Sparse Generalized Principal Components Analysis with Applications to Neuroimaging

Genevera I. Allen

Department of Statistics, Rice University,
Department of Pediatrics-Neurology, Baylor College of Medicine,
& Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital.

March 11, 2013

Review: Principal Components Analysis

Principal Components Analysis (PCA):

- Dimension reduction.
- Exploratory data analysis.
- PCA Problem:

$$\underset{\mathbf{v}_k}{\text{maximize}} \quad \text{Var}(\mathbf{X}\,\mathbf{v}_k) = \mathbf{v}_k^T\,\mathbf{X}^T\,\mathbf{X}\,\mathbf{v}_k$$
$$\text{subject to} \quad \mathbf{v}_k^T\,\mathbf{v}_k = 1 \;\;\&\;\; \mathbf{v}_k^T\,\mathbf{v}_{k'} = 0 \;\;\forall\;\; k' < k.$$

  PC: $\mathbf{z}_k = \mathbf{X}\,\mathbf{v}_k$.

- Given by the singular value decomposition (SVD) of the data matrix: $\mathbf{X} = \mathbf{U}\,\mathbf{D}\,\mathbf{V}^T$, then $\mathbf{Z} = \mathbf{X}\,\mathbf{V}$.

When does PCA (SVD) fail?

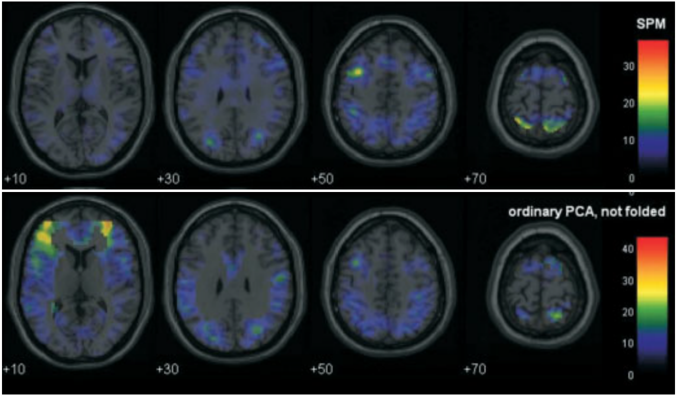
1. High-dimensional data.
   - Fix: Sparsity - Sparse PCA (Johnstone and Lu, 2004).
2. Structured Factors.
   - Fix: Smoothness - Functional PCA (Rice and Silverman, 1991).
   - Fix: Sparsity - Sparse PCA (Jolliffe et. al, 2003).

When does PCA (SVD) fail?

1. High-dimensional data.
   - Fix: Sparsity - Sparse PCA (Johnstone and Lu, 2004).
2. Structured Factors.
   - Fix: Smoothness - Functional PCA (Rice and Silverman, 1991).
   - Fix: Sparsity - Sparse PCA (Jolliffe et. al, 2003).
3. Strong dependencies among row and/or column variables? Structured Data?
   - *Transposable Data*: Dependencies among the rows and/or column of a data matrix.

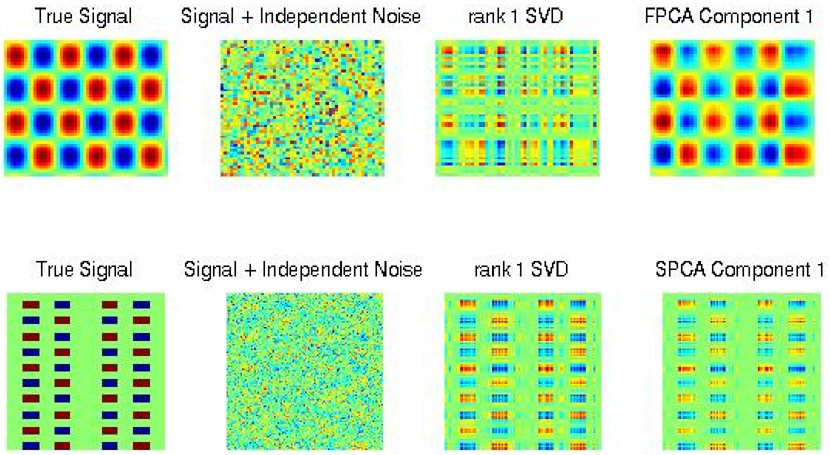
PCA in Neuroimaging

Multivariate analysis techniques used for finding Regions of Interest and Activation Patterns, understanding Functional Connectivity, but ...

