# A Deflation Method for Probabilistic PCA

Rajiv Khanna[1], Joydeep Ghosh[1], Russel Poldrack[2], Oluwasanmi Koyejo[2]
1: UT Austin, 2: Stanford

## Introduction

- Goal: Sequential estimation of factors for Probabilistic PCA
- Lot of prior work factors for deterministic PCA, including sparse variants
- For PPCA, factors are estimated all at once as a subspace
- Sequential estimation benefits include better scalability, runtime model selection, and interpretation of individual factors as directions maximizing variance while retaining the capability to incorporate rich prior structure
- Idea: Use recent results for information projection to restrict support of subsequent factors.
- We present empirical results on fMRI datasets to illustrate interpretability and performance in terms of variance explained.

## Background

- First principal component:

$$\max_{\|\mathbf{w}\|_2=1} \mathbf{w}^\dagger \mathbf{\Sigma} \mathbf{w}$$

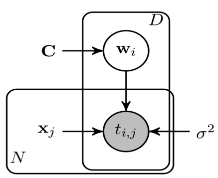
- Deflation (remove effect of previous component(s)):

$$\mathbf{\Sigma}_{i+1} = \mathbf{\Sigma}_i - \mathbf{w}_i \mathbf{w}_i^\dagger \mathbf{\Sigma}_i \mathbf{w}_i \mathbf{w}_i^\dagger$$

- For the data matrix:

$$\mathbf{T}_{i+1} = \mathbf{T}_i - \mathbf{x}_i \mathbf{w}_i^\dagger$$

- Probabilistic PCA plate model




- Information Projection of a density p onto the set of all densities on X:

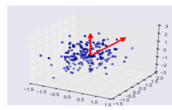
$$\mathrm{KL}(q\|p) = \int_{x\in \mathsf{X}} q(x) \log \frac{q(x)}{p(x)} dx$$

- Variational characterization of restriction of density $p$ : Information projection is equivalent to domain restriction [1].

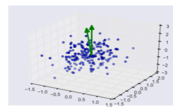
## Deflation for PPCA

Motivating example:

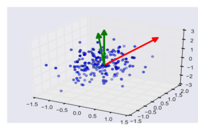
$$\mathbf{W} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \ \mathbf{x}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \ \mathbf{e}_n \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & .9 \\ 0 & .9 & 1 \end{bmatrix}\right)$$



(a) Ground Truth



(b) Estimated model



(c) Combined Plot

## Variational EM

- Energy function viewpoint of the EM

$$\mathscr{F}(q(\mathbf{W}), \Theta) = -\mathrm{KL}(q(\mathbf{W}) \| p(\mathbf{W}|\mathbf{T};\Theta)) + \log p(\mathbf{T};\Theta)$$

$$\text{E-step:} \quad \max_q \mathscr{F}(q(\mathbf{W}), \Theta),$$

$$\text{M-step:} \quad \max_\Theta \mathscr{F}(q(\mathbf{W}), \Theta).$$

- Modify the variational E step to restrict support of factors by information projection
- For Gaussian case, E-step equations are

$$\mathbf{\Sigma}_i^{-1} = \mathrm{P}_{\mathcal{M}_\perp^{(i-1)}} \left( \frac{1}{\sigma^2} (\mathbf{X}_{.,i}^\dagger \mathbf{X}_{.,i}) \mathbf{I} + \mathbf{C}^{-1} \right) \mathrm{P}_{\mathcal{M}_\perp^{(i-1)}},$$

$$\mathbf{m}_i = \frac{1}{\sigma^2} \mathbf{\Sigma}_i \mathrm{P}_{\mathcal{M}_\perp^{(i-1)}} \mathbf{Z}_i^\dagger \mathbf{X}_{.,i}.$$

## Sparse PPCA

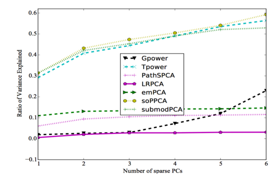
- Information projection can be used to enforce sparsity as well [2]. The support selection set function is submodular.
- For orthogonal sparse PPCA, we can do project onto a sparse set orthogonal to already chosen factors.
- This can be shown to be equivalent to doing iterated projections.
- Support selection optimization function is written as

$$\max_{\mathcal{S}\in\mathcal{S}_{k_i}} (\mathrm{P}_\mathcal{S} \mathbf{r}_i)^\dagger (\mathrm{P}_\mathcal{S} \mathbf{\Sigma}_i^{-1} \mathrm{P}_\mathcal{S})^{-1} (\mathrm{P}_\mathcal{S} \mathbf{r}_i) - \log\det \mathrm{P}_\mathcal{S} \mathbf{\Sigma}_i^{-1} \mathrm{P}_\mathcal{S}.$$

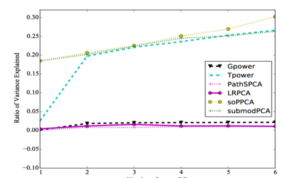
## Results

- We compare on fMRI datasets against several known sparse PCA methods. Of special note is comparison against emPCA and submodPCA both of which use naïve deflation

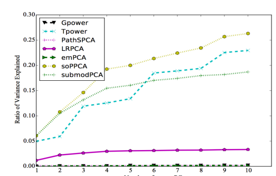
D=100




D=1000




D=10000

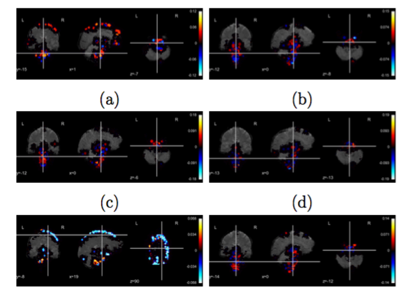



## Deterministic reductions

- The Orthogonal PPCA reduces to standard deterministic PCA if C=I
- The Sparse PPCA reduces to Truncated Power Method with Orthogonal projection deflation.

## Brain plots

Brain plots indicating extraction of motion artifacts from resting state data.



(a) (b) (c) (d) (e) (f)

## References

- [1] Oluwasanmi Koyejo, Rajiv Khanna, Joydeep Ghosh, and Poldrack Russell. On prior distributions and approximate inference for structured variables. In NIPS, 2014.
- [2] Rajiv Khanna, Joydeep Ghosh, Russell A. Pol- drack, and Oluwasanmi Koyejo. Sparse submod- ular probabilistic PCA. In AISTATS 2015