High-dimensional eigen-analysis, spiked models, and regularization

Debashis Paul

(*University of California at Davis, Dept. of Statistics*)

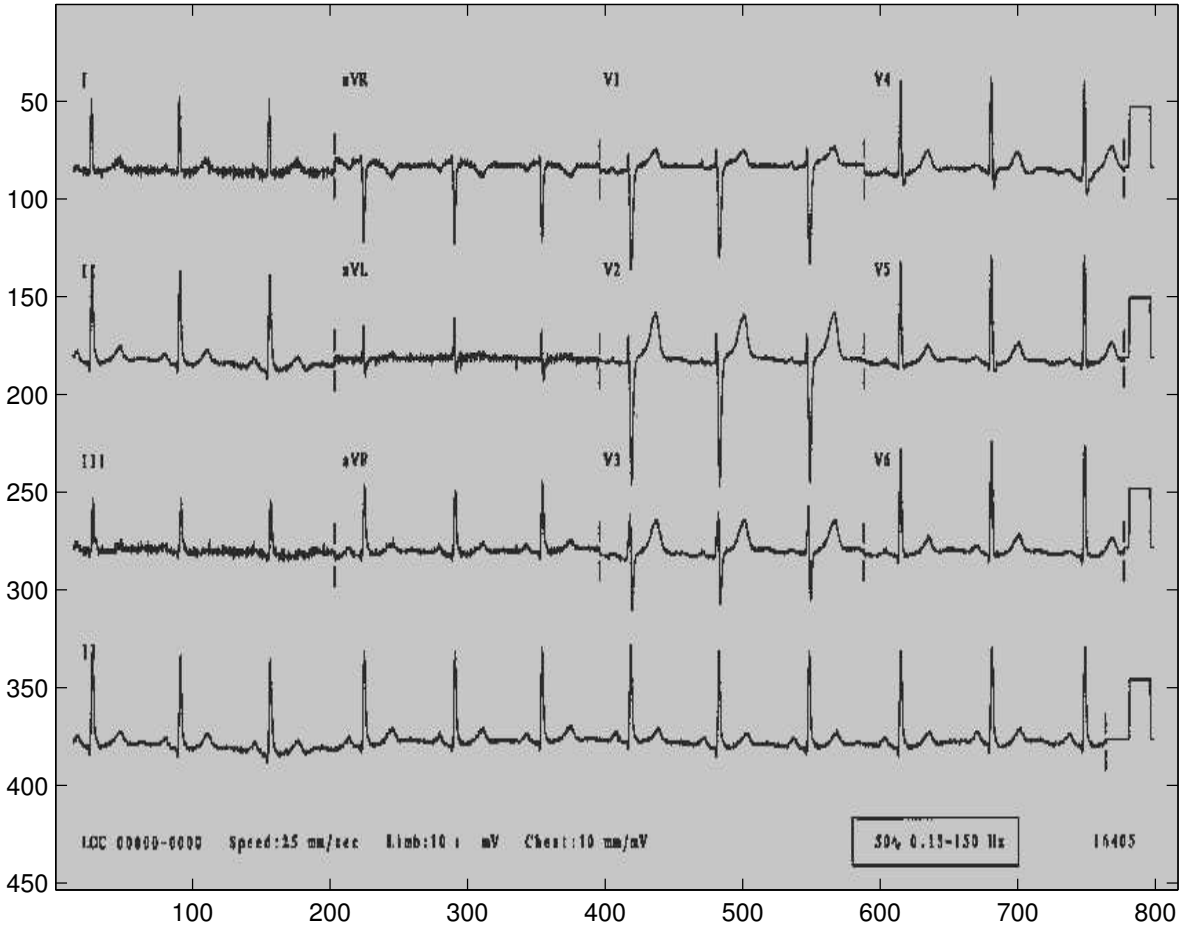
NCAR workshop

May 07, 2007

Agenda

- High dimensional multivariate data : examples
- **Review of standard PCA**
  - Objective
  - Drawbacks
- **Sparsity**
  - Motivation
  - Sparse PCA algorithm
  - Augmented Sparse PCA
  - Simulation study
  - Spatio-temporal extension

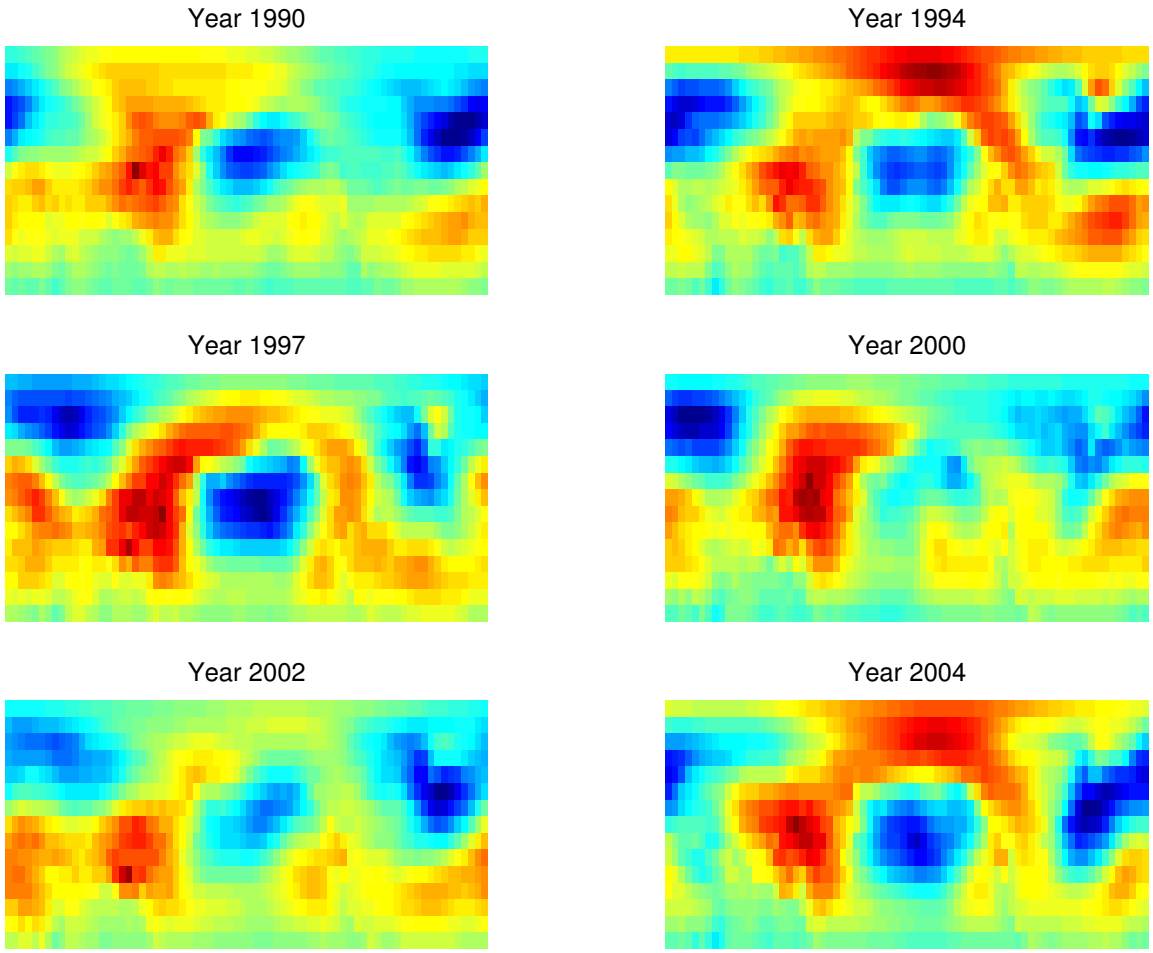
ECG data



Goal : Structural analysis

Monthly Average Sea Level Pressure : January

$p = 15 \times 72$, $n \approx 100$ (NCAR database)



Goal : Model diagnostics

Other examples

- Analysis of gene expression array (prediction of survival)
- Speech and image recognition (data compression, denoising)
- Financial data - stock prices (structural analysis)

Common characteristics

- High dimension (p)
- Comparatively low sample size (n)
- Noisy observations

A standard approach

- Represent the observation vectors in a "good" basis.
- Preserve the covariance structure.
- Accomplish this (approximately) with fewer basis elements.

