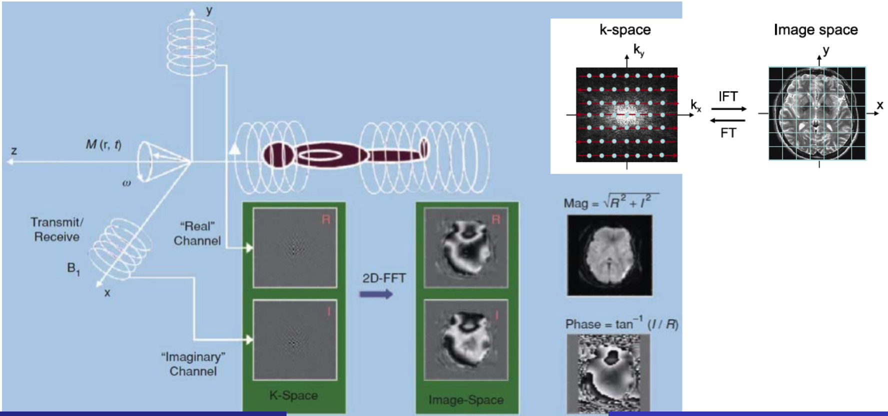
Bayesian Modeling of Complex-valued fMRI Signals for Brain Activation

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

Joint Statistical Meetings, August 1, 2017

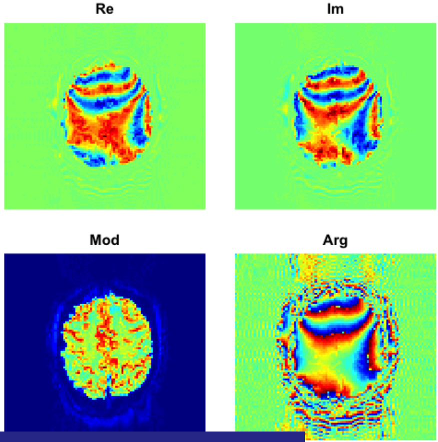
Motivation of using a complex-valued model

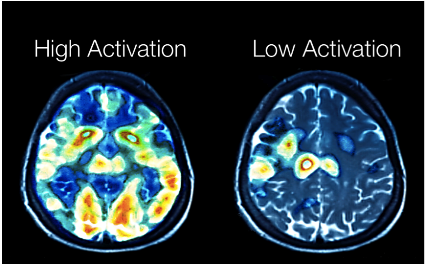
- Many kinds of data sets are complex-valued (CV), for example, imaging, radar, and sonar. fMRI data are **complex-valued** after FT and IFT image reconstruction. But most fMRI studies use **magnitude-only** (MO) data and **phase information is discarded.**



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.





Complex-valued linear regression: Rowe (2005)

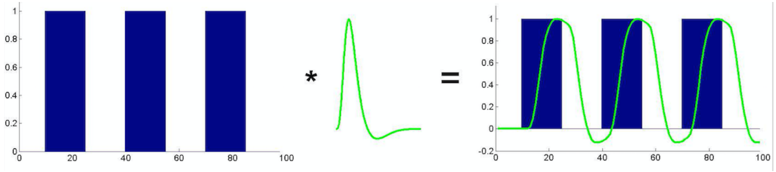
- In fMRI studies, given time $t = 1, \dots, T$ and at voxel $v = 1, \dots, N$, we have (Rowe-Logan constant phase)

$$
\begin{aligned}
y_t^v &= \rho_t^v \cos(\phi^v) + i\rho_t^v \sin(\phi^v) + \epsilon_t^v, \\
\rho_t^v &= \alpha_0^v + \alpha_1^v x_{1,t} + \alpha_2^v x_{2,t} + \cdots + \alpha_p^v x_{p,t},
\end{aligned}
$$

