Localized Functional Principal Component Analysis

Kehui Chen,
Department of Statistics, University of Pittsburgh

ICSA, Fort Collins, June 16, 2015

Outline

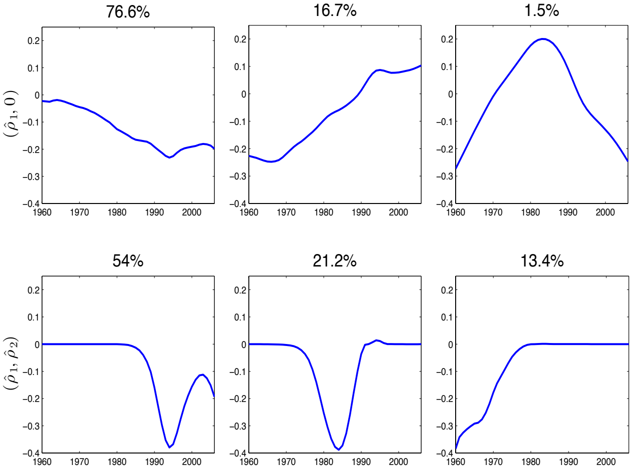
- Motivation and Challenges
- Formulation and Algorithm
- Theoretical and Numerical Properties.

Based on K. Chen and J. Lei (2015) Localized Functional Principal Component Analysis, *JASA*.

Motivation

- FPCA is a nice tool for dimension reduction
- Also a major tool to explore the source of variability in a sample of random curves: Canadian weather data.
- LFPCA is to look for orthogonal basis functions that have localized support regions.
- Meanwhile we still hope a few components can explain a large proportion of the variance.

The Mortality Rate Data



Challenges

- Adding some penalties to the eigen components.

$$\max_{v \in R^p} v^T S v - \mathscr{P}_1(v) - \mathscr{P}_2(v), \quad \text{s.t. } \|v\|_2 = 1,$$

- Two main challenges remain unsolved
  - Non-convex problem $\rightarrow$ Fantope Projection
  - Can not guarantee orthogonality of the eigenfunctions
    (a) sequential covariance deflation (White 1958, Mackey 2008)
    (b) k-subspace methods (Vu et al 2013, Lei and Vu 2015): difference between sparsity and localization
    $\rightarrow$ Deflated Fantope formulation.

Covariance Estimator

- Consider $X(t) \in L_2(t)$ for $t \in \mathscr{T} \subset \mathscr{R}$ with covariance operator $\Gamma(s,t)$. The starting point of our method is a sup-norm consistent estimator of $\Gamma(s,t)$, up to a constant shift on the diagonal.
- For dense and regularly observed functional data, the sample covariance S will be a good start point. ($\sqrt{\log p/n}$ with a constant shift σ^2 on the diagonal)
- For other designs of functional, obtain a sup-norm consistent estimator using for example two-dimensional smoothing.

Convex relaxation

- The non-convex problem

$$\max_{v \in R^p} v^T S v - \mathscr{P}_1(v) - \mathscr{P}_2(v), \;\; \text{s.t. } \|v\|_2 = 1,$$

- Let $H = vv^T$ and P_1 is the set of one-dimensional projection matrix.

$$\max_{v \in R^p} v^T S v, \;\; \text{s.t. } \|v\|_2 = 1 \Leftrightarrow \max_{H \in P_1} \langle S, H \rangle,$$

- Because $\langle S, H \rangle$ is linear in H, it is equivalent to

$$\max_H \langle S, H \rangle, \; \text{s.t. } H \in \text{Convhull}(P_1)\,.$$

Fantope

***Theorem** (Fillmore & Williams 71)*

$$\text{Convhull}(\text{all } d\text{-dim projection matrices})$$
$$= \{H : H = H^T, 0 \preceq Z \preceq I_p, \ \text{Trace}(H) = d\}$$
$$:= \mathscr{F}_{p,d} \quad (\text{the Fantope})$$

$$\max_{v \in R^p} v^T S v, \ \ \text{s.t. } \|v\|_2 = 1 \Leftrightarrow \max_{H \in \mathscr{F}_{p,1}} \langle S, H \rangle,$$

LFPCA for the first component

- Consider sparse penalty: $\mathscr{P}(H) = \|H\|_{1,1} = \sum_{i,j} |H_{ij}|$.
- Consider smooth penalty: $\mathscr{P}(H) = \langle \Delta^T \Delta, H \rangle = \langle D, H \rangle$ [Rice & Silverman 91], where $\Delta \in \mathbb{R}^{(p-2)\times p}$ is a second-differencing operator:

$$\Delta = \begin{pmatrix} -1 & 2 & -1 & 0 & \cdots \\ 0 & -1 & 2 & -1 & \cdots \\ & & \cdots & & \end{pmatrix}.$$

