Bayesian Modeling of Complex-valued fMRI Signals for Brain Activation

Cheng-Han Yu, Raquel Prado, Hernando Ombao, and Daniel B. Rowe

Joint Statistical Meetings, August 1, 2017
Motivation of using a complex-valued model

Many kinds of data sets are complex-valued (CV), for example, imaging, radar, and sonar. fMRI data are complex-valued after FT and IFT image reconstruction. But most fMRI studies use magnitude-only (MO) data and phase information is discarded.
Goal and Result

We propose a model that takes real and imaginary parts into account, then **fast detect which voxels are activated**. We find that the proposed Bayesian complex-valued model has **higher power** and **lower type I error rate** than traditional magnitude models.
In fMRI studies, given time \( t = 1, \ldots, T \) and at voxel \( v = 1, \ldots, N \), we have (Rowe-Logan constant phase)

\[
y^v_t = \rho^v_t \cos(\phi^v) + i \rho^v_t \sin(\phi^v) + \epsilon^v_t,
\]

\[
\rho^v_t = \alpha_0^v + \alpha_1^v x_{1,t} + \alpha_2^v x_{2,t} + \cdots + \alpha_p^v x_{p,t},
\]

\[
\begin{bmatrix}
y_1^v \\
\vdots \\
y_T^v
\end{bmatrix} = 
\begin{bmatrix}
1 & x_{11} & \cdots & x_{p1} \\
\vdots & \vdots & \cdots & \vdots \\
1 & x_{1T} & \cdots & x_{pT}
\end{bmatrix} 
\begin{bmatrix}
\alpha_0^v \cos(\phi^v) + i \alpha_0^v \sin(\phi^v) \\
\vdots \\
\alpha_p^v \cos(\phi^v) + i \alpha_p^v \sin(\phi^v)
\end{bmatrix}
\begin{bmatrix}
\beta^v_{Re} \\
\beta^v_{Im}
\end{bmatrix}
+ 
\begin{bmatrix}
\epsilon_1^v \\
\vdots \\
\epsilon_T^v
\end{bmatrix}
\]

\[
y^v = X \beta^v_{Re} + i X \beta^v_{Im} + \epsilon^v, \text{ or } y^v = X \beta^v + \epsilon^v
\]
Complex-valued linear regression: $y^v = X\beta^v + \epsilon^v$

- $X$ is a designed matrix formed by Blood Oxygen Level Dependent contrasts (BOLD), which is convolution of an stimulus function from an experiment and a hemodynamic response function (HRF).

- In general, noise $\epsilon^v_t = \epsilon^v_{t, re} + i\epsilon^v_{t, im} \sim \mathcal{CN}_1(0, \sigma^2, \tau^2)$
Approach to identifying activations for \( y^v = X\beta^v + \epsilon^v \)

- **Variable Selection**: \( \beta^v_j \neq 0 \) iff voxel \( v \) at task \( j \) is activated.

- **Complex normal spike-and-slab** prior on \( \beta^v \):
  \[
  \beta^v_j \sim (1 - \gamma^v_j) \text{CN}(0, \omega_0, \lambda_0) + \gamma^v_j \text{CN}(0, \omega_1, \lambda_1), \quad \omega_0 < \omega_1, \lambda_0 < \lambda_1.
  \]

- If \( \gamma^v_j = 0 \), treat \( \beta^v_j \) as zero, and if \( \gamma^v_j = 1 \), \( \beta^v_j \) is non-zero.

- \( \text{CN}(\mu, \sigma^2, 0) \): Circular normal that real and imaginary parts are independent.

- Activation is inferred by borrowing information across voxels through a Bernoulli prior on \( \gamma^v_j \sim \text{Ber}(\theta^v_j = \theta_j) \) with a common probability of activation for all voxels.
Model setup

We develop an complex-valued EM variable selection (C-EMVS) algorithm for fast detecting activation at the lowest voxel level.

