SUPPLEMENT TO

"AUTOMATIC SPARSE PCA FOR HIGH-DIMENSIONAL DATA"

Kazuyoshi Yata[1] and Makoto Aoshima[1]

[1] *University of Tsukuba*

Supplementary Material

In this supplement, we give an estimator of the intrinsic part in $\boldsymbol{\Sigma}$ using A-SPCA, asymptotic results for A-SPCA under a milder assumption than (A-ii), comparisons between TSPCA and RSPCA, examples of the strongly spiked eigenstructures, asymptotic properties of the conventional PCA, an R-code for A-SPCA and proofs of the theoretical results in the main work together with additional theoretical results and the proofs. The equation numbers and the mathematical symbols used in the supplement are the same as those which are made reference to in the main document.

Appendix A: Estimation of the intrinsic part in the covariance matrix

In this section, we consider estimating the intrinsic part, $\boldsymbol{\Sigma}_1 = \sum_{s=1}^{m} \lambda_s \boldsymbol{h}_s \boldsymbol{h}_s^T$. Fan, Liao and Mincheva (2013) proposed a covariance matrix estimation

procedure called the POET. The key point of this procedure is the estimation of $\boldsymbol{\Sigma}_1$. However, they considered an estimation based on $(\hat{\lambda}_j, \hat{\boldsymbol{h}}_j)$s such that the estimation does not hold consistency properties unless $\delta/\lambda_m = o(1)$. See Proposition A.2 in Section A.2. We apply A-SPCA to the estimator of $\boldsymbol{\Sigma}_1$.

A.1 Estimation of scaled PC directions

Let $\boldsymbol{\beta}_j = \lambda_j^{1/2}\boldsymbol{h}_j$ for $j = 1, ..., m$. Note that $\|\boldsymbol{\beta}_j\|^2 = \lambda_j$ for $j = 1, ..., m$. Let $\tilde{\boldsymbol{\beta}}_j = \tilde{\lambda}_j^{1/2}\tilde{\boldsymbol{h}}_{j*}$ for $j = 1, ..., m$. By combining Proposition 1 and Theorem 1, we have the following result.

Corollary A.1. *Assume (A-i), (A-ii), and (C-i) to (C-iii). Under ($\star$), it holds for $j = 1, ..., m$ that $\|\tilde{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\|^2/\lambda_j = O_P(\eta_j + n^{-1}) = o_P(1)$ as $d \to \infty$ and $n \to \infty$.*

Let $\hat{\boldsymbol{\beta}}_j = \hat{\lambda}_j^{1/2}\hat{\boldsymbol{h}}_j$ for $j = 1, ..., m$.

Proposition A.1. *Assume (A-i) and (C-i). Then, it holds for $j = 1, ..., m$ that $\|\hat{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\|^2/\lambda_j = \delta/\lambda_j + O_P(n^{-1})$ as $d \to \infty$ and $n \to \infty$.*

Therefore, $\hat{\boldsymbol{\beta}}_j$ does not hold consistency unless $\delta/\lambda_j = o(1)$. In contrast, $\tilde{\boldsymbol{\beta}}_j$ holds consistency even when $\delta/\lambda_j \to \infty$.

A.2 Estimation of $\boldsymbol{\Sigma}_1$

The conventional estimator of $\boldsymbol{\Sigma}_1$ is given by $\widehat{\boldsymbol{\Sigma}}_1 = \sum_{s=1}^{m} \hat{\boldsymbol{\beta}}_s \hat{\boldsymbol{\beta}}_s^T$. Let $\|\cdot\|_F$ be the Frobenius norm.

Proposition A.2. *Assume (A-i) and (C-i). Then, it holds that*

$$\|\widehat{\boldsymbol{\Sigma}}_1 - \boldsymbol{\Sigma}_1\|_F^2 = m\delta^2 + 2\sum_{s=1}^{m} \lambda_s \delta\{1 + o_P(1)\} + O_P(\lambda_1^2/n) \quad \textit{as } d \to \infty \textit{ and } n \to \infty.$$

Here, λ_m is the smallest non-zero eigenvalue of $\boldsymbol{\Sigma}_1$. In addition, $\lambda_1\delta/\lambda_m^2 \geq \delta/\lambda_m$ and $\mathrm{tr}(\boldsymbol{\Sigma}_2)/d \in (0, \infty)$ as $d \to \infty$ under (C-ii). Therefore, from (2.3), we have consistency under the conditions in Proposition A.2 in the sense that $\|\widehat{\boldsymbol{\Sigma}}_1 - \boldsymbol{\Sigma}_1\|_F^2 = o_P(\lambda_m^2)$ if $\lambda_1\delta/\lambda_m^2 = o(1)$. However, the consistency does not hold unless δ/λ_m approaches 0.

Here, we consider the following estimator of $\boldsymbol{\Sigma}_1$ using A-SPCA:

$$\widetilde{\boldsymbol{\Sigma}}_1 = \sum_{s=1}^{m} \tilde{\boldsymbol{\beta}}_s \tilde{\boldsymbol{\beta}}_s^T.$$

See Section S2 in Aoshima and Yata (2018) for a consistent estimator of m.

Theorem A.1. *Assume (A-i), (A-ii), and (C-i) to (C-iii). Under ($\star$), it holds that*

$$\|\widetilde{\boldsymbol{\Sigma}}_1 - \boldsymbol{\Sigma}_1\|_F^2 = O_P\Big(\lambda_1^2 n^{-1} + \sum_{s=1}^{m} \lambda_s^2 \eta_s\Big) \quad \textit{as } d \to \infty \textit{ and } n \to \infty.$$

Thus, under the conditions in Theorem A.1, we have consistency in the

sense that

$$\|\widetilde{\mathbf{\Sigma}}_1 - \mathbf{\Sigma}_1\|_F^2 = o_P(\lambda_m^2) \tag{A.1}$$

if $n^{-1}(\lambda_1^2/\lambda_m^2) = o(1)$ and $\sum_{s=1}^m (\lambda_s^2/\lambda_m^2)\eta_s = o(1)$. Further, if $\limsup_{d\to\infty} \lambda_1/\lambda_m < \infty$, then (A.1) holds under the conditions in Theorem A.1. Therefore, $\widetilde{\mathbf{\Sigma}}_1$ holds the consistency even when $\delta/\lambda_m \to \infty$.

Appendix B: Asymptotic results for A-SPCA under a milder assumption than (A-ii)

We consider the following assumption instead of (A-ii):

(A-ii') $\limsup_{d\to\infty} E(x_{i(j),2}^8) < \infty$ for all j; and $\limsup_{d\to\infty} E\{(x_{i(j),2} z_{ij'})^8\} < \infty$ for all j and $j' = 1, ..., m$.

Note that (A-ii) implies (A-ii'). Also, if $x_{i(j),2}$s and $z_{ij'}$s are independent, the second condition of (A-ii') is met under the first condition of (A-ii') and $\limsup_{d\to\infty} E(z_{ij'}^8) < \infty$ for all j. We consider the following divergence condition instead of ($\star$):

($\star$') $\dfrac{d}{n^4} = o(1)$ as $d \to \infty$ and $n \to \infty$.

Note that ($\star$') holds even when $d/n \to \infty$.

Proposition B.1. *After replacing (A-ii) and ($\star$) with (A-ii') and ($\star$'), the results in Theorems 1, 2, A.1 and Corollaries 1, A.1 are still justified.*

Appendix C: Comparisons between TSPCA and RSPCA

In this section, we present several comparisons between threshold-based SPCA (TSPCA) and regularized SPCA (RSPCA).

C.1 Asymptotic property

The key of A-SPCA is the following asymptotic property for the PC-direction by the NR method:

$$\tilde{\boldsymbol{h}}_j = \{1 + o_P(1)\}\boldsymbol{h}_j + \tilde{\boldsymbol{v}}_j \quad \text{as } d \to \infty \text{ and } n \to \infty, \tag{C.1}$$

where $\boldsymbol{h}_j^T \tilde{\boldsymbol{v}}_j = 0$; i.e., the coefficient of $\boldsymbol{h}_j$ in $\tilde{\boldsymbol{h}}_j$ is asymptotically 1. See (2.6) for additional details. However, to the best of my knowledge, the estimation of PC-directions by RSPCAs does not hold such a result. Based on (C.1), we can obtain accurate results for the PC-directions by a thresholded method without requiring threshold values. See Section 3.1 for further details. Thus the threshold-based estimation has a significant theoretical advantage over RSPCAs.

C.2 Computational cost

Zou and Hastie (2006) considered an RSPCA under high-dimensional settings. The first g PC-directions estimated as follows: Let $\boldsymbol{B}_1 = (\boldsymbol{\beta}_1, ..., \boldsymbol{\beta}_g)$ and $\boldsymbol{B}_2$ be $d \times g$ matrices. They considered the following optimization problem:

$$(\widehat{\boldsymbol{B}}_1, \widehat{\boldsymbol{B}}_2) = \underset{\boldsymbol{B}_1, \boldsymbol{B}_2}{\operatorname{argmin}} \sum_{i=1}^{n} \|\boldsymbol{x}_i - \boldsymbol{B}_2 \boldsymbol{B}_1^T \boldsymbol{x}_i\|^2 + \psi \sum_{j=1}^{g} \|\boldsymbol{\beta}_j\|^2 + \sum_{j=1}^{g} \psi_j \|\boldsymbol{\beta}_j\|_1$$

$$\text{subject to } \boldsymbol{B}_2^T \boldsymbol{B}_2 = \boldsymbol{I}_g,$$

where ψ and ψ_js are tuning parametors, $\|\boldsymbol{\beta}_j\|_1$ is the L1 norm of $\boldsymbol{\beta}_j$, and $\widehat{\boldsymbol{B}}_1 = (\hat{\boldsymbol{\beta}}_1, ..., \hat{\boldsymbol{\beta}}_g)$. Then, $\hat{\boldsymbol{\beta}}_j / \|\hat{\boldsymbol{\beta}}_j\|$ is the estimator of $\boldsymbol{h}_j$ for $j = 1, ..., g$. They also expressed the computational cost of the RSPCA as

$$O\{T(dnJ + J^3)\} \text{ when } d > n,$$

where J is the number of nonzero coefficients in the PC-directions and T is the number of iterations before convergence. See Section 3.5 in Zou and Hastie (2006) for further details. On the other hand, by using the singular value decomposition of $\boldsymbol{X} - \overline{\boldsymbol{X}}$, the computational cost of the TSPCA by (1.5) becomes $O(dn)$. The computational cost of the A-SPCA by the R-code in Appendix F is also $O(dn)$. We note that because A-SPCA is a TSPCA, it does not require iterations before convergence. Overall, TSPCAs are easier to handle than RSPCAs in terms of computational complexity.

