Response to the referee reports on the manuscript "Testing Simultaneous Diagonalizability"

Yuchen Xu, Marie-C. Düker, David S. Matteson

November 8, 2022

Thank you for the very careful reading of our manuscript, and for your helpful comments. Below you will find our responses to the report and the changes we made. Your comments are included in italics.

Response to Referee

This manuscript follows the work of Flury and others in estimation and testing of common eigenvalues. However, this manuscript considerably generalizes classical treatments of this problem. On the positive side, the manuscript includes: a diagonalization estimation procedure, a testing procedure for common a partially common diagonalization, a simulation study and a data application.

General comment: In response to the reviewer's comments, we added Appendix A: "Complementary simulation results". We included the following subsections

- A.1 Empirical type I and II errors,
- A.2 Sequential application of partial tests,
- A.3 High-dimensional data.

Appendix A entails all our responses to the reviewer's questions. We provide more details below where we reply to each individual comment. Note also that we attached Appendix A to our responds file right after our individual responses. We refer to Appendix A at several points in the main paper.

1. *Many modern problems of this sort have large d (in your notation) and small n. Can you comment on estimation / testing issues in this case or for cases where the asymptotic assumptions are suspect? Also, it might be worth varying d in the simulation studies.*

   Response: Appendix A.3: "High-dimensional data" provides a discussion on whether our methods are applicable for possibly high-dimensional data. In particular, we added Table 5 to illustrate for different but fixed sample sizes how the empirical size increases with the dimension. The illustration is based on the commutator based test but similar results are expected for all proposed test statistics.

In the main paper, we extended our conclusions (Section 8) to refer to the discussion in the supplementary material:

"In this work, we considered the classical "fixed d, large n" regime. However, many contemporary data go beyond the low dimensional setting and require the dimension d to be of the same order as, or possibly even larger than, the sample size n. While the high-dimensional setting goes beyond the scope of this work, we added a discussion and simulation study in Section A.3 of the supplementary material to emphasize that the methodology in this paper is not sufficient to do testing on high-dimensional data."

2. *On the simulation studies. In the appendix could formal type I, II tables for standard cutoffs be created. I understand this information is given in the manuscript. However, I think it would be useful (readable) in this restricted form.*

   Response: We provide tables with the type I and II errors for all proposed tests. The results are collected in Tables 1,2 and 3.

   We added the following sentence to the first paragraph in Section 6 in the main paper to refer to the additional simulation results:

   "Complementary to the plots presented in this section, we refer the reader to Section A.1 in the supplementary material for tables providing type I and II errors for all our simulation studies."

3. *Am I missing how the number of partial components is selected in the partial test? Is k assumed known? If I'm not mistaken, then can the simulation studies reflect the consequence of misspecifying k?*

   Response: The number of partial components is assumed to be known. However, one can apply our partial test sequentially. In a numerical study stated in Table 4, we illustrate that the method works well.

   We added the following paragraph at the end of Section 6.4 in the main paper to refer to the additional results:

   "As pointed out in Section 5, we assume that the number of partial common eigenvectors in known. Since this assumption is not feasible in practice, we propose a sequential testing procedure. We refer to Section A.2 in the supplementary material for a detailed description of the testing procedure and a corresponding simulation study to access its performance."

4. *Can some discussion be added on testing diagonality versus partial diagonality in practice? For example, how do the two tests relate when applied to the same dataset? Can you give guidance as to some sort of procedure, such as testing partial commonality then commonality? Can you discuss the mathematical relationships (if test A rejects then test B rejects ...)?*

   Response: Our proposed testing statistics guarantees that, if the number of common components equals the full rank, the partial test from Section 5 is equivalent to the multi-sample test from Section 4. Please, see our response to your previous comment that addresses the remaining concerns.

5. *Finally, can some discussion of how results change if the joint diagonalization estimation changes (since minimizing the sum of the off diagonal squares is just one of a few methods)?*

   Response: Our test design for the multi-sample test, especially the one in Corollary 4.1, is developed based on the squared off-diagonal norms, hence the objective function for optimization does matter to some extent. However, referring to André et al. [2020] and references therein, most of the existing joint diagonalization algorithms settle the cost functions, at least intermediately, to the similar formats despite their different novelties on estimation routines. Such formulation of minimizing the squared off-diagonal norms is usually preferred due to its outstanding computational efficiency as addressed earlier by Ziehe et al. [2004], while the potential weakness on trivial convergence (concerning orthogonal eigenvector matrix under symmetric cases) mentioned there is automatically mitigated under asymmetric setup due to invertibility.

Additional changes

Due to some new developments in the literature, we decided to update the algorithm to estimate the joint eigenvectors in the multi sample setting. To be more precise, we updated '*JDTE*' to '*(W)JDTE*' according to André et al. [2020] with the additional line search step for a weighted update per iteration to improve convergence performance.

These changes effect mostly our simulation results for Propositions 4.1 and 5.1. The improved algorithm let to a test size close to the nominal test level and a test power which is now similar to those of Corollary 4.2 and 5.2.

Appendix A

For completeness of our responds, we added Appendix A. It includes all the changes we made in responds to your report. In particular, it presents the additional simulation results.

