Supplementary Materials for "Common Reducing Subspace Model and Network Alternation Analysis"

by Wenjing Wang, Xin Zhang and Lexin Li

A1 Additional Simulations

A1.1 Computation time

We report the computation time of our proposed method in covariance estimation. In addition, we report the running time of two variants of our method, as well as the methods of Franks and Hoff (2016) and Cook and Forzani (2008). We use the simulation examples in Section 5.1. Table S1 reports the average running time, in seconds, over 20 data replications. It is seen that our method runs reasonably fast. It is comparable in terms of running time to the method of Cook and Forzani (2008), and is about 10 to 50 fold faster than that of Franks and Hoff (2016). All simulations have been carried out on a x86 64-redhat-linux-gnu (64-bit) computing platform, with Intel Xeon CPU E5-2670 2.60GHz processors.

Table S1: Running time for covariance estimation. The average running time, in seconds, over 20 data replications for the covariance estimation examples in Section 5.1. CRS: the proposed method with the one-direction-at-a-time (1D) initialization plus the full Grassmann manifold optimization; CRS(1D): the proposed method with the 1D initialization only; CRS(rdm+FG): the proposed method with a random initialization plus the full Grassmann manifold optimization; FH16: the shared subspace model of Franks and Hoff (2016); CF08: the covariance reducing model of Cook and Forzani (2008).

	$p = 100, n_k = 40$	$p = 100, n_k = 100$	$p = 200, n_k = 40$	$p = 200, n_k = 100$	max SE
			Model I		
CRS	38.5	37.1	274	273	1.85
CRS(1D)	32.3	31.2	237	236	1.62
CRS(rdm+FG)	6.02	5.90	36.7	36.8	0.25
FH16	550	559	551	592	3.00
CF08	8.40	42.1	38.8	34.7	13.8
			Model II		
CRS	8.26	8.70	83.2	83.2	1.90
CRS(1D)	8.16	8.59	82.8	82.6	1.90
CRS(rdm+FG)	6.28	5.63	39.9	38.3	0.81
FH16	578	538	571	549	4.18
CF08	3.77	3.28	36.0	36.3	4.52
			Model III		
CRS	7.97	8.61	81.9	86.13	3.42
CRS(1D)	7.90	8.52	81.3	85.7	3.41
CRS(rdm+FG)	4.14	4.24	31.2	27.6	1.78
FH16	594	590	596	531	2.04
CF08	13.2	14.2	143	180	12.9

A1.2 Effect of initialization

Our method has employed the one-direction-at-a-time (1D) approach of Cook and Zhang (2016) for initialization, followed by a full Grassmann manifold optimization. Our initialization step produces a $\sqrt{n}$-consistent initial estimator, which is important for the subsequent manifold optimization. Based on our empirical experience, we have found that, after our initialization step, the manifold optimization usually converges after only a few iterations. To further evaluate the effect of the initialization, we consider two variants of our method. One is our method with the 1D initialization only, and the other is with a random initialization plus the full Grassmann manifold optimization. Table S1 reports the computation time, and Table S2 reports the corresponding estimation error. It is first observed in Table S1 that the computation time with and without the full Grassmann manifold optimization step after the 1D initialization is close. This suggests that the manifold optimization converges quickly after the 1D initialization. Moreover, it is seen in Table S2 that the method with a random initialization usually results in a much larger estimation error. This suggests the importance of a proper initialization.

A1.3 Number of groups

We further study the effect of the number of groups in the context of sparse precision matrix estimation. We adopt the simulation setup in Section 5.2 of the paper, and add the number of groups $K = 3$ and $K = 5$.

Specifically, when $K = 3$, our model is of the form,

$$\begin{aligned}
&\text{Model I:} && \boldsymbol{\Phi}_k = \boldsymbol{B}_k + \delta \boldsymbol{I}, \quad \text{where } \boldsymbol{B}_k = \boldsymbol{\Gamma}\boldsymbol{\Psi}_k\boldsymbol{\Gamma}^T, \delta = 0.5.\\
&\text{Model II:} && \boldsymbol{\Phi}_k = \boldsymbol{B}_k + \delta \boldsymbol{I}, \quad \text{where } \boldsymbol{B}_k = \sigma_k^2\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T, \sigma_1 = 1, \sigma_2 = 2, \sigma_3 = 3, \delta = 0.5.\\
&\text{Model III:} && \boldsymbol{\Phi}_k = \boldsymbol{B}_k + \delta \boldsymbol{I}, \quad \text{where } \boldsymbol{B}_k = \sigma_k\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T, \sigma_1 = -2, \sigma_2 = -3, \sigma_3 = -4, \delta = 5.
\end{aligned}$$

When $K = 5$, our model is of the form,

Table S2: Effect of initialization. The average estimation error over 100 data replications for the covariance estimation examples in Section 5.1. CRS: the proposed method with the one-direction-at-a-time (1D) initialization plus the full Grassmann manifold optimization; CRS(1D): the proposed method with the 1D initialization only; CRS(rdm+FG): the proposed method with a random initialization plus the full Grassmann manifold optimization.

	$p = 100, n_k = 40$	$p = 100, n_k = 100$	$p = 200, n_k = 40$	$p = 200, n_k = 100$	max SE
			Model I		
CRS	1·469	0·913	2·388	1·387	0·061
CRS(1D)	1·721	1.049	3.231	1.989	0.070
CRS(rdm+FG)	6·503	6.511	6.802	6.726	0.059
			Model II		
CRS	0·004	0·002	0·004	0·003	0·000
CRS(1D)	0.004	0.002	0.004	0.003	0.000
CRS(rdm+FG)	0.755	0.481	1.097	0.700	0.006
			Model III		
CRS	0·413	0·263	0·471	0·293	0·006
CRS(1D)	0.415	0.260	0.469	0.295	0.006
CRS(rdm+FG)	3.289	3.320	3.593	3.352	0.100

Table S3: Sparse precision matrix estimation when $K = 3$. Reported are the average estimation error, the sensitivity (%) and specificity (%) of nonzero elements selection, over 100 data replications. CRS: the proposed common reducing subspace model; FH16: the shared subspace model of Franks and Hoff (2016); CF08: the covariance reducing model of Cook and Forzani (2008); DWW14: the fused graphical Lasso estimator of Danaher et al. (2014); LL15: the joint prcesion estimator of Lee and Liu (2015); Sample: the sample mean estimator. For CRS, FH16, CF08 and Sample estimators, we further combine with a fast sparse estimation method of Liu and Luo (2015) to obtain a sparse precision matrix estimator. The last column, max SE: the maximum standard error. The left panel reports the estimation error. The right panel reports the sensitivity (the first number) and specificity (the second number).

	Estimation error					Selection specificity and sensitivity			
	$p = 100$ $n_k = 40$	$p = 100$ $n_k = 100$	$p = 200$ $n_k = 40$	$p = 200$ $n_k = 100$	max SE	$p = 100$ $n_k = 40$	$p = 100$ $n_k = 100$	$p = 200$ $n_k = 40$	$p = 200$ $n_k = 100$
			Model I					Model I	
CRS	0·647	0·361	1·120	0·625	0·009	90·0, 72·7	94·0, 72·3	75·7, 77·1	87·2, 74·0
FH16	2·943	2·941	2·955	2·955	0·000	4·76, 97·4	4·81, 97·4	2·48, 98·6	2·38, 98·8
CF08	3·070	3·001	3·115	3·122	0·022	88·4, 64·1	94·3, 57·1	70·6, 74·7	85·3, 67·3
DWW14	2·253	1·440	2·519	2·175	0·010	90·3, 49·1	93·8, 48·1	61·2, 76·5	80·8, 36·5
LL15	2.246	2.166	3.273	2.658	0.001	80.1,73.6	67.8, 97.6	69.3, 70.3	56.1, 97.5
Sample	1·509	1·045	2·402	1·791	0·004	82·4, 66·8	90·9, 60·2	49·9, 83·8	77·0, 69·7
			Model II					Model II	
CRS	2·498	1·665	4·202	2·848	0·037	98·2, 52·1	98·9, 51·0	96·6, 57·8	98·0, 55·9
FH16	35·74	35·74	35·76	35·76	0·001	6·18, 96·0	1·00, 91·2	2·95, 98·3	2·26, 99·0
CF08	3·059	2·972	3·119	3·087	0·019	97·9, 38·1	99·2, 27·6	94·8, 51·9	97·8, 41·2
DWW14	9·263	4·919	27·66	7·267	0·027	99·8, 3·05	99·7, 4·44	85·5, 15·3	99·7, 4·17
LL15	11·48	11·96	33·20	33·44	0·004	92·9, 75·1	92·1, 97·5	66·4, 98·6	58·3, 100·0
Sample	3·897	2·527	8·631	5·253	0·052	96·7, 42·1	98·1, 39·3	93·4, 50·4	94·6, 44·7
			Model III					Model III	
CRS	1·993	1·188	4·244	2·606	0·033	86·2, 74·3	91·5, 72·1	72·9, 79·2	82·7, 74·4
FH16	1·040	0·644	1·310	0·919	0·015	97·7, 79·1	98·0, 84·3	95·8, 83·1	96·8, 85·9
CF08	1·649	1·272	2·114	1·724	0·078	88·3, 70·4	93·0, 67·7	74·3, 76·2	84·8, 71·0
DWW14	5·753	4·586	17·58	6·343	0·090	84·5, 52·6	83·4, 70·1	88·2, 40·4	73·8, 70·9
LL15	10.57	6.107	10.10	9.265	0.002	64.4, 73.7	50.4, 97.4	50.5, 70.2	29.5, 97.3
Sample	11·56	8·104	17·78	13·59	0·028	81·5, 67·9	89·0, 61·3	62·3, 79·4	76·3, 71·2

Model I: $\boldsymbol{\Phi}_k = \boldsymbol{B}_k + \delta \boldsymbol{I}$, where $\boldsymbol{B}_k = \boldsymbol{\Gamma}\boldsymbol{\Psi}_k\boldsymbol{\Gamma}^T$, $\delta = 0.5$.

Model II: $\boldsymbol{\Phi}_k = \boldsymbol{B}_k + \delta \boldsymbol{I}$, where $\boldsymbol{B}_k = \sigma_k^2\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T$,
$\sigma_1 = 0.5, \sigma_2 = 1, \sigma_3 = 1.5, \sigma_4 = 2, \sigma_5 = 3, \delta = 0.5$.

Model III: $\boldsymbol{\Phi}_k = \boldsymbol{B}_k + \delta \boldsymbol{I}$, where $\boldsymbol{B}_k = \sigma_k\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T$,
$\sigma_1 = -0.5, \sigma_2 = -1, \sigma_3 = -2, \sigma_4 = -3, \sigma_5 = -4, \delta = 5$.

When $K > 2$, we define the estimation error as,

$$\sum_{k=2}^{K}\sum_{j=1}^{k-1} \|\boldsymbol{\Phi}_k - \boldsymbol{\Phi}_j - (\widehat{\boldsymbol{\Phi}}_k - \widehat{\boldsymbol{\Phi}}_j)\|_F$$

Let $\boldsymbol{\Psi}_s = \boldsymbol{\Phi}_k - \boldsymbol{\Phi}_j$, $k = 2, \ldots, K, j = 1, \ldots, k-1, s = 1, \ldots, m = \binom{K}{2}$, then we define the specificity and sensitivity as,

Table S4: Sparse precision matrix estimation when $K = 5$. The rest of setup is the same as in Table S3.

	Estimation error					Selection specificity and sensitivity			
	$p = 100$ $n_k = 40$	$p = 100$ $n_k = 100$	$p = 200$ $n_k = 40$	$p = 200$ $n_k = 100$	max SE	$p = 100$ $n_k = 40$	$p = 100$ $n_k = 100$	$p = 200$ $n_k = 40$	$p = 200$ $n_k = 100$
			Model I					Model I	
CRS	2·172	1·440	3·286	2·094	0·165	93·7, 72·2	96·3, 71·0	85·6, 74·8	91·7, 73·5
FH16	14·75	14·75	14·79	14·79	0·002	4·93, 96·9	5·40, 96·8	2·53, 98·6	2·50, 98·7
CF08	10·82	9·394	11·69	10·81	0·234	93·0, 65·4	96·4, 62·5	84·4, 70·7	91·8, 66·2
DWW14	11·29	11·01	12·91	14·83	0·148	78·0, 56·5	75·6, 78·8	35·9, 78·3	15·8, 93·8
LL15	10·51	9·686	14·92	14·03	0·075	82·2, 56·6	75·9, 88·2	21·1, 96·9	45·2, 94·8
Sample	5·958	4·032	10·35	7·062	0·519	84·6, 65·5	91·9, 59·2	56·0, 81·4	80·4, 67·0
			Model II					Model II	
CRS	6·199	4·087	9·571	6·710	0·089	96·8, 58·2	98·1, 57·2	93·2, 61·7	97·4, 62·3
FH16	91·51	91·51	91·59	91·60	0·002	6·91, 95·0	8·09, 93·4	3·07, 98·0	3·22, 97·8
CF08	76·24	70·05	79·51	77·44	0·815	98·1, 39·8	99·0, 38·6	94·3, 50·0	97·4, 42·2
DWW14	79·12	79·45	85·03	84·92	0·058	82·9, 54·6	78·8, 79·2	71·9, 61·8	69·9, 80·0
LL15	51·45	39·71	85·53	85·41	0·223	83·2, 94·1	88·6, 95·5	53·2, 96·3	45·7, 99·9
Sample	11·29	7·347	25·60	15·34	0·116	92·8, 54·6	95·6, 50·7	82·8, 63·7	91·4, 57·6
			Model III					Model III	
CRS	9·104	5·334	18·21	11·26	0·055	83·4, 76·4	89·5, 74·5	67·9, 81·1	78·6, 78·3
FH16	5·865	3·210	6·464	3·885	0·040	88·4, 89·1	85·2, 94·2	83·2, 93·3	65·7, 97·8
CF08	17·67	13·01	25·10	18·92	0·575	76·7, 74·0	78·9, 73·5	62·6, 78·5	60·0, 81·4
DWW14	31·35	30·82	37·14	37·66	0·159	41·2, 99·5	41·5, 99·6	19·0, 99·7	20·1, 99·8
LL15	32·32	31·56	42·08	39·78	0·056	44·2, 97·0	42·9, 99·8	272, 97·1	26·0, 99·9
Sample	32·29	23·05	48·27	36·60	0·058	66·6, 78·4	78·3, 72·5	45·5, 87·1	46·4, 83·6

$$\text{specificity} = \frac{1}{m}\sum_{s=1}^{m}\frac{\#(\boldsymbol{\Psi}_s \neq 0, \widehat{\boldsymbol{\Psi}}_s \neq 0)}{\#(\boldsymbol{\Psi}_s \neq 0)}$$

$$\text{sensitivity} = \frac{1}{m}\sum_{s=1}^{m}\frac{\#(\boldsymbol{\Psi}_s = 0, \widehat{\boldsymbol{\Psi}}_s = 0)}{\#(\boldsymbol{\Psi}_s = 0)}$$

We report the simulation results over 100 data replications in Tables S3 and S4. It is seen that, for $K > 2$, the performance of our method remains competitive. Moreover, the estimation error of all the methods increases as K increases, since it is defined as the sum of K-choose-2 pairs of differences and a larger K implies more difference terms involved. On the other hand, the selection accuracy in terms of sensitivity and specificity remain about the same as K increases.