Here, we compared the computational cost of the A-SPCA with that of the RSPCA proposed by Zou and Hastie (2006) for the (S-i) and (S-ii) settings in Section 4. We used the R-code of the RSPCA in the "elasticnet", which is available from CRAN (`https://cran.r-project.org/web/packages/elasticnet`). We set $K = 2$ and para=c(0.05,0.05) in the R-code and calculated the ratios of computational costs for the RSPCA over the A-SPCA by 2000 iterations for both settings, with the results indicating lower computational costs for the latter, as shown in Fig. C.1. We emphasize that the RSPCA and TSPCA by (1.5) heavily depend on threshold (tuning) values determined by specific cross-validation or information criteria. In contrast, because the A-SPCA does not depend on any threshold (tuning) values, it quickly obtains an accurate result at a lower computational cost.

Appendix D: Examples of the strongly spiked eigenstructures

We provide examples of (C-i) and (C-iii).

First, we consider an intraclass correlation model given by

$$\boldsymbol{\Gamma}_q = \beta(\alpha \boldsymbol{I}_q + (1-\alpha)\mathbf{1}_q\mathbf{1}_q^T), \tag{D.1}$$

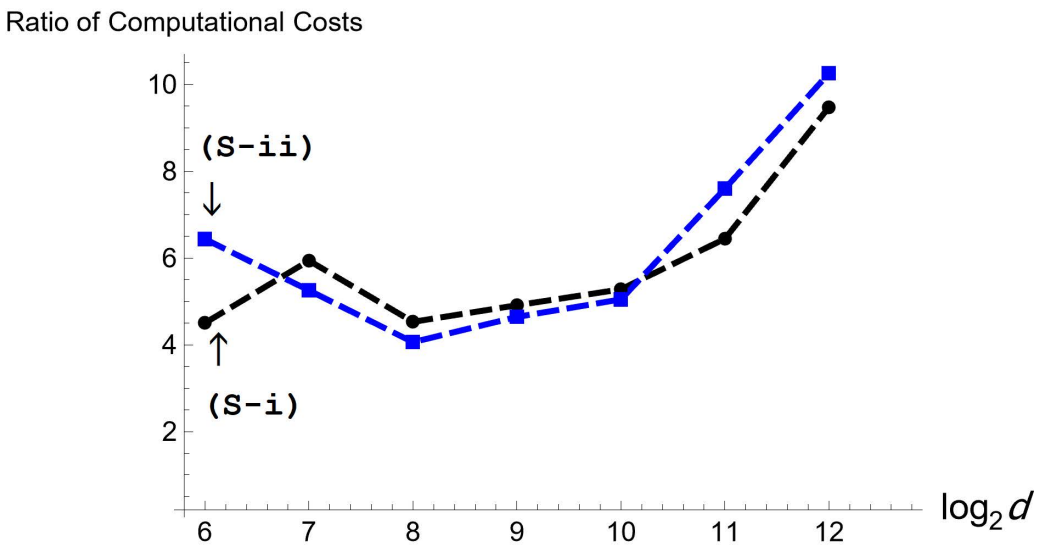

Figure C.1: The ratios of the computational costs for the RSPCA over the A-SPCA for (S-i) $N_d(\mathbf{0}, \mathbf{\Sigma})$, $d = 2^s$ $(s = 6, ..., 12)$, $n = \lceil d^{1/2} \rceil$, where $\mathbf{\Sigma}$ has $\lambda_1 = d^{2/3}$, $\lambda_2 = d^{1/2}$, and $\lambda_3 = \cdots = \lambda_d = 1$ together with $\boldsymbol{h}_1 = (1, 0, ..., 0)^T$ and $\boldsymbol{h}_2 = (0, 1, 0, ..., 0)^T$, and for (S-ii) $N_d(\mathbf{0}, \mathbf{\Sigma})$, $d = 2^s$ $(s = 6, ..., 12)$, $n = \lceil d^{1/2} \rceil$, where $\mathbf{\Sigma}$ has $\lambda_1 \approx d^{2/3}$ and $\lambda_2 \approx d^{1/2}$ together with $\boldsymbol{h}_1 = (1, ..., 1, ..., 0)^T$, whose $\lceil d^{2/3} \rceil$ elements are 1 and $\boldsymbol{h}_2 = (0, ..., 0, 1, ..., 1, 0, ..., 0)^T$, whose $\lceil d^{1/2} \rceil$ elements are 1.

where $\alpha \in (0,1)$ and β (>0) are fixed constants. For the model, $\lambda_{\max}(\boldsymbol{\Gamma}_q) = \beta\{(1-\alpha)q+\alpha\}$ and the other eigenvalues are $\alpha\beta$. If $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}_d$, (C-i) and (C-iii) with $(m, k_{1*}) = (1, d)$ are satisfied because $\boldsymbol{h}_1 = d^{-1/2}\mathbf{1}_d$ when $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}_d$. Then, $\boldsymbol{h}_1$ is a non-sparse vector in the sense that all elements of $\boldsymbol{h}_1$ are nonzero.

Next, we consider the following model.

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Gamma}_{d_1} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{\Gamma}_{d_2} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{\Omega}_{d_3} \end{pmatrix}, \tag{D.2}$$

where $d_1 > d_2 > d_3 > 0$, $d_1 + d_2 + d_3 = d$ and $\boldsymbol{\Omega}_{d_3}$ is a d_3-dimensional non-negative definite matrix. If

$$d_2 \geq d^{1/2},\ \lambda_{\max}(\boldsymbol{\Omega}_{d_3})/d_2 \to 0, \quad \text{and} \quad \{\lambda_{\max}(\boldsymbol{\Omega}_{d_3})\}^2/\mathrm{tr}(\boldsymbol{\Omega}_{d_3}^2) \to 0 \ \text{ as } d \to \infty,$$

(C-i) and (C-iii) with $(m, k_{1*}, k_{2*}) = (2, d_1, d_2)$ are met from the fact that $\boldsymbol{h}_1 = (d_1^{-1/2}\mathbf{1}_{d_1}^T, 0, ..., 0)^T$ and $\boldsymbol{h}_2 = (0, ..., 0, d_2^{-1/2}\mathbf{1}_{d_2}^T, 0, ..., 0)^T$.

Appendix E: Asymptotic properties of the conventional PCA

Let $v_j = \sum_{i=1}^n (z_{ij} - \bar{z}_j)^2/(n-1)$ for all j, where $\bar{z}_j = \sum_{i=1}^n z_{ij}/n$. For the conventional PCA, we obtain the following results from Proposition 2 and

(S6.1) in Aoshima and Yata (2018).

Proposition E.1 (Aoshima and Yata, 2018). *Assume that (A-i) and (C-i). It holds for $j = 1, ..., m$ that*

$$\frac{\hat{\lambda}_j}{\lambda_j} = v_j + \frac{\delta}{\lambda_j} + O_P(n^{-1}) = 1 + \frac{\delta}{\lambda_j} + O_P(n^{-1/2}) \quad \textit{and}$$
$$\textit{Angle}(\hat{\boldsymbol{h}}_j, \boldsymbol{h}_j) = \textit{Arccos}\left(\frac{1}{\sqrt{1+\delta/\lambda_j}} + O_P(n^{-1/2}) \right) \quad \textit{as } d \to \infty \textit{ and } n \to \infty. \tag{E.1}$$

Remark E.1. Equation (E.1) is equivalent to

$$\hat{\boldsymbol{h}}_j^T \boldsymbol{h}_j = (1+\delta/\lambda_j)^{-1/2} + O_P(n^{-1/2}) \quad \text{or}$$
$$\|\hat{\boldsymbol{h}}_j - \boldsymbol{h}_j\|^2 = 2\{1 - (1+\delta/\lambda_j)^{-1/2}\} + O_P(n^{-1/2}).$$

Yata and Aoshima (2012) proposed the noise-reduction (NR) methodology to reduce the bias term; it was introduced using a geometric representation of high-dimensional noise.

Appendix F: R-code for the A-SPCA

We give the following R-code for A-SPCA:

Input ASPCA(X, r); d (≥ 2) by n (≥ 4) matrix X as $X = (\boldsymbol{x}_1, ..., \boldsymbol{x}_n)$, and $r \in [2, \min\{d, n-2\}]$ (the number of components to be computed).

