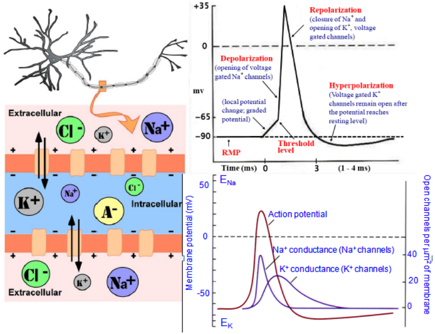
Bayesian EEG Source Estimation

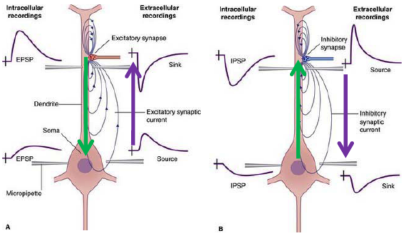
Lecture Notes

Ahmet Ademoglu, *PhD*,
Bogazici University,
Institute of Biomedical Engineering

Some concepts and illustrations in this lecture are adapted from the textbook,
Statistical Parametric Mapping: The Analysis of Functional Brain Images, Editors: K. Friston, J. Ashburner, S. Kiebel, T. Nichols and W. Penny, *Academic Press*, 2006.

What do we measure by EEG/MEEG?



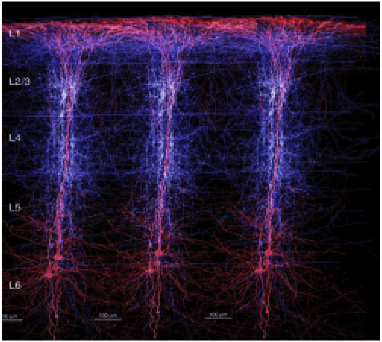


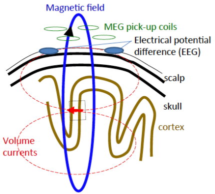
Action potentials are biphasic : not ideal for summation due to cancellation

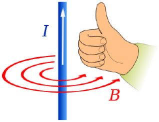
Post-synaptic potentials are monophasic and slower : ideal for summation

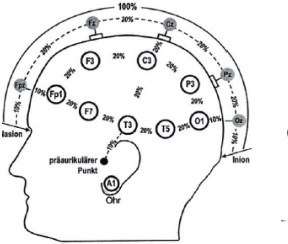
Pyramidal cells are ideal as current generators because they are:

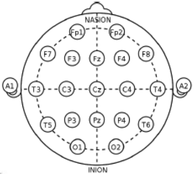
- Spatially Aligned
- Perpendicular to cortical surface
- Recurrently connected
- Receive synchronous inputs



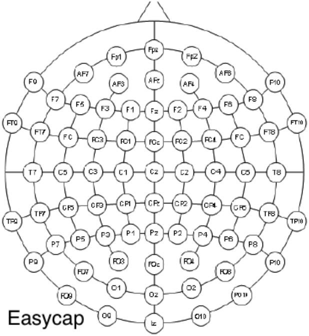






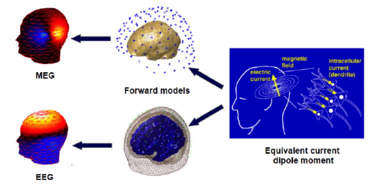


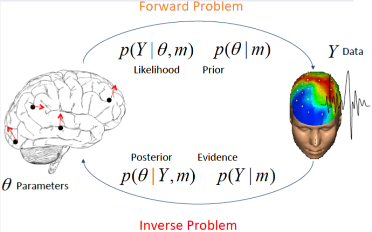
10-20 Electrode System



Easycap

Source Reconstruction





Forward Problem : $\mathbf{y} = \mathbf{X}\theta + \epsilon$

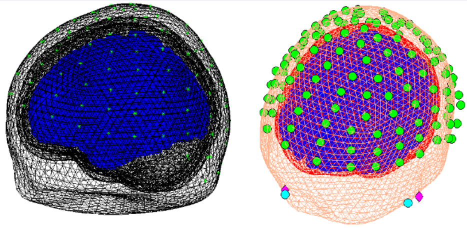
$\mathbf{y}$: $(n_e \times 1)$ vector of measured signal from sensors

θ : $(n_v \times 1)$ electrical current dipole vector at the vertices of the cortex model

$\mathbf{X}$: $(n_e \times n_v)$ transformation matrix called *Lead Field Matrix*

Inverse Mapping from Sensor Space (EEG/MEG) to Source Space

Realistic Head Model for Boundary Element Method



- The *Lead Field Matrix*, $\mathbf{X}$ is computed using the primary current sources on the cortex and the secondary current sources due to the boundaries over the tesselated structure.
- The conductivities of the brain/CSF, the skull and the scalp are assumed to be homogeneous.
- The tesselated structure is obtained from the segmented anatomical MRI and the sensor locations are registered to the appropriate vertices on the scalp surface.

Distributed Source Reconstruction Problem

$$\mathbf{y} = \mathbf{X}\theta + \epsilon,$$

- Determination of source distribution θ from the measurements $\mathbf{y}$ is an ill-posed problem because $n_e << n_v$ leading to multiple solutions.
- Assuming $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_\epsilon)$, the solution is obtained by imposing constraints on the source space otherwise known as *regularization*:

$$\hat{\theta} = \min_j \left(|\mathbf{C}_\epsilon^{-1/2}(\mathbf{y} - \mathbf{X}\theta)|^2 + \lambda |\mathbf{H}\theta|^2 \right)$$

It is called weighted minimum norm solution is expressed as

$$\hat{\theta} = \mathbf{T}\mathbf{y}$$

$$\mathbf{T} = [\mathbf{X}^T\mathbf{C}_\epsilon^{-1}\mathbf{X} + \lambda\mathbf{H}^T\mathbf{H}]^{-1}\mathbf{X}^T\mathbf{C}_\epsilon^{-1} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{X}^T[\mathbf{X}(\mathbf{H}^T\mathbf{H}^{-1}) + \lambda\mathbf{C}_\epsilon]^{-1}$$

The latter expression is obtained by using the Matrix Inversion Lemma,

$$(\mathbf{A} + \mathbf{BCD})^{-1}\mathbf{BC} = \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}$$

- λ is the regularization constant and it has to be optimized by plotting the $|\mathbf{H}\hat{\theta}|$ vs $|\mathbf{C}_\epsilon^{-1/2}(\mathbf{y} - \mathbf{X}\hat{\theta})|$ *i.e.* the $L-$curve and by finding its deflection point.

```
% SIMULATION OF A 2D INVERSE PROBLEM WITH SMOOTHNESS + DEPTH CONSTRAINT
DIM = [32 32 1]; S = 40; % number of sensors
Gain=1; X = zeros(DIM(1),DIM(2));
RANGE = round ( [DIM(1)/6+[1:5]   DIM(1)/1.5+[1:6] ] );
X( RANGE,RANGE) = 1; % create an image
X = spm_conv(X,3,3); % smooth the image
L = Gain*randn(S,DIM(1)*DIM(2)); % Generate Lead Field Matrix
V = L*X(:); % Calculate the forward problem to obtain the sensor data
Vn = V + randn(S,1); % V + E : Add noise to sensor data
C_e = eye(S,S); % Noise covariance
% Inversion : V = LX + E,  Minimize { (V - L*X)^T*C_e_inv*(V - L*X)^T + lambda*X^T*HTH*X }
% HTH : a priori constraint, C_e_inv : Inverse of the noise covariance matrix E distributed as N(0,C_e)
% T = inv(L'*C_e_inv*L + lambda*HTH)*L'*C_e_inv
% HTH : Laplacian Operator
% Calculate the Adjacency matrix of the 2D data
mat = reshape([1:DIM(1)*DIM(2)],DIM(1),DIM(2));              % neigboring graph
[r,c] = size(mat);                             %# Get the matrix size
diagVec1 = repmat([ones(c-1,1); 0],r,1);  %# Make the first diagonal vector
%#   (for horizontal connections)
diagVec1 = diagVec1(1:end-1);                  %# Remove the last value
diagVec2 = ones(c*(r-1),1);                    %# Make the second diagonal vector
%#   (for vertical connections)
adj = diag(diagVec1,1)+...                     %# Add the diagonals to a zero matrix
diag(diagVec2,c);
adj = adj+adj.' ;                              %'# Add the matrix to a transposed
M = (adj - 4*eye(size(adj)));  % Laplace operator
W=  diag((diag(L'*L).^( 0.5)));   % Depth Weighting   %W = eye(size(M)); % no depth weighting
MW = M*W; % combined prior : Laplace + Depth Weighting
HTH_inv = pinv (MW'*MW);  lambda = 1.5;
T = HTH_inv*L'*inv(L*HTH_inv*L' + lambda * C_e); % Using Matrix Inversion Lemma
X_e = reshape(T*Vn,DIM(1),DIM(2)); X_ls = reshape (pinv(L)*Vn,DIM(1),DIM(2));
subplot(3,1,1); imagesc(X);subplot(3,1,2); imagesc(X_e); subplot(3,1,3); imagesc(X_ls);
```

