Adjusting systematic bias in high dimensional principal component scores

Sungkyu Jung

Seoul National University

Supplementary Material

This supplementary material contains technical details and proofs, and a table summarizing simulation results.

S1 Proofs of Theorem 1 and Lemma 2

For reference, we restate the theorems and formulas in the main article that are used in the proof.

Theorem 1. *Assume the m-component model under Conditions (A1)–(A4) and let $n > m \geqslant 0$ be fixed and $d \to \infty$. Then, the first m sample and prediction scores are systematically biased:*

$$\widehat{W}_1 = SR^{\mathrm{T}}W_1 + O_p(d^{-1/4}), \tag{S1.1}$$

$$\widehat{W}_* = S^{-1}R^{\mathrm{T}}W_* + O_p(d^{-1/2}), \tag{S1.2}$$

where $R = [v_1(\mathcal{W}), \ldots, v_m(\mathcal{W})]$, $S = diag(\rho_1, \ldots, \rho_m)$, and $\rho_k = \sqrt{1 + \tau^2/\lambda_k(\mathcal{W})}$.

Moreover, for $k > m$,

$$\hat{w}_{kj} = O_p(d^{1/2}), \quad j = 1, \ldots, n, \tag{S1.3}$$

$$\hat{w}_{k*} = O_p(1). \tag{S1.4}$$

Lemma 1. *[Theorem S2.1, Jung et al. (2018)] Assume the conditions of Theorem 1. (i) the sample principal component variances converge in probability as $d \to \infty$;*

$$d^{-1} n \hat{\lambda}_i = \begin{cases} \lambda_i(\mathcal{W}) + \tau^2 + O_p(d^{-1/2}), & i = 1, \ldots, m; \\ \tau^2 + O_p(d^{-1/2}), & i = m+1, \ldots, n. \end{cases}$$

(ii) The inner product between sample and population PC directions converges in probability as $d \to \infty$;

$$\hat{u}_i^{\mathrm{T}} u_j = \begin{cases} \rho_i^{-1} v_{ij}(\mathcal{W}) + O_p(d^{-1/2}), & i, j = 1, \ldots, m; \\ O_p(d^{-1/2}), & \text{otherwise.} \end{cases}$$

Lemma 2. *Assume the m-component model with (A1)–(A4) and let $n > m \geqslant 0$ be fixed. For $k = 1, \ldots, n$, $E(\epsilon_{k*} | W_1) = 0$, and*

$$\lim_{d \to \infty} Var(\epsilon_{k*} \mid W_1) = \upsilon_O^2 / (\lambda_k(\mathcal{W}) + \tau^2), \quad \textit{for } k \leqslant m; \tag{S1.5}$$

$$\lim_{d \to \infty} \frac{1}{n-m} \sum_{k=m+1}^{n} Var(\epsilon_{k*} \mid W_1) = \upsilon_O^2 / \tau^2, \tag{S1.6}$$

where $\upsilon_O^2 = \lim_{d \to \infty} d^{-1} \sum_{i=m+1}^{d} \lambda_i^2$. As $d \to \infty$, $\epsilon_{k*} = O_p(1)$.

Proof of Lemma 2. Fix $k = 1, \ldots, n$. Let $Y_i = \sqrt{\lambda_i} z_{i*} p_{ki}$, where $p_{ki} = \hat{u}_k^{\mathrm{T}} u_i$. Then $\epsilon_{k*} = \sum_{i=m+1}^{d} Y_i$. Since z_{i*} and p_{ki} are independent, for each $i > m$,

$\mathrm{E}(Y_i \mid W_1) = 0$ and

$$\mathrm{Var}(\sum_{i=m+1}^{d} Y_i \mid W_1) = \mathrm{E}(\sum_{i=m+1}^{d} \lambda_i z_{i*}^2 p_{ki}^2 \mid W_1) = \sum_{i=m+1}^{d} \lambda_i \mathrm{E}(p_{ki}^2 \mid W_1),$$

where we use the fact that $\mathrm{E}(z_{i*}) = 0$, $\mathrm{E}(z_{i*}^2) = 1$.