- $$\begin{bmatrix} y_1^v \\ \vdots \\ y_T^v \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{p1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1T} & \cdots & x_{pT} \end{bmatrix} \begin{bmatrix} \alpha_0^v \cos(\phi^v) + i\alpha_0^v \sin(\phi^v) \\ \vdots \\ \underbrace{\alpha_p^v \cos(\phi^v)}_{\boldsymbol{\beta}_{Re}^v} + i \underbrace{\alpha_p^v \sin(\phi^v)}_{\boldsymbol{\beta}_{Im}^v} \end{bmatrix} + \begin{bmatrix} \epsilon_1^v \\ \vdots \\ \epsilon_T^v \end{bmatrix}$$
- $\mathbf{y}^v = \mathbf{X}\boldsymbol{\beta}_{Re}^v + i\mathbf{X}\boldsymbol{\beta}_{Im}^v + \boldsymbol{\epsilon}^v$, or $\mathbf{y}^v = \mathbf{X}\boldsymbol{\beta}^v + \boldsymbol{\epsilon}^v$

Complex-valued linear regression: $\mathbf{y}^v = \mathbf{X}\beta^v + \boldsymbol{\epsilon}^v$

- $\mathbf{X}$ 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).



Source: Martin Lindquist (2008)

- In general, noise $\epsilon_t^v = \epsilon_{t,re}^v + i\epsilon_{t,im}^v \sim CN_1(0, \sigma^2, \tau^2)$

Approach to identifying activations for $\mathbf{y}^v = \mathbf{X}\boldsymbol{\beta}^v + \boldsymbol{\epsilon}^v$

- **Variable Selection**: $\beta_j^v \neq 0$ iff voxel v at task j is activated.
- **Complex normal spike-and-slab** prior on $\boldsymbol{\beta}^v$:

$$\beta_j^v \sim (1-\gamma_j^v)\underbrace{CN(0,\omega_0,\lambda_0)}_{spike} + \gamma_j^v \underbrace{CN(0,\omega_1,\lambda_1)}_{slab}, \quad \omega_0 < \omega_1, \lambda_0 < \lambda_1.$$

- If $\gamma_j^v = 0$, treat β_j^v as zero, and if $\gamma_j^v = 1$, β_j^v is non-zero.
- $CN(\mu, \sigma^2, 0)$: Circular normal that real and imaginary parts are independent.
- Activation is inferred by borrowing information across voxels through a Bernoulli prior on $\gamma_j^v \sim Ber(\theta_j^v = \theta_j)$ 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**.

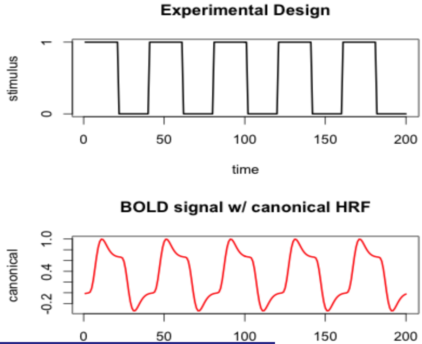
$$\begin{aligned}
\mathbf{y}^v &= \mathbf{X}\boldsymbol{\beta}^v + \boldsymbol{\epsilon}^v, \quad \boldsymbol{\epsilon}^v \sim CN_T(\mathbf{0}, 2\sigma_v^2\mathbf{I}, \mathbf{0}), v = 1, ..., N \\
\beta_j^v|\gamma_j^v &\overset{\text{indep}}{\sim} (1-\gamma_j^v)CN_1(0, d_j\sigma_v^2\Gamma_j, e_j\sigma_v^2 C_j) + \gamma_j^v CN_1(0, \sigma_v^2\Gamma_j, \sigma_v^2 C_j), \\
d_j &<< 1, e_j << 1, j = 1, ..., p \\
\gamma_j^v|\theta_j &\overset{\text{IID}}{\sim} Ber(\theta_j) \\
\theta_j &\overset{\text{IID}}{\sim} Beta(a_\theta, b_\theta) \\
\sigma_v^2 &\sim IG(a_\sigma, b_\sigma)
\end{aligned}$$