A Complementary simulation results

We provide here some empirical results complementary to the numerical analysis presented in the main paper. Section A.1 gives empirical sizes and powers for the proposed test, Section A.2 studies sequential application of our partial tests and Section A.3 discusses application to possibly high-dimensional data.

A.1 Empirical type I and II errors

In addition to the p-values in the main paper, we provide here tables with type I and II errors for our tests to assess their performances. Tables A.1, A.2 and A.3 show respectively the errors for the two-sample, multi-sample and partial tests.

Test Type	Statistics Type	Sample Size	Type I Error	Type II Error	
				SNR=1000	SNR=10
Commutator-based test	Chi-test	50	0.218	0.216	0.000
		250	0.056	0.000	0.000
		1000	0.054	0.000	0.000
LLR test (Appendix B)	Oracle Chi-Test with (B.5)	50	0.182	0.000	0.000
		250	0.140	0.000	0.000
		1000	0.060	0.000	0.000
	Plugin Chi-test with (B.5)	50	1.000	0.000	0.000
		250	1.000	0.000	0.000
		1000	1.000	0.000	0.000
	Oracle Chi-Test with (B.6)	50	0.182	0.058	0.000
		250	0.140	0.000	0.000
		1000	0.060	0.000	0.000
	Plugin Chi-test with (B.6)	50	0.760	0.000	0.000
		250	0.842	0.000	0.000
		1000	0.774	0.000	0.000

Table A.1: Two-sample test results on simulated $\mathcal{M}_2(\rho, 5; 5)$ for $\rho^2 = \frac{1}{\text{SNR}} \in \{0, \frac{1}{1000}, \frac{1}{10}\}$.

Statistics Type	Sample Size	Type I Error	Type II Error		
			SNR = 1000	SNR = 100	SNR = 10
Oracle Chi-test (Proposition 4.1)	100	0.230	NA	NA	NA
	1000	0.060	NA	NA	NA
	10000	0.045	NA	NA	NA
	100000	0.070	NA	NA	NA
Plugin Chi-test (Proposition 4.2)	100	0.175	0.000	0.000	0.000
	1000	0.095	0.000	0.000	0.000
	10000	0.090	0.000	0.000	0.000
	100000	0.075	0.000	0.000	0.000
Plugin Gamma-test (Corollary 4.1)	100	0.015	0.005	0.000	0.000
	1000	0.025	0.000	0.000	0.000
	10000	0.005	0.000	0.000	0.000
	100000	0.015	0.000	0.000	0.000

Table A.2: Multi-sample test results on simulated $\mathcal{M}_8(\rho, 4; 4)$ for $\rho^2 = \frac{1}{\text{SNR}} \in \{0, \frac{1}{1000}, \frac{1}{100}\ \frac{1}{10}\}$.

Statistics Type	Sample Size	Type I Error	Type II Error		
			SNR = 1000	SNR = 100	SNR = 10
Chi-test (Proposition 5.2)	100	0.020	0.000	0.000	0.000
	1000	0.020	0.000	0.000	0.000
	10000	0.025	0.000	0.000	0.000
Gamma-test (Corollary 5.1)	100	0.010	0.000	0.000	0.000
	1000	0.015	0.000	0.000	0.000
	10000	0.015	0.000	0.000	0.000

Table A.3: Partial test results on simulated $\mathcal{M}_8(\rho, 2; 4)$ for $\rho = \frac{1}{\text{SNR}} \in \{0, \frac{1}{1000}, \frac{1}{100} \frac{1}{10}\}$.

A.2 Sequential application of partial tests

As pointed out in Section 5, we assume that the number of partial common eigenvectors in known. Since this assumption is not feasible in practice, we propose a sequential testing procedure. The hypothesis testing problem (2) can be stated for $k \in \{1, \ldots, d\}$. The sequential testing starts with $k = d$, then $k = d - 1$ and so on, till the null hypothesis is not rejected. The performance of this procedure is accessed through a simulation study in Table A.4.

Statistics Type	Sample Size	Rejection Rate		
		$k = 2$	$k = 3$	$k = 4$
Chi-test (Proposition 5.2)	100	0.020	1.000	1.000
	1000	0.020	1.000	1.000
	10000	0.025	1.000	1.000
Gamma-test (Corollary 5.1)	100	0.010	1.000	1.000
	1000	0.015	1.000	1.000
	10000	0.015	1.000	1.000

Table A.4: Partial test results on simulated $\mathcal{M}_8(0, 2; 4)$ and potentially mis-specified $k \in \{2, 3, 4\}$.

A.3 High-dimensional data

In this work, we consider the classical "fixed d, large n" regime. However, many contemporary data go beyond the low dimensional setting and require the dimension d to be of the same order as, or possibly even larger than, the sample size n. While the high-dimensional setting goes beyond the scope of this work, we would like to point out why our methodology is not sufficient to do testing on high-dimensional data.

