Eigenvalue-free risk bounds for PCA projectors

Andre Mas
IMAG, Univ. Montpellier
joint work with many people (G. Biau, E. Brunel, C. Crambes, N. Hilgert, A. Roche, F. Ruymgaart, N. Verzelen)

CIRM, February 2016

Random Projections vs PCA

Two kinds of projections usually encountered in ML :

- Principal components/PCA and variants : projection space depends on the data. Real projections i.e. $P^2 = P$.

Random Projections vs PCA

Two kinds of projections usually encountered in ML :

- Principal components/PCA and variants : projection space depends on the data. Real projections i.e. $P^2 = P$.
- "Random projections" (RP) : projection space independent from the data. Not real projections but rectangle matrices with nrows<ncols filled with i.i.d. (Gaussian/Rademacher) entries.

RP are used in sparse recovery, compressed sensing, etc. Allow effective dimension reduction for sparse signals through Johnson-Lindenstrauss Lemma and Restricted Isometry Property.

Focus on functional PCA

- fPCA=PCA adapted to samples of random functions. Adapted to high-dimensional data as well.
- Principle does not change : retrieve the spectrum (eigenvalues-eigenvectors) of a covariance operator

Focus on functional PCA

- fPCA=PCA adapted to samples of random functions. Adapted to high-dimensional data as well.
- Principle does not change : retrieve the spectrum (eigenvalues-eigenvectors) of a covariance operator
- Main feature : optimal linear reconstruction. Consider X random with cov. matrix Σ and $\mathcal{P}_k$ the set of orth. projections with rank k. Then, under general conditions :

$$\min_{P_k \in \mathcal{P}_k} \mathbb{E} \left\| X - P_k X \right\|^2 = \min_{P_k \in \mathcal{P}_k} \mathrm{tr} \left[P_k^{\perp} \Sigma P_k^{\perp} \right]$$

is obtained for principal components of X (= s.v.d. of Σ here).

Focus on functional PCA

- fPCA=PCA adapted to samples of random functions. Adapted to high-dimensional data as well.
- Principle does not change : retrieve the spectrum (eigenvalues-eigenvectors) of a covariance operator
- Main feature : optimal linear reconstruction. Consider X random with cov. matrix Σ and $\mathcal{P}_k$ the set of orth. projections with rank k. Then, under general conditions :

$$\min_{P_k \in \mathcal{P}_k} \mathbb{E} \|X - P_k X\|^2 = \min_{P_k \in \mathcal{P}_k} \operatorname{tr}\left[P_k^{\perp} \Sigma P_k^{\perp}\right]$$

  is obtained for principal components of X (= s.v.d. of Σ here).
- Numerous variants in ML : sparse PCA, kernel/non linear PCA...

Notations and Set-up

$\mathbf{H} = \mathbf{L}^2([0,1])$ or $\mathbf{W}^{2,m}([0,1])$ with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$
$X_1, X_2, ...X_n$ sample of i.i.d random elements with values in $\mathbf{H}$ and $\mathbb{E}X_1 = 0$:

$$\Sigma = \mathbb{E}(X_1 \otimes X_1) = \mathbb{E}\langle X_1, \cdot \rangle X_1 = \sum_{i=1}^{+\infty} \lambda_k (\varphi_k \otimes \varphi_k),$$

$$\Sigma_n = \frac{1}{n} \sum_{k=1}^{n} X_k \otimes X_k \quad \Sigma_n \widehat{\varphi}_i = \widehat{\lambda}_i \widehat{\varphi}_i$$

- $(\lambda_k)_{k \in \mathbb{N}^*} \in l_1^+$,with $\lambda_1 > ... > \lambda_k > ... > 0$
- Rank one projector : $\pi_i = \varphi_i \otimes \varphi_i$.
- Projection onto $\mathrm{span}\{\varphi_1, ..., \varphi_k\}$: $\mathbf{P}_k = \sum_{i=1}^k \pi_i$.
- Empirical counterparts : $\widehat{\pi}_i = \widehat{\varphi}_i \otimes \widehat{\varphi}_i$ and $\widehat{\mathbf{P}}_k = \sum_{i=1}^k \widehat{\pi}_i$.