A hierarchical approach

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\theta + \epsilon^{(1)} \\ \theta &= \mathbf{0} + \epsilon^{(2)} \end{aligned}$$

where $\epsilon^{(1)} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\epsilon^{(1)}})$ and $\epsilon^{(2)} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\epsilon^{(2)}})$.
$\mathbf{C}_{\epsilon}^{(1)}$ and $\mathbf{C}_{\epsilon}^{(2)}$ can be modeled in terms of covariance components:

$$\begin{aligned} \mathbf{C}_{\epsilon^{(1)}} &= \lambda_1^{(1)}\mathbf{Q}_1^{(1)} + \lambda_2^{(1)}\mathbf{Q}_2^{(1)} + \ldots \\ \mathbf{C}_{\epsilon^{(2)}} &= \lambda_1^{(2)}\mathbf{Q}_1^{(2)} + \lambda_2^{(2)}\mathbf{Q}_2^{(2)} + \ldots \end{aligned}$$

Choice of the priors

At the sensor level, a single noise component can be defined as $\mathbf{Q}_1^{(1)} = \mathbf{I}_{n_e}$. At the source level, several different priors can be chosen as;

1. Smoothness Constraint: $\mathbf{Q}_s^{(1)}(i,j) = \exp(-d_{ij}^2/2s^2)$ where d_{ij} is the geodesic distance between dipoles i and j. s is the smoothness parameter, typically chosen as 8 mm enforcing correlation among sources.
2. Extrinsic Functional Constraints: The activity map from other imaging modalities like fMRI or PET can be used as a binary mask to distinguish *a priori*, the active and inactive cortical areas. The corresponding source variance component can be defined as $\mathbf{Q}_e^{(1)}(i,i) = 1$, when the source is part of the i^{th} vertex, and zero otherwise.
3. Depth priors: To ensure that the sources contribute to the solution equally irrespective of their depth, deeper sources are given a larger prior variance than superficial sources. The depth is indexed by the norm of the gain vector for each source. This covariance component is defined by $diag(\mathbf{X}^\mathbf{T}\mathbf{X})^{-\mathbf{1}}$.

Restricted log-likelihood function $\mathcal{L}$ for a model

$$\mathbf{y} = \mathbf{X}\theta + \epsilon,\ \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_\epsilon)$$

$\mathcal{L} = \frac{1}{2}\log|\mathbf{C}_\epsilon^{-1}(\lambda)| - \frac{1}{2}\log|\mathbf{X}^\mathbf{T}\mathbf{C}_\epsilon^{-1}(\lambda)\mathbf{X}| - \frac{1}{2}Tr[\mathbf{C}_\epsilon^{-1}(\lambda)(\mathbf{y}-\mathbf{X}\theta)(\mathbf{y}-\mathbf{X}\theta)^T] + Const$

Optimizing $\mathcal{L}$ w.r.t. λ i.e. $\frac{\partial\mathcal{L}}{\partial\lambda} = 0$ yields the ReML algorithm

REML algorithm

Input : $\mathbf{yy}^T$, $\{\mathbf{Q}_1, \mathbf{Q}_2, \ldots\}$ Output : $\mathbf{C}_\epsilon$

$\mathbf{C}_\epsilon = \sum_i \lambda_i \mathbf{Q}_i = \lambda_1^{(1)}\mathbf{Q}_1^{(1)} + \lambda_2^{(1)}\mathbf{Q}_2^{(1)} + \ldots$

$\mathbf{C}_{\theta|y} = (\mathbf{X}^T\mathbf{C}_\epsilon^{-1}\mathbf{X})^{-1}$

$\mathbf{P} = \mathbf{C}_\epsilon^{-1} - \mathbf{C}_\epsilon^{-1}\mathbf{X}\mathbf{C}_{\theta|y}^{-1}\mathbf{X}^T\mathbf{C}_\epsilon^{-1}$

$g_i = -\frac{1}{2}Tr[\mathbf{PQ_i}] + \frac{1}{2}Tr[\mathbf{Pyy}^T\mathbf{P}^T\mathbf{Q}_i]$ $H_{ij} = \frac{1}{2}Tr[\mathbf{PQ}_i\mathbf{PQ_j}]$

$\lambda \longleftarrow \lambda + \mathbf{H}^{-1}\mathbf{g}$

The posterior covariance and mean of θ can be determined as

$\mathbf{C}_{\theta|y} = (\mathbf{X}^T\mathbf{C}_\epsilon^{-1}\mathbf{X})^{-1}$ and $\eta_{\theta|y} = \mathbf{C}_{\theta|y}(\mathbf{X}^T\mathbf{C}_\epsilon^{-1}\mathbf{y})$

ReML solution

Using the two-level hierarchical model

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\theta + \epsilon^{(1)} \\ \theta &= \mathbf{0} + \epsilon^{(2)} \end{aligned}$$

where $\epsilon^{(1)} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\epsilon^{(1)}})$ and $\epsilon^{(2)} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\epsilon^{(2)}})$.
and the augmented form
$\bar{\mathbf{y}} = [\mathbf{y}\ \mathbf{0}]^T$ and $\bar{\mathbf{X}} = \begin{bmatrix} \mathbf{X} \\ \mathbf{I} \end{bmatrix}$ and $\mathbf{C}_\epsilon = \begin{bmatrix} \mathbf{C}_{\epsilon^{(1)}} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{\epsilon^{(2)}} \end{bmatrix}$
$\mathbf{C}_{\theta|y} = (\bar{\mathbf{X}}^T \mathbf{C}_\epsilon^{-1} \bar{\mathbf{X}})^{-1}$ and $\eta_{\theta|y} = \mathbf{C}_{\theta|y}(\bar{\mathbf{X}}^T \mathbf{C}_\epsilon^{-1} \bar{\mathbf{y}})$

REML algorithm (Augmented Model)

Input : $\bar{\mathbf{y}}\bar{\mathbf{y}}^T, \{\mathbf{Q}_1^{(1)}, \mathbf{Q}_2^{(1)}, \ldots, \mathbf{Q}_1^{(2)}, \mathbf{Q}_2^{(2)}, \ldots\}$ Output : $\mathbf{C}_\epsilon$
$\mathbf{C}_\epsilon = \sum_i \lambda_i \mathbf{Q}_i = \lambda_1^{(1)}\mathbf{Q}_1^{(1)} + \lambda_2^{(1)}\mathbf{Q}_2^{(1)} + \cdots + \lambda_1^{(2)}\mathbf{Q}_1^{(2)} + \lambda_2^{(2)}\mathbf{Q}_2^{(2)} + \ldots$
$\mathbf{C}_{\theta|y} = (\bar{\mathbf{X}}^T \mathbf{C}_\epsilon^{-1} \bar{\mathbf{X}})^{-1}$
$\mathbf{P} = \mathbf{C}_\epsilon^{-1} - \mathbf{C}_\epsilon^{-1}\bar{\mathbf{X}}\mathbf{C}_{\theta|y}^{-1}\bar{\mathbf{X}}^T\mathbf{C}_\epsilon^{-1}$
$g_i = -\frac{1}{2}Tr[\mathbf{P}\mathbf{Q_i}] + \frac{1}{2}Tr[\mathbf{P}\bar{\mathbf{y}}\bar{\mathbf{y}}^T\mathbf{P}^T\mathbf{Q}_i]$ $H_{ij} = \frac{1}{2}Tr[\mathbf{P}\mathbf{Q}_i\mathbf{P}\mathbf{Q_j}]$
$\lambda \longleftarrow \lambda + \mathbf{H}^{-1}\mathbf{g}$

ReML solution on sensor space

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\theta + \epsilon^{(1)} \\ \theta &= \mathbf{0} + \epsilon^{(2)} \end{aligned}$$ where $\epsilon^{(1)} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\epsilon^{(1)}})$ and $\epsilon^{(2)} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\epsilon^{(2)}})$.

$\mathbf{C}_\epsilon^{(1)}$ and $\mathbf{C}_\epsilon^{(2)}$ can be modeled in terms of covariance components:

$$\begin{aligned} \mathbf{C}_{\epsilon^{(1)}} &= \lambda_1^{(1)}\mathbf{Q}_1^{(1)} + \lambda_2^{(1)}\mathbf{Q}_2^{(1)} + \dots \\ \mathbf{C}_{\epsilon^{(2)}} &= \lambda_1^{(2)}\mathbf{Q}_1^{(2)} + \lambda_2^{(2)}\mathbf{Q}_2^{(2)} + \dots \end{aligned}$$

The collapsed model: $\mathbf{y} = \mathbf{X}\epsilon^{(2)} + \epsilon^{(1)}$ yields the sensor space covariance $\mathbf{C}_y$ as
$\mathbf{C}_y = \mathbf{C}_{\epsilon^{(1)}} + \mathbf{X}\mathbf{C}_{\epsilon^{(2)}}\mathbf{X}^T = \lambda_1^{(1)}\mathbf{Q}_1^{(1)} + \lambda_2^{(1)}\mathbf{Q}_2^{(1)} + \dots + \lambda_1^{(2)}\mathbf{X}\mathbf{Q}_1^{(2)}\mathbf{X}^T + \lambda_2^{(2)}\mathbf{X}\mathbf{Q}_2^{(2)}\mathbf{X}^T + \dots$

REML algorithm (Collapsed Model)

