Supplementary material for "Determining the number of canonical correlation pairs for high-dimensional vectors"

Jiasen Zheng*and Lixing Zhu†

Abstract

This supplementary material is to complete the mathematical techniques of the proof involved in the main text and additional numerical conclusions are presented for illustrating the theoretical results more explicitly.

*2017000815@ruc.edu.cn, School of Statistics, Renmin University of China, Beijing 100872, China.

†The corresponding author, lzhu@hkbu.edu.hk. The research described herewith was supported by a grant (HKBU12303419) from The University Grants Council of Hong Kong, and a grant from The National Natural Science Foundation of China (NSFC11671042). The authors thank Associate editor and two referees for their constructive suggestions and comments that led to the improvement of an early manuscript.

S1 Proofs of Theorems in Section 3

Proof of Theorem 1. Recall the definition of k in (2.3). From the construction of $\hat{k}_1$, to derive that $\mathbb{P}(\hat{k}_1 = k) \to 1$ as $n \to \infty$, all needed to do are to prove that, in probability,

$$\left| \frac{\hat{\lambda}_{k+1} - \hat{\lambda}_{k+2} + c_n}{\hat{\lambda}_k - \hat{\lambda}_{k+1} + c_n} \right| \le \tau_1,$$

and uniformly over all $i > k$,

$$\left| \frac{\hat{\lambda}_{i+1} - \hat{\lambda}_{i+2} + c_n}{\hat{\lambda}_i - \hat{\lambda}_{i+1} + c_n} \right| > \tau_1.$$

These inequalities show that by the definition, $\hat{k}_1$ is not possible to be smaller than k in probability. By Lemma 2 for any $k < i \le L-2$, $\hat{\lambda}_i - d_+ = O_p(n^{-2/3})$ and $\hat{\lambda}_i - \hat{\lambda}_{i+1} = O_p(n^{-2/3})$. Thus, by the rate of c_n with $n^{-2/3}c_n^{-1} \to 0$, we have that, in probability,

$$\left| \frac{\hat{\lambda}_{i+1} - \hat{\lambda}_{i+2} + c_n}{\hat{\lambda}_i - \hat{\lambda}_{i+1} + c_n} \right| = \left| \frac{(\hat{\lambda}_{i+1} - \hat{\lambda}_{i+2}) \cdot c_n^{-1} + 1}{(\hat{\lambda}_i - \hat{\lambda}_{i+1}) \cdot c_n^{-1} + 1} \right| = \frac{o_P(1) + 1}{o_P(1) + 1} \to 1 > \tau_1,$$

As L is a fixed integer, this convergence can then be uniformly for all i with $k < i \le L-2$. Consider in the case of $i = k$. Again by Lemma 2, $\hat{\lambda}_k - \gamma_k = O_p(n^{-1/2})$ and when $i = k, k+1$, $\hat{\lambda}_{i+1} - d_+ = O_p(n^{-2/3})$. In other words, $\hat{\lambda}_k - \hat{\lambda}_{k+1} \to \gamma_k - d_+$.

Notice, for $i \ge k$, $\gamma_i = \lambda_i(1 - c_1 + c_1\lambda_i{}^{-1})(1 - c_2 + c_2\lambda_i{}^{-1})$ and $d_+ = c_1(1-c_2) + c_2(1-c_1) + 2\sqrt{c_1c_2(1-c_1)(1-c_2)}$, then we have

$$\begin{aligned}
\gamma_i - d_+ &= \lambda_i[(1-c_1)(1-c_2) + c_2(1-c_1)\lambda_i^{-1} + c_1(1-c_2)\lambda_i^{-1} + c_1c_2\lambda_i{}^{-2}] \\
&\quad - [c_1(1-c_2) + c_2(1-c_1) + 2\sqrt{c_1c_2(1-c_1)(1-c_2)}] \\
&= \lambda_i(1-c_1)(1-c_2) + \frac{c_1c_2}{\lambda_i} - 2\sqrt{c_1c_2(1-c_1)(1-c_2)}.
\end{aligned}$$

From Assumption 1, we also have $c_1 > 0, c_2 > 0, 1 - c_1 > 0, 1 - c_2 > 0$, and Assumption 2 reveals $\lambda_i > 0$, hence,

$$\lambda_i(1-c_1)(1-c_2) + \frac{c_1c_2}{\lambda_i} \ge 2\sqrt{c_1c_2(1-c_1)(1-c_2)},$$

and the equality holds if and only if $\lambda_i = r_c = \sqrt{\frac{c_1c_2}{(1-c_1)(1-c_2)}}$. So, under the model in Assumption 2, if $\lambda_i > r_c$, then $\hat{\lambda}_k - \hat{\lambda}_{k+1} \to \gamma_k - d_+ > 0$.