Principal component analysis

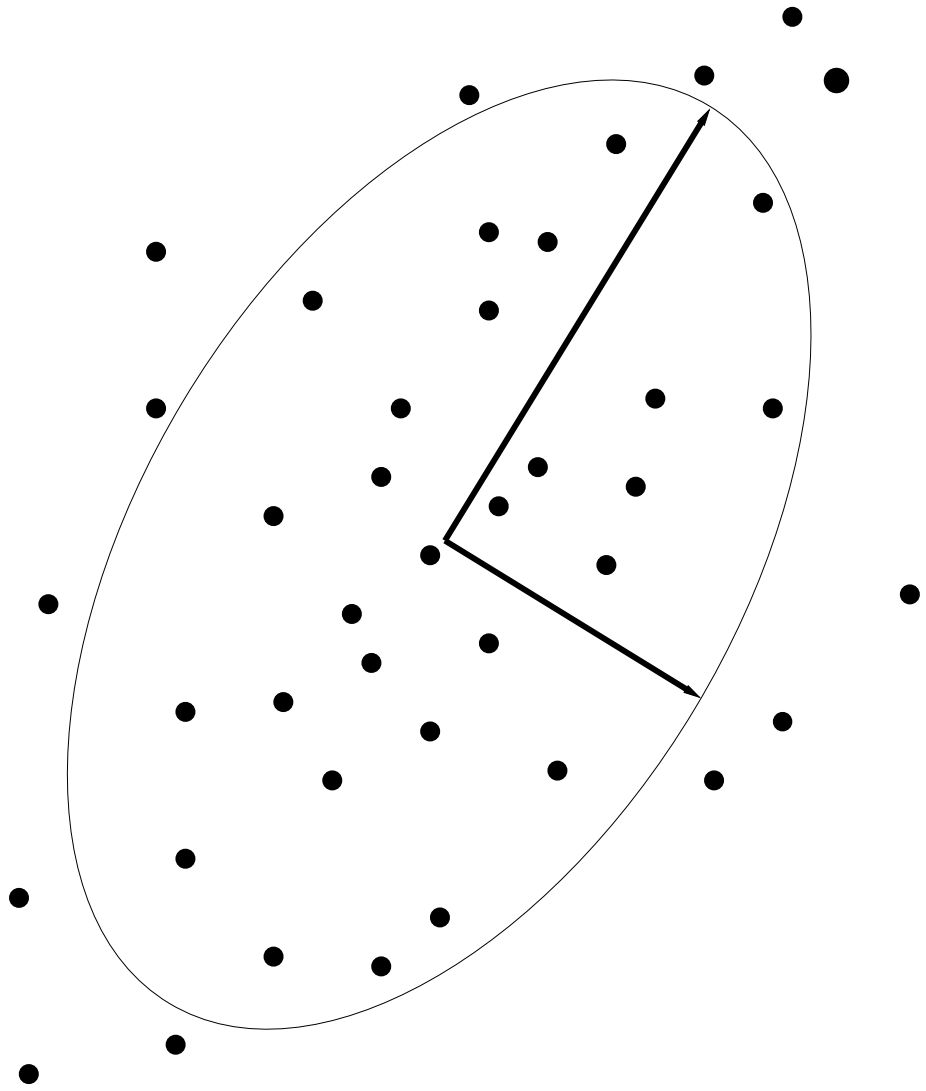
Data : X ($p \times n$ matrix) where

$$X_i,\ i = 1, \ldots, n, \qquad i.i.d.\ (\mu, \Sigma)$$

- **Strategy :** Find orthonormal basis vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots$ such that Variance($\mathbf{u}_\nu^t X$) is maximized subject to $\mathbf{u}_\nu \perp \mathbf{u}_1, \ldots, \mathbf{u}_{\nu-1}$.
- **Solution :** $\mathbf{u}_1, \mathbf{u}_2, \ldots$ are eigenvectors of Σ corresponding to eigenvalues $\ell_1 \geq \ell_2 \geq \ldots \geq \ell_p \geq 0$

$$\Sigma \mathbf{u}_\nu = \ell_\nu \mathbf{u}_\nu, \qquad \nu = 1, 2, \ldots, \text{rank}(\Sigma)$$

Principal component analysis : illustration



Standard PCA

- Define *sample covariance matrix*

$$\mathbf{S} := \frac{1}{n}(\mathbf{X} - \hat{\mu}\mathbf{1}^t)(\mathbf{X} - \hat{\mu}\mathbf{1}^t)^t, \qquad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Estimate eigenvectors and eigenvalues of Σ by eigenvectors and eigenvalues of $\mathbf{S}$.

Model I : High dimensional i.i.d. data

- **Large dimension :** e.g. p (dimension) and n (sample size) are such that $\frac{p}{n} \approx \gamma > 0$.
- **Model :** $X_1, \ldots, X_n$ i.i.d. $N(\mu, \Sigma)$, where $\Sigma = \Sigma_0 + \sigma^2 I$.
- **Components of variation :**

$$X_i = \mu + \sum_{\nu=1}^{M} \sqrt{\lambda_\nu} v_{\nu i} \mathbf{u}_\nu + \sigma Z_i, \qquad i = 1, \ldots, n$$

where $\lambda_1 \geq \ldots \geq \lambda_M > 0$ are the eigenvalues of Σ_0, and $\mathbf{u}_1, \ldots, \mathbf{u}_M$ are corresponding eigenvectors.

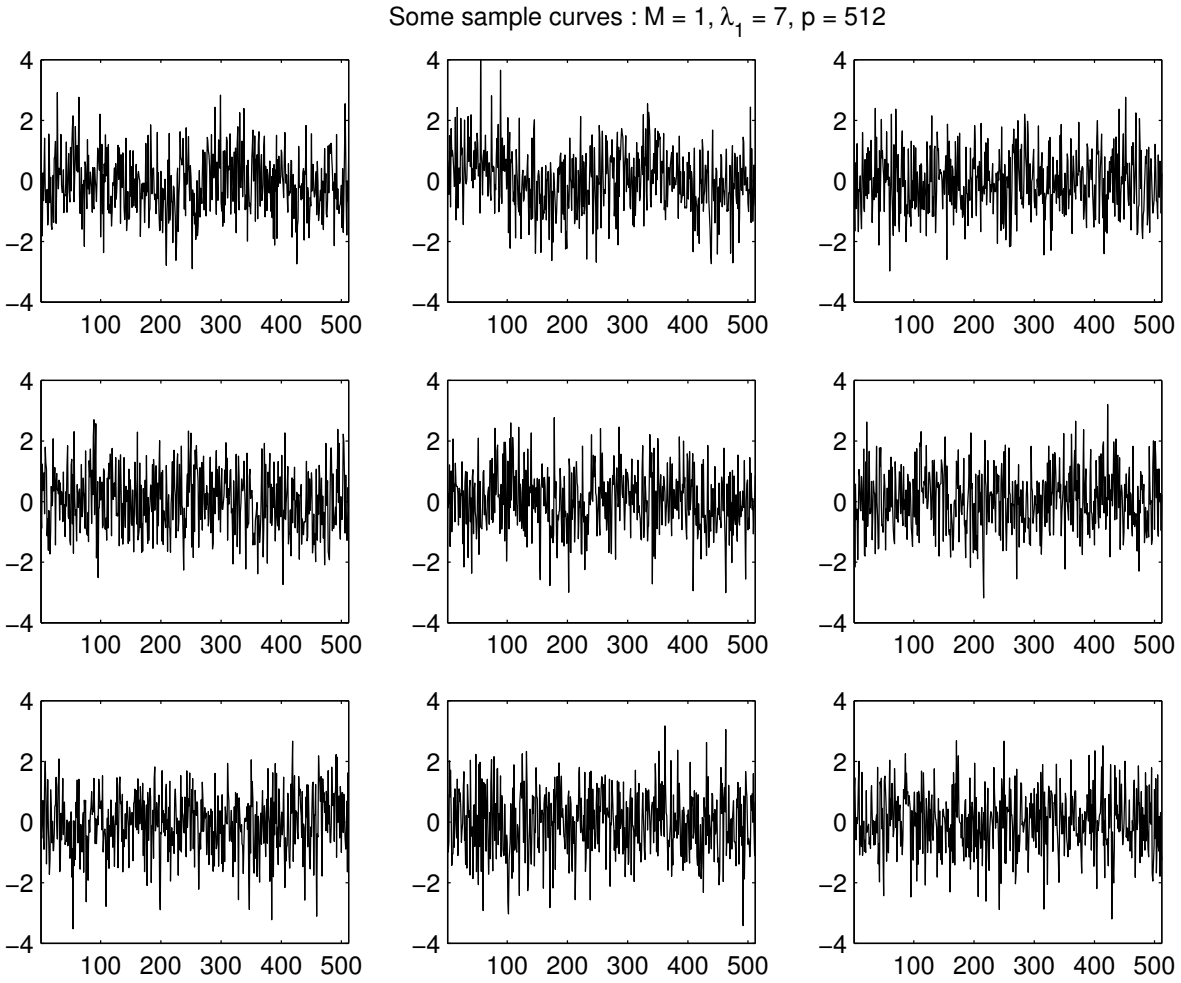
$\{v_{\nu i} : 1 \leq \nu \leq M\}$ i.i.d. $N(0,1)$ independent of $Z_i \sim N_p(0, I_p)$

