James-Stein estimator of moderately-spiked leading eigenvector of high-dimensional covariance matrix

Sungkyu Jung

Seoul National University

August 2023
Table of Contents

1 Motivation from a financial application

2 JS estimation of \( u \) in the HL regime

3 Extensions to moderately spiked eigenvalues
Portfolio selection

Markowitz (1952)'s mean-variance optimal portfolio:

Minimum-variance portfolio

- $\mathbf{w} = (w_1, \ldots, w_p)^\top$: a (fully invested) portfolio, defined over $p$ securities (with $\mathbf{1}^\top \mathbf{w} = \sum_{i=1}^p w_i = 1$)
- $\Sigma$: the $p \times p$ matrix of volatility of the $p$ securities

- Minimum-variance portfolio: The solution $\mathbf{w}^*$ of the optimization problem

$$\min_{\mathbf{w} \in \mathbb{R}^p} \mathbf{w}^\top \Sigma \mathbf{w},$$

subject to

$$\mathbf{1}^\top \mathbf{w} = 1.$$
A motivating research question

Can we accurately estimate the minimum-variance portfolio?

\[ w^* = \arg \min_{w \in \mathbb{R}^p} w^\top \Sigma w \quad \text{subject to} \quad 1^\top w = 1 \]

\[ = \Sigma^{-1} 1 / (1^\top \Sigma^{-1} 1) \]

- Since \( \Sigma \) is unavailable, we need to work with the estimate \( \hat{\Sigma} \) (from security returns data).
- ... translates into the problem of estimating the eigenvalues and eigenvectors of \( \Sigma \).
A working model for volatility matrix
(Goldberg et al., 2020)

• A single-index “market” model for the excess return $y$ to $p$ securities: For random and unobservable $X \in \mathbb{R}$, $z \in \mathbb{R}^p$,

$$y = \beta X + z,$$

where the betas $\beta = (\beta_1, \ldots, \beta_p)^\top$ are the relative volatility of the securities, compared to the market.

• This model induces a simple one-spike model for $\Sigma$:

$$\Sigma = \sigma^2 \beta \beta^\top + \tau^2 \mathbb{I}.$$
Natural assumptions on the “betas”

(i) The betas satisfy $\beta_i = \text{Cov}(y_i, X) \in (-\infty, \infty)$ (if the market volatility $\sigma = 1$)

(ii) Typically

$\beta_i \approx 1$ for most $i$,

and

$\beta_1^2 + \cdots + \beta_p^2 \asymp p$.

Taking the direction $u = \beta / \|\beta\|$ from $\beta$,

- the working covariance model becomes $\Sigma = \sigma^2 p^\alpha uu^\top + \tau^2 \mathbb{I}$, where $\alpha = 1$, and

- $w^* \propto c_1 1 + c_2 u$, where $c_1, c_2$ depend on $\sigma^2 p^\alpha, \tau^2$. 
# Table of Contents

1. Motivation from a financial application

2. JS estimation of \( u \) in the HL regime

3. Extensions to moderately spiked eigenvalues
Principal Component Analysis

- Dimension Reduction
- Visualization of Data Structure
- Estimation of factors (eigenvectors, $\mathbf{u}$) and the variances (eigenvalues of $\Sigma$)
Principal Component Analysis

From the working covariance model $\Sigma = \sigma^2 \beta \beta^\top + \tau^2 \mathbb{I}$,

- $\mathbf{u} \propto \beta \in \mathbb{R}^p$: (First) principal component direction,
- $\lambda := \sigma^2 p^\alpha + \tau^2$: The (largest) variance of the first principal component.

Requirements for estimators $(\hat{\mathbf{u}}, \hat{\lambda})$ of PC directions and variances:

- Consistency. $\text{Angle}(\mathbf{u}, \hat{\mathbf{u}}) \to 0$ and $\hat{\lambda}/\lambda \to 1$ as $p \to \infty$.
- Or, minimum bias (even better if unbiased).

We will investigate asymptotic performance of $(\hat{\mathbf{u}}, \hat{\lambda})$ and its JS shrinkage

- first along the intuitive “HL” (High-dimension, Low-sample-size; HDLSS) regime (Hall et al., 2005; ?)
- then by allowing smaller signal strength (?)
PCA inconsistency, $p \gg n$

- For $\Sigma = \sigma^2 p^\alpha u u^T + \tau^2 I_d$, $\alpha \geq 0$,
- $X = [X_1, \ldots, X_n]$: $p \times n$ data matrix, where each column vector has zero mean and covariance matrix $\Sigma$.
- For the classical estimates of PC direction and variance pairs $(\hat{u}, \hat{\lambda})$ by eigen-decomposition of $S$, as $p \to \infty$ ($n$ fixed)
  
  $$\text{Angle}(\hat{u}, u) \to \begin{cases} 
  0, & \text{if } \alpha > 1, \quad \text{(consistent)} \\
  \text{random}, & \text{if } \alpha = 1, \quad \text{(inconsistent)} \\
  \pi/2, & \text{if } \alpha < 1, \quad \text{(strongly inconsistent)}
  \end{cases}$$