Then,

$$\left|\frac{\hat{\lambda}_{k+1}-\hat{\lambda}_{k+2}+c_n}{\hat{\lambda}_k-\hat{\lambda}_{k+1}+c_n}\right| = \left|\frac{\hat{\lambda}_{k+1}-d_+-(\hat{\lambda}_{k+2}-d_+)+c_n}{\hat{\lambda}_k-\gamma_k+\gamma_k-d_+-(\hat{\lambda}_{k+1}-d_+)+c_n}\right| = \frac{o_p(c_n)+c_n}{\gamma_k-d_++o_p(1)+c_n} \to 0 < \tau_1.$$

The proof is completed. □

Proof of Theorem 2. Recall the definition of k in (2.3). From the construction of $\hat{k}_2$, to derive that $\mathbb{P}(\hat{k}_2 = k) \to 1$ as $n \to \infty$, all needed to do are to prove that, in probability,

$$\left|\frac{\hat{\lambda}_{k+1}-d_++c_n}{\hat{\lambda}_k-d_++c_n}\right| \le \tau_2,$$

and uniformly over all $i > k$,

$$\left|\frac{\hat{\lambda}_{i+1}-d_++c_n}{\hat{\lambda}_i-d_++c_n}\right| > \tau_2.$$

These inequalities show that by the definition, $\hat{k}_2$ is not possible to be smaller than k in probability. Note that for any $k < i \le L-1$, we have, in probability,

$$\left|\frac{\hat{\lambda}_{i+1}-d_++c_n}{\hat{\lambda}_i-d_++c_n}\right| = \left|\frac{(\hat{\lambda}_{i+1}-d_+)\cdot c_n^{-1}+1}{(\hat{\lambda}_i-d_+)\cdot c_n^{-1}+1}\right| = \frac{o_P(1)+1}{o_P(1)+1} \to 1 > \tau_2,$$

by applying Lemma 2 and $n^{-2/3}c_n^{-1} \to 0$. As L is a fixed integer, this convergence can then be uniformly for all i with $k < i \le L-1$. Further, when $i = k$, again by Lemma 2, $\hat{\lambda}_{k+1} - d_+ = O_p(n^{-2/3})$ and in probability $\hat{\lambda}_k - d_+ \to \gamma_k - d_+$.

Notice, for $i \ge k$, $\gamma_i = \lambda_i(1-c_1+c_1\lambda_i{}^{-1})(1-c_2+c_2\lambda_i{}^{-1})$ and $d_+ = c_1(1-c_2)+c_2(1-c_1)+2\sqrt{c_1c_2(1-c_1)(1-c_2)}$, then we have

$$\begin{aligned}\gamma_i - d_+ &= \lambda_i[(1-c_1)(1-c_2)+c_2(1-c_1)\lambda_i^{-1}+c_1(1-c_2)\lambda_i^{-1}+c_1c_2\lambda_i{}^{-2}] \\ &\quad - [c_1(1-c_2)+c_2(1-c_1)+2\sqrt{c_1c_2(1-c_1)(1-c_2)}] \\ &= \lambda_i(1-c_1)(1-c_2)+\frac{c_1c_2}{\lambda_i}-2\sqrt{c_1c_2(1-c_1)(1-c_2)}.\end{aligned}$$

From Assumption 1, we also have $c_1 > 0, c_2 > 0, 1-c_1 > 0, 1-c_2 > 0$, and Assumption 2 reveals $\lambda_i > 0$, hence,

$$\lambda_i(1-c_1)(1-c_2)+\frac{c_1c_2}{\lambda_i} \ge 2\sqrt{c_1c_2(1-c_1)(1-c_2)},$$

and the equality holds if and only if $\lambda_i = r_c = \sqrt{\frac{c_1c_2}{(1-c_1)(1-c_2)}}$. So, under the model in

Assumption 2, if $\lambda_i > r_c$, then $\hat{\lambda}_k - \hat{\lambda}_{k+1} \to \gamma_k - d_+ > 0$.