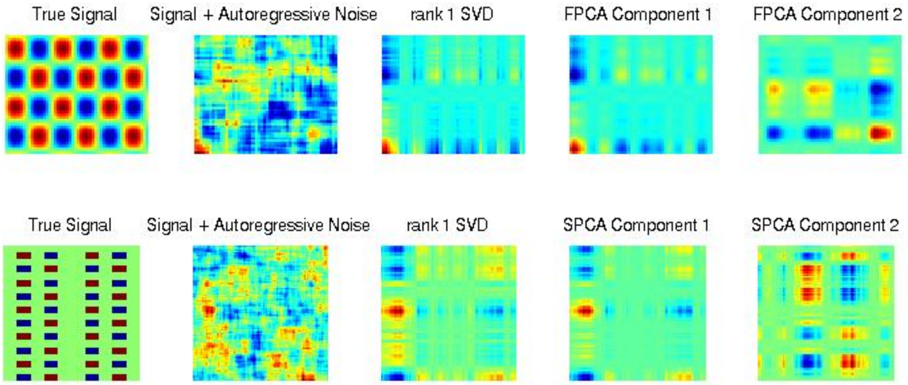


(Viviani et. al, 2005)

PCA and Correlated Noise



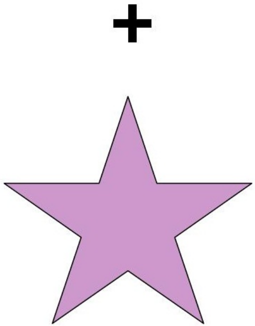
PCA and Correlated Noise



StarPlus fMRI Data

StarPlus Data (Subject 04847): (Mitchell et al., 2004)

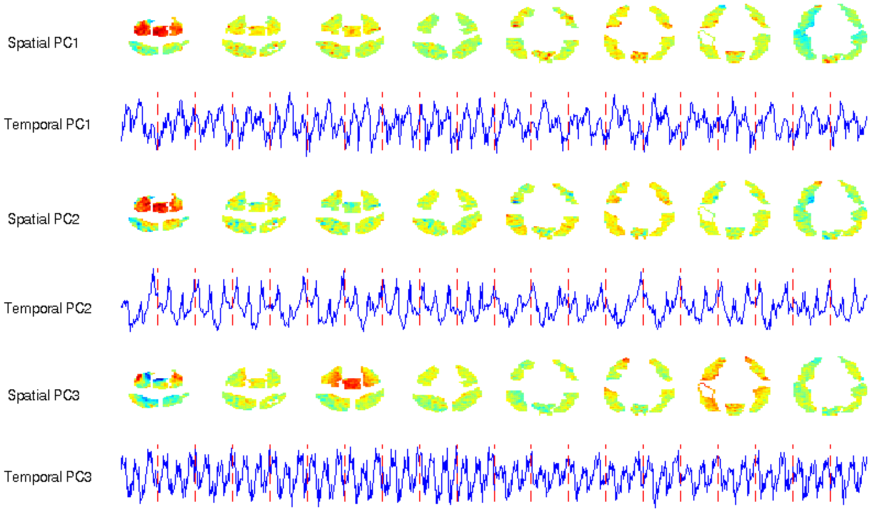
- Task: Object identification.
- 20 tasks in which sentence agrees with image.
- 20 tasks in which sentence opposes image.
- Each task lasted 27 seconds (55 time points).
- Images: $64 \times 64 \times 8$.
- Data Set: 4,698 voxels $\times$ 40 tasks $\times$ 55 time points.



Goal: Use pattern recognition techniques to find regions of interest and activation patterns related to object identification.