A2 Additional Theory

A2.1 Connection with envelope

We present an additional theorem that connects our common reducing subspace model with the notion of the envelope and the central subspace in the covariance reducing model of Cook and Forzani (2008).

Theorem S1. *The following statements are true.*

(i) The smallest common reducing subspace that satisfies (1) *is* $\mathcal{E}_{\boldsymbol{\Delta}}(\boldsymbol{\Upsilon})$.

(ii) $\mathcal{E}_{\boldsymbol{\Delta}}(\boldsymbol{\Upsilon}) = \mathcal{E}_{\boldsymbol{\Delta}}(\mathcal{C}) = \mathcal{E}_{\boldsymbol{\Delta}}(\mathcal{P})$.

(iii) For all $k = 1, \ldots, K$, $\mathcal{E}_{\boldsymbol{\Delta}}(\mathcal{C}) = \mathcal{E}_{\boldsymbol{\Delta}_k}(\mathcal{C})$ and $\mathcal{E}_{\boldsymbol{\Delta}}(\mathcal{P}) = \mathcal{E}_{\boldsymbol{\Delta}_k}(\mathcal{P})$.

Proof. (i) First, we show that, if $\mathrm{span}(\boldsymbol{\Gamma}) = \mathcal{E}_{\boldsymbol{\Delta}}(\boldsymbol{\Upsilon})$, then it satisfies our common reducing subspace model (1); that is,

$$\boldsymbol{\Delta}_k = \boldsymbol{\Gamma}\boldsymbol{\Omega}_k\boldsymbol{\Gamma}^{\mathrm{T}} + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}}, \quad k = 1, \ldots, K,$$

for some symmetric positive definite matrices $\boldsymbol{\Omega}_k$ and $\boldsymbol{\Omega}_0$. We achieve this by verifying: (a) $\boldsymbol{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{Q}_{\boldsymbol{\Gamma}} = 0$, and (b) $\boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{Q}_{\boldsymbol{\Gamma}} = \boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}\boldsymbol{Q}_{\boldsymbol{\Gamma}}$. Recall that $\boldsymbol{P}_{\boldsymbol{\Upsilon}(\boldsymbol{\Delta})} = \boldsymbol{\Upsilon}(\boldsymbol{\Upsilon}^{\mathrm{T}}\boldsymbol{\Delta}\boldsymbol{\Upsilon})^{-1}\boldsymbol{\Upsilon}^{\mathrm{T}}\boldsymbol{\Delta}$ and

$$\boldsymbol{P}_{\boldsymbol{\Upsilon}(\boldsymbol{\Delta})}\boldsymbol{Q}_{\boldsymbol{\Gamma}} \;=\; \boldsymbol{\Upsilon}(\boldsymbol{\Upsilon}^{\mathrm{T}}\boldsymbol{\Delta}\boldsymbol{\Upsilon})^{-1}\boldsymbol{\Upsilon}^{\mathrm{T}}\boldsymbol{\Delta}\boldsymbol{Q}_{\boldsymbol{\Gamma}} = \boldsymbol{\Upsilon}(\boldsymbol{\Upsilon}^{\mathrm{T}}\boldsymbol{\Delta}\boldsymbol{\Upsilon})^{-1}\boldsymbol{\Upsilon}^{\mathrm{T}}\boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}\boldsymbol{Q}_{\boldsymbol{\Gamma}} = 0,$$

where the equality $\boldsymbol{\Delta}\boldsymbol{Q}_{\boldsymbol{\Gamma}} = \boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}\boldsymbol{Q}_{\boldsymbol{\Gamma}}$ is due to the fact that $\mathrm{span}(\boldsymbol{\Gamma})$ reduces $\boldsymbol{\Delta}$, and the last equality is due to that $\boldsymbol{\Upsilon}^{\mathrm{T}}\boldsymbol{Q}_{\boldsymbol{\Gamma}} = 0$ as $\mathrm{span}(\boldsymbol{\Upsilon}) \subseteq \mathrm{span}(\boldsymbol{\Gamma})$. Then by (5), we have

$$\boldsymbol{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{Q}_{\boldsymbol{\Gamma}} \;=\; \boldsymbol{P}_{\boldsymbol{\Gamma}}\{\boldsymbol{\Delta} + \boldsymbol{P}_{\boldsymbol{\Upsilon}(\boldsymbol{\Delta})}^{\mathrm{T}}(\boldsymbol{\Delta}_k - \boldsymbol{\Delta})\boldsymbol{P}_{\boldsymbol{\Upsilon}(\boldsymbol{\Delta})}\}\boldsymbol{Q}_{\boldsymbol{\Gamma}} = \boldsymbol{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}\boldsymbol{Q}_{\boldsymbol{\Gamma}} = 0,$$

where the last equality again follows from the fact that $\mathrm{span}(\boldsymbol{\Gamma})$ is a reducing subspace of $\boldsymbol{\Delta}$. Similarly, we have

$$\boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{Q}_{\boldsymbol{\Gamma}} \;=\; \boldsymbol{Q}_{\boldsymbol{\Gamma}}\{\boldsymbol{\Delta} + \boldsymbol{P}_{\boldsymbol{\Upsilon}(\boldsymbol{\Delta})}^{\mathrm{T}}(\boldsymbol{\Delta}_k - \boldsymbol{\Delta})\boldsymbol{P}_{\boldsymbol{\Upsilon}(\boldsymbol{\Delta})}\}\boldsymbol{Q}_{\boldsymbol{\Gamma}} = \boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}\boldsymbol{Q}_{\boldsymbol{\Gamma}}.$$

Second, we show that $\mathcal{E}_{\boldsymbol{\Delta}}(\boldsymbol{\Upsilon})$ is the smallest subspace that satisfies model (1). We prove by contradiction. Suppose there is a subspace $\mathrm{span}(\boldsymbol{G})$ that satisfies,

$$\boldsymbol{\Delta}_k = \boldsymbol{G}\boldsymbol{\Phi}_k\boldsymbol{G}^{\mathrm{T}} + \boldsymbol{G}_0\boldsymbol{\Phi}_0\boldsymbol{G}_0^{\mathrm{T}}, \quad k = 1, \ldots, K,$$

for some symmetric positive definite matrices $\boldsymbol{\Phi}_k$ and $\boldsymbol{\Phi}_0$. Then by the definition of $\boldsymbol{\Delta} = \sum_k \pi_k\boldsymbol{\Delta}_k$,

$$\boldsymbol{\Delta} = \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{G}^{\mathrm{T}} + \boldsymbol{G}_0\boldsymbol{\Phi}_0\boldsymbol{G}_0^{\mathrm{T}}, \quad \boldsymbol{\Phi} = \sum_k \pi_k\boldsymbol{\Phi}_k,$$

which implies that $\mathrm{span}(\boldsymbol{G})$ is a reducing subspace of $\boldsymbol{\Delta}$. Furthermore, we can now write $\mathrm{span}(\boldsymbol{\Delta}_k - \boldsymbol{\Delta}) = \mathrm{span}(\boldsymbol{G}\boldsymbol{\Phi}_k\boldsymbol{G}^{\mathrm{T}} - \boldsymbol{G}\boldsymbol{\Phi}\boldsymbol{G}^{\mathrm{T}}) \subseteq \mathrm{span}(\boldsymbol{G})$, and $\boldsymbol{P}_{\boldsymbol{G}} = \boldsymbol{P}_{\boldsymbol{G}(\boldsymbol{\Delta})}$. It is thus straightforward to verify that $\mathrm{span}(\boldsymbol{G})$ satisfies the sufficient dimension reduction condition (5). Henceforth, $\mathrm{span}(\boldsymbol{\Upsilon}) \subseteq \mathrm{span}(\boldsymbol{G})$. Together, we have shown that $\mathrm{span}(\boldsymbol{G})$ is a reducing subspace of $\boldsymbol{\Delta}$ that contains $\mathrm{span}(\boldsymbol{\Upsilon})$. By the definition of the envelope, $\mathcal{E}_{\boldsymbol{\Delta}}(\boldsymbol{\Upsilon})$ is the smallest reducing subspace of $\boldsymbol{\Delta}$ that contains $\mathrm{span}(\boldsymbol{\Upsilon})$. This proves the first statement (i).

(ii) First, we show that $\mathcal{E}_{\boldsymbol{\Delta}}(\mathcal{C}) \subseteq \mathcal{E}_{\boldsymbol{\Delta}}(\boldsymbol{\Upsilon})$. Let $\mathrm{span}(\boldsymbol{\Gamma}) = \mathcal{E}_{\boldsymbol{\Delta}}(\boldsymbol{\Upsilon})$. Then by Theorem S1 (i), we have $\boldsymbol{\Delta}_k = \boldsymbol{\Gamma}\boldsymbol{\Omega}_k\boldsymbol{\Gamma}^{\mathrm{T}} + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}}$. Henceforth, $\mathcal{C} \subseteq \mathrm{span}(\boldsymbol{\Gamma})$, and $\mathcal{E}_{\boldsymbol{\Delta}}(\mathcal{C}) \subseteq \mathrm{span}(\boldsymbol{\Gamma})$, because $\mathrm{span}(\boldsymbol{\Gamma})$ is a reducing subspace of $\boldsymbol{\Delta}$.

Second, we show that $\mathcal{E}_{\boldsymbol{\Delta}}(\boldsymbol{\Upsilon}) \subseteq \mathcal{E}_{\boldsymbol{\Delta}}(\mathcal{C})$. Let $\mathrm{span}(\boldsymbol{\Gamma}) = \mathcal{E}_{\boldsymbol{\Delta}}(\mathcal{C})$. Then we can write $\boldsymbol{\Delta} = \boldsymbol{\Gamma}\boldsymbol{\Omega}\boldsymbol{\Gamma}^{\mathrm{T}} + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}}$ for some $\boldsymbol{\Omega}$ and $\boldsymbol{\Omega}_0$, because $\mathrm{span}(\boldsymbol{\Gamma})$ is a reducing subspace of $\boldsymbol{\Delta}$. Since $\mathcal{C} \subseteq \mathrm{span}(\boldsymbol{\Gamma})$, we have that $\mathrm{span}(\boldsymbol{\Delta}_k - \boldsymbol{\Delta}) \subseteq \mathcal{C} \subseteq \mathrm{span}(\boldsymbol{\Gamma})$. Therefore, we can write $\boldsymbol{\Delta}_k = \boldsymbol{\Delta} + \boldsymbol{\Gamma}\boldsymbol{\Phi}_k\boldsymbol{\Gamma}^{\mathrm{T}}$ for some symmetric matrix $\boldsymbol{\Phi}_k$. It follows that $\boldsymbol{\Delta}_k = \boldsymbol{\Gamma}(\boldsymbol{\Omega} + \boldsymbol{\Phi}_k)\boldsymbol{\Gamma}^{\mathrm{T}} + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}}$, and henceforth $\mathrm{span}(\boldsymbol{\Gamma})$ is a common reducing subspace. The conclusion follows from the fact that $\mathcal{E}_{\boldsymbol{\Delta}}(\boldsymbol{\Upsilon})$ is the smallest common reducing subspace.

Combining the above two steps, we have $\mathcal{E}_{\boldsymbol{\Delta}}(\boldsymbol{\Upsilon}) = \mathcal{E}_{\boldsymbol{\Delta}}(\mathcal{C})$. The proof for $\mathcal{E}_{\boldsymbol{\Delta}}(\boldsymbol{\Upsilon}) = \mathcal{E}_{\boldsymbol{\Delta}}(\mathcal{P})$ follows a similar argument, and is omitted. This proves the second statement (ii).

(iii) We show that, for all $k = 1, \ldots, K$, $\mathcal{E}_{\boldsymbol{\Delta}}(\mathcal{C}) = \mathcal{E}_{\boldsymbol{\Delta}_k}(\mathcal{C})$. It follows from a similar argument that $\mathcal{E}_{\boldsymbol{\Delta}}(\mathcal{P}) = \mathcal{E}_{\boldsymbol{\Delta}_k}(\mathcal{P})$.