Smoothness of processes and eigenvalues decay

Karhunen-Loeve development : $X = \sum_{j=1}^{+\infty} \sqrt{\lambda_j}\xi_j\varphi_j$
[Kind of random generalized Fourier series]

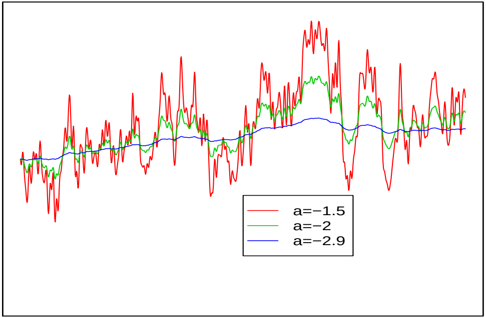


Figure: Three examples of random curves with $\lambda_j = j^a$ and the same ξ_j's.

Motivation-Goals

- $\Sigma \approx$ infinite size matrix, Σ trace-class, Σ^{-1} unbounded (not coercive)
- No specific model on X (e.g: signal$+$ noise) or Σ (spiked covariance structure not suited to functional setting)

Motivation-Goals

- $\Sigma \approx$ infinite size matrix, Σ trace-class, Σ^{-1} unbounded (not coercive)
- No specific model on X (e.g: signal+ noise) or Σ (spiked covariance structure not suited to functional setting)
- Bound with high probability the norms of : $\widehat{\pi}_i - \pi_i$ and $\widehat{\mathbf{P}}_k - \mathbf{P}_k$

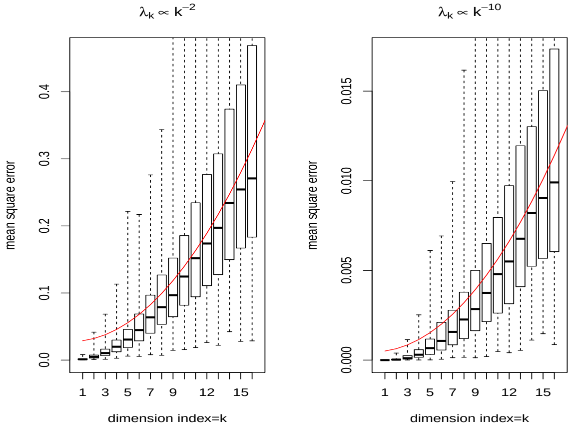
Motivation-Goals

- $\Sigma \approx$ infinite size matrix, Σ trace-class, Σ^{-1} unbounded (not coercive)
- No specific model on X (e.g: signal+ noise) or Σ (spiked covariance structure not suited to functional setting)
- Bound with high probability the norms of : $\widehat{\pi}_i - \pi_i$ and $\widehat{\mathbf{P}}_k - \mathbf{P}_k$
- Usually upper bounds are (roughly) $\|\widehat{\pi}_i - \pi_i\|_\infty^2 \prec 1/\left(n\lambda_i\right)$ or $1/\left(n\delta_i\right)$ with $\delta_i = \min\left(|\lambda_i - \lambda_{i+1}|, |\lambda_{i-1} - \lambda_i|\right)$
- Previous bound degenerates if $\lambda_i = e^{-mi}$ or $\lambda_i = i^{-\alpha}$ for large α
- Not that shocking if eig.vec. estimation seen as fixed point pb $\Sigma_n \hat{\varphi}_i / \hat{\lambda}_i = \hat{\varphi}_i$. If $i \uparrow \infty$ pb no more well posed $\mapsto$ price to pay...

Motivation-Goals

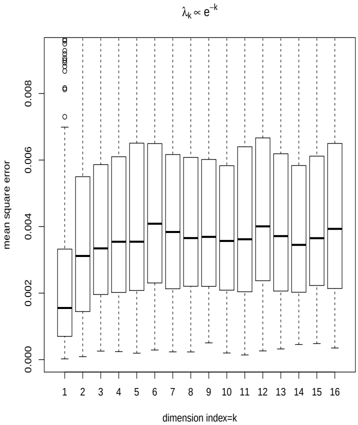
- $\Sigma \approx$ infinite size matrix, Σ trace-class, Σ^{-1} unbounded (not coercive)
- No specific model on X (e.g: signal+ noise) or Σ (spiked covariance structure not suited to functional setting)
- Bound with high probability the norms of : $\widehat{\pi}_i - \pi_i$ and $\widehat{\mathbf{P}}_k - \mathbf{P}_k$
- Usually upper bounds are (roughly) $\|\widehat{\pi}_i - \pi_i\|_\infty^2 \prec 1/\left(n\lambda_i\right)$ or $1/\left(n\delta_i\right)$ with $\delta_i = \min\left(\left|\lambda_i - \lambda_{i+1}\right|, \left|\lambda_{i-1} - \lambda_i\right|\right)$
- Previous bound degenerates if $\lambda_i = e^{-mi}$ or $\lambda_i = i^{-\alpha}$ for large α
- Not that shocking if eig.vec. estimation seen as fixed point pb $\Sigma_n \hat{\varphi}_i / \hat{\lambda}_i = \hat{\varphi}_i$. If $i \uparrow \infty$ pb no more well posed $\mapsto$ price to pay...
- But paradox ? X may be concentrated on low dimensional space $\rightarrow$ convergence of $\widehat{\pi}_i$ could be slow even for i in low dimensions.
- Why not some simulations now ?