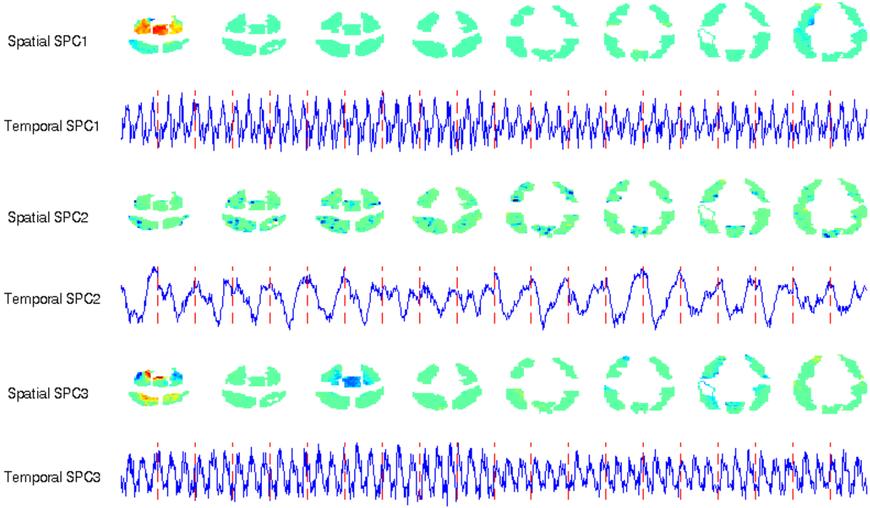
Starplus PCA Results

Classical PCA:



Starplus PCA Results

Sparse PCA:



Objectives

Goals

1. Incorporate known noise structure and/or dependencies into PCA problems.
2. Develop a framework for regularization of PCA factors.
3. Provide computationally feasible solutions and algorithms in ultra high-dimensional settings.

PCA Model

$$\mathbf{X}_{n \times p} = \sum_{k=1}^{K} d_k \, \mathbf{u}_k \, \mathbf{v}_k^T + \mathbf{E}_{n \times p} \, .$$

- Random: d_k & Fixed: $\mathbf{U} = [\mathbf{u}_1, \ldots \mathbf{u}_K]$ and $\mathbf{V} = [\mathbf{v}_1, \ldots \mathbf{v}_K]$.
- Independent Noise: $\mathbf{E}_{ij} \overset{iid}{\sim} (0, \sigma^2)$, or $\mathrm{Cov}(\mathrm{vec}(\mathbf{E})) = \sigma^2 \mathbf{I}_{(p)} \otimes \mathbf{I}_{(n)}$.

SVD Loss Function: $|| \, \mathbf{X} - \mathbf{U} \, \mathbf{D} \, \mathbf{V}^T \, ||_F^2$.

- $|| \cdot ||_F$ is the Frobenius norm (sums of squared errors).
- Error terms weighted equally.
- Cross-product errors between elements ij and $i'j'$ are ignored.

Our Generative Model

$$\mathbf{X}_{n \times p} = \sum_{k=1}^{K} d_k \, \mathbf{u}_k \, \mathbf{v}_k^T + \mathbf{E}_{n \times p} \, .$$

- Random: d_k & Fixed: $\mathbf{U} = [\mathbf{u}_1, \dots \mathbf{u}_K]$ and $\mathbf{V} = [\mathbf{v}_1, \dots \mathbf{v}_K]$.
- Noise: Two-way (separable) dependencies:

$$\mathrm{Cov}(\mathrm{vec}(\mathbf{E})) = \mathbf{\Delta} \otimes \mathbf{\Sigma},$$

with $\mathbf{\Delta} \in \Re^{p \times p}$ the column covariance and $\mathbf{\Sigma} \in \Re^{n \times n}$ the row covariance.

$$\mathbf{\Delta} \otimes \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Delta}_{11} \mathbf{\Sigma} & \mathbf{\Delta}_{12} \mathbf{\Sigma} & \dots & \mathbf{\Delta}_{1p} \mathbf{\Sigma} \\ \mathbf{\Delta}_{21} \mathbf{\Sigma} & \mathbf{\Delta}_{22} \mathbf{\Sigma} & & \\ \vdots & & \ddots & \vdots \\ \mathbf{\Delta}_{p1} \mathbf{\Sigma} & & \dots & \mathbf{\Delta}_{pp} \mathbf{\Sigma} \end{pmatrix}$$

- Signal factors assumed to be orthogonal to the noise covariance: $\mathbf{U}^T \mathbf{\Sigma} \mathbf{U} = \mathbf{I}$ & $\mathbf{V}^T \mathbf{\Delta} \mathbf{V} = \mathbf{I}$.

