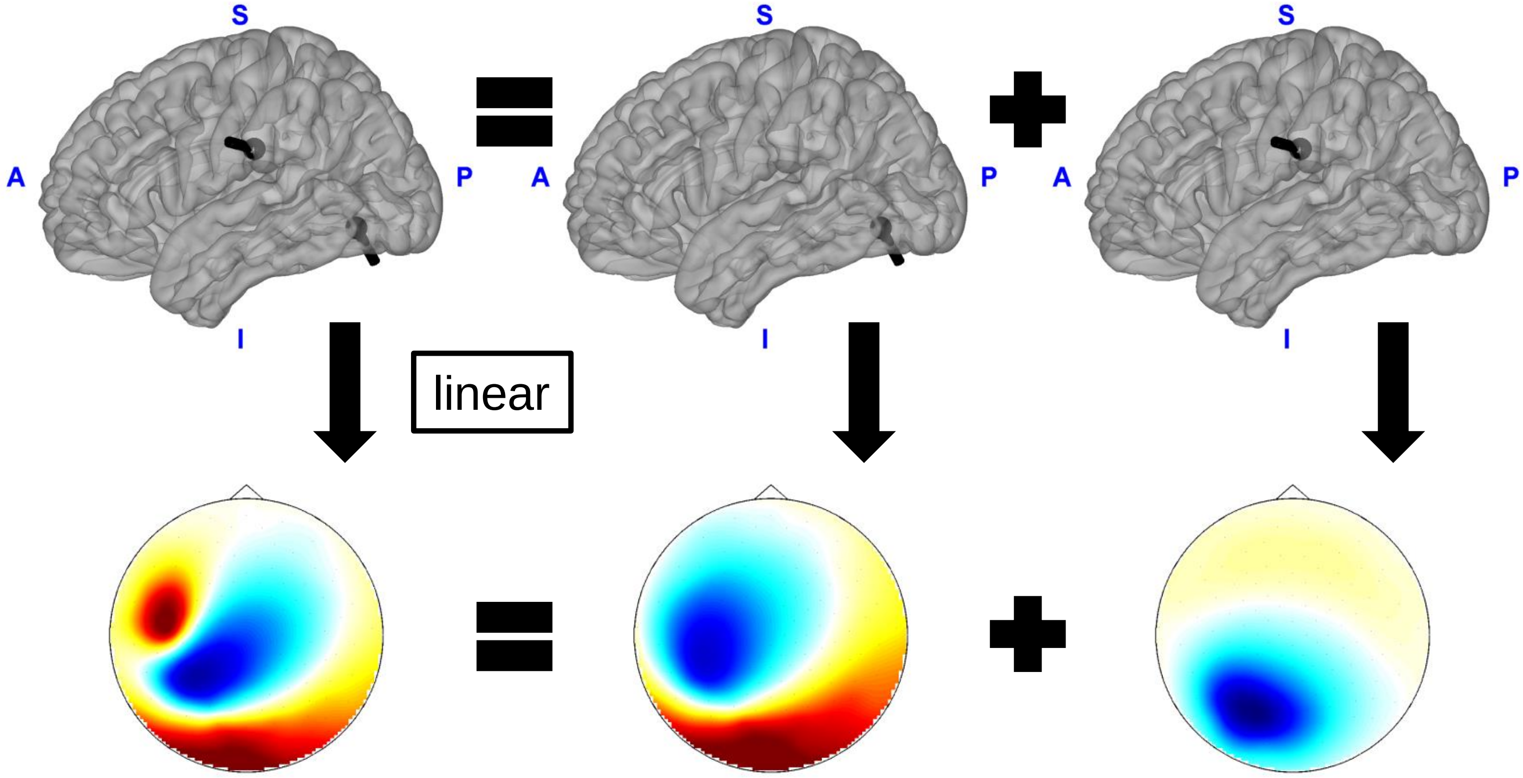
Primary current generator (dipole)



Resulting EEG scalp potential: ~5-10 μV

Volume conduction: superposition of activity

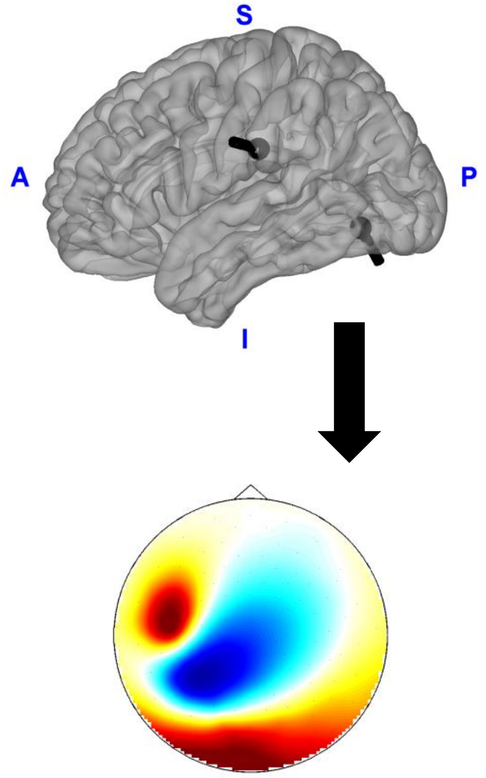
Primary current generator (dipole)

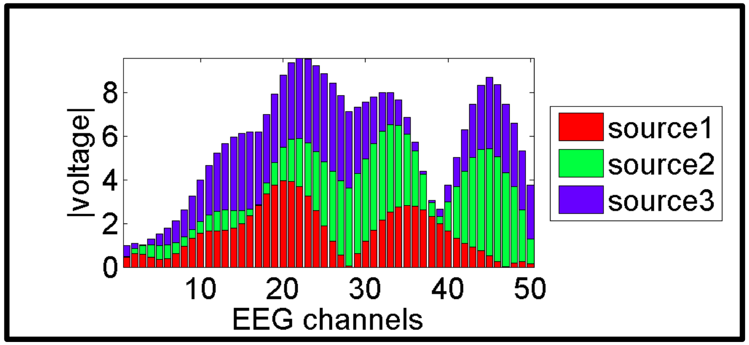


Resulting EEG scalp potential: ~5-10 μV

Volume conduction: superposition of activity

Primary current generator (dipole)

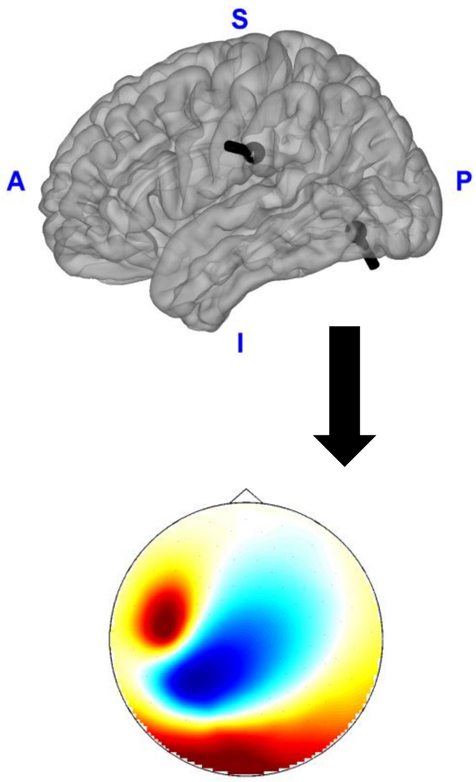


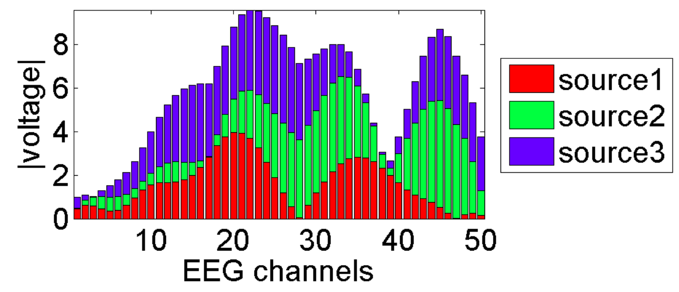


Resulting EEG scalp potential: ~5-10 μV

Volume conduction: difficulties caused by

Primary current generator (dipole)





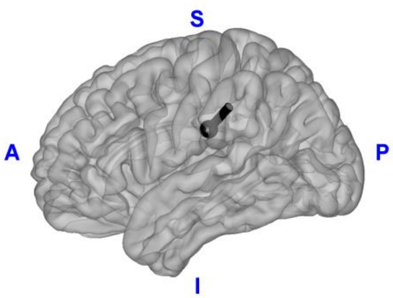
Difficulties for data analysis:

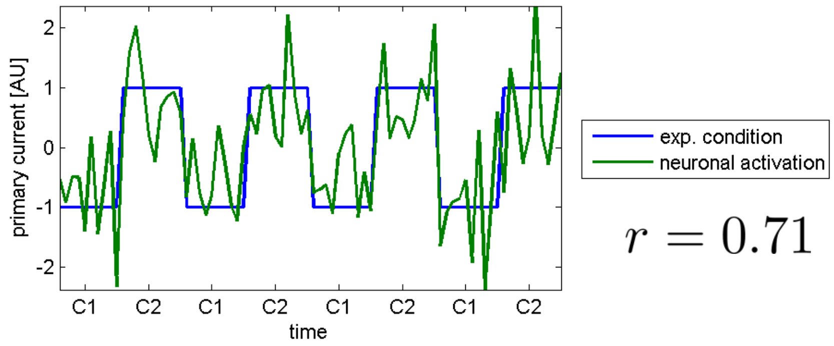
- Low SNR (small effect sizes, high p values)
- Localization/interpretation

Resulting EEG scalp potential: ~5-10 μV

Illustration: sensor-space analysis

Assume there is a brain area modulated by, e.g., the experimental condition.





$$r = 0.71$$

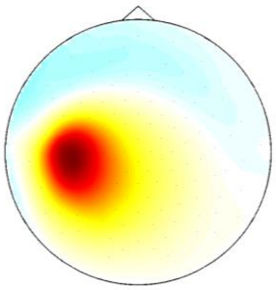
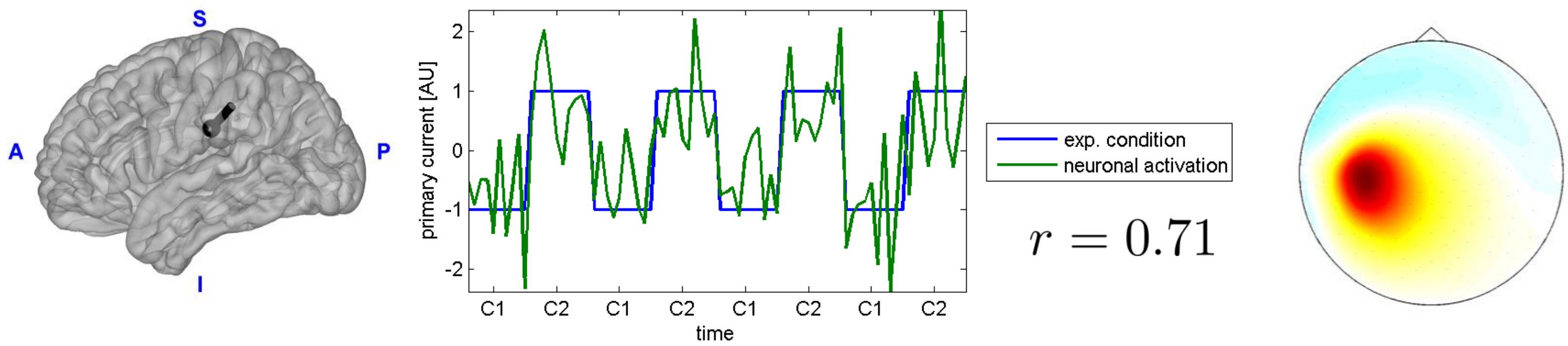
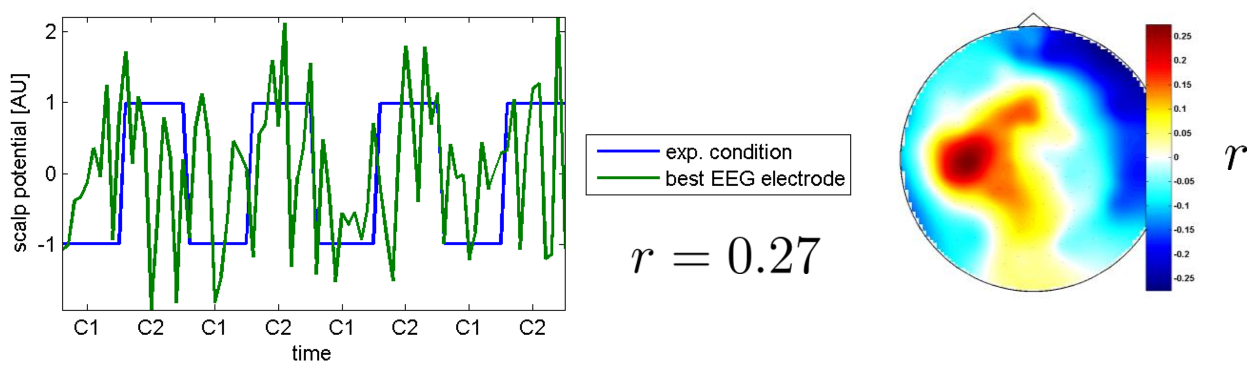


Illustration: sensor-space analysis

Assume there is a brain area modulated by, e.g., the experimental condition.



Due to contributions from other brain areas + noise, we will observe lower correlations and distorted correlation patterns in the EEG.