Some simulations before theory...



MC simulation of $\mathbb{E}\,\|\widehat{\pi}_k - \pi_k\|_\infty^2$ with $M = 500$ replications, sample size $n = 500$ and $X = \sum_{j=1}^{kmax} \sqrt{\lambda_j}\xi_j e_j$ with $kmax = 20$, ξ_j Gaussian, e_j Fourier basis and $\mathrm{tr}\Sigma = 1$. Red curve : k^2 fit

Some simulations before theory



MC simulation of $\mathbb{E}\left\|\widehat{\pi}_k - \pi_k\right\|_\infty^2$ with $M = 500$ replications, sample size $n = 500$ and $X = \sum_{j=1}^{kmax} \sqrt{\lambda_j}\xi_j e_j$ with $kmax = 20$, ξ_j Gaussian, e_j Fourier basis and $\mathrm{tr}\Sigma = 1$.

From covariance to projectors : perturbation theory and functional calculus

Notice $\left|\widehat{\lambda}_k - \lambda_k\right| \leq \|\Sigma_n - \Sigma\|_\infty \leq \|\Sigma_n - \Sigma\|_2 .$

- Is it possible to get $\widehat{\pi}_k = f_k(\Sigma_n)$ and $\pi_k = f_k(\Sigma)$?
- So that $\widehat{\pi}_k - \pi_k = f_k(\Sigma_n) - f_k(\Sigma) \simeq g_k(\Sigma_n - \Sigma)$ with 'nice' g_k

From covariance to projectors : perturbation theory and functional calculus

Notice $\left|\widehat{\lambda}_k - \lambda_k\right| \leq \|\Sigma_n - \Sigma\|_\infty \leq \|\Sigma_n - \Sigma\|_2$.

- Is it possible to get $\widehat{\pi}_k = f_k(\Sigma_n)$ and $\pi_k = f_k(\Sigma)$?
- So that $\widehat{\pi}_k - \pi_k = f_k(\Sigma_n) - f_k(\Sigma) \simeq g_k(\Sigma_n - \Sigma)$ with 'nice' g_k

Functional calculus defines $f(M)$ where M and $f(M)$ matrices or operators for certain classes of functions f. Two kinds:

- **Spectral FC** $\diamond$ M selfadjoint $\diamond$ f Borel bounded on $\mathbb{R}$ $\diamond$ **Main tool :** Resolution of the identity (Spectral Theorem)
- **Perturbation FC** $\diamond$ any M $\diamond$ f holomorphic $\diamond$ **Main tool** : Resolvent (Kato's theory)

Projection via perturbation in a nutshell

- Take $\lambda_k \in \mathbb{C}$ and let Ω_k a closed oriented contour around λ_k (say a circle) separating λ_k from the other eigenvalues
- Use Cauchy integral properties to compute :

$$\frac{1}{2\pi\iota}\oint_{\Omega_k}\frac{dz}{z-\lambda_l}=\begin{cases}1 & k=l\\ 0 & k\neq l\end{cases}$$

Projection via perturbation in a nutshell

- Take $\lambda_k \in \mathbb{C}$ and let Ω_k a closed oriented contour around λ_k (say a circle) separating λ_k from the other eigenvalues
- Use Cauchy integral properties to compute :

$$\frac{1}{2\pi\iota}\oint_{\Omega_k}\frac{dz}{z-\lambda_l} = \begin{cases} 1 & k = l \\ 0 & k \neq l \end{cases}$$