- Interpretation as **Random Signal** (M dimensional) + **Isotropic Noise**

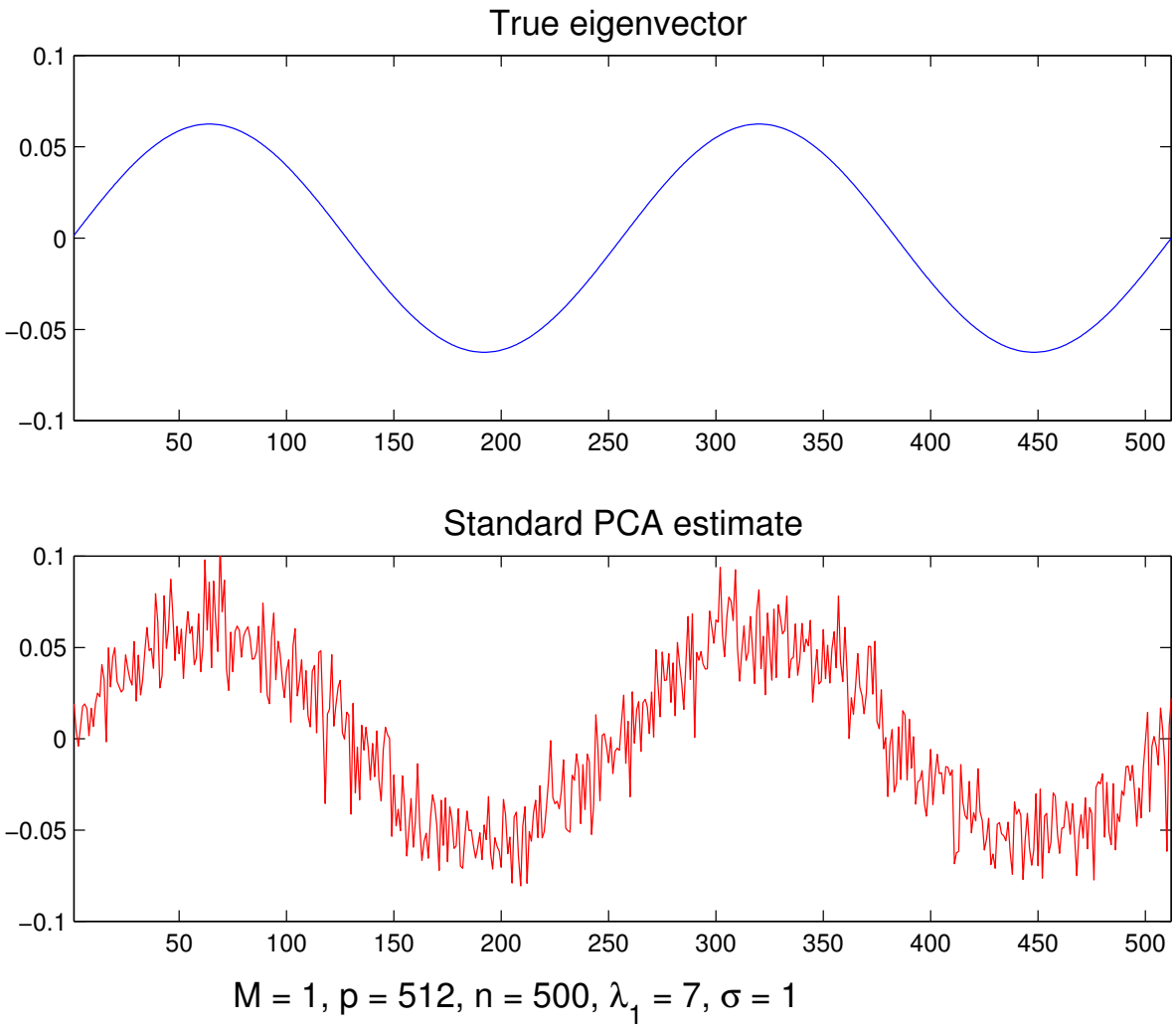
Drawback of standard PCA

- Noisy estimates for large dimension.
- High computational cost.

A toy example



Standard PCA estimate



Notion of consistency

An estimator $\widehat{\mathbf{u}}_\nu$ is *consistent* for *true* eigenvector $\mathbf{u}_\nu$ if

$$\angle(\mathbf{u}_\nu, \widehat{\mathbf{u}}_\nu) \to 0, \quad \text{in probability as } n \to \infty$$

Results on sample eigenvalues

(Paul (2005), Baik and Silverstein (2005), Onatski (2005)

Suppose eigenvalues of Σ $(p \times p)$ are $\ell_1 > \ldots > \ell_M > 1 = \ldots = 1$. (Note, $\ell_\nu = \lambda_\nu + \sigma^2$ and $\sigma = 1$). Let $\frac{p}{n} \to \gamma \in (0, 1)$.

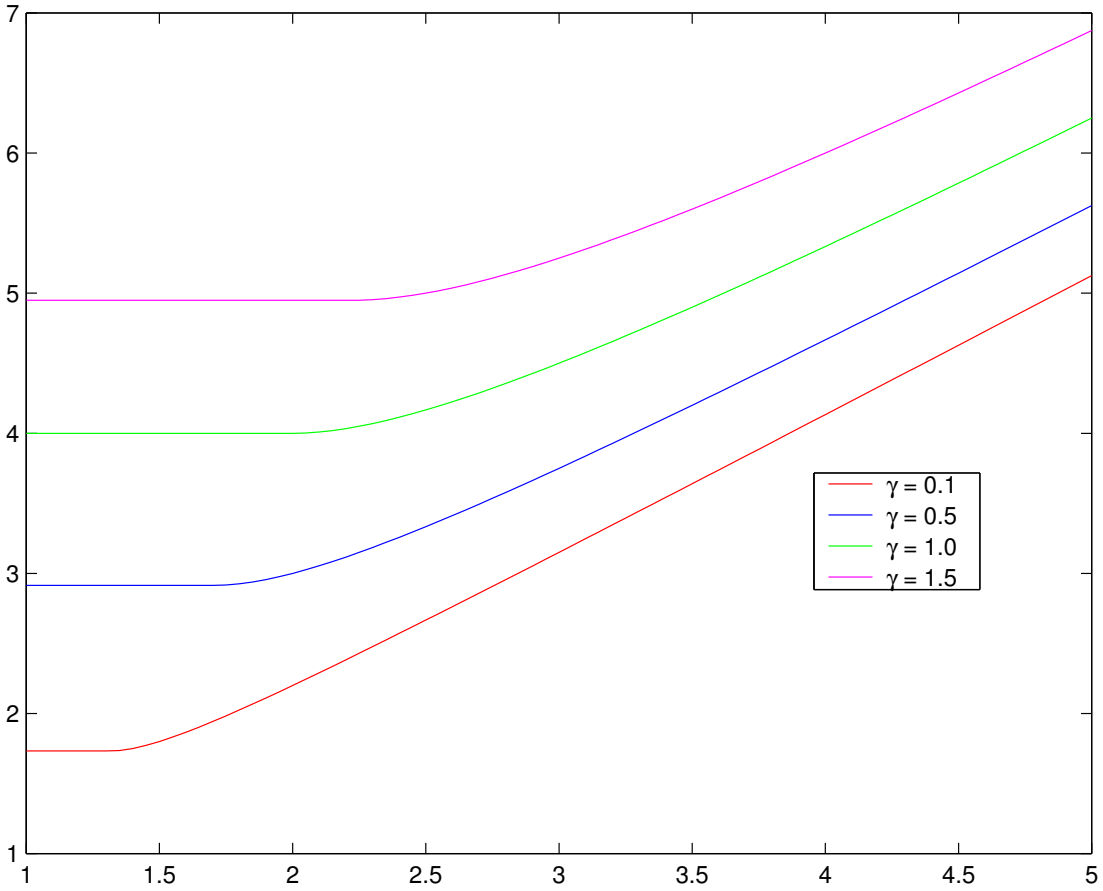
- **Phase transition :**

$$\widehat{\ell}_\nu \overset{a.s.}{\to} \begin{cases} (1+\sqrt{\gamma})^2 & \text{if } \ell_\nu \in (1, 1+\sqrt{\gamma}] \\ \rho_\nu := \ell_\nu \left(1 + \frac{\gamma}{\ell_\nu - 1}\right) & \text{if } \ell_\nu > 1+\sqrt{\gamma} \end{cases}$$

- **Asymptotic distribution :** If $\ell_\nu > 1+\sqrt{\gamma}$ and of multiplicity 1, and $|\frac{p}{n} - \gamma| = o(n^{-1/2})$,

$$\sqrt{n}(\widehat{\ell}_\nu - \rho_\nu) \Longrightarrow N(0, \widetilde{\sigma}_\nu^2), \quad \text{where } \widetilde{\sigma}_\nu^2 = 2\ell_\nu^2 \left(1 - \frac{\gamma}{(\ell_\nu - 1)^2}\right)$$

Phase transition of eigenvalues



$\ell_\nu \longrightarrow$

Asymptotics of sample eigenvectors

(Paul (2005), Hoyle and Rattray (2004), Onatski (2005)

Eigenvector associated with $\ell_\nu > 1$ of multiplicity one.