Output values[j]: The estimator of the j-th eigenvalue by A-SPCA (the NR method).

vectors[, j]: The estimator of the j-th PC direction by A-SPCA.

```
ASPCA <- function(X, r){
  d <- dim(X)[1]
  n <- dim(X)[2]
  q <- min(n-2, d, r)
  X <- sweep(X, 1, apply(X, 1, mean), '-')
  X0 <- X
  svd0 <- svd(X0 / (n-1)^(1/2), nu = q, nv = q)
  sval <- svd0$d[1:q]
  svec <- svd0$v[,1:q]
  trSd <- norm(X0, "F")^2 / (n-1)
  nrmval <- numeric(q)
  nrmvec <- matrix(0, d, q)
  aspca <- matrix(0, d, q)
    for (i in 1:q){
      nrmval[i] <- sval[i]^2 - ( trSd
                    - sum(sval[1:i]^2)) / (n-i-1)
      nrmvec[, i] <- X %*% svec[, i]
```

```
                    / sqrt((n-1) * nrmval[i])
      ord <- order(abs(nrmvec[, i]), decreasing=T)
      cri <- 0
      for (j in 1:d){
        cri <- cri + nrmvec[ord[j], i]^2
        aspca[ord[j], i] <- nrmvec[ord[j], i]
        if (cri >= 1){
          break
      }
    }
  }
  return(list(values=nrmval, vectors=aspca))
}
```

Remark F.1. One can calculate the shrinkage PC direction $\tilde{\boldsymbol{h}}_{j\omega}$ with a given constant $\omega_j \in (0, 1]$ by using the above code after replacing "cri >= 1" with "cri >= ω_j".

Appendix G: Proofs

Throughout all the proofs, we assume $\boldsymbol{\mu} = \mathbf{0}$ for the sake of simplicity. Let $\boldsymbol{u}_j = (z_{1j}, ..., z_{nj})^T/(n-1)^{1/2}$ and $\dot{\boldsymbol{u}}_j = \|\boldsymbol{u}_j\|^{-1}\boldsymbol{u}_j$ for all j. From (S6.1) to (S6.5) in Appendix B of Aoshima and Yata (2018), under (A-i) and (C-i), we have that as $d \to \infty$ and $n \to \infty$

$$\tilde{\lambda}_j/\lambda_j = \|\boldsymbol{u}_j\|^2 + O_P(n^{-1}) = 1 + O_P(n^{-1/2})$$

$$\text{and} \quad \hat{\boldsymbol{u}}_j^T \dot{\boldsymbol{u}}_j = 1 + O_P(n^{-1}) \ \text{ for } j = 1, ..., m; \tag{G.1}$$

$$\hat{\boldsymbol{u}}_{j'}^T \boldsymbol{u}_j = O_P(n^{-1/2} \min\{1, \lambda_{j'}/\lambda_j\}) \ \text{ for } j \neq j' \ (\leq m). \tag{G.2}$$

Note that $\|\boldsymbol{u}_j\|^2 = v_j + O_P(n^{-1})$ as $n \to \infty$ for $j = 1, ..., m$. Let $\boldsymbol{P}_n = \boldsymbol{I}_n - \mathbf{1}_n\mathbf{1}_n^T/n$. Note that $\mathbf{1}_n^T\hat{\boldsymbol{u}}_j = 0$ and $\boldsymbol{P}_n\hat{\boldsymbol{u}}_j = \hat{\boldsymbol{u}}_j$ when $\hat{\lambda}_j > 0$ since $\mathbf{1}_n^T\boldsymbol{S}_D\mathbf{1}_n = 0$. Also, when $\hat{\lambda}_j > 0$, note that

$$(\boldsymbol{X} - \overline{\boldsymbol{X}})\hat{\boldsymbol{u}}_j = \boldsymbol{X}\boldsymbol{P}_n\hat{\boldsymbol{u}}_j = \boldsymbol{X}\hat{\boldsymbol{u}}_j = (n-1)^{1/2}\sum_{s=1}^{d}\lambda_s^{1/2}\boldsymbol{h}_s\boldsymbol{u}_s^T\hat{\boldsymbol{u}}_j. \tag{G.3}$$

Let $z_{ij(j'),x} = z_{ij}x_{i(j'),2}$ for $j = 1, ..., m$; $j' = 1, ..., d$ and all i. Let $\bar{z}_{j(j'),x} = \sum_{i=1}^n z_{ij(j'),x}/n$ and $\boldsymbol{x}_{(j'),2} = (x_{1(j'),2}, ..., x_{n(j'),2})^T/(n-1)^{1/2}$ for $j = 1, ..., m$; $j' = 1, ..., d$.

Proof of Lemma 1. Assume (A-i), (A-ii), (C-i), (C-ii) and ($\star$). Let $w_{i(j)} = x_{i(j),2}^2 - \sigma_{(j),2}$ and $\bar{w}_{(j)} = \sum_{i'=1}^n w_{i'(j)}/n$ for all i, j. From (A-ii) and (C-ii),

we note that

$$\limsup_{d\to\infty} E\big\{\exp(tw_{i(j)})\big\} = \limsup_{d\to\infty} \frac{E\big\{\exp(tx_{i(j),2}^2)\big\}}{\exp(t\sigma_{(j),2})} < \infty$$

for $|t| \leq t_1$ and all j. Then, from (A-ii), for any $t > 0$ satisfying $t = o(1)$ as $d \to \infty$ and $n \to \infty$, we have that as $d \to \infty$ and $n \to \infty$

$$P(|\bar{z}_{j(j'),x}| \geq t) \leq 2\exp(-nt^2/\psi_1) \ \text{ for all } j, j'; \text{ and} \tag{G.4}$$

$$P(|\bar{w}_{(j)}| \geq t) \leq 2\exp(-nt^2/\psi_2) \ \text{ for all } j. \tag{G.5}$$

for some fixed constants $\psi_1 > 0$ and $\psi_2 > 0$. Refer to Section 2.1.3 in Wainwright (2019) for the details of this results. Then, from (G.4), it holds that for $j = 1, ..., m$,

$$\sum_{j'=1}^{d} P\{|\bar{z}_{j(j'),x}| \geq (2\psi_1 n^{-1}\log d)^{1/2}\} \leq \sum_{j'=1}^{d} 2d^{-2} \to 0,$$

so that

$$\bar{z}_{j(j'),x} = O_P\{(n^{-1}\log d)^{1/2}\} \ \text{ for all } j, j'. \tag{G.6}$$

Similar to (G.6), from (G.5), we can claim that

$$\bar{w}_{(j)} = O_P\{(n^{-1}\log d)^{1/2}\}$$

for all j, so that

$$\sum_{i=1}^{n} x_{i(j),2}^2/n = \sigma_{(j),2}\{1 + o_P(1)\} \ \text{ for all } j. \tag{G.7}$$

From (G.6) and (G.7), we have that

$$\boldsymbol{u}_j^T \boldsymbol{x}_{(j'),2} = O_P\{(n^{-1}\log d)^{1/2}\} \text{ for } j = 1, ..., m, \text{ and all } j'; \text{ and}$$

$$\|\boldsymbol{x}_{(j),2}\|^2/\sigma_{(j),2} = 1 + o_P(1) \text{ for all } j. \tag{G.8}$$

Here, from (G.1), there exists a unit random vector $\boldsymbol{\varepsilon}_j$ such that $\dot{\boldsymbol{u}}_j^T \boldsymbol{\varepsilon}_j = 0$ and

$$\hat{\boldsymbol{u}}_j = \{1 + O_P(n^{-1})\}\dot{\boldsymbol{u}}_j + \boldsymbol{\varepsilon}_j \times O_P(n^{-1/2}) \text{ for } j = 1, ..., m. \tag{G.9}$$

Note that $\|\boldsymbol{u}_j\| = 1 + o_P(1)$ as $n \to \infty$ for $j = 1, ..., m$. Then, by combining (G.1), (G.8) and (G.9), we have for $j = 1, ..., m$, that

$$\tilde{\lambda}_j^{-1/2} \boldsymbol{x}_{(j'),2}^T \hat{\boldsymbol{u}}_j = O_P\{(\lambda_j^{-1} n^{-1}\log d)^{1/2}\} \text{ for all } j'. \tag{G.10}$$

On the other hand, from (G.1) and (G.2), we have for $j = 1, ..., m$, that

$$\begin{aligned}\frac{\boldsymbol{A}_1 \boldsymbol{X} \hat{\boldsymbol{u}}_j}{\{(n-1)\tilde{\lambda}_j\}^{1/2}} &= \tilde{\lambda}_j^{-1/2} \sum_{j'=1}^{m} \lambda_{j'}^{1/2} \|\boldsymbol{u}_{j'}\| \boldsymbol{h}_{j'} \dot{\boldsymbol{u}}_{j'}^T \hat{\boldsymbol{u}}_j \\ &= \boldsymbol{h}_j\{1 + O_P(n^{-1})\} \\ &\quad + O_P\{(n\lambda_j)^{-1/2}\} \times \sum_{j'=1(\neq j)}^{m} \frac{\lambda_{j'}^{1/2} \boldsymbol{h}_{j'}}{\max\{1, \lambda_{j'}/\lambda_j\}}.\end{aligned} \tag{G.11}$$

Note that $\sigma_{(j''),1} \leq \sigma_{(j'')}$ and $\sigma_{(j''),1} = \sum_{j'=1}^{m} \lambda_{j'} h_{j'(j'')}^2$ for all j'', so that

$$\lambda_{j'}^{1/2} h_{j'(j'')} = O(\sigma_{(j'')}^{1/2}) = O(1) \tag{G.12}$$

for all j'' and $j' = 1, ..., m$. Also, note that $\{(n-1)\tilde{\lambda}_j\}^{-1/2} \boldsymbol{A}_2 \boldsymbol{X} \hat{\boldsymbol{u}}_j = \tilde{\lambda}_j^{-1/2} (\boldsymbol{x}_{(1),2}, ..., \boldsymbol{x}_{(d),2})^T \hat{\boldsymbol{u}}_j$. Then, from (G.3), by combining (G.10) and

(G.11), we have for $j = 1, ..., m$, that

$$\tilde{h}_{j(j')} = h_{j(j')} + O_P\{(\lambda_j^{-1} n^{-1} \log d)^{1/2}\} \ \text{ for all } j'.$$