- *Switch from scalar to operator form* and apply eigenvectors :

$$\frac{1}{2\pi\iota}\oint_{\Omega_k}(zI-\Sigma)^{-1}(\varphi_l)\,dz = \frac{1}{2\pi\iota}\oint_{\Omega_k}\frac{dz}{z-\lambda_l}\varphi_l = \begin{cases} \varphi_k & k = l \\ 0 & k \neq l \end{cases}$$

Then
$\pi_k = \frac{1}{2\pi\iota}\int_{\Omega_k}(zI-\Sigma)^{-1}\,dz$ and $\widehat{\pi}_k = \frac{1}{2\pi\iota}\int_{\widehat{\Omega}_k}(zI-\Sigma_n)^{-1}\,dz,$

Projection via perturbation in a nutshell

- Take $\lambda_k \in \mathbb{C}$ and let Ω_k a closed oriented contour around λ_k (say a circle) separating λ_k from the other eigenvalues
- Use Cauchy integral properties to compute :

$$\frac{1}{2\pi\iota}\oint_{\Omega_k} \frac{dz}{z-\lambda_l} = \begin{cases} 1 & k=l \\ 0 & k \neq l \end{cases}$$

- *Switch from scalar to operator form* and apply eigenvectors :

$$\frac{1}{2\pi\iota}\oint_{\Omega_k} (zI-\Sigma)^{-1}(\varphi_l)\,dz = \frac{1}{2\pi\iota}\oint_{\Omega_k} \frac{dz}{z-\lambda_l}\varphi_l = \begin{cases} \varphi_k & k=l \\ 0 & k \neq l \end{cases}$$

Then
$\pi_k = \frac{1}{2\pi\iota}\int_{\Omega_k}(zI-\Sigma)^{-1}\,dz$ and $\widehat{\pi}_k = \frac{1}{2\pi\iota}\int_{\widehat{\Omega}_k}(zI-\Sigma_n)^{-1}\,dz$,

!!! $\widehat{\Omega}_k$ is a random contour of $\mathbb{C}$. Control of the location of eigenvalues (concentration inequalities) then $\widehat{\pi}_k = \frac{1}{2\pi\iota}\int_{\Omega_k}(zI-\Sigma_n)^{-1}\,dz + r_n$

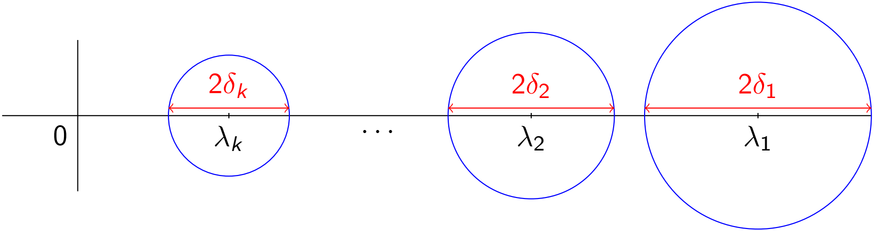


Figure: Contour made of disjoint circles (court. A. Roche)

Typical contour to define one or more π_k's

Two examples of classical contours

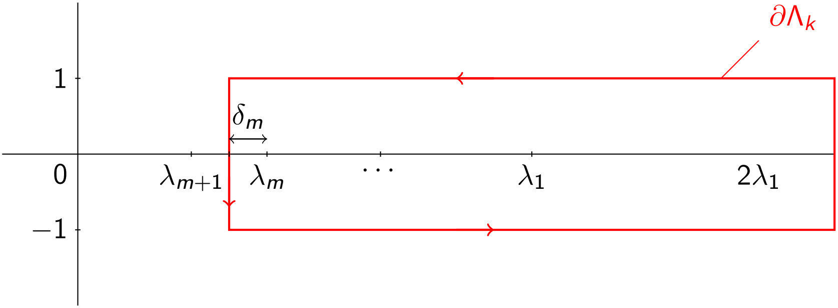


Figure: Rectangular contour (court. A. Roche)

Typical contour to define P_m, projector associated with the m first eigenvalues.