True: $\mathbf{u}_\nu$ Estimated: $\widehat{\mathbf{u}}_\nu$

- **Phase transition :**

$$|\langle \mathbf{u}_\nu, \widehat{\mathbf{u}}_\nu \rangle| \overset{a.s.}{\to} \begin{cases} 0 & \text{if } \ell_\nu \in (1, 1+\sqrt{\gamma}] \\ \sqrt{\left(1 - \frac{\gamma}{(\ell_\nu - 1)^2}\right) / \left(1 + \frac{\gamma}{\ell_\nu - 1}\right)} & \text{if } \ell_\nu > 1+\sqrt{\gamma} \end{cases}$$

Issues

- How to reduce noise from the estimates of eigenvectors, and achieve consistency ?
- How many components are to be used to represent the data ?
- Can these be done at a *low* computational cost ?

Smoothing based approach

- Assume that the observations are **noisy versions of smooth curves**
- While estimating the principal components, **add a penalty** to ensure that the estimate is reasonable smooth.
- Boils down to essentially applying a linear smoothing of the curves and then a further smoothing of eigenvectors of the covariance matrix of the smoothed curves.
- Needs specification of a smoothing parameter.

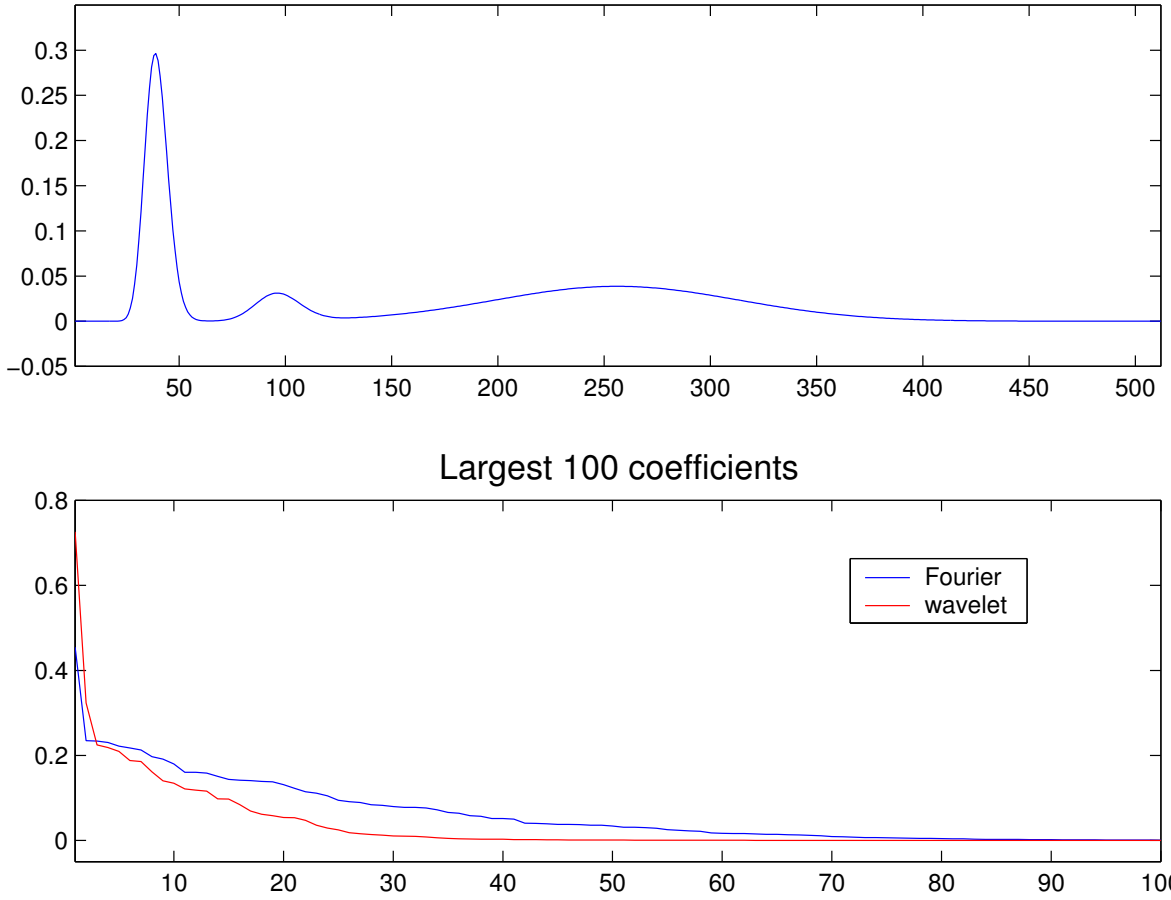
Reference : Rice and Silverman (1991), Kneip (1994), Silverman (1996), Cardot (2002), Ramsey and Silverman (1997).

Sparsity based approach

- Assume that the observations are **noisy versions of some random curves**
- Assume that the noiseless curves can be **sparsely represented** in some known orthogonal basis
- Data first transformed in an **orthogonal basis**
- Sparseness of the observation vectors in the transformed basis used for reducing noise

Reference : Johnstone and Lu (2004), Joliffe (2004), Zou and Hastie (2004), Paul (2005)

Sparsity : an illustration



$\sum |\textbf{wavelet coefficients}| = 3.732, \sum |\textbf{Fourier coefficients}| = 6.348$

Suitable basis $\equiv$ sparse representation

- **Heuristics :**

  If the eigenvectors $\mathbf{u}_\nu$ are **sparse** (in an appropriate basis), *i.e. have few significantly large coefficients*, then a smaller submatrix of Σ contains most of the useful information.

- **Motivation :**
  - Observations are smooth random functions $\Longrightarrow$ eigenvectors are smooth and can be sparsely represented in a wavelet/spline basis.
  - In many applications (e.g. microarray experiments) very few coordinates (genes) may carry any useful information.

Sparse PCA : Idea

With $\Sigma = \Sigma_0 + \sigma^2 I$,

$$\Sigma_{kk} = \sum_{\nu=1}^{M} \lambda_\nu \mathbf{u}_{\nu k}^2 + \sigma^2 = \Sigma_{0,kk} + \sigma^2$$

The sequence $(\Sigma_{0,kk}, k = 1, \ldots, p)$ is **sparse** if and only if the **eigenvectors** $\{\mathbf{u}_\nu, \nu = 1, \ldots, M\}$ **are sparse**.

Sparse PCA Algorithm (Lu, 2002)

Hereafter we assume that we already have a **sparse representation** of the observations in some **known basis**.

- Set a threshold $t_n > \sigma^2$. Select coordinate k if $\mathbf{S}_{kk} > t_n$. Call this set B_n.
- Perform PCA on the data $\{X_{li} : l \in B_n, i = 1, \ldots, n\}$.
- Extend the eigenvectors to full dimension by adding zeros to coordinates not in B_n.

Consistency

Result (Johnstone and Lu, 2004) :

Under assumption of Gaussianity, and if $\| \mathbf{u}_\nu \|_q \leq C_\nu < \infty$, for some $0 < q < 2$, then taking $t_n = \sigma^2(1 + c_0\sqrt{\frac{\log(p \vee n)}{n}})$, the estimate $\widehat{\mathbf{u}}_\nu$ obtained from SPCA scheme satisfies

$$\angle(\widehat{\mathbf{u}}_\nu, \mathbf{u}_\nu) \to 0 \qquad \text{a.s. as } n \to \infty$$

Here $\| \mathbf{u}_\nu \|_q \stackrel{def}{=} [\sum_{k=1}^{p} |\mathbf{u}_{\nu k}|^q]^{1/q}$

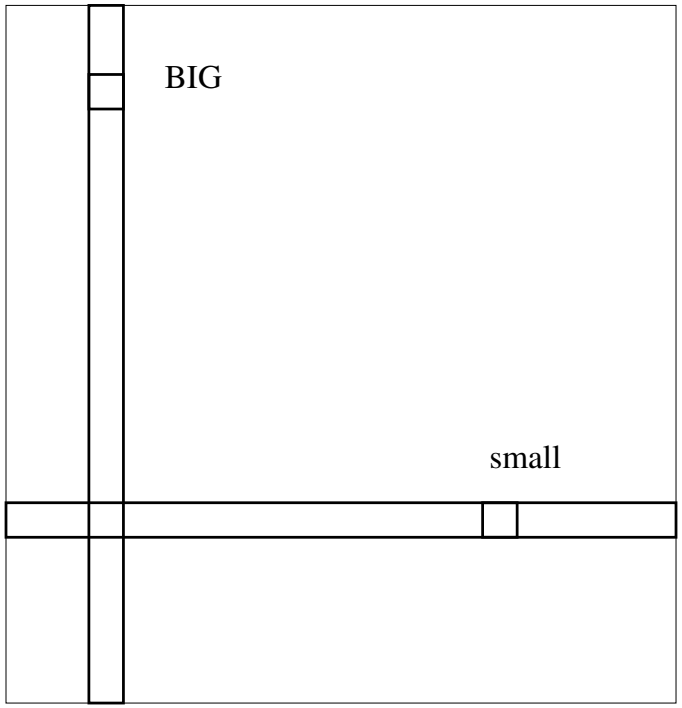
Drawback of Sparse PCA

- Sparse PCA does not use the covariance structure for selecting coordinates.
- Sparse PCA can only recover coordinates that are *rather large* (of size at least $c_0'\sigma(\frac{\log(p\vee n)}{n\lambda_\nu^2})^{1/4}$).
- Hence, if the eigenvectors are not *very sparse*, Sparse PCA scheme can have *large* bias.