It concludes the result. □

Proof of Theorem 1. Assume (A-i), (A-ii), (C-i) to (C-iii) and ($\star$). We first consider the proof for $\tilde{\boldsymbol{h}}_{1*}$. We assume $|h_{1(1)}| \geq \cdots \geq |h_{1(d)}|$ for the sake of simplicity. From Lemma 1 and (C-iii), it holds for all j', that

$$\tilde{h}_{1(j')}^2 = h_{1(j')}^2 + o_P\Big(h_{o1(k_{1*})} \max\{|h_{o1(k_{1*})}|, |h_{1(j')}|\}\Big)$$

as $d \to \infty$ and $n \to \infty$. Then, we have that

$$\tilde{h}_{1(j)}^2 = h_{1(j)}^2\{1 + o_P(1)\} \ \text{ for } j = 1, ..., k_{1*}; \ \text{ and}$$
$$\tilde{h}_{1(j)}^2 = h_{1(j)}^2 + o_P\big(h_{o1(k_{1*})}^2\big) \ \text{ for } j = k_{1*} + 1, ..., d. \tag{G.13}$$

From $\sigma_{(j),1} = \sum_{j'=1}^m \lambda_{j'} h_{j'(j)}^2$, (C-ii) and (C-iii), we note that

$$\lambda_1 h_{1(j)}^2 \in (0, \infty) \ \text{ as } d \to \infty \text{ for } j = 1, ..., k_{1*}; \ \text{ and}$$
$$h_{1(j)}^2 = O(\lambda_1^{-1}) \ \text{ for } j = k_{1*} + 1, ..., d. \tag{G.14}$$

Also, we note that

$$k_{j'*} \to \infty \ \text{ and } \ k_{j'*}/\lambda_{j'} \in (0, \infty) \ \text{ as } d \to \infty \text{ for } j' = 1, ..., m. \tag{G.15}$$

Let $\boldsymbol{h}_{1,1} = (h_{1(1)}, ..., h_{1(k_{1*})}, 0, ..., 0)^T$ and $\boldsymbol{h}_{1,2} = (0, ..., 0, h_{1(k_{1*}+1)}, ..., h_{1(d)})^T$. From $\boldsymbol{\Sigma}_2^{1/2} \boldsymbol{h}_1 = \boldsymbol{0}$, we note that $\boldsymbol{\Sigma}_2^{1/2} \boldsymbol{h}_{1,1} = -\boldsymbol{\Sigma}_2^{1/2} \boldsymbol{h}_{1,2}$, so that $\boldsymbol{h}_{1,1}^T \boldsymbol{\Sigma}_2 \boldsymbol{h}_{1,1} =$

$\boldsymbol{h}_{1,2}^T\boldsymbol{\Sigma}_2\boldsymbol{h}_{1,2} \le \|\boldsymbol{h}_{1,2}\|^2\lambda_{\max}(\boldsymbol{\Sigma}_2) = \eta_1\lambda_{m+1}$. Here, we have that

$$E\Big\{\Big(\sum_{s=1}^{k_{1*}} h_{1(s)}\boldsymbol{u}_1^T\boldsymbol{x}_{(s),2}\Big)^2\Big\} = O\Big(\frac{\boldsymbol{h}_{1,1}^T\boldsymbol{\Sigma}_2\boldsymbol{h}_{1,1}}{n}\Big) = O(\eta_1\lambda_{m+1}/n) \quad \text{and}$$

$$E\Big(\Big\|\sum_{s=1}^{k_{1*}} h_{1(s)}\boldsymbol{x}_{(s),2}\Big\|^2\Big) = O(\boldsymbol{h}_{1,1}^T\boldsymbol{\Sigma}_2\boldsymbol{h}_{1,1}) = O(\eta_1\lambda_{m+1}).$$

Then, by using Markov's inequality, for any $c > 0$, we have that $P\{|\sum_{s=1}^{k_{1*}} h_{1(s)} \times\boldsymbol{u}_1^T\boldsymbol{x}_{(s),2}|^2 \ \ge\ c(\eta_1\lambda_{m+1}/n)\} \ \le\ E\{(\sum_{s=1}^{k_{1*}} h_{1(s)}\boldsymbol{u}_1^T\boldsymbol{x}_{(s),2})^2\}n/(\eta_1\lambda_{m+1}c) = O(c^{-1})$ and $P(\|\sum_{s=1}^{k_{1*}} h_{1(s)}\boldsymbol{x}_{(s),2}\|^2 \ge c\eta_1\lambda_{m+1}) = O(c^{-1})$, so that

$$\sum_{s=1}^{k_{1*}} h_{1(s)}\boldsymbol{u}_1^T\boldsymbol{x}_{(s),2} = O_P\big(\sqrt{\eta_1\lambda_{m+1}/n}\big) \ \text{ and } \ \Big\|\sum_{s=1}^{k_{1*}} h_{1(s)}\boldsymbol{x}_{(s),2}\Big\| = O_P(\sqrt{\eta_1\lambda_{m+1}}). \tag{G.16}$$

By combining (G.1), (G.9) and (G.16), we have that

$$\frac{\boldsymbol{h}_{1,1}^T\boldsymbol{A}_2\boldsymbol{X}\hat{\boldsymbol{u}}_1}{\{(n-1)\tilde{\lambda}_1\}^{1/2}} = \tilde{\lambda}_1^{-1/2}\sum_{s=1}^{k_{1*}} h_{1(s)}\hat{\boldsymbol{u}}_1^T\boldsymbol{x}_{(s),2} = O_P(\eta_1^{1/2}n^{-1/2}) = O_P(\eta_1 + n^{-1}). \tag{G.17}$$

Note that $|\boldsymbol{h}_{1,1}^T\boldsymbol{h}_j| = |\boldsymbol{h}_{1,2}^T\boldsymbol{h}_j| \le \eta_1^{1/2}$ for $j \ge 2$. Then, from (G.11), it holds that

$$\begin{aligned}\frac{\boldsymbol{h}_{1,1}^T\boldsymbol{A}_1\boldsymbol{X}\hat{\boldsymbol{u}}_1}{\{(n-1)\tilde{\lambda}_1\}^{1/2}} &= (1-\eta_1)\{1 + O_P(n^{-1})\} + O_P\big(\sqrt{\eta_1/n}\big)\\ &= 1 + O_P(\eta_1 + n^{-1}).\end{aligned} \tag{G.18}$$

From (G.17) and (G.18), it holds that

$$\sum_{s=1}^{k_{1*}} h_{1(s)}\tilde{h}_{1(s)} = 1 + O_P(\eta_1 + n^{-1}). \tag{G.19}$$

We note that $E(\sum_{s=1}^{k_{1*}} \bar{z}_{1(s),x}^2) = O(\sum_{s=1}^{k_{1*}} \sigma_{(s),2}/n) = O(k_{1*}/n)$ and $E(\sum_{s=1}^{k_{1*}} \|\boldsymbol{x}_{(s),2}\|^2) = O(k_{1*})$. Then, from (G.1) and (G.15), we have that $\tilde{\lambda}_1^{-1} \sum_{s=1}^{k_{1*}} (\boldsymbol{x}_{(s),2}^T \boldsymbol{u}_1)^2 = O_P(1/n)$ and $\tilde{\lambda}_1^{-1} \sum_{s=1}^{k_{1*}} (\boldsymbol{x}_{(s),2}^T \boldsymbol{\varepsilon}_1/n^{-1/2})^2 \leq \tilde{\lambda}_1^{-1} \sum_{s=1}^{k_{1*}} \|\boldsymbol{x}_{(s),2}\|^2/n = O_P(1/n)$, where $\boldsymbol{\varepsilon}_1$ is defined in (G.9). Thus it holds that

$$\tilde{\lambda}_1^{-1} \|(\boldsymbol{x}_{(1),2}, ..., \boldsymbol{x}_{(k_{1*}),2})^T \hat{\boldsymbol{u}}_1\|^2 = O_P(1/n). \tag{G.20}$$

Let $\boldsymbol{x}_{(j),1} = (x_{1(j),1}, ..., x_{n(j),1})^T/(n-1)^{1/2}$ for all j. From (G.11), (G.12) and (G.14), it holds that

$$\begin{aligned} \tilde{\lambda}_1^{-1} \|(\boldsymbol{x}_{(1),1}, ..., \boldsymbol{x}_{(k_{1*}),1})^T \hat{\boldsymbol{u}}_1 - (h_{1(1)}, ..., h_{1(k_{1*})})^T\|^2 &= O_P\left(\sum_{s=1}^{k_{1*}} (n\lambda_1)^{-1}\right) \\ &= O_P(1/n). \end{aligned} \tag{G.21}$$

From (G.20) and (G.21), it holds that $\sum_{s=1}^{k_{1*}} (\tilde{h}_{1(s)} - h_{1(s)})^2 = O_P(1/n)$. Then, from (G.19), it holds that

$$\begin{aligned} \sum_{s=1}^{k_{1*}} \tilde{h}_{1(s)}^2 &= \sum_{s=1}^{k_{1*}} \{(\tilde{h}_{1(s)} - h_{1(s)})^2 + 2h_{1(s)}\tilde{h}_{1(s)} - h_{1(s)}^2\} \\ &= 1 + O_P(\eta_1 + n^{-1}). \end{aligned} \tag{G.22}$$