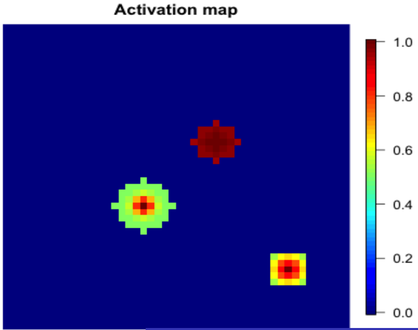
We determine if the real and imaginary parts of β_j^v are zero **jointly**: $\beta_j^v \neq 0$ if $Pr(\gamma_j^v = 1 \mid \boldsymbol{\beta}^*, \theta^*, \sigma^*, y) > 0.5$, where * means the posterior mode and 0.5 is the threshold value.

Simulation study: data generating

$$\left.\begin{array}{ll} y^v_{t,re} = (\beta_0 + \beta^v_1 x_{bold,t}) \cos(\pi/4) + \epsilon^v_{t,re}, & \epsilon^v_{t,re} \overset{\text{IID}}{\sim} N(0, 0.25) \\ y^v_{t,im} = (\beta_0 + \beta^v_1 x_{bold,t}) \sin(\pi/4) + \epsilon^v_{t,im}, & \epsilon^v_{t,im} \overset{\text{IID}}{\sim} N(0, 0.25) \end{array}\right\} \epsilon^v_t \sim CN(0, 0.5, 0)$$

$$y^v_{t,mag} = \sqrt{(y^v_{t,re})^2 + (y^v_{t,im})^2}, \quad \text{SNR} = \beta_0/\sigma, \quad \text{CNR}_v = \beta^v_1/\sigma$$





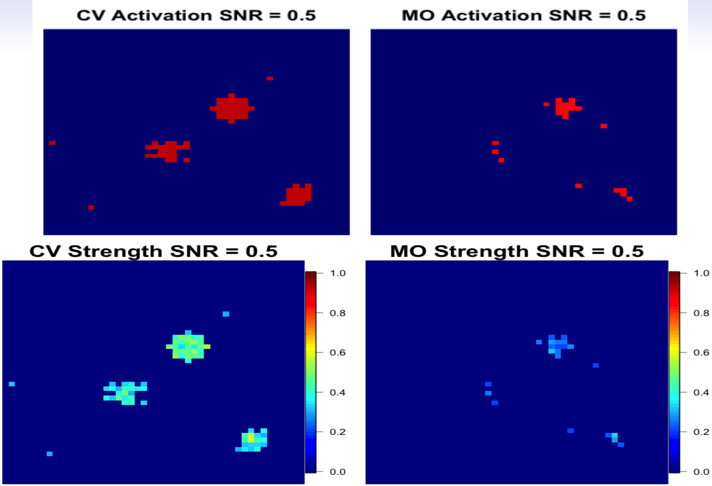
Simulation study: modeling

- The model:

$$
\begin{aligned}
\mathbf{y}^v &= \mathbf{X}\boldsymbol{\beta}^v + \boldsymbol{\epsilon}^v, \quad \boldsymbol{\epsilon}^v \sim CN_T(\mathbf{0}, 2\sigma^2 I, \mathbf{0}), v = 1, ..., V \\
\beta_j^v \mid \gamma_j^v &\overset{\text{indep}}{\sim} (1-\gamma_j^v) CN_1(0, v_0 2\sigma^2, 0) + \gamma_j^v CN_1(0, 2\sigma^2, 0), \\
\sigma^2 &\propto 1/\sigma^2 \\
\gamma_j^v \mid \theta_j &\overset{\text{IID}}{\sim} Ber(\theta_j) \\
\theta_j &\overset{\text{IID}}{\sim} Beta(a_\theta = 1, b_\theta = 1)
\end{aligned}
$$