For $k \leqslant m$, if the following claim,

$$\mathrm{E}(p_{ki}^2 \mid W_1) = d^{-1} \frac{\lambda_i}{(\lambda_k(\mathcal{W}) + \tau^2)} + O(d^{-3/2}), \tag{S1.7}$$

is true for any $i > m$, then it is easy to check (S1.5).

To show (S1.7), we first post-multiply $\hat{v}_i$ to

$$\mathcal{X} = \sqrt{n} \sum_{i=1}^{n} \sqrt{\hat{\lambda}_i} \hat{u}_i \hat{v}_i^{\mathrm{T}}, \tag{S1.8}$$

to obtain $\hat{u}_i = (n\hat{\lambda}_i)^{-1/2} \mathcal{X} \hat{v}_i$. By writing $z_i^{\mathrm{T}} = \lambda_i^{-1/2} w_i^{\mathrm{T}} = (z_{i1}, \ldots, z_{in})$, we have

$$\begin{aligned} p_{ki} &= u_i^{\mathrm{T}} \hat{u}_k \\ &= (n\hat{\lambda}_k)^{-1/2} u_i^{\mathrm{T}} \mathcal{X} \hat{v}_k \\ &= (n\hat{\lambda}_k)^{-1/2} \lambda_i^{1/2} z_i^{\mathrm{T}} \hat{v}_k. \end{aligned}$$

Thus,

$$\begin{aligned} p_{ki}^2 &= d^{-1} \frac{\lambda_i}{nd^{-1}\hat{\lambda}_k} (z_i^{\mathrm{T}} \hat{v}_k)^2 \\ &= d^{-1} \frac{\lambda_i}{\lambda_k(\mathcal{W}) + \tau^2 + O_p(d^{-1/2})} (z_i^{\mathrm{T}} \hat{v}_k)^2 \\ &= d^{-1} \frac{\lambda_i}{\lambda_k(\mathcal{W}) + \tau^2} (z_i^{\mathrm{T}} \hat{v}_k)^2 + O_p(d^{-3/2}). \end{aligned} \tag{S1.9}$$

In (S1.9), we used Lemma 1(i) and that $(1+x)^{-1} = 1 + O(x)$, and the fact that $|z_i^{\mathrm{T}} \hat{v}_k|^2 \leqslant \|z_i\|_2^2 \, \|\hat{v}_k\|_2^2 = \|z_i\|_2^2 = O_p(1)$.

Write $(z_i^{\mathrm{T}} \hat{v}_k)^2 = [z_i^{\mathrm{T}} v_k(W_1^{\mathrm{T}} W_1) + z_i^{\mathrm{T}}(\hat{v}_k - v_k(W_1^{\mathrm{T}} W_1))]^2$. Note that $W_1^{\mathrm{T}} W_1$ is an $n \times n$ matrix, and is different from the $m \times m$ matrix $\mathcal{W} = W_1 W_1^{\mathrm{T}}$. It can be shown that the right singular vector $\hat{v}_k$ converges in probability to $v_k(W_1^{\mathrm{T}} W_1)$ (see, e.g., Lemma S1.1 of Jung et al., 2018): For $k = 1, \ldots, m$,

$$\hat{v}_k = v_k(W_1^{\mathrm{T}} W_1) + O_p(d^{-1/2}). \tag{S1.10}$$

Thus we get $|z_i^{\mathrm{T}}(\hat{v}_k - v_k(W_1^{\mathrm{T}} W_1))| \leqslant \|z_i\|_2 \, \|\hat{v}_k - v_k(W_1^{\mathrm{T}} W_1))\|_2 = O_p(d^{-1/2})$. Therefore,

$$\begin{aligned} \mathrm{E}((z_i^{\mathrm{T}} \hat{v}_k)^2 \mid W_1) &= \mathrm{E}((z_i^{\mathrm{T}} v_k(W_1^{\mathrm{T}} W_1))^2 \mid W_1) + O(d^{-1/2}) \\ &= \sum_{\ell=1}^{n} \mathrm{E}(z_{i\ell}^2) v_{k\ell}^2(W_1^{\mathrm{T}} W_1) + O(d^{-1/2}) \\ &= 1 + O(d^{-1/2}). \end{aligned} \tag{S1.11}$$

Combing (S1.9) and (S1.11), we get (S1.7) for $k \leqslant m$ as desired.