Let $\mathcal{D} = \{j | \, \tilde{h}_{1*(j)} = 0 \text{ for } j = 1, ..., d\}$, $\mathcal{D}_1 = \{j | \, \tilde{h}_{1*(j)} = 0 \text{ for } j = 1, ..., k_{1*}\}$, $\mathcal{D}_2 = \{j | \, \tilde{h}_{1*(j)} = 0 \text{ for } j = k_{1*}+1, ..., d\}$ and $\mathcal{D}_* = \{k_{1*}+1, ..., d\}$. From (G.13) and (C-iii), it holds that

$$\max_{j \in \{k_{1*}+1, ..., d\}} \tilde{h}_{1(j)}^2 < \min_{j \in \{1, ..., k_{1*}\}} \tilde{h}_{1(j)}^2 \tag{G.23}$$

with probability tending to 1. Here, we assume

$$\liminf_{d\to\infty,n\to\infty} P(\tilde{k}_1 > k_{1*}) > 0 \ \text{ and } \ \liminf_{d\to\infty,n\to\infty} P(\tilde{k}_1 \le k_{1*}) > 0$$

for the sake of simplicity. Then, from (G.23), we have that

$$\mathcal{D} = \mathcal{D}_2 \subset \mathcal{D}_* \ \text{ and } \ \tilde{h}_{o1(\tilde{k}_1)} \in \{\tilde{h}_{1(k_{1*}+1)}, ..., \tilde{h}_{1(d)}\} \ \text{ if } \tilde{k}_1 > k_{1*}; \ \text{ and}$$

$$\mathcal{D}_2 = \mathcal{D}_* \ \text{ if } \tilde{k}_1 \le k_{1*} \tag{G.24}$$

with probability tending to 1. If $\tilde{k}_1 > k_{1*}$, from (2.3), Lemma 1 and (G.24), we have that

$$\begin{aligned} 1 \le \|\tilde{\boldsymbol{h}}_{1*}\|^2 \le 1 + \tilde{h}^2_{o1(\tilde{k}_1)} &= 1 + O_P(h^2_{1(k_{1*}+1)} + \lambda_1^{-1} n^{-1} \log d) \\ &= 1 + O_P(\eta_1 + n^{-1}). \end{aligned} \tag{G.25}$$

Note that $\sum_{s\in\mathcal{D}_*\setminus\mathcal{D}_2} h^2_{1(s)} \le \sum_{s\in\mathcal{D}_*} h^2_{1(s)} = \eta_1$. Also, from (G.24), note that $\sum_{s\in\mathcal{D}_*\setminus\mathcal{D}_2} \tilde{h}^2_{1(s)} = \|\tilde{\boldsymbol{h}}_{1*}\|^2 - \sum_{s=1}^{k_{1*}} \tilde{h}^2_{1(s)}$ with probability tending to 1 if $\tilde{k}_1 > k_{1*}$. Then, if $\tilde{k}_1 > k_{1*}$, from (G.22) and (G.25), we have that

$$\begin{aligned} &\sum_{s\in\mathcal{D}_*\setminus\mathcal{D}_2} \tilde{h}^2_{1(s)} = O_P(\eta_1 + n^{-1}) \ \text{ and} \\ &\Big|\sum_{s\in\mathcal{D}_*\setminus\mathcal{D}_2} h_{1(s)}\tilde{h}_{1(s)}\Big| \le \Big(\sum_{s\in\mathcal{D}_*\setminus\mathcal{D}_2} h^2_{1(s)}\Big)^{1/2}\Big(\sum_{s\in\mathcal{D}_*\setminus\mathcal{D}_2} \tilde{h}^2_{1(s)}\Big)^{1/2} \\ &\qquad = O_P\{\eta_1^{1/2}(\eta_1 + n^{-1})^{1/2}\} = O_P(\eta_1 + n^{-1}), \end{aligned} \tag{G.26}$$

so that from (G.19),

$$\boldsymbol{h}_1^T \tilde{\boldsymbol{h}}_{1*} = \sum_{s=1}^{k_{1*}} h_{1(s)}\tilde{h}_{1(s)} + \sum_{s\in\mathcal{D}_*\setminus\mathcal{D}_2} h_{1(s)}\tilde{h}_{1(s)} = 1 + O_P(\eta_1 + n^{-1}). \tag{G.27}$$

When $\tilde{k}_1 \leq k_{1*}$, from (G.24), we note that $\|\tilde{\boldsymbol{h}}_{1*}\|^2 + \sum_{s \in \mathcal{D}_1} \tilde{h}_{1(s)}^2 = \sum_{s=1}^{k_{1*}} \tilde{h}_{1(s)}^2$ with probability tending to 1. Then, from $\|\tilde{\boldsymbol{h}}_{1*}\|^2 \geq 1$, (G.13) and (G.22) if $\tilde{k}_1 \leq k_{1*}$, we have that

$$\sum_{s \in \mathcal{D}_1} \tilde{h}_{1(s)}^2 = \sum_{s=1}^{k_{1*}} \tilde{h}_{1(s)}^2 - \|\tilde{\boldsymbol{h}}_{1*}\|^2 \leq \sum_{s=1}^{k_{1*}} \tilde{h}_{1(s)}^2 - 1 = O_P(\eta_1 + n^{-1})$$

$$\text{and} \quad \sum_{s \in \mathcal{D}_1} |h_{1(s)} \tilde{h}_{1(s)}| = \sum_{s \in \mathcal{D}_1} \tilde{h}_{1(s)}^2 \{1 + o_P(1)\} = O_P(\eta_1 + n^{-1}). \tag{G.28}$$

Thus from (G.19) and (G.22), if $\tilde{k}_1 \leq k_{1*}$, we have that

$$\boldsymbol{h}_1^T \tilde{\boldsymbol{h}}_{1*} = \sum_{s=1}^{k_{1*}} h_{1(s)} \tilde{h}_{1(s)} - \sum_{s \in \mathcal{D}_1} h_{1(s)} \tilde{h}_{1(s)} = 1 + O_P(\eta_1 + n^{-1})$$

$$\text{and} \quad \|\tilde{\boldsymbol{h}}_{1*}\|^2 = \sum_{s=1}^{k_{1*}} \tilde{h}_{1(s)}^2 - \sum_{s \in \mathcal{D}_1} \tilde{h}_{1(s)}^2 = 1 + O_P(\eta_1 + n^{-1}). \tag{G.29}$$

If $P(\tilde{k}_1 \leq k_{1*}) = o(1)$ or $P(\tilde{k}_1 > k_{1*}) = o(1)$, we can obtain (G.25) and (G.27) or (G.29). Thus from (G.25), (G.27) and (G.29), we can conclude the results for $\tilde{\boldsymbol{h}}_{1*}$. As for $\tilde{\boldsymbol{h}}_{j*}$ with $j \geq 2$, we obtain the results similarly. It concludes the results of Theorem 1. □

Proof of Corollary 1. Assume (A-i) and (A-ii), (C-i) to (C-iii) and ($\star$). From Theorem 1, we can claim the first result of Corollary 1. Next, we consider the second result of Corollary 1. Note that $|\tilde{\boldsymbol{h}}_{j*}^T \boldsymbol{h}_{j'}| = |(\tilde{\boldsymbol{h}}_{j*} - \boldsymbol{h}_j)^T \boldsymbol{h}_{j'}| \leq \|\tilde{\boldsymbol{h}}_{j*} - \boldsymbol{h}_j\|$ for $j < j'$ $(\leq m)$. Thus from the first result of Corollary 1, it holds that as $d \to \infty$ and $n \to \infty$

$$\tilde{\boldsymbol{h}}_{j*}^T \boldsymbol{h}_{j'} = O_P(\sqrt{\eta_j + n^{-1}}) = O_P(\eta_j^{1/2} + n^{-1/2}) \ \text{ for } j < j' \ (\leq m). \tag{G.30}$$

Here, we consider the case of $\tilde{\boldsymbol{h}}_{2*}^T \boldsymbol{h}_1$. We assume $|h_{2(1)}| \geq \cdots \geq |h_{2(d)}|$ for the sake of simplicity. Let $\boldsymbol{h}_{1,2*} = (h_{1(1)}, ..., h_{1(k_{2*})}, 0, ..., 0)^T$. Similar to (G.14), we note that

$$h_{j''(s)}^2 = O(\lambda_{j''}^{-1}) \text{ for } s = 1, ..., d;\ j'' = 1, ..., m. \tag{G.31}$$

From (G.15) and (G.31), it holds that $\|\boldsymbol{h}_{1,2*}\|^2 \leq \sum_{s=1}^{k_{2*}} h_{1(s)}^2 = O(\lambda_2/\lambda_1)$, so that

$$\boldsymbol{h}_{1,2*}^T \boldsymbol{\Sigma}_2 \boldsymbol{h}_{1,2*} \leq \|\boldsymbol{h}_{1,2*}\|^2 \lambda_{\max}(\boldsymbol{\Sigma}_2) = O(\lambda_{m+1}\lambda_2/\lambda_1).$$

Then, similar to (G.16)-(G.17), we can claim that

$$\frac{\boldsymbol{h}_{1,2*}^T \boldsymbol{A}_2 \boldsymbol{X} \hat{\boldsymbol{u}}_2}{\{(n-1)\tilde{\lambda}_2\}^{1/2}} = \tilde{\lambda}_2^{-1/2} \sum_{s=1}^{k_{2*}} h_{1(s)} \hat{\boldsymbol{u}}_2^T \boldsymbol{x}_{(s),2} = O_P(\lambda_2^{1/2}/(n\lambda_1)^{1/2}). \tag{G.32}$$

From (G.15) and (G.31), we note that $|\boldsymbol{h}_{1,2*}^T \boldsymbol{h}_{j''}| \leq \sum_{s=1}^{k_{2*}} |h_{1(s)} h_{j''(s)}| = \{\lambda_2/(\lambda_1\lambda_{j''})^{1/2}\}$ for $j''(\neq 2) \leq m$. Also, note that $|\boldsymbol{h}_{1,2*}^T \boldsymbol{h}_2| = |(\boldsymbol{h}_1 - \boldsymbol{h}_{1,2*})^T \boldsymbol{h}_2| \leq \eta_2^{1/2}$. Then, from (G.11), we have that

$$\begin{aligned}\frac{\boldsymbol{h}_{1,2*}^T \boldsymbol{A}_1 \boldsymbol{X} \hat{\boldsymbol{u}}_2}{\{(n-1)\tilde{\lambda}_2\}^{1/2}} =& O_P(\eta_2^{1/2}) \\ &+ O_P\left((n\lambda_2)^{-1/2} \sum_{j''=1(\neq 2)}^{m} \frac{\lambda_2/\lambda_1^{1/2}}{\max\{1, \lambda_{j''}/\lambda_2\}}\right) \\ =& O_P\{\eta_2^{1/2} + \lambda_2^{1/2}/(n\lambda_1)^{1/2}\}. \end{aligned} \tag{G.33}$$