\[ \mathbf{y}^\nu = \mathbf{X}\beta^\nu + \mathbf{\epsilon}^\nu, \quad \mathbf{\epsilon}^\nu \sim \text{CN}_T(\mathbf{0}, 2\sigma^2_v \mathbf{I}, \mathbf{0}), \nu = 1, ..., N \]

\[ \beta_j^\nu | \gamma_j^\nu \text{ indep} (1 - \gamma_j^\nu) \text{CN}_1(0, d_j \sigma^2_v \Gamma_j, e_j \sigma^2_v C_j) + \gamma_j^\nu \text{CN}_1(0, \sigma^2_v \Gamma_j, \sigma^2_v C_j), \]

\[ d_j << 1, e_j << 1, j = 1, ..., p \]

\[ \gamma_j^\nu | \theta_j \text{ IID } \text{Ber}(\theta_j) \]

\[ \theta_j \text{ IID } \text{Beta}(a_\theta, b_\theta) \]

\[ \sigma^2_v \sim \text{IG}(a_\sigma, b_\sigma) \]

We determine if the real and imaginary parts of \( \beta_j^\nu \) are zero jointly:

\[ \beta_j^\nu \neq 0 \text{ if } \Pr(\gamma_j^\nu = 1 | \beta^*, \theta^*, \sigma^*, \mathbf{y}) > 0.5, \text{ where } * \text{ means the posterior mode and 0.5 is the threshold value.} \]
Simulation study: data generating

\[ y_{t,re}^v = (\beta_0 + \beta_1 v x_{bold,t}) \cos(\pi/4) + \epsilon_{t, re}^v, \quad \epsilon_{t, re}^v \sim \text{iid} N(0, 0.25) \]
\[ y_{t,im}^v = (\beta_0 + \beta_1 v x_{bold,t}) \sin(\pi/4) + \epsilon_{t, im}^v, \quad \epsilon_{t, im}^v \sim \text{iid} N(0, 0.25) \]
\[ y_{t,mag}^v = \sqrt{(y_{t,re}^v)^2 + (y_{t,im}^v)^2}, \quad \text{SNR} = \beta_0/\sigma, \quad \text{CNR}_v = \beta_1^v/\sigma \]
Simulation study: modeling

- The model:

\[ y^v = X\beta^v + \epsilon^v, \quad \epsilon^v \sim CN_T(0, 2\sigma^2 I, 0), \quad v = 1, \ldots, V \]

\[ \beta^v_j | \gamma^v_j \sim (1 - \gamma^v_j)CN(0, \nu_0 2\sigma^2, 0) + \gamma^v_j CN(0, 2\sigma^2, 0), \]

\[ \sigma^2 \propto 1/\sigma^2 \]

\[ \gamma^v_j | \theta_j \sim Ber(\theta_j) \]

\[ \theta_j \sim Beta(a_\theta = 1, b_\theta = 1) \]

- Voxel \( v \) is active if \( Pr(\gamma^v_{BOLD} = 1 | \beta^*, \theta^*, \sigma^*) > 0.5 \)

<table>
<thead>
<tr>
<th>SNR</th>
<th>0.5</th>
<th>1</th>
<th>5</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNR</td>
<td>0.5</td>
<td>1</td>
<td>1.5</td>
<td>0.5</td>
</tr>
<tr>
<td>Data No.</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
• The CV model detects more true positives than the MO when the SNR is small, leading to higher sensitivity, precision and accuracy.

• The CV model performs better in estimating strength.
- EM performs better than lasso and adaptive lasso.
- CV-EM performs consistently across different SNRs.
- CV-EM dominates when SNR is small; MO-EM is better when SNR is large.
How to choose $\nu_0$?

- From Rockova and George (2014) and Wang et al. (2015), choose $\nu_0^*$ that maximizes the $Pr(\gamma | y)$ that evaluates $\gamma$ containing only those variables for which $\gamma^\nu = 1$.

- Suggestion: Create a grid between $1/\sqrt{100T_p}$ and $1/\sqrt{10T_p}$
How to choose $v_0$?

- From Rockova and George (2014) and Wang et al. (2015), choose $v_0^*$ that maximizes the $Pr(\gamma | y)$ that evaluates $\gamma$ containing only those variables for which $\gamma^v = 1$.