We prove the equality $\mathcal{E}_{\boldsymbol{\Delta}}(\mathcal{C}) = \mathcal{E}_{\boldsymbol{\Delta}_k}(\mathcal{C})$, by first assuming $\mathrm{span}(\mathcal{C}) \subseteq \mathrm{span}(\boldsymbol{\Gamma})$, and showing that: (a) if in addition $\mathrm{span}(\boldsymbol{\Gamma})$ is a reducing subspace of $\boldsymbol{\Delta}$, then it is also a reducing subspace of $\boldsymbol{\Delta}_k$, and (b) if

$\mathrm{span}(\boldsymbol{\Gamma})$ is a reducing subspace of $\boldsymbol{\Delta}$, then it is also a reducing subspace of $\boldsymbol{\Delta}_k$. Therefore, (a) implies that $\mathcal{E}_{\boldsymbol{\Delta}}(\mathcal{C}) \subseteq \mathcal{E}_{\boldsymbol{\Delta}_k}(\mathcal{C})$, and (b) implies that $\mathcal{E}_{\boldsymbol{\Delta}_k}(\mathcal{C}) \subseteq \mathcal{E}_{\boldsymbol{\Delta}}(\mathcal{C})$.

If $\mathrm{span}(\mathcal{C}) \subseteq \mathrm{span}(\boldsymbol{\Gamma})$, then $\boldsymbol{\Delta}_k - \boldsymbol{\Delta} = \boldsymbol{P}_{\boldsymbol{\Gamma}}^{\mathrm{T}}(\boldsymbol{\Delta}_k - \boldsymbol{\Delta})\boldsymbol{P}_{\boldsymbol{\Gamma}}$. If $\mathrm{span}(\boldsymbol{\Gamma})$ is a reducing subspace of $\boldsymbol{\Delta}$, then we can write

$$\boldsymbol{\Delta}_k \;=\; \boldsymbol{\Delta} + \boldsymbol{P}_{\boldsymbol{\Gamma}}(\boldsymbol{\Delta}_k - \boldsymbol{\Delta})\boldsymbol{P}_{\boldsymbol{\Gamma}} = \boldsymbol{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}\boldsymbol{P}_{\boldsymbol{\Gamma}} + \boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}\boldsymbol{Q}_{\boldsymbol{\Gamma}} + \boldsymbol{P}_{\boldsymbol{\Gamma}}(\boldsymbol{\Delta}_k - \boldsymbol{\Delta})\boldsymbol{P}_{\boldsymbol{\Gamma}} = \boldsymbol{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{P}_{\boldsymbol{\Gamma}} + \boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}\boldsymbol{Q}_{\boldsymbol{\Gamma}},$$

which implies that $\mathrm{span}(\boldsymbol{\Gamma})$ is also a reducing subspace of $\boldsymbol{\Delta}_k$. Similarly, by symmetry, we can show if $\mathrm{span}(\boldsymbol{\Gamma})$ is a reducing subspace of $\boldsymbol{\Delta}_k$ then it is also a reducing subspace of $\boldsymbol{\Delta}$. This proves the third statement (iii), and completes the proof of Theorem S1. □

A2.2 Theory with a fixed sample size

We also study the theoretical properties of our estimator when the sample size $n = \sum_{k=1}^{K} n_k$ is fixed. For simplicity, we assume that the common reducing subspace $\mathrm{span}(\boldsymbol{\Gamma})$ is known. We note that it is very difficult to fully characterize the estimation variation in the estimated basis matrix $\widehat{\boldsymbol{\Gamma}}$, and hence difficult to provide a full theoretical analysis under the fixed sample size. Nevertheless, we find that, even the result under the reduced setting with a known $\mathrm{span}(\boldsymbol{\Gamma})$ offers some useful understanding of our method.

Specifically, we define $\boldsymbol{I}_{(p,q)}$ as the vec-permutation matrix such that $\mathrm{vec}(\boldsymbol{A}_{p\times q}) = \boldsymbol{I}_{(p,q)}\mathrm{vec}(\boldsymbol{A}_{p\times q}^{\mathrm{T}})$ for any $p \times q$ matrix $\boldsymbol{A}_{p\times q}$ (Henderson and Searle, 1979). An important property of $\boldsymbol{I}_{(p,p)}$ is that $\boldsymbol{I}_{(p,p)}(\boldsymbol{B}_{p\times p} \otimes \boldsymbol{A}_{p\times p}) = (\boldsymbol{A}_{p\times p} \otimes \boldsymbol{B}_{p\times p})\boldsymbol{I}_{(p,p)}$. We have the following result for a fixed sample size n.

Proposition S1. *Assume that the observed data matrices $\boldsymbol{D}_{i,k}$, $i = 1, \ldots, n_k$, $k = 1, \ldots, K$, are i.i.d. samples from a Wishart distribution $W_p(\boldsymbol{\Delta}_k/d, d)$ with d degrees of freedom, and that the basis matrix $\boldsymbol{\Gamma} \in \mathbb{R}^{p\times u}$ is known. Then the variance reduction of the proposed estimator over the sample covariance estimator is $0 \le \mathrm{cov}\{\mathrm{vec}(\overline{\boldsymbol{D}}_k)\} - \mathrm{cov}\{\mathrm{vec}(\widehat{\boldsymbol{\Delta}}_k)\} = \boldsymbol{V}_1 + \boldsymbol{V}_2 + \boldsymbol{V}_3$, where*

$$\begin{aligned}
\boldsymbol{V}_1 &= \frac{n - n_k}{nn_kd}\left\{(\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}} \otimes (\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}})\right\}\left\{\boldsymbol{I}_{p^2} + \boldsymbol{I}_{(p,p)}\right\},\\
\boldsymbol{V}_2 &= \frac{1}{n_kd}\left\{(\boldsymbol{\Gamma}\boldsymbol{\Omega}_k\boldsymbol{\Gamma}^{\mathrm{T}}) \otimes (\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}})\right\}\left\{\boldsymbol{I}_{p^2} + \boldsymbol{I}_{(p,p)}\right\},\\
\boldsymbol{V}_3 &= \frac{1}{n_kd}\left\{(\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}}) \otimes (\boldsymbol{\Gamma}\boldsymbol{\Omega}_k\boldsymbol{\Gamma}^{\mathrm{T}})\right\}\left\{\boldsymbol{I}_{p^2} + \boldsymbol{I}_{(p,p)}\right\}.
\end{aligned}$$

This proposition offers some additional insight on the efficiency gain of our estimator, while it does not require n_k to diverge to infinity. Specifically, our proposed estimator based on the common reducing subspace model is of the form, $\widehat{\boldsymbol{\Delta}}_k = \boldsymbol{P}_{\boldsymbol{\Gamma}}\overline{\boldsymbol{D}}_k\boldsymbol{P}_{\boldsymbol{\Gamma}} + \boldsymbol{Q}_{\boldsymbol{\Gamma}}\overline{\boldsymbol{D}}\boldsymbol{Q}_{\boldsymbol{\Gamma}}$. The sample covariance estimator $\overline{\boldsymbol{D}}_k$ can be written as $\overline{\boldsymbol{D}}_k = \boldsymbol{P}_{\boldsymbol{\Gamma}}\overline{\boldsymbol{D}}_k\boldsymbol{P}_{\boldsymbol{\Gamma}} + \boldsymbol{Q}_{\boldsymbol{\Gamma}}\overline{\boldsymbol{D}}_k\boldsymbol{Q}_{\boldsymbol{\Gamma}} + \boldsymbol{Q}_{\boldsymbol{\Gamma}}\overline{\boldsymbol{D}}_k\boldsymbol{P}_{\boldsymbol{\Gamma}} + \boldsymbol{P}_{\boldsymbol{\Gamma}}\overline{\boldsymbol{D}}_k\boldsymbol{Q}_{\boldsymbol{\Gamma}}$. Then the efficiency gain of $\widehat{\boldsymbol{\Delta}}_k$ over $\overline{\boldsymbol{D}}_k$, under the fixed n_k, includes two parts. The first part, quantified as $\boldsymbol{V}_1$, comes from pooling the shared information across groups. That is, we estimate $\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0$ by $\boldsymbol{Q}_{\boldsymbol{\Gamma}}\overline{\boldsymbol{D}}\boldsymbol{Q}_{\boldsymbol{\Gamma}}$, instead of $\boldsymbol{Q}_{\boldsymbol{\Gamma}}\overline{\boldsymbol{D}}_k\boldsymbol{Q}_{\boldsymbol{\Gamma}}$. Interestingly, if we have $n_1 = \cdots = n_K = n/K$, then the factors $(n - n_k)(nn_kd) = (K - 1)(nd)$ increases as the number of groups increases. In other words, we gain more efficiency with more groups. The second part of efficiency gain, quantified as $\boldsymbol{V}_2 + \boldsymbol{V}_3$, comes from the reducing subspace structure. Under our model assumption, the proposed estimator identifies and eliminates $\boldsymbol{Q}_{\boldsymbol{\Gamma}}\overline{\boldsymbol{D}}_k\boldsymbol{P}_{\boldsymbol{\Gamma}}$ and $\boldsymbol{P}_{\boldsymbol{\Gamma}}\overline{\boldsymbol{D}}_k\boldsymbol{Q}_{\boldsymbol{\Gamma}}$ in the estimation. These two terms are essentially zero matrices, and their estimation would only bring extraneous variability. Our method, however, avoids estimating these terms.

Proof. Under a fix n_k and the true $\boldsymbol{\Gamma}$, we have $\widehat{\boldsymbol{\Delta}}_k = \boldsymbol{P}_{\boldsymbol{\Gamma}}\overline{\boldsymbol{D}}_k\boldsymbol{P}_{\boldsymbol{\Gamma}} + \boldsymbol{Q}_{\boldsymbol{\Gamma}}\overline{\boldsymbol{D}}\boldsymbol{Q}_{\boldsymbol{\Gamma}}$, and

$$\mathrm{vec}(\widehat{\boldsymbol{\Delta}}_k) = (\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}})\mathrm{vec}(\overline{\boldsymbol{D}}_k) + (\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}})\mathrm{vec}(\overline{\boldsymbol{D}}).$$

Since $\boldsymbol{D}_{i,k} \sim W_p(\boldsymbol{\Delta}_k/d, d)$ are i.i.d. Wishart samples, $i = 1, \ldots, n_k$, $k = 1, \ldots, K$, we have $n_k\overline{\boldsymbol{D}}_k = \sum_{i=1}^{n_k} \boldsymbol{D}_{i,k} \sim W_p(\boldsymbol{\Delta}_k/d, n_k d)$. Denote $\boldsymbol{\Sigma}_k = \text{cov}\{\text{vec}(\overline{\boldsymbol{D}}_k)\}$, and $\boldsymbol{\Sigma} = \text{cov}\{\text{vec}(\overline{\boldsymbol{D}})\}$. Then $\boldsymbol{\Sigma} = \sum_{k=1}^{K} n_k^2/n^2\, \boldsymbol{\Sigma}_k$, and $\text{cov}\{\text{vec}(\overline{\boldsymbol{D}}), \text{vec}(\overline{\boldsymbol{D}}_k)\} = n_k/n\, \boldsymbol{\Sigma}_k$. Furthermore, based on Henderson and Searle (1979), we have

$$\boldsymbol{\Sigma}_k = \frac{1}{n_k d}(\boldsymbol{\Delta}_k \otimes \boldsymbol{\Delta}_k)\left\{\boldsymbol{I}_{p^2} + \boldsymbol{I}_{(p,p)}\right\},$$

where $\boldsymbol{I}_{p^2}$ is an identity matrix with dimension $p^2 \times p^2$, and $\boldsymbol{I}_{(p,p)}$ is the vec-permutation matrix. Then,

$$\begin{aligned}
\text{cov}\{\text{vec}(\widehat{\boldsymbol{\Delta}}_k)\} &= (\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}}) + (\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}}) \\
&\quad + \frac{n_k}{n}(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}}) + \frac{n_k}{n}(\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}}) \\
&= (\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}}) + (\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}}), \\
\text{cov}\{\text{vec}(\overline{\boldsymbol{D}}_k)\} &= \{(\boldsymbol{P}_{\boldsymbol{\Gamma}} + \boldsymbol{Q}_{\boldsymbol{\Gamma}}) \otimes (\boldsymbol{P}_{\boldsymbol{\Gamma}} + \boldsymbol{Q}_{\boldsymbol{\Gamma}})\}\boldsymbol{\Sigma}_k\{(\boldsymbol{P}_{\boldsymbol{\Gamma}} + \boldsymbol{Q}_{\boldsymbol{\Gamma}}) \otimes (\boldsymbol{P}_{\boldsymbol{\Gamma}} + \boldsymbol{Q}_{\boldsymbol{\Gamma}})\}.
\end{aligned}$$

