A Kernel Random Matrix-Based Approach for Sparse PCA

Mohamed El Amine Seddik [1,2] Mohamed Tamaazousti [1] Romain Couillet [2,3]

[1]CEA List [2]CentraleSupélec [3]GIPSA-Lab University of GrenobleAlpes

Abstract

Context: Sparse PCs recovery given a set of data observations of **high dimension**.

Objective: Study of **kernel random matrices** of the form $f(\hat{C})$ where $\hat{C}$ is the **sample covariance matrix** of the data.

Result: f such that $f'(0) = f''(0) = 0$ allows for **powerful recovery** of sparse PCs.

Model & Assumptions

Data matrix:

$$X \equiv \Sigma_p^{1/2} Z = \left(I_p + \sum_{i=1}^{k} \omega_i u_i u_i^\intercal \right)^{1/2} Z \quad \text{(of size } p \times n)$$

- Z random matrix with $\mathcal{N}(0,1)$ *i.i.d.* entries.
- $U = [u_1, \ldots, u_k] \in \mathbb{R}^{p \times k}$ isometric.

Kernel matrix:

$$F = f\left(\frac{1}{n} X X^\intercal\right) \underset{\text{entry-wise}}{\equiv} \left\{ f\left(\left[\frac{1}{n} X X^\intercal \right]_{ij} \right) \right\}_{i,j=1}^{p}$$

(A1) As $n \to \infty$; $p/n \to c \in (0, \infty)$ and $\limsup_n \max_i \omega_i < \infty$.

(A2) f is three-times continuously **differentiable**.

Concentration of Measure

Definition 1. A random vector V is said to be **normally** $(q = 2)$ or **exponentially** $(q = 1)$ **concentrated** if for any 1-**Lipschitz real** function $\mathbb{P}\{|f(V) - \mathbb{E}f(V)| \geq t\} \leq C\, e^{-(t/\sigma)^q} \Rightarrow V \in \mathcal{N}(\sigma)$ or $V \in \mathcal{E}(\sigma)$.

(P1) If $V \in \mathbb{R}^p$ Gaussian then $V \in \mathcal{N}(\sigma)$ [1].

(P2) If $V \in \mathbb{R}$ with finite expectation and $V \in \mathcal{N}(\sigma)$ then $V^2 \in \mathcal{E}(\sigma) + \mathcal{N}(\sigma)$.

General Result: Asymptotic Equivalent

$$\tilde{F} \equiv f(\Sigma_p) + \sum_{k=1}^{2} \frac{f^{(k)}(\Sigma_p)}{k} \odot \left[\Sigma_p^{1/2} \left(\frac{1}{n} Z Z^\intercal - I_p \right) \Sigma_p^{1/2} \right]^{\odot k}$$

Proposition 1. Under (A1) and (A2), for $\eta > 0$

$$F = \tilde{F} + \mathcal{O}_\eta(n^{-\frac{1}{2}})$$

where $X = \mathcal{O}_\eta^m(n^{-\alpha})$ means $\mathbb{P}\left\{ \|X\|_{op} \geq C\, n^{-\alpha} \eta^{-\frac{1}{2m}} \right\} \leq \eta$.

Sketch of Proof:

- Taylor expansion of F around $f(\Sigma_p)$.
- Control of the operator norm of $f^{(3)}(\xi^n) \odot \left[\Sigma_p^{1/2} \left(\frac{1}{n} Z Z^\intercal - I_p \right) \Sigma_p^{1/2} \right]^{\odot 3}$.
- Key lemma: $g(Z) = v^\intercal (Z Z^\intercal / n - I_p) v \in \mathcal{E}\left(n^{-\frac{1}{2}}\right) + \mathcal{N}\left(n^{-\frac{1}{2}}\right)$

$$v^\intercal Z \overset{(P1)}{\longrightarrow} \|v^\intercal Z / \sqrt{n}\| \in \mathcal{N}\left(n^{-\frac{1}{2}}\right) \overset{(P2)}{\longrightarrow} \|v^\intercal Z / \sqrt{n}\|^2 \in \mathcal{E}\left(n^{-\frac{1}{2}}\right) + \mathcal{N}\left(n^{-\frac{1}{2}}\right).$$

Sparse Matrices

Notations:

