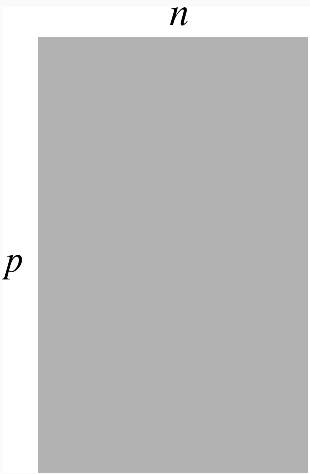
Beyond James-Stein estimation for PCA.

Alex Shkolnik (shkolnik@ucsb.edu)
(joint with Youhong Lee)

6th International Conference on Statistics & Econometrics,
Waseda University, Tokyo, Japan.
August 1, 2023.

Department of Statistics & Applied Probability
University of California, Santa Barbara

Let $\langle x, y\rangle = \sum_{\ell=1}^{m} x_\ell y_\ell$ for $x, y \in \mathbb{R}^m$ and $|x| = \sqrt{\langle x, x\rangle}$.



$Y = \theta X^\top + \mathcal{E}$

(only Y is observed)

$\theta \in \mathbb{R}^p$ (signal)

$X \in \mathbb{R}^n$ (latent factor)

$\mathcal{E}$ ($p \times n$ "noise" matrix)

Largest singular value decomposition involves $(\mathscr{s}, w, v)$ with

$$\mathrm{Y}w = (\theta \mathrm{X} + \mathcal{E})w = \mathscr{s}\, v$$

such that $\mathscr{s} = |\mathrm{Y}w|$ is maximized over all $w \in \mathbb{R}^{n}$ with $|w| = 1$.

Largest singular value decomposition involves $(\mathcal{s}, w, v)$ with

$$\mathrm{Y}w = (\theta \mathrm{X} + \mathcal{E})w = \mathcal{s}\, v$$

such that $\mathcal{s} = |\mathrm{Y}w|$ is maximized over all $w \in \mathbb{R}^n$ with $|w| = 1$.

- $v \in \mathbb{R}^p$ *is the left singular vector* ($|v| = 1$).
- $w \in \mathbb{R}^n$ *is the right singular vector* ($|w| = 1$).
- $\mathcal{s} \in \mathbb{R}$ *is the largest singular value.*

Largest singular value decomposition involves $(\mathscr{s}, w, u)$ with

$$\mathrm{Y}w = (\theta \mathrm{X} + \mathcal{E})w = \mathscr{s}\, v$$

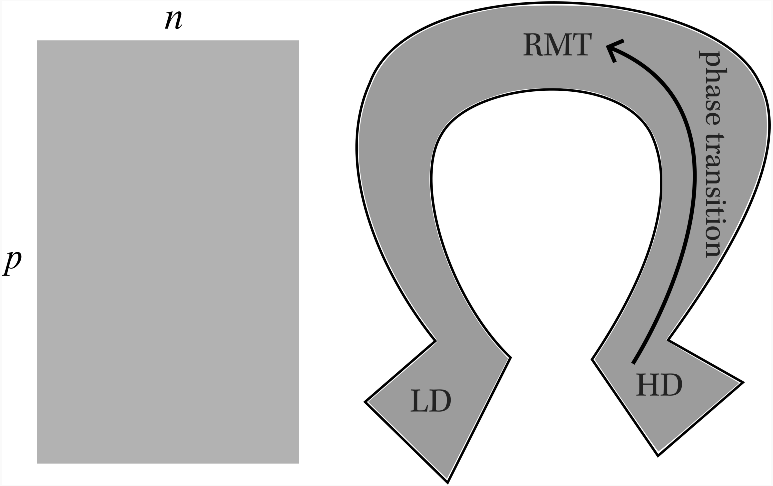
such that $\mathscr{s} = |\mathrm{Y}w|$ is maximized over all $w \in \mathbb{R}^n$ with $|w| = 1$.

Goal. Estimate θ and X (upto unidentifiability).

- *Note,* $v = \theta/|\theta|$ *and* $w = \mathrm{X}/|\mathrm{X}|$ *when* $\mathcal{E} = 0$.
- *And* $\mathscr{s} = |\theta|$ *provided* $|\mathrm{X}| = 1$ *(w.l.o.g.).*

"Behaviour" of $\mathcal{E}$ determines the accuracy otherwise.