Directly expanding $\text{cov}\{\text{vec}(\overline{\boldsymbol{D}}_k)\}$ leads to 16 additive terms, of which 10 are zero; i.e.,

$$\begin{aligned}
(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}}) &= (\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}}) = 0, \\
(\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}}) &= (\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}}) = 0, \\
(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}}) &= (\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}}) = 0, \\
(\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}}) &= (\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}}) = 0, \\
(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}}) &= (\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}}) = 0,
\end{aligned}$$

since $\boldsymbol{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{Q}_{\boldsymbol{\Gamma}} = \boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{P}_{\boldsymbol{\Gamma}} = 0$. There are six nonzero terms left in $\text{cov}\{\text{vec}(\overline{\boldsymbol{D}}_k)\}$, i.e., $(\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}})$, $(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}})$, and

$$\begin{aligned}
(\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}}) &= \frac{1}{n_k d}(\boldsymbol{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{P}_{\boldsymbol{\Gamma}}) \otimes (\boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{Q}_{\boldsymbol{\Gamma}}), \\
(\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}}) &= \frac{1}{n_k d}\left\{(\boldsymbol{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{P}_{\boldsymbol{\Gamma}}) \otimes (\boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{Q}_{\boldsymbol{\Gamma}})\right\}\boldsymbol{I}_{(p,p)}, \\
(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}}) &= \frac{1}{n_k d}\left\{(\boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{Q}_{\boldsymbol{\Gamma}}) \otimes (\boldsymbol{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{P}_{\boldsymbol{\Gamma}})\right\}\boldsymbol{I}_{(p,p)}, \\
(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}}) &= \frac{1}{n_k d}\left\{(\boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{Q}_{\boldsymbol{\Gamma}}) \otimes (\boldsymbol{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{P}_{\boldsymbol{\Gamma}})\right\}.
\end{aligned}$$

Therefore, we have

$$\begin{aligned}
\text{cov}\{\text{vec}(\overline{\boldsymbol{D}}_k)\} &= (\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{P}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{P}_{\boldsymbol{\Gamma}}) + (\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}}) \\
&\quad + \frac{1}{n_k d}\left\{(\boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{Q}_{\boldsymbol{\Gamma}}) \otimes (\boldsymbol{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{P}_{\boldsymbol{\Gamma}})\right\}\left\{\boldsymbol{I}_{p^2} + \boldsymbol{I}_{(p,p)}\right\} \\
&\quad + \frac{1}{n_k d}\left\{(\boldsymbol{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{P}_{\boldsymbol{\Gamma}}) \otimes (\boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{Q}_{\boldsymbol{\Gamma}})\right\}\left\{\boldsymbol{I}_{p^2} + \boldsymbol{I}_{(p,p)}\right\}.
\end{aligned}$$

The difference of the variance of the two estimators is,

$$\begin{aligned}
\text{cov}(\text{vec}\{\overline{\boldsymbol{D}}_k)\} - \text{cov}\{\text{vec}(\widehat{\boldsymbol{\Delta}}_k)\} &= (\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}})(\boldsymbol{\Sigma}_k - \boldsymbol{\Sigma})(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}}) \\
&\quad + \frac{1}{n_k d}\left\{(\boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{Q}_{\boldsymbol{\Gamma}}) \otimes (\boldsymbol{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{P}_{\boldsymbol{\Gamma}})\right\}\left\{\boldsymbol{I}_{p^2} + \boldsymbol{I}_{(p,p)}\right\} \\
&\quad + \frac{1}{n_k d}\left\{(\boldsymbol{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{P}_{\boldsymbol{\Gamma}}) \otimes (\boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{Q}_{\boldsymbol{\Gamma}})\right\}\left\{\boldsymbol{I}_{p^2} + \boldsymbol{I}_{(p,p)}\right\}.
\end{aligned}$$

In addition, we note that,

$$(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}_k(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}}) = \frac{1}{n_k d}\left\{(\boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{Q}_{\boldsymbol{\Gamma}}) \otimes (\boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{Q}_{\boldsymbol{\Gamma}})\right\}\left\{\boldsymbol{I}_{p^2} + \boldsymbol{I}_{(p,p)}\right\},$$

$$(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}})\boldsymbol{\Sigma}(\boldsymbol{Q}_{\boldsymbol{\Gamma}} \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}}) = \frac{1}{nd}\left\{(\boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{Q}_{\boldsymbol{\Gamma}}) \otimes (\boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{Q}_{\boldsymbol{\Gamma}})\right\}\left\{\boldsymbol{I}_{p^2} + \boldsymbol{I}_{(p,p)}\right\}.$$

The conclusion then follows by noting that $\boldsymbol{Q}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{Q}_{\boldsymbol{\Gamma}} = \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}}$, and $\boldsymbol{P}_{\boldsymbol{\Gamma}}\boldsymbol{\Delta}_k\boldsymbol{P}_{\boldsymbol{\Gamma}} = \boldsymbol{\Gamma}\boldsymbol{\Omega}_k\boldsymbol{\Gamma}^{\mathrm{T}}$. This completes the proof of Proposition S1. □

A3 Proofs

A3.1 Proof of Proposition 1

Proof. The equivalence between (1) and (3), and between (2) and (4), directly follow by noting that $\boldsymbol{\Delta}_k\boldsymbol{\Delta}_k^{-1} = \boldsymbol{I}_p$. Next we prove the equivalence between (1) and (2).

First we show (1) implies (2). It is sufficient to show that $\boldsymbol{P}_{\mathcal{S}}\boldsymbol{\Delta}_k\boldsymbol{Q}_{\mathcal{S}} = 0$. Since $\mathcal{S} = \mathrm{span}(\boldsymbol{\Gamma})$, we have $\boldsymbol{P}_{\mathcal{S}}\boldsymbol{\Delta}_k\boldsymbol{Q}_{\mathcal{S}} = \boldsymbol{P}_{\mathcal{S}}(\boldsymbol{\Gamma}\boldsymbol{\Omega}_k\boldsymbol{\Gamma}^{\mathrm{T}} + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}})\boldsymbol{Q}_{\mathcal{S}} = 0$, and $\boldsymbol{Q}_{\mathcal{S}}(\boldsymbol{\Delta}_k - \boldsymbol{\Delta}_j)\boldsymbol{Q}_{\mathcal{S}} = \boldsymbol{Q}_{\mathcal{S}}\left\{\boldsymbol{\Gamma}(\boldsymbol{\Omega}_k - \boldsymbol{\Omega}_j)\boldsymbol{\Gamma}^{\mathrm{T}}\right\}\boldsymbol{Q}_{\mathcal{S}} = 0$.

Next we show (2) implies (1). Since $\mathcal{S} = \mathrm{span}(\boldsymbol{\Gamma})$, we rewrite $\boldsymbol{P}_{\mathcal{S}} = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathrm{T}}$, and $\boldsymbol{Q}_{\mathcal{S}} = \boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^{\mathrm{T}}$. Then we have $\boldsymbol{\Delta}_k = \boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Delta}_k\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathrm{T}} + \boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^{\mathrm{T}}\boldsymbol{\Delta}_k\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^{\mathrm{T}}$. Denote $\boldsymbol{\Omega}_k = \boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Delta}_k\boldsymbol{\Gamma}$, and $\boldsymbol{\Omega}_{k0} = \boldsymbol{\Gamma}_0^{\mathrm{T}}\boldsymbol{\Delta}_k\boldsymbol{\Gamma}_0$. We have $\boldsymbol{\Gamma}_0(\boldsymbol{\Omega}_{k0} - \boldsymbol{\Omega}_{j0})\boldsymbol{\Gamma}_0^{\mathrm{T}} = \boldsymbol{Q}_{\mathcal{S}}(\boldsymbol{\Delta}_k - \boldsymbol{\Delta}_j)\boldsymbol{Q}_{\mathcal{S}} = 0$, which holds for all $1 \le k, j \le K$. Therefore, for any k, j, $\boldsymbol{\Omega}_{k0} = \boldsymbol{\Omega}_{j0} = \boldsymbol{\Omega}_0$. Then $\boldsymbol{\Delta}_k = \boldsymbol{\Gamma}\boldsymbol{\Omega}_k\boldsymbol{\Gamma} + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}}$ follows. This completes the proof of Proposition 1. □

A3.2 Proof of Proposition 2

Proof. Proposition 2 is a direct consequence of Theorem S1. □

A3.3 Proof of Proposition 3

Proof. First we describe the joint likelihood-based objective function. For i.i.d. Wishart samples $\boldsymbol{D}_{i,k} \sim W_p(\boldsymbol{\Delta}_k/d, d)$, $i = 1, \ldots, n_k$, $k = 1, \ldots, K$, we have $n_k\overline{\boldsymbol{D}}_k = \sum_{i=1}^{n_k}\boldsymbol{D}_{i,k} \sim W_p(\boldsymbol{\Delta}_k/d, n_k d)$. Then

$$f(n_1\overline{\boldsymbol{D}}_1, ..., n_K\overline{\boldsymbol{D}}_K|\boldsymbol{\Lambda}_1, \ldots, \boldsymbol{\Lambda}_K, \boldsymbol{\beta}, \boldsymbol{\Delta}_0, d) \propto \prod_{k=1}^{K}\left|\frac{\boldsymbol{\Delta}_k}{d}\right|^{-\frac{n_k d}{2}}\mathrm{etr}\left\{-\left(\frac{\boldsymbol{\Delta}_k}{d}\right)^{-1}n_k\overline{\boldsymbol{D}}_k/2)\right\}.$$

Therefore we have the negative log-likelihood as the objective function, which is of the form,

$$\ell(\boldsymbol{\Delta}_1, ..., \boldsymbol{\Delta}_K) = \sum_{k=1}^{K}\frac{n_k d}{2}\left\{\log|\boldsymbol{\Delta}_k| + \mathrm{tr}(\boldsymbol{\Delta}_k^{-1}\overline{\boldsymbol{D}}_k)\right\}.$$

For our model, we have $\boldsymbol{\Delta}_k = \boldsymbol{\Gamma}\boldsymbol{\Omega}_k\boldsymbol{\Gamma}^{\mathrm{T}} + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}}$, and $\boldsymbol{\Delta}_k^{-1} = \boldsymbol{\Gamma}\boldsymbol{\Omega}_k^{-1}\boldsymbol{\Gamma}^{\mathrm{T}} + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Gamma}_0^{\mathrm{T}}$, $k = 1, \ldots, K$. Then the joint log-likelihood-based objective function is

$$\begin{aligned}\ell(\boldsymbol{\Gamma}, \boldsymbol{\Omega}_0, \boldsymbol{\Omega}_1, ..., \boldsymbol{\Omega}_K) &= \sum_{k=1}^{K}\frac{n_k d}{2}\left[\log|\boldsymbol{\Gamma}\boldsymbol{\Omega}_k\boldsymbol{\Gamma}^{\mathrm{T}} + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}}| + \mathrm{tr}\left\{(\boldsymbol{\Gamma}\boldsymbol{\Omega}_k^{-1}\boldsymbol{\Gamma}^{\mathrm{T}} + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Gamma}_0^{\mathrm{T}})\overline{\boldsymbol{D}}_k\right\}\right] \\ &= \sum_{k=1}^{K}\frac{n_k d}{2}\left[\log|\boldsymbol{\Omega}_k| + \log|\boldsymbol{\Omega}_0| + \mathrm{tr}\left\{(\boldsymbol{\Gamma}\boldsymbol{\Omega}_k^{-1}\boldsymbol{\Gamma}^{\mathrm{T}} + \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Gamma}_0^{\mathrm{T}})\overline{\boldsymbol{D}}_k\right\}\right].\end{aligned}$$

Furthermore, we have,

$$
\begin{aligned}
\frac{\partial \ell}{\partial \Omega_0} &= \sum_{k=1}^{K} \frac{n_k d}{2} \left\{ \frac{\partial \log |\boldsymbol{\Omega}_0|}{\partial \boldsymbol{\Omega}_0} + \frac{\mathrm{tr}\left(\boldsymbol{\Gamma}_0 \boldsymbol{\Omega}_0^{-1} \boldsymbol{\Gamma}_0^{\mathrm{T}} \overline{\boldsymbol{D}}_k\right)}{\partial \boldsymbol{\Omega}_0} \right\}, \\
\frac{\partial \ell}{\partial \Omega_k} &= \sum_{k=1}^{K} \frac{n_k d}{2} \left\{ \frac{\partial \log |\boldsymbol{\Omega}_k|}{\partial \boldsymbol{\Omega}_k} + \frac{\mathrm{tr}\left(\boldsymbol{\Gamma} \boldsymbol{\Omega}_k^{-1} \boldsymbol{\Gamma}^{\mathrm{T}} \overline{\boldsymbol{D}}_k\right)}{\partial \boldsymbol{\Omega}_k} \right\}, \\
\frac{\partial \log |\boldsymbol{\Omega}_0|}{\partial \boldsymbol{\Omega}_0} &= \boldsymbol{\Omega}_0^{-1}, \ \frac{\partial \log |\boldsymbol{\Omega}_k|}{\partial \boldsymbol{\Omega}_k} = \boldsymbol{\Omega}_k^{-1}, \\
\frac{\partial \mathrm{tr}\left(\boldsymbol{\Gamma}_0 \boldsymbol{\Omega}_0^{-1} \boldsymbol{\Gamma}_0^{\mathrm{T}} \overline{\boldsymbol{D}}_k\right)}{\partial \boldsymbol{\Omega}_0} &= \frac{\partial \mathrm{tr}\left(\boldsymbol{\Omega}_0^{-1} \boldsymbol{\Gamma}_0^{\mathrm{T}} \overline{\boldsymbol{D}}_k \boldsymbol{\Gamma}_0\right)}{\partial \boldsymbol{\Omega}_0} = -\boldsymbol{\Omega}_0^{-1} \boldsymbol{\Gamma}_0^{\mathrm{T}} \overline{\boldsymbol{D}}_k \boldsymbol{\Gamma}_0 \boldsymbol{\Omega}_0^{-1}, \\
\frac{\partial \mathrm{tr}\left(\boldsymbol{\Gamma} \boldsymbol{\Omega}_k^{-1} \boldsymbol{\Gamma}^{\mathrm{T}} \overline{\boldsymbol{D}}_k\right)}{\partial \boldsymbol{\Omega}_k} &= \frac{\partial \mathrm{tr}\left(\boldsymbol{\Omega}_k^{-1} \boldsymbol{\Gamma}^{\mathrm{T}} \overline{\boldsymbol{D}}_k \boldsymbol{\Gamma}\right)}{\partial \boldsymbol{\Omega}_k} = -\boldsymbol{\Omega}_k^{-1} \boldsymbol{\Gamma}^{\mathrm{T}} \overline{\boldsymbol{D}}_k \boldsymbol{\Gamma} \boldsymbol{\Omega}_k^{-1}.
\end{aligned}
$$