To show (S1.6), note that $\mathcal{W} = W_1 W_1^{\mathrm{T}}$ is of rank m. For $k > m$, with $\lambda_k(\mathcal{W}) = 0$, (S1.9) holds. Thus,

$$\begin{aligned} \frac{1}{n-m} \sum_{k=m+1}^{n} \mathrm{Var}(\epsilon_{k*} \mid W_1) &= \frac{1}{n-m} \sum_{k=m+1}^{n} \sum_{i=m+1}^{d} \lambda_i \mathrm{E}(p_{ki}^2 \mid W_1) \qquad \text{(S1.12)} \\ &= \frac{1}{d(n-m)} \sum_{i=m+1}^{d} \lambda_i^2/\tau^2 \sum_{k=m+1}^{n} \mathrm{E}((z_i^{\mathrm{T}} \hat{v}_k)^2 \mid W_1). \end{aligned}$$

To simplify the expression $\mathrm{E}((z_i^{\mathrm{T}}\hat{v}_k)^2 \mid W_1)$, one should not naively try (S1.11). This is because that (S1.11) does not apply for $k > m$ due to the non-unique kth eigenvector $v_k(W_1^{\mathrm{T}}W_1)$ of the rank-m matrix $W_1^{\mathrm{T}}W_1$. Instead, from

$$\sum_{k=m+1}^{n}(z_i^{\mathrm{T}}\hat{v}_k)^2 = z_i^{\mathrm{T}}z_i - \sum_{k=1}^{m}(z_i^{\mathrm{T}}\hat{v}_k)^2,$$

and (S1.11) for $k \leqslant m$, we get

$$\sum_{k=m+1}^{n}\mathrm{E}((z_i^{\mathrm{T}}\hat{v}_k)^2 \mid W_1) = n - m + O(d^{-1/2}). \tag{S1.13}$$

Taking the limit $d \to \infty$ to (S1.12), combined with (S1.13), leads to (S1.6).

The last statement, $\epsilon_{k*} = O_p(1)$, easily follows from the fact $\lim_{d\to\infty}\mathrm{Var}(\epsilon_{k*}) \leqslant v_O^2/\tau^2(n-m) < \infty$, which is obtained by (S1.5) and (S1.6). □

We are now ready to show Theorem 1. Note that the results on the sample scores, (S1.1) and (S1.3), can be easily shown, using the decomposition $d^{-1/2}\hat{w}_k = \sqrt{d^{-1}n\hat{\lambda}_k}\hat{v}_k$, together with Lemma 1(i) and (S1.10). We show (S1.2) and (S1.4).

Proof of Theorem 1. Proof of (S1.2). Recall the decomposition

$$\hat{w}_{k*} = \hat{u}_k^{\mathrm{T}}X_* = \sum_{i=1}^{m} w_{i*}\hat{u}_k^{\mathrm{T}}u_i + \epsilon_{k*}, \tag{S1.14}$$

where $\epsilon_{k*} = \sum_{i=m+1}^{d} w_{i*}\hat{u}_k^{\mathrm{T}}u_i$. Using the notation $p_{ki} = \hat{u}_k^{\mathrm{T}}u_i$, we write $\hat{w}_{k*} =$