Input : $\mathbf{y}\mathbf{y}^T$, $\{\mathbf{Q}_1^{(1)}, \mathbf{Q}_2^{(1)}, \dots, \mathbf{X}\mathbf{Q}_1^{(2)}\mathbf{X}^T, \mathbf{X}\mathbf{Q}_2^{(2)}\mathbf{X}^T, \dots\}$ Output : $\mathbf{C}_y$

$\mathbf{C}_y = \sum_i \lambda_i \mathbf{Q}_i = \lambda_1^{(1)}\mathbf{Q}_1^{(1)} + \lambda_2^{(1)}\mathbf{Q}_2^{(1)} + \lambda_1^{(2)}\mathbf{X}\mathbf{Q}_1^{(2)}\mathbf{X}^T + \lambda_2^{(2)}\mathbf{X}\mathbf{Q}_2^{(2)}\mathbf{X}^T + \dots$

$\mathbf{C}_{\theta|y} = (\mathbf{X}^T\mathbf{C}_y^{-1}\mathbf{X})^{-1}$ and $\eta_{\theta|y} = \mathbf{C}_{\theta|y}(\mathbf{X}^T\mathbf{C}_y^{-1}\mathbf{y})$

$\mathbf{P} = \mathbf{C}_y^{-1} - \mathbf{C}_y^{-1}\mathbf{X}\mathbf{C}_{\theta|y}^{-1}\mathbf{X}^T\mathbf{C}_y^{-1}$

$g_i = -\frac{1}{2}Tr[\mathbf{P}\mathbf{Q_i}] + \frac{1}{2}Tr[\mathbf{P}\mathbf{y}\mathbf{y}^T\mathbf{P}^T\mathbf{Q}_i]$ $H_{ij} = \frac{1}{2}Tr[\mathbf{P}\mathbf{Q}_i\mathbf{P}\mathbf{Q_j}]$

$\lambda \longleftarrow \lambda + \mathbf{H}^{-1}\mathbf{g}$

Bayesian source reconstruction using spatio-temporal models

Using the two-level hierarchical model

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\theta + \epsilon^{(1)} \\ \theta &= \mathbf{0} + \epsilon^{(2)} \end{aligned}$$

- $\mathbf{Y}$: $(n_e \times n_t)$ matrix of measured signal from sensors
- θ : $(n_v \times n_t)$ electrical current dipole vector at the vertices of the cortex model
- $\mathbf{X}$: $(n_e \times n_v)$ transformation matrix called *Lead Field Matrix*
- $vec(\epsilon^{(1)}) \sim \mathcal{N}(\mathbf{0}, \mathbf{V}^{(1)} \otimes \mathbf{C}_{\epsilon^{(1)}}) \longleftrightarrow$ parametric noise covariance model
- $vec(\epsilon^{(2)}) \sim \mathcal{N}(\mathbf{0}, \mathbf{V}^{(2)} \otimes \mathbf{C}_{\epsilon^{(2)}}) \longleftrightarrow$ parametric source covariance model
- $\mathbf{V}^{(1)}$ and $\mathbf{V}^{(2)}$: temporal correlation matrices of noise and signal, respectively
- $\mathbf{C}_{\epsilon^{(1)}}$ and $\mathbf{C}_{\epsilon^{(2)}}$: spatial covariance matrices of noise and sources, respectively.
- $\mathbf{C}_{\epsilon^{(1)}} = \sum \lambda_i^{(1)} \mathbf{Q}_i^{(1)}$
- $\mathbf{C}_{\epsilon^{(2)}} = \sum \lambda_i^{(2)} \mathbf{Q}_i^{(2)}$

Tensor Properties

- $(\mathbf{V} \otimes \mathbf{C})^{-1} = \mathbf{V}^{-1} \otimes \mathbf{C}^{-1}$
- $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$
- $(\mathbf{A} \otimes \mathbf{B})^T = (\mathbf{A}^T \otimes \mathbf{B}^T)$
- $Tr(\mathbf{A}^T\mathbf{B}) = vec(\mathbf{A})^T vec(\mathbf{B})$
- $vec(\mathbf{Y})^T(\mathbf{V} \otimes \mathbf{C})^{-1}vec(\mathbf{Y}) = vec(\mathbf{Y})^T(\mathbf{V}^{-1} \otimes \mathbf{C}^{-1})vec(\mathbf{Y}) = Tr[\mathbf{C}^{-1}\mathbf{Y}\mathbf{V}^{-1}\mathbf{Y}^T]$
- $|\mathbf{V} \otimes \mathbf{C}| = |\mathbf{V}|^{rank(\mathbf{C})}|\mathbf{C}|^{rank(\mathbf{V})}$
- $vec(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})vec(\mathbf{B})$
- $Tr[\mathbf{A} \otimes \mathbf{B}] = Tr[\mathbf{A}]Tr[\mathbf{B}]$

The posterior covariance $\mathbf{C}_{\theta|Y}$ and its mean $vec(\eta_{\theta|Y})$ are

$$\begin{aligned}
\mathbf{C}_{\theta|Y} &= \left(\mathbf{V}^{(1)-1} \otimes \mathbf{X}^T \mathbf{C}_{\epsilon(1)}^{-1} \mathbf{X} + \mathbf{V}^{(2)-1} \otimes \mathbf{C}_{\epsilon(1)}^{-1}\right)^{-1} \\
vec(\eta_{\theta|Y}) &= \mathbf{C}_{\theta|Y}(\mathbf{V}^{(1)-1} \otimes \mathbf{X}^T \mathbf{C}_{\epsilon(1)}^{-1}) vec(\mathbf{Y}) \\
&= (\mathbf{V}^{(2)} \otimes \mathbf{C}_{\epsilon(2)} \mathbf{X}^T)\left(\mathbf{V}^{(2)} \otimes \mathbf{X}\mathbf{C}_{\epsilon(2)}\mathbf{X}^T + \mathbf{V}^{(1)} \otimes \mathbf{C}_{\epsilon(1)}\right)^{-1} vec(\mathbf{Y})
\end{aligned}$$

The last line is obtained by using matrix inversion lemma

$$(\mathbf{A} + \mathbf{BCD})^{-1}\mathbf{BC} = \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{D}\mathbf{A}^{-1}\mathbf{B})^{-1}$$

Computation of $\mathbf{C}_{\theta|Y}$ and $vec(\eta_{\theta|Y})$ becomes simpler if we choose $\mathbf{V}^{(1)}$ and $\mathbf{V}^{(2)}$ to be the same *i.e.* $\mathbf{V}^{(1)} = \mathbf{V}^{(2)} = \mathbf{V}$:

$$\begin{aligned}
\mathbf{C}_{\theta|Y} &= \mathbf{V} \otimes (\mathbf{X}^T \mathbf{C}_{\epsilon(1)}^{-1} \mathbf{X} + \mathbf{C}_{\epsilon(2)}^{-1})^{-1} \\
\eta_{\theta|Y} &= \mathbf{C}_{\theta|Y} \mathbf{X}^T (\mathbf{C}_{\epsilon(2)}^{-1} + \mathbf{X}^T \mathbf{C}_{\epsilon(1)}^{-1} \mathbf{X}) \mathbf{Y}
\end{aligned}$$

Self-Study Problem

Show that for the hierarchical model,

a) $vec(\hat{\theta}) = (\mathbf{V}^{(2)} \otimes \mathbf{C}^{(2)}\mathbf{X}^T)\left(\mathbf{V}^{(2)} \otimes \mathbf{X}\mathbf{C}^{(2)}\mathbf{X}^T + \mathbf{V}^{(1)} \otimes \mathbf{C}^{(1)}\right)^{-1} vec(\mathbf{Y})$

b) $\hat{\theta} = \mathbf{C}^2\mathbf{X}^T\mathbf{C}^{-1}\mathbf{Y}$

where

$\mathbf{C} = \mathbf{C}^{(1)} + \mathbf{X}\mathbf{C}^{(2)}\mathbf{X}^T$ and $\hat{\mathbf{C}} = (\mathbf{X}^T\mathbf{C}^{(1)-1}\mathbf{X} + \mathbf{C}^{(2)-1})^{-1}$

ReML Solution on sensor space for spatiotemporal model

Using the collapsed model: $\mathbf{Y} = \mathbf{X}\epsilon^{(2)} + \epsilon^{(1)}$
the sensor space covariance $\mathbf{C}_Y = \mathbf{V} \otimes \mathbf{C}$ where $\mathbf{C} = \mathbf{X}\mathbf{C}_{\epsilon^{(2)}}\mathbf{X}^T + \mathbf{C}_{\epsilon^{(1)}}$
Once again we have a problem to estimate the hyperparameters of a covariance matrix
$\mathbf{C} = \lambda_1^{(1)}\mathbf{Q}_1^{(1)} + \lambda_2^{(1)}\mathbf{Q}_2^{(1)} + \cdots + \lambda_1^{(2)}\mathbf{X}\mathbf{Q}_1^{(2)}\mathbf{X}^T + \lambda_2^{(2)}\mathbf{X}\mathbf{Q}_2^{(2)}\mathbf{X}^T + \ldots$
to determine the posterior mean and covariance of θ as
$\mathbf{C}_{\theta|y} = (\mathbf{X}^T\mathbf{C}_Y^{-1}\mathbf{X})^{-1}$ and $\eta_{\theta|y} = \mathbf{C}_{\theta|y}\mathbf{X}^T\mathbf{C}_Y^{-1}vec(\mathbf{Y})$