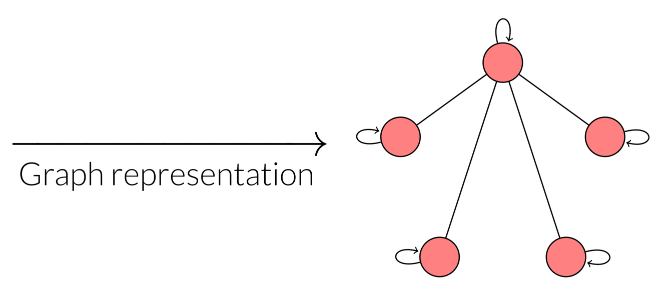
- $\mathcal{A}(M) = \{1_{M_{ij} \neq 0}\}_{i,j=1}^{p}$ defines a **graph** $\mathcal{G}_p$.
- $\mathcal{C}_p(k)$ the set of **closed walks** of length k on $\mathcal{G}_p$.

Definition 2 (ε-sparse matrices [2]). M is ε-sparse if $\mathcal{G}_p$ satisfies, for all $k \in 2\mathbb{N}$

$$|\mathcal{C}_p(k)| \leq C_k\, p^{\varepsilon(k-1)+1}$$

Example:

$$M = \begin{pmatrix} 1 & \frac{1}{\sqrt{p}} & \frac{1}{\sqrt{p}} & \frac{1}{\sqrt{p}} & \frac{1}{\sqrt{p}} \\ \frac{1}{\sqrt{p}} & 1 & 0 & 0 & 0 \\ \frac{1}{\sqrt{p}} & 0 & 1 & 0 & 0 \\ \frac{1}{\sqrt{p}} & 0 & 0 & 1 & 0 \\ \frac{1}{\sqrt{p}} & 0 & 0 & 0 & 1 \end{pmatrix}$$



(P3) If M is ε-**sparse** and $f(0) = 0$, then $f(M)$ is also ε-**sparse**.

(P4) If M is ε-**sparse** with **bounded entries** then $\forall k \in 2\mathbb{N}$ $\|M\|_{op} = \mathcal{O}(p^{\varepsilon(1-1/k)+1/k})$.

Special Case: Sparse PCA

(A3) Σ_p is $\frac{1}{2+\mu}$-sparse for $\mu > 0$.

Proposition 2. Under (A1), (A2) and (A3), for all $\epsilon \in (0, \frac{\mu}{2(3+2\mu)})$

$$F = f(\Sigma_p) + \mathcal{O}_\eta^{\lfloor 1/\epsilon \rfloor}\left(n^{\frac{-\mu}{2(2+\mu)} + \epsilon \left(2 - \frac{1}{2+\mu}\right)} \right) \quad \text{s.t. } f'(0) = f''(0) = 0$$

Sketch of Proof:
Control the operator norm of $f^{(k)}(\Sigma_p) \odot \left[\Sigma_p^{1/2} \left(\frac{1}{n} Z Z^\intercal - I_p \right) \Sigma_p^{1/2} \right]^{\odot k}$

- $f^{(k)}(\Sigma_p)$ is ε-sparse by (P3), its operator norm is controlled by (P4).
- $\left[\Sigma_p^{1/2} \left(\frac{1}{n} Z Z^\intercal - I_p \right) \Sigma_p^{1/2} \right]^{\odot k}$ has entries of order $\mathcal{O}(n^{-k/2})$.

Example of function with $f'(0) = f''(0) = 0$

Remark 1. For the spiked model

$$\Sigma_p = I_p + \sum_{i=1}^{k} \omega_i u_i u_i^\intercal \quad \Rightarrow \quad \Sigma_p \ \ 0\text{-sparse}$$

$\forall i$, u_i **sparse** $\Rightarrow F = f(\Sigma_p) + \mathcal{O}_\eta^{\lfloor 1/\epsilon \rfloor}(n^{-\frac{1}{2}+2\epsilon})$

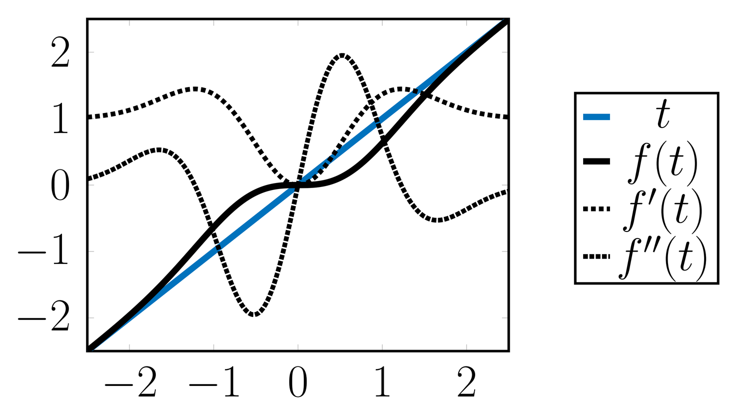
$\Rightarrow$ Choose f such that $f'(0) = f''(0) = 0$.