$(p_{k1},\dots,p_{km})(w_{1*},\dots,w_{m*})^{\mathrm{T}} + \epsilon_{k*}$. Putting all parts together, we have

$$\widehat{W}_* = d^{-1/2}(\hat{w}_{1*},\dots,\hat{w}_{m*})^{\mathrm{T}} = \begin{pmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & \ddots & \vdots \\ p_{m1} & \cdots & p_{mm} \end{pmatrix} W_* + \tilde{\boldsymbol{\epsilon}}_{k*},$$

where $\tilde{\boldsymbol{\epsilon}}_{k*} = d^{-1/2}(\epsilon_{1*},\dots,\epsilon_{m*})^{\mathrm{T}}$. By Lemma 2, as $d \to \infty$, $\tilde{\boldsymbol{\epsilon}}_{k*} = O_p(d^{-1/2})$. Since $p_{ki} = \rho_k^{-1} v_{ki}(\mathcal{W}) + O_p(d^{-1/2})$, by Lemma 1(ii), we have

$$\widehat{W}_*^{\mathrm{T}} = S^{-1} R^{\mathrm{T}} W_*^{\mathrm{T}} + O_p(d^{-1/2}).$$

Proof of (S1.4). Using the decomposition (S1.14), and by the fact $\epsilon_{k*} = O_p(1)$, from Lemma 2, it is enough to show $\sum_{i=1}^m w_{i*} p_{ki} = O_p(1)$. But, since Lemma 1 implies $d^{\frac{1}{2}} p_{ki} = O_p(1)$ for any pair of (k,i) such that $k > m, i \leqslant m$, we have $\sum_{i=1}^m w_{i*} p_{ki} = \sigma_i(d^{\frac{1}{2}} p_{k1},\dots,d^{\frac{1}{2}} p_{km})(z_{1*},\dots,z_{m*}) = O_p(1)$, $\square$

S2 Proof of Theorem 2

Theorem 2. *Let* $\zeta_{kj} = \lambda_k(\mathcal{W})/(\sum_{\ell=1}^m v_{\ell j}^2(\mathcal{W})\lambda_\ell(\mathcal{W}))$ *and* $\bar{\zeta}_{kj} = \sigma_k^2/(\sum_{\ell=1}^m v_{\ell j}^2(\mathcal{W})\sigma_\ell^2)$. *Under the assumptions of Theorem 1, as* $d \to \infty$, *for* $k, j = 1\dots,m$,

(i) $r(\hat{w}_k, w_j) \to v_{kj}(\mathcal{W})\zeta_{kj}^{1/2}$ *in probability ;*

(ii) $\lim_{d\to\infty} Corr(\hat{w}_{k*}, w_{j*} \mid W_1) = v_{kj}(\mathcal{W})\bar{\zeta}_{kj}^{1/2}$.

Proof of Theorem 2. Proof of (i). Write the singular value decomposition of the $m \times n$ matrix of scaled scores W_1 as

$$W_1 = R\mathrm{diag}(\sqrt{\lambda_1(\mathcal{W})}, \ldots, \sqrt{\lambda_1(\mathcal{W})})G^{\mathrm{T}}, \tag{S2.1}$$

where $G = [g_1, \ldots, g_m]$ is the $n \times m$ matrix consisting of right singular vectors of W_1. The left singular vector matrix $R = [v_1(\mathcal{W}), \ldots, v_m(\mathcal{W})]$ is exactly the matrix R appearing in Theorem 1. Since

$$W_1 = \sum_{\ell=1}^{m} \sqrt{\lambda_\ell(\mathcal{W})} v_\ell(\mathcal{W}) g_\ell^{\mathrm{T}},$$

the jth row of W_1 is, for $j \leqslant m$,

$$d^{-\frac{1}{2}} w_j^{\mathrm{T}} = \sum_{\ell=1}^{m} \sqrt{\lambda_\ell(\mathcal{W})} v_{\ell j}(\mathcal{W}) g_\ell^{\mathrm{T}}.$$

