Distribution-Independent Confidence Intervals for the Eigendecomposition of Covariance Matrices via the Eigenvalue-Eigenvector Identity

Teodora Popordanoska[1], Aleksei Tiulpin[2], Wacha Bounliphone[3], and Matthew B. Blaschko[1]

[1] ESAT-PSI, KU Leuven, Belgium, firstname.lastname@kuleuven.be
[2] Dept. of Computer Science, Aalto University, Finland, aleksei.tiulpin@aalto.fi
[3] IPSOS, Social Intelligence Analytics, Paris, France, wacha.bounliphone@gmail.com

Abstract. The eigendecomposition of a matrix is the central procedure in probabilistic models based on matrix factorization, for instance PCA and topic models. Quantifying the uncertainty of such a decomposition based on a finite sample estimate is essential to reasoning under uncertainty when employing such models. This paper tackles the challenge of computing confidence bounds on the individual entries of eigenvectors of a covariance matrix of fixed dimension. The assumptions behind our method are minimal and require that the covariance matrix exists, and its empirical estimator converges to the true covariance. We make use of the theory of U-statistics to probabilistically bound the L_2 perturbation of the empirical covariance matrix. Subsequently, we leverage Weyl's theorem and the recently introduced eigenvalue-eigenvector identity [5] to probabilistically bound the eigenvectors via the bounds on eigenvalues. We demonstrate our theoretical results on real-world data from medical and physics domains.

1 Introduction

Estimating confidence intervals on the eigendecomposition of an empirical covariance matrix $\hat{\Sigma}$ is significant in various applications. Given a finite sample $\{x_i\}_{i=1}^n$ drawn i.i.d. from some unknown distribution with $x_i \in \mathbb{R}^p$, the task is to find confidence intervals on the individual entries of the eigenvectors u_i such that $\Sigma u_i = \lambda_i u_i$, Σ being the *population* covariance matrix to which the empirical estimate $\hat{\Sigma}$ converges as n goes to infinity. That such confidence intervals are possible is due to the recently codified eigenvalue-eigenvector identity [5], a fundamental relationship between eigenvalues and individual entries of eigenvectors.

In this article, we propose a novel, computationally efficient framework for estimating the bounds on the eigenvalues and eigenvectors. We focus on the case when $n \gg p$, and in particular ensure that the procedure has computational complexity *linear* in n, while making two *minimal* distributional assumptions:

1. The covariance matrix Σ exists.

2. An unbiased estimator $\hat{\Sigma}$ converges to Σ.

The proposed intervals are asymptotically consistent without the need to set a computationally expensive bootstrap procedure or any hyperparameters.

Multiple studies focused on studying the limiting behavior of eigenvalues and eigenvectors of covariance matrices. As such, [19] established a convergence rate of the eigenvector empirical spectral distribution, which was recently improved by [18], assuming a finite 8^{th} moment condition of the underlying data distribution. Prior to that line of work, [8] quantified the deviation of the sample eigenvectors from their population values by generalizing the Marčenko-Pastur equation [11] and assuming 12^{th} finite moment and extra constraints. Additional results with similar assumptions can be found in [2,14,16,15].

For obtaining error bounds on the entries of an eigendecomposition, the only available distribution-independent method until now, to the best of our knowledge, is a bootstrap technique [6], which becomes computationally expensive in practice when n is large.

2 Bounding the Perturbation of the Covariance Matrix

2.1 A U-statistic Estimator of the Cross-Covariance

Most of the materials in this section can be found in [7], [13, Chap. 5], [10, Chap. 6] and [9]. Suppose we have a sample $\mathrm{X_q} = \{X_1, ..., X_q\}$ of size q drawn i.i.d. from a distribution P_X. A U-statistic concerns an unbiased estimator of a parameter θ of P_X using $\mathrm{X_q}$. That is, θ may be represented as

$$\theta = \mathrm{E}[h(X_1, ..., X_q)] \tag{1}$$

for some function h, called a kernel of order q.

Definition 1 *(U-statistic, [13, Chap. 5]) Given a kernel h of order q and a sample $\mathrm{X_n} = \{X_1, ..., X_n\}$ of size $n \geq q$, the corresponding U-statistic for estimation of θ is obtained by averaging the kernel h symetrically over the observations:*

$$\hat{U} := \frac{1}{(n)_q} \sum_{i_q^n} h(X_{i_1}, ..., X_{i_q}) \tag{2}$$

where the summation ranges over the set i_q^n of all $\dfrac{n!}{(n-q)!}$ permutations of size q chosen from $(1, ..., n)$ and $(n)_q$ is the Pochhammer symbol $(n)_q := \dfrac{n!}{(n-q)!}$.

Definition 2 *(U-statistic estimator of the covariance) Let $u_r = (X_{i_r}, X_{j_r})^T$ be ordered pairs of samples, with $1 \leq r \leq n$ and $1 \leq i, j \leq p$. Consider $\Sigma = \mathrm{Cov}(X_i, X_j)$, the covariance functional between X_i and X_j and h, the kernel of order 2 for the functional Σ such that*

$$h(u_1, u_2) = \frac{1}{2}(X_{i_1} - X_{i_2})(X_{j_1} - X_{j_2}). \tag{3}$$

The corresponding U-statistic estimator of the covariance Σ is

$$\hat{\Sigma} = \frac{1}{n-1} \sum_{r=1}^{n} (X_{i_r} - \bar{X}_i)(X_{j_r} - \bar{X}_j), \tag{4}$$

where $\bar{X}_i = \frac{1}{n} \sum_{s=1}^{n} X_{is}$. $\hat{\Sigma}$ can be computed in linear time.