REML algorithm

Input : $vec(\mathbf{Y})vec(\mathbf{Y})^T, \{\mathbf{Q}_1^{(1)}, \mathbf{Q}_2^{(1)}, \ldots, \mathbf{X}\mathbf{Q}_1^{(2)}\mathbf{X}^T, \mathbf{X}\mathbf{Q}_2^{(2)}\mathbf{X}^T, \ldots\}$ Output : $\mathbf{C}_Y$

$\mathbf{C} = \sum_i \lambda_i\mathbf{Q}_i = \lambda_1^{(1)}\mathbf{Q}_1^{(1)} + \lambda_2^{(1)}\mathbf{Q}_2^{(1)} + \lambda_1^{(2)}\mathbf{X}\mathbf{Q}_1^{(2)}\mathbf{X}^T + \lambda_2^{(2)}\mathbf{X}\mathbf{Q}_2^{(2)}\mathbf{X}^T + \ldots$
$\mathbf{C}_Y = \mathbf{V} \otimes \mathbf{C}$,
$\mathbf{C}_{\theta|y} = (\mathbf{X}^T\mathbf{C}_Y^{-1}\mathbf{X})^{-1}$
$\mathbf{P} = \mathbf{C}_Y^{-1} - \mathbf{C}_Y^{-1}\mathbf{X}\mathbf{C}_{\theta|y}^{-1}\mathbf{X}^T\mathbf{C}_Y^{-1}$
$g_i = -\frac{1}{2}Tr[\mathbf{P}\mathbf{Q_i}] + \frac{1}{2}Tr[\mathbf{P}vec(\mathbf{Y})vec(\mathbf{Y})^T\mathbf{P}^T\mathbf{Q}_i]$ $H_{ij} = \frac{1}{2}Tr[\mathbf{P}\mathbf{Q}_i\mathbf{P}\mathbf{Q_j}]$
$\lambda \longleftarrow \lambda + \mathbf{H}^{-1}\mathbf{g}$

A Modification of the ReML scheme for spatio-temporal models

Restricted log-likelihood function $\mathcal{L}$ for a model

$$\mathbf{y} = \mathbf{X}\theta + \epsilon,\ \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_\epsilon)$$

$$\mathcal{L} = \frac{1}{2}\log|\mathbf{C}_\epsilon(\lambda)^{-1}| - \frac{1}{2}\log|\mathbf{X}^\mathbf{T}\mathbf{C}_\epsilon^{-1}(\lambda)\mathbf{X}| - \frac{1}{2}Tr[\mathbf{C}_\epsilon^{-1}(\lambda)(\mathbf{y}-\mathbf{X}\theta)(\mathbf{y}-\mathbf{X}\theta)^T] + Const$$

Using the collapsed model for the hierarchical structure

$$vec(\mathbf{Y}) = (\mathbf{I}\otimes\mathbf{X})vec(\epsilon^{(2)}) + vec(\epsilon^{(1)})$$

where $vec(\epsilon^{(1)}) \sim \mathcal{N}(\mathbf{0}, \mathbf{V}\otimes\mathbf{C}_{\epsilon^{(1)}})$, $vec(\epsilon^{(2)}) \sim \mathcal{N}(\mathbf{0}, \mathbf{V}\otimes\mathbf{C}_{\epsilon^{(2)}})$,

the log-likelihood $\mathcal{L}$ becomes

$$\frac{1}{2}\log|(\mathbf{V}\otimes\mathbf{C})^{-1}| - \frac{1}{2}\log|(\mathbf{I}\otimes\mathbf{X}^\mathbf{T})(\mathbf{V}\otimes\mathbf{C})^{-1}(\mathbf{I}\otimes\mathbf{X})|$$
$$-\frac{1}{2}(vec(\mathbf{Y}) - (\mathbf{I}\otimes\mathbf{X})vec(\theta))^T(\mathbf{V}\otimes\mathbf{C})^{-1}(vec(\mathbf{Y}) - (\mathbf{I}\otimes\mathbf{X})vec(\theta)) + Const$$

where $\mathbf{C}_\epsilon = \mathbf{V}\otimes\mathbf{C}$ and $\mathbf{C} = \mathbf{C}_{\epsilon^{(1)}} + \mathbf{X}\mathbf{C}_{\epsilon^{(2)}}\mathbf{X}^T$

Using $(\mathbf{C}^T \otimes \mathbf{A})vec(\mathbf{B}) = vec(\mathbf{ABC})$

$-\frac{1}{2}(vec(\mathbf{Y}) - (\mathbf{I} \otimes \mathbf{X})vec(\theta))^T(\mathbf{V} \otimes \mathbf{C})^{-1}(vec(\mathbf{Y}) - (\mathbf{I} \otimes \mathbf{X})vec(\theta))$
becomes

$$-\frac{1}{2}vec(\mathbf{Y} - \mathbf{X}\theta)^T(\mathbf{V} \otimes \mathbf{C})^{-1}vec(\mathbf{Y} - \mathbf{X}\theta)$$

Using $vec(\mathbf{Y})^T(\mathbf{V} \otimes \mathbf{C})^{-1}vec(\mathbf{Y}) = Tr[\mathbf{C}^{-1}\mathbf{Y}^T\mathbf{V}^{-1}\mathbf{Y}]$

$-\frac{1}{2}vec(\mathbf{Y} - \mathbf{X}\theta)^T(\mathbf{V} \otimes \mathbf{C})^{-1}vec(\mathbf{Y} - \mathbf{X}\theta) = -\frac{1}{2}Tr[\mathbf{C}^{-1}(\mathbf{Y} - \mathbf{X}\theta)^T\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\theta)]$
Also is the fact that
$-\frac{1}{2}\log|(\mathbf{I} \otimes \mathbf{X^T})(\mathbf{V} \otimes \mathbf{C})^{-1}(\mathbf{I} \otimes \mathbf{X})| = -\frac{1}{2}\log|\mathbf{V}^{-1} \otimes \mathbf{X}^T\mathbf{C}^{-1}\mathbf{X}|$

Using $|\mathbf{V} \otimes \mathbf{C}| = |\mathbf{V}|^{rank(\mathbf{C})}|\mathbf{C}|^{rank(\mathbf{V})}$

$= \frac{1}{2}\log|(\mathbf{V}^{-1} \otimes \mathbf{C})^{-1}| - \frac{1}{2}\log|(\mathbf{V} \otimes \mathbf{X}^T\mathbf{C}^{-1}\mathbf{X})|$
$= \frac{1}{2}rank(\mathbf{V})\log|(\mathbf{C}^{-1})| - \frac{1}{2}rank(\mathbf{V})\log|(\mathbf{X}^T\mathbf{C}^{-1}\mathbf{X})| + Const$
Finally, the $\mathcal{L}$ functional becomes
$\frac{1}{2}\log|(\mathbf{C}^{-1})| - \frac{1}{2}\log|(\mathbf{X}^T\mathbf{C}^{-1}\mathbf{X})| - \frac{1}{2}Tr[\mathbf{C}^{-1}(\mathbf{Y} - \mathbf{X}\theta)\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\theta)^T]/rank(\mathbf{V}) + Const$

The maximum of the $\mathcal{L}$ can be found by replacing

$$vec(\mathbf{Y})vec(\mathbf{Y})^T \longrightarrow \mathbf{Y}\mathbf{V}^{-1}\mathbf{Y}^T/rank(\mathbf{V}) \text{ and } \mathbf{V}\otimes\mathbf{Q} \longrightarrow \mathbf{Q}$$

$$\lambda = \max_{\lambda} \mathcal{L}(\mathbf{Y}|\lambda,\mathbf{Q}) = ReML(vec(\mathbf{Y}), vec(\mathbf{Y})^T, \mathbf{V}\otimes\mathbf{Q}) = ReML(\mathbf{Y}\mathbf{V}^{-1}\mathbf{Y}^T/rank(\mathbf{V}), \mathbf{Q})$$

which means we work with much smaller $n_e \times n_e$ matrices instead of $n_e n_t \times n_e n_t$.