GPCA Optimization Problem

SVD Problem

$$\underset{\mathbf{U},\mathbf{D},\mathbf{V}}{\text{minimize}} \quad ||\,\mathbf{X} - \mathbf{U}\,\mathbf{D}\,\mathbf{V}^T\,||_F^2$$

$$\text{subject to} \quad \mathbf{U}^T\,\mathbf{U} = \mathbf{I}_{(K)}, \quad \mathbf{V}^T\,\mathbf{V} = \mathbf{I}_{(K)} \quad \& \quad \text{diag}(\mathbf{D}) \geq 0.$$

How can we modify the Frobenius norm?

GPCA Optimization Problem

Generalized Least Squares Matrix Decomposition (GMD / GPCA) Problem

$$\underset{\mathbf{U},\mathbf{D},\mathbf{V}}{\text{minimize}} \quad ||\, \mathbf{X} - \mathbf{U}\,\mathbf{D}\,\mathbf{V}^T \,||^2_{\mathbf{Q},\mathbf{R}}$$

$$\text{subject to} \quad \mathbf{U}^T\,\mathbf{Q}\,\mathbf{U} = \mathbf{I}_{(K)}, \quad \mathbf{V}^T\,\mathbf{R}\,\mathbf{V} = \mathbf{I}_{(K)} \quad \& \quad \text{diag}(\mathbf{D}) \geq 0.$$

- Quadratic Operators: $\mathbf{Q} \in \Re^{n \times n}$ and $\mathbf{R} \in \Re^{p \times p}$ positive semi-definite.
- $\mathbf{Q}, \mathbf{R}$-norm: $||\, \mathbf{X} \,||_{\mathbf{Q},\mathbf{R}} = \sqrt{\text{tr}(\mathbf{Q}\,\mathbf{X}\,\mathbf{R}\,\mathbf{X}^T)}$.
- Generalization of the Frobenius norm: If $\mathbf{Q}, \mathbf{R} = \mathbf{I}$, then GMD equivalent to the SVD.

Quadratic Operators

Interpretations:

1. Matrix-variate Normal:
   - $|| \mathbf{X} - \mathbf{U}\,\mathbf{D}\,\mathbf{V}^T ||^2_{\mathbf{Q},\mathbf{R}} \propto \ell_{n,p}(\mathbf{U}\,\mathbf{D}\,\mathbf{V}^T, \mathbf{Q}^{-1}, \mathbf{R}^{-1})$.
   - $\mathbf{Q}$ and $\mathbf{R}$ behave like inverse row and column covariances.
2. Covariance Decomposition:
   Under certain model assumptions:

   $$\mathrm{Cov}(\mathrm{vec}(\mathbf{X})) = \sum \mathrm{Var}(d_k)(\mathbf{v}_k\,\mathbf{v}_k^T) \otimes (\mathbf{u}_k\,\mathbf{u}_k^T) + \mathbf{R} \otimes \mathbf{Q}\,.$$

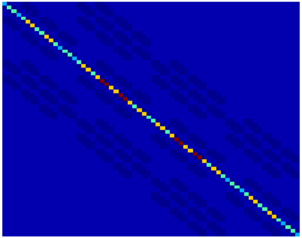
   where $\mathbf{V}^T\,\mathbf{R}\,\mathbf{V} = \mathbf{I}$ and $\mathbf{U}^T\,\mathbf{Q}\,\mathbf{U} = \mathbf{I}$.
3. Smoothing Matrices: Factors $\mathbf{U}$ and $\mathbf{V}$ as smooth as the smallest eigenvectors of $\mathbf{Q}$ and $\mathbf{R}$.
4. Weighting Matrices: Up-weight cross-product errors in the loss according to the covariance between variables.

Quadratic Operators

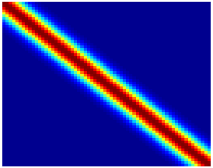
Classes of Quadratic Operators:

1. Model-Based Operators.
   - Random Field Covariances.
   - Temporal Processes Covariance or Inverse Covariances.
   - Gaussian Markov Random Fields.
2. Smoothing Operators.
   - Functional Data Analysis.
3. Graphical Operators.
   - Graph Laplacians.