From (G.32) and (G.33), it holds that

$$\sum_{s=1}^{k_{2*}} h_{1(s)} \tilde{h}_{2(s)} = O_P\{\eta_2^{1/2} + \lambda_2^{1/2}/(n\lambda_1)^{1/2}\}. \tag{G.34}$$

Here, we assume

$$\liminf_{d\to\infty,n\to\infty} P(\tilde{k}_2 > k_{2*}) > 0 \quad \text{and} \quad \liminf_{d\to\infty,n\to\infty} P(\tilde{k}_2 \leq k_{2*}) > 0$$

for the sake of simplicity. Let $\mathcal{G} = \{j|\ \tilde{h}_{2*(j)} = 0 \text{ for } j = 1, ..., d\}$, $\mathcal{G}_1 = \{j|\ \tilde{h}_{2*(j)} = 0 \text{ for } j = 1, ..., k_{2*}\}$, $\mathcal{G}_2 = \{j|\ \tilde{h}_{2*(j)} = 0 \text{ for } j = k_{2*}+1, ..., d\}$ and $\mathcal{G}_* = \{k_{2*}+1, ..., d\}$. Then, similar to (G.24), we can claim that

$$\mathcal{G} = \mathcal{G}_2 \subset \mathcal{G}_* \ \text{ if } \tilde{k}_2 > k_{2*}; \ \text{ and } \ \mathcal{G}_2 = \mathcal{G}_* \ \text{ if } \tilde{k}_2 \leq k_{2*} \tag{G.35}$$

with probability tending to 1. If $\tilde{k}_2 \leq k_{2*}$, similar to (G.28), we can claim that

$$\sum_{s\in\mathcal{G}_1} \tilde{h}_{2(s)}^2 = O_P(\eta_2 + n^{-1}) \ \text{ and}$$
$$\tilde{\boldsymbol{h}}_{2*}^T \boldsymbol{h}_1 = \sum_{s=1}^{k_{2*}} h_{1(s)}\tilde{h}_{2(s)} - \sum_{s\in\mathcal{G}_1} h_{1(s)}\tilde{h}_{2(s)}. \tag{G.36}$$

with probability tending to 1. From (G.15) and (G.31), we note that $\sum_{s\in\mathcal{G}_1} h_{1(s)}^2 \leq \sum_{s=1}^{k_{2*}} h_{1(s)}^2 = O(\lambda_2/\lambda_1)$. Then, from (G.36), if $\tilde{k}_2 \leq k_{2*}$, it holds that

$$\sum_{s\in\mathcal{G}_1} |h_{1(s)}\tilde{h}_{2(s)}| \leq \left(\sum_{s\in\mathcal{G}_1} h_{1(s)}^2 \sum_{s\in\mathcal{G}_1} \tilde{h}_{2(s)}^2\right)^{1/2} = O_P\{(\eta_2^{1/2} + n^{-1/2})\lambda_2^{1/2}/\lambda_1^{1/2}\},$$

so that from (G.34) and (G.36),

$$\tilde{\boldsymbol{h}}_{2*}^T \boldsymbol{h}_1 = \sum_{s=1}^{k_{2*}} h_{1(s)}\tilde{h}_{2(s)} - \sum_{s\in\mathcal{G}_1} h_{1(s)}\tilde{h}_{2(s)} = O_P\{\eta_2^{1/2} + \lambda_2^{1/2}/(n\lambda_1)^{1/2}\}. \tag{G.37}$$

If $\tilde{k}_2 > k_{2*}$, similar to (G.26), we can claim that

$$\sum_{s \in \mathcal{G}_* \backslash \mathcal{G}_2} \tilde{h}_{2(s)}^2 = O_P(\eta_2 + n^{-1}). \tag{G.38}$$

Then, by noting that $\sum_{s \in \mathcal{G}_* \backslash \mathcal{G}_2} h_{1(s)}^2 \leq 1$, it holds that

$$\begin{aligned} \Big| \sum_{s \in \mathcal{G}_* \backslash \mathcal{G}_2} h_{1(s)} \tilde{h}_{2(s)} \Big| &\leq \Big(\sum_{s \in \mathcal{G}_* \backslash \mathcal{G}_2} h_{1(s)}^2 \Big)^{1/2} \Big(\sum_{s \in \mathcal{G}_* \backslash \mathcal{G}_2} \tilde{h}_{2(s)}^2 \Big)^{1/2} \\ &= O_P\{(\eta_2 + n^{-1})^{1/2}\}. \end{aligned} \tag{G.39}$$

Here, similar to (G.22), from (G.35), if $\tilde{k}_2 > k_{2*}$, we can claim that

$$\sum_{s=1}^{k_{2*}} \tilde{h}_{2(s)}^2 = \sum_{s=1}^{k_{2*}} \tilde{h}_{o2(s)}^2 = 1 + O_P(\eta_2 + n^{-1}). \tag{G.40}$$

Then, from (2.5), it holds that

$$\sum_{s=k_{2*}+1}^{d} \tilde{h}_{o2(s)}^2 = \|\tilde{\boldsymbol{h}}_2\|^2 - \sum_{s=1}^{k_{2*}} \tilde{h}_{o2(s)}^2 = (\delta/\lambda_2)\{1 + o_P(1)\} + O_P(\eta_2 + n^{-1}).$$

Thus from $\liminf_{d \to \infty} \mathrm{tr}(\boldsymbol{\Sigma}_2)/d > 0$, (G.15) and (G.40), if $\tilde{k}_2 > k_{2*}$, $\lambda_2 = o(d)$ and $\eta_2 = o(n^{-1})$, we have that

$$\begin{aligned} &\sum_{s=k_{2*}+1}^{\tilde{k}_2} \tilde{h}_{o2(s)}^2 \geq (\tilde{k}_2 - k_{2*})\{\mathrm{tr}(\boldsymbol{\Sigma}_2)/(d\lambda_2 n)\}\{1 + o_P(1)\} \ \text{ and} \\ &\sum_{s=1}^{k_{2*}} \tilde{h}_{o2(s)}^2 = 1 + O_P(n^{-1}). \end{aligned} \tag{G.41}$$

Let $\#(A)$ denote the cardinality of the set A. Note that

$$\#(\mathcal{G}_* \backslash \mathcal{G}_2) = \tilde{k}_2 - k_{2*}$$

with probability tending to 1 if $\tilde{k}_2 > k_{2*}$. From $\sum_{s=1}^{\tilde{k}_2} \tilde{h}_{o2(s)}^2 \geq 1$ and (G.41), if $\tilde{k}_2 > k_{2*}$, $\lambda_2 = o(d)$ and $\eta_2 = o(n^{-1})$, it holds that $\tilde{k}_2 - k_{2*} = O_P(\lambda_2)$, so that from (G.15) and (G.38),

$$\Big| \sum_{s \in \mathcal{G}_* \setminus \mathcal{G}_2} h_{1(s)} \tilde{h}_{2(s)} \Big| \leq \Big(\sum_{s \in \mathcal{G}_* \setminus \mathcal{G}_2} h_{1(s)}^2 \Big)^{1/2} \Big(\sum_{s \in \mathcal{G}_* \setminus \mathcal{G}_2} \tilde{h}_{2(s)}^2 \Big)^{1/2}$$
$$= O_P\{(\lambda_2/\lambda_1)^{1/2}(\eta_2 + n^{-1})^{1/2}\}. \tag{G.42}$$

From (2.3), note that $\lambda_1/\lambda_2 = O(1)$ if $\liminf_{d\to\infty} \lambda_2/d > 0$. Thus it holds that

$$\eta_2^{1/2} + \lambda_2^{1/2}/(n\lambda_1)^{1/2} = O(\eta_2^{1/2}) \ \text{ if } \liminf_{d\to\infty, n\to\infty} n\eta_2 > 0; \text{ and}$$
$$\eta_2^{1/2} + \lambda_2^{1/2}/(n\lambda_1)^{1/2} = O(\eta_2^{1/2} + n^{-1/2}) \ \text{ if } \liminf_{d\to\infty} \lambda_2/d > 0. \tag{G.43}$$

Thus from (G.34), (G.39), (G.42) and (G.43), if $\tilde{k}_2 > k_{2*}$, we have that

$$\tilde{\boldsymbol{h}}_{2*}^T \boldsymbol{h}_1 = \sum_{s=1}^{k_{2*}} h_{1(s)} \tilde{h}_{2(s)} + \sum_{s \in \mathcal{G}_* \setminus \mathcal{G}_2} h_{1(s)} \tilde{h}_{2(s)} = O_P\{\eta_2^{1/2} + \lambda_2^{1/2}/(n\lambda_1)^{1/2}\}. \tag{G.44}$$

If $P(\tilde{k}_2 > k_{2*}) = o(1)$ or $P(\tilde{k}_2 \leq k_{2*}) = o(1)$, we can obtain (G.37) or (G.44). Thus from (G.37) and (G.44), we can conclude the second result for $\tilde{\boldsymbol{h}}_{2*}^T \boldsymbol{h}_1$. As for $\tilde{\boldsymbol{h}}_{j*}^T \boldsymbol{h}_{j'}$ with $j' < j$; $(j', j) \neq (1, 2)$, we obtain the result similarly. From (G.30), we can claim the second result of Corollary 1.