$\Rightarrow$ $f(t) \approx t$ for large values of t.

For $\alpha > 0$

$$f(t) \equiv t(1 - e^{-\alpha t^2})$$

$$f'(t) = 1 + e^{-\alpha t^2}(2\alpha t^2 - 1) \Rightarrow f'(0) = 0,$$
$$f''(t) = -2\alpha t e^{-\alpha t^2}\left(2\alpha t^2 - 3\right) \Rightarrow f''(0) = 0.$$



Experiments

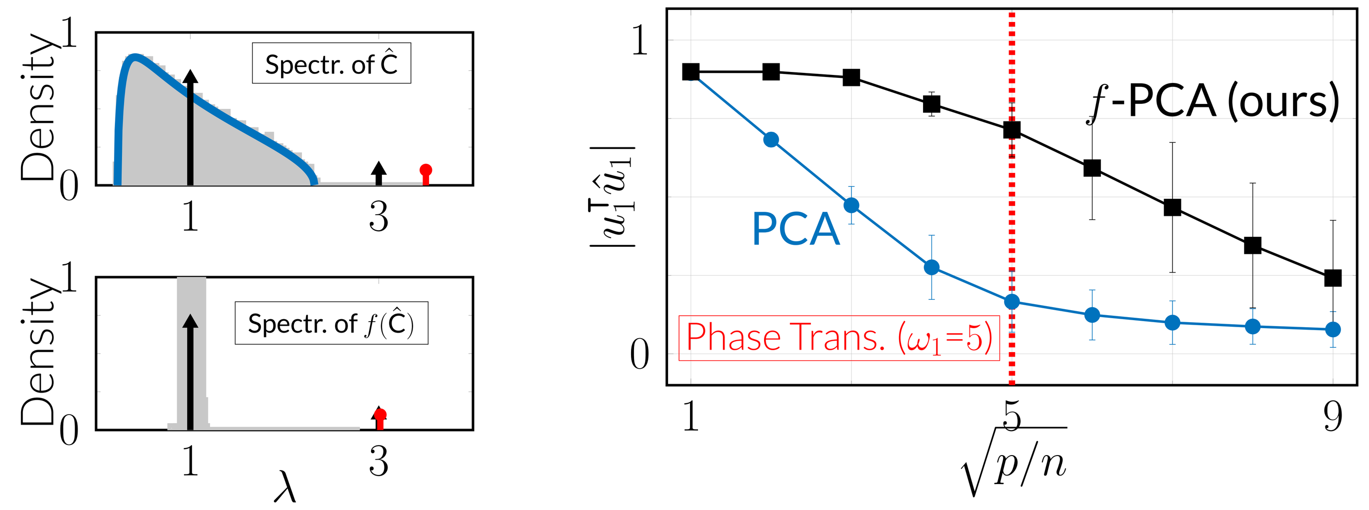


Figure 1. **(Left)** Spectrum of $\hat{C}$ *(up)* and $f(\hat{C})$ *(bottom)* for $p = 2048$ and $n = 7500$. Limiting Marenko-Pastur density versus **spectrum** of Σ_p. **(Right)** Alignment between estimated PC and GT.