Asymptotics' Horseshoe



Let $Y = \theta X^\top + \mathcal{E}$ regarding the length $|\theta|$ as the signal strength.

Limit behaviour of the following quantity is insightful.

$$\tau_p = |\theta|\sqrt{n/p}$$

e.g., $\tau_p \to \infty$ implies consistency of many important estimates.

Let $Y = \theta X^\top + \mathcal{E}$ regarding the length $|\theta|$ as the signal strength.

Limit behaviour of the following quantity is insightful.

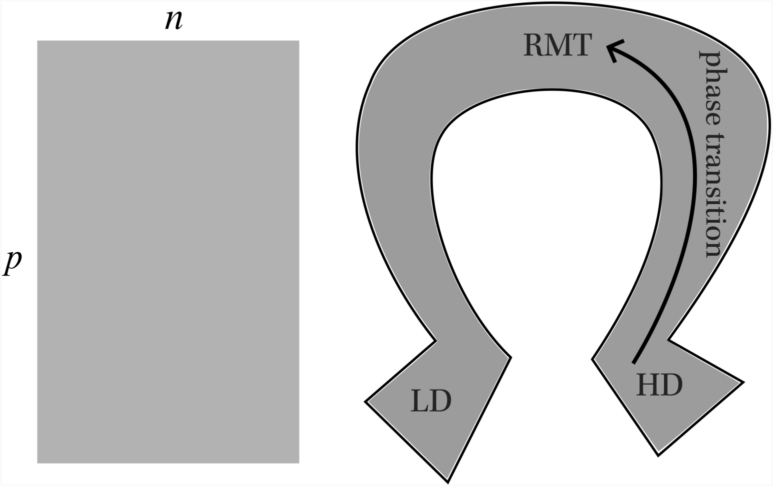
$$\tau_p = |\theta| \sqrt{n/p}$$

e.g., $\tau_p \to \infty$ implies consistency of many important estimates.

CONDITION 1. τ_p converges in $(0, \infty)$ as $p \to \infty$.

Borrowed from Wang & Fan (2017).

Note that $n \geq 2$ may remain finite or tend to ∞ as well.



Our results will fail at RMT ($\sup_{p} |\theta| < \infty$).

Inconsistency of the left singular vector

Let $Y = \theta X + \mathcal{E}$ and for the noise matrix $\mathcal{E}$, define

$$\Gamma = \mathcal{E}^\top \mathcal{E} / p \, .$$

Assumption A. Condition 1 holds with $\lim_{p \uparrow \infty} |\theta| = \infty$. Additionally,

(1) There exists ν which converges in $(0, \infty)$ and satisfies

$$|\Gamma - \nu^2 I| \leq C_p \sqrt{n/p}$$

for random sequence C_p with $\limsup_p C_p < \infty$ almost surely,

(2) We have $\limsup_p |\mathcal{E}^\top \theta|^2 / p < \infty$ almost surely,

(3) We have $\lim_p \langle \mathcal{E}^\top \theta, X \rangle / p = 0$ almost surely,

(4) $\langle X, X \rangle / n \neq 0$ almost surely for every $n \leq \infty$.

Let $\mathcal{s}_j$ be the jth singular value of $Y (= \theta X^\top + \mathcal{E})$.

Let ψ be a function of the singular values specified as,

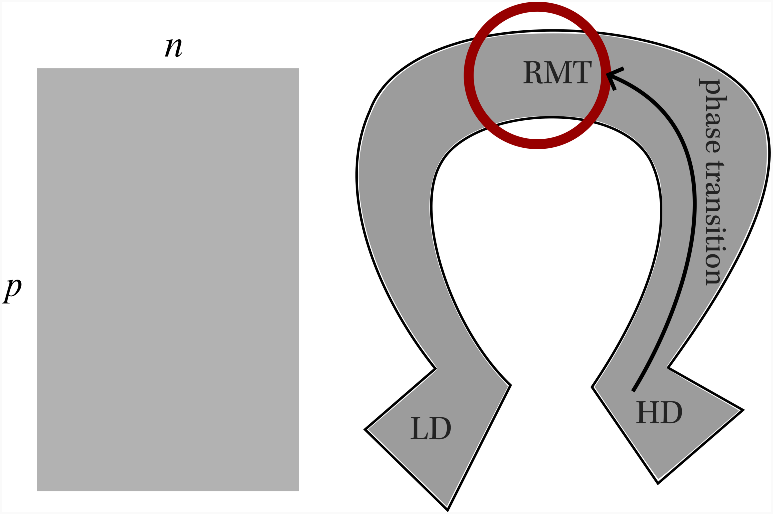
$$\psi^2 = 1 - \Big(\frac{\sum_{j<p} \mathcal{s}_j^2}{\mathcal{s}_p^2}\Big)\Big(\frac{1 + n/p}{n - n/p - 1}\Big).$$