Model for EEG data

$$\mathbf{x}(t) = \sum_{\mathbf{u}_i \in \mathcal{B}} \mathbf{L}_i \mathbf{j}_i(t) + \boldsymbol{\epsilon}(t) = \mathbf{L}\mathbf{j}(t) + \boldsymbol{\epsilon}(t)$$

$\mathcal{B}$: discretized brain

Model for EEG data

$$\mathbf{x}(t) = \sum_{\mathbf{u}_i \in \mathcal{B}} \mathbf{L}_i \mathbf{j}_i(t) + \boldsymbol{\epsilon}(t) = \mathbf{L}\mathbf{j}(t) + \boldsymbol{\epsilon}(t)$$

$\mathcal{B}$: discretized brain

EEG scalp potential $\mathbf{x}(t) \in \mathbb{R}^M$ at M electrodes is a function of

Model for EEG data

$$\mathbf{x}(t) = \sum_{\mathbf{u}_i \in \mathcal{B}} \mathbf{L}_i \mathbf{j}_i(t) + \boldsymbol{\epsilon}(t) = \mathbf{L}\mathbf{j}(t) + \boldsymbol{\epsilon}(t)$$

$\mathcal{B}$: discretized brain

EEG scalp potential $\mathbf{x}(t) \in \mathbb{R}^M$ at M electrodes is a function of

$\mathbf{j}_i(t) \in \mathbb{R}^3$: primary current at brain location $\mathbf{u}_i$

Model for EEG data

$$\mathbf{x}(t) = \sum_{\mathbf{u}_i \in \mathcal{B}} \mathbf{L}_i \mathbf{j}_i(t) + \boldsymbol{\epsilon}(t) = \mathbf{L}\mathbf{j}(t) + \boldsymbol{\epsilon}(t)$$

$\mathcal{B}$: discretized brain

EEG scalp potential $\mathbf{x}(t) \in \mathbb{R}^M$ at M electrodes is a function of

$\mathbf{j}_i(t) \in \mathbb{R}^3$: primary current at brain location $\mathbf{u}_i$

$\mathbf{L}_i \in \mathbb{R}^{M \times 3}$: mapping describing the propagation of secondary currents to sensors for unit primary currents at $\mathbf{u}_i$

Model for EEG data

$$\mathbf{x}(t) = \sum_{\mathbf{u}_i \in \mathcal{B}} \mathbf{L}_i \mathbf{j}_i(t) + \boldsymbol{\epsilon}(t) = \mathbf{L}\mathbf{j}(t) + \boldsymbol{\epsilon}(t)$$

$\mathcal{B}$: discretized brain

EEG scalp potential $\mathbf{x}(t) \in \mathbb{R}^M$ at M electrodes is a function of

$\mathbf{j}_i(t) \in \mathbb{R}^3$: primary current at brain location $\mathbf{u}_i$

$\mathbf{L}_i \in \mathbb{R}^{M \times 3}$: mapping describing the propagation of secondary currents to sensors for unit primary currents at $\mathbf{u}_i$

$\mathbf{L} \in \mathbb{R}^{M \times 3N}$: **lead field**

Model for EEG data

$$\mathbf{x}(t) = \sum_{\mathbf{u}_i \in \mathcal{B}} \mathbf{L}_i \mathbf{j}_i(t) + \boldsymbol{\epsilon}(t) = \mathbf{L}\mathbf{j}(t) + \boldsymbol{\epsilon}(t)$$

$\mathcal{B}$: discretized brain

EEG scalp potential $\mathbf{x}(t) \in \mathbb{R}^M$ at M electrodes is a function of

$\mathbf{j}_i(t) \in \mathbb{R}^3$: primary current at brain location $\mathbf{u}_i$

$\mathbf{L}_i \in \mathbb{R}^{M \times 3}$: mapping describing the propagation of secondary currents to sensors for unit primary currents at $\mathbf{u}_i$

$\mathbf{L} \in \mathbb{R}^{M \times 3N}$: **lead field** (forward mapping for N brain locations)

$\mathbf{j}(t) \in \mathbb{R}^{3N}$: **primary current density**

Model for EEG data

$$\mathbf{x}(t) = \sum_{\mathbf{u}_i \in \mathcal{B}} \mathbf{L}_i \mathbf{j}_i(t) + \boldsymbol{\epsilon}(t) = \mathbf{L}\mathbf{j}(t) + \boldsymbol{\epsilon}(t)$$

$\mathcal{B}$: discretized brain

EEG scalp potential $\mathbf{x}(t) \in \mathbb{R}^M$ at M electrodes is a function of

$\mathbf{j}_i(t) \in \mathbb{R}^3$: primary current at brain location $\mathbf{u}_i$

$\mathbf{L}_i \in \mathbb{R}^{M \times 3}$: mapping describing the propagation of secondary currents to sensors for unit primary currents at $\mathbf{u}_i$

$\mathbf{L} \in \mathbb{R}^{M \times 3N}$: **lead field** (forward mapping for N brain locations)

$\mathbf{j}(t) \in \mathbb{R}^{3N}$: **primary current density**

$\boldsymbol{\epsilon}(t) \in \mathbb{R}^M$: electrical activity of no interest (sensor noise, artifacts)

The lead field (forward mapping)

$\mathbf{L}$ depends on

- Conductivities of brain/skull/skin etc.
- Head geometry obtained from structural MRI
- Electrode positions (3D scanner)

Figure from Litvak et al., 2011

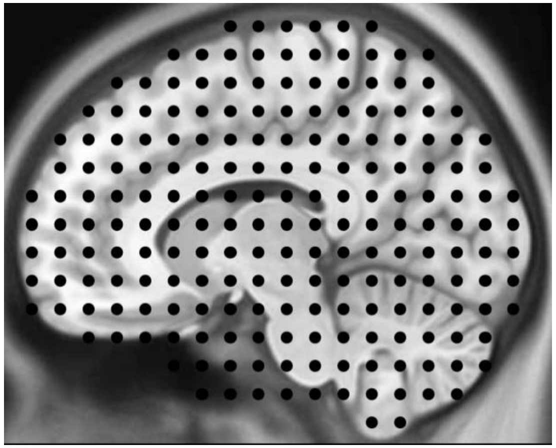
The lead field (forward mapping)

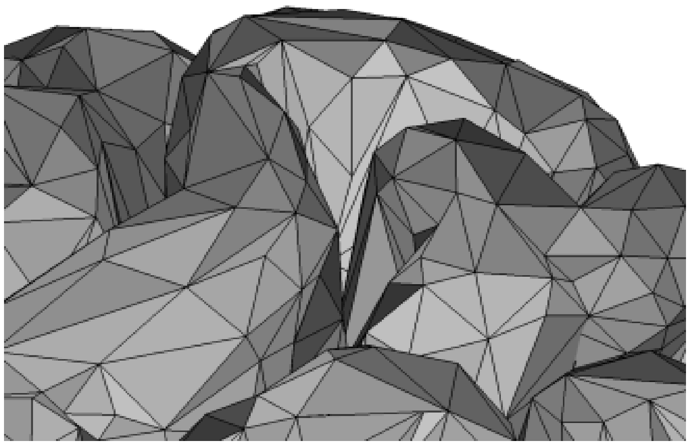
$\mathbf{L}$ depends on

- Conductivities of brain/skull/skin etc.
- Head geometry obtained from structural MRI
- Electrode positions (3D scanner)

Figure from Litvak et al., 2011

$\mathbf{L}$ is evaluated at $N \gg M$ brain locations $\mathbf{u}_i$ in 3D volume or on cortical surface.

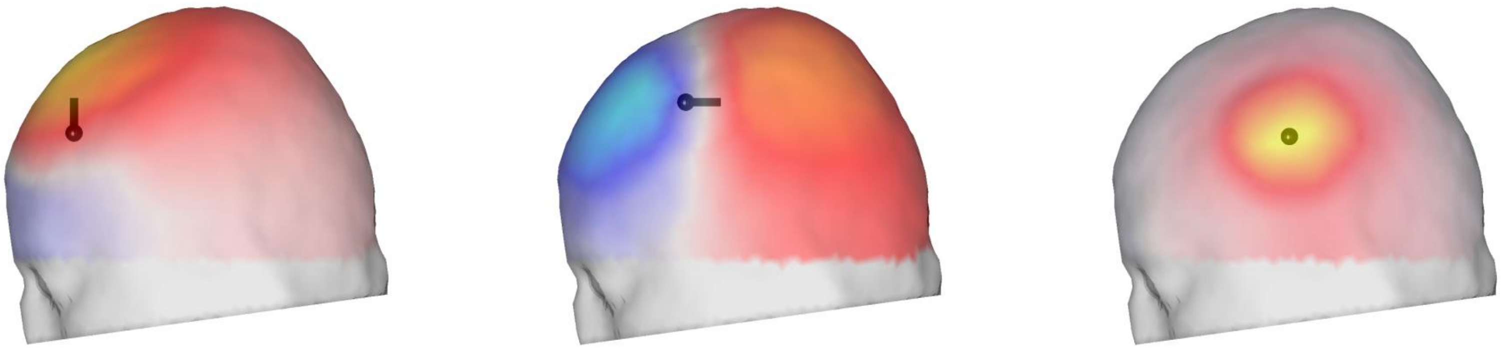




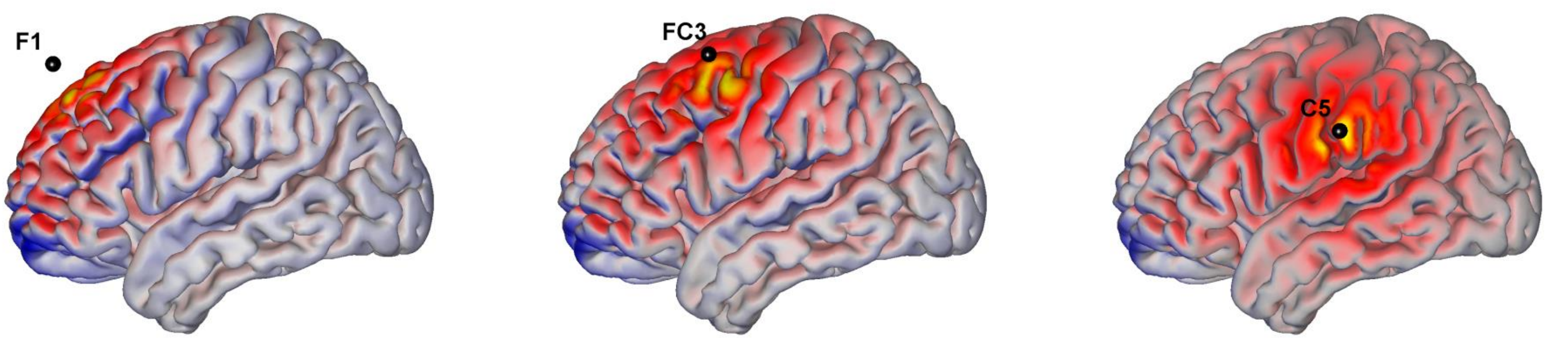
The lead field (forward mapping)

$\mathbf{L} \in \mathbb{R}^{M \times 3N}$, pre-calculated

Columns :



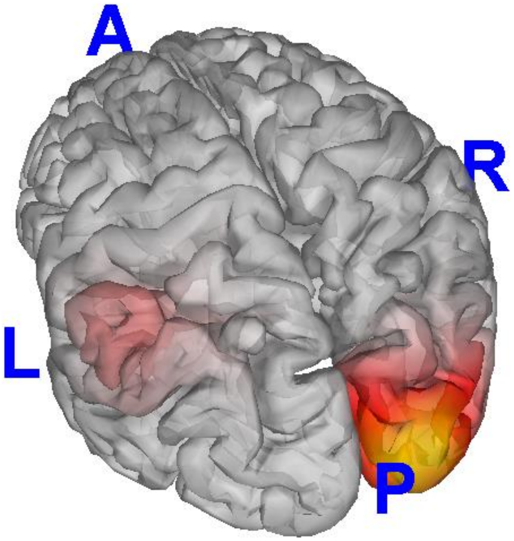
Rows:



The current density

$\mathbf{j}(t) \in \mathbb{R}^{3N}$, to be estimated from $\mathbf{L}$ and $\mathbf{x}(t)$

$t = 1$

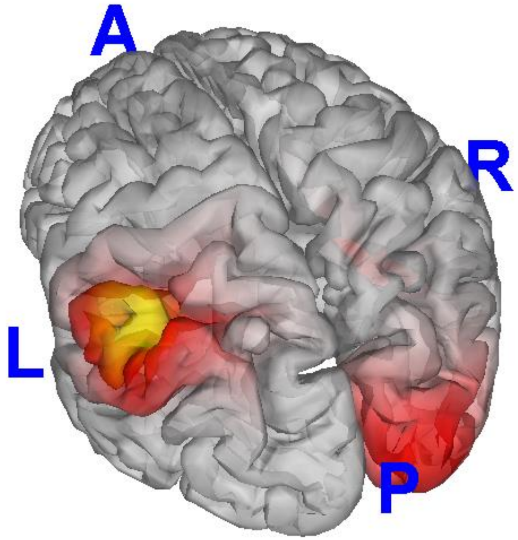


Vectorfield, plotting magnitudes here.