Improved estimation scheme

Idea : If we have a "reasonably good" estimator of the eigenvectors $\mathbf{u}_\nu$'s then use that to utilize the **structure of covariance**.

$$M = 1,\ \Sigma_{0,bb} = \lambda_1 \mathbf{u}_{1b}^2,\ \Sigma_{0,ss} = \lambda_1 \mathbf{u}_{1s}^2,\ \Sigma_{0,bs} = \lambda_1 \mathbf{u}_{1b}\mathbf{u}_{1s}$$



Structure of covariance

- Three **classes of coordinates** (based on the size of their variance): $B \equiv \textit{big}$, $S \equiv \textit{small}$, $T \equiv \textit{tiny}$.
- $S \cup T = B^c$, $B \cup B^c = \{1, \ldots, p\}$.
- Partition Σ_0 and $\mathbf{u}_\nu$:

$$\Sigma_0 = \begin{bmatrix} \Sigma_{0,BB} & \Sigma_{0,BB^c} \\ \Sigma_{0,B^cB} & \Sigma_{0,B^cB^c} \end{bmatrix} \qquad \mathbf{u}_\nu = \begin{bmatrix} \mathbf{u}_{\nu,B} \\ \mathbf{u}_{\nu,B^c} \end{bmatrix} \qquad \mathbf{u}_{\nu,B^c} = \begin{bmatrix} \mathbf{u}_{\nu,S} \\ \mathbf{u}_{\nu,T} \end{bmatrix}$$

- Want to retain coordinates that are in $B \cup S$, but don't want to select coordinates that are in T.

Augmented SPCA

Stage I

- Get preliminary estimates $\widehat{\mathbf{u}}_\nu$'s by SPCA scheme. $\widehat{B}$ = set of selected coordinates. Let q_n be the size of the set $\widehat{B}$.
- Estimate M by retaining only those eigenvalues $\widehat{\ell}_1 \geq \ldots \geq \widehat{\ell}_{\min\{q_n,n\}}$ of $\mathbf{S}_{\widehat{B}\widehat{B}}$ that are above a threshold Λ_n.

Stage II

- Compute $\quad \widetilde{\mathbf{u}}_{\nu,\widehat{B}^c} := \widehat{\ell}_\nu^{-1} \mathbf{S}_{\widehat{B}^c \widehat{B}} \widehat{\mathbf{u}}_{\nu,\widehat{B}}, \quad \nu = 1, \ldots, \widehat{M}$
- Select a coordinate k from $\widehat{B}^c$ if

$$\sum_{\nu=1}^{\widehat{M}} \widehat{\ell}_\nu \widetilde{\mathbf{u}}_{\nu,k}^2 > \overline{t}_n$$

  for some threshold $\overline{t}_n$.
- $\widehat{S} :=$ coordinates selected at the second stage. Perform standard PCA on the submatrix of $\mathbf{S}$ (full data) corresp. to $\widehat{C} = \widehat{B} \cup \widehat{S}$.

Choice of threshold

- Stage I (Same as Sparse PCA) :

$$t_n = \sigma^2(1 + c_0\sqrt{\frac{\log(p \vee n)}{n}})$$

- Threshold for the eigenvalues (for estimating M) :

$$\Lambda_n = \sigma^2[(1 + \sqrt{\frac{q_n}{n}})^2 + c_1\sqrt{\frac{\log(p \vee n)q_n}{n}}]$$

- Stage II :

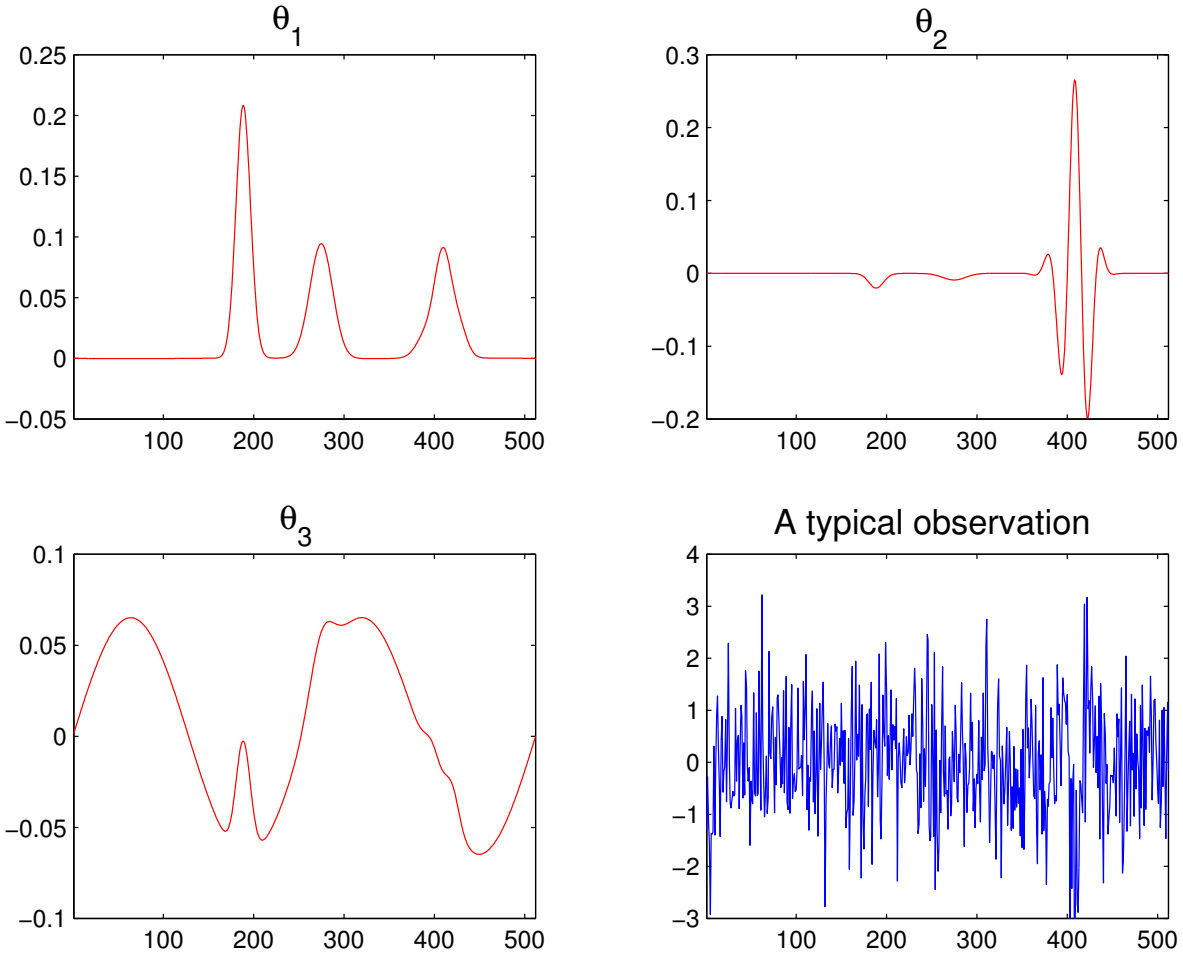
$$\bar{t}_n = \sigma^2 \; c_2\frac{\log(p \vee n)}{n}$$