ReML Algorithm

Input : $\mathbf{Y}\mathbf{V}^{-1}\mathbf{Y}^T/rank(\mathbf{V})$, $\{\mathbf{Q}_1^{(1)}, \mathbf{Q}_2^{(1)}, \ldots, \mathbf{X}\mathbf{Q}_1^{(2)}\mathbf{X}^T, \mathbf{X}\mathbf{Q}_2^{(2)}\mathbf{X}^T, \ldots\}$ Output : $\mathbf{C}_Y$

$$\mathbf{C} = \sum_i \lambda_i \mathbf{Q}_i = \lambda_1^{(1)}\mathbf{Q}_1^{(1)} + \lambda_2^{(1)}\mathbf{Q}_2^{(1)} + \lambda_1^{(2)}\mathbf{X}\mathbf{Q}_1^{(2)}\mathbf{X}^T + \lambda_2^{(2)}\mathbf{X}\mathbf{Q}_2^{(2)}\mathbf{X}^T + \ldots$$

$$\mathbf{C}_Y = \mathbf{V}\otimes\mathbf{C},$$

$$\mathbf{C}_{\theta|Y} = (\mathbf{X}^T\mathbf{C}_Y^{-1}\mathbf{X})^{-1}$$

$$\mathbf{P} = \mathbf{C}_Y^{-1} - \mathbf{C}_Y^{-1}\mathbf{X}\mathbf{C}_{\theta|Y}^{-1}\mathbf{X}^T\mathbf{C}_Y^{-1}$$

$$g_i = -\frac{1}{2}Tr[\mathbf{P}\mathbf{Q_i}] + \frac{1}{2}Tr[\mathbf{P}\mathbf{Y}\mathbf{V}^{-1}\mathbf{Y}^T/rank(\mathbf{V})\mathbf{P}^T\mathbf{Q}_i] \qquad H_{ij} = \frac{1}{2}Tr[\mathbf{P}\mathbf{Q}_i\mathbf{P}\mathbf{Q_j}]$$

Update λ until convergence

$$\lambda \longrightarrow \lambda + \mathbf{H}^{-1}\mathbf{g}$$

Temporally Informed Scheme

- Although not a full spatio-temporal model, a relaxation of the assumption of $\mathbf{V}^{(1)} = \mathbf{V}^{(2)}$, is to determine a projection matrix $\mathbf{S}$ so that $\mathbf{S}^T\mathbf{V}^{(2)}\mathbf{S} = \mathbf{S}^T\mathbf{V}^{(1)}\mathbf{S}$ which makes the temporal correlation of signal and noise are equivalent by projecting them onto a subspace spanned by the columns of $\mathbf{S}$.
- This projection removes high-frequency noise components so that the remaining smooth components exhibit the same correlations as signal.

Assuming $\mathbf{V}^{(1)} = \mathbf{V}$ for simplicity of notation

$$\begin{aligned}
\mathbf{Y} &= \mathbf{X}\beta\mathbf{S}^T + \epsilon^{(1)} \\
\beta &= \epsilon^{(2)} \\
Cov(vec(\epsilon(1))) &= \mathbf{V} \otimes \mathbf{C}^{(1)} \\
Cov(vec(\epsilon(2))) &= \mathbf{S}^T\mathbf{V}\mathbf{S} \otimes \mathbf{C}^{(2)}
\end{aligned}$$

where $\beta = \theta\mathbf{S}$

If we assume that the temporal covariance of sensors and sources are the same then

$Cov(\beta) = Cov(\theta\mathbf{S}) = E[\mathbf{S}^T\theta^T\theta\mathbf{S}] = \mathbf{S}^T Cov(\theta)\mathbf{S} = \mathbf{S}^T\mathbf{V}\mathbf{S}$

The sources are estimated in terms of the activity β of temporal modes.

The orthonormal columns of the temporal basis set $\mathbf{S}$ define these modes, where $\mathbf{S}^T\mathbf{S} = \mathbf{I}_r$ and $r < n_t$.

Temporal priors on the sources are $\mathbf{V}^{(2)} = \mathbf{S}\mathbf{S}^T\mathbf{V}^{(1)}\mathbf{S}\mathbf{S}^T$ ensuring that $\mathbf{S}^T\mathbf{V}^{(2)}\mathbf{S} = \mathbf{S}^T\mathbf{V}^{(1)}\mathbf{S}$.

$$
\begin{aligned}
\mathbf{YS} &= \mathbf{X}\beta + \epsilon^{(S)} \\
\beta &= \epsilon^{(2)} \\
Cov(vec(\epsilon^{(S)})) &= \mathbf{S}^T\mathbf{VS} \otimes \mathbf{C}^{(1)} \\
Cov(vec(\epsilon^{(2)})) &= \mathbf{S}^T\mathbf{VS} \otimes \mathbf{C}^{(2)}
\end{aligned}
\quad \text{where } \epsilon^{(S)} = \epsilon^{(1)}\mathbf{S}
$$

Temporal correlations of signal and noise are the same.

This is a problem that can be solved by the REML algorithm using the data covariance $\mathbf{YS}(\mathbf{S}^T\mathbf{VS})^{-1}\mathbf{S}^T\mathbf{Y}^T$ and $\mathbf{C} = \sum_i \lambda_i \mathbf{Q}_i$

$\lambda = ReML(\frac{1}{r}\mathbf{YS}(\mathbf{S}^T\mathbf{VS})^{-1}\mathbf{S}^T\mathbf{Y}^T, \mathbf{Q})$

Remembering that the posterior mean and covariance of θ are

$$\begin{aligned}
\hat{\Sigma} &= \mathbf{V} \otimes \hat{\mathbf{C}} \\
\hat{\theta} &= \mathbf{MY} \\
\mathbf{M} &= \mathbf{C}^{(2)}\mathbf{X}^T\mathbf{C}^{-1} \\
\mathbf{C} &= \mathbf{C}^{(1)} + \mathbf{XC}^{(2)}\mathbf{X}^T \\
\hat{\mathbf{C}} &= (\mathbf{X}^T\mathbf{C}^{(1)-1}\mathbf{X} + \mathbf{C}^{(2)-1})^{-1}
\end{aligned}$$

we can use λ_i determined from ReML algorithm to compute the conditional moments of the sources as a function of time

$$\begin{aligned}
\hat{\theta} &= \hat{\beta}\mathbf{S}^T = \mathbf{MYSS}^T \\
\mathbf{C} &= \hat{\mathbf{V}} \otimes \hat{\mathbf{C}} \\
\hat{\mathbf{V}} &= \mathbf{SS}^T\mathbf{VSS}^T
\end{aligned}$$

$\hat{\mathbf{V}}$ is the temporal covariance of source space and is
$\hat{\mathbf{V}} = Cov[\hat{\theta}] = Cov[\beta\mathbf{S}^T] = E[\mathbf{S}\beta^T\beta\mathbf{S}^T] = \mathbf{S}E[\beta^T\beta]\mathbf{S}^\mathbf{T} = \mathbf{SS}^T\mathbf{VSS}^\mathbf{T}$
The temporal correlations $\hat{\mathbf{V}}$ are rank deficient and non-stationary, because the conditional responses do not span the null space of $\mathbf{S}$.

Estimating Response Energy

If the response is projected onto a basis defined by the columns of $\mathbf{W}$, the energy of the i^{th} source is

$$\theta_{i,\bullet}\mathbf{W}\mathbf{W}^T\theta_{i,\bullet}^T$$

The columns of $\mathbf{W}$ may be wavelet functions like windowed sine-cosine pairs at particular frequencies or Fourier components.

The conditional density of $\theta_{i,\bullet}$ is

$$\begin{aligned} p(\theta_{i,\bullet}|\mathbf{Y},\lambda) &= \mathcal{N}(\hat{\theta}_{i,\bullet},\hat{\mathbf{C}}_{ii}\hat{\mathbf{V}}) \\ \hat{\theta}_{i,\bullet} &= \mathbf{M}_{i,\bullet}\mathbf{Y}\mathbf{S}\mathbf{S}^T \end{aligned}$$

and the conditional expectation of the energy is:

$$\begin{aligned} E_p[\theta_{i,\bullet}\mathbf{W}\mathbf{W}^T\theta_{i,\bullet}^T] &= Tr[\mathbf{W}\mathbf{W}^T E_p[\theta_{i,\bullet}^T\theta_{i,\bullet}]] = Tr[\mathbf{W}\mathbf{W}^T[\hat{\theta}_{i,\bullet}^T\hat{\theta}_{i,\bullet} + \hat{\mathbf{C}}_{ii}\hat{\mathbf{V}}]] \\ &= \mathbf{M}_{i,\bullet}\mathbf{Y}\mathbf{G}\mathbf{Y}^T\mathbf{M}_{i,\bullet}^T + \hat{\mathbf{C}}_{ii}Tr[\mathbf{G}\mathbf{V}] \\ \text{where } \mathbf{G} &= \mathbf{S}\mathbf{S}^T\mathbf{W}\mathbf{W}^T\mathbf{S}\mathbf{S}^T \end{aligned}$$

This is a function of $\mathbf{Y}\mathbf{G}\mathbf{Y}^T$ *i.e.* energy in the channel space

Conditional expectation of energy over all sources

$$\hat{\mathbf{E}} = E_p[\theta \mathbf{W}\mathbf{W}^T\theta^T] = \mathbf{MYGY}^T\mathbf{M}^T + \hat{\mathbf{C}}\,Tr[\mathbf{GV}]$$

1. Diagonal terms of $\hat{\mathbf{E}}$: Energy at the corresponding source
2. Off-diagonal terms of $\hat{\mathbf{E}}$: Cross energy (e.g. cross-spectral density or coherence).
3. Conditional energy has two components, one attributable to the energy in the conditional mean and one related to the conditional covariance. When conditional uncertainty is high, the priors shrink the conditional mean of the sources towards zero. This results in an underestimate of energy based solely on the conditional expectations of the sources. By including the second term, the energy estimator becomes unbiased.