Spatial Graphical Operator



Temporal Smoothing Operator



GPCA Solution

GMD Solution

Let $\tilde{\mathbf{X}} = \mathbf{Q}^{1/2}\,\mathbf{X}\,\mathbf{R}^{1/2}$ and $\tilde{\mathbf{X}} = \tilde{\mathbf{U}}\tilde{\mathbf{D}}\tilde{\mathbf{V}}^T$ be the SVD of $\tilde{\mathbf{X}}$. Then the GMD solution, $\hat{\mathbf{X}} = \mathbf{U}^*\,\mathbf{D}^*(\mathbf{V}^*)^T$, is given by:

$$\mathbf{U}^* = \tilde{\mathbf{Q}}^{-1/2}\tilde{\mathbf{U}}, \quad \mathbf{V}^* = \tilde{\mathbf{R}}^{-1/2}\tilde{\mathbf{V}}, \quad \& \quad \mathbf{D}^* = \tilde{\mathbf{D}}.$$

$\mathbf{U}^*$ sample GPCs (scores) and $\mathbf{V}^*$ GPCA loadings (directions)

Alternative Computational Approaches:

- Linear algebra tricks when $n << p$.
- Deflation via the Generalized Power Method: Performs alternating generalized least squares regression.

Regularized GPCA

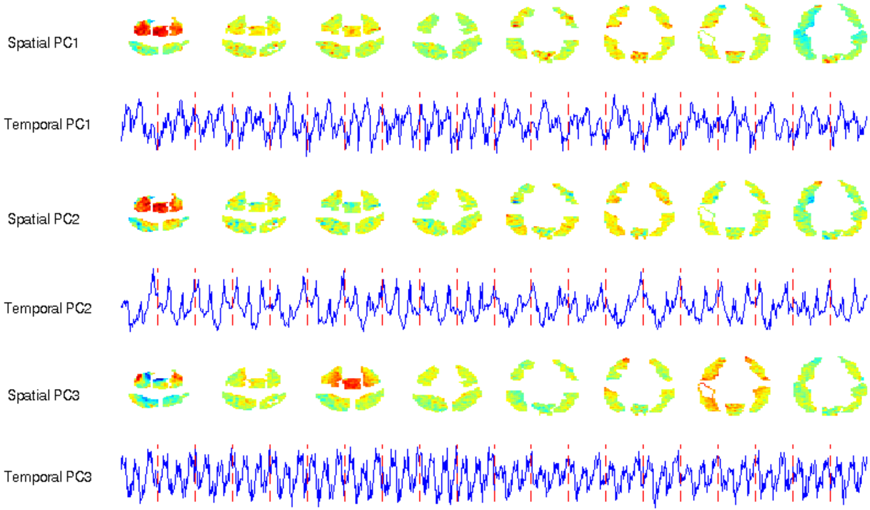
Regularized GPCA Optimization Problem

$$\underset{\mathbf{v},\mathbf{u}}{\text{maximize}} \quad \mathbf{u}^T \mathbf{Q} \mathbf{X} \mathbf{R} \mathbf{v} - \lambda_{\mathbf{v}} P_1(\mathbf{v}) - \lambda_{\mathbf{u}} P_2(\mathbf{u})$$
$$\text{subject to} \quad \mathbf{u}^T \mathbf{Q} \mathbf{u} \leq 1 \;\&\; \mathbf{v}^T \mathbf{R} \mathbf{v} \leq 1.$$

- **Theorem**: If $P()$ is any norm or semi-norm, then factor-wise solutions given by solving a penalized regression problem and re-scaling.
- Options: $P(x) = ||x||_1$ - sparsity, group sparsity, ℓ_q balls, total variation, and etc.
- Multiple factors computed via deflation.