If we set $\partial \ell / \partial \boldsymbol{\Omega}_0 = 0$ and $\partial \ell / \partial \boldsymbol{\Omega}_k = 0$, then the maximum likelihood estimators of $\boldsymbol{\Omega}_0$, $\boldsymbol{\Omega}_k$ and $\boldsymbol{\Delta}_k$ satisfy that $\widehat{\boldsymbol{\Omega}}_0 = \widehat{\boldsymbol{\Gamma}}_0^{\mathrm{T}} \overline{\boldsymbol{D}} \widehat{\boldsymbol{\Gamma}}_0$, $\widehat{\boldsymbol{\Omega}}_k = \widehat{\boldsymbol{\Gamma}}^{\mathrm{T}} \overline{\boldsymbol{D}}_k \widehat{\boldsymbol{\Gamma}}$, and $\widehat{\boldsymbol{\Delta}}_k = \widehat{\boldsymbol{\Gamma}} \widehat{\boldsymbol{\Gamma}}^{\mathrm{T}} \overline{\boldsymbol{D}}_k \widehat{\boldsymbol{\Gamma}} \widehat{\boldsymbol{\Gamma}}^{\mathrm{T}} + \widehat{\boldsymbol{\Gamma}}_0 \widehat{\boldsymbol{\Gamma}}_0^{\mathrm{T}} \overline{\boldsymbol{D}} \widehat{\boldsymbol{\Gamma}}_0 \widehat{\boldsymbol{\Gamma}}_0^{\mathrm{T}}$, where $\overline{\boldsymbol{D}} = \sum_{k=1}^{K} (n_k/n) \overline{\boldsymbol{D}}_k$, and $\widehat{\boldsymbol{\Gamma}}$ is the minimizer of the following partially optimized negative log-likelihood,

$$
\ell(\boldsymbol{\Gamma}) = \sum_{k=1}^{K} \frac{n_k d}{2} \left(\log |\boldsymbol{\Gamma}^{\mathrm{T}} \overline{\boldsymbol{D}}_k \boldsymbol{\Gamma}| + \log |\boldsymbol{\Gamma}_0^{\mathrm{T}} \overline{\boldsymbol{D}} \boldsymbol{\Gamma}_0| \right),
$$

in which $(\boldsymbol{\Gamma}, \boldsymbol{\Gamma}_0)$ forms an orthogonal basis matrix of $\mathbb{R}^p$. To further simplify $\ell(\boldsymbol{\Gamma})$, we apply the identity $|\boldsymbol{\Gamma}_0^{\mathrm{T}} \overline{\boldsymbol{D}} \boldsymbol{\Gamma}_0| = |\overline{\boldsymbol{D}}| |\boldsymbol{\Gamma}^{\mathrm{T}} (\overline{\boldsymbol{D}})^{-1} \boldsymbol{\Gamma}|$ from Proposition A2 of Cook and Forzani (2008). Omitting an additive constant $\log |\overline{\boldsymbol{D}}|$, $\ell(\boldsymbol{\Gamma})$ can be rewritten as,

$$
\ell(\boldsymbol{\Gamma}) = \frac{nd}{2} \log |\boldsymbol{\Gamma}^{\mathrm{T}} (\overline{\boldsymbol{D}})^{-1} \boldsymbol{\Gamma}| + \sum_{k=1}^{K} \frac{n_k d}{2} \log |\boldsymbol{\Gamma}^{\mathrm{T}} \overline{\boldsymbol{D}}_k \boldsymbol{\Gamma}|.
$$

Therefore the likelihood-based estimator $\widehat{\boldsymbol{\Gamma}}$ is obtained from minimizing the above objective function $\ell(\boldsymbol{\Gamma})$ subject to the constraint $\boldsymbol{\Gamma}^{\mathrm{T}} \boldsymbol{\Gamma} = \mathrm{I}_u$. This completes the proof of Proposition 3. □

A3.4 Proof of Theorem 1

Proof. By direct calculation, we can obtain both the asymptotic covariance, $\boldsymbol{V} = \boldsymbol{J}_h^{-1}$, of the standard maximum likelihood estimator, as well as the asymptotic covariance, $\boldsymbol{U} = \boldsymbol{H}(\boldsymbol{H}^{\mathrm{T}} \boldsymbol{J}_h \boldsymbol{H})^{\dagger} \boldsymbol{H}^{\mathrm{T}}$, of our common reducing subspace-based estimator. Specifically, the Fisher information matrix $\boldsymbol{J}_h$ and the gradient matrix $\boldsymbol{H}$ is consistent with Cook et al. (2010).

To show $\boldsymbol{U} \leq \boldsymbol{V}$, we observe that

$$
\boldsymbol{V} - \boldsymbol{U} = \boldsymbol{J}_h^{-1} - \boldsymbol{H}(\boldsymbol{H}^{\mathrm{T}} \boldsymbol{J}_h \boldsymbol{H})^{\dagger} \boldsymbol{H}^{\mathrm{T}} = \boldsymbol{J}_h^{-\frac{1}{2}} \left\{ \boldsymbol{I} - \boldsymbol{J}_h^{\frac{1}{2}} \boldsymbol{H}(\boldsymbol{H}^{\mathrm{T}} \boldsymbol{J}_h \boldsymbol{H})^{\dagger} \boldsymbol{H}^{\mathrm{T}} \boldsymbol{J}_h^{\frac{1}{2}} \right\} \boldsymbol{J}_h^{-\frac{1}{2}} = \boldsymbol{J}_h^{-\frac{1}{2}} \boldsymbol{Q}_{\boldsymbol{J}_h^{\frac{1}{2}} \boldsymbol{H}} \boldsymbol{J}_h^{-\frac{1}{2}},
$$

which is a symmetric positive semi-definite matrix. This completes the proof of Theorem 1. □

A3.5 Proof of Theorem 2

Proof. We first derive an alternative form for $\boldsymbol{U} = \boldsymbol{H}(\boldsymbol{H}^{\mathrm{T}} \boldsymbol{J}_h \boldsymbol{H})^{\dagger} \boldsymbol{H}^{\mathrm{T}}$, which facilitates the interpretation. Since $\boldsymbol{U}$ only depends on the columnn space of $\boldsymbol{H}$, we can replace $\boldsymbol{H}$ by $\boldsymbol{H}_1$ that has the same column space as $\boldsymbol{H}$, whereas $\boldsymbol{H}_1^{\mathrm{T}} \boldsymbol{J}_h \boldsymbol{H}_1$ is block-diagonal, and

$$
\begin{aligned}
\boldsymbol{H}_1 &= \begin{pmatrix} 2\boldsymbol{C}_p(\boldsymbol{\Gamma}\boldsymbol{\Omega}_1 \otimes \boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}_0 \boldsymbol{\Omega}_0) & \boldsymbol{C}_p(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}) \boldsymbol{E}_u & \boldsymbol{0} & \boldsymbol{C}_p(\boldsymbol{\Gamma}_0 \otimes \boldsymbol{\Gamma}_0) \boldsymbol{E}_{p-u} \\ 2\boldsymbol{C}_p(\boldsymbol{\Gamma}\boldsymbol{\Omega}_2 \otimes \boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}_0 \boldsymbol{\Omega}_0) & \boldsymbol{0} & \boldsymbol{C}_p(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}) \boldsymbol{E}_u & \boldsymbol{C}_p(\boldsymbol{\Gamma}_0 \otimes \boldsymbol{\Gamma}_0) \boldsymbol{E}_{p-u} \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{H}_{11} & \boldsymbol{H}_{12} & \boldsymbol{H}_{13} & \boldsymbol{H}_{14} \end{pmatrix}.
\end{aligned}
$$

Moreover, let

$$\boldsymbol{H}_2 = \begin{pmatrix} \boldsymbol{I}_u \otimes \boldsymbol{\Gamma}_0^{\mathrm{T}} & 0 & 0 & 0 \\ 2\boldsymbol{C}_u(\boldsymbol{\Omega}_1 \otimes \boldsymbol{\Gamma}^{\mathrm{T}}) & \boldsymbol{I}_{\frac{u(u+1)}{2}} & 0 & 0 \\ 2\boldsymbol{C}_u(\boldsymbol{\Omega}_2 \otimes \boldsymbol{\Gamma}^{\mathrm{T}}) & 0 & \boldsymbol{I}_{\frac{u(u+1)}{2}} & 0 \\ 0 & 0 & 0 & \boldsymbol{I}_{\frac{(r-u)(r-u+1)}{2}} \end{pmatrix}.$$

Since $\boldsymbol{H} = \boldsymbol{H}_1\boldsymbol{H}_2$, and $\boldsymbol{H}_2$ is of the full row rank, we have span$(\boldsymbol{H})$ = span$(\boldsymbol{H}_1)$. Furthermore, by straightforward multiplication, we see that $\boldsymbol{H}_1^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_1$ is block-diagonal with

$$\begin{aligned} \boldsymbol{H}_{11}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_{11} &= d\sum_{k=1}^{2}\pi_k\left(\boldsymbol{\Omega}_k \otimes \boldsymbol{\Omega}_0^{-1} - 2\boldsymbol{I}_{u\times(r-u)} + \boldsymbol{\Omega}_k^{-1} \otimes \boldsymbol{\Omega}_0\right), \\ \boldsymbol{H}_{12}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_{12} &= \frac{d\pi_1}{2}\left\{\boldsymbol{E}_u^{\mathrm{T}}(\boldsymbol{\Omega}_1^{-1} \otimes \boldsymbol{\Omega}_1^{-1})\boldsymbol{E}_u\right\}, \\ \boldsymbol{H}_{13}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_{13} &= \frac{d\pi_2}{2}\left\{\boldsymbol{E}_u^{\mathrm{T}}(\boldsymbol{\Omega}_2^{-1} \otimes \boldsymbol{\Omega}_2^{-1})\boldsymbol{E}_u\right\}, \\ \boldsymbol{H}_{14}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_{14} &= \frac{d}{2}\left\{\boldsymbol{E}_{p-u}^{\mathrm{T}}(\boldsymbol{\Omega}_0^{-1} \otimes \boldsymbol{\Omega}_0^{-1})\boldsymbol{E}_{p-u}\right\}. \end{aligned}$$

Thus, we can now write $\boldsymbol{U} = \boldsymbol{H}_1(\boldsymbol{H}_1^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_1)^{\dagger}\boldsymbol{H}_1^{\mathrm{T}} = \sum_{j=1}^{4}\boldsymbol{H}_{1j}(\boldsymbol{H}_{1j}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_{1j})^{\dagger}\boldsymbol{H}_{1j}^{\mathrm{T}}$.

Our interest is avar$\{\sqrt{n}$vech$(\widehat{\boldsymbol{\Delta}}_1)\}$, which is the upper left block of $\boldsymbol{U}$. Since $\boldsymbol{H}_{13}^{\mathrm{T}}\boldsymbol{J}\boldsymbol{H}_{13}$ does not contribute to the upper left block, we focus only on the upper left block of $\boldsymbol{H}_{1j}(\boldsymbol{H}_{1j}^{\mathrm{T}}\boldsymbol{J}\boldsymbol{H}_{1j})^{\dagger}\boldsymbol{H}_{1j}^{\mathrm{T}}$, for $j = 1, 2, 4$. Let $\boldsymbol{H}_{1,ij}$ denote the (i,j)th block of the matrix $\boldsymbol{H}_1$, then the upper left block of $\boldsymbol{U}$ can be decomposed in the following way:

$$\begin{aligned} \boldsymbol{H}_{1,11}(\boldsymbol{H}_{11}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_{11})^{\dagger}\boldsymbol{H}_{1,11}^{\mathrm{T}} &= \frac{4}{d}\boldsymbol{C}_p(\boldsymbol{\Gamma}\boldsymbol{\Omega}_1 \otimes \boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0) \\ &\times \left\{\sum_{k=1}^{2}\pi_k(\boldsymbol{\Omega}_k \otimes \boldsymbol{\Omega}_0^{-1} - 2\boldsymbol{I}_{u\times(r-u)} + \boldsymbol{\Omega}_k^{-1} \otimes \boldsymbol{\Omega}_0)\right\}^{\dagger} \\ &\times (\boldsymbol{\Omega}_1\boldsymbol{\Gamma}^{\mathrm{T}} \otimes \boldsymbol{\Omega}_0^{\mathrm{T}} - \boldsymbol{\Gamma}^{\mathrm{T}} \otimes \boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}})\boldsymbol{C}_p^{\mathrm{T}}, \\ \boldsymbol{H}_{1,12}(\boldsymbol{H}_{12}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_{12})^{\dagger}\boldsymbol{H}_{1,12}^{\mathrm{T}} &= \boldsymbol{C}_p(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})\boldsymbol{E}_u\left\{\frac{d\pi_1}{2}\boldsymbol{E}_u^{\mathrm{T}}(\boldsymbol{\Omega}_1^{-1} \otimes \boldsymbol{\Omega}_1^{-1})\boldsymbol{E}_u\right\}^{\dagger}\boldsymbol{E}_u^{\mathrm{T}}(\boldsymbol{\Gamma}^{\mathrm{T}} \otimes \boldsymbol{\Gamma}^{\mathrm{T}})\boldsymbol{C}_p^{\mathrm{T}} \\ &= \frac{2}{d\pi_1}\boldsymbol{C}_p(\boldsymbol{\Gamma}\boldsymbol{\Omega}_1\boldsymbol{\Gamma}^{\mathrm{T}} \otimes \boldsymbol{\Gamma}\boldsymbol{\Omega}_1\boldsymbol{\Gamma}^{\mathrm{T}})\boldsymbol{C}_p^{\mathrm{T}}, \\ \boldsymbol{H}_{1,14}(\boldsymbol{H}_{14}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_{14})^{\dagger}\boldsymbol{H}_{1,14}^{\mathrm{T}} &= \boldsymbol{C}_p(\boldsymbol{\Gamma}_0 \otimes \boldsymbol{\Gamma}_0)\boldsymbol{E}_{p-u}\left\{\frac{d}{2}\boldsymbol{E}_{p-u}^{\mathrm{T}}(\boldsymbol{\Omega}_0^{-1} \otimes \boldsymbol{\Omega}_0^{-1})\boldsymbol{E}_{p-u}\right\}^{\dagger}\boldsymbol{E}_{p-u}^{\mathrm{T}}(\boldsymbol{\Gamma}_0^{\mathrm{T}} \otimes \boldsymbol{\Gamma}_0^{\mathrm{T}})\boldsymbol{C}_p^{\mathrm{T}} \\ &= \frac{2}{d}\boldsymbol{C}_p(\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}} \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}})\boldsymbol{C}_p^{\mathrm{T}}. \end{aligned}$$