Multi-trial Models

If the trials are concatenated as $\bar{\mathbf{Y}} = [\mathbf{Y}_1 \;\; \mathbf{Y}_2 \;\; \ldots \mathbf{Y}_n]$,
the model for n trials:

$$
\begin{aligned}
\bar{\mathbf{Y}} &= \mathbf{X}\beta^{(1)}(\mathbf{I_n} \otimes \mathbf{S})^T + \epsilon^{(1)} \\
\beta^{(1)} &= (\mathbf{1}_n \otimes \beta^{(2)}) + \epsilon^{(2)} \\
\beta^{(2)} &= \epsilon^{(3)} \\
Cov[vec(\epsilon^{(1)})] &= \mathbf{I}_n \otimes \mathbf{V} \otimes \mathbf{C}^{(1)} \\
Cov[vec(\epsilon^{(2)})] &= \mathbf{I}_n \otimes \mathbf{S}^T\mathbf{V}\mathbf{S} \otimes \mathbf{C}^{(2)} \\
Cov[vec(\epsilon^{(3)})] &= \mathbf{S}^T\mathbf{V}\mathbf{S} \otimes \mathbf{C}^{(3)}
\end{aligned}
$$

$\mathbf{1}_n = [1\,1\ldots\,1] : 1 \times n$ vector.

- $\beta^{(2)}$: A component common to all trials related to evoked response
- $\epsilon^{(2)}$: A trial specific component related to induced response

$\beta^{(1)}$: Source space response can be partitioned into an evoked

$$\beta^e = \beta^{(1)}(\mathbf{1}_n^- \otimes \mathbf{I}_r) = \beta^{(2)} + \epsilon^{(2)}(\mathbf{1}_n^- \otimes \mathbf{I}_r)$$

and orthogonal to it, an induced response

$$\beta^i = \beta^{(1)}((\mathbf{I}_n - \mathbf{1}_n^-\mathbf{1}_n) \otimes \mathbf{I}_r)$$

Working Example

```
n=3; r =2;
beta2 = [10 20 ; 10 20];
epsil  = [2 2 3 3 4 4; 2 2 3 3 4 4 ]
%evoked part
kron(ones(1,n),beta2)*kron(pinv(ones(1,n)), eye(r,r))
epsil*kron(pinv(ones(1,n)), eye(r,r))
% induced part
kron(ones(1,n),beta2)*kron((eye(n,n)-pinv(ones(1,n))*ones(1,n)),eye(r,r))
epsil*kron((eye(n,n)-pinv(ones(1,n))*ones(1,n)),eye(r,r))
% Trial Averaging Ytilde = Ybar (1_n^- otimes I_t)
t=2;
Y_bar = [1 2 3 4 ; 1 2 3 4 ;1 2 3 4 ; 1 2 3 4 ; 1 2 3 4]
Y_bar = ones(5,1)*kron([1:n],[1 1 ])
Y_tilde = Y_bar*kron(pinv(ones(1,n)),eye(t))
```

$$
\begin{aligned}
\bar{\mathbf{Y}} &= \mathbf{X}\beta^{(1)}(\mathbf{I_n} \otimes \mathbf{S})^T + \epsilon^{(1)} \\
\beta^{(1)} &= (\mathbf{1}_n \otimes \beta^{(2)}) + \epsilon^{(2)} \\
\beta^{(2)} &= \epsilon^{(3)} \\
Cov[vec(\epsilon^{(1)})] &= \mathbf{I}_n \otimes \mathbf{V} \otimes \mathbf{C}^{(1)} \\
Cov[vec(\epsilon^{(2)})] &= \mathbf{I}_n \otimes \mathbf{S}^T\mathbf{V}\mathbf{S} \otimes \mathbf{C}^{(2)} \\
Cov[vec(\epsilon^{(3)})] &= \mathbf{S}^T\mathbf{V}\mathbf{S} \otimes \mathbf{C}^{(3)}
\end{aligned}
$$

$$
\beta^e = \beta^{(1)}(\mathbf{1}_n^- \otimes \mathbf{I}_r) = \beta^{(2)} + \epsilon^{(2)}(\mathbf{1}_n^- \otimes \mathbf{I}_r)
$$

Evoked Response

The multi-trial data is transformed into a spatiotemporally separable form by averaging the data and projecting onto the signal subspace $\tilde{\mathbf{Y}} = \bar{\mathbf{Y}}(\mathbf{1}_n^- \otimes \mathbf{I}_t)$

$$
\begin{aligned}
\bar{\mathbf{Y}}(\mathbf{1}_n^- \otimes \mathbf{S}) = \tilde{\mathbf{Y}}\mathbf{S} &= \mathbf{X}\beta^{(e)} + \tilde{\epsilon}^{(1)} & \tilde{\epsilon}^{(1)} &= \epsilon^{(1)}(\mathbf{1}_n^- \otimes \mathbf{S}) \\
\beta^{(e)} &= \epsilon^{(e)} & \beta^{(e)} &= \beta^{(1)}(\mathbf{I}_n \otimes \mathbf{S})^T(\mathbf{1}_n^- \otimes \mathbf{S}) = \beta^{(1)}(\mathbf{1}_n^- \otimes \mathbf{I}_r) \\
Cov[vec(\tilde{\epsilon}^{(1)})] &= \mathbf{S}^T\mathbf{V}\mathbf{S} \otimes \tilde{\mathbf{C}}^{(1)} & \beta^{(e)} &= \epsilon^{(2)}(\mathbf{1}_n^- \otimes \mathbf{I}_r) + \beta^{(2)} \\
Cov[vec(\epsilon^{(e)})] &= \mathbf{S}^T\mathbf{V}\mathbf{S} \otimes \mathbf{C}^{(e)} \\
\tilde{\mathbf{C}}^{(1)} &= \frac{1}{n}\mathbf{C}^{(1)} \\
\mathbf{C}^{(e)} &= \frac{1}{n}\underbrace{\mathbf{C}^{(2)}}_{\text{trial specific}} + \underbrace{\mathbf{C}^{(3)}}_{\text{nonspecific}}
\end{aligned}
$$

REML estimation of $\tilde{\mathbf{C}}^{(1)}$ and $\mathbf{C}^{(e)}$ is possible by providing the data covariance $\tilde{\mathbf{Y}}\mathbf{S}(\mathbf{S}^T\mathbf{V}\mathbf{S})^{-1}\mathbf{S}^T\tilde{\mathbf{Y}}^T$ and the covariance components $\mathbf{Q}_i$ for $\tilde{\mathbf{C}}^{(1)} = \sum \lambda_i^{(1)}\mathbf{Q}_i^{(1)}$ and $\mathbf{C}^{(e)} = \sum \lambda_i^{(e)}\mathbf{Q}_i^{(e)}$, *i.e.*

$$\lambda = ReML(\frac{1}{r}\mathbf{Y}(\mathbf{S}(\mathbf{S}^T\mathbf{V}\mathbf{S})^{-1}\mathbf{S}^T)\mathbf{Y}^T, \mathbf{Q})$$

$$\begin{aligned} \hat{\beta}^{(e)} &= \mathbf{M}\tilde{\mathbf{Y}}\mathbf{S}\mathbf{S}^T \\ \mathbf{M} &= \mathbf{C}^{(e)}\mathbf{X}^T(\mathbf{X}\mathbf{C}^{(e)}\mathbf{X}^T + \bar{\mathbf{C}}^{(1)})^{-1} \end{aligned}$$

The conditional expectation of evoked power:

$$\begin{aligned} \hat{\mathbf{E}}^{(e)} &= \mathbf{M}\mathbf{E}_y^{(e)}\mathbf{M}^T + \hat{\mathbf{C}}\,Tr[\mathbf{G}\mathbf{V}] \\ \mathbf{E}_y^{(e)} &= \tilde{\mathbf{Y}}\mathbf{G}\tilde{\mathbf{Y}} \text{ and } \hat{\mathbf{C}} = (\mathbf{X}^T\tilde{\mathbf{C}}^{(1)-1}\mathbf{X} + \mathbf{C}^{(e)-1})^{-1} \end{aligned}$$

$$
\begin{aligned}
\bar{\mathbf{Y}} &= \mathbf{X}\beta^{(1)}(\mathbf{I_n} \otimes \mathbf{S})^T + \epsilon^{(1)} \\
\beta^{(1)} &= (\mathbf{1}_n \otimes \beta^{(2)}) + \epsilon^{(2)} \\
\beta^{(2)} &= \epsilon^{(3)} \\
Cov[vec(\epsilon^{(1)})] &= \mathbf{I}_n \otimes \mathbf{V} \otimes \mathbf{C}^{(1)} \\
Cov[vec(\epsilon^{(2)})] &= \mathbf{I}_n \otimes \mathbf{S}^T\mathbf{V}\mathbf{S} \otimes \mathbf{C}^{(2)} \\
Cov[vec(\epsilon^{(3)})] &= \mathbf{S}^T\mathbf{V}\mathbf{S} \otimes \mathbf{C}^{(3)}
\end{aligned}
$$

$$\beta^i = \beta^{(1)}((\mathbf{I}_n - \mathbf{1}_n^- \mathbf{1}_n) \otimes \mathbf{I}_r)$$

Induced Response

Induced response is obtained by subtracting the evoked response from all trials: $\hat{\mathbf{Y}} = \bar{\mathbf{Y}}((\mathbf{I}_n - \mathbf{1}_n^- \mathbf{1}_n) \otimes \mathbf{I}_t)$ and projecting it onto the signal subspace.