The current density

$\mathbf{j}(t) \in \mathbb{R}^{3N}$, to be estimated from $\mathbf{L}$ and $\mathbf{x}(t)$

$t = 2$

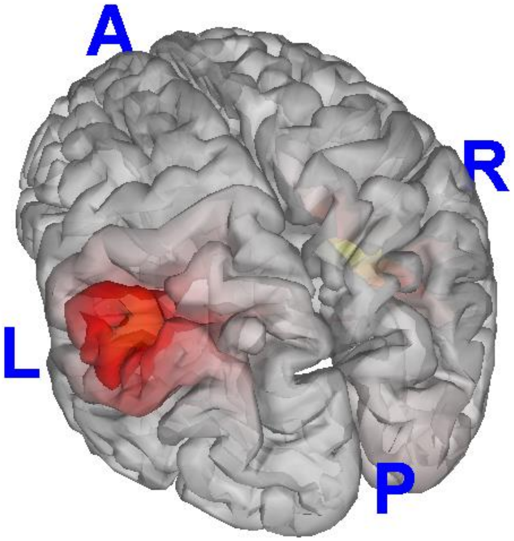


Vectorfield, plotting magnitudes here.

The current density

$\mathbf{j}(t) \in \mathbb{R}^{3N}$, to be estimated from $\mathbf{L}$ and $\mathbf{x}(t)$

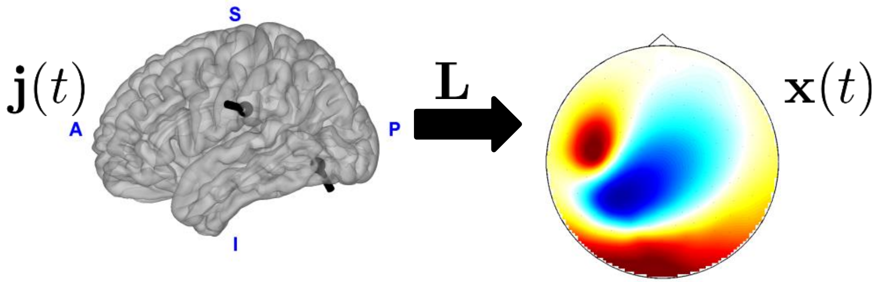
$t = 3$



Vectorfield, plotting magnitudes here.

The Inverse Problem

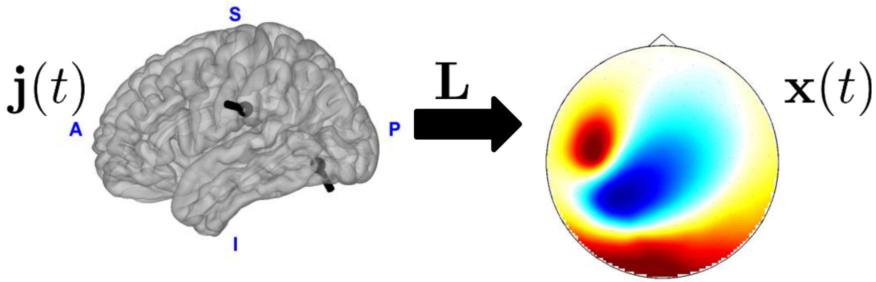
$$\mathbf{x}(t) = \mathbf{L}\mathbf{j}(t) + \boldsymbol{\epsilon}(t)$$



We would like to invert the mapping $\mathbf{L}$ to obtain the current sources $\mathbf{j}(t)$.

The Inverse Problem

$$\mathbf{x}(t) = \mathbf{L}\mathbf{j}(t) + \boldsymbol{\epsilon}(t)$$



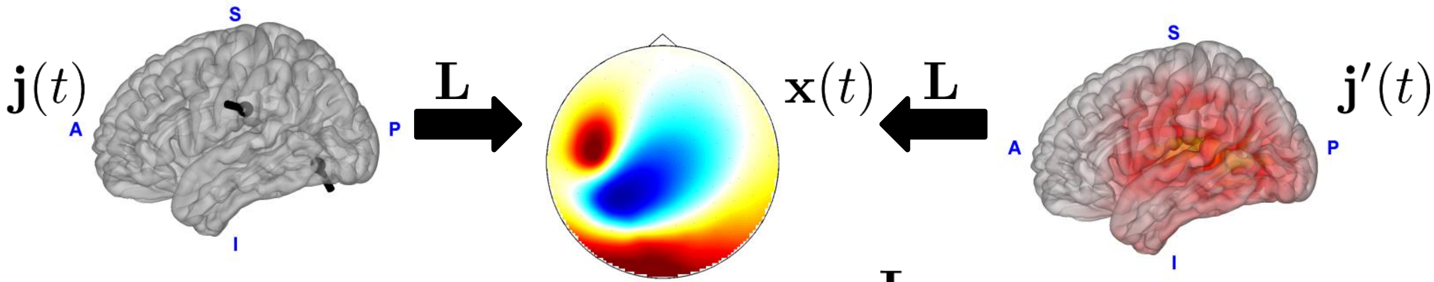
We would like to invert the mapping $\mathbf{L}$ to obtain the current sources $\mathbf{j}(t)$.

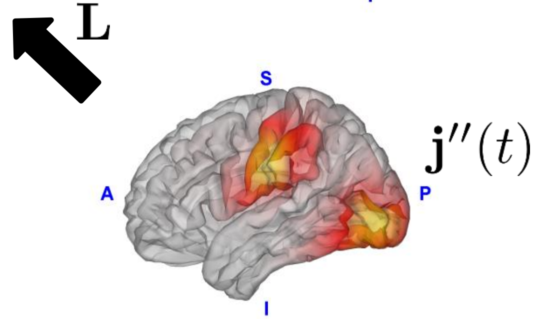
Potential benefits:

- Increase in SNR
- Localization/interpretation

The Inverse Problem

$$\mathbf{x}(t) = \mathbf{L}\mathbf{j}(t) + \boldsymbol{\epsilon}(t)$$

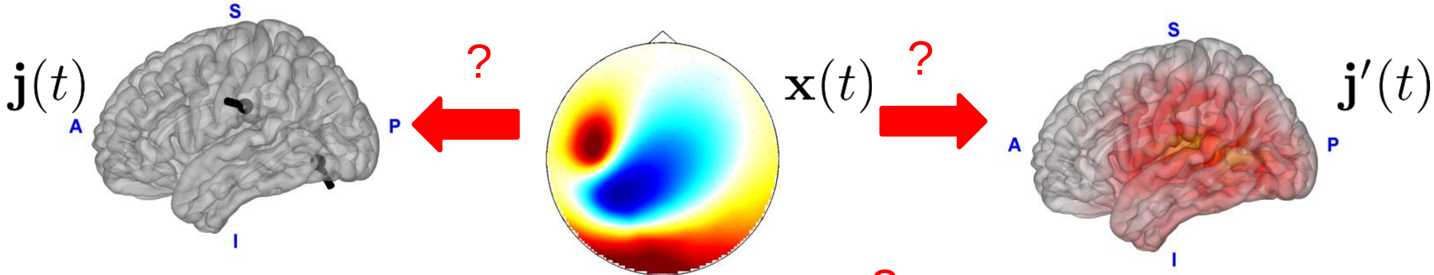


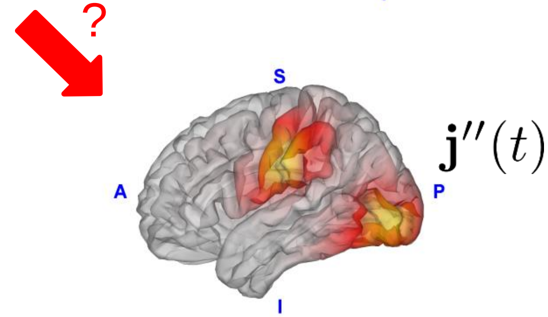


However, the inverse problem (estimating $\mathbf{j}$ from $\mathbf{x}$)
has infinitely many solutions.

The Inverse Problem

$$\mathbf{x}(t) = \mathbf{L}\mathbf{j}(t) + \boldsymbol{\epsilon}(t)$$

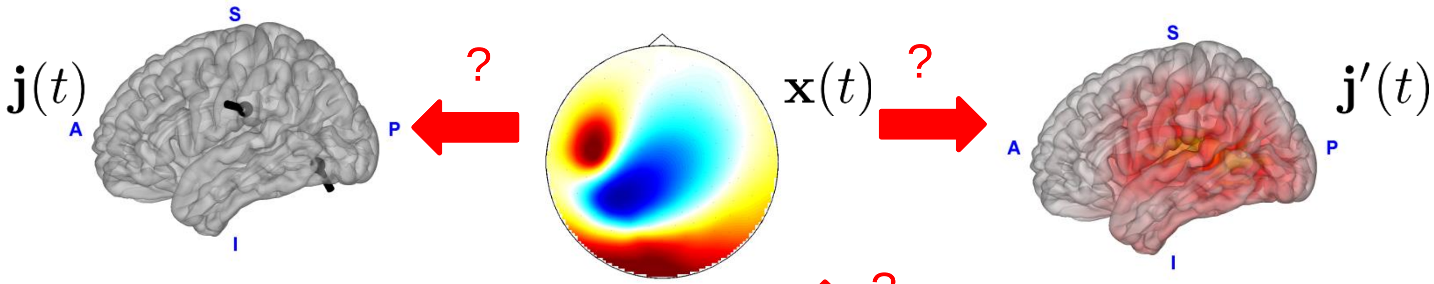


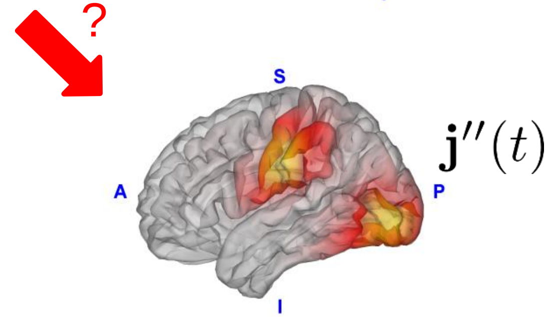


However, the inverse problem (estimating $\mathbf{j}$ from $\mathbf{x}$) has infinitely many solutions.

The Inverse Problem

$$\mathbf{x}(t) = \mathbf{L}\mathbf{j}(t) + \boldsymbol{\epsilon}(t)$$





However, the inverse problem
(estimating $\mathbf{j}$ from $\mathbf{x}$)
has infinitely many solutions.

Solving the inverse problem = selecting the sources that best match prior expectations (assumptions), while explaining the data.

Inverse methods

MNE
MCE
WMNE
Loreta
sLORETA
eLORETA
Laura
Electra
WROP
S-FLEX
Champagne

Aquavit
DICS
LCMV Beamformer
Nulling Beamformer
FOCUSS
Minimum Entropy
Dipole Modeling
Multipole Modeling
MUSIC/RAP-MUSIC
DCM

Every method performs well if its specific assumptions are met.

No method can perform well in all realistic situations.

Maximum-likelihood and maximum a-posteriori estimation

$$\mathbf{x}(t) = \mathbf{L}\mathbf{j}(t) + \boldsymbol{\epsilon}(t)$$

Assuming the noise $\boldsymbol{\epsilon}(t)$ is Gaussian distributed with covariance $\mathbf{Q}$, the maximum-likelihood approach to estimating the source current density is

$$\hat{\mathbf{j}}_{\mathrm{ML}}(t) = \arg\min_{\mathbf{j}(t)} l(\mathbf{j}(t))$$ with $$l(\mathbf{j}(t)) = \|\mathbf{x}(t) - \mathbf{L}\mathbf{j}(t)\|^2_{\mathbf{Q}^{-1}} \ .$$

Maximum-likelihood and maximum a-posteriori estimation

$$\mathbf{x}(t) = \mathbf{L}\mathbf{j}(t) + \boldsymbol{\epsilon}(t)$$

Assuming the noise $\boldsymbol{\epsilon}(t)$ is Gaussian distributed with covariance $\mathbf{Q}$,
the maximum-likelihood approach to estimating the source current density is

$$\hat{\mathbf{j}}_{\mathrm{ML}}(t) = \arg\min_{\mathbf{j}(t)} l(\mathbf{j}(t))$$ with $$l(\mathbf{j}(t)) = \|\mathbf{x}(t) - \mathbf{L}\mathbf{j}(t)\|^2_{\mathbf{Q}^{-1}} \ .$$

However, since $N \gg M$ (the system $\mathbf{x} = \mathbf{L}\mathbf{j}$ is underdetermined),

$l(\mathbf{j}(t))$ is zero for infinitely many choices of $\mathbf{j}(t)$.

Maximum-likelihood and maximum a-posteriori estimation

$$\mathbf{x}(t) = \mathbf{L}\mathbf{j}(t) + \boldsymbol{\epsilon}(t)$$

Assuming the noise $\boldsymbol{\epsilon}(t)$ is Gaussian distributed with covariance $\mathbf{Q}$,
the maximum-likelihood approach to estimating the source current density is

$$\hat{\mathbf{j}}_{\mathrm{ML}}(t) = \arg\min_{\mathbf{j}(t)} l(\mathbf{j}(t)) \qquad \text{with} \qquad l(\mathbf{j}(t)) = \|\mathbf{x}(t) - \mathbf{L}\mathbf{j}(t)\|^2_{\mathbf{Q}^{-1}} \ .$$

However, since $N \gg M$ (the system $\mathbf{x} = \mathbf{L}\mathbf{j}$ is underdetermined),

$l(\mathbf{j}(t))$ is zero for infinitely many choices of $\mathbf{j}(t)$.

Need to impose additional penalty/constraint $g(\mathbf{j}(t))$ on the sources.