Main results 1/2:

Let $\mathbf{a}_k = \sum_{i \neq k} \frac{\lambda_i}{|\lambda_i - \lambda_k|} + \frac{\lambda_k}{\delta_k}$.
Theorem (lower bound) Take X Gaussian and $\mathbf{a}_k/\sqrt{n} < 0.5$ then :

$$\mathbb{E}\left\|\widehat{\pi}_k - \pi_k\right\|_\infty^2 \geq \underbrace{\frac{1}{2n}\sum_{j \neq k} \lambda_k \lambda_j / \left(\lambda_j - \lambda_k\right)^2}_{\mathrm{M}_k} - 4\frac{\mathbf{a}_k^4}{n^2}$$

Main results 1/2:

Let $\mathbf{a}_k = \sum_{i \neq k} \frac{\lambda_i}{|\lambda_i - \lambda_k|} + \frac{\lambda_k}{\delta_k}$.
Theorem (lower bound) Take X Gaussian and $\mathbf{a}_k/\sqrt{n} < 0.5$ then :

$$\mathbb{E} \left\| \widehat{\pi}_k - \pi_k \right\|_\infty^2 \geq \underbrace{\frac{1}{2n} \sum_{j \neq k} \lambda_k \lambda_j / \left(\lambda_j - \lambda_k\right)^2}_{\mathrm{M}_k} - 4 \frac{\mathbf{a}_k^4}{n^2}$$

- If $\lambda_k \propto k^{-\alpha}$ then $\mathrm{M}_k \succ k^2/n$ and $\mathbf{a}_k \prec k \log k$
- If $\lambda_k \propto \exp(-\alpha k)$ then $\mathrm{M}_k \succ 1/n$ and $\mathbf{a}_k \prec k$ ... but k^3/n should be small.

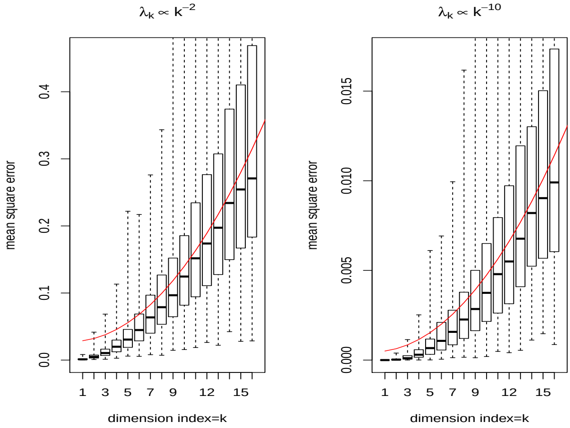
Main results 1/2:

Let $\mathbf{a}_k = \sum_{i \neq k} \frac{\lambda_i}{|\lambda_i - \lambda_k|} + \frac{\lambda_k}{\delta_k}$.
Theorem (lower bound) Take X Gaussian and $\mathbf{a}_k/\sqrt{n} < 0.5$ then :

$$\mathbb{E}\left\|\widehat{\pi}_k - \pi_k\right\|_\infty^2 \geq \underbrace{\frac{1}{2n}\sum_{j \neq k} \lambda_k \lambda_j / \left(\lambda_j - \lambda_k\right)^2}_{\mathrm{M}_k} - 4\frac{\mathbf{a}_k^4}{n^2}$$

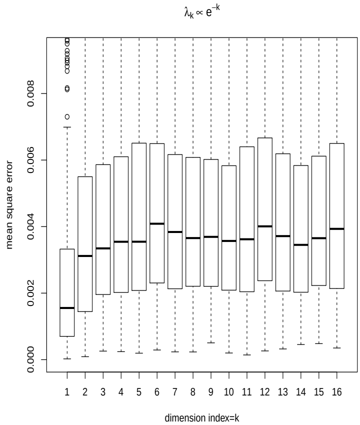
- If $\lambda_k \propto k^{-\alpha}$ then $\mathrm{M}_k \succ k^2/n$ and $\mathbf{a}_k \prec k \log k$
- If $\lambda_k \propto \exp\left(-\alpha k\right)$ then $\mathrm{M}_k \succ 1/n$ and $\mathbf{a}_k \prec k$ ... but k^3/n should be small.
- N.B.1 : the constants depend on the eigenvalues, the rate *not really*.
- N.B.2 : If $\exists \gamma > 0, j\lambda_j \log^{1+\gamma}(j) \downarrow 0$, then $\mathbf{a}_k \leq c\left(\gamma\right) k \log k$.