Henceforth,

$$\text{avar}\{\sqrt{n}\text{vech}(\widehat{\boldsymbol{\Delta}}_1)\} = \boldsymbol{H}_{1,11}(\boldsymbol{H}_{11}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_{11})^{\dagger}\boldsymbol{H}_{1,11}^{\mathrm{T}} + \boldsymbol{H}_{1,12}(\boldsymbol{H}_{12}^{\mathrm{T}}\boldsymbol{J}\boldsymbol{H}_{12})^{\dagger}\boldsymbol{H}_{1,12}^{\mathrm{T}} + \boldsymbol{H}_{1,14}(\boldsymbol{H}_{14}^{\mathrm{T}}\boldsymbol{J}\boldsymbol{H}_{14})^{\dagger}\boldsymbol{H}_{1,14}^{\mathrm{T}}.$$

The Fisher information matrix of $\boldsymbol{\theta}$ is simply $\boldsymbol{H}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}$; that is,

$$\boldsymbol{J}_{\boldsymbol{\theta}} = \boldsymbol{H}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H} = \begin{pmatrix} \boldsymbol{J}_{\boldsymbol{\Gamma}} & \boldsymbol{J}_{\boldsymbol{\Gamma}\boldsymbol{\Omega}_1} & \boldsymbol{J}_{\boldsymbol{\Gamma}\boldsymbol{\Omega}_2} & 0 \\ \boldsymbol{J}_{\boldsymbol{\Omega}_1\boldsymbol{\Gamma}} & \boldsymbol{J}_{\boldsymbol{\Omega}_1} & 0 & 0 \\ \boldsymbol{J}_{\boldsymbol{\Omega}_2\boldsymbol{\Gamma}} & 0 & \boldsymbol{J}_{\boldsymbol{\Omega}_2} & 0 \\ 0 & 0 & 0 & \boldsymbol{J}_{\boldsymbol{\Omega}_0} \end{pmatrix}.$$

Each nonzero block is of the form,

$$\boldsymbol{J}_{\boldsymbol{\Gamma}} = 2d\sum_{k=1}^{2}\pi_k(\boldsymbol{\Omega}_k^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}\otimes\boldsymbol{I}_p - \boldsymbol{\Gamma}^{\mathrm{T}}\otimes\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}})\boldsymbol{C}_p^{\mathrm{T}}\boldsymbol{E}_p^{\mathrm{T}}(\boldsymbol{\Delta}_k^{-1}\otimes\boldsymbol{\Delta}_k^{-1})\boldsymbol{E}_p\boldsymbol{C}_p(\boldsymbol{\Gamma}\boldsymbol{\Omega}_k\otimes\boldsymbol{I}_p - \boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}}),$$

$$\begin{aligned}
\boldsymbol{J}_{\boldsymbol{\Gamma}\boldsymbol{\Omega}_1} &= \boldsymbol{J}_{\boldsymbol{\Omega}_1\boldsymbol{\Gamma}}^{\mathrm{T}} = d\pi_1(\boldsymbol{I}_u\otimes\boldsymbol{\Delta}_1^{-1}\boldsymbol{\Gamma})\boldsymbol{E}_u,\\
\boldsymbol{J}_{\boldsymbol{\Gamma}\boldsymbol{\Omega}_2} &= \boldsymbol{J}_{\boldsymbol{\Omega}_2\boldsymbol{\Gamma}}^{\mathrm{T}} = d\pi_2(\boldsymbol{I}_u\otimes\boldsymbol{\Delta}_2^{-1}\boldsymbol{\Gamma})\boldsymbol{E}_u,\\
\boldsymbol{J}_{\boldsymbol{\Omega}_1} &= \frac{d\pi_1}{2}\left\{\boldsymbol{E}_u^{\mathrm{T}}(\boldsymbol{\Omega}_1^{-1}\otimes\boldsymbol{\Omega}_1^{-1})\boldsymbol{E}_u\right\},\\
\boldsymbol{J}_{\boldsymbol{\Omega}_2} &= \frac{d\pi_2}{2}\left\{\boldsymbol{E}_u^{\mathrm{T}}(\boldsymbol{\Omega}_2^{-1}\otimes\boldsymbol{\Omega}_2^{-1})\boldsymbol{E}_u\right\},\\
\boldsymbol{J}_{\boldsymbol{\Omega}_0} &= \frac{d}{2}\left\{\boldsymbol{E}_{p-u}^{\mathrm{T}}(\boldsymbol{\Omega}_0^{-1}\otimes\boldsymbol{\Omega}_0^{-1})\boldsymbol{E}_{p-u}\right\}.
\end{aligned}$$

Considering avar$\{\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\Gamma}})\}$, the asymptotic variance is the upper left block of $\boldsymbol{J}_\theta^{-1}$:

$$\left\{\boldsymbol{J}_{\boldsymbol{\Gamma}} - \begin{pmatrix}\boldsymbol{J}_{\boldsymbol{\Gamma}\boldsymbol{\Omega}_1} & \boldsymbol{J}_{\boldsymbol{\Gamma}\boldsymbol{\Omega}_2}\end{pmatrix}\begin{pmatrix}\boldsymbol{J}_{\boldsymbol{\Omega}_1}^{-1} & 0\\ 0 & \boldsymbol{J}_{\boldsymbol{\Omega}_2}^{-1}\end{pmatrix}\begin{pmatrix}\boldsymbol{J}_{\boldsymbol{\Omega}_1\boldsymbol{\Gamma}}\\ \boldsymbol{J}_{\boldsymbol{\Omega}_2\boldsymbol{\Gamma}}\end{pmatrix}\right\}^{-1}.$$

We have

$$\begin{aligned}
\boldsymbol{J}_{\boldsymbol{\Gamma}} &= 2d\sum_{k=1}^{2}\pi_k(\boldsymbol{\Omega}_k^{\mathrm{T}}\boldsymbol{\Gamma}^{\mathrm{T}}\otimes\boldsymbol{I}_p - \boldsymbol{\Gamma}^{\mathrm{T}}\otimes\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}})\boldsymbol{C}_p^{\mathrm{T}}\boldsymbol{E}_p^{\mathrm{T}}(\boldsymbol{\Delta}_k^{-1}\otimes\boldsymbol{\Delta}_k^{-1})\boldsymbol{E}_p\boldsymbol{C}_p(\boldsymbol{\Gamma}\boldsymbol{\Omega}_k\otimes\boldsymbol{I}_p - \boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}})\\
&= d\sum_{k=1}^{2}\pi_k\left\{(\boldsymbol{\Omega}_k\otimes\boldsymbol{\Delta}_k^{-1}) + (\boldsymbol{\Omega}_k\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Delta}_k^{-1}\otimes\boldsymbol{\Delta}_k^{-1}\boldsymbol{\Gamma}\boldsymbol{\Omega}_k)\boldsymbol{K}_{ru} - 2(\boldsymbol{I}_u\otimes\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^{\mathrm{T}}) + (\boldsymbol{\Omega}_k^{-1}\otimes\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}})\right\}.
\end{aligned}$$

Through calculation, we have

$$\begin{pmatrix}\boldsymbol{J}_{\boldsymbol{\Gamma}\boldsymbol{\Omega}_1} & \boldsymbol{J}_{\boldsymbol{\Gamma}\boldsymbol{\Omega}_2}\end{pmatrix}\begin{pmatrix}\boldsymbol{J}_{\boldsymbol{\Omega}_1}^{-1} & 0\\ 0 & \boldsymbol{J}_{\boldsymbol{\Omega}_2}^{-1}\end{pmatrix}\begin{pmatrix}\boldsymbol{J}_{\boldsymbol{\Omega}_1\boldsymbol{\Gamma}}\\ \boldsymbol{J}_{\boldsymbol{\Omega}_2\boldsymbol{\Gamma}}\end{pmatrix} = d\sum_{k=1}^{2}\pi_k\left\{\boldsymbol{\Omega}_k\otimes\boldsymbol{\Gamma}\boldsymbol{\Omega}_k^{-1}\boldsymbol{\Gamma}^{\mathrm{T}} + (\boldsymbol{\Omega}_k\boldsymbol{\Gamma}^{\mathrm{T}}\boldsymbol{\Delta}_k^{-1}\otimes\boldsymbol{\Delta}_k^{-1}\boldsymbol{\Gamma}\boldsymbol{\Omega}_k)\boldsymbol{K}_{ru}\right\},$$

where $\boldsymbol{K}_{pu}\in\mathbb{R}^{pu\times pu}$ is the commutation matrix, such that for $\boldsymbol{M}\in\mathbb{R}^{p\times u}$, $\mathrm{vec}(\boldsymbol{M}^T) = \boldsymbol{K}_{pu}\mathrm{vec}(\boldsymbol{M})$. Then we have

$$\left[\mathrm{avar}\left\{\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\Gamma}})\right\}\right]^{-1} = d\sum_{k=1}^{2}\pi_k\left\{(\boldsymbol{\Omega}_k\otimes\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0^{-1}\boldsymbol{\Gamma}_0^{\mathrm{T}}) - 2(\boldsymbol{I}_u\otimes\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^{\mathrm{T}}) + (\boldsymbol{\Omega}_k^{-1}\otimes\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}})\right\}.$$

In addition, we have avar$\left\{\sqrt{n}\mathrm{vech}(\widehat{\boldsymbol{\Omega}}_{1,\boldsymbol{\Gamma}})\right\} = \boldsymbol{J}_{\boldsymbol{\Omega}_1}^{-1}$, avar$\left\{\sqrt{n}\mathrm{vech}(\widehat{\boldsymbol{\Omega}}_{0,\boldsymbol{\Gamma}})\right\} = \boldsymbol{J}_{\boldsymbol{\Omega}_0}^{-1}$. By comparing these with avar$\{\sqrt{n}\mathrm{vech}(\widehat{\boldsymbol{\Delta}}_1)\}$, we see that

$$\begin{aligned}
\boldsymbol{H}_{1,12}(\boldsymbol{H}_{12}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_{12})^{\dagger}\boldsymbol{H}_{1,12}^{\mathrm{T}} &= \boldsymbol{C}_p(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\boldsymbol{E}_u\cdot\mathrm{avar}\left\{\sqrt{n}\mathrm{vech}(\widehat{\boldsymbol{\Omega}}_{1,\boldsymbol{\Gamma}})\right\}\cdot\boldsymbol{E}_u^{\mathrm{T}}(\boldsymbol{\Gamma}^{\mathrm{T}}\otimes\boldsymbol{\Gamma}^{\mathrm{T}})\boldsymbol{C}_p^{\mathrm{T}}\\
&= \mathrm{avar}\left[\sqrt{n}\left\{\boldsymbol{C}_p(\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma})\boldsymbol{E}_u\right\}\mathrm{vech}(\widehat{\boldsymbol{\Omega}}_{1,\boldsymbol{\Gamma}})\right]\\
&= \mathrm{avar}\left\{\sqrt{n}\mathrm{vech}(\boldsymbol{\Gamma}\widehat{\boldsymbol{\Omega}}_{1,\boldsymbol{\Gamma}}\boldsymbol{\Gamma}^{\mathrm{T}})\right\}.
\end{aligned}$$

Similarly, we can show that $\boldsymbol{H}_{1,14}(\boldsymbol{H}_{14}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_{14})^{\dagger}\boldsymbol{H}_{1,14}^{\mathrm{T}} = \mathrm{avar}\left\{\sqrt{n}\mathrm{vech}(\boldsymbol{\Gamma}_0\widehat{\boldsymbol{\Omega}}_{0,\boldsymbol{\Gamma}}\boldsymbol{\Gamma}_0^{\mathrm{T}})\right\}$. Therefore we have $\boldsymbol{H}_{1,12}(\boldsymbol{H}_{12}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_{12})^{\dagger}\boldsymbol{H}_{1,12}^{\mathrm{T}} + \boldsymbol{H}_{1,14}(\boldsymbol{H}_{14}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_{14})^{\dagger}\boldsymbol{H}_{1,14}^{\mathrm{T}} = \mathrm{avar}\left\{\sqrt{n}\mathrm{vech}(\widehat{\boldsymbol{\Delta}}_{1,\boldsymbol{\Gamma}})\right\}$. Also, comparing $\boldsymbol{H}_{11}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_{11}$ with avar$\left\{\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\Gamma}})\right\}$, we see that

$$\boldsymbol{H}_{11}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_{11} = (\boldsymbol{I}_u\otimes\boldsymbol{\Gamma}_0^{\mathrm{T}})\mathrm{avar}\left\{\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\Gamma}})\right\}^{-1}(\boldsymbol{I}_u\otimes\boldsymbol{\Gamma}_0).$$