Maximum-a-posteriori estimate: $\hat{\mathbf{j}}_{\mathrm{MAP}}(t) = \arg\min_{\mathbf{j}(t)} l(\mathbf{j}(t)) + \lambda g(\mathbf{j}(t))$

Maximum-likelihood and maximum a-posteriori estimation

$$\mathbf{x}(t) = \mathbf{L}\mathbf{j}(t) + \boldsymbol{\epsilon}(t)$$

Assuming the noise $\boldsymbol{\epsilon}(t)$ is Gaussian distributed with covariance $\mathbf{Q}$, the maximum-likelihood approach to estimating the source current density is

$$\hat{\mathbf{j}}_{\mathrm{ML}}(t) = \arg\min_{\mathbf{j}(t)} l(\mathbf{j}(t))$$

with

$$\boxed{l(\mathbf{j}(t)) = \|\mathbf{x}(t) - \mathbf{L}\mathbf{j}(t)\|^2_{\mathbf{Q}^{-1}} \; .}$$

However, since $N \gg M$ (the system $\mathbf{x} = \mathbf{L}\mathbf{j}$ is underdetermined),

$l(\mathbf{j}(t))$ is zero for infinitely many choices of $\mathbf{j}(t)$.

Need to impose additional penalty/constraint $g(\mathbf{j}(t))$ on the sources.

Maximum-a-posteriori estimate: $$\boxed{\hat{\mathbf{j}}_{\mathrm{MAP}}(t) = \arg\min_{\mathbf{j}(t)} l(\mathbf{j}(t)) + \lambda g(\mathbf{j}(t))}$$

Spatial constraints

Since $\mathbf{L}$ links source activity to brain locations, constraints on the spatial structure of the current density can be imposed.

Spatial constraints: smoothness

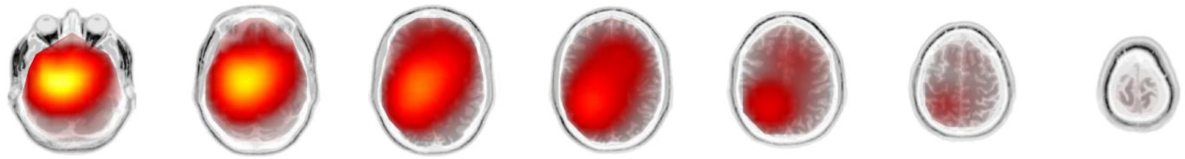
Since $\mathbf{L}$ links source activity to brain locations, constraints on the spatial structure of the current density can be imposed.

Smoothness

- Assumption: neighboring voxels show similar activity
- Examples: (weighted) minimum norm estimate, LORETA

[Jeffs et al., 1987; Pascual-Marqui et al., 1994]

LORETA



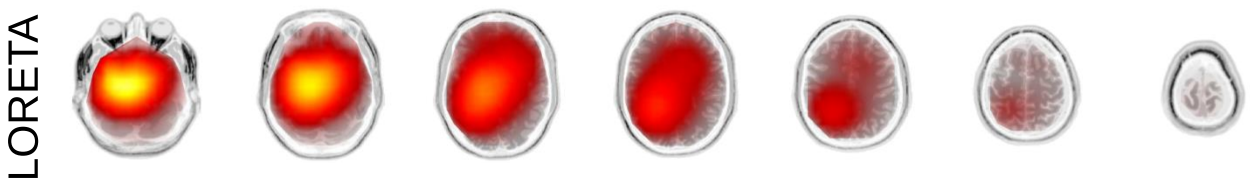
Spatial constraints: smoothness

Since $\mathbf{L}$ links source activity to brain locations, constraints on the spatial structure of the current density can be imposed.

Smoothness

- Assumption: neighboring voxels show similar activity
- Examples: (weighted) minimum norm estimate, LORETA

[Jeffs et al., 1987; Pascual-Marqui et al., 1994]



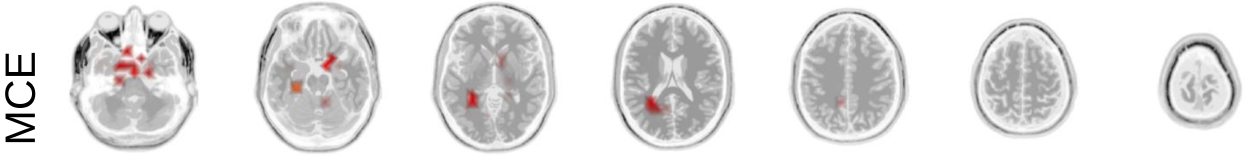
- Technically: L_2-norm penalty $g(\mathbf{j}(t)) = \|\mathbf{\Gamma}\mathbf{j}(t)\|_2^2$
- Solution linear in data: $\hat{\mathbf{j}}(t) = \underbrace{\left(\mathbf{L}^\top\mathbf{L} + \lambda\mathbf{\Gamma}^\top\mathbf{\Gamma}\right)^{-1}\mathbf{L}^\top}_{\mathbf{P}}\mathbf{x}(t)$
- $\mathbf{P}$ is precomputable → very efficient

Spatial constraints: sparsity

Sparsity

- Assumption: only a small part of the brain is active during task
- E.g., minimum current estimate (MCE), FOCUSS

[Matsuura et al., 1995; Gorodnitsky et al., 1995]



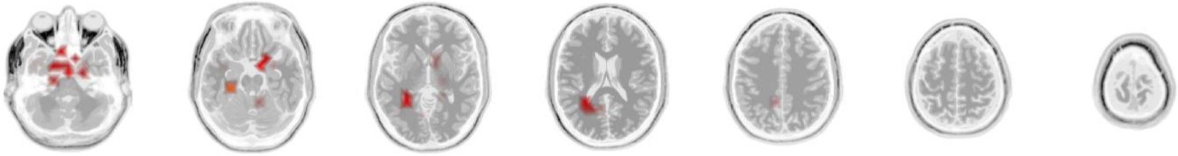
Spatial constraints: sparsity

Sparsity

- Assumption: only a small part of the brain is active during task
- E.g., minimum current estimate (MCE), FOCUSS

[Matsuura et al., 1995; Gorodnitsky et al., 1995]

MCE



- Technically: L_1-norm $g(\mathbf{j}(t)) = \|\mathbf{j}(t)\|_1$ leads to sparsity
- Solution nonlinear in data, iterative optimization required

Limitations of smooth and sparse inverses

Smooth inverses

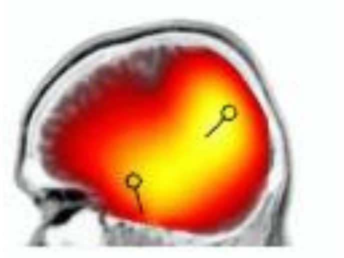
- Difficulty to distinguish sources

- Occurence of „ghost sources“

Limitations of smooth and sparse inverses

Smooth inverses

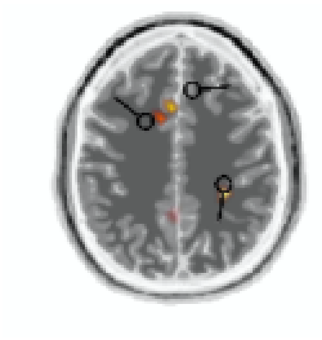
- Difficulty to distinguish sources



- Occurence of „ghost sources“

Sparse inverses

- Scattered sources in the presence of noise



Combining sparsity and smoothness

1. **Mixed-norm penalties**, e.g., $g(\mathbf{j}) = \|\mathbf{j}(t)\|_1 + \gamma\|\mathbf{\Gamma}\mathbf{j}(t)\|_2$

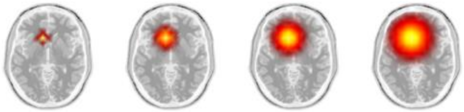
[Haufe et al., 2008; Vega-Hernández et al., 2008]

Combining sparsity and smoothness

1. **Mixed-norm penalties**, e.g., $g(\mathbf{j}) = \|\mathbf{j}(t)\|_1 + \gamma\|\mathbf{\Gamma}\mathbf{j}(t)\|_2$

[Haufe et al., 2008; Vega-Hernández et al., 2008]

2. **Sparsity in different spatial basis**

E.g. $g(\mathbf{j}(t)) = \|\tilde{\mathbf{j}}_s(t)\|_1$ with $\mathbf{j} = \mathbf{\Phi}_s\tilde{\mathbf{j}}_s$ and $\mathbf{\Phi}_s =$ 

→ Solution has simple spatial structure

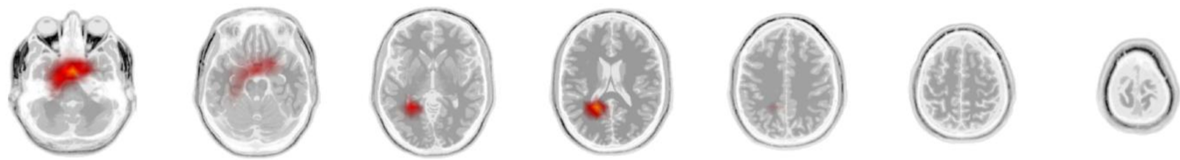
[Haufe et al., 2011]

Combining sparsity and smoothness

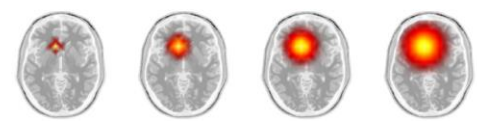
1. **Mixed-norm penalties**, e.g., $g(\mathbf{j}) = \|\mathbf{j}(t)\|_1 + \gamma\|\mathbf{\Gamma}\mathbf{j}(t)\|_2$

[Haufe et al., 2008; Vega-Hernández et al., 2008]

FVR



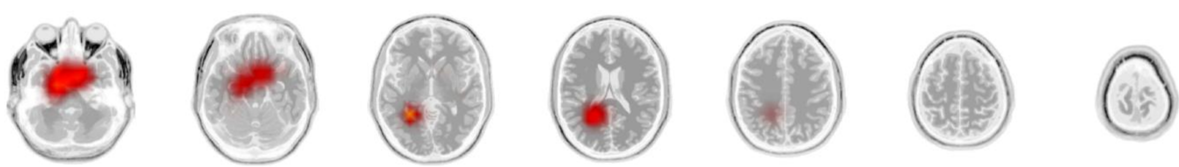
2. **Sparsity in different spatial basis**

E.g. $g(\mathbf{j}(t)) = \|\tilde{\mathbf{j}}_s(t)\|_1$ with $\mathbf{j} = \mathbf{\Phi}_s\tilde{\mathbf{j}}_s$ and $\mathbf{\Phi}_s =$ 

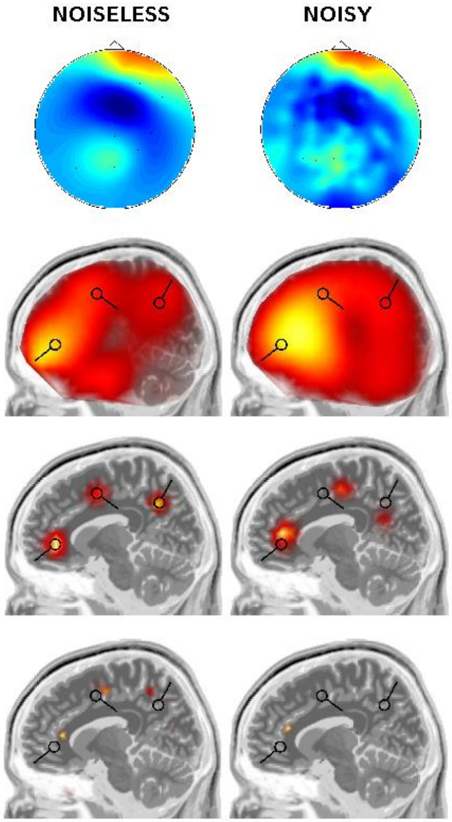
→ Solution has simple spatial structure

[Haufe et al., 2011]

S-FLEX



Comparison



Smooth

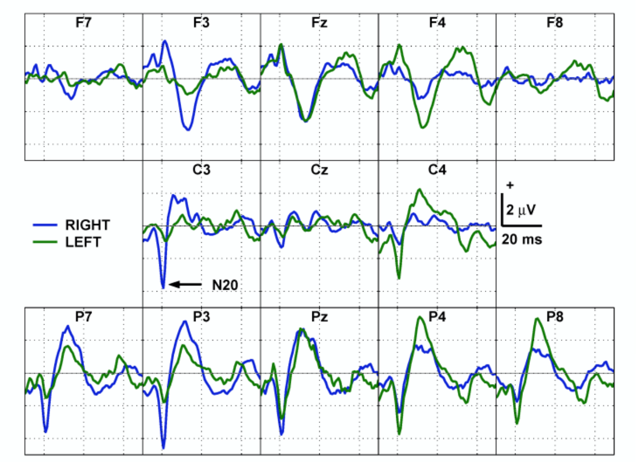
Both

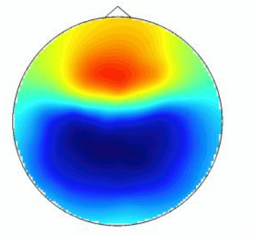
Sparse

Localization of hand areas in somatosensory cortex

Electrical stimulation at both thumbs
(Median nerves)

→ N20 event-related potential in the EEG





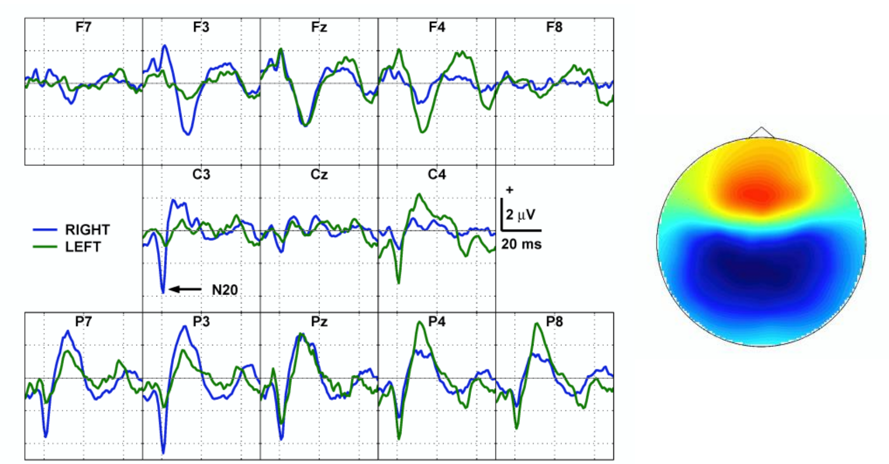
There should be two lateralized symmetric sources in the somatosensory cortices.