For the scaled sample score $d^{-1/2}\hat{w}_k$, $k \leqslant m$, we obtain from Theorem 1 and (S2.1) that $\widehat{W}_1 = S\mathrm{diag}(\sqrt{\lambda_1(\mathcal{W})}, \ldots, \sqrt{\lambda_1(\mathcal{W})})G^{\mathrm{T}} + O_p(d^{-1/4})$ and its kth row $d^{-1/2}\hat{w}_k = \sqrt{\lambda_k(\mathcal{W}) + \tau^2} g_k + O_p(d^{-1/4})$. Since g_ℓ's are orthonormal,

$$\|d^{-\frac{1}{2}} \hat{w}_k\|_2 = \sqrt{\lambda_k(\mathcal{W}) + \tau^2} + O_p(d^{-1/4}),$$

and

$$\begin{aligned} d^{-1}\hat{w}_k^{\mathrm{T}} w_j &= (d^{-1/2}\hat{w}_k)^{\mathrm{T}}(d^{-1/2}w_j) \\ &= \sqrt{\lambda_k(\mathcal{W})}\sqrt{\lambda_k(\mathcal{W}) + \tau^2} v_{kj}(\mathcal{W}) + O_p(d^{-1/4}). \end{aligned}$$

Since $d^{-1}w_j^{\mathrm{T}} w_j = \sum_{\ell=1}^{m} v_{\ell j}^2(\mathcal{W})\lambda_\ell(\mathcal{W})$, we have

$$r(\hat{w}_k, w_j) = \frac{d^{-1}\hat{w}_k^{\mathrm{T}} w_j}{\|d^{-1/2}\hat{w}_k\|_2 \cdot \|d^{-1/2}w_j\|_2} \to v_{kj}(\mathcal{W})\zeta_{kj}^{1/2}$$

in probability, as $d \to \infty$.

Proof of (ii). From Theorem 1, write

$$d^{-1/2}\hat{w}_{k*} = \rho_k^{-1}\sum_{\ell=1}^{m} v_{k\ell}(\mathcal{W}) d^{-1/2} w_{\ell*} + O_p(d^{-1/2}), \tag{S2.2}$$

and note that $\mathrm{E}(w_{k*}) = \mathrm{E}(\hat{w}_{k*}) = 0$. Then for $k = 1, \ldots, m$, we have

$$\mathrm{Var}(d^{-1/2} w_{k*}) = d^{-1}\mathrm{E}(w_{k*})^2 = \sigma_k^2 \mathrm{E}(z_{k*})^2 = \sigma_k^2,$$

and, by (S2.2),

$$\mathrm{Var}(d^{-1/2}\hat{w}_{k*} \mid W_1) = \rho_k^{-2}\sum_{\ell=1}^{m} \left(v_{k\ell}(\mathcal{W})\right)^2 \sigma_\ell^2 + O(d^{-1/2}).$$

The independence of $w_{\ell*}$ and w_{k*} for $k \neq \ell$ and (S2.2) give

$$\begin{aligned}\mathrm{Cov}(d^{-1/2}\hat{w}_{k*}, d^{-1/2} w_{j*} \mid W_1) &= \mathrm{E}(d^{-1}\hat{w}_{k*} w_{j*} \mid W_1)\\ &= \rho_k^{-1} v_{kj}(\mathcal{W})\sigma_j^2 + O(d^{-1/2}),\end{aligned}$$

which in turn leads to

$$\begin{aligned}\mathrm{corr}(\hat{w}_{k*}, w_{j*} \mid W_1) &= \frac{\mathrm{Cov}(d^{-1/2}\hat{w}_{k*}, d^{-1/2} w_{j*} \mid W_1)}{\left(\mathrm{Var}(d^{-1/2} w_{j*})\mathrm{Var}(d^{-1/2}\hat{w}_{k*} \mid W_1)\right)^{1/2}}\\ &= v_{kj}(\mathcal{W})\frac{\sigma_j}{\left[\sum_{\ell=1}^{m} \left(v_{k\ell}(\mathcal{W})\right)^2 \sigma_\ell^2\right]^{1/2}} + O(d^{-1/2}).\end{aligned}$$