- $$\hat{H} = \arg \max_{H \in \mathscr{F}_{p,1}} \langle S, H \rangle - \rho_1 \langle D, H \rangle - \rho_2 \|H\|_{1,1}$$

  and $\hat{v}$ is the first eigenvector of $\hat{H}$

LFPCA through Deflated Fantope Localization

For any $p \times p$ projection matrix Π, define the deflated Fantope as

$$\mathscr{D}_{\Pi} := \{H : 0 \preceq H \preceq I, \ \text{trace}(H) = 1, \ \text{and} \ \langle H, \Pi \rangle = 0\},$$

then the sequential estimator is as follows

$$H_j = \arg\max \langle S - \rho_1 D, H \rangle - \rho_2 \|H\|_{1,1}, \ \text{s.t.} \ H \in \mathscr{D}_{\hat{\Pi}_{j-1}},$$

$$\hat{v}_j = \text{ the first eigenvector of } H_j,$$

$$\hat{\Pi}_j = \hat{\Pi}_{j-1} + \hat{v}_j \hat{v}_j^T.$$

ADMM Algorithm

- $$\max_H \langle S - \rho_1 D, H \rangle - \rho_2 \|H\|_{1,1}, \text{ s.t. } H \in \mathscr{D}_\Pi$$
- An efficient algorithm based on ADMM [Boyd et al 11] is used
- separate the l_1 penalty and deflated Fantope constraint:

$$\min_{H,Z} \mathbb{I}_{\mathscr{D}_\Pi}(H) - \langle S - \rho_1 D, H \rangle + \rho_2 \|H\|_{1,1},$$
$$\text{s.t. } H - Z = 0,$$

- The augmented Lagrangian (with scaled dual variable W) is given by

$$L(H,Z,W) = \mathbb{I}_{\mathscr{D}_\Pi}(H) - \langle S - \rho_1 D, H \rangle + \rho_2 \|H\|_{1,1} + \frac{\tau}{2} \|H - Z + W\|_2$$

ADMM iteration

Given current values H^{old}, Z^{old}, W^{old}, the variables are updated by iteratively optimizing the Lagrangian over H and Z.

$$H^{new} = \mathscr{P}_{\mathscr{D}_\Pi}(Z^{old} - W^{old} - (S - \rho_1 D)/\tau), \quad \text{deflated Fantope projection}$$

$$Z^{new} = \mathscr{S}_{\rho_2/\tau}(H^{new} + W^{old}), \quad \text{entry-wise soft thresholding}$$

$$W^{new} = W^{old} + (H^{new} - Z^{new}), \quad \text{dual update}$$

Then $(H^{old}, Z^{old}, W^{old}) \leftarrow (H^{new}, Z^{new}, W^{new})$ and repeat until convergence is observed.

Deflated Fantope Projection

(i) Soft-thresholding operator: for any $a > 0$,

$$\mathscr{S}_a(x) = \text{sign}(x) \max(|x| - a, 0)\,.$$

(ii) Deflated-Fantope-projection operator: For any $p \times p$ symmetric matrix A and projection matrix Π,

$$\mathscr{P}_{\mathscr{D}_\Pi}(A) := \arg\min_{B \in \mathscr{D}_\Pi} \|A - B\|_F^2$$

Lemma (Chen & Lei 15)

Let $\Pi = VV^T$, where V is a $p \times d$ matrix with orthonormal columns. Let U be a $p \times (p-d)$ matrix that forms an orthogonal complement basis of V. Then

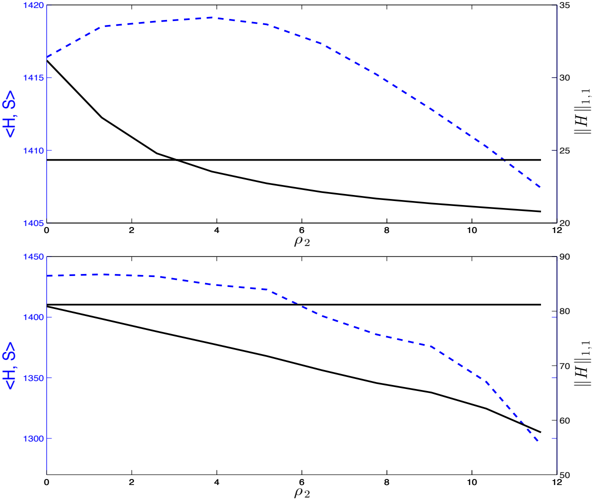
$$\mathscr{P}_{\mathscr{D}_\Pi}(A) = U \left[\sum_{i=1}^{p-d} \gamma_i^+(\theta) \eta_i \eta_i^T \right] U^T,$$

where $(\gamma_i, \eta_i)_{i=1}^{p-d}$ are eigenvalue-eigenvector pairs of $U^T AU$: $U^T AU = \sum_{i=1}^{p-d} \gamma_i \eta_i \eta_i^T$, and $\gamma_i^+(\theta) = \min(\max(\gamma_i - \theta, 0), 1)$, with θ chosen such that $\sum_{i=1}^{p-d} \gamma_i^+(\theta) = 1$.