Therefore,

$$\begin{aligned}\boldsymbol{H}_{1,11}(\boldsymbol{H}_{11}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_{11})^{\dagger}\boldsymbol{H}_{1,11}^{\mathrm{T}} &= 4\boldsymbol{C}_p(\boldsymbol{\Gamma}\boldsymbol{\Omega}_1\otimes\boldsymbol{I}_p-\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}})(\boldsymbol{I}_u\otimes\boldsymbol{\Gamma}_0)\\ &\times \left[(\boldsymbol{I}_u\otimes\boldsymbol{\Gamma}_0^{\mathrm{T}})\mathrm{avar}\left\{\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\Gamma}})\right\}^{-1}(\boldsymbol{I}_u\otimes\boldsymbol{\Gamma}_0)\right]^{\dagger}\\ &\times (\boldsymbol{I}_u\otimes\boldsymbol{\Gamma}_0^{\mathrm{T}})(\boldsymbol{\Omega}_1\boldsymbol{\Gamma}^{\mathrm{T}}\otimes\boldsymbol{I}_p-\boldsymbol{\Gamma}^{\mathrm{T}}\otimes\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}}).\end{aligned}$$

From Cook et al. (2010), since $\boldsymbol{P}_{\boldsymbol{G}}=\boldsymbol{I}_u\otimes\boldsymbol{\Gamma}_0\boldsymbol{\Gamma}_0^{\mathrm{T}}$, and $\left[\mathrm{avar}\left\{\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\Gamma}})\right\}\right]^{-1}$ commutes, we have

$$\begin{aligned}\boldsymbol{H}_{1,11}(\boldsymbol{H}_{11}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_{11})^{\dagger}\boldsymbol{H}_{1,11}^{\mathrm{T}} &= 4\boldsymbol{C}_p(\boldsymbol{\Gamma}\boldsymbol{\Omega}_1\otimes\boldsymbol{I}_p-\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}})(\boldsymbol{I}_u\otimes\boldsymbol{Q}_{\boldsymbol{\Gamma}})\times\mathrm{avar}\left\{\sqrt{n}\mathrm{vec}(\widehat{\boldsymbol{\Gamma}})\right\}\\ &\times (\boldsymbol{I}_u\otimes\boldsymbol{Q}_{\boldsymbol{\Gamma}})(\boldsymbol{\Omega}_1\boldsymbol{\Gamma}^{\mathrm{T}}\otimes\boldsymbol{I}_p-\boldsymbol{\Gamma}^{\mathrm{T}}\otimes\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}})\boldsymbol{C}_p^{\mathrm{T}}\\ &= \boldsymbol{A}\cdot\mathrm{avar}\left\{\sqrt{n}\mathrm{vec}(\boldsymbol{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Gamma}})\right\}\cdot\boldsymbol{A}^{\mathrm{T}}.\end{aligned}$$

We see that

$$\begin{aligned}\boldsymbol{A} &= 2\boldsymbol{C}_p(\boldsymbol{\Gamma}\boldsymbol{\Omega}_1\otimes\boldsymbol{I}_p-\boldsymbol{\Gamma}\otimes\boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}})=\frac{\partial\mathrm{vech}(\boldsymbol{\Delta}_1)}{\partial\mathrm{vec}^{\mathrm{T}}(\boldsymbol{\Gamma})}\\ &= \left[\frac{\partial\mathrm{vech}\left\{\widehat{\boldsymbol{\Delta}}_{1,\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_0}(\widehat{\boldsymbol{\Gamma}})\right\}}{\partial\mathrm{vec}^{\mathrm{T}}(\widehat{\boldsymbol{\Gamma}})}\right]_{\widehat{\boldsymbol{\Gamma}}=\boldsymbol{\Gamma}}=\left[\frac{\partial\mathrm{vech}\left\{\widehat{\boldsymbol{\Delta}}_{1,\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_0}(\boldsymbol{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Gamma}})\right\}}{\partial\mathrm{vec}^{\mathrm{T}}(\boldsymbol{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Gamma}})}\right]_{\widehat{\boldsymbol{\Gamma}}=\boldsymbol{\Gamma}},\end{aligned}$$

where

$$\widehat{\boldsymbol{\Delta}}_{1,\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_0}(\boldsymbol{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Gamma}})=\boldsymbol{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Gamma}}\boldsymbol{\Omega}_1\widehat{\boldsymbol{\Gamma}}^{\mathrm{T}}\boldsymbol{Q}_{\boldsymbol{\Gamma}}+\boldsymbol{P}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Gamma}}_0\boldsymbol{\Omega}_0\widehat{\boldsymbol{\Gamma}}_0^{\mathrm{T}}\boldsymbol{P}_{\boldsymbol{\Gamma}}=\boldsymbol{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Delta}}_{1,\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_0}\boldsymbol{Q}_{\boldsymbol{\Gamma}}+\boldsymbol{P}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Delta}}_{1,\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_0}\boldsymbol{P}_{\boldsymbol{\Gamma}}.$$

By the delta method, we have that

$$\boldsymbol{A}\cdot\mathrm{avar}\left\{\sqrt{n}\mathrm{vec}(\boldsymbol{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Gamma}})\right\}\cdot\boldsymbol{A}^{\mathrm{T}}=\mathrm{avar}\left\{\sqrt{n}\mathrm{vech}(\widehat{\boldsymbol{\Delta}}_{1,\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_0}(\boldsymbol{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Gamma}})\right\}.$$

Henceforth,

$$\begin{aligned}\boldsymbol{H}_{1,11}(\boldsymbol{H}_{11}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_{11})^{\dagger}\boldsymbol{H}_{1,11}^{\mathrm{T}} &= \mathrm{avar}\left\{\sqrt{n}\mathrm{vech}(\boldsymbol{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Delta}}_{1,\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_0}\boldsymbol{Q}_{\boldsymbol{\Gamma}}+\boldsymbol{P}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Delta}}_{1,\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_0}\boldsymbol{P}_{\boldsymbol{\Gamma}})\right\}\\ &= \mathrm{avar}\left\{\sqrt{n}\mathrm{vech}(\boldsymbol{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Delta}}_{1,\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_0}\boldsymbol{Q}_{\boldsymbol{\Gamma}})\right\}+\mathrm{avar}\left\{\sqrt{n}\mathrm{vech}(\boldsymbol{P}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Delta}}_{1,\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_0}\boldsymbol{P}_{\boldsymbol{\Gamma}})\right\}.\end{aligned}$$

Then the result in (8) of Theorem 2 follows.

Next we consider

$$\widehat{\boldsymbol{g}}_{\mathrm{crs}}=\mathrm{vech}(\widehat{\boldsymbol{\Delta}}_1)-\mathrm{vech}(\widehat{\boldsymbol{\Delta}}_2)=\boldsymbol{G}\begin{pmatrix}\mathrm{vech}\,\widehat{\boldsymbol{\Delta}}_1\\ \mathrm{vech}\,\widehat{\boldsymbol{\Delta}}_2\end{pmatrix},$$

where $\boldsymbol{G}=\begin{pmatrix}\boldsymbol{I}_{p(p+1)/2} & -\boldsymbol{I}_{p(p+1)/2}\end{pmatrix}$. Then, we have $\mathrm{avar}(\sqrt{n}\widehat{\boldsymbol{g}}_{\mathrm{crs}})=\boldsymbol{G}\boldsymbol{U}\boldsymbol{G}^{\mathrm{T}}$, and

$$\boldsymbol{U}=\boldsymbol{H}_1(\boldsymbol{H}_1^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_1)^{\dagger}\boldsymbol{H}_1^{\mathrm{T}}=\begin{pmatrix}\boldsymbol{U}_1 & \boldsymbol{U}_2\\ \boldsymbol{U}_2^{\mathrm{T}} & \boldsymbol{U}_3,\end{pmatrix},$$

following the proof of Theorem 1. Since $\boldsymbol{H}_1^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_1$ is block-diagonal, we can see that $(\boldsymbol{H}_1^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_1)^{\dagger}$ is a block-diagonal matrix with blocks $\boldsymbol{A}_j^{\dagger}=(\boldsymbol{H}_{1j}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_{1j})^{\dagger}$, $j=1,2,3,4$. Then by straightforward calculation, we have

$$\begin{aligned}\boldsymbol{U}_1 &= \boldsymbol{H}_{1,11}\boldsymbol{A}_1^{\dagger}\boldsymbol{H}_{1,11}^{\mathrm{T}}+\boldsymbol{H}_{1,12}\boldsymbol{A}_2^{\dagger}\boldsymbol{H}_{1,12}^{\mathrm{T}}+\boldsymbol{H}_{1,14}\boldsymbol{A}_4^{\dagger}\boldsymbol{H}_{1,14}^{\mathrm{T}},\\ \boldsymbol{U}_2 &= \boldsymbol{H}_{1,11}\boldsymbol{A}_1^{\dagger}\boldsymbol{H}_{1,21}^{\mathrm{T}}+\boldsymbol{H}_{1,14}\boldsymbol{A}_4^{\dagger}\boldsymbol{H}_{1,14}^{\mathrm{T}},\\ \boldsymbol{U}_3 &= \boldsymbol{H}_{1,21}\boldsymbol{A}_1^{\dagger}\boldsymbol{H}_{1,21}^{\mathrm{T}}+\boldsymbol{H}_{1,23}\boldsymbol{A}_3^{\dagger}\boldsymbol{H}_{1,23}^{\mathrm{T}}+\boldsymbol{H}_{1,24}\boldsymbol{A}_4^{\dagger}\boldsymbol{H}_{1,24}^{\mathrm{T}}.\end{aligned}$$

$$
\begin{aligned}
\boldsymbol{GUG}^{\mathrm{T}} &= \boldsymbol{H}_{1,11}\boldsymbol{A}_1^{\dagger}\boldsymbol{H}_{1,11}^{\mathrm{T}} + \boldsymbol{H}_{1,12}\boldsymbol{A}_2^{\dagger}\boldsymbol{H}_{1,12}^{\mathrm{T}} - \boldsymbol{H}_{1,21}\boldsymbol{A}_1^{\dagger}\boldsymbol{H}_{1,11}^{\mathrm{T}} - \boldsymbol{H}_{1,11}\boldsymbol{A}_1^{\dagger}\boldsymbol{H}_{1,21}^{\mathrm{T}} \\
&+ \boldsymbol{H}_{1,21}\boldsymbol{A}_1^{\dagger}\boldsymbol{H}_{1,21}^{\mathrm{T}} + \boldsymbol{H}_{1,23}\boldsymbol{A}_3^{\dagger}\boldsymbol{H}_{1,23}^{\mathrm{T}}.
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
\boldsymbol{H}_{1,11}\boldsymbol{A}_1^{\dagger}\boldsymbol{H}_{1,11}^{\mathrm{T}} &= 4\boldsymbol{C}_p(\boldsymbol{\Gamma}\boldsymbol{\Omega}_1 \otimes \boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0)\left(\boldsymbol{H}_{11}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_{11}\right)^{\dagger}(\boldsymbol{\Omega}_1\boldsymbol{\Gamma}^{\mathrm{T}} \otimes \boldsymbol{\Gamma}_0^{\mathrm{T}} - \boldsymbol{\Gamma}^{\mathrm{T}} \otimes \boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}})\boldsymbol{C}_p^{\mathrm{T}}, \\
\boldsymbol{H}_{1,12}\boldsymbol{A}_2^{\dagger}\boldsymbol{H}_{1,12}^{\mathrm{T}} &= \boldsymbol{C}_p(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})\boldsymbol{E}_u\left\{\frac{d\pi_1}{2}(\boldsymbol{E}_u^{\mathrm{T}}(\boldsymbol{\Omega}_1^{-1} \otimes \boldsymbol{\Omega}_1^{-1})\boldsymbol{E}_u)\right\}^{\dagger}\boldsymbol{E}_u^{\mathrm{T}}(\boldsymbol{\Gamma}^{\mathrm{T}} \otimes \boldsymbol{\Gamma}^{\mathrm{T}})\boldsymbol{C}_p^{\mathrm{T}}, \\
\boldsymbol{H}_{1,23}\boldsymbol{A}_3^{\dagger}\boldsymbol{H}_{1,23}^{\mathrm{T}} &= \boldsymbol{C}_p(\boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma})\boldsymbol{E}_u\left\{\frac{d\pi_2}{2}(\boldsymbol{E}_u^{\mathrm{T}}(\boldsymbol{\Omega}_2^{-1} \otimes \boldsymbol{\Omega}_2^{-1})\boldsymbol{E}_u)\right\}^{\dagger}\boldsymbol{E}_u^{\mathrm{T}}(\boldsymbol{\Gamma}^{\mathrm{T}} \otimes \boldsymbol{\Gamma}^{\mathrm{T}})\boldsymbol{C}_p^{\mathrm{T}}, \\
\boldsymbol{H}_{1,21}\boldsymbol{A}_1^{\dagger}\boldsymbol{H}_{1,11}^{\mathrm{T}} &= 4\boldsymbol{C}_p(\boldsymbol{\Gamma}\boldsymbol{\Omega}_2 \otimes \boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0)\left[\boldsymbol{H}_{11}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_{11}\right]^{\dagger}(\boldsymbol{\Omega}_1\boldsymbol{\Gamma}^{\mathrm{T}} \otimes \boldsymbol{\Gamma}_0^{\mathrm{T}} - \boldsymbol{\Gamma}^{\mathrm{T}} \otimes \boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}})\boldsymbol{C}_p^{\mathrm{T}}, \\
\boldsymbol{H}_{1,21}\boldsymbol{A}_1^{\dagger}\boldsymbol{H}_{1,21}^{\mathrm{T}} &= 4\boldsymbol{C}_p(\boldsymbol{\Gamma}\boldsymbol{\Omega}_2 \otimes \boldsymbol{\Gamma}_0 - \boldsymbol{\Gamma} \otimes \boldsymbol{\Gamma}_0\boldsymbol{\Omega}_0)\left[\boldsymbol{H}_{11}^{\mathrm{T}}\boldsymbol{J}_h\boldsymbol{H}_{11}\right]^{\dagger}(\boldsymbol{\Omega}_2\boldsymbol{\Gamma}^{\mathrm{T}} \otimes \boldsymbol{\Gamma}_0^{\mathrm{T}} - \boldsymbol{\Gamma}^{\mathrm{T}} \otimes \boldsymbol{\Omega}_0\boldsymbol{\Gamma}_0^{\mathrm{T}})\boldsymbol{C}_p^{\mathrm{T}}.
\end{aligned}
$$