Back to simulations



MC simulation of $\mathbb{E}\,\|\widehat{\pi}_k - \pi_k\|_\infty^2$ with $M = 500$ replications, sample size $n = 500$ and $X = \sum_{j=1}^{kmax} \sqrt{\lambda_j}\xi_j e_j$ with $kmax = 20$, ξ_j Gaussian, e_j Fourier basis and $\mathrm{tr}\Sigma = 1$. Red curve : k^2 fit

Back to simulations



MC simulation of $\mathbb{E}\,\|\widehat{\pi}_k - \pi_k\|_\infty^2$ with $M = 500$ replications, sample size $n = 500$ and $X = \sum_{j=1}^{kmax} \sqrt{\lambda_j}\xi_j e_j$ with $kmax = 20$, ξ_j Gaussian, e_j Fourier basis and $\mathrm{tr}\Sigma = 1$.

Main results 2/2

Theorem (Upper bound for reconstruction) : Take $\mathbf{u} = X_i$ or $\mathbf{u} = X_{n+1}$ or $\mathbf{u}$ nonrandom with $\sup_i |\langle \mathbf{u}, \varphi_i \rangle|^2 / \lambda_i < 1$ then for all $n \geq 2$ and $k \geq 2$:

$$\mathbb{E} \left\| \left(\widehat{\mathbf{P}}_k - \mathbf{P}_k \right) \mathbf{u} \right\|^2 \leq \underbrace{\kappa c_0 \exp \left(-c_1 \frac{n}{\mathbf{a}_k^2} \right)}_{\mathcal{E}_{k,n}} + c_2 \cdot \frac{\mathbf{a}_k^2}{n} \log^2 n \cdot \left(\sum_{j=1}^{k} \sqrt{j \delta_j} \right)^2$$

where $\kappa = (2b-1) \sum_i \lambda_i^2 + (\mathrm{tr}\Sigma)^2$. No assumption on λ_k.

Main results 2/2

Theorem (Upper bound for reconstruction) : Take $\mathbf{u} = X_i$ or $\mathbf{u} = X_{n+1}$ or $\mathbf{u}$ nonrandom with $\sup_i |\langle \mathbf{u}, \varphi_i \rangle|^2 / \lambda_i < 1$ then for all $n \geq 2$ and $k \geq 2$:

$$\mathbb{E} \left\| \left(\widehat{\mathbf{P}}_k - \mathbf{P}_k \right) \mathbf{u} \right\|^2 \leq \underbrace{\kappa c_0 \exp\left(-c_1 \frac{n}{\mathbf{a}_k^2} \right)}_{\mathcal{E}_{k,n}} + c_2 \cdot \frac{\mathbf{a}_k^2}{n} \log^2 n \cdot \left(\sum_{j=1}^{k} \sqrt{j \delta_j} \right)^2$$

where $\kappa = (2b-1) \sum_i \lambda_i^2 + (\mathrm{tr}\Sigma)^2$. No assumption on λ_k.

- Polynomial decay : Upper bound $\prec k^2 [\log k \log n]^2 / n$
- Exp. decay : Upper bound $\prec k^2 \log^2 n / n$ but $k^2 \leq \exp(\alpha k) \approx 1/\delta_k$

Main results 2/2

Theorem (Upper bound for reconstruction) : Take $\mathbf{u} = X_i$ or $\mathbf{u} = X_{n+1}$ or $\mathbf{u}$ nonrandom with $\sup_i |\langle \mathbf{u}, \varphi_i \rangle|^2 / \lambda_i < 1$ then for all $n \geq 2$ and $k \geq 2$:

$$\mathbb{E} \left\| \left(\widehat{\mathbf{P}}_k - \mathbf{P}_k \right) \mathbf{u} \right\|^2 \leq \underbrace{\kappa c_0 \exp\left(-c_1 \frac{n}{\mathbf{a}_k^2} \right)}_{\mathcal{E}_{k,n}} + c_2 \cdot \frac{\mathbf{a}_k^2}{n} \log^2 n \cdot \left(\sum_{j=1}^{k} \sqrt{j \delta_j} \right)^2$$

where $\kappa = (2b-1) \sum_i \lambda_i^2 + (\mathrm{tr}\Sigma)^2$. No assumption on λ_k.