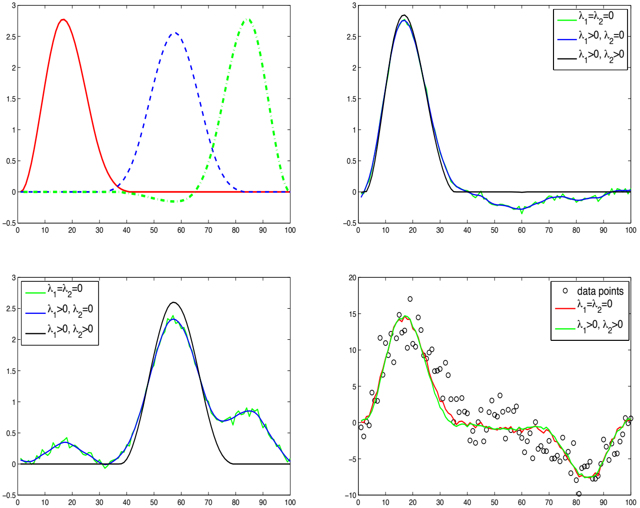
Choosing the Localization Penalty

- Suppose ρ_1 has been chosen, ρ_2 can be chosen using cross-validation.
- Divide the sample into V folds. For $1 \leq l \leq V$, let $S^{(l)}$ be the sample covariance on fold l, and $\hat{H}^{(-l)}$ be the output from data in folds other than l.
  - $\rho_{2,j} = \arg\max_\rho \sum_{l=1}^{V} \langle S^{(l)}, \hat{H}_j^{(-l)}(\rho_1, \rho) \rangle$.

Simulation: cross-validation method



Simulation: Localized eigenfunctions



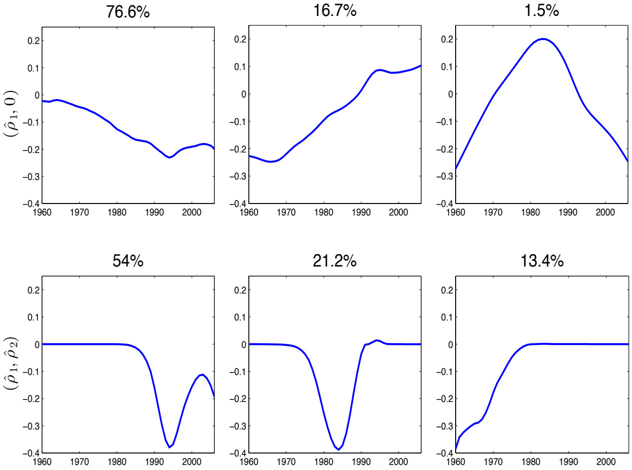
An Alternative Tuning of Localization Penalty

- Even when the true eigenvector is not sparse, we may still want sparsity for better interpretability.
- Given $\alpha \in (0,1)$, we can use the sparse penalty parameter such that the regularized eigenvector loses at most α proportion of explained variance compared to the non-regularized estimator.

$$\rho_{2,j} = \max \left\{ \rho : \frac{\hat{v}_j^T(\rho_1,\rho) S \hat{v}_j(\rho_1,\rho)}{\hat{v}_j^T(\rho_1,0) S \hat{v}_j(\rho_1,0)} \geq 1-\alpha \right\}.$$

- We still need to choose α, which is more interpretable than ρ_2.

The Mortality Rate Data



Mortality Data: different values of a

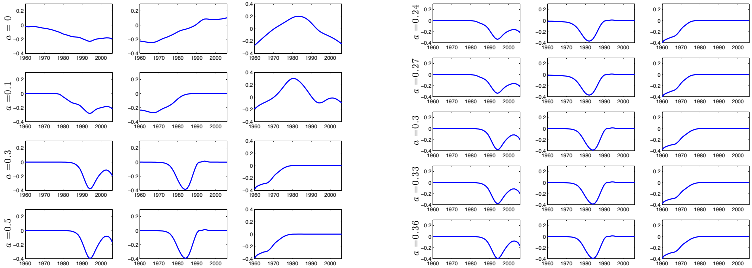


Figure : The estimated $\phi_j(t), j = 1, 2, 3$ for the mortality data with different values of a.

Summary of theoretical and numerical properties

- The proposed LFPCA converges to the original FPCA when the tuning parameters are chosen appropriately (Thm 4.1 & 4.2)
- The proposed LFPCA significantly improve the estimation accuracy when the eigenfunctions are truly supported on some subdomains.
- In the scenario that the original eigenfunctions are not localized, the proposed LFPCA also serves as a nice tool in finding orthogonal basis functions that balance between interpretability and the capability of explaining variability of the data.

Future & Ongoing Work

- Generalization to image data
- Use a differencing operator that ensures row-smoothness and column-smoothness
- Faster algorithms

Thank You!

Questions?