New Tools for Decoding Mental Representations from Neuroimaging Data

Samuel J. Gershman,[1,2] Jeremy R. Manning,[2,3] David M. Blei,[3] & Kenneth A. Norman[1,2]

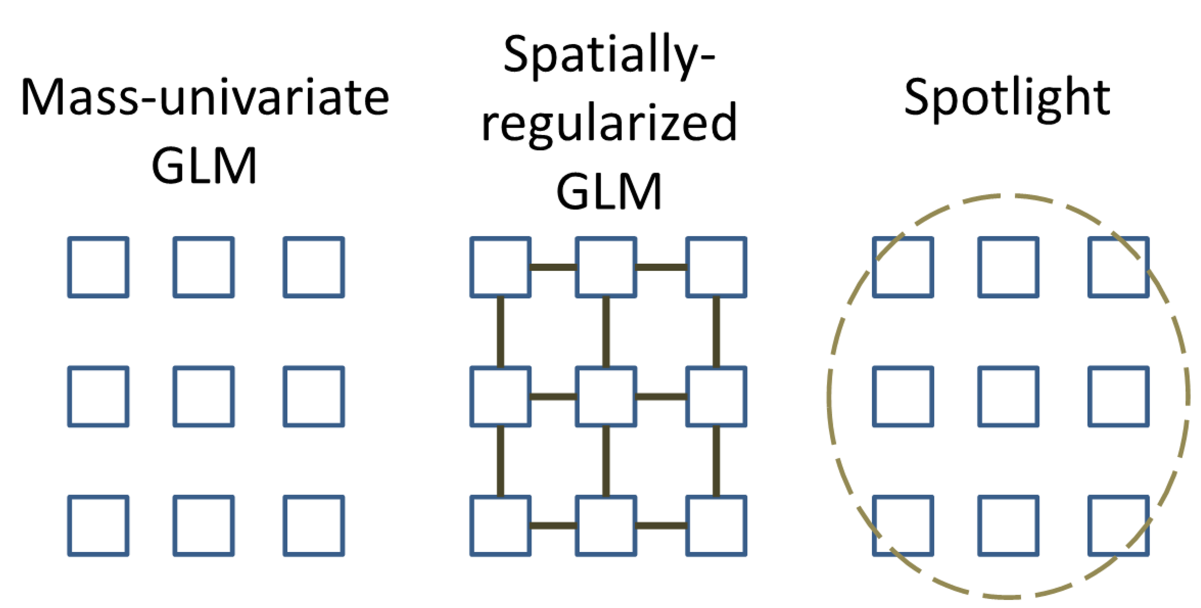
[1]Department of Psychology, [2] Princeton Neuroscience Institute, and [3]Department of Computer Science, Princeton University

ABSTRACT

We introduce a probabilistic model that decomposes multi-subject neuroimaging data into topographic latent sources. These sources are anatomically interpretable, intrinsically low-dimensional and shared across subjects. A fast variational inference algorithm makes fitting this model to large data sets feasible. We show how this model can be effectively applied to the task of decoding mental states and reconstructing brain images.