The multi-trial model

$$
\begin{aligned}
\hat{\mathbf{Y}}(\mathbf{I}_n \otimes \mathbf{S}) &= \mathbf{X}\beta^{(i)} + \hat{\epsilon}^{(1)}, \quad \hat{\epsilon}^{(1)} = \epsilon^{(1)}((\mathbf{I}_n - \mathbf{1}_n^- \mathbf{1}_n) \otimes \mathbf{I}_t)(\mathbf{I}_n \otimes \mathbf{S}) \\
& \qquad \hat{\epsilon}^{(1)} = \epsilon^{(1)}(\mathbf{I}_n - \mathbf{1}_n^- \mathbf{1}_n) \otimes \mathbf{S} \\
\beta^{(i)} &= \epsilon^{(i)}, \quad \beta^i = \beta^{(1)}(\mathbf{I}_n \otimes \mathbf{S})^T((\mathbf{I}_n - \mathbf{1}_n^- \mathbf{1}_n) \otimes \mathbf{I}_t)(\mathbf{I}_n \otimes \mathbf{S}) \\
& \qquad \beta^i = \beta^{(1)}((\mathbf{I}_n - \mathbf{1}_n^- \mathbf{1}_n) \otimes \mathbf{I}_r) \\
Cov[vec(\hat{\epsilon}^{(1)})] &= \mathbf{I}_n \otimes \mathbf{S}^T\mathbf{V}\mathbf{S} \otimes \hat{\mathbf{C}}^{(1)} \\
Cov[vec(\epsilon^{(i)})] &= \mathbf{I}_n \otimes \mathbf{S}^T\mathbf{V}\mathbf{S} \otimes \mathbf{C}^{(i)}
\end{aligned}
$$

$\beta^{(i)}$ is a large $n_s \times n \cdot r$ matrix that covers all trials.

ReML algorithm can be used by providing the covariance of mean-corrected data, $\hat{\mathbf{Y}}$, whitened and averaged over trials, $\hat{\mathbf{Y}}(\mathbf{I}_n \otimes \mathbf{S}(\mathbf{S}^T\mathbf{V}\mathbf{S})^{-1}\mathbf{S}^T)\hat{\mathbf{Y}}$ and the covariance components $\mathbf{Q}_i$ of $\hat{\mathbf{C}}^{(1)} = \sum \lambda_i^{(1)}\mathbf{Q}_i^{(1)}$ and $\mathbf{C}^{(i)} = \sum \lambda_i^{(2)}\mathbf{Q}_i^{(2)}$ to estimate the $\lambda_i^{(1)}$ and $\lambda_i^{(2)}$

$ReML(\frac{1}{n \cdot r}\hat{\mathbf{Y}}(\mathbf{I}_n \otimes \mathbf{S}(\mathbf{S}^T\mathbf{V}\mathbf{S})^{-1}\mathbf{S}^T)\hat{\mathbf{Y}}, \mathbf{Q})$
and to determine the conditional mean and covariance of sources

$$\begin{aligned} \hat{\beta}^{(i)} &= \mathbf{M}\hat{\mathbf{Y}}\mathbf{S}\mathbf{S}^T \\ \mathbf{M} &= \mathbf{C}^{(i)}\mathbf{X}^T(\mathbf{X}\mathbf{C}^{(i)}\mathbf{X}^T + \hat{\mathbf{C}}^{(1)})^{-1} \end{aligned}$$

The conditional expectation of induced energy, per trial,
$\hat{\mathbf{E}}^{(i)} = \frac{1}{n}\mathbf{M}\hat{\mathbf{Y}}(\mathbf{I}_n \otimes \mathbf{G})\hat{\mathbf{Y}}\mathbf{M}^T) + \frac{1}{n}\hat{\mathbf{C}}\,Tr[\mathbf{I}_n \otimes \mathbf{G}\mathbf{V}] = \mathbf{M}\mathbf{E}_y^{(i)}\mathbf{M} + \hat{\mathbf{C}}\,Tr[\mathbf{G}\mathbf{V}]$
$\mathbf{E}_y^{(i)} = \frac{1}{n}\hat{\mathbf{Y}}(\mathbf{I}_n \otimes \mathbf{G})\hat{\mathbf{Y}}^T$ where $\mathbf{E}_y^{(i)}$ is the induced cross-energy per trial, in channel space and
$\hat{\mathbf{C}} = (\mathbf{X}^T\hat{\mathbf{C}}^{(1)-1}\mathbf{X} + \mathbf{C}^{(i)-1})^{-1}$

A Toy Example

- Data is generated with $n_v = 128$ sources, $n_e = 16$ electrode channels and $n_t = 64$ time points.
- Lead field matrix $\mathbf{X}$ is a 2-D $n_e \times n_v$ Gaussian random matrix.
- The spatial covariance component of sensors is $\mathbf{Q}_1^{(1)} = \mathbf{I}_{n_e}$.
- The spatial covariance component of sources are $\mathbf{Q}_1^{(2)} = \mathbf{D}\mathbf{D}^T$ and $\mathbf{Q}_2^{(2)} = \mathbf{D}\mathbf{F}\mathbf{D}^T$ where $\mathbf{D}$ is a spatial convolution using a Gaussian kernel with a standard deviation of 4. This is a smoothness constraint. $\mathbf{F}$ represents structural or functional MRI constraints with a leading diagonal matrix encoding the prior probability of a source at each voxel. It is chosen randomly by smoothing a random Gaussian sequence raised to the power 4.
- Sensor space temporal noise covariance is *i.i.d.*, $\mathbf{V}^{(1)} = \mathbf{V} = \mathbf{I}_{n_t}$.
- Signal subspace in time $\mathbf{S}$ is obtained by $r = 8$ principle eigenvectors of a Gaussian autocorrelation matrix of standard deviation 2, windowed with a function of peristimulus time $t^2 e^{-t/8}$.
- Prior temporal correlation structure of the sources $\mathbf{V}^{(2)} = \mathbf{S}\mathbf{S}^T\mathbf{V}\mathbf{S}\mathbf{S}^T$ which are smooth and restricted to earlier time bins by the window function.
- The hyperparameters are chosen to emphasize the MRI priors $\lambda = [\lambda^{(1)}{}_1 \lambda^{(2)}{}_1 \lambda_2^{(2)}] = [1\ 0\ 8]$ providing an $SNR = \sum_{n_e} \frac{\sigma_y}{\sigma_e} \approx 1$

Toy example for EEG Source Reconstruction

```
nv=128; ne=16; nt=64; r=8 % dimensions of source, sensor, time and signal subspace
X = randn(ne,nv); % Lead field matrix
C1 = eye(ne); % spatial covariance of sensor space data
V1 = eye(nt); % temporal covariance of sensor space data
T = [0:nv-1]'*ones(1,nv) - ones(nv,1)*([0:nv-1]) ;
D = exp(-(1/(2*16))*(T).^2);
F = randn(nv,1) ;   % MRI priors
F = (sgolayfilt(F.^4,7,9)); % filtered version
F = diag(F);
Q2_1 = D*D';     Q2_1 = Q2_1/max(Q2_1(:)); Q2_2 = D*F*D';  Q2_2 = Q2_2/max(Q2_2(:));
subplot(3,2,1);  imagesc(Q2_1); colormap gray
subplot(3,2,3);  imagesc(Q2_2); colormap gray
t  =  [0:nt-1]'; T =  t*ones(1,nt) - ones(nt,1)*t';
S = exp(-(1/(2*4))*(T).^2).*( t.^2.*exp(-t/8)*ones(1,nt) ); % Temporal subspace for the sources
[v,d] = eig(S);  S = v(:,1:r);
V2 = S*S'*V1*S*S'; % temporal covariance of source space data
subplot(3,2,5);  imagesc(V2); colormap gray
lambda = [1 0 1]; % lambda1_1 lambda_2_1 lambda_2_2
C2 = lambda(2)*Q2_1 + lambda(3)*Q2_2 ; % spatial covariance matrix of sources
I_A_C2 = spm_sqrtm(C2); % spatial coloring matrix for source space
I_A_V2 = spm_sqrtm(S'*V1*S); % temporal  coloring matrix for source space
theta = I_A_C2 * randn(nv,r) * I_A_V2'; % spatio-temporal data in source space
I_A_C1 = spm_sqrtm(C1); % spatial coloring matrix for sensor space
I_A_V1 = spm_sqrtm(V1);  % temporal coloring matrix for sensor space
e = I_A_C1 * randn(ne,nt) * I_A_V1'; % spatio-temporal noise in sensor space
theta = theta *S'; Y0 = X*theta ; Y = Y0 + e ;
subplot(3,2,2); plot(theta'); subplot(3,2,4); plot(Y0');
subplot(3,2,6); plot(Y'); SNR = mean(std(Y0,'',2)./std(e,'',2)) % SNR
text(50,10,['SNR = ' num2str(SNR)]);
```

ReML Algorithm (Augmented Spatiotemporal Model)

Remembering that the augmented form of the two level hierarchical spatial model

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\theta + \epsilon^{(1)} \\ \theta &= \mathbf{0} + \epsilon^{(2)} \end{aligned}$$

where $\epsilon^{(1)} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\epsilon^{(1)}})$ and $\epsilon^{(2)} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\epsilon^{(2)}})$
is $\bar{\mathbf{y}} = [\mathbf{y}\ \mathbf{0}]^T$ and $\bar{\mathbf{X}} = \begin{bmatrix} \mathbf{X} \\ \mathbf{I} \end{bmatrix}$ and $\mathbf{C}_\epsilon = \begin{bmatrix} \mathbf{C}_{\epsilon^{(1)}} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_{\epsilon^{(2)}} \end{bmatrix}$
we can determine the posterior covariance and mean of θ as,

$$\mathbf{C}_{\theta|y} = (\bar{\mathbf{X}}^T \mathbf{C}_\epsilon^{-1} \bar{\mathbf{X}})^{-1} \text{ and } \eta_{\theta|y} = \mathbf{C}_{\theta|y}(\bar{\mathbf{X}}^T \mathbf{C}_\epsilon^{-1} \bar{\mathbf{y}})$$