- The *inconsistent* case $\alpha = 1$ may be perceived as more natural (as in the *market beta* example). (Leek, 2011; Jung et al., 2012; Hellton and Thoresen, 2014, 2017; Shen et al., 2016)
PCA inconsistency, $p \gg n$

- For $\alpha = 1$, $\hat{u}$ is biased;

$$\text{Angle}(\hat{u}, u) \to \cos^{-1} \left( \left\{ 1 + \frac{\tau^2}{\sigma^2 \chi_n^2} \right\}^{-1/2} \right),$$

and the direction to which $\hat{u}$ is biased is completely random.
Eigenvector bias adjustment

- Recall that for betas typically
  \[ \beta_i \approx 1 \quad \text{for most } i \]
- This translates to that the eigenvector \( \mathbf{u} = (u_1, \ldots, u_p)\top \)
  \[ u_i \approx \frac{1}{\sqrt{p}}, \quad u_i \neq 0 \quad \text{for most } i \]
- Geometrically, the vector \( \mathbf{u} \) is close to \( \mathbf{1} = (\frac{1}{\sqrt{p}}, \ldots, \frac{1}{\sqrt{p}})\top \).
• The idea is to shrink $\hat{u}$ towards $\vec{1}$ (closer to $u$ than $\hat{u}$)

$$\hat{u}^{JS}(c) = \frac{P_1(\hat{u}) + c\{\hat{u} - P_1(\hat{u})\}}{\|P_1(\hat{u}) + c\{\hat{u} - P_1(\hat{u})\}\|_2},$$

for a carefully chosen $c$. 
James-Stein estimation of eigenvectors

Comparing the naive PC direction estimate

\[ \hat{u} = (\hat{u}_1, \ldots, \hat{u}_p)^\top, \]

with the shrinked (or, rather, rotated) estimate

\[ \hat{u}^{JS} = (\hat{u}^{JS}_1, \ldots, \hat{u}^{JS}_p)^\top, \]

\( \hat{u}^{JS}_i \)'s are closer to \( \frac{1}{\sqrt{p}} \) than \( \hat{u}_i \) for any \( i \).

Credits: Goldberg, Papanicolaou and Shkolnik (GPS) (Goldberg et al., 2022) and Goldberg and Kercheval (2022); Shkolnik (2021)
James and Stein (1961)’s shrinkage estimator for multivariate parameters:

Geometry of $\hat{\mathbf{u}}^{JS}$ in the HL regime

Let $k = \lim_{p \to \infty} \mathbf{u}^\top \mathbf{1}$ and $M_n = \tau^2 / (\sigma^2 \chi_n^2)$
Eigenvector bias adjustment

It can be shown that if $\text{Angle}(\hat{\mathbf{v}}, \mathbf{u}) \neq \frac{\pi}{2}$, then asymptotically as $p \rightarrow \infty$, with $c_n$ being a consistent estimator of the optimal shrinkage parameter $c_0$,

$$\text{Angle}(\hat{\mathbf{u}}^{\text{JS}}(c_n), \mathbf{u}) < \text{Angle}(\hat{\mathbf{u}}, \mathbf{u}),$$

and the inequality is strict.
# Table of Contents

1. Motivation from a financial application

2. JS estimation of $u$ in the HL regime

3. Extensions to moderately spiked eigenvalues
What happens if $n \to \infty$ as well?

From the results in the HL regime ($p \to \infty$, $n$ fixed), take $n \to \infty$ as well. Then, as $p \to \infty$, $n \to \infty$ (successively),

- $M_n = \frac{\tau^2}{\sigma^2 \chi^2_n} \to 0$,
- $\text{Angle}(\hat{u}, u) \to 0$
- $\text{Angle}(\hat{u}^{JS}(c), u) \to 0$ only if $c = 1$, at which $\hat{u}^{JS}(1) = \hat{u}$. 
Moderately-spiked eigenvalue regimes

- Calls for an investigation for smaller signal strength (while allowing for $n \rightarrow \infty$ as well)
- Let $n, p \rightarrow \infty$ (simultaneously, possibly at different rates)
- Models for signal strength (Aoshima and Yata; see Aoshima et al. (2018) and references therein)

$$
\lambda_1 = \sigma^2 p^{\alpha}, \quad \lambda_2 = \ldots, \lambda_p = \tau^2
$$

1. Case I: $\alpha = 1$ “Extremely spiked case”
2. Case II: $\frac{1}{2} \leq \alpha < 1$ “Strongly spiked case”
3. Case III: $\alpha < \frac{1}{2}$ “Moderately spiked case”
Case I: Extremely spiked case

If $\alpha = 1$ and $p, n \to \infty$ at any rate, then

1. $\text{Angle}(\hat{u}, u) \to 0$.