Table A.5 presents the empirical rejection rates and average degrees of freedom for the two-sample test in Proposition 3.1, considering different sample sizes $n = 50, 100, 500$ and letting d grow. We present results assuming that the limiting covariance matrix in Equation 7 is estimated and known. The existence of a consistent estimator is stated in Assumption 2 and makes our procedure feasible in practice.

The results in Table A.5 show that the classical theory suffers a α test size much higher than the nominal test level once we consider high-dimensional data and estimate the limiting covariance matrix. Intuitively, the results are expected to break down once the sample size does not satisfy $n > r_1(d^2 + d^2)$. This can be easily seen by counting the degrees of freedom required to specify a rank-r_1 matrix of size $d^2 \times d^2$. Roughly speaking, we need r_1 numbers to specify the matrix's singular values, and $r_1 d^2$ and $r_1 d^2$ numbers to specify its left and right singular vectors.

The α test size much higher than the nominal test level is also due to Assumption 2 no longer being satisfied in a high-dimensional regime. In particular, the difference between estimator and true matrix is incorrectly normalized once the dimension grows with the sample size. It is expected to require results from random matrix theory to get convergence under suitable assumptions on the ratio between d and n.

Thanks again for all your suggestions and your careful reading of our manuscript.

References

Rémi André, Xavier Luciani, and Eric Moreau. Joint eigenvalue decomposition algorithms based on first-order Taylor expansion. *IEEE Transactions on Signal Processing*, 68:1716–1727, 2020. doi: 10.1109/TSP.2020.2976580.

Ziehe, Andreas, Laskov, Pavel, Nolte, Guido, ller, and KlausRobert. A fast algorithm for joint diagonalization with non-orthogonal transformations and its application to blind source separation. *Journal of Machine Learning Research*, 5(3):777–800, 2004.

d	Sample Size=50				Sample Size=100				Sample Size = 500			
	Empirical Cov		True Cov		Empirical Cov		True Cov		Empirical Cov		True Cov	
	Size	Avg DF	Size	Avg DF	Size	Avg DF	Size	Avg DF	Size	Avg DF	Size	Avg DF
2	0.036	2.05	0.022	2.04	0.054	2.00	0.058	2.00	0.044	2.00	0.044	2.00
3	0.066	6.10	0.034	6.13	0.074	6.02	0.054	6.02	0.062	6.00	0.060	6.00
4	0.088	12.33	0.014	12.38	0.092	12.06	0.036	12.06	0.066	12.00	0.054	12.00
5	0.186	20.15	0.018	20.42	0.116	20.05	0.050	20.08	0.062	20.00	0.056	20.00
6	0.358	29.75	0.020	30.65	0.176	29.97	0.020	30.09	0.058	30.00	0.046	30.00
7	0.698	40.85	0.018	42.79	0.310	40.46	0.026	40.97	0.092	40.00	0.050	40.00
8	0.960	55.82	0.024	57.78	0.548	56.04	0.034	56.30	0.092	56.00	0.044	56.00
9	1.000	69.11	0.014	74.19	0.778	70.54	0.044	71.97	0.098	70.44	0.052	70.57
10	1.000	86.27	0.020	92.94	0.962	89.62	0.034	90.74	0.162	90.00	0.054	90.00
11	1.000	96.93	0.024	114.55	0.996	108.25	0.030	111.15	0.238	108.79	0.050	109.15
12	1.000	98.00	0.030	136.88	1.000	131.96	0.042	133.48	0.340	132.00	0.056	132.00
13	1.000	98.00	0.022	161.63	1.000	153.87	0.030	157.93	0.468	156.00	0.044	156.00
14	1.000	98.00	0.028	188.91	1.000	174.85	0.026	184.56	0.622	182.00	0.044	182.00
15	1.000	98.00	0.040	217.63	1.000	193.89	0.042	212.96	0.774	209.88	0.052	209.99
16	0.998	98.00	0.042	248.27	1.000	198.00	0.034	243.25	0.874	239.89	0.042	240.00
17	0.998	98.00	0.050	281.03	1.000	198.00	0.042	275.62	0.974	270.00	0.046	270.14
18	0.978	98.00	0.050	315.85	1.000	198.00	0.036	310.37	0.986	303.18	0.056	304.74
19	0.994	98.00	0.046	352.90	1.000	198.00	0.050	347.06	0.998	340.60	0.044	341.50
20	0.994	98.00	0.056	391.32	1.000	198.00	0.046	384.37	1.000	373.18	0.044	375.67
21	0.460	98.00	0.054	433.27	1.000	198.00	0.040	426.84	1.000	418.44	0.044	419.91
22	0.982	98.00	0.044	475.39	1.000	198.00	0.040	468.52	1.000	455.78	0.054	459.21
23	0.986	98.00	0.044	520.58	1.000	198.00	0.054	513.53	1.000	499.67	0.042	503.17
24	0.284	98.00	0.032	567.74	1.000	198.00	0.030	560.31	1.000	544.63	0.040	551.13
25	0.702	98.00	0.056	617.31	1.000	198.00	0.046	609.14	1.000	590.03	0.054	595.95

Table A.5: Two-sample test results on simulated $\mathcal{M}_2(0, d; d)$ for dimensions $d \in \{2, \ldots, 25\}$.