In the two level hierarchical spatio-temporal model

$$\begin{aligned} vec(\mathbf{Y}) &= (\mathbf{I} \otimes \mathbf{X}) vec(\theta) + vec(\epsilon^{(1)}) \\ vec(\theta) &= vec(\epsilon^{(2)}) \end{aligned}$$

$vec(\epsilon^{(1)}) \sim \mathcal{N}(\mathbf{0}, \mathbf{V}^{(1)} \otimes \mathbf{C}_{\epsilon^{(1)}})$, $vec(\epsilon^{(2)}) \sim \mathcal{N}(\mathbf{0}, \mathbf{V}^{(2)} \otimes \mathbf{C}_{\epsilon^{(2)}})$,
the augmented form yields $\mathbf{X} \rightarrow \mathbf{I} \otimes \bar{\mathbf{X}}$ and $\mathbf{C}_\epsilon \rightarrow \begin{bmatrix} \mathbf{V}^{(1)} \otimes \mathbf{C}_{\epsilon^{(1)}} & \mathbf{0} \\ \mathbf{0} & \mathbf{V}^{(2)} \otimes \mathbf{C}_{\epsilon^{(2)}} \end{bmatrix}$

Augmented Spatiotemporal Model ReML Formulation

Input : $\{vec(\mathbf{Y})vec(\mathbf{Y}^T)\}$ and $\{\mathbf{Q}_1^{(1)}, \mathbf{Q}_2^{(1)}, \ldots, \mathbf{Q}_1^{(2)}, \mathbf{Q}_2^{(2)}, \ldots\}$ Output : $\mathbf{C}_\epsilon$

$$\mathbf{C}_\epsilon = \begin{bmatrix} \mathbf{V}^{(1)} \otimes (\lambda_1^{(1)}\mathbf{Q}_1^{(1)} + \lambda_2^{(1)}\mathbf{Q}_2^{(1)} + \ldots) & \mathbf{0} \\ \mathbf{0} & \mathbf{V}^{(2)} \otimes (\lambda_1^{(2)}\mathbf{Q}_1^{(2)} + \lambda_2^{(2)}\mathbf{Q}_2^{(2)} + \ldots) \end{bmatrix}$$

$\mathbf{C}_{\theta|y} = (\bar{\mathbf{X}}^T\mathbf{C}_\epsilon^{-1}\bar{\mathbf{X}})^{-1}$ and $\eta_{\theta|y} = \mathbf{C}_{\theta|y}(\bar{\mathbf{X}}^T\mathbf{C}_\epsilon^{-1}\bar{\mathbf{y}})$

$\mathbf{P} = \mathbf{C}_\epsilon^{-1} - \mathbf{C}_\epsilon^{-1}\bar{\mathbf{X}}\mathbf{C}_{\theta|y}^{-1}\bar{\mathbf{X}}^T\mathbf{C}_\epsilon^{-1}$

$g_i = -\frac{1}{2}Tr[\mathbf{P}\mathbf{Q_i}] + \frac{1}{2}Tr[\mathbf{P}vec(\mathbf{Y})vec(\mathbf{Y})^T\mathbf{P}^T\mathbf{Q}_i]$

$= -\frac{1}{2}Tr[\mathbf{F}_i\mathbf{R}(vec(\mathbf{Y})vec(\mathbf{Y})^T\mathbf{R}^T - \mathbf{C}_\epsilon)]$

$H_{ij} = \frac{1}{2}Tr[\mathbf{P}\mathbf{Q}_i\mathbf{P}\mathbf{Q_j}] = \frac{1}{2}Tr[\mathbf{F}_i\mathbf{R}\mathbf{C}_\epsilon\mathbf{F}_j\mathbf{R}\mathbf{C}_\epsilon]$

$\lambda \longleftarrow \lambda + \mathbf{H}^{-1}\mathbf{g}$

where $\mathbf{F}_i = -\frac{\partial \mathbf{C}_\epsilon^{-1}}{\partial \lambda_i} = -\mathbf{C}_\epsilon^{-1}\mathbf{Q}_i\mathbf{C}_\epsilon^{-1}$

and $\mathbf{R} = \mathbf{I} - \bar{\mathbf{X}}(\bar{\mathbf{X}}^T\mathbf{C}_\epsilon^{-1}\bar{\mathbf{X}})^{-1}\bar{\mathbf{X}}^T\mathbf{C}_\epsilon^{-1}$

Not necessary

If multiple independent observations are arranged as $\mathbf{Y} = [\mathbf{y}_1\ \mathbf{y}_2\ \ldots \mathbf{y}_r]$

$$\begin{aligned} vec(\mathbf{Y}) &= (\mathbf{I} \otimes \mathbf{X})vec(\theta) + vec(\epsilon^{(1)}) \\ vec(\theta) &= vec(\epsilon^{(2)}) \end{aligned}$$

$\mathbf{F}_i \rightarrow \mathbf{I} \otimes \mathbf{F}_i, \quad \mathbf{X} \rightarrow \mathbf{I} \otimes \mathbf{X}, \quad \mathbf{y} \rightarrow vec(\mathbf{Y}) \quad \mathbf{R} \rightarrow \mathbf{I} \otimes \mathbf{R},$

Input : $vec(\mathbf{Y})vec(\mathbf{Y})^T,\ \{\mathbf{Q}_1^{(1)}, \mathbf{Q}_2^{(1)}, \ldots, \mathbf{Q}_1^{(2)}, \mathbf{Q}_2^{(2)}, \ldots\}$ Output : $\mathbf{C}_\epsilon$

- $g_i = -\frac{1}{2} Tr[(\mathbf{I} \otimes \mathbf{F}_i)(\mathbf{I} \otimes \mathbf{R})(vec(\mathbf{Y})vec(\mathbf{Y})^T(\mathbf{I} \otimes \mathbf{R})^T - \mathbf{I} \otimes \mathbf{C}_\epsilon)]$
  $= -\frac{r}{2} Tr[\mathbf{F}_i\mathbf{R}(\frac{1}{r}\mathbf{Y}\mathbf{Y}^T\mathbf{R}^T - \mathbf{C}_\epsilon)]$
- $\mathbf{H}_{ij} = \frac{1}{2} Tr[(\mathbf{I} \otimes \mathbf{F}_i)(\mathbf{I} \otimes \mathbf{R})(\mathbf{I} \otimes \mathbf{C}_\epsilon)(\mathbf{I} \otimes \mathbf{F}_j)(\mathbf{I} \otimes \mathbf{R})(\mathbf{I} \otimes \mathbf{C}_\epsilon)]$
  $= \frac{1}{2} Tr[\mathbf{I} \otimes \mathbf{F}_i\mathbf{R}\mathbf{C}_\epsilon\mathbf{F}_j\mathbf{R}\mathbf{C}_\epsilon] = \frac{r}{2} Tr[\mathbf{F}_i\mathbf{R}\mathbf{C}_\epsilon\mathbf{F}_j\mathbf{R}\mathbf{C}_\epsilon]$
- $\lambda \longleftarrow \lambda + \mathbf{H}^{-1}\mathbf{g}$

Self-Study Question

If multiple independent observations model is used, show that the following assignments are valid:

$$a)\ \mathbf{F}_i \longrightarrow \mathbf{I} \otimes \mathbf{F}_i,\ b)\ \mathbf{R} \longrightarrow \mathbf{I} \otimes \mathbf{R}.$$

ReML Algorithm

$\mathbf{C} = \sum_i \lambda_i \mathbf{Q}_i$, $\mathbf{Q} = \{\mathbf{Q}_1^{(1)}, \mathbf{Q}_2^{(1)}, \ldots, \mathbf{X}\mathbf{Q}_1^{(2)}\mathbf{X}^T, \mathbf{X}\mathbf{Q}_2^{(2)}\mathbf{X}^T, \ldots\}$

$\mathbf{F}_i = -\mathbf{C}^{-1}\mathbf{Q}_i\mathbf{C}^{-1}$

$\mathbf{C}_{\theta|y} = (\mathbf{X}^T\mathbf{C}_y^{-1}\mathbf{X})^{-1}$ and $\eta_{\theta|y} = \mathbf{C}_{\theta|y}(\mathbf{X}^T\mathbf{C}_y^{-1}\mathbf{y})$

$g_i = -\frac{r}{2} Tr[\mathbf{F}_i\mathbf{R}(\frac{1}{r}\mathbf{Y}\mathbf{V}^{-1}\mathbf{Y}^T - \mathbf{C})\mathbf{R}^T]$

$H_{ij} = \frac{r}{2} Tr[\mathbf{F}_i\mathbf{R}\mathbf{C}\mathbf{F}_j\mathbf{R}\mathbf{C}]$

Update λ until convergence

$\lambda \longrightarrow \lambda + \mathbf{H}^{-1}\mathbf{g}$