2. $\text{Angle}(\hat{u}^{\text{JS}}(c), u) \to 0$ only if $c = 1$, at which $\hat{u}^{\text{JS}}(1) = \hat{u}$. 
Case II: Strongly spiked case

If $\frac{1}{2} \leq \alpha < 1$ and $p, n \to \infty$ at any rate, then for

$$\delta_\infty := \lim_{n,p \to \infty} \frac{\tau^2 p^{1-\alpha}}{\sigma^2 n} \in [0, \infty],$$

we have

1. $\text{Angle}(\hat{u}, u) \to \cos^{-1} \left[ \sqrt{\frac{1}{1+\delta_\infty}} \right],$

2. $\text{Angle}(\hat{u}^{JS}(c), u) \to \cos^{-1} \left[ \frac{k^2 + c(1-k^2)}{\sqrt{k^2 + c(1+\delta_\infty - k^2)}} \right] =: \zeta(c), \text{ and}$

$\zeta(c)$ is minimized at $c_0 = \frac{1-k^2}{1-k^2+\delta_\infty}$.

Note that $k = \lim_{p \to \infty} u^\top \tilde{\mathbf{I}}$, and for $\hat{k} := \hat{u}^\top \tilde{\mathbf{I}},$

$$\hat{c}_{n,p} := 1 - \frac{1}{n-1} \sum_{i=2}^{n} \frac{\hat{\lambda}_i}{(1 - \hat{k}^2) \hat{\lambda}_1} \to c_0$$
Case III: Moderately spiked case

If \( \alpha < \frac{1}{2} \) and \( p, n \to \infty \) satisfy

\[
\frac{p^{1-2\alpha}}{n} \to 0,
\]

then

1. \( \angle(\hat{u}, u) \to \cos^{-1}\left[\frac{1}{\sqrt{1+\delta_\infty}}\right] \),

2. \( \angle(\hat{u}^{JS}(c), u) \to \cos^{-1}\left[\frac{k^2+c(1-k^2)}{\sqrt{k^2+c(1+\delta_\infty-k^2)}}\right] =: \zeta(c), \text{ and} \)

\( \zeta(c) \) is minimized at \( c_0 = \frac{1-k^2}{1-k^2+\delta_\infty} \).
In a nutshell...

If $\alpha \in [0, 1]$ and $p, n \to \infty$ satisfy

$$\frac{\max\{p^{1-2\alpha}, 1\}}{n} \to 0,$$

then,

$$\text{Angle}(\hat{u}^{JS}(c), u) \to \cos^{-1}\left[\frac{k^2 + c(1 - k^2)}{\sqrt{k^2 + c(1 + \delta_\infty - k^2)}}\right].$$

1. If $\delta_\infty = 0$ (Strong signal), $c_0 = 1$ and no advantage in using $\hat{u}^{JS}(c)$.

2. If $\delta_\infty = c \in (0, \infty)$ (Moderate signal), $c_0 \in (0, 1)$

$$\text{Angle}(\hat{u}^{JS}(c_{n,p}), u) < \text{Angle}(\hat{u}, u).$$

3. If $\delta_\infty = \infty$ (Weak signal), nothing works (theoretically)
Numerical demonstration

- \( k = \lim_{p \to \infty} \mathbf{u}^{\top} \mathbf{1} = 1/2 \)
- \( n = p^{2/3}. \) [We will only see the case \( n = 1,600 \) at \( p = 64,000 \)]
- Three cases
  1. Strong signal \( \alpha = \frac{2}{3} \) so that \( \delta_\infty = 0. \)
  2. Moderate signal \( \alpha = \frac{1}{3} \) so that \( \delta_\infty = 1. \)
  3. Weak signal \( \alpha = \frac{1}{4} \) so that \( \delta_\infty = \infty. \)
- For all cases, \( \frac{\max\{p^{1-2\alpha},1\}}{n} \to 0 \) as required.
Strong signal

Empirical \((u^\top \hat{u}^{JS}(c))^2\) vs Theoretical \(\lim_{n,p \to \infty} (u^\top \hat{u}^{JS}(c))^2\)
Moderate signal

Empirical \( (\mathbf{u}^\top \hat{\mathbf{u}}^{JS}(c))^2 \) vs Theoretical \( \lim_{n,p \to \infty} (\mathbf{u}^\top \hat{\mathbf{u}}^{JS}(c))^2 \)
Weak signal

Empirical $(\mathbf{u}^\top \hat{\mathbf{u}}^{JS}(c))^2$ vs Theoretical $\lim_{n,p \to \infty} (\mathbf{u}^\top \hat{\mathbf{u}}^{JS}(c))^2$