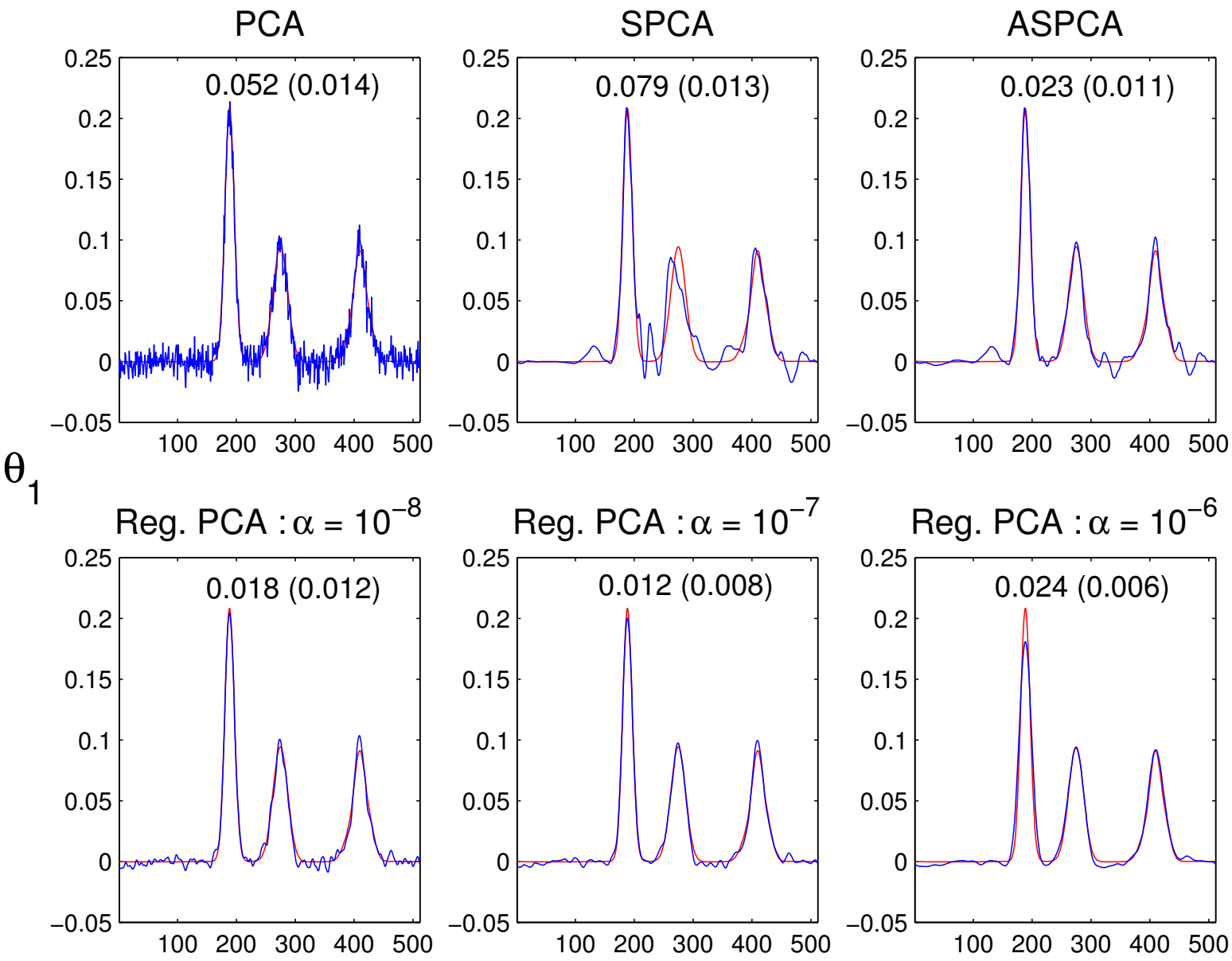
Computational efficiency

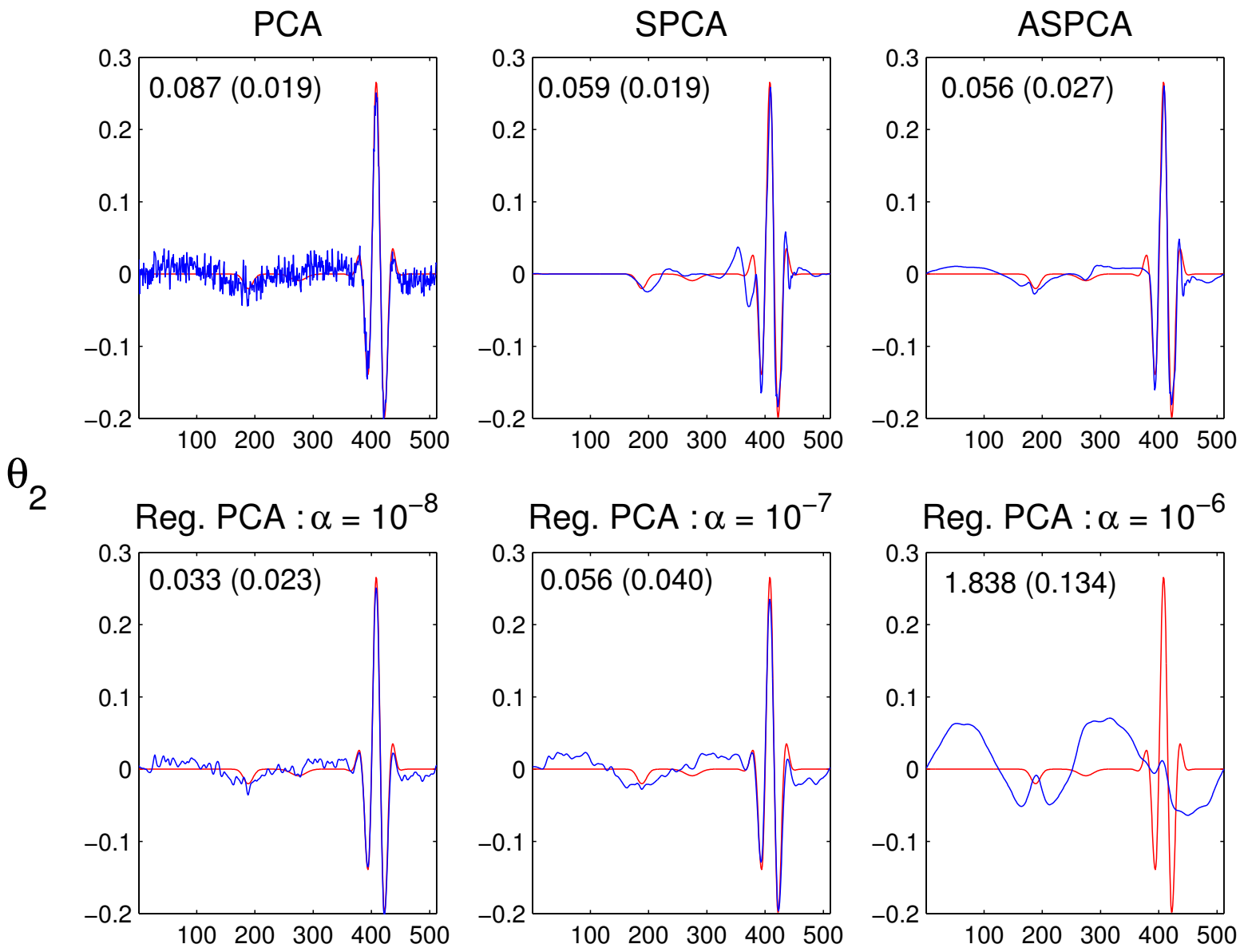
- ASPCA requires $O(\min\{nk_n^2, n^2k_n\})$ computations, where k_n is the number of selected coordinate.
- Typically k_n is a lot smaller than p (due to **sparsity**).
- Much faster than standard PCA or smoothed (regularized) PCA (Ramsay and Silverman, 1997).