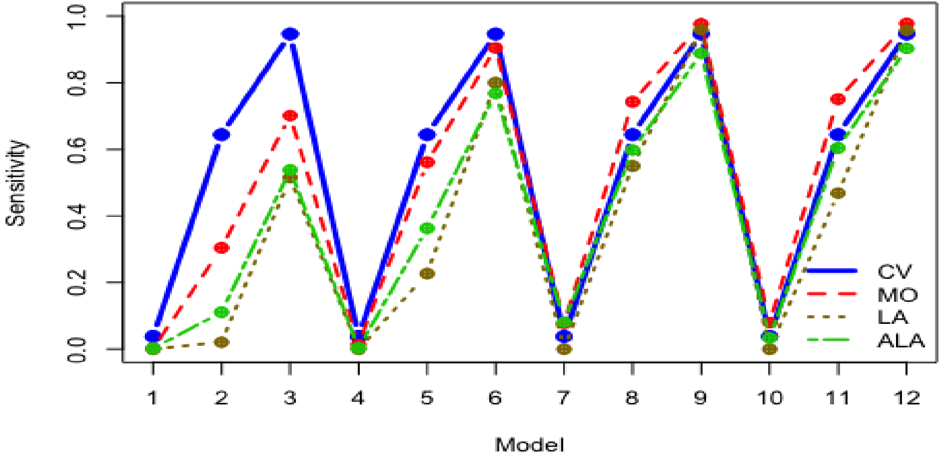
- Voxel v is active if $Pr(\gamma_{BOLD}^v = 1 \mid \boldsymbol{\beta}^*, \theta^*, \sigma^*) > 0.5$

SNR	0.5			1			5			10		
CNR	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
Data No.	1	2	3	4	5	6	7	8	9	10	11	12



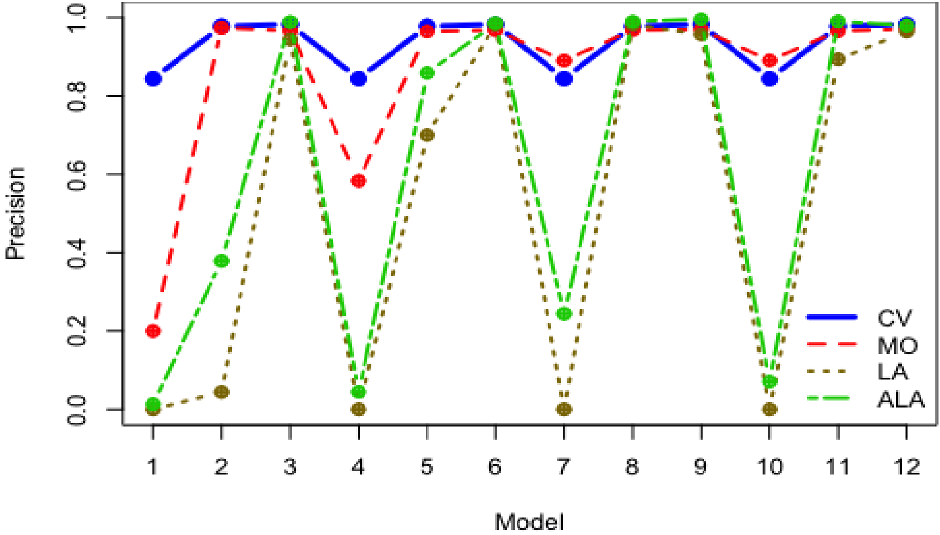
- The CV model detects more true positives than the MO when the SNR is small, leading to higher sensitivity, precision and accuracy.
- The CV model performs better in estimating strength.

SNR	0.5			1			5			10		
CNR	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
Data No.	1	2	3	4	5	6	7	8	9	10	11	12