Left singular vector v estimates $u = \theta/|\theta|$ inconsistently.

Theorem. *Suppose Assumption A holds. Then, almost surely,*

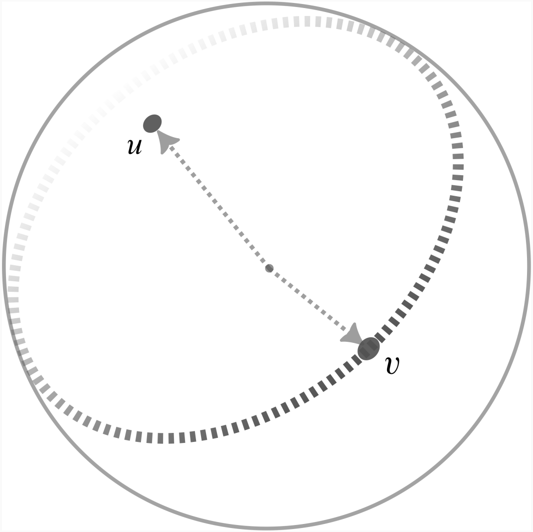
$$\langle v, u \rangle^2 \sim \psi^2 \qquad (p \uparrow \infty)$$

and eventually ψ is in the interval $(0, 1)$. *Moreover, the singular value $\mathcal{s}$ is also inconsistent, but the right singular vector is consistent.*



Theorem 1 holds if $\tau_p = |\theta|\sqrt{n/p}$ *converges, but fails at RMT (where* $\sup_p |\theta| < \infty$*, i.e., Johnstone's spike model (Johnstone & Lu 2009)).*

Inconsistency of the left singular vector.



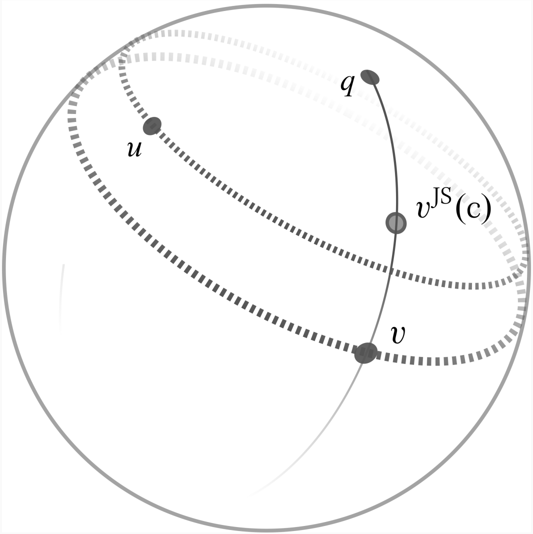
Some James-Stein results

The James-Stein estimator for the left singular vector is

$$v^{\mathrm{JS}} \propto m + \mathrm{c}\,(v - m) \qquad \left(\mathrm{c} = 1 - \frac{1 - \psi^2}{1 - \langle v, q\rangle^2}\right)$$

normalized to $|v^{\mathrm{JS}}| = 1$ for $m = \langle v, q\rangle q$ and any unit vector q.

The James-Stein left singular vector



Let $q \in \mathbb{R}^p$ be any (sequence) of unit length vectors such that assumption A still holds with q replacing $u = \theta/|\theta|$.

- Let $\mathrm{d}^2 = \frac{1-\psi^2}{\psi^2}$.