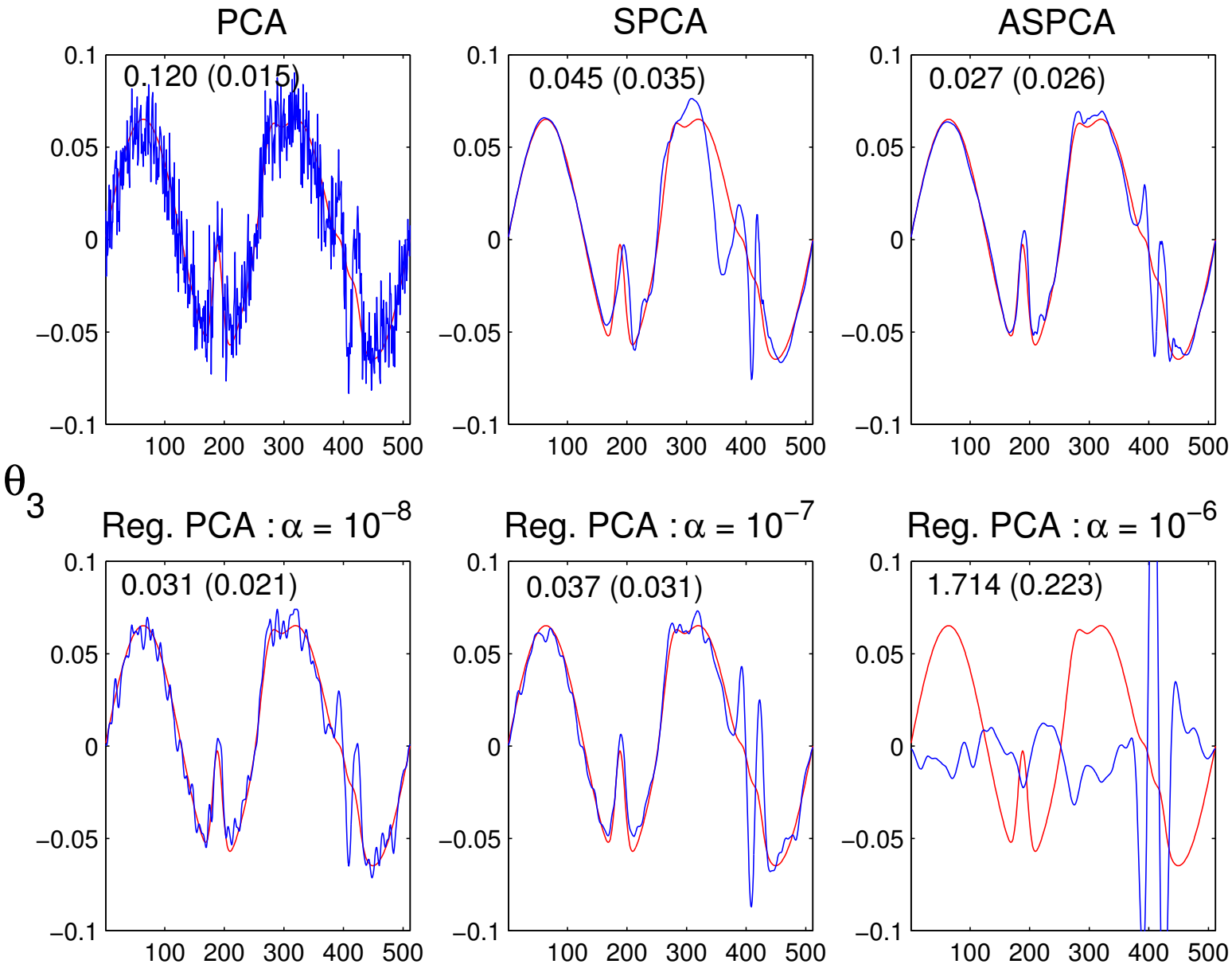
Simulation study

Dimension $p = 512$, sample size $n = 500$. $M = 3$, $\lambda_1 = 25$, $\lambda_2 = 15$, $\lambda_3 = 10$, $\sigma = 1$.




PCA
SPCA
ASPCA
0.052 (0.014)
0.079 (0.013)
0.023 (0.011)
θ1
Reg. PCA : α = 10−8
Reg. PCA : α = 10−7
Reg. PCA : α = 10−6
0.018 (0.012)
0.012 (0.008)
0.024 (0.006)
0.25
0.2
0.15
0.1
0.05
0
−0.05
100 200 300 400 500


PCA
SPCA
ASPCA
0.087 (0.019)
0.059 (0.019)
0.056 (0.027)
θ_2
Reg. PCA : $\alpha = 10^{-8}$
Reg. PCA : $\alpha = 10^{-7}$
Reg. PCA : $\alpha = 10^{-6}$
0.033 (0.023)
0.056 (0.040)
1.838 (0.134)
0.3
0.2
0.1
0
−0.1
−0.2
100 200 300 400 500


PCA
SPCA
ASPCA
0.120 (0.015)
0.045 (0.035)
0.027 (0.026)
θ3
Reg. PCA : α = 10−8
Reg. PCA : α = 10−7
Reg. PCA : α = 10−6
0.031 (0.021)
0.037 (0.031)
1.714 (0.223)
0.1
0.05
0
−0.05
−0.1
100 200 300 400 500

Analysis

Loss function :

Squared error type loss

$$L(u,v) = 2(1 - |\langle u, v\rangle|) = \| u - sign(\langle u, v\rangle)v \|^2 = 4\sin^2 \frac{1}{2}\angle(u,v)$$

for $u, v \in \mathbb{S}^{p-1}$.

Assumptions :

- $\lambda_1 \to \lambda_{1,\infty} \in [0,\infty]$, $\frac{\lambda_\nu}{\lambda_1} \to r_\nu$ for $\nu = 1, \ldots, M$ with $1 = r_1 > r_2 > \ldots > r_M > 0$.
- $c_1 \log n \leq \log p \leq c_2 \log n$ for $c_2 \geq 1 \geq c_1 > 0$.

Two important functions

$$h(x) = \frac{x^2}{1+x}, \quad x > 0, \qquad g(x,y) = \frac{(x-y)^2}{xy}, \quad x \neq y > 0$$

(**nonparametric component**) (**nuisance parameters**)

Risk behavior : standard PCA

Let $\Theta = \{\mathbf{u}_\nu \text{ orthornormal } : \mathbf{u}_\nu \in \mathbb{S}^{p-1}, \nu = 1, \ldots, M\}$

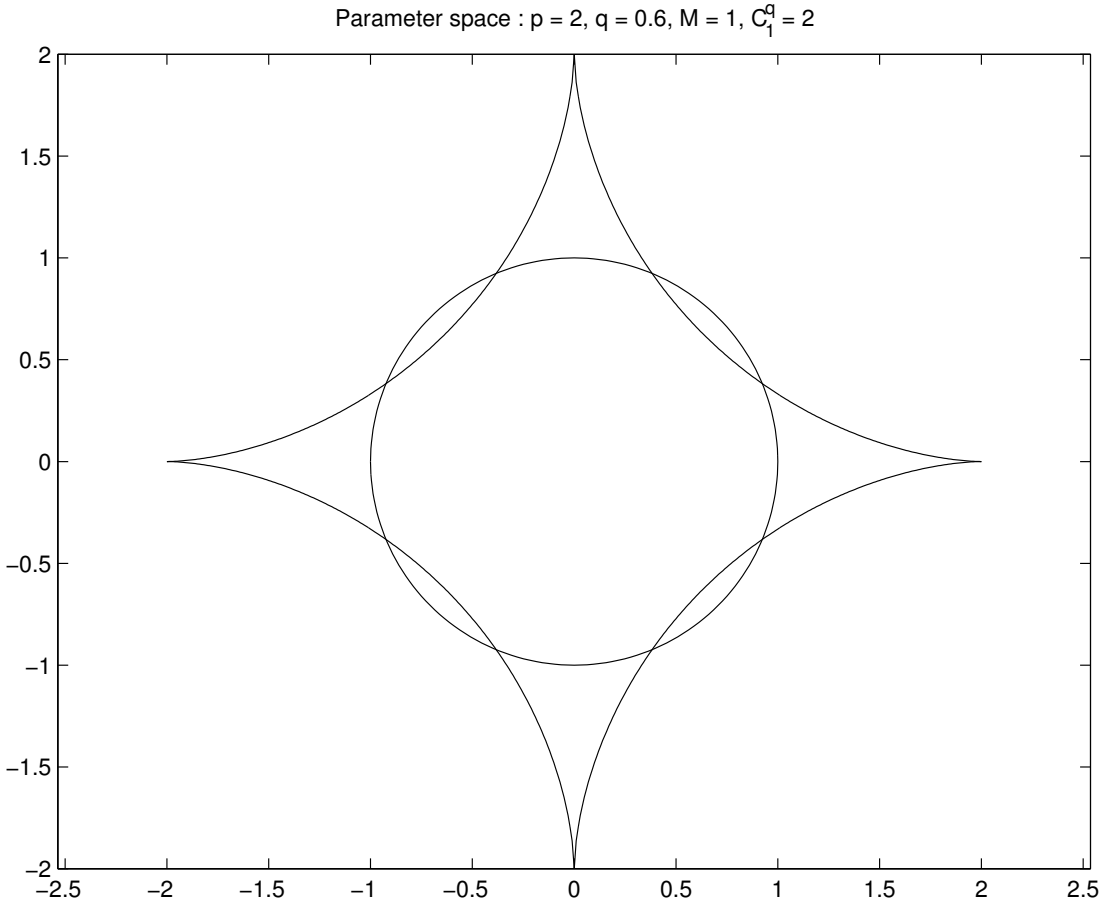
W.l.o.g. take $\sigma = 1$.

$$\sup_{\mathbf{u} \in \Theta} \mathbb{E} L(\widehat{\mathbf{u}}_\nu, \mathbf{u}_\nu) \leq [\frac{p}{nh(\lambda_\nu)} + \frac{1}{n} \sum_{\nu \neq \nu'} \frac{1}{g(1+\lambda_\nu, 1+\lambda_{\nu'})}](1 + o(1))$$

whenever RHS converges to 0.

Parameter space

Assume $\Theta_q = \{\mathbf{u}_\nu \in \mathbb{S}^{p-1} \text{ orthonormal } : \ \| \mathbf{u}_\nu \|_q^q \leq C_\nu^q, \ \nu = 1, \ldots, M\}$.