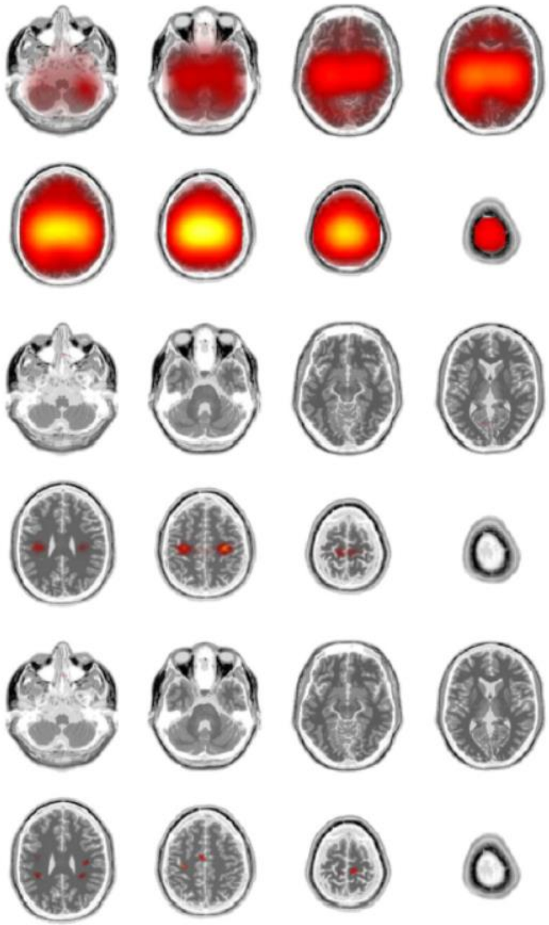
[Haufe et al., 2008]

Localization of hand areas in somatosensory cortex

Electrical stimulation at both thumbs
(Median nerves)

→ N20 event-related potential in the EEG





There should be two lateralized symmetric sources in the somatosensory cortices.

[Haufe et al., 2008]

Technicalities

- Compensating for bias towards superficial sources
- Fixing current orientations in cortically-constrained estimation
- Measuring distances on the cortical manifold
- Achieving sparsity for vectorial currents
- Dealing with time series data

Reconstruction of time series

Problem with L_1-norm penalties: sparsity pattern may differ for each sample, causing jumps in the source time series between voxels.

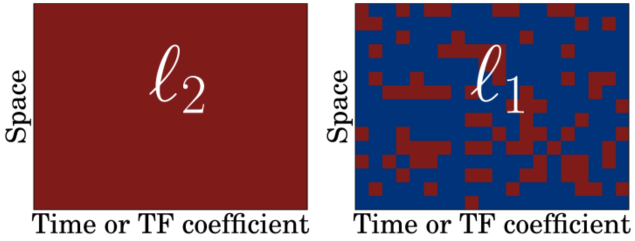


Figure from Gramfort et al., 2013

Reconstruction of time series

Problem with L_1-norm penalties: sparsity pattern may differ for each sample, causing jumps in the source time series between voxels.

Remedy for stationary time series: select the same set of active voxels/basis functions for all samples.

$$g\left(\tilde{\mathbf{j}}(1), \ldots, \tilde{\mathbf{j}}(T)\right) = \sum_i \left\| \left(\tilde{\mathbf{j}}_i^\top(1), \ldots, \tilde{\mathbf{j}}_i^\top(T)\right)^\top \right\|_2 = \left\|\tilde{\mathbf{J}}\right\|_{21}$$

[Haufe et al., 2008; Ou et al., 2009]

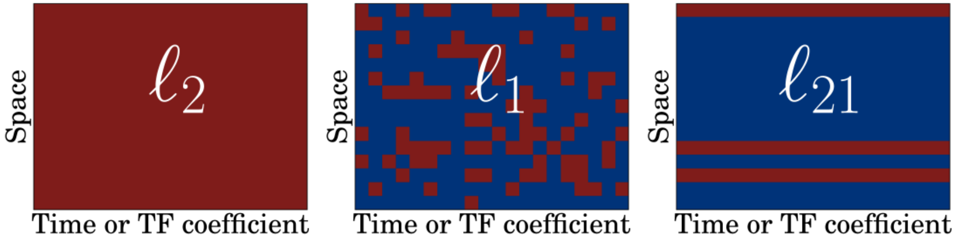


Figure from Gramfort et al., 2013

Reconstruction of time series

To model nonstationarity:

- Decompose time series into time-frequency atoms

$$\mathbf{j} = \tilde{\mathbf{j}}_t \boldsymbol{\Phi}_t \qquad \boldsymbol{\Phi}_t =$$

- Mixed-norm penalty $g(\mathbf{J}) = \|\tilde{\mathbf{J}}_t\|_{21} + \gamma \|\tilde{\mathbf{J}}_t\|_1$

[Gramfort et al., 2013]

Reconstruction of time series

To model nonstationarity:

- Decompose time series into time-frequency atoms

$$\mathbf{j} = \tilde{\mathbf{j}}_t \mathbf{\Phi}_t \qquad \mathbf{\Phi}_t =$$

- Mixed-norm penalty $g(\mathbf{J}) = \|\tilde{\mathbf{J}}_t\|_{21} + \gamma \|\tilde{\mathbf{J}}_t\|_1$

[Gramfort et al., 2013]

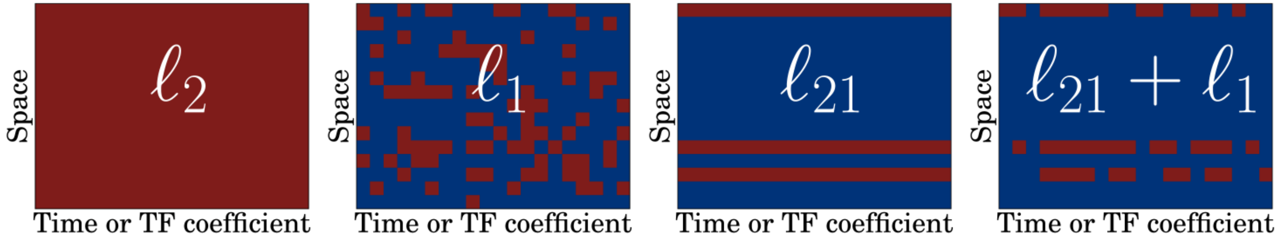


Figure from Gramfort et al., 2013

Reconstruction of time series

To model nonstationarity:

- Decompose time series into time-frequency atoms

$$\mathbf{j} = \tilde{\mathbf{j}}_t \boldsymbol{\Phi}_t \qquad \boldsymbol{\Phi}_t =$$

- Mixed-norm penalty $g(\mathbf{J}) = \|\tilde{\mathbf{J}}_t\|_{21} + \gamma \|\tilde{\mathbf{J}}_t\|_1$

[Gramfort et al., 2013]

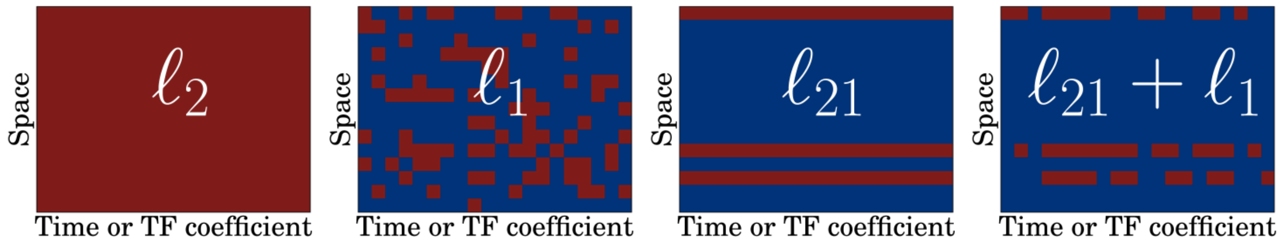


Figure from Gramfort et al., 2013

Other dynamical constraints: Random walk model, Kalman filter, …

[Schmitt et al., 2002; Galka et al., 2004]

Other source localization paradigms

Dipole fits: instead of estimating currents of $N \gg M$ dipoles with fixed locations, estimate current+location of $K \ll M$ dipoles.

Nonconvex cost function, danger of local minima.

[e.g., Scherg, 1992]

Other source localization paradigms

Dipole fits: instead of estimating currents of $N \gg M$ dipoles with fixed locations, estimate current+location of $K \ll M$ dipoles.

Nonconvex cost function, danger of local minima.

[e.g., Scherg, 1992]

Scanning Techniques:

- **Subspace methods** (MUSIC, RapMUSIC): for each voxel, compute angle between space spanned by dipole at that voxel and space spanned by data. The angle is taken as an index of activation at that voxel.

[Schmitt, 1986; Mosher and Leahy, 1999]

Other source localization paradigms

Dipole fits: instead of estimating currents of $N \gg M$ dipoles with fixed locations, estimate current+location of $K \ll M$ dipoles.

Nonconvex cost function, danger of local minima.

[e.g., Scherg, 1992]

Scanning Techniques:

- **Subspace methods** (MUSIC, RapMUSIC): for each voxel, compute angle between space spanned by dipole at that voxel and space spanned by data. The angle is taken as an index of activation at that voxel.

[Schmitt, 1986; Mosher and Leahy, 1999]

- **Beamformers:** for each voxel, find a spatial filter which maximizes the SNR of signals originating at that voxel. The SNR at each voxel is taken as an activity index.

[van Veen et al., 1997]

Activity indices of scanning techniques do not explain the data.

(Blind) source separation

If temporal constraints are available, one might drop spatial constraints.

→ Useful if no accurate leadfield (e.g., no individual structural MRI) exists.

(Blind) source separation

If temporal constraints are available, one might drop spatial constraints.
→ Useful if no accurate leadfield (e.g., no individual structural MRI) exists.

Factorize current density into $\mathbf{j}(t) = \mathbf{F}\mathbf{s}(t) + \boldsymbol{\epsilon}_{\mathbf{j}}(t)$, where

$\mathbf{s}(t) \in \mathbb{R}^K$ are $K \leq M \ll N$ **latent factors** (sources, components) of brain activity with specific temporal dynamics, and

$\mathbf{F} \in \mathbb{R}^{3N \times K}$ are their corresponding **source space activation patterns**.

(Blind) source separation

If temporal constraints are available, one might drop spatial constraints.
→ Useful if no accurate leadfield (e.g., no individual structural MRI) exists.

Factorize current density into $\mathbf{j}(t) = \mathbf{F}\mathbf{s}(t) + \boldsymbol{\epsilon}_{\mathbf{j}}(t)$, where

$\mathbf{s}(t) \in \mathbb{R}^K$ are $K \leq M \ll N$ **latent factors** (sources, components) of brain activity with specific temporal dynamics, and

$\mathbf{F} \in \mathbb{R}^{3N \times K}$ are their corresponding **source space activation patterns**.

The overall decomposition of the EEG becomes

$$\mathbf{x}(t) = \underbrace{\mathbf{L}\mathbf{F}}_{\mathbf{A}}\mathbf{s}(t) + \underbrace{\mathbf{L}\boldsymbol{\epsilon}_{\mathbf{j}}(t) + \boldsymbol{\epsilon}(t)}_{\boldsymbol{\varepsilon}(t)} = \mathbf{A}\mathbf{s}(t) + \boldsymbol{\varepsilon}(t)$$, where

$\mathbf{A} \in \mathbb{R}^{M \times K}$ are **sensor-space activation patterns** to be estimated.