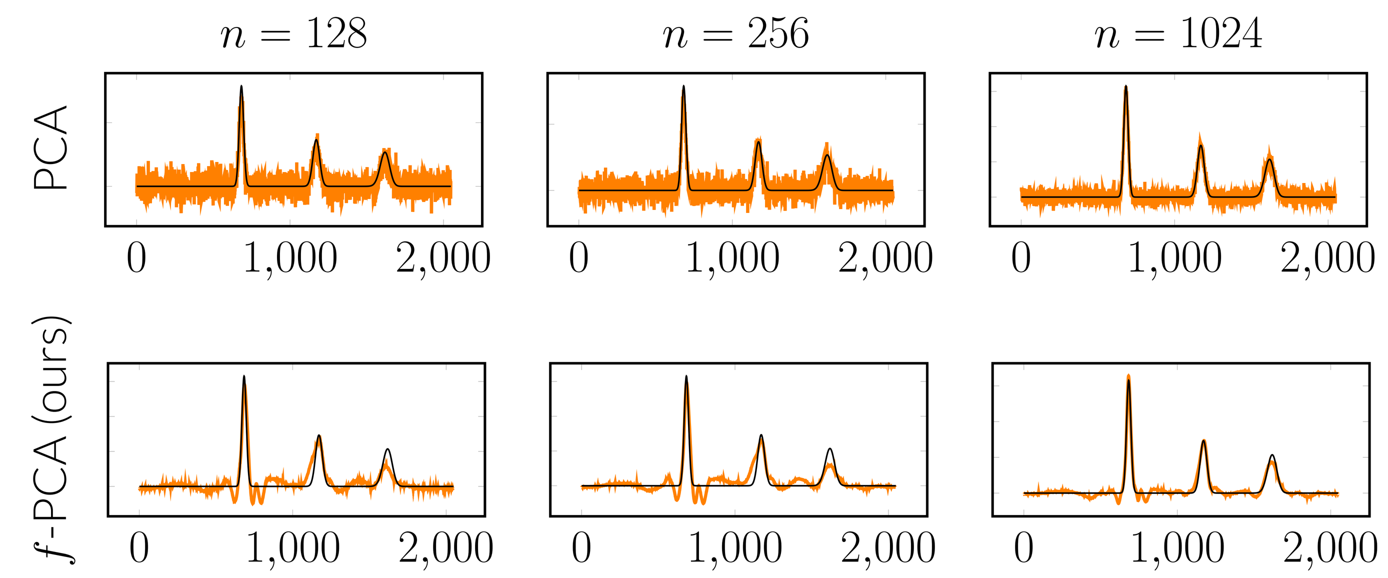


Figure 2. PC recovery by PCA **(up)** and our method **(down)**. $p = 2048$ and $\omega_1 = 5$.

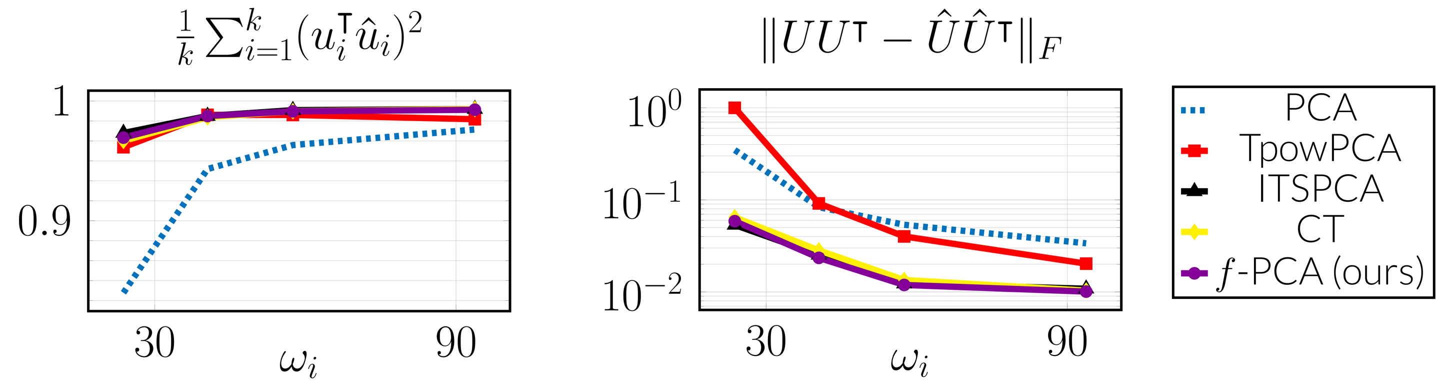


Figure 3. Performances of standard PCA, different state-of-the-art sparse PCA methods and our method. $p = 2048$, $n = 1024$ and $k = 4$.

Perspectives

- Consider the **non-trivial setting** of Cheng *et al.* [3].
- Estimate **optimal hyper-parameter choices** based on this setting.

References

[1] Terence Tao. Topics in random matrix theory, volume 132. American Mathematical Society Providence, RI, 2012.

[2] Noureddine El Karoui. Operator norm consistent estimation of large-dimensional sparse covariance matrices. The annals of Statistics, 2008.

[3] Xiuyuan Cheng and Amit Singer. The spectrum of random inner-product kernel matrices. Random Matrices: Theory and Applications, 2013.

ICLR'19, New Orleans, LA, US - **Contact:** mohamedelamine.seddik@cea.fr