As shown earlier, we have

$$
\begin{aligned}
\boldsymbol{H}_{1,12}\boldsymbol{A}_2^{\dagger}\boldsymbol{H}_{1,12}^{\mathrm{T}} + \boldsymbol{H}_{1,23}\boldsymbol{A}_3^{\dagger}\boldsymbol{H}_{1,23}^{\mathrm{T}} &= \text{avar}\left\{\sqrt{n}\text{vech}(\boldsymbol{\Gamma}\widehat{\boldsymbol{\Omega}}_{1,\boldsymbol{\Gamma}}\boldsymbol{\Gamma}^{\mathrm{T}})\right\} + \text{avar}\left\{\sqrt{n}\text{vech}(\boldsymbol{\Gamma}\widehat{\boldsymbol{\Omega}}_{2,\boldsymbol{\Gamma}}\boldsymbol{\Gamma}^{\mathrm{T}})\right\} \\
&= \text{avar}(\sqrt{n}\text{vech}\left\{(\widehat{\boldsymbol{\Delta}}_{1,\boldsymbol{\Gamma}} - \widehat{\boldsymbol{\Delta}}_{2,\boldsymbol{\Gamma}})\right\}.
\end{aligned}
$$

In addition,

$$
\begin{aligned}
&\boldsymbol{H}_{1,11}\boldsymbol{A}_1^{\dagger}\boldsymbol{H}_{1,11}^{\mathrm{T}} - \boldsymbol{H}_{1,11}\boldsymbol{A}_1^{\dagger}\boldsymbol{H}_{1,21}^{\mathrm{T}} + \boldsymbol{H}_{1,21}\boldsymbol{A}_1^{\dagger}\boldsymbol{H}_{1,21}^{\mathrm{T}} - \boldsymbol{H}_{1,21}\boldsymbol{A}_1^{\dagger}\boldsymbol{H}_{1,11}^{\mathrm{T}} \\
= \;& 4\boldsymbol{C}_p\left\{\boldsymbol{\Gamma}(\boldsymbol{\Omega}_1 - \boldsymbol{\Omega}_2) \otimes \boldsymbol{I}_p\right\}(\boldsymbol{I}_u \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}}) \times \text{avar}\left\{\sqrt{n}\text{vec}(\widehat{\boldsymbol{\Gamma}})\right\} \times (\boldsymbol{I}_u \otimes \boldsymbol{Q}_{\boldsymbol{\Gamma}})\left\{(\boldsymbol{\Omega}_1 - \boldsymbol{\Omega}_2)\boldsymbol{\Gamma}^{\mathrm{T}} \otimes \boldsymbol{I}_p\right\} \\
= \;& \boldsymbol{B} \cdot \text{avar}\left\{\sqrt{n}\text{vec}(\boldsymbol{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Gamma}})\right\} \cdot \boldsymbol{B}^{\mathrm{T}},
\end{aligned}
$$

where $\boldsymbol{B} = 2\boldsymbol{C}_p\left[\boldsymbol{\Gamma}(\boldsymbol{\Omega}_1 - \boldsymbol{\Omega}_2) \otimes \boldsymbol{I}_p\right]$. Define $\boldsymbol{\Delta}_{\text{diff}} = \boldsymbol{\Delta}_1 - \boldsymbol{\Delta}_2$, we have

$$
\boldsymbol{B} = \frac{\partial\text{vech}(\boldsymbol{\Delta}_{\text{diff}})}{\partial\text{vec}^{\mathrm{T}}(\boldsymbol{\Gamma})} = \left[\frac{\partial\text{vech}\left\{\widehat{\boldsymbol{\Delta}}_{\text{diff},\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_2}(\widehat{\boldsymbol{\Gamma}})\right\}}{\partial\text{vec}^{\mathrm{T}}(\widehat{\boldsymbol{\Gamma}})}\right]_{\widehat{\boldsymbol{\Gamma}}=\boldsymbol{\Gamma}} = \left[\frac{\partial\text{vech}\left\{\widehat{\boldsymbol{\Delta}}_{\text{diff},\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_2}(\boldsymbol{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Gamma}})\right\}}{\partial\text{vec}^{\mathrm{T}}(\boldsymbol{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Gamma}})}\right]_{\widehat{\boldsymbol{\Gamma}}=\boldsymbol{\Gamma}},
$$

where $\widehat{\boldsymbol{\Delta}}_{\text{diff},\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_2}(\boldsymbol{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Gamma}}) = \boldsymbol{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Gamma}}(\boldsymbol{\Omega}_1 - \boldsymbol{\Omega}_2)\widehat{\boldsymbol{\Gamma}}^{\mathrm{T}}\boldsymbol{Q}_{\boldsymbol{\Gamma}} = \boldsymbol{Q}_{\boldsymbol{\Gamma}}(\widehat{\boldsymbol{\Delta}}_{1,\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_0} - \widehat{\boldsymbol{\Delta}}_{2,\boldsymbol{\Omega}_2,\boldsymbol{\Omega}_0})\boldsymbol{Q}_{\boldsymbol{\Gamma}}$. Again by the delta method, we have that

$$
\begin{aligned}
\boldsymbol{B} \cdot \text{avar}(\sqrt{n}\text{vec}(\boldsymbol{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Gamma}})) \cdot \boldsymbol{B}^{\mathrm{T}} &= \text{avar}\left\{\sqrt{n}\text{vech}(\widehat{\boldsymbol{\Delta}}_{\text{diff},\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_2}(\boldsymbol{Q}_{\boldsymbol{\Gamma}}\widehat{\boldsymbol{\Gamma}})\right\} \\
&= \text{avar}\left\{\sqrt{n}\text{vech}(\boldsymbol{Q}_{\boldsymbol{\Gamma}}(\widehat{\boldsymbol{\Delta}}_{1,\boldsymbol{\Omega}_1,\boldsymbol{\Omega}_0} - \widehat{\boldsymbol{\Delta}}_{2,\boldsymbol{\Omega}_2,\boldsymbol{\Omega}_0})\boldsymbol{Q}_{\boldsymbol{\Gamma}})\right\}
\end{aligned}
$$

Henceforth, the result in (9) of Theorem 2 follows. This completes the proof of Theorem 2. □

A3.6 Proof of Theorem 3

Proof. Recall that

$$
\ell(\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2) = -\sum_{k=1}^{2}\frac{n_k d}{2}\left\{\log|\boldsymbol{\Delta}_k| + \text{tr}(\boldsymbol{\Delta}_k^{-1}\overline{\boldsymbol{D}}_k)\right\}.
$$

We define a minimum discrepancy function,

$$\begin{aligned}\boldsymbol{F}(\boldsymbol{h},\widehat{\boldsymbol{h}}) &= -\frac{2}{n}\left[\ell(\boldsymbol{\Delta}_1,\boldsymbol{\Delta}_2)-\ell(\overline{\boldsymbol{D}}_1,\overline{\boldsymbol{D}}_2)\right] \\ &= \sum_{k=1}^{2}\pi_k d\left\{\log|\boldsymbol{\Delta}_k|+\mathrm{tr}(\boldsymbol{\Delta}_k^{-1}\overline{\boldsymbol{D}}_k)-\log|\overline{\boldsymbol{D}}_k|-\mathrm{tr}(\overline{\boldsymbol{D}}_k^{-1}\overline{\boldsymbol{D}}_k)\right\},\end{aligned}$$

with $\boldsymbol{h}=\{\boldsymbol{\Delta}_1(\boldsymbol{\theta}),\boldsymbol{\Delta}_2(\boldsymbol{\theta})\}$, and $\widehat{\boldsymbol{h}}=(\overline{\boldsymbol{D}}_1,\overline{\boldsymbol{D}}_2)$.

Then the Fisher information is calculated as

$$\boldsymbol{J}_h=\left\{\frac{1}{2}\frac{\partial\boldsymbol{F}(\boldsymbol{h},\widehat{\boldsymbol{h}})}{\partial\boldsymbol{h}\partial\boldsymbol{h}^{\mathrm{T}}}\right\}_{\widehat{\boldsymbol{h}}=\boldsymbol{h}=\boldsymbol{h}_{\text{true}}}.$$

Also, we have that: (a) $\boldsymbol{F}(\boldsymbol{h},\widehat{\boldsymbol{h}})\geq 0$ for all $\boldsymbol{h}$ and $\widehat{\boldsymbol{h}}$, (b) $\boldsymbol{F}(\boldsymbol{h},\widehat{\boldsymbol{h}})=0$ if and only if $\widehat{\boldsymbol{h}}=\boldsymbol{h}$, and (c) $\boldsymbol{F}(\boldsymbol{h},\widehat{\boldsymbol{h}})$ is twice continously differentiable in $\boldsymbol{h}$ and $\widehat{\boldsymbol{h}}$. Since $\overline{\boldsymbol{D}}_k$ is the sample covariance matrix, by the Central Limit Theorem, we have avar$\left\{\sqrt{n}\mathrm{vech}(\widehat{\boldsymbol{h}})\right\}=\boldsymbol{W}$, for some positive definite covariance matrix $\boldsymbol{W}$. Then by Proposition 4.1 of Shapiro (1986), we have the consistency result of $\widehat{\boldsymbol{h}}_{\text{crs}}$.

Moreover, we have

$$\begin{aligned}\boldsymbol{W}-\boldsymbol{K} &= \boldsymbol{W}-\boldsymbol{J}_h^{-\frac{1}{2}}\boldsymbol{P}_{\boldsymbol{J}_h^{\frac{1}{2}}\boldsymbol{H}}\boldsymbol{J}_h^{\frac{1}{2}}\boldsymbol{W}\boldsymbol{J}_h^{\frac{1}{2}}\boldsymbol{P}_{\boldsymbol{J}_h^{\frac{1}{2}}\boldsymbol{H}}\boldsymbol{J}_h^{-\frac{1}{2}} \\ &= \boldsymbol{J}_h^{-\frac{1}{2}}\left[\boldsymbol{J}_h^{\frac{1}{2}}\boldsymbol{W}\boldsymbol{J}_h^{\frac{1}{2}}-\boldsymbol{P}_{\boldsymbol{J}_h^{\frac{1}{2}}\boldsymbol{H}}\boldsymbol{J}_h^{\frac{1}{2}}\boldsymbol{W}\boldsymbol{J}_h^{\frac{1}{2}}\boldsymbol{P}_{\boldsymbol{J}_h^{\frac{1}{2}}\boldsymbol{H}}\right]\boldsymbol{J}_h^{-\frac{1}{2}}.\end{aligned}$$

This completes the proof of Theorem 3. □

A4 Computer Code

The computer code associated with this paper is available at the Biometrics website on Wiley Online Library.

References

Cook, R. D. and Forzani, L. (2008). Covariance reducing models: An alternative to spectral modelling of covariance matrices. *Biometrika*, 95(4):799–812.

Cook, R. D., Li, B., and Chiaromonte, F. (2010). Envelope models for parsimonious and efficient multivariate linear regression. *Statist. Sinica*, 20(3):927–960.

Cook, R. D. and Zhang, X. (2016). Algorithms for envelope estimation. *Journal of Computational and Graphical Statistics*, 25(1):284–300.

Danaher, P., Wang, P., and Witten, D. M. (2014). The joint graphical lasso for inverse covariance estimation across multiple classes. *Journal of the Royal Statistical Society: Series B.*, 76(2):373–397.

Franks, A. and Hoff, P. (2016). Shared subspace models for multi-group covariance estimation. *arXiv preprint arXiv:1607.03045*.

Henderson, H. V. and Searle, S. (1979). Vec and vech operators for matrices, with some uses in jacobians and multivariate statistics. *Canadian Journal of Statistics*, 7(1):65–81.

Lee, W. and Liu, Y. (2015). Joint estimation of multiple precision matrices with common structures. *The Journal of Machine Learning Research*, 16(1):1035–1062.

Liu, W. and Luo, X. (2015). Fast and adaptive sparse precision matrix estimation in high dimensions. *Journal of Multivariate Analysis*, 135:153–162.

Shapiro, A. (1986). Asymptotic theory of overparameterized structural models. *Journal of the American Statistical Association*, 81(393):142–149.