Thus,

$$\left|\frac{\hat{\lambda}_{k+1} - d_+ + c_n}{\hat{\lambda}_k - d_+ + c_n}\right| = \left|\frac{\hat{\lambda}_{k+1} - d_+ + c_n}{\hat{\lambda}_k - \gamma_k + \gamma_k - d_+ + c_n}\right| = \frac{o_p(c_n) + c_n}{\gamma_k - d_+ + o_p(1) + c_n} \to 0 < \tau_2.$$

The proof is completed. □

S2 Additional numerical results

Additional simulations are operated to further illustrate the performance for all methodologies mentioned in the main text.

Model S1: This model has higher eigenvalue multiplicity on the boundary: $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7) = (0.8, 0.6, 0.4, 0.2, 0.2, 0.2, 0.2)$, and $\lambda_8 = \cdots = \lambda_q = 0$. The results are presented in Table S1. As the multiplicity on the boundary is 4 and thus the true rank is much more difficult to detect. The results obviously suggest that all methods cannot work as well as them for the previous models. Further comparisons also indicate that AIC, BIC and C_p basically fail to work in rank determination and the scree plot method tends to have more serious under-estimation issue than our criteria. Again, the original eigenvalue-based ratio criterion works better than eigenvalue difference-based one. When $p = 210$, although $k = 7 = q_1$, it is not easy to be determined.

Table S1: The proportions of estimated rank in 1000 replications for Model S1.

	p=60, $r_c = 0.0445, \gamma_7 = 0.2778, d_+ = 0.1674$						p=110, $r_c = 0.0854, \gamma_7 = 0.3514, d_+ = 0.2956$					
	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_{BM}$	$\hat{k}_A$	$\hat{k}_B$	$\hat{k}_C$	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_{BM}$	$\hat{k}_A$	$\hat{k}_B$	$\hat{k}_C$
≤ 4	34	0	0	0	1000	0	467	0	0	0	1000	0
5	1	0	0	0	0	0	36	17	237	0	0	0
6	0	6	11	0	0	0	24	607	657	5	0	0
7	**922**	**994**	**989**	**904**	**0**	**651**	**395**	**376**	**106**	**703**	**0**	**4**
≥ 8	43	0	0	96	0	349	78	1	0	292	0	996
	p=160, $r_c = 0.1305, \gamma_7 = 0.4330, d_+ = 0.4133$						p=210, $r_c = 0.1809, \gamma_7 = 0.5226, d_+ = 0.5206$					
≤ 4	734	9	625	0	1000	0	942	769	1000	0	1000	0
5	68	401	361	0	0	0	23	227	0	0	0	0
6	55	582	14	17	0	0	13	4	0	17	0	0
7	**80**	**8**	**0**	**276**	**0**	**0**	**10**	**0**	**0**	**72**	**0**	**0**
≥ 8	63	0	0	607	0	1000	12	0	0	607	0	1000
	p=260, $r_c = 0.2379, \gamma_4 = 0.6644, d_+ = 0.6174$						p=310, $r_c = 0.2871, \gamma_4 = 0.7222, d_+ = 0.7037$					
≤ 4	949	809	1000	0	1000	0	986	924	1000	0	1000	0
5	21	189	0	0	0	0	8	75	0	0	0	0
6	13	2	0	0	0	0	1	1	0	0	0	0
7	**8**	**0**	**0**	**0**	**0**	**0**	**1**	**0**	**0**	**0**	**0**	**0**
≥ 8	9	0	0	1000	0	1000	4	0	0	1000	0	1000

As the boundary eigenvalues take small value $\lambda_4 = \cdots = \lambda_7 = 0.2$, with higher dimension, $k = 4 < q_1 = 7$, the situation gets worse. But when $p = 260$, γ_4 can still be larger than d_+ with a reasonable margin. $\hat{k}_2$ can still determine a value of $k = 5$ with higher probability than the others. Comparably, $\hat{k}_2$ still uniformly outperforms the other competitors with less serious underestimation problem.