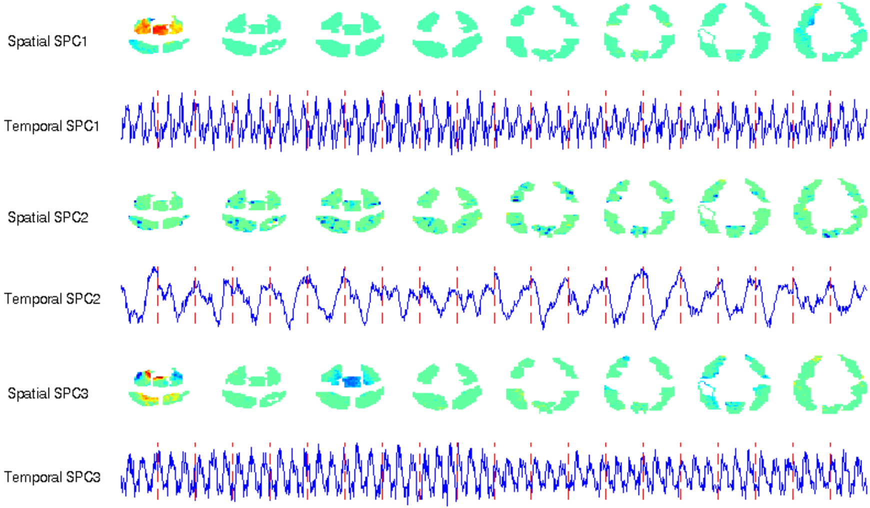
Starplus PCA Results

Classical PCA:



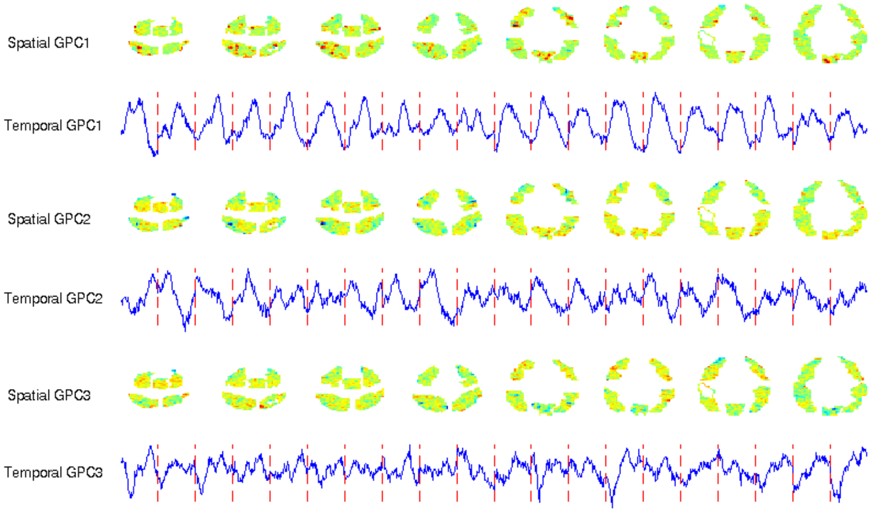
Starplus PCA Results

Sparse PCA:



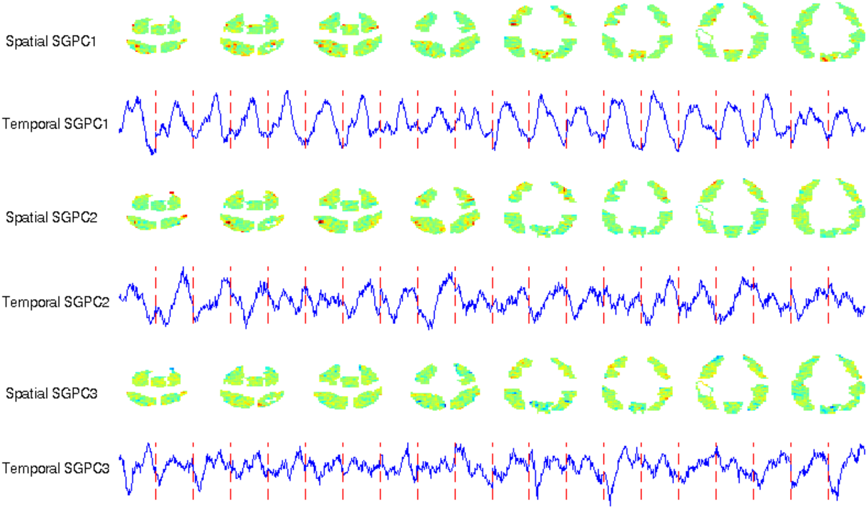
Starplus GPCA Results

Generalized PCA:



Starplus GPCA Results

Sparse Generalized PCA:



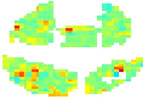
Starplus GPCA Results

Results:

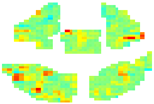
- Identified "Ventral Stream" (brain regions associated with object identification).
- Anatomical Regions:
  - Bilateral occipital.
  - Left-lateralized inferior temporal.
  - Inferior frontal.

(Pennick & Kana, 2012)

SGPCA 1, Axial Slice 2

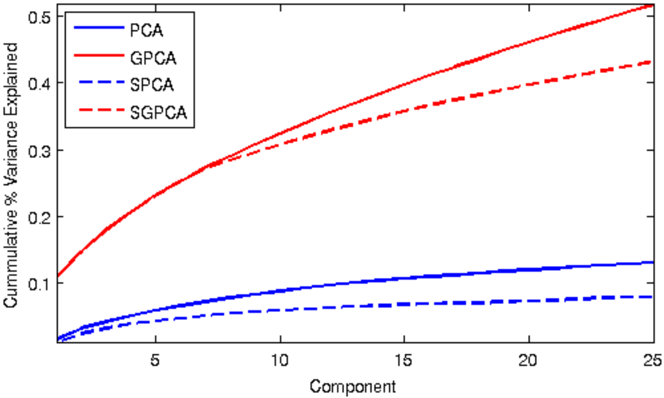


SGPCA 1, Axial Slice 3



PCA Comparisons

Extent of dimension reduction achieved:



Other Extensions

- Non-negative (and sparse) GPCA.
- Tensor GPCA or Higher-Order GPCA.
  - Based on the Tucker Decomposition.
- Sparse Higher-Order GPCA.
  - Computes sequential rank-one decompositions that are a relaxation of the CANDENCOMP/PARAFAC decomposition.

Applications: Multi-subject neuroimaging data.

Concluding Remarks

When to use GPCA vs. PCA:

- **Structured Data** (variables are associated with a specific location).
- Smooth or functional data.
- Data with low signal to noise ratio.

Future Work:

- Statistical Work: How to choose **Q** and **R**, the rank of the decomposition, the level of sparsity, and consistency studies.
- Applications in Neuroimaging: Comparisons to ICA methods.
- Other Applications: Genomcis, proteomics, image data, time series and longitudinal data, spatio-temporal data, climate studies, remote sensing.

Concluding Remarks

R Package & Matlab Toolbox

Coming Soon . . .

Code available from www.stat.rice.edu∼/gallen.

Acknowledgments

Funding:

National Science Foundation DMS-1209017

Collaborators:

- Logan Grosenick, Center for Mind and Brain, Stanford University.
- Jonathan Taylor, Statistics, Stanford University.
- Mirjana Maletic-Savatic, Jan and Dan Duncan Neurological Research Institute & Baylor College of Medicine.
- Frederick Campbell, PhD Candidate, Statistics, Rice University.

References

G. I. Allen, L. Grosenick & J. Taylor, A generalized least squares decomposition, arXiv: 1102.3074, Rice University Technical Report No. TR2011-03, 2011.

G. I. Allen, Regularized tensor decompositions and higher-order principal components analysis, arXiv:1202.2476, 2012.

G. I. Allen, Sparse higher-order principal components analysis, In *Artificial Intelligence and Statistics*, 2012.

G. I. Allen & M. Maletic-Savatic, Sparse non-negative generalized PCA with applications to metabolomics, *Bioinformatics*, **27**:21, 3029-3035, 2011.

F. D. Campbell & G. I. Allen, Algorithms and approaches for analyzing massive structured data with Sparse Generalized PCA, In Preparation.

G. I. Allen, C. Peterson, M. Vannucci, and M. Maletic-Savatic, "Regularized Partial Least Squares with an Application to NMR Spectroscopy", (To Appear) *Statistical Analysis and Data Mining*, 2013.