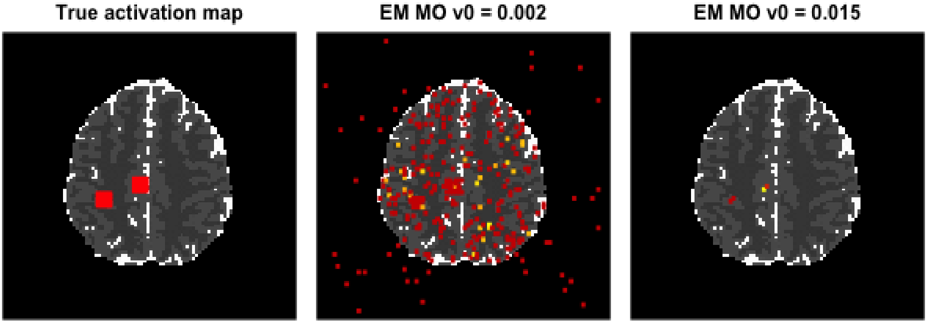
- EM performs better than lasso and adaptive lasso.
- CV-EM performs consistently across different SNRs.
- CV-EM dominates when SNR is small; MO-EM is better when SNR is large.

SNR	0.5			1			5			10		
CNR	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
Data No.	1	2	3	4	5	6	7	8	9	10	11	12



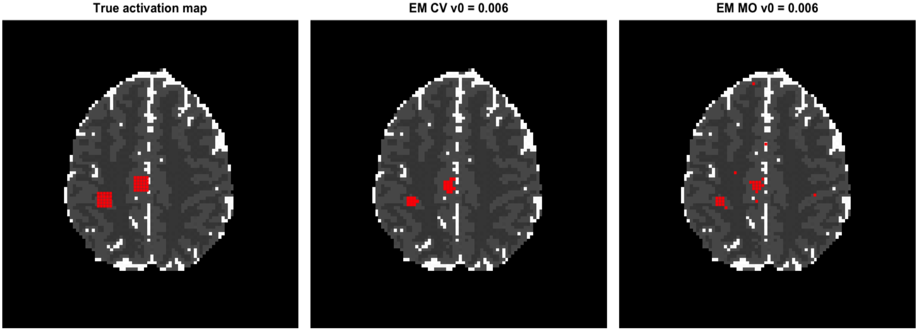
How to choose v_0?

- From Rockova and George (2014) and Wang et al. (2015), choose v_0^* that maximizes the $Pr(\gamma \mid \mathbf{y})$ that evaluates γ containing only those variables for which $\gamma^v = 1$.
- Suggestion: Create a grid between $1/\sqrt{100Tp}$ and $1/\sqrt{10Tp}$

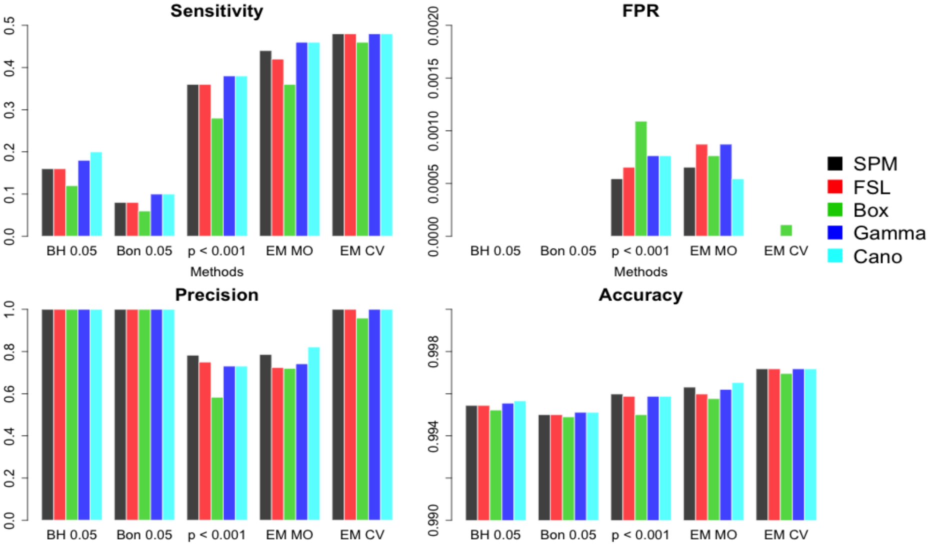


How to choose v_0?