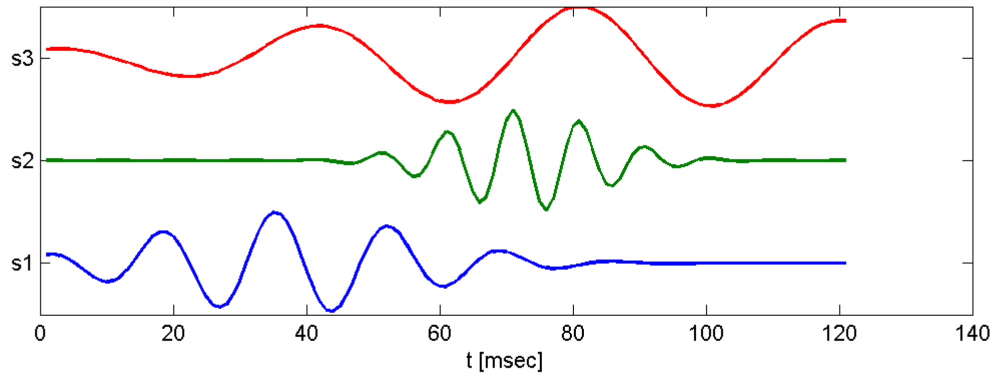
The factor/component time series

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \boldsymbol{\varepsilon}(t)$$

The factor/component time series

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \boldsymbol{\varepsilon}(t)$$

$\mathbf{s}(t) \in \mathbb{R}^K$, to be estimated



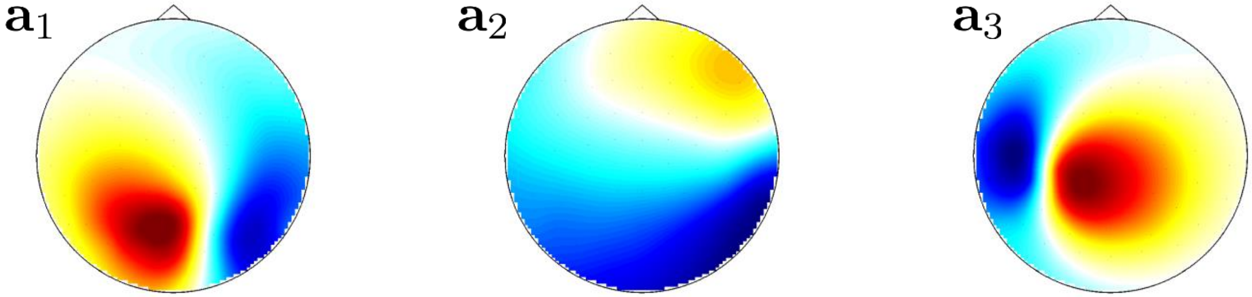
Each $s_k(t)$ is linked to a static sensor-space activation pattern $\mathbf{a}_k$ rather than to a brain location.

The sensor space activation patterns

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \boldsymbol{\varepsilon}(t)$$

$\mathbf{A} \in \mathbb{R}^{M \times K}$, also to be estimated

Columns:



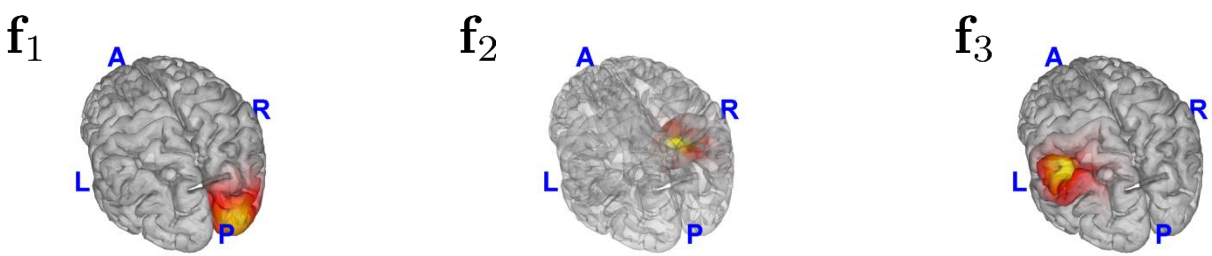
The activation patterns $\mathbf{a}_k$ represent the time invariant current density of the component $s_k(t)$.

Source localization of activation patterns

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \boldsymbol{\varepsilon}(t) = \mathbf{L}\mathbf{F}\mathbf{s}(t) + \boldsymbol{\varepsilon}(t)$$

Recall that $\mathbf{a}_k = \mathbf{L}\mathbf{f}_k$.

→ Using the techniques described in the first part, the estimated $\mathbf{a}_k$ can be source-localized by estimating $\mathbf{f}_k$ using a precomputed leadfield $\mathbf{L}$.



The source space activation patterns $\mathbf{f}_k$ represent the time invariant current density of the component $s_k(t)$.

Forward and backward models

A BSS method may either directly fit the **forward model** $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \boldsymbol{\varepsilon}(t)$

(that is, estimate $\mathbf{A}$ and $\mathbf{s}(t)$ jointly),

Forward and backward models

A BSS method may either directly fit the **forward model** $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \boldsymbol{\varepsilon}(t)$

(that is, estimate $\mathbf{A}$ and $\mathbf{s}(t)$ jointly),

or fit a **backward model** $\mathbf{W}^\top \mathbf{x}(t) = \mathbf{s}(t)$

parameterized only by the **extraction filters** $\mathbf{W} \in \mathbb{R}^{M \times K}$.

Forward and backward models

A BSS method may either directly fit the **forward model** $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \boldsymbol{\varepsilon}(t)$

(that is, estimate $\mathbf{A}$ and $\mathbf{s}(t)$ jointly),

or fit a **backward model** $\mathbf{W}^\top \mathbf{x}(t) = \mathbf{s}(t)$

parameterized only by the **extraction filters** $\mathbf{W} \in \mathbb{R}^{M \times K}$.

If $\boldsymbol{\varepsilon}(t)$ and $\mathbf{s}(t)$ are uncorrelated, both approaches are equivalent,

and related through $\mathbf{A} = \boldsymbol{\Sigma}_{\mathbf{x}} \mathbf{W} \boldsymbol{\Sigma}_{\mathbf{s}}^{-1}$, [Parra et al., 2005; Haufe et al., 2014]

where $\boldsymbol{\Sigma}_{\mathbf{x}}$ and $\boldsymbol{\Sigma}_{\mathbf{s}}$ are the covariance matrices of $\mathbf{x}(t)$ and $\mathbf{s}(t)$.

Forward and backward models

A BSS method may either directly fit the **forward model** $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \boldsymbol{\varepsilon}(t)$

(that is, estimate $\mathbf{A}$ and $\mathbf{s}(t)$ jointly),

or fit a **backward model** $\mathbf{W}^\top \mathbf{x}(t) = \mathbf{s}(t)$

parameterized only by the **extraction filters** $\mathbf{W} \in \mathbb{R}^{M \times K}$.

If $\boldsymbol{\varepsilon}(t)$ and $\mathbf{s}(t)$ are uncorrelated, both approaches are equivalent,

and related through $\mathbf{A} = \boldsymbol{\Sigma}_{\mathbf{x}} \mathbf{W} \boldsymbol{\Sigma}_{\mathbf{s}}^{-1}$, [Parra et al., 2005; Haufe et al., 2014]

where $\boldsymbol{\Sigma}_{\mathbf{x}}$ and $\boldsymbol{\Sigma}_{\mathbf{s}}$ are the covariance matrices of $\mathbf{x}(t)$ and $\mathbf{s}(t)$.

Both forward and backward models provide solutions of the inverse problem, as long as $\mathbf{x}(t)$ is "raw" (not nonlinearly preprocessed) EEG data.

Parameter interpretation

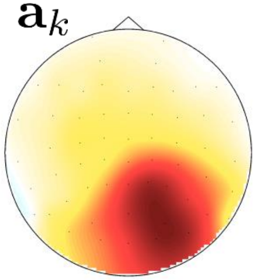
Both filters and patterns can be visualized as scalp maps. However, their meanings are completely different.

[Haufe et al., 2014]

Parameter interpretation

Both filters and patterns can be visualized as scalp maps. However, their meanings are completely different.

$\mathbf{a}_k$



Patterns tell us how brain activity $s_k(t)$ is expressed in each sensor.

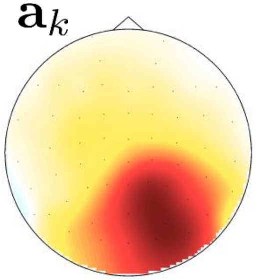
→ $\mathbf{a}_k$ depends only on $s_k(t)$.

[Haufe et al., 2014]

Parameter interpretation

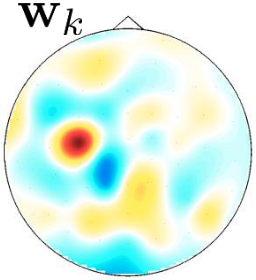
Both filters and patterns can be visualized as scalp maps. However, their meanings are completely different.

$\mathbf{a}_k$



Patterns tell us how brain activity $s_k(t)$ is expressed in each sensor.

$\mathbf{w}_k$



Filters tell us how to weight sensors to extract the brain activity $s_k(t)$.

→ $\mathbf{a}_k$ depends only on $s_k(t)$.

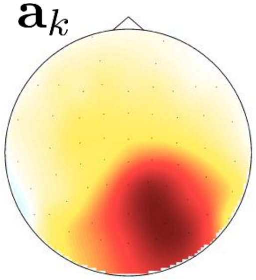
→ $\mathbf{w}_k$ depends on $s_k(t)$ and all noise sources.

[Haufe et al., 2014]

Parameter interpretation

Both filters and patterns can be visualized as scalp maps. However, their meanings are completely different.

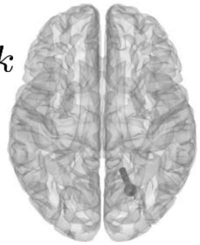
$\mathbf{a}_k$



Patterns tell us how brain activity $s_k(t)$ is expressed in each sensor.

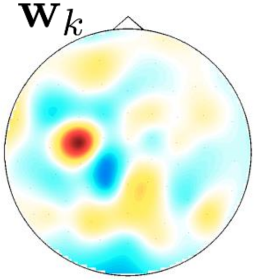
→ $\mathbf{a}_k$ depends only on $s_k(t)$.

$\mathbf{f}_k$



Only patterns can be source localized by virtue of $\mathbf{a}_k = \mathbf{L}\mathbf{f}_k$.

$\mathbf{w}_k$



Filters tell us how to weight sensors to extract the brain activity $s_k(t)$.

→ $\mathbf{w}_k$ depends on $s_k(t)$ and all noise sources.

[Haufe et al., 2014]

BSS methods

For model fitting, a backward modeling approach is typically adopted,

$$\boxed{\mathbf{W} = \arg\min_{\mathbf{W}'} f\left(\mathbf{W}'^{\top}\mathbf{x}(t)\right)}$$

where f encodes assumptions on the sources $\mathbf{s}(t) = \mathbf{W}^{\top}\mathbf{x}(t)$.

BSS methods

For model fitting, a backward modeling approach is typically adopted,

$$\boxed{\mathbf{W} = \arg\min_{\mathbf{W}'} f\left(\mathbf{W}'^{\top}\mathbf{x}(t)\right)}$$

where f encodes assumptions on the sources $\mathbf{s}(t) = \mathbf{W}^{\top}\mathbf{x}(t)$.

PCA
ICA
TDSEP
xDAWN
CCA
CSP
SPoC
cSPoC
SSD
DSS

LDA
SVM
LLR
SSA
SCSA
MVARICA
CICAAR
PISA
MOCA

BSS methods by assumption

Brain activity differs between experimental conditions.

(ERP studies)

→ Linear classifiers (LDA, SVM, LLR)

[e.g., Blankertz et al., 2010]

BSS methods by assumption

Brain activity differs between experimental conditions.

(ERP studies)

→ Linear classifiers (LDA, SVM, LLR)

[e.g., Blankertz et al., 2010]

Brain activity correlates with behaviour or stimulus variables.

(ERP studies)

→ Linear regression (OLS, Ridge regression, LASSO)

[e.g., Parra et al., 2005]

BSS methods by assumption

Brain activity differs between experimental conditions.

(ERP studies)

→ Linear classifiers (LDA, SVM, LLR)

[e.g., Blankertz et al., 2010]

Brain activity correlates with behaviour or stimulus variables.

(ERP studies)

→ Linear regression (OLS, Ridge regression, LASSO)

[e.g., Parra et al., 2005]

Brain activity of interest is the strongest component of the EEG.

(e.g. for artifact removal, dimensionality reduction)

→ Principal component analysis (PCA)

[e.g., Parra et al., 2005]

BSS methods by assumption

Brain activity differs between experimental conditions.

(ERP studies)

→ Linear classifiers (LDA, SVM, LLR)

[e.g., Blankertz et al., 2010]

Brain activity correlates with behaviour or stimulus variables.

(ERP studies)

→ Linear regression (OLS, Ridge regression, LASSO)

[e.g., Parra et al., 2005]

Brain activity of interest is the strongest component of the EEG.

(e.g. for artifact removal, dimensionality reduction)

→ Principal component analysis (PCA)

[e.g., Parra et al., 2005]

Brain activity of interest correlates across subjects/stimulus repetitions.

(Hyperscanning ERP studies)

→ Canonical correlation analysis (CCA) [e.g., Dmochowski et al., 2011]

BSS methods by assumption (2)

Brain components are mutually independent.

(many uses including artifact removal)

→ Independent component analysis (ICA)

[e.g., Makeig et al.]

BSS methods by assumption (2)

Brain components are mutually independent.

(many uses including artifact removal)

→ Independent component analysis (ICA)

[e.g., Makeig et al.]

Brain components are Granger-causally interacting.

(brain connectivity studies)

→ SCSA, MVARICA

[Gomez-Herrero et al., 2008; Haufe et al., 2010]

BSS methods by assumption (2)

Brain components are mutually independent.

(many uses including artifact removal)

→ Independent component analysis (ICA)

[e.g., Makeig et al.]

Brain components are Granger-causally interacting.

(brain connectivity studies)

→ SCSA, MVARICA

[Gomez-Herrero et al., 2008; Haufe et al., 2010]

Brain activity is (non-) stationary.

(dimensionality reduction)

→ Stationary subspace analysis (SSA)

[von Bünau et al., 2009]

BSS methods by assumption (2)

Brain components are mutually independent.

(many uses including artifact removal)

→ Independent component analysis (ICA)

[e.g., Makeig et al.]

Brain components are Granger-causally interacting.

(brain connectivity studies)

→ SCSA, MVARICA

[Gomez-Herrero et al., 2008; Haufe et al., 2010]

Brain activity is (non-) stationary.