For the third result of Corollary 1, by noting that.

$$\begin{aligned}\tilde{\boldsymbol{h}}_{j*}^T\tilde{\boldsymbol{h}}_{j'*} =& \boldsymbol{h}_j^T\tilde{\boldsymbol{h}}_{j'*} + (\tilde{\boldsymbol{h}}_{j*} - \boldsymbol{h}_j)^T\tilde{\boldsymbol{h}}_{j'*} \\ =& \boldsymbol{h}_j^T\tilde{\boldsymbol{h}}_{j'*} + \boldsymbol{h}_{j'}^T\tilde{\boldsymbol{h}}_{j*} + (\tilde{\boldsymbol{h}}_{j*} - \boldsymbol{h}_j)^T(\tilde{\boldsymbol{h}}_{j'*} - \boldsymbol{h}_{j'})\end{aligned}$$

for $j \neq j'$, from the second result of Corollary 1, it concludes the result. □

Proof of Proposition 3. From (G.3), it holds that $\tilde{\boldsymbol{h}}_j^T\boldsymbol{h}_{j'} = \lambda_{j'}^{1/2}\boldsymbol{u}_{j'}^T\hat{\boldsymbol{u}}_j/\tilde{\lambda}_j^{1/2}$ for $j \neq j'$. Thus from (G.1) and (G.2), we can conclude the result. □

Proof of Theorem 2. Assume (A-i), (A-ii), (C-i), (C-ii), (C-iii') and ($\star$). We first consider the proof for $\tilde{\boldsymbol{h}}_{1\omega}$. We assume $|h_{1(1)}| \geq \cdots \geq |h_{1(d)}|$ for the sake of simplicity. Similar to (G.13), we can claim that

$$\begin{aligned}&\tilde{h}_{1(j)}^2 = h_{1(j)}^2\{1 + o_P(1)\} \ \text{ for } j = 1, ..., k_{1\omega} + r_1; \ \text{ and} \\ &\tilde{h}_{1(j)}^2 = h_{1(j)}^2 + o_P\big(h_{o1(k_{1\omega}+r_1)}^2\big) \ \text{ for } j = k_{1\omega} + r_1 + 1, ..., d \end{aligned} \tag{G.45}$$

as $d \to \infty$ and $n \to \infty$. Similar to (G.14), from (C-iii'), we note for $j' = 1, ..., m$, that

$$\begin{aligned}&\lambda_1 h_{1(j)}^2 \in (0, \infty) \ \text{ as } d \to \infty \text{ for } j = 1, ..., k_{1\omega} + r_1; \\ &\text{and } \ h_{1(j)}^2 = O(\lambda_1^{-1}) \ \text{ for } j = k_{1\omega} + r_1 + 1, ..., d. \end{aligned} \tag{G.46}$$

Also, we note that

$$k_{j'\omega} \to \infty \ \text{ and } \ k_{j'\omega}/(\omega_{j'}\lambda_{j'}) \in (0,\infty) \ \text{ as } d \to \infty \text{ for } j' = 1, ..., m. \tag{G.47}$$

Then, from (G.45), we have that

$$\omega_1 \leq \|\boldsymbol{h}_{1\omega}\|^2 \leq \omega_1 + h^2_{o1(k_{1\omega})} = \omega_1 + O(\lambda_1^{-1}) = \omega_1 + O(\omega_1/k_{1\omega}) \ \text{ and}$$

$$\omega_1 \leq \|\tilde{\boldsymbol{h}}_{1\omega}\|^2 \leq \omega_1 + \tilde{h}^2_{o1(\tilde{k}_{1\omega})} = \omega_1 + O_P(\omega_1/k_{1\omega}). \tag{G.48}$$

From (C-ii), all the elements of $\boldsymbol{\Sigma}_2$ are bounded. Thus, from (G.46) and $\boldsymbol{h}_{1\omega} = (h_{1(1)}, ..., h_{1(k_{1\omega})}, 0, ..., 0)^T$, we note that $\boldsymbol{h}_{1\omega}^T \boldsymbol{\Sigma}_2 \boldsymbol{h}_{1\omega} = O(k_{1\omega}^2/\lambda_1)$. Then, similar to (G.16)-(G.17), from (G.47), we can claim that

$$\frac{\boldsymbol{h}_{1\omega}^T \boldsymbol{A}_2 \boldsymbol{X} \hat{\boldsymbol{u}}_1}{\{(n-1)\tilde{\lambda}_1\}^{1/2}} = O_P(k_{1\omega}\lambda_1^{-1} n^{-1/2}) = O_P(\omega_1 n^{-1/2}). \tag{G.49}$$

From (G.31), we note that $|\boldsymbol{h}_{1\omega}^T \boldsymbol{h}_j| = |\sum_{s=1}^{k_{1\omega}} h_{1(s)} h_{j(s)}| = O\{k_{1\omega}/(\lambda_1\lambda_j)^{1/2}\}$ for $j = 2, ..., m$. Then, from (G.11) and (G.48), it holds that

$$\begin{aligned}\frac{\boldsymbol{h}_{1\omega}^T \boldsymbol{A}_1 \boldsymbol{X} \hat{\boldsymbol{u}}_1}{\{(n-1)\tilde{\lambda}_1\}^{1/2}} =& \boldsymbol{h}_{1\omega}^T \boldsymbol{h}_1 \{1 + O_P(n^{-1})\} + O_P(k_{1\omega}\lambda_1^{-1} n^{-1/2}) \\ =& \omega_1 + O_P(\omega_1/k_{1\omega} + \omega_1 n^{-1/2}).\end{aligned} \tag{G.50}$$

From (G.49) and (G.50), it holds that

$$\sum_{s=1}^{k_{1\omega}} \tilde{h}_{1(s)} h_{1(s)} = \omega_1 + O_P(\omega_1/k_{1\omega} + \omega_1 n^{-1/2}). \tag{G.51}$$

From (G.47), note that $E(\sum_{s=1}^{k_{1\omega}} \bar{z}^2_{1(s),x}) = O(\omega_1\lambda_1/n)$ and $E(\sum_{s=1}^{k_{1\omega}} \|\boldsymbol{x}_{(s),2}\|^2) = O(\omega_1\lambda_1)$. Then, similar to (G.20)-(G.21), we have that $\sum_{s=1}^{k_{1\omega}} \{(\tilde{h}_{1(s)} -$

$h_{1(s)})^2 = O_P(\omega_1/n)$, so that from (G.45)-(G.48) and (G.51),

$$\sum_{s=1}^{k_{1\omega}+r_1} \tilde{h}_{1(s)}^2 = \sum_{s=1}^{k_{1\omega}} \{(\tilde{h}_{1(s)} - h_{1(s)})^2 + 2h_{1(s)}\tilde{h}_{1(s)} - h_{1(s)}^2\} + \sum_{s=k_{1\omega}+1}^{k_{1\omega}+r_1} \tilde{h}_{1(s)}^2$$
$$= \omega_1 + O_P(\omega_1/k_{1\omega} + \omega_1 n^{-1/2}). \tag{G.52}$$

From (G.45), it holds that

$$\max_{j\in\{k_{1\omega}+r_1+1,\ldots,d\}} \tilde{h}_{1(j)}^2 < \min_{j\in\{1,\ldots,k_{1*}+r_1\}} \tilde{h}_{1(j)}^2 \tag{G.53}$$

with probability tending to 1. Here, we assume

$$\liminf_{d\to\infty,n\to\infty} P(\tilde{k}_{1\omega} \geq k_{1\omega} + r_1) > 0 \quad \text{and} \quad \liminf_{d\to\infty,n\to\infty} P(\tilde{k}_{1\omega} < k_{1\omega} + r_1) > 0$$

for the sake of simplicity. Then, if $\tilde{k}_{1\omega} \geq k_{1\omega} + r_1$, from (G.51) and (G.53), it holds that

$$\tilde{\boldsymbol{h}}_{1\omega}^T \boldsymbol{h}_{1\omega} = \sum_{s=1}^{k_{1\omega}} \tilde{h}_{1(s)} h_{1(s)} = \omega_1 + O_P(\omega_1/k_{1\omega} + \omega_1 n^{-1/2}). \tag{G.54}$$

Let $\mathcal{D}_\omega = \{j|\ \tilde{h}_{1\omega(j)} = 0 \text{ for } j = 1, ..., k_{1\omega} + r_1\}$ and $\mathcal{G}_\omega = \{j|\ \tilde{h}_{1\omega(j)} = 0 \text{ for } j = 1, ..., k_{1\omega}\}$. When $\tilde{k}_{1\omega} < k_{1\omega} + r_1$, from (G.53), we note that $\|\tilde{\boldsymbol{h}}_{1\omega}\|^2 + \sum_{s\in\mathcal{D}_\omega} \tilde{h}_{1(s)}^2 = \sum_{s=1}^{k_{1\omega}+r_1} \tilde{h}_{1(s)}^2$ with probability tending to 1. Then, if $\tilde{k}_{1\omega} < k_{1\omega} + r_1$, from (G.45), (G.48) and (G.52), we have that $\sum_{s\in\mathcal{G}_\omega} \tilde{h}_{1(s)}^2 \leq \sum_{s\in\mathcal{D}_\omega} \tilde{h}_{1(s)}^2 = \sum_{s=1}^{k_{1\omega}+r_1} \tilde{h}_{1(s)}^2 - \|\tilde{\boldsymbol{h}}_{1\omega}\|^2 = O_P(\omega_1/k_{1\omega} + \omega_1 n^{-1/2})$ and $\sum_{j\in\mathcal{G}_\omega} \tilde{h}_{1(j)} h_{1(j)} = \sum_{j\in\mathcal{G}_\omega} \tilde{h}_{1(j)}^2 \{1 + o_P(1)\}$. Thus, if $\tilde{k}_{1\omega} < k_{1\omega} + r_1$, from (G.51), it

holds that

$$\tilde{\boldsymbol{h}}_{1\omega}^T \boldsymbol{h}_{1\omega} = \sum_{s=1}^{k_{1\omega}} \tilde{h}_{1(s)} h_{1(s)} - \sum_{s \in \mathcal{G}_\omega} \tilde{h}_{1(s)} h_{1(s)}$$
$$= \omega_1 + O_P(\omega_1 / k_{1\omega} + \omega_1 n^{-1/2}). \tag{G.55}$$