$\square$

S3 Proof of Corollary 1

Corollary 1. *Suppose the assumptions of Lemma 1 are satisfied. Let $d \to \infty$. For $i = 1, \ldots, m$, conditional to W_1, $\tilde{\tau}^2$, $\tilde{\lambda}_i(\mathcal{W})$ and $\tilde{\rho}_i$ are consistent estimators of τ^2, $\lambda_i(\mathcal{W})$ and ρ_i, respectively.*

Proof of Corollary 1. Lemma 1 is used to show that $\tilde{\tau}^2$ and $\tilde{\lambda}_i(\mathcal{W})$ converge in probability to τ^2 and $\lambda_i(\mathcal{W})$ as $d \to \infty$, respectively. By continuous mapping theorem, $\tilde{\rho}_i$ converges in probability to ρ_i. □

S4 Complete Table 2

Bibliography

Jung, S., Ahn, J. & Lee, M. H. (2018). On the number of principal components in high dimensions. *Biometrika* **105**, 389–402.

Lee, S., Zou, F. & Wright, F. A. (2010). Convergence and prediction of principal component scores in high-dimensional settings. *Ann. Stat.* **38**, 3605.

	d	n	ρ_1 Theory	Best	Asymp.	Jackknife	LZW
	5000	50	1.41 (0.07)	1.42	1.40	1.43	1.41
Spike model	10000	50	1.42 (0.06)	1.43	1.42	1.44	1.42
$\beta = 0.3$	10000	100	1.23 (0.03)	1.23	1.23	1.24	1.23
	20000	100	1.23 (0.02)	1.23	1.23	1.24	1.23
	5000	50	1.42 (0.08)	1.45	1.41	1.45	1.40
Spike model	10000	50	1.43 (0.07)	1.45	1.43	1.46	1.42
$\beta = 0.5$	10000	100	1.22 (0.02)	1.23	1.22	1.23	1.21
	20000	100	1.23 (0.02)	1.23	1.23	1.24	1.22
	5000	50	2.06 (0.06)	2.22	1.92	2.14	2.00
Mixture model	10000	50	2.09 (0.06)	2.17	1.98	2.14	2.02
$a = 0.15$	10000	100	1.63 (0.02)	1.67	1.61	1.65	1.63
	20000	100	1.64 (0.02)	1.66	1.62	1.66	1.63

	d	n	ρ_2 Theory	Best	Asymp.	Jackknife	LZW
	5000	50	1.79 (0.11)	1.86	1.75	1.78	1.79
Spike model	10000	50	1.79 (0.11)	1.82	1.77	1.77	1.79
$\beta = 0.3$	10000	100	1.43 (0.06)	1.44	1.43	1.42	1.43
	20000	100	1.43 (0.05)	1.44	1.43	1.42	1.43
	5000	50	1.79 (0.11)	1.99	1.72	1.81	1.71
Spike model	10000	50	1.80 (0.11)	1.88	1.76	1.79	1.74
$\beta = 0.5$	10000	100	1.44 (0.05)	1.47	1.43	1.44	1.41
	20000	100	1.42 (0.05)	1.44	1.42	1.41	1.40
	5000	50	2.62 (0.21)	5.44	2.20	2.68	2.46
Mixture model	10000	50	2.68 (0.19)	3.20	2.35	2.68	2.50
$a = 0.15$	10000	100	2.00 (0.09)	2.13	1.90	2.00	1.99
	20000	100	1.99 (0.10)	2.05	1.93	1.97	1.97

Table 1: Simulation results from 100 repetitions. "Theory" is mean (standard deviation) of ρ_i; "Best" is $\check{\rho}_i$; "Asymp." is $\tilde{\rho}_i$; "Jackknife" is $\hat{\rho}_i^{(1)}$; "LZW" is from Lee et al. (2010). Averages are shown for the latter four columns. The standard errors of the quantities in estimation of ρ_i are at most 0.04.