- Polynomial decay : Upper bound $\prec k^2 \left[\log k \log n \right]^2 / n$
- Exp. decay : Upper bound $\prec k^2 \log^2 n / n$ but $k^2 \leq \exp(\alpha k) \approx 1/\delta_k$
- Same kind of bounds for $\mathbb{E} \left\| \widehat{\pi}_k - \pi_k \right\|_\infty^2$ and for $\mathbb{E} \left\| \hat{\varphi}_k - \varphi_k \right\|^2$
- Previous rates may be improved with projection MS error:

$$\underbrace{\|P_k - P_k\|_\infty}_{uniform} \gg \underbrace{\|(P_k - P_k)\, u\|}_{strong} \gg \underbrace{\langle (P_k - P_k)\, u, v \rangle}_{weak}$$

Application to high-dimensional kernel estimation

Take X in high dim. space, $Y \in \mathbb{R}$ and consider modified non parametric regression : $r_k(x) = \mathbb{E}(Y|\mathbf{P}_k X = x)$ with $x \in \mathrm{Im}(\mathbf{P}_k)$

$$\widehat{r}(x) = \frac{\sum_{i=1}^n Y_i K\left(\left\|\widehat{\mathbf{P}}_k(X_i - x)\right\|/h\right)}{\sum_{i=1}^n K\left(\left\|\widehat{\mathbf{P}}_k(X_i - x)\right\|/h\right)}.$$

$$r^*(x) = \frac{\sum_{i=1}^n Y_i K\left(\left\|\mathbf{P}_k(X_i - x)\right\|/h\right)}{\sum_{i=1}^n K\left(\left\|\mathbf{P}_k(X_i - x)\right\|/h\right)}.$$

Question : is it possible to get $\mathbb{E}\left[\widehat{r}(x) - r^*(x)\right]^2 \prec \tau(n,k)$? with $\tau(n,k)$= Minimax rate in NP regression in $\mathbb{R}^k$ (and with an almost free choice of h, please)

Application to high-dimensional kernel estimation

Take X in high dim. space, $Y \in \mathbb{R}$ and consider modified non parametric regression : $r_k(x) = \mathbb{E}(Y|\mathbf{P}_k X = x)$ with $x \in \mathrm{Im}(\mathbf{P}_k)$

$$\widehat{r}(x) = \frac{\sum_{i=1}^n Y_i K\left(\left\|\widehat{\mathbf{P}}_k(X_i - x)\right\| / h\right)}{\sum_{i=1}^n K\left(\left\|\widehat{\mathbf{P}}_k(X_i - x)\right\| / h\right)}.$$

$$r^*(x) = \frac{\sum_{i=1}^n Y_i K(\|\mathbf{P}_k(X_i - x)\| / h)}{\sum_{i=1}^n K(\|\mathbf{P}_k(X_i - x)\| / h)}.$$

Question : is it possible to get $\mathbb{E}[\widehat{r}(x) - r^*(x)]^2 \prec \tau(n,k)$? with $\tau(n,k)$= Minimax rate in NP regression in $\mathbb{R}^k$ (and with an almost free choice of h, please)

Prop : Non asymptotic bound for all h and k subject to $k > 2$, $0 \leq h \leq h_{\max} < 1$

$$\mathbb{E}[\widehat{r}(x) - r^*(x)]^2 \leq c_6 \frac{k^4}{nh^2} \log^2 n \log\left(\frac{h\sqrt{n}}{k^2}\right).$$

Application to linear inference for functional data

Model generalizes linear regression to functional input/ouput

- Functional with scalar output : $y = \langle \beta, X \rangle + \varepsilon = \int \beta(s) X(s) \, ds + \varepsilon$

Application to linear inference for functional data

Model generalizes linear regression to functional input/ouput

- Functional with scalar output : $y = \langle \beta, X \rangle + \varepsilon = \int \beta(s) X(s) \, ds + \varepsilon$
- $\Delta = E(yX) = \Sigma \beta \Longrightarrow \Delta_n = \Sigma_n \beta_n$ implicit equation.
- Estimation of β requires inversion/regularization of $\Sigma_n \mapsto$ linear inverse problem with unknown operator