Model S2: This model we let q_1 vary with a higher order of q to demonstrate the accuracy of estimates braced a number of nonzero eigenvalues. Specifically, let $Card(q_1) = \lfloor q^{2/3} \rfloor$, which is the largest integer less than or equal to $q^{2/3}$. In this setting, the population canonical pairs being more larger as the size of dimension diverse. Likewise, we use the vector $q_1 = (n_1, n_2, n_3, n_4)$ to present the multiplicity of nonzero eigenvalues of $0.8, 0.6, 0.4, 0.2$ respectively. The value of number $n_i, i = 1, 2, 3, 4$ is selected based on the size of q_1 which to be specified in Table S2. The information from Table S2 indicate that $\hat{k}_B$ and $\hat{k}_C$ almost bankrupt with the magnitude of $Card(q_1)$ increasing. The former forth methods seem display better performance when $Card(q_1) = 9$. To be surprised, as we increase the size of $Card(q_1)$, the AIC becomes outperform over others, but as continue larger $Card(q_1)$ it is tent to overestimate. This method is more sensitive than our criterion $\hat{k}_2$ and method $\hat{k}_{BM}$. Overall, all the methodology not get out the dilemma to avoid information loss under the high-dimensional setting, but $\hat{k}_2$ still can be viewed as a reliable criterion for statistical inference.

Table S2: The proportions of estimated rank in 1000 replications for Model S2.

	p=60, $Card(q_1) = 9$, $M_{q_1} = (2,2,3,2)$							p=110, $Card(q_1) = 14$, $M_{q_1} = (4,4,3,3)$					
	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_{BM}$	$\hat{k}_A$	$\hat{k}_B$	$\hat{k}_C$		$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_{BM}$	$\hat{k}_A$	$\hat{k}_B$	$\hat{k}_C$
≤ 7	5	0	0	0	1000	0	≤ 12	210	58	354	0	1000	0
8	0	4	5	0	0	0	13	12	913	586	0	0	0
9	**875**	**996**	**995**	**879**	**0**	**643**	14	**601**	**29**	**60**	**756**	**0**	**28**
10	71	0	0	119	0	331	15	85	0	0	237	0	309
≥ 11	49	0	0	2	0	26	≥ 16	92	0	0	7	0	663
	p=160, $Card(q_1) = 18$, $M_{q_1} = (5,5,5,3)$							p=210, $Card(q_1) = 22$, $M_{q_1} = (7,7,7,2)$					
≤ 15	442	297	866	0	1000	0	≤ 19	343	58	398	0	1000	0
16	67	661	134	0	0	0	20	11	754	543	0	0	0
17	92	42	0	17	0	0	21	295	188	59	0	0	0
18	**215**	**0**	**0**	**446**	**0**	**0**	22	**132**	**0**	**0**	**0**	**0**	**0**
≥ 19	184	0	0	537	0	1000	23	219	0	0	1000	0	1000
	p=260, $Card(q_1) = 25$, $M_{q_1} = (7,9,7,2)$							p=310, $Card(q_1) = 28$, $M_{q_1} = (8,8,8,4)$					
≤ 20	559	212	997	0	1000	0	≤ 18	807	15	999	0	1000	0
21	27	694	3	0	0	0	19	24	398	1	0	0	0
22	43	94	0	0	0	0	20-23	18	523	0	0	0	0
23	**177**	**0**	**0**	**0**	**0**	**0**	24	**58**	**36**	**0**	**0**	**0**	**0**
$24-25$	136	0	0	0	0	0	25-28	76	28	0	0	0	0
≥ 26	58	0	0	1000	0	1000	≥ 29	17	0	0	1000	0	1000