Theorem 2. *Suppose Assumption A holds. Then,*

$$\langle v^{\mathrm{JS}}, u \rangle^2 - \langle v, u \rangle^2 \sim \langle v, q \rangle^2 \mathrm{d}^2 (1 - \mathrm{c})$$

and c *converges in* $(0, 1)$ *almost surely as* $p \to \infty$.

Remarks.

- $\sharp$ v^{JS} improves upon the left singular vector if $\langle v, q \rangle \neq 0$ ev.
- The result fails as we transition into the RMT regime.

Stepping stone to a CLT theorem for James-Stein.

- Let $\kappa^2 = \lim_{p\to\infty} \langle v^{\mathrm{JS}}, u\rangle^2$.
- Let $\mathrm{d}^2 = \frac{1-\psi^2}{\psi^2}$.

Theorem 3. Suppose Assumption A holds with n fixed. Then,

$$\langle v^{\mathrm{JS}}, u\rangle^2 - \kappa^2 \sim (\mathrm{c}\,\mathrm{d})^2 \mathscr{R}_1 + 2(\mathrm{cd})(1-\mathrm{c})\,\mathscr{R}_2 + (1-\mathrm{c})^2\,\mathscr{R}_3$$

where every rate $\mathscr{R}_j \to 0$ as $p \to \infty$ almost surely.

Remark. Each rate $\mathscr{R}_j$ may be reasonably assumed to decay at the rate $\sqrt{p}$ (weakly) or, more finely as $\sqrt{p \log\log p}$ (LIL).

Recall, $Y = \theta X + \mathcal{E}$ with $\Gamma = \mathcal{E}^\top \mathcal{E}/p \sim \nu^2 I$ and $\tau_p = |\theta|\sqrt{n/p}$.

The first fundamental rate with coefficient $(\mathrm{cd})^2$ is given by,

$$\mathscr{R}_1 = \left(\frac{\tau_p^2 - \tau_\infty^2}{\tau_\infty^2}\right) + \left(\frac{\langle X, (\Gamma - \nu^2 I)X\rangle}{|X|^2 \nu^2}\right).$$

Recall $\mathrm{Y} = \theta\mathrm{X} + \mathcal{E}$ and $u = \theta/|\theta|$.

The second fundamental rate with coefficent $2(\mathrm{cd})(1-\mathrm{c})$ is given by,

$$\mathscr{R}_2 = \frac{\langle \mathrm{X}, \mathcal{E}^\top u_q \rangle}{|\mathrm{X}|\nu\sqrt{p}} \qquad \left(u_q = \frac{u - \langle u, q\rangle q}{\sqrt{1 - \langle u, q\rangle^2}}\right).$$

Recall $Y = \theta X + \mathcal{E}$ and $u = \theta/|\theta|$.

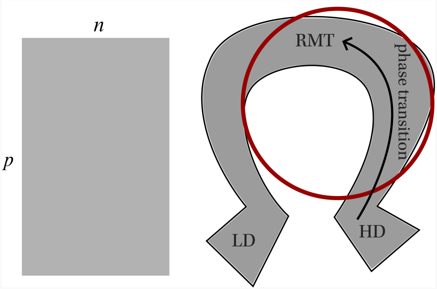
The third fundamental rate with coeffcient $(1 - c)^2$ is given by,

$$\mathscr{R}_3 = \langle u, q \rangle - \lim_{p \to \infty} \langle u, q \rangle .$$

Theorem 3. Suppose Assumption A holds with n fixed. Then,

$$\langle v^{\mathrm{JS}}, u\rangle^2 - \kappa^2 \sim (\mathrm{c\,d})^2 \mathscr{R}_1 + 2(\mathrm{cd})(1-\mathrm{c})\,\mathscr{R}_2 + (1-\mathrm{c})^2 \mathscr{R}_3$$

where every rate $\mathscr{R}_j \to 0$ as $p \to \infty$ almost surely.



Working out the covariances of the $\mathscr{R}_j$ allows for a CLT.

References.

Johnstone, I. M. & Lu, A. Y. (2009), 'On consistency and sparsity for principal components analysis in high dimensions', *Journal of the American Statistical Association* **104**(486), 682–693.

Wang, W. & Fan, J. (2017), 'Asymptotics of empirical eigenstructure for high dimensional spiked covariance', *The Annals of Statistics* **45**(3), 1342–1374.