If $P(\tilde{k}_{1\omega} < k_{1\omega} + r_1) = o(1)$ or $P(\tilde{k}_{1\omega} \geq k_{1\omega} + r_1) = o(1)$, we can obtain (G.54) or (G.55). Thus from (G.48), (G.54) and (G.55), we can conclude the result for $\tilde{\boldsymbol{h}}_{1\omega}$. As for $\tilde{\boldsymbol{h}}_{j\omega}$ with $j \geq 2$, we obtain the results similarly. It concludes the result of Theorem 2. □

Proofs of Corollary A.1 and Proposition A.1. For Corollary A.1, from Theorem 1 and (G.1), under the conditions in Corollary A.1, we have that for $j = 1, ..., m$, that $\|\tilde{\boldsymbol{\beta}}_j - \boldsymbol{\beta}_j\|^2 / \lambda_j = v_j \|\tilde{\boldsymbol{h}}_{j*}\|^2 + 1 - 2 v_j^{1/2} \tilde{\boldsymbol{h}}_{j*}^T \boldsymbol{h}_j + O_P(n^{-1})$ as $d \to \infty$ and $n \to \infty$. Then, by noting that $v_j^{1/2} = 1/2 + v_j/2 + O_P(n^{-1})$, from Theorem 1, we can conclude the result in Corollary A.1.

For Proposition A.1, from Proposition 2 and (G.1), under (A-i) and (C-i), it holds for $j = 1, ..., m$, that

$$\hat{\lambda}_j^{1/2} \hat{\boldsymbol{h}}_j^T \boldsymbol{h}_j = \tilde{\lambda}_j^{1/2} \tilde{\boldsymbol{h}}_j^T \boldsymbol{h}_j = \lambda_j^{1/2} \{ v_j^{1/2} + O_P(n^{-1}) \}$$
$$= \lambda_j^{1/2} \{ 1/2 + v_j/2 + O_P(n^{-1}) \}. \tag{G.56}$$

Then, from Proposition E.1, we can conclude the result in Proposition A.1. □

Proofs of Theorem A.1 and Proposition A.2. We first consider the proof of Theorem A.1. From Theorem 1, (G.1) and (G.15), under the conditions in Theorem A.1, we have that for $j = 1, ..., m$, that

$$\begin{aligned}\|\tilde{\boldsymbol{\beta}}_j\tilde{\boldsymbol{\beta}}_j^T - \boldsymbol{\beta}_j\boldsymbol{\beta}_j^T\|_F^2 =& \lambda_j^2\{1 + 2(v_j - 1) + O_P(n^{-1} + \eta_j)\} + \lambda_j^2 \\ & - 2\lambda_j^2\{1 + (v_j - 1) + O_P(n^{-1} + \eta_j)\} \\ =& O_P\{\lambda_j^2(n^{-1} + \eta_j)\} \qquad \text{(G.57)}\end{aligned}$$

as $d \to \infty$ and $n \to \infty$. Note that $|\mathrm{tr}\{(\tilde{\boldsymbol{\beta}}_j\tilde{\boldsymbol{\beta}}_j^T - \boldsymbol{\beta}_j\boldsymbol{\beta}_j^T)(\tilde{\boldsymbol{\beta}}_{j'}\tilde{\boldsymbol{\beta}}_{j'}^T - \boldsymbol{\beta}_{j'}\boldsymbol{\beta}_{j'}^T)\}| \leq \|\tilde{\boldsymbol{\beta}}_j\tilde{\boldsymbol{\beta}}_j^T - \boldsymbol{\beta}_j\boldsymbol{\beta}_j^T\|_F \cdot \|\tilde{\boldsymbol{\beta}}_{j'}\tilde{\boldsymbol{\beta}}_{j'}^T - \boldsymbol{\beta}_{j'}\boldsymbol{\beta}_{j'}^T\|_F \leq \|\tilde{\boldsymbol{\beta}}_j\tilde{\boldsymbol{\beta}}_j^T - \boldsymbol{\beta}_j\boldsymbol{\beta}_j^T\|_F^2 + \|\tilde{\boldsymbol{\beta}}_{j'}\tilde{\boldsymbol{\beta}}_{j'}^T - \boldsymbol{\beta}_{j'}\boldsymbol{\beta}_{j'}^T\|_F^2$ for $j \neq j'$. Then, it holds that

$$\|\widetilde{\boldsymbol{\Sigma}}_1 - \boldsymbol{\Sigma}_1\|_F^2 \leq m\sum_{s=1}^{m}\|\tilde{\boldsymbol{\beta}}_s\tilde{\boldsymbol{\beta}}_s^T - \boldsymbol{\beta}_s\boldsymbol{\beta}_s^T\|_F^2.$$

Thus from (G.57), we can conclude the result in Theorem A.1.

For Proposition A.2, from Proposition E.1 and (G.56), under (A-i) and (C-i), we have for $j = 1, ..., m$, that $\|\hat{\boldsymbol{\beta}}_j\hat{\boldsymbol{\beta}}_j^T - \boldsymbol{\beta}_j\boldsymbol{\beta}_j^T\|_F^2 = 2\{1 + o_P(1)\}\delta\lambda_j + \delta^2 + O_P(\lambda_j^2 n^{-1})$. Here, from (G.2) and (G.3), it holds that $\hat{\lambda}_j^{1/2}\hat{\boldsymbol{h}}_j^T\boldsymbol{h}_{j'} = \lambda_{j'}^{1/2}\boldsymbol{u}_{j'}^T\hat{\boldsymbol{u}}_j = O_P(\lambda_{j'}^{1/2}/n^{1/2})$. Thus, we can conclude the result in Proposition A.2. □

Proof of Proposition B.1. Assume (A-i), (C-i), (C-ii) and ($\star$'). By using Markov's inequality, for any $c > 0$, under (A-ii'), we have as $d \to \infty$ and

$n \to \infty$ that

$$\sum_{j'=1}^{d} P(|\bar{z}_{j(j'),x}| \geq cd^{1/8}/n^{1/2}) = \sum_{j'=1}^{d} P(|\bar{z}_{j(j'),x}|^8 \geq c^8 d/n^4)$$
$$\leq \sum_{j'=1}^{d} \frac{n^4 E(|\bar{z}_{j(j'),x}|^8)}{dc^8} = O(c^{-8}) \ \text{ for } j = 1, ..., m; \ \text{ and}$$
$$\sum_{j'=1}^{d} P(\|\boldsymbol{x}_{(j'),2}\| \geq cd^{1/8}) = \sum_{j'=1}^{d} P(\|\boldsymbol{x}_{(j'),2}\|^8 \geq c^8 d)$$
$$\leq \sum_{j'=1}^{d} \frac{E(\|\boldsymbol{x}_{(j'),2}\|^8)}{dc^8} = O(c^{-8}),$$

so that

$$\bar{z}_{j(j'),x} = O_P(d^{1/8}/n^{1/2}) \ \text{ for all } j, j'; \ \text{ and}$$
$$\|\boldsymbol{x}_{(j'),2}\| = O_P(d^{1/8}) \ \text{ for all } j'.$$

Then, from (G.1) and (G.9), it holds for $j = 1, ..., m$, that

$$\tilde{\lambda}_j^{-1/2} \boldsymbol{x}_{(j'),2}^T \hat{\boldsymbol{u}}_j = O_P\{(\lambda_j^{-1} n^{-1} d^{1/4})^{1/2}\} \ \text{ for all } j'.$$

Thus from (G.3), by combining (G.10) and (G.11), we have for $j = 1, ..., m$, that

$$\tilde{h}_{j(j')} = h_{j(j')} + O_P\{(\lambda_j^{-1} n^{-1} d^{1/4})^{1/2}\} \ \text{ for all } j'. \tag{G.58}$$

From (G.58), under (C-iii), it holds for $j = 1, ..., m$ and all j', that

$$\tilde{h}_{j(j')}^2 = h_{j(j')}^2 + o_P\Big(h_{oj(k_{j*})} \max\{|h_{oj(k_{j*})}|, |h_{j(j')}|\}\Big). \tag{G.59}$$

From (2.3), note that $\lambda_j^{-1} d^{1/4} = o(1)$ for $j = 1, ..., m$. Then, from (G.59), similar to the proofs of Theorem 1 and Corollary 1, we can obtain the

results in Theorem 1 and Corollary 1 after replacing (A-ii) and ($\star$) with (A-ii') and ($\star$'). By using the results in Theorem 1 and Corollary 1, similar to the proofs of Theorem A.1 and Corollary A.1, we can obtain the results in Theorem A.1 and Corollary A.1 after replacing (A-ii) and ($\star$) with (A-ii') and ($\star$'). For Theorem 2, from (G.58), similar to the proofs of Theorem 2, we can obtain the result after replacing (A-ii) and ($\star$) with (A-ii') and ($\star$'). $\square$

Bibliography

Aoshima, M. and Yata, K. (2018). Two-sample tests for high-dimension, strongly spiked eigenvalue models. *Statist. Sinica* **28**, 43-62.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society: Series B* **75**, 603-680.

Wainwright, M.J. (2019). High-Dimensional Statistics: A Non-Asymptotic Viewpoint. Cambridge University Press, Cambridge.

Yata, K. and Aoshima, M. (2012). Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations. *J. Multivariate Anal.* **105**, 193-215.

Zou, H. and Hastie, R. (2006). Sparse principal component analysis. *Journal of Computational and Graphical Statistics* **15**, 265-286.