Risk behavior : Lower bound

$$\inf_{\widehat{\mathbf{u}}_\nu} \sup_{\mathbf{u}\in\Theta_q} \mathbb{E}L(\widehat{\mathbf{u}}_\nu, \mathbf{u}_\nu) \geq K \left[\max\{\delta_n, \frac{1}{n}\max_{\nu'\neq\nu} \frac{1}{g(1+\lambda_\nu, 1+\lambda_{\nu'})}\} \wedge 1\right] \quad (1)$$

where

$$\delta_n = \begin{cases} \frac{C_\nu^q (\log p)^{1-q/2}}{(nh(\lambda_\nu))^{1-q/2}} & \text{if } \frac{C_\nu^q (nh(\lambda_\nu))^{q/2}}{(\log p)^{q/2}} = O(p^{1-\alpha}), \quad \textbf{highly sparse} \\ \frac{C_\nu^q}{(nh(\lambda_\nu))^{1-q/2}} & \text{if } C_\nu^q (nh(\lambda_\nu))^{q/2} \leq p, \qquad \textbf{sparse} \\ \frac{p}{nh(\lambda_\nu)} & \text{if } C_\nu^q (nh(\lambda_\nu))^{q/2} \geq p, \qquad \textbf{dense} \end{cases}$$

where $\alpha \in (0, 1)$.

Understanding lower bound

- Note that δ_n is of the form $m_n \tau_n^2$, where

$$m_n = \begin{cases} [\frac{C^2 nh(\lambda_\nu)}{\log p}]^{q/2} \\ [C^2 nh(\lambda_\nu)]^{q/2} \\ p \end{cases} \qquad \tau_n = \begin{cases} \frac{\sqrt{\log p}}{\sqrt{nh(\lambda_\nu)}} & \textbf{highly sparse} \\ \frac{1}{\sqrt{nh(\lambda_\nu)}} & \textbf{sparse} \\ \frac{1}{\sqrt{nh(\lambda_\nu)}} & \textbf{dense} \end{cases}$$

- **Interpretation :**

  $m_n =$ **effective dimension**

  $\tau_n = (\sqrt{\log p}) \times$ **asymptotic noise per parameter**

Noteworthy features

- Three regimes of sparsity - *highly sparse*, *sparse*, *dense*.
- *Dense* $\equiv$ *no knowledge* about the eigenvectors $\Longrightarrow$ no estimator *uniformly consistent* over $\mathbb{S}^{p-1}$ when $\frac{p}{nh(\lambda_\nu)} \not\to 0$.
- Standard PCA has *optimal rate of convergence* in the *dense* case.
- Dependence on the *spacings* between eigenvalues

Risk behavior : Augmented SPCA estimates

$$\sup_{\mathbf{u}\in\Theta_q} \mathbb{E}L(\widehat{\mathbf{u}}_\nu, \mathbf{u}_\nu) \quad \leq \quad K(q)[\overline{r}_\nu^{-q/2} \frac{(\sum_{\nu'=1}^{M} \overline{r}_{\nu'}^{q/2} C_{\nu'}^q)(\log p)^{1-q/2}}{(nh(\lambda_\nu))^{1-q/2}}$$
$$+\frac{1}{n}\sum_{\nu\neq\nu'} \frac{1}{g(1+\lambda_\nu, 1+\lambda_{\nu'})} \ ] \ (1+o(1))$$

provided the risk of for SPCA estimates converge to zero. Here $\overline{r}_\nu = h(\lambda_\nu)/h(\lambda_1)$.

- Does not depend on p except for a $\log p$ factor.
- In the case $M = 1$, upper bound of the form $K(q)m_n\overline{\tau}_n^2(1+o(1))$, where

$$m_n = \left[\frac{C^2 nh(\lambda_\nu)}{\log p}\right]^{q/2}, \qquad \overline{\tau}_n = \frac{\sqrt{\log p}}{\sqrt{nh(\lambda_\nu)}}$$

Summary of the results

When sparsity is measured over parameter spaces that impose constraint on the l^q norm $(0 < q < 2)$ of the eigenvectors,

- There are different regimes of sparsity that give rise to different rates of convergence
- ASPCA achieves the optimal rate in the highly sparse case
- Usual PCA can achieve the optimal rate if there are sufficiently many observations
- Without enough sparsity it may not be possible to estimate the eigenvectors uniformly consistently

Beyond i.i.d. setting

- Separable space-time covariance.
- Dynamic spatio-temporal systems.

Model II : Separable Spatio-temporal process

- Apart from spatial (coordinate-wise) dependence, there is also a "time"-dependence - observations $X_1, \ldots, X_n$ are **not** independent.
- **Separable spatio-temporal model :** assume that spatial variability and temporal variability do not depend on each other.

  i.e., the $N \times n$ data matrix $\mathbf{X}_n = [X_1 : \ldots : X_n]$ has covariance $\Sigma_N \otimes \Delta_n$ where Σ_N is $N \times N$ and Δ_n is $n \times n$.

Separable "Spiked" model

Σ_N has eigenvalues $\ell_1 \geq \ldots \geq \ell_M > 1 = \ldots = 1$

Empirical distribution of the eigenvalues of Δ_n converges to a distribution G compactly supported on $\mathbb{R}^+$ (and satisfies some technical condition).

Let F^{Δ} denote the limiting ESD of the matrix $\frac{1}{n}\mathbf{Z}\Delta_n\mathbf{Z}^T$ where $\mathbf{Z}$ is $(N-M)\times n$ with i.i.d. $N(0,1)$ entries.

Result : There is a threshold $\tau_{c,\Delta}$ such that, if $\ell_k > \tau_{c,\Delta}$ then $\widehat{\ell}_k \overset{a.s.}{\to} \rho_k$ where ρ_k satisfies

$$\rho_k = \ell_k \int \frac{t}{1 - t\alpha_1(\rho_k, c)} dG(t)$$

where

$$\alpha_j(\rho, c) = c \int \frac{1}{(\rho - x)^j} dF^{\Delta}(x), \quad j = 1, 2, \ldots.$$

for $\rho > \max\ supp(F^{\Delta})$.

Asymptotic normality

Result : Under the stated assumptions (plus a few regularity conditions), if ℓ_k is of multiplicity one, then

$$\sqrt{n}(\widehat{\ell}_k - \rho_k) \Longrightarrow N(0, \sigma^2(\ell_k))$$

where

$$\sigma^2(\ell_k) = 2\ell_k^2 \left[\frac{\beta(\rho_k, c)(1 + \alpha_2(\rho_k, c)\beta(\rho_k, c))}{(1 + \ell_k \alpha_2(\rho_k, c)\beta(\rho_k, c))^2} \right]$$

with

$$\beta(\rho_k, c) := \int \frac{t^2}{(1 - t\alpha_1(\rho_k, c))^2} dG(t).$$

Analogous results for sample eigenvectors (a.s. limit, asymptotic normality).

Model III : Autoregressive process

- A *Vector Autoregressive Process of order 1 (VAR(1))* in dimension N is given by

$$X_t = AX_{t-1} + \varepsilon_t, \qquad t = \ldots, -2, -1, 0, 1, 2, \ldots,$$

  where $\varepsilon_t \in \mathbb{R}^N$ are i.i.d. with mean 0 and covariance matrix Ψ, $A \in \mathbb{R}^{N \times N}$.

- **Special case :** $A = A^T$ and $\Psi = I_N$. If $\| A \| < 1$, then $\{X_t\}$ is stationary and $\Sigma_N := \mathrm{Var}(X_t) = (I - A^2)^{-1}$.

- **Spiked model :** If $A = A^T$, $\Psi = I_N$, $\| A \| < 1$ and $\mathrm{rank}(A) = r$ fixed and $< N$, then Σ has eigenvalues $\ell_j = \frac{1}{1-\sigma_j^2(A)}$, $j = 1, \ldots, M$, and $\ell_j = 1$ for $j = M+1, \ldots, N$.

Asymptotics for eigenvalues

- We observe $X_1, \ldots, X_n$. Assume that $n = n(N)$ is such that $\frac{N}{n} \to c \in (0, 1)$.
- If $j \in \{1, \ldots, r\}$ is such that $\ell_j > \tau_c$, for some $\tau_c > 0$, then

$$\widehat{\ell}_j \overset{a.s.}{\to} \rho_j \qquad \text{with} \quad \rho_j = \frac{\ell_j}{1 - g_c(\rho_j)},$$

  with $g_c(\rho) = c \int \frac{1}{\rho - x} dF_c^{MP}(x)$.

- Asymptotic normality of j-th sample eigenvalue and corresponding eigenvector if $\ell_j > \tau_c$ and has multiplicity 1.

Research directions

- Exploration of the relationship between nonparametric regression and estimation of eigenvectors
- Different approach to regularization - e.g. direct penalization
- Confidence balls for eigenvectors for high-dimensional covariances
- Extension to spatio-temporal problems (for example, those arising in geophysical and meteorological applications)
- Application to signal estimation for sensor networks under noisy background
- Application to spectro-temporal problems in chemometrics