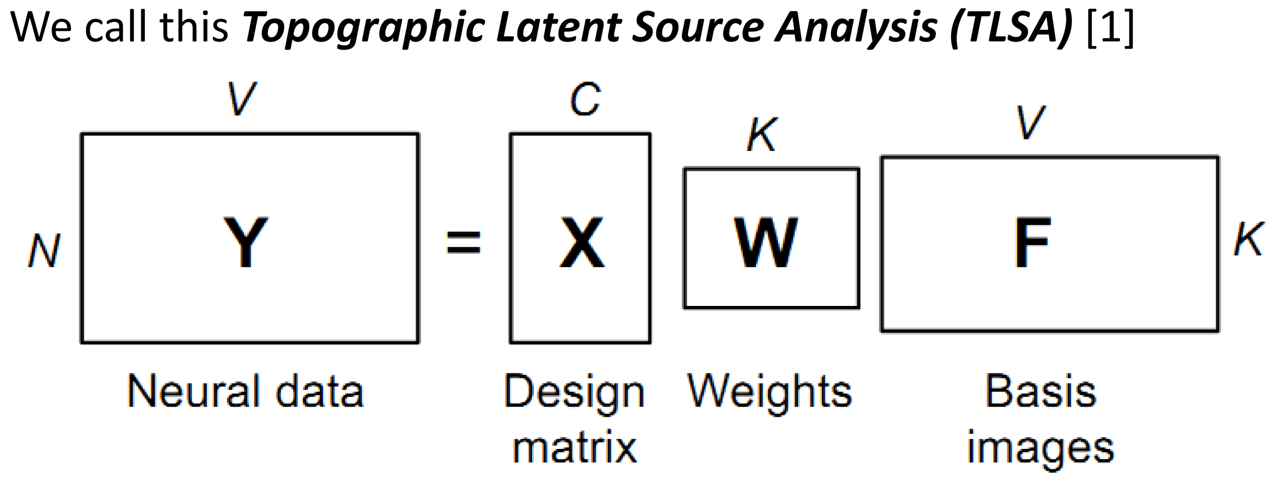
1. Modeling fMRI data

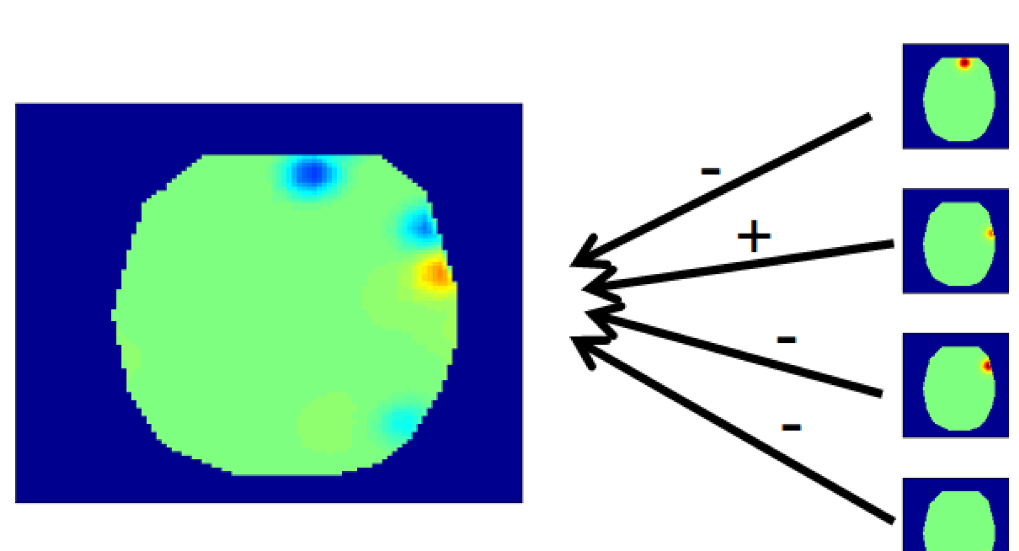
Standard approaches to spatial modeling of fMRI data:



Problems:
- arbitrary discretization
- sensitive to noise
- explosion of parameters
- hard to align results across subjects

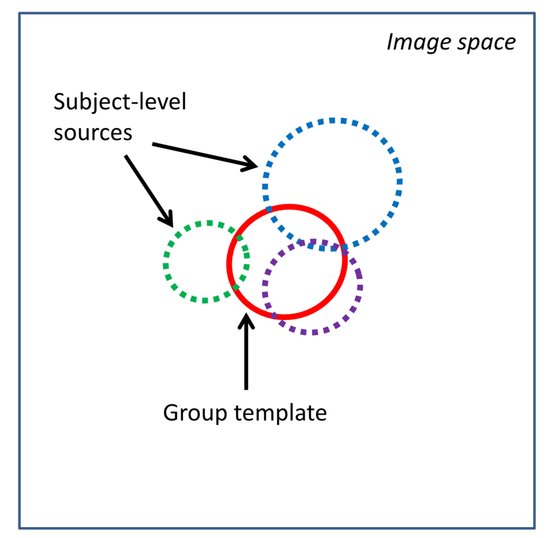
The current approach: model neural activation as a covariate-dependent superposition of latent sources:
We call this ***Topographic Latent Source Analysis (TLSA)*** [1]