Model S3: This model has the eigenvalue multiplicity in the middle part as $(\lambda_1, \lambda_2, \lambda_3, \lambda_4, \lambda_5, \lambda_6, \lambda_7) = (0.8, 0.6, 0.6, 0.4, 0.4, 0.4, 0.2)$, $\lambda_8 = \cdots = \lambda_q = 0$. It is used to depict the consequence under asymmetric non-normal case. Specifically, we generate the deviants from chi-square distribution with two degrees of freedom. The results are tubulated in Table S3. Similarly, the information can be inferred from Table S3 in accordance with the symmetric Student's t distribution with ten degrees of freedom. $\hat{k}_A, \hat{k}_B$ and $\hat{k}_C$ cause deviation for estimation especially under high dimension. $\hat{k}_{BM}$ is to some extent credible, but its still lose information obviously with dimension diverse. Our criterion $\hat{k}_2$ can be viewed as a reliable rule for order determination under this symmetric and asymmetric situation. Though there is lack persuasive evidence in theoretical to support all methodologies for non-normal situation, the simulation results imply some heuristic sparks for the pursuit of truth.

Table S3: The proportions of estimated rank in 1000 replications for Model S3 under asymmetric $\chi^2_{df=2}$.

	p=60, $r_c = 0.0445, \gamma_7 = 0.2778, d_+ = 0.1674$						p=110, $r_c = 0.0854, \gamma_7 = 0.3514, d_+ = 0.2956$					
	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_{BM}$	$\hat{k}_A$	$\hat{k}_B$	$\hat{k}_C$	$\hat{k}_1$	$\hat{k}_2$	$\hat{k}_{BM}$	$\hat{k}_A$	$\hat{k}_B$	$\hat{k}_C$
≤ 4	0	0	0	0	0	0	0	0	0	0	1000	0
5	0	0	0	0	1000	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0
7	**1000**	**1000**	**1000**	**1000**	**0**	**1000**	**0**	**1000**	**1000**	**0**	**0**	**0**
≥ 8	0	0	0	0	0	0	1000	0	0	1000	0	1000
	p=160, $r_c = 0.1305, \gamma_7 = 0.4330, d_+ = 0.4133$						p=210, $r_c = 0.1809, \gamma_7 = 0.5226, d_+ = 0.5206$					
≤ 4	0	0	0	0	1000	0	0	0	0	0	1000	0
5	0	0	0	0	0	0	0	0	0	0	0	0
6	1000	0	1000	0	0	0	1000	1000	1000	0	0	0
7	**0**	**1000**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**
≥ 8	0	0	0	1000	0	1000	0	0	0	1000	0	1000
	p=260, $r_c = 0.2379, \gamma_6 = 0.6644, d_+ = 0.6174$						p=310, $r_c = 0.2871, \gamma_6 = 0.7222, d_+ = 0.7037$					
≤ 4	1000	0	0	0	1000	0	1000	0	1000	0	1000	0
5	0	0	1000	0	0	0	0	0	0	0	0	0
6	**0**	**1000**	**0**	**0**	**0**	**0**	**0**	**1000**	**0**	**0**	**0**	**0**
7	0	0	0	0	0	0	0	0	0	0	0	0
≥ 8	0	0	0	1000	0	1000	0	0	0	1000	0	1000

References

Bao, Z., Hu, J., Pan, G., and Zhou, W. (2019). "Canonical correlation coefficients of high-dimensional Gaussian vectors: Finite rank case." *The Annals of Statistics*, **47(1)**, 612–640.

Fujikoshi, Y. (2017b). "High-dimensional properties of AIC, BIC and C_p for estimation of dimensionality in canonical correlation analysis." *SUT journal of Mathematics*, **53(1)**, 59–72.

Headrick, T.C. (2002). "Fast fifth-order polynomial transforms for generating univariate and multivariate nonnormal distributions." *Computational Statistics and Data Analysis*, **40(4)**, 685–711.

Kendall, M. and Stuart, A. (1977). "*The advanced theory of statistics*, 4th edition," *Macmillan, New York.*

Jiasen Zheng
School of Statistics, Renmin University of China, Beijing 100872, China.
E-mail: 2017000815@ruc.edu.cn
Lixing Zhu
School of Statistics, Beijing Normal University, Beijing 100875, China. and
Department of Mathematics, Hong Kong Baptist University, Kowloon Tong, Hong Kong
E-mail: lzhu@hkbu.edu.cn