- From Rockova and George (2014) and Wang et al. (2015), choose v_0^* that maximizes the $Pr(\boldsymbol{\gamma} \mid \mathbf{y})$ that evaluates $\boldsymbol{\gamma}$ containing only those variables for which $\gamma^v = 1$.
- Suggestion: Create a grid between $1/\sqrt{100Tp}$ and $1/\sqrt{10Tp}$



Comparison to traditional GLM and different HRFs



Extension 1: AR(1) noises to human subject data

$$\epsilon^v_{t,k} = \varphi_v \epsilon^v_{t-1,k} + z^v_{t,k}, \quad z^v_{t,k} \sim N(0, \sigma^2), k = Re, Im$$

$$p(\varphi_v) \sim Unif(-1, 1)$$

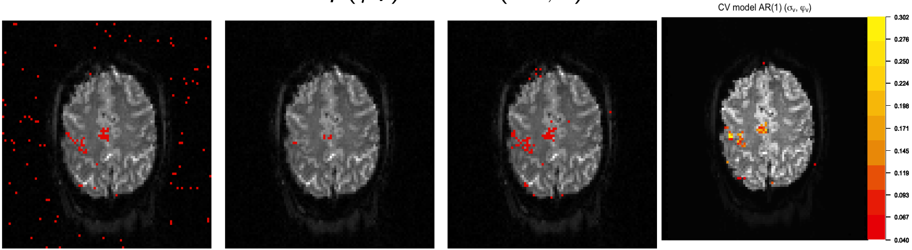
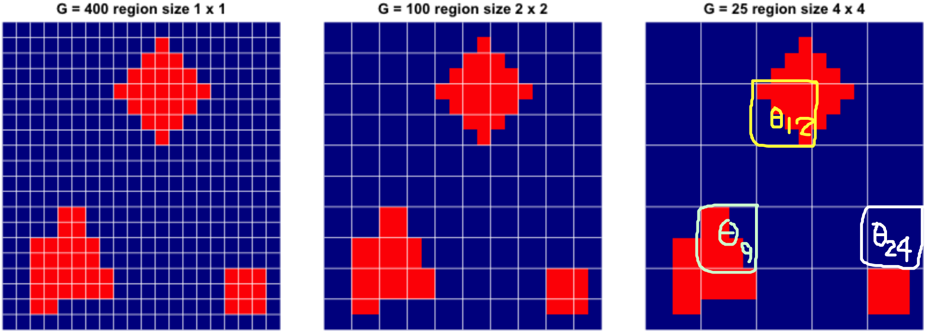


Figure : Data from Karaman, Bruce and Rowe (KBR, 2014) Left to right: KBR-CV, KBR-MO, DeTeCT-ING, EM-CV

- 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 allow prior anatomical information is used to model the spatial dependence.



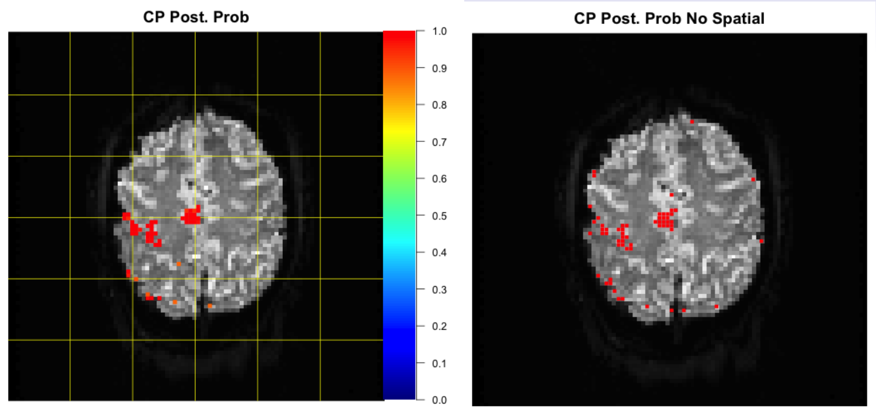


Figure : Left: Spatial MCMC. Right: Non-spatial EMVS

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.