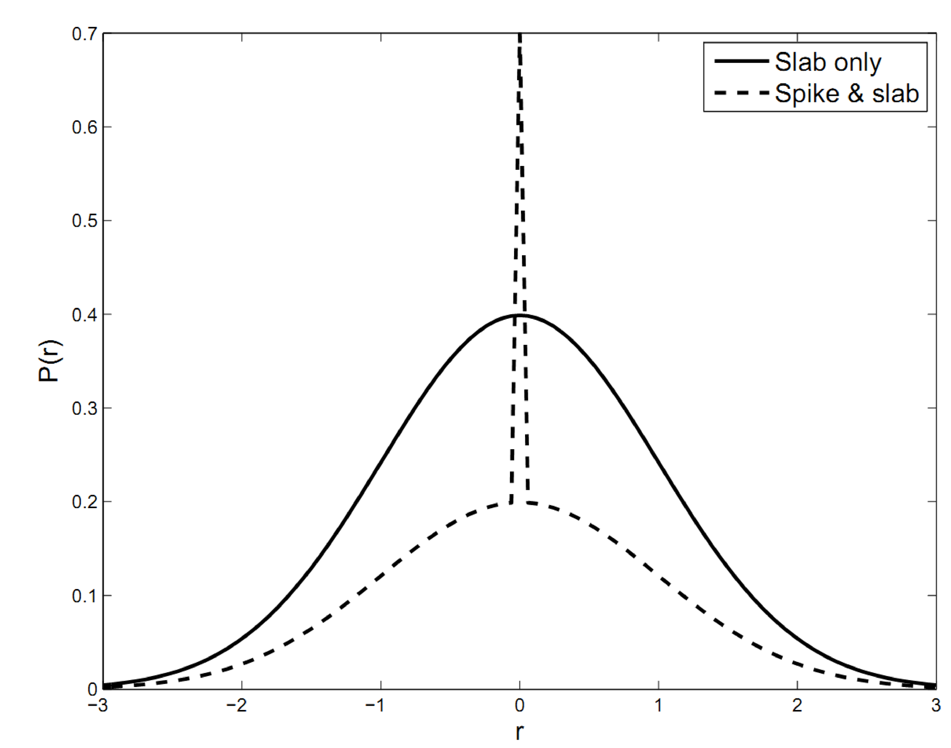
Each latent source is a continuous function over space (in this case, a radial basis function)

2. Hierarchy and sparsity



- Each latent source is assumed to have separate representations at the subject and group level
- Each subject's source is a small translation and dilation of the group template
- This allows sharing of statistical strength across subjects while allowing intersubject variability
- A similar hierarchy is defined for the covariate-source weights

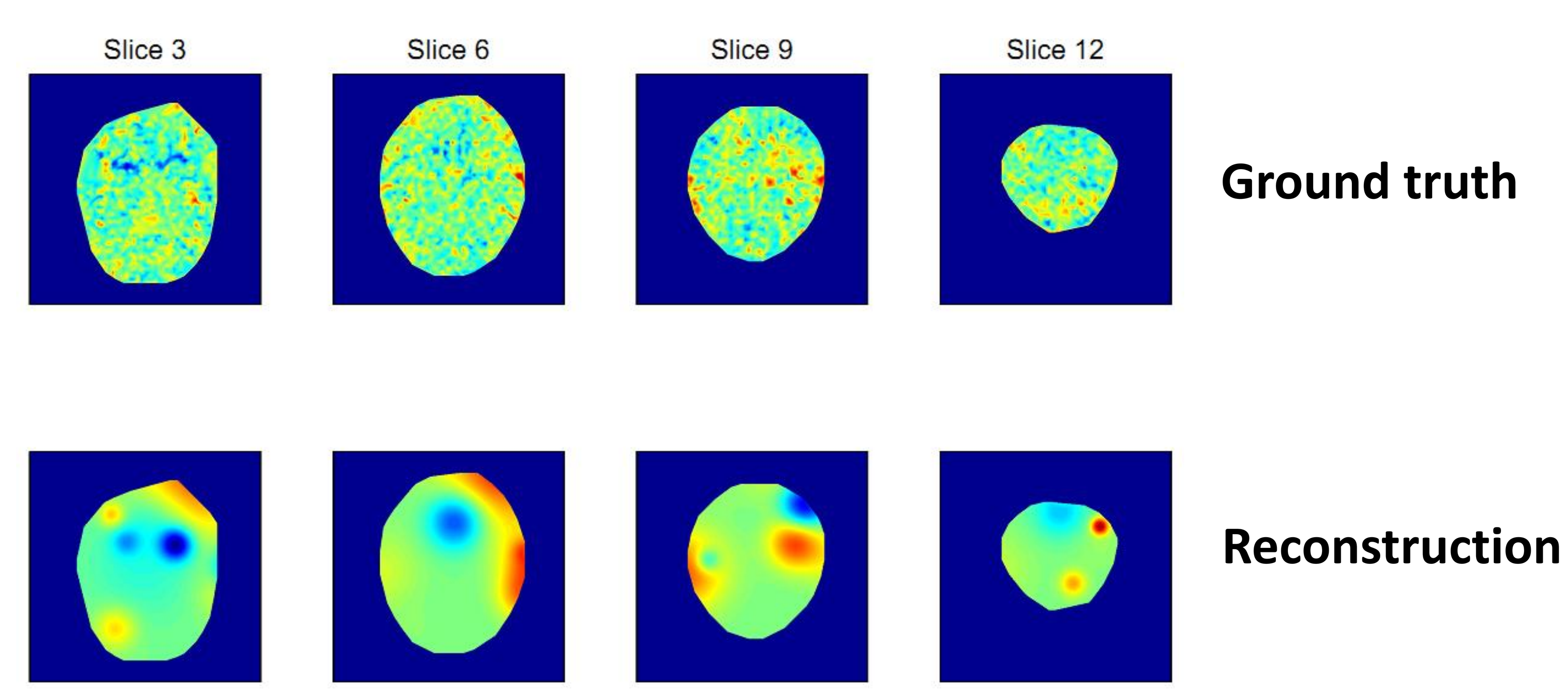
- In practice we don't know how many sources to use
- We can infer this by placing a sparse "spike and slab" prior on the weights
- The model adaptively adjusts the sparsity level
- As the number of sources grows to infinity, this model defines a nonparametric prior (the beta process)
- Only a small number of sources will be active for any given data set
- The model thus automatically infers the "effective" number of sources