Application to linear inference for functional data

Model generalizes linear regression to functional input/ouput

- Functional with scalar output : $y = \langle \beta, X \rangle + \varepsilon = \int \beta(s) X(s)\, ds + \varepsilon$
- $\Delta = E(yX) = \Sigma\beta \Longrightarrow \Delta_n = \Sigma_n \beta_n$ implicit equation.
- Estimation of β requires inversion/regularization of $\Sigma_n \mapsto$ linear inverse problem with unknown operator
- Many recipes : penalization/Tikhonov/thresholding...
- PCA=spectral cut : $\Sigma_n^\dagger = \sum_{j=1}^k \widehat{\lambda}_j^{-1} \left(\widehat{e}_j \otimes \widehat{e}_j\right)$

Application to linear inference for functional data

Model generalizes linear regression to functional input/ouput

- Functional with scalar output : $y = \langle \beta, X \rangle + \varepsilon = \int \beta(s) X(s)\, ds + \varepsilon$
- $\Delta = E(yX) = \Sigma\beta \Longrightarrow \Delta_n = \Sigma_n \beta_n$ implicit equation.
- Estimation of β requires inversion/regularization of $\Sigma_n \mapsto$ linear inverse problem with unknown operator
- Many recipes : penalization/Tikhonov/thresholding...
- PCA=spectral cut : $\Sigma_n^{\dagger} = \sum_{j=1}^{k} \widehat{\lambda}_j^{-1} (\widehat{e}_j \otimes \widehat{e}_j)$
- Main goal not β_n but predictor $\langle \beta_n, X_{n+1} \rangle$

Why PCA ?

- Consider bias=approx error for prediction after projection with $\mathcal{P}_k$ (which is any k-dimensional projector)

$$\mathbb{E}\left\langle \mathcal{P}_k\beta - \beta, X\right\rangle^2 = \mathbb{E}\left\langle \beta, \left(I - \mathcal{P}_k\right) X\right\rangle^2 \leq \left\|\beta\right\|^2 \mathbb{E}\left\|\left(I - \mathcal{P}_k\right) X\right\|^2$$

- PCA: good candidate for projection basis with no prior on β

Why PCA ?

- Consider bias=approx error for prediction after projection with $\mathcal{P}_k$ (which is any k-dimensional projector)

$$\mathbb{E}\left\langle \mathcal{P}_k\beta - \beta, X \right\rangle^2 = \mathbb{E}\left\langle \beta, (I - \mathcal{P}_k)\, X \right\rangle^2 \leq \|\beta\|^2\, \mathbb{E}\left\|(I - \mathcal{P}_k)\, X\right\|^2$$

- PCA: good candidate for projection basis with no prior on β
- Avoids decoupling assumptions : smoothness of X / smoothness of β
- Provide minimax adaptive estimators
- Allows minimax adaptive testing as well
- Mean Square Pred. Risk : $\log n/n$ even if eigenvalues AND β Fourier coeff decay at exponential rate $\approx$ highly ill-posed.

Example of estimation results : Set $L = \left\| \Sigma^{1/2} \beta \right\|$, $\varphi(j) = \lambda_j \langle \beta, \varphi_j \rangle^2 / L^2$ and k_n^* as the integer part of the unique solution of the integral equation (in x) :

$$\frac{1}{x} \int_x^{+\infty} \varphi(x)\, dx = \frac{1}{n} \frac{\sigma_\varepsilon^2}{L^2}. \tag{1}$$

then with $\mathcal{R}_n(L, \varphi) = \sup_{\Sigma^{1/2}\beta \in \mathcal{L}_2(L,\varphi)} \mathbb{E} \langle \beta_n - \beta, X_{n+1} \rangle^2$

$$\lim \sup_{n \to +\infty} \frac{n}{k_n^*} \mathcal{R}_n(L, \varphi) = 2\sigma_\varepsilon^2$$

If $\varphi_a(j) \propto \left(j^{2+\alpha} (\log j)^\beta \right)^{-1}$,$\mathcal{R}_n(L, \varphi_a) \propto \frac{(\log n)^{\beta/(2+\alpha)}}{n^{(1+\alpha)/(2+\alpha)}}$,
If $\varphi_b(j) \propto \exp(-\alpha j)$,$\mathcal{R}_n(L, \varphi_b) \leq \frac{\log n}{\alpha n}$.
Besides $\inf_{\widehat{T}} \sup_{\Sigma^{1/2}\beta \in \mathcal{L}_2(\varphi, L)} \mathbb{E} \left\| \widehat{T}(X_{n+1}) - S(X_{n+1}) \right\|^2 \asymp \frac{k_n^*}{n}$

Thank you for your attention.