(dimensionality reduction)

→ Stationary subspace analysis (SSA)

[von Bünau et al., 2009]

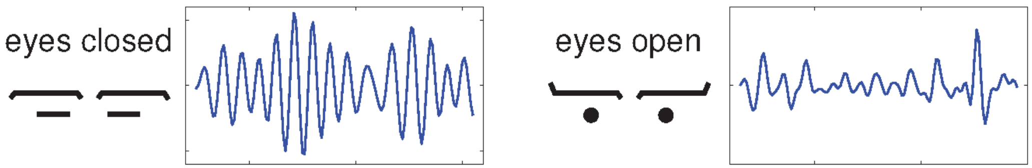
If the brain activity of interest can be characterized in several ways, multiple BSS methods may lead to the same solution.

BSS for oscillations

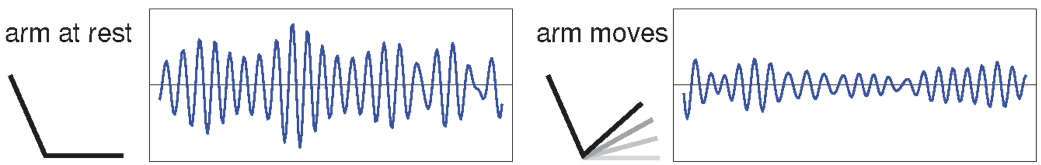
Not all EEG phenomena are phase-locked to certain events. There are also **rhythms**, the amplitude of which modulates depending on the mental state.

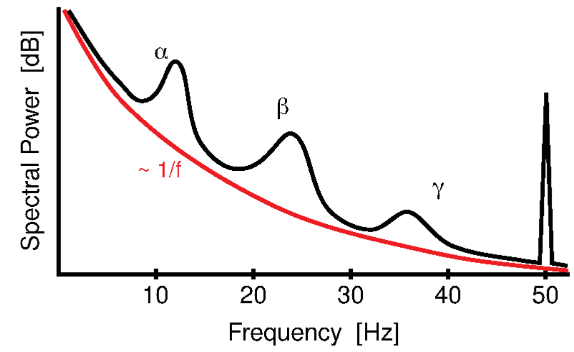
Most rhythms are **idle** rhythms, i.e., are attenuated during activation.

- α-rhythm (around 10 Hz) in visual cortex:



- μ-rhythm (around 10 Hz) in motor and sensory cortex:



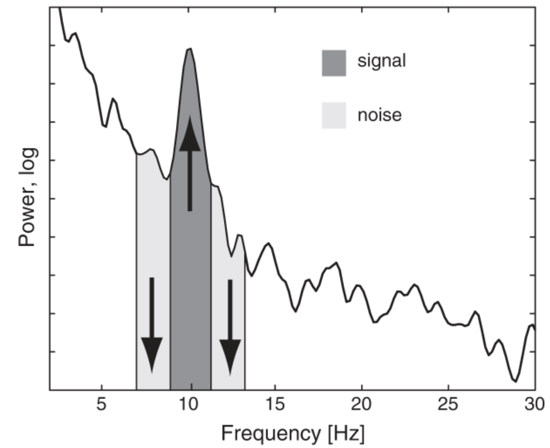


Figures by Benjamin Blankertz

Spatio-spectral decomposition (SSD)

Signal of interest is narrow-band oscillation.

$$\mathbf{w} = \arg\max_{\mathbf{w}'} \mathrm{SNR}(\mathbf{w}') = \arg\max_{\mathbf{w}'} \frac{\mathbf{w}'^{\top} \boldsymbol{\Sigma}_{\mathrm{signal}} \mathbf{w}'}{\mathbf{w}'^{\top} \left(\boldsymbol{\Sigma}_{\mathrm{noise}}\right) \mathbf{w}'}$$



$\boldsymbol{\Sigma}_{\mathrm{signal}}$ and $\boldsymbol{\Sigma}_{\mathrm{noise}}$ are the covariances of the data filtered in the central and flanking frequency bands.

$\mathbf{w}$ is obtained as the solution to the generalized eigenvalue equation

$\boldsymbol{\Sigma}_{\mathrm{signal}} \mathbf{w} = \lambda \boldsymbol{\Sigma}_{\mathrm{noise}} \mathbf{w}$ (in Matlab: $\mathbf{W} = \mathrm{eig}(\boldsymbol{\Sigma}_{\mathrm{signal}}, \boldsymbol{\Sigma}_{\mathrm{noise}});$) . [Nikulin et al., 2011]

Common spatial patterns (CSP)

Power of oscillations differs between two experimental conditions C1 and C2.

$$\mathbf{w}_1 = \arg\min_{\mathbf{w}} \frac{\mathbf{w}^\top \mathbf{\Sigma}_1 \mathbf{w}}{\mathbf{w}^\top (\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2) \mathbf{w}} \qquad \mathbf{w}_2 = \arg\min_{\mathbf{w}} \frac{\mathbf{w}^\top \mathbf{\Sigma}_2 \mathbf{w}}{\mathbf{w}^\top (\mathbf{\Sigma}_1 + \mathbf{\Sigma}_2) \mathbf{w}}$$

[Koles et al., 1990]

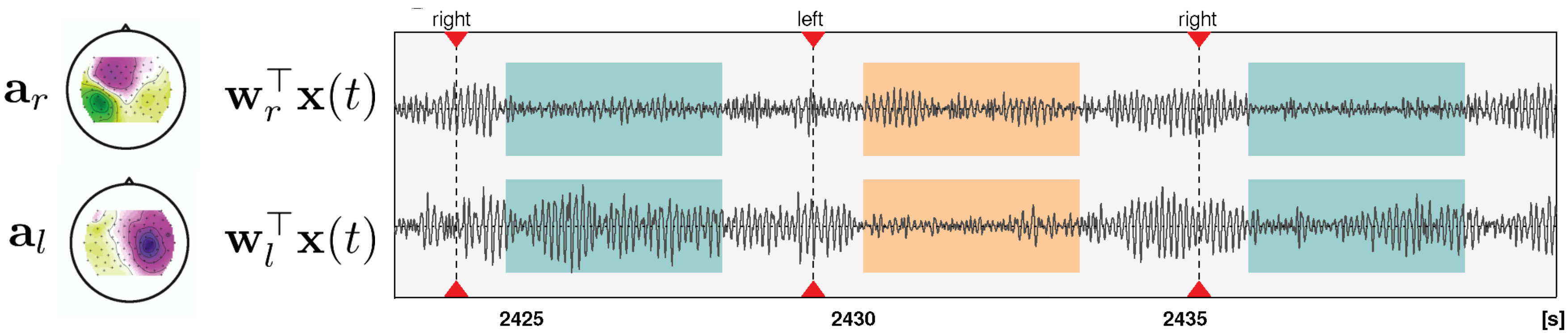
Common spatial patterns (CSP)

Power of oscillations differs between two experimental conditions C1 and C2.

$$\mathbf{w}_1 = \arg\min_{\mathbf{w}} \frac{\mathbf{w}^\top \boldsymbol{\Sigma}_1 \mathbf{w}}{\mathbf{w}^\top (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)\mathbf{w}} \qquad \mathbf{w}_2 = \arg\min_{\mathbf{w}} \frac{\mathbf{w}^\top \boldsymbol{\Sigma}_2 \mathbf{w}}{\mathbf{w}^\top (\boldsymbol{\Sigma}_1 + \boldsymbol{\Sigma}_2)\mathbf{w}}$$

[Koles et al., 1990]

Example: BCI based on motor imagery of left and right hand.



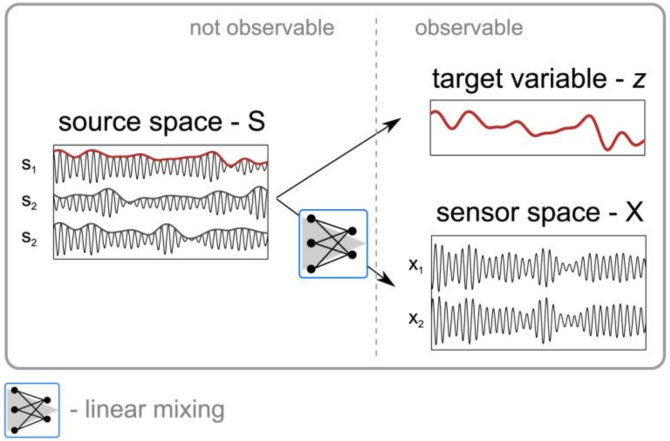
Figures by Benjamin Blankertz

Source power correlation analysis (SPoC)

Instantaneous amplitude (envelope) of oscillations correlates with continuous variable (behaviour, stimulus properties, etc.) .

[Dähne et al., 2014]

$$\mathbf{w} = \arg\max_{\mathbf{w}'} \operatorname{corr}\left(\operatorname{env}\left(\mathbf{w}'^{\top}\mathbf{x}(t)\right), z(t)\right)$$

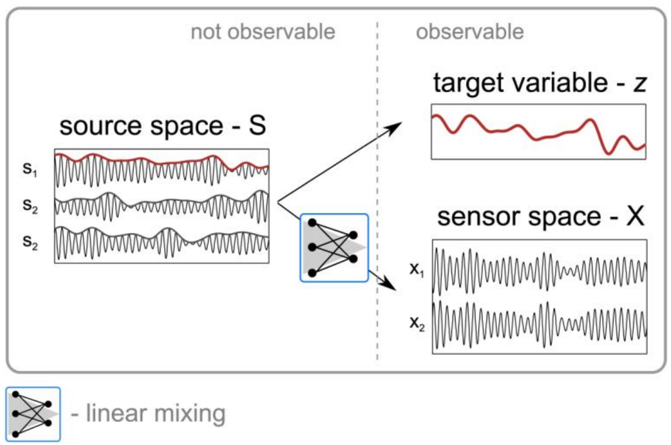


Source power correlation analysis (SPoC)

Instantaneous amplitude (envelope) of oscillations correlates with continuous variable (behaviour, stimulus properties, etc.) .

[Dähne et al., 2014]

$$\mathbf{w} = \arg\max_{\mathbf{w}'} \operatorname{corr}\left(\operatorname{env}\left({\mathbf{w}'}^\top \mathbf{x}(t)\right), z(t)\right)$$



Instantaneous amplitude correlates across subjects/stimulus repetitions

→ Canonical SPoC (cSPoC) .

[Dähne et al., 2014, Submitted]

Extraction of steady-state auditory evoked potentials

Rhythmic auditory stimulation elicits phase-locked rhythmic activity in auditory cortex = SSAEP (same as for visual stimulation and SSVEP).

[e.g., Galambos et al., 1981]

Linear relationship between loudness (in dB) and SSAEP amplitude.

[Picton et al., 2003]

Extraction of steady-state auditory evoked potentials

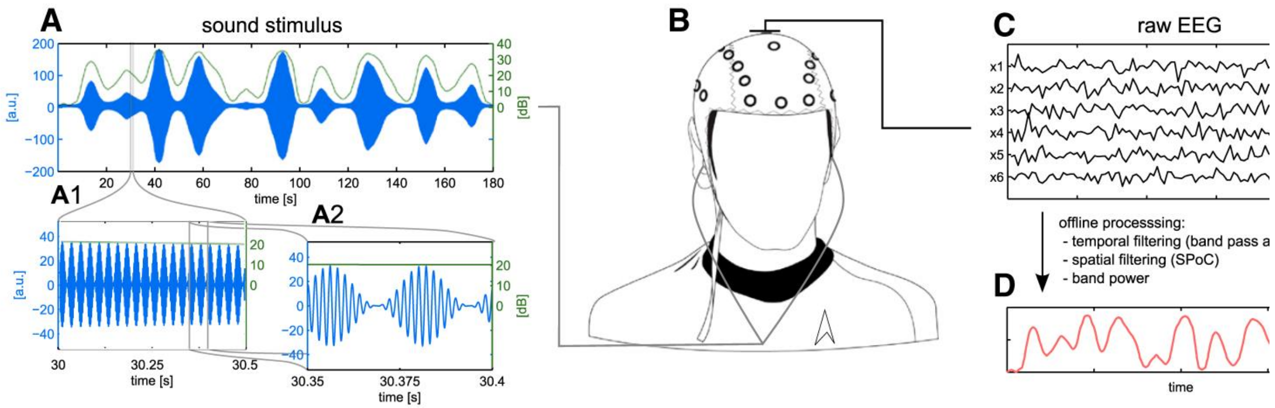
Rhythmic auditory stimulation elicits phase-locked rhythmic activity in auditory cortex = SSAEP (same as for visual stimulation and SSVEP).

[e.g., Galambos et al., 1981]

Linear relationship between loudness (in dB) and SSAEP amplitude.

[Picton et al., 2003]

Experiment: apply 40Hz artifical sound stimulus modulating loudness.