3. Inference

- Our goal is to invert the generative model using Bayes' rule to recover the posterior over parameters (θ): $P(\theta|\mathbf{X},\mathbf{Y}) = P(\mathbf{Y},\mathbf{X},\theta)/P(\mathbf{Y},\mathbf{X})$
- This computation is intractable, however
- We therefore approximate the posterior with a tractable distribution that is close (in Kullback-Leibler divergence) to the true posterior. This is called *variational Bayes* [2].
- Can be understood as maximizing a lower bound on the log probability of the data
- Updates are in closed form and computationally efficient

Example brain map

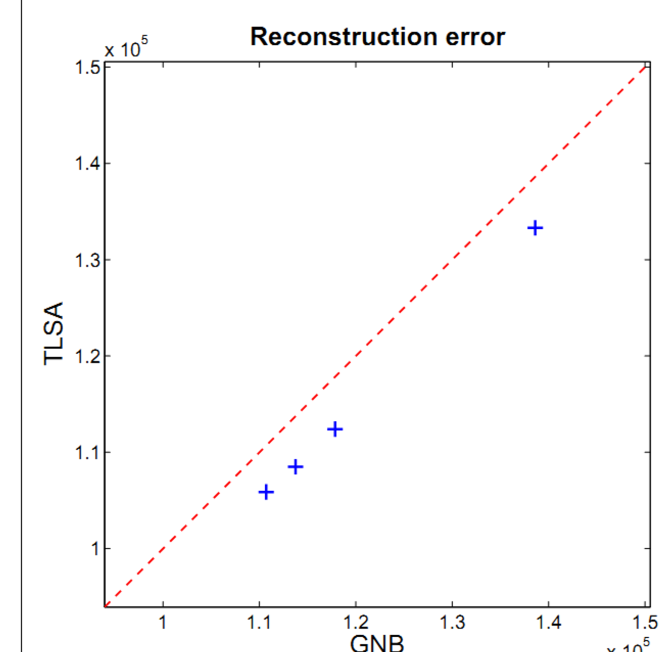


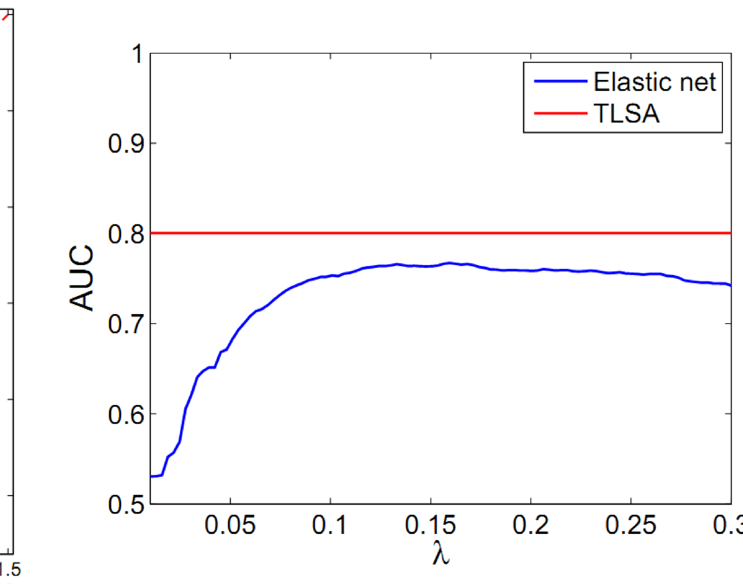
4. Results

- Data set from Mason & Just (unpublished)
- Subjects viewed images of tools and buildings

How well can we reconstruct neural data when we only know the class?

How well can we decode the class when we only know the neural data?

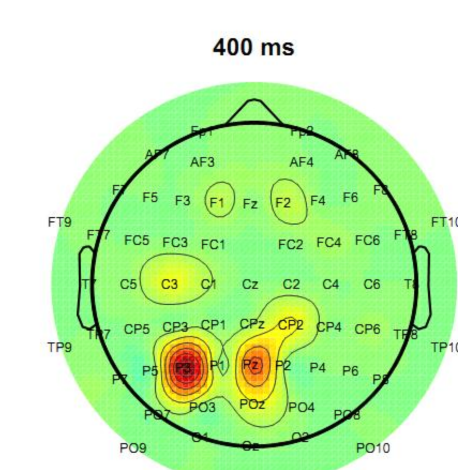


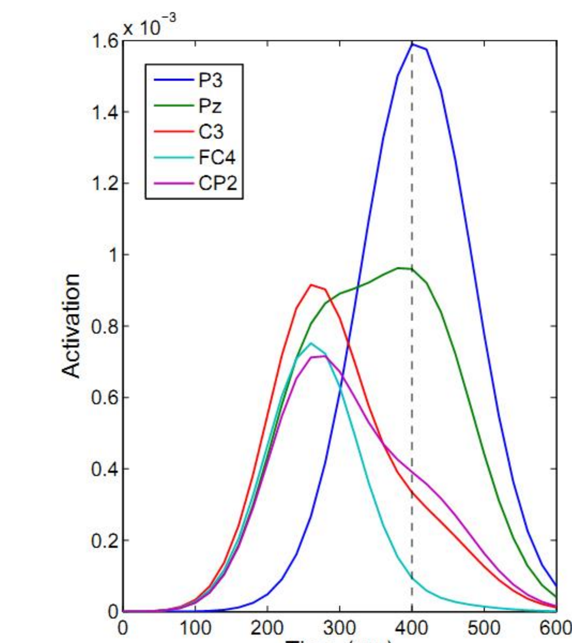


Gaussian naïve Bayes

5. Extension to EEG data

- Data set from Sederberg (unpublished)
- Subjects performed an audiovisual oddball task
- Spatiotemporal sources: blobs in space and time
- Model discovers the classic P300 waveform over Parietal electrodes:





6. Conclusions

- TLSA effectively models the structure of brain images
- The generative modeling approach gives us a flexible means of constraining inferences about latent sources; it is easy to build in new spatial and temporal constraints
- Can be applied to multiple brain imaging techniques
- We are working on decoding high-dimensional semantic representations and tracking these as subjects search through memory

ACKNOWLEDGEMENTS: This research was supported by the NSF/NIH Collaborative Research in Computational Neuroscience Program, grant number NSF IIS-1009542

References:

[1] Gershman, S.J., Blei, D.M., Pereira, F., & Norman, K.A. (2011). A topographic latent source model for fMRI data. *NeuroImage, 57,* 89-100.

[2] Chappell, M.A., Groves, A.R., Whitcher, B., Woolrich, M.W. (2009). Variational Bayesian Inference for a non-linear forward model. *IEEE Transactions on Signal Processing, 57*, 223-236.