- Suggestion: Create a grid between $1/\sqrt{100T_p}$ and $1/\sqrt{10T_p}$
Comparison to traditional GLM and different HRFs
Extension 1: AR(1) noises to human subject data

\[
\epsilon^v_{t,k} = \varphi^v \epsilon^v_{t-1,k} + z^v_{t,k}, \quad z^v_{t,k} \sim N(0, \sigma^2), \quad k = \text{Re}, \text{Im}
\]

\[
p(\varphi^v) \sim \text{Unif}(-1, 1)
\]

Figure: Data from Karaman, Bruce and Rowe (KBR, 2014) Left to right: KBR-CV, KBR-MO, DeTeCT-ING, EM-CV

- removes false positives outside the brain area, comparing to KBR-CV.
- has higher detecting power than KBR-MO.
- is comparable to the nonlinear sophisticated DeTeCT-ING model.
Extension 2: IID to Spatial dependence (Bezener (2015))

- The spatial dependence is governed by an underlying areal model.
- Parcellating the images into clusters of voxels.
- A spatial hierarchical prior that allows prior anatomical information is used to model the spatial dependence.
1. The spatial model removes false positives outside the brain and is comparable to the nonlinear sophisticated DeTeCT-ING model.
2. The spatial model further eliminates single isolated active voxels, and encourage grouping active voxels, comparing to the non-spatial models.

Figure : Left: Spatial MCMC. Right: Non-spatial EMVS
Summary

• C-EMVS fast detects activations at the voxel level.

• Bayesian modeling does not have multiple testing issues, for example, Bonferroni or FDR correction.

• Using complex-valued data detects more true positives (active voxels) and/or less false positives, especially when SNR is small, while magnitude data can be used if SNR is large.

• C-EMVS is based on linear model and does not use sophisticated spatio-temporal or nonlinear models, but its detecting performance is comparable to those models.

• The CV algorithm can be applied to any complex-valued data.

Thank you! cheyu@soe.ucsc.edu
• C-EMVS fast detects activations at the voxel level.

• Bayesian modeling does not have multiple testing issues, for example, Bonferroni or FDR correction.

• Using complex-valued data detects more true positives (active voxels) and/or less false positives, especially when SNR is small, while magnitude data can be used if SNR is large.

• C-EMVS is based on linear model and does not use sophisticated spatio-temporal or nonlinear models, but its detecting performance is comparable to those models.

• The CV algorithm can be applied to any complex-valued data.

• Thank you! cheyu@soe.ucsc.edu
Complex Normal Distribution

- $Z = X + iY \sim CN_n(\mu_z, \Gamma, C)$ iff
  \[
  \begin{pmatrix} X \\ Y \end{pmatrix} \sim N_{2n} \left( \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \Sigma = \begin{pmatrix} \Sigma_X & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_Y \end{pmatrix} \right). \quad \text{With } \mu_z = 0,
  \]
  \[
  \Gamma := E(Z\overline{Z'}) = \Sigma_X + \Sigma_Y + i(\Sigma_{XY} - \Sigma_{YX}),
  \]
  \[
  C := E(Z\overline{Z'}) = \Sigma_X - \Sigma_Y + i(\Sigma_{XY} + \Sigma_{YX}).
  \]

- $\Gamma$: covariance matrix. $C$: relation matrix

- $C = 0$ iff $\Sigma_X = \Sigma_Y$ and $\Sigma_{XY} = -\Sigma_{YX}$, and $Z$ is called proper or second order circular. Its real and imaginary parts have the same covariance, and are uncorrelated.
• Given a complex normal r.v. $Z = X + iY$
  (1) the amplitude $R = \sqrt{X^2 + Y^2}$ follows (marginal) Rician distribution,
  (2) the phase $\phi$ conditional on $R$, $(\phi \mid R)$, follows Tikhonov distribution (ODonoughue and Moura (IEEE 2012)).

• The model $y^\nu = X\beta^\nu + \epsilon^\nu$, with $\beta^\nu = \beta^\nu_{Re} + i\beta^\nu_{Im}$ is linear and easy to work with.
• Develop a corresponding approximate inference, say variational inference that includes spatio-temporal structure and fast detect activations.

• Find a way to adaptively determine the region shapes and sizes of images.

• Examine the effects of different shrinkage priors, for example, spike-slab lasso, horseshoe, etc.

• Build a more general model that includes connectivity issue.

• Study the noncircular complex normal behavior in detail.