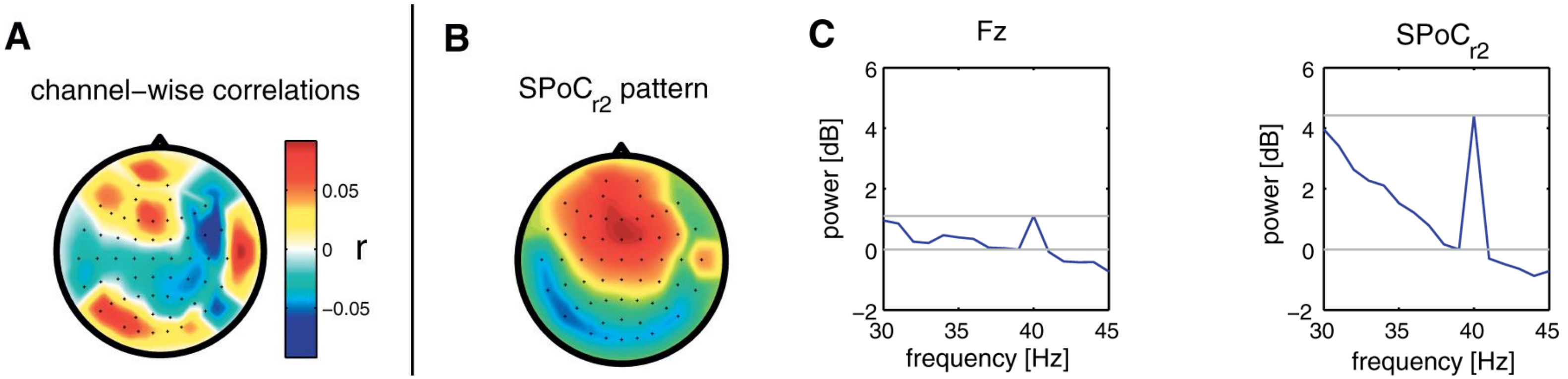
Task: identify SSAEP component.

[Dähne et al., 2014]

Extraction of steady-state auditory evoked potentials

Results:

- Compared to single sensors, SPoC leads to higer SNR (peak height) and higher correlation with the sound volume (r=0.6 vs. r=0.1)
- SPoC activation pattern localizes to left and right auditory cortices
- Similar results for SSD instead of SPoC



[Dähne et al., 2014]

Summary

- EEG data are mixed due to volume conduction in the head

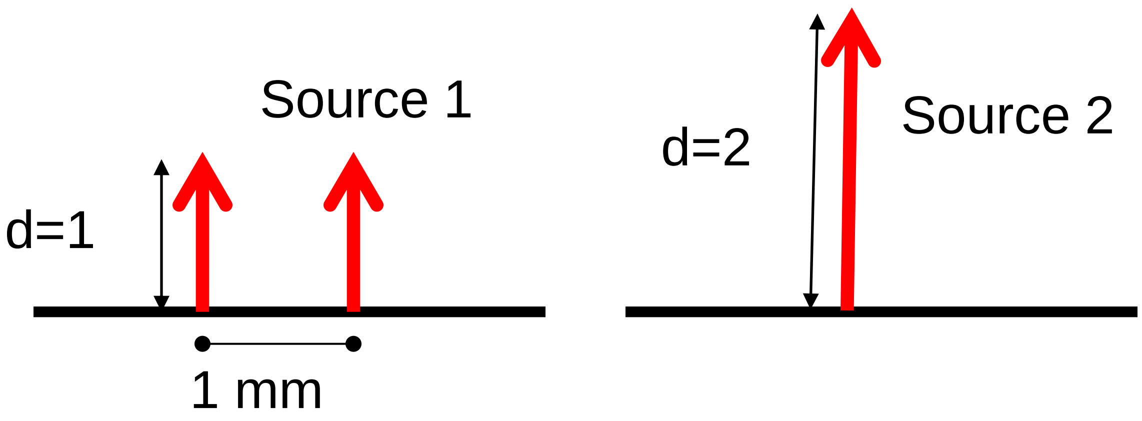
- To increase SNR, and achieve interpretability, the inverse problem needs to be "solved"

- Can be done using a physical model of volume conduction (inverse source reconstruction) or using purely statistical models (source separation)

- In any case, a unique solution is only obtained if prior assumptions/constraints are imposed

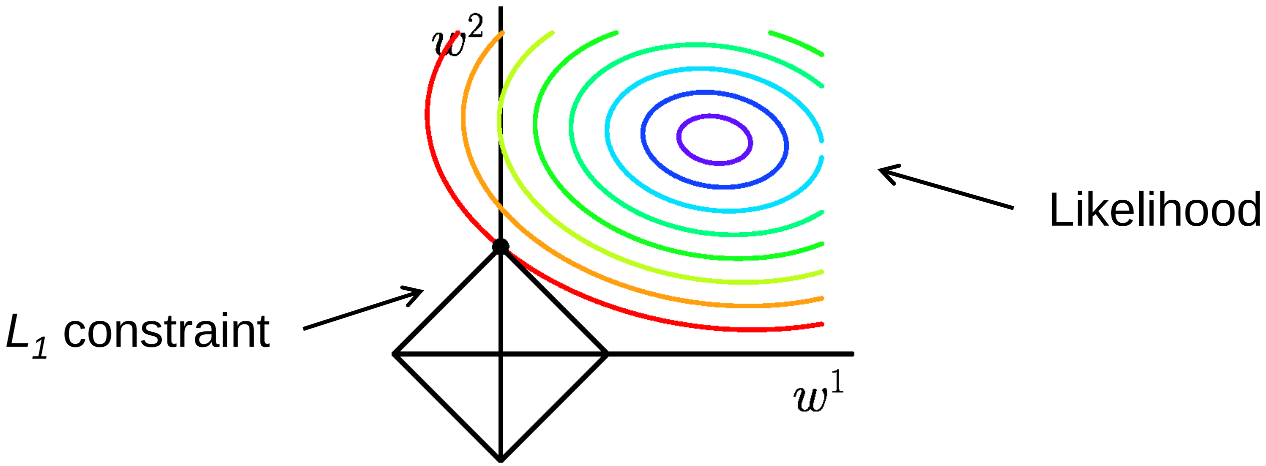
- Correctness of the solution relies on correctness of assumptions

Origin of blurring



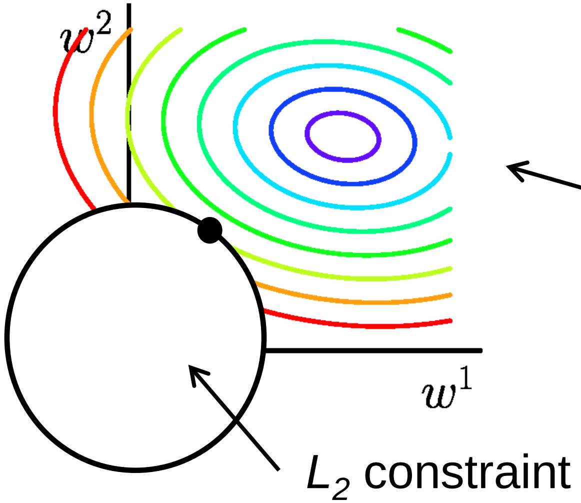
- Both sources explain data equally well
- Source 1 has L_2-norm: $\sqrt{1^2 + 1^2} = \sqrt{2}$
- Source 2 has L_2-norm: $\sqrt{2^2} = 2$

Origin of sparsity



The level sets of Likelihood and constraint **almost always** intersect at the coordinate axes.

No sparsity using L_2-norm



The level sets of Likelihood and constraint **almost never** intersect at the coordinate axes.

Depth compensation

Superficial sources contribute more to the EEG than deep ones.

→ Many superficial sources „cost less“ than one deep source.

→Location bias towards superficial sources.

Depth compensation

Superficial sources contribute more to the EEG than deep ones.

→ Many superficial sources „cost less“ than one deep source.

→Location bias towards superficial sources.

Countermeasure: minimize norm of weighted sources

$$g(\mathbf{s}) = \|\mathbf{W}\mathbf{s}\|$$

with diagonal or blockdiagonal $\mathbf{W}$ encoding a voxel-specific penalty

Depth compensation

1. Norm of the columns of the lead field

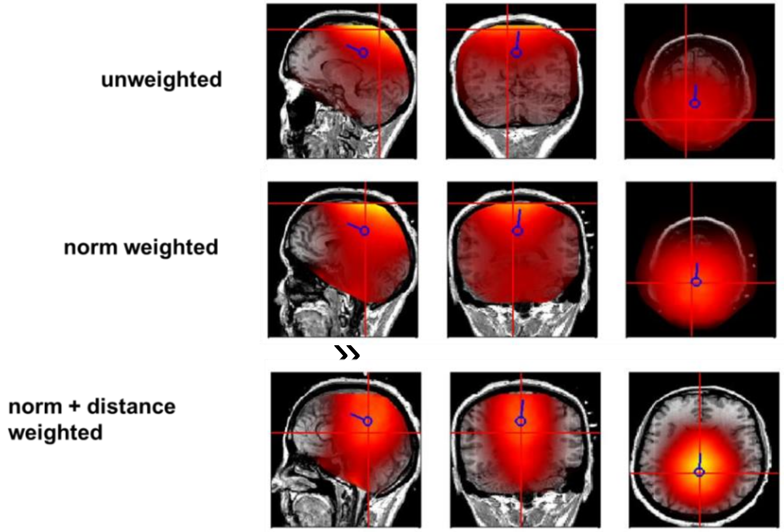
[Jeffs et al., 1987]

2. Voxel-wise (co-) variance of the minimum-norm solution

[Pascual-Marqui, 2002; Haufe et al., 2008]

3. Norm + distance from EEG sensors

[Marzetti et al., 2008]



Choice of $\mathbf{W}$ is crucial.

Sparsity of Vector Fields

Dipole orientations are 3D vectors,
current distributions are 3D vectorfields

Technicality: L_1-norm sets single dimensions to 0

→ Estimated sources are not physiologically plausible (parallel to coordinate axes)

[Haufe et al., 2008; Ding et al., 2008; Ou et al., 2009]

Sparsity of Vector Fields

Dipole orientations are 3D vectors,
current distributions are 3D vectorfields

Technicality: L_1-norm sets single dimensions to 0

→ Estimated sources are not physiologically plausible (parallel to coordinate axes)

Solution: $L_{1,2}$-norm penalty $\sum_i \|\mathbf{s}_i\|_2$

→ Dipole dimensions can only be pruned jointly

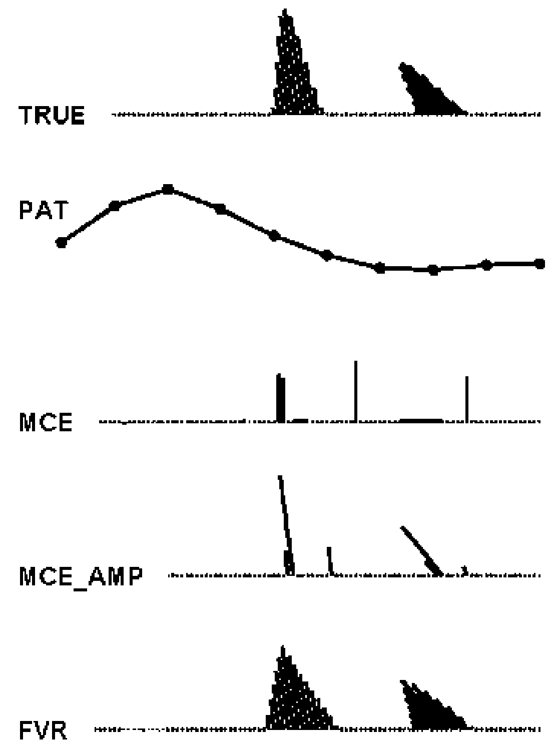
[Haufe et al., 2008; Ding et al., 2008; Ou et al., 2009]

Sparsity of Vector Fields

Dipole orientations are 3D vectors,
current distributions are 3D vectorfields

Technicality: L_1-norm sets single dimensions to 0

→ Estimated sources are not physiologically plausible (parallel to coordinate axes)



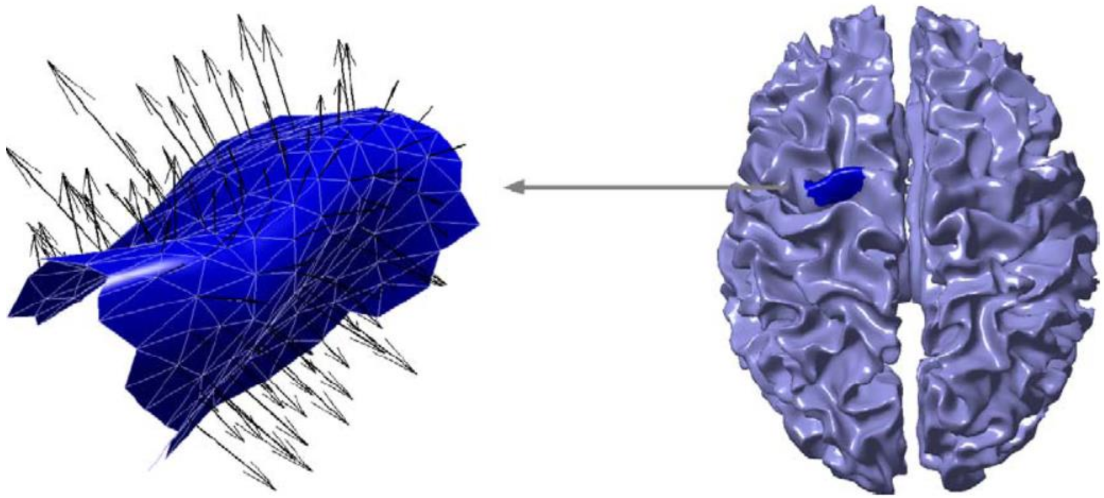
Solution: $L_{1,2}$-norm penalty $\sum_i \|\mathbf{s}_i\|_2$

→ Dipole dimensions can only be pruned jointly.

[Haufe et al., 2008; Ding et al., 2008; Ou et al., 2009]

More „physiological“ constraints

K. Jerbi et al. / NeuroImage 22 (2004) 779–793



1. Sources on cortex, arbitrary orientation
2. Sources on cortex, orientation normal to surface (dangerous!)
3. Regions of interest
4. Symmetric configurations