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

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

Complex Normal Distribution

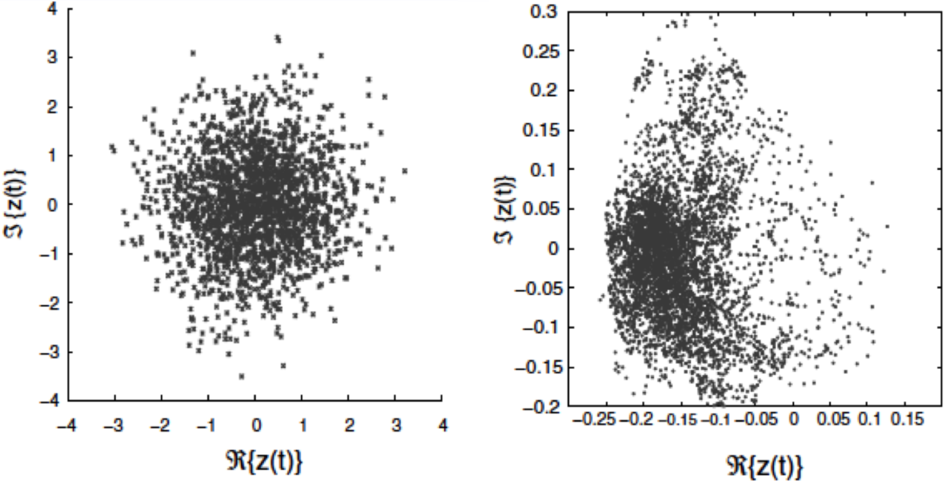
- $\mathbf{Z} = \mathbf{X} + i\mathbf{Y} \sim CN_n(\boldsymbol{\mu}_z, \boldsymbol{\Gamma}, \mathbf{C})$ iff

  $$\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix} \sim N_{2n}\left(\begin{pmatrix}\boldsymbol{\mu}_x\\ \boldsymbol{\mu}_y\end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix}\Sigma_X & \Sigma_{XY}\\ \Sigma_{YX} & \Sigma_Y\end{pmatrix}\right).$$ With $\boldsymbol{\mu}_z = \mathbf{0}$,

  $$\boldsymbol{\Gamma} := \mathrm{E}(\mathbf{Z}\bar{\mathbf{Z}}') = \Sigma_X + \Sigma_Y + i(\Sigma_{XY} - \Sigma_{YX}),$$
  $$\mathbf{C} := \mathrm{E}(\mathbf{Z}\mathbf{Z}') = \Sigma_X - \Sigma_Y + i(\Sigma_{XY} + \Sigma_{YX}).$$

- $\boldsymbol{\Gamma}$: covariance matrix. $\mathbf{C}$: relation matrix

- $\mathbf{C} = \mathbf{0}$ iff $\Sigma_X = \Sigma_Y$ and $\Sigma_{XY} = -\Sigma_{YX}$, and $\mathbf{Z}$ is called proper or second order circular. Its **real and imaginary parts have the same covariance**, and **are uncorrelated**.


ℑ{z(t)}
ℜ{z(t)}
4
3
2
1
0
−1
−2
−3
−4
−4 −3 −2 −1 0 1 2 3 4
ℑ{z(t)}
ℜ{z(t)}
0.3
0.25
0.2
0.15
0.1
0.05
0
−0.05
−0.1
−0.15
−0.2
−0.25 −0.2 −0.15 −0.1 −0.05 0 0.05 0.1 0.15

Distribution of Magnitude and Phase

- 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 ϕ conditional on R, $(\phi \mid R)$, follows Tikhonov distribution (ODonoughue and Moura (IEEE 2012)).
- The model $\mathbf{y}^v = \mathbf{X}\beta^v + \boldsymbol{\epsilon}^v$, with $\boldsymbol{\beta}^v = \boldsymbol{\beta}^v_{Re} + i\boldsymbol{\beta}^v_{Im}$ is linear and easy to work with.

Work in process and future work

- 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.