2.2 U-statistic based convergence of the empirical covariance matrix

We focus here on U-statistic estimates of $\hat{\Sigma}$ and its asymptotic normal distribution. We develop the full covariance between the elements of $\hat{\Sigma}$, which we denote $\mathrm{Cov}(\hat{\Sigma}) \in \mathbb{R}^{\frac{p(p+1)}{2} \times \frac{p(p+1)}{2}}$, where the size is due to the symmetry of $\hat{\Sigma}$. We denote $U(A)$ the function returning the upper triangular part and diagonal of a matrix A.

Theorem 1. *(Joint asymptotic normality distribution of the covariance matrix, [7]) For all (i, j, k, l) range over each of the p variates in a covariance matrix $\hat{\Sigma}$, if $\mathrm{Var}(\hat{\Sigma}_{ij}) > 0$ and $\mathrm{Var}(\hat{\Sigma}_{kl}) > 0$, then $\left[\hat{\Sigma}_{ij}; \hat{\Sigma}_{kl}\right]^T$ converges in distribution (as $n \to \infty$) to a Gaussian random variable*

$$n^{\frac{1}{2}} \begin{pmatrix} \hat{\Sigma}_{ij} - \Sigma_{ij} \\ \hat{\Sigma}_{kl} - \Sigma_{kl} \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, K, \right), \tag{5}$$

where

$$K = \begin{pmatrix} \mathrm{Var}(\hat{\Sigma}_{ij}) & \mathrm{Cov}(\hat{\Sigma}_{ij}, \hat{\Sigma}_{kl}) \\ \mathrm{Cov}(\hat{\Sigma}_{ij}, \hat{\Sigma}_{kl}) & \mathrm{Var}(\hat{\Sigma}_{kl}) \end{pmatrix} \tag{6}$$

We note respectively h and g the corresponding kernels of order 2 for the two unbiased estimates $\hat{\Sigma}_{ij}$ and $\hat{\Sigma}_{kl}$, where $h(u_1, u_2) = \dfrac{1}{2}(X_{i_1} - X_{i_2})(X_{j_1} - X_{j_2})$ with $u_r = (X_{i_r},\ X_{j_r})^T$ and $g(v_1, v_2) = \dfrac{1}{2}(X_{k_1} - X_{k_2})(X_{l_1} - X_{l_2})$ with $v_r = (X_{k_r}, X_{l_r})^T$. Then, we state the following theorem.

Theorem 2. *(Covariance of the U-statistic for the covariance matrix) The low variance, unbiased estimates of the covariance between two U-statistics estimates $\hat{\Sigma}_{ij}$ and $\hat{\Sigma}_{kl}$, where (i, j, k, l) range over each of the p variates in a covariance matrix of $\hat{\Sigma}$ $\mathrm{Cov}(\hat{\Sigma}) := \mathrm{Cov}(\hat{\Sigma}_{ij}, \hat{\Sigma}_{kl})$ is*

$$\mathrm{Cov}(\hat{\Sigma}_{ij}, \hat{\Sigma}_{kl}) = \binom{n}{2}^{-1} (2(n-2)\zeta_1) + \mathcal{O}(n^{-2}), \tag{7}$$

where $\zeta_1 = \mathrm{Cov}\left(\mathrm{E}_{u_2}[h(u_1, u_2)], \mathrm{E}_{v_2}[g(v_1, v_2)]\right)$. There are seven exhaustive cases, derived in Appendix A.3, which are used to estimate Eq. (7) for all $1 \leq i, j, k, l \leq p$ through simple variable substitution. Each of these cases has computation linear in n.

Lemma 1. *With probability at least* $1-\delta$*, we have the two following inequalities*

$$\|\Sigma - \hat{\Sigma}\|_2 \leq \sqrt{2\lambda_{\max}}\Phi^{-1}\left(1-\delta/2\right) \tag{8}$$

$$\leq \sqrt{2\,\mathrm{Tr}[\mathrm{Cov}(\hat{\Sigma})]}\Phi^{-1}\left(1-\delta/2\right) \tag{9}$$

where $\Phi(\cdot)$ *is the CDF of* $\mathcal{N}(0,1)$ *and* $\lambda_{\max}$ *is the largest eigenvalue of* $\mathrm{Cov}(\hat{\Sigma})$.

Proof. As $\hat{\Sigma}$ is a U-statistic, we have that $U(\hat{\Sigma})$, a vector containing its upper diagonal component (including the diagonal), is Gaussian distributed with covariance $\mathrm{Cov}(\hat{\Sigma})$ (cf. Thm 1, 2). Therefore, with probability at least $1-\delta$,

$$\|U(\Sigma) - U(\hat{\Sigma})\|_2 \leq \sqrt{\lambda_{\max}}\Phi^{-1}\left(1-\delta/2\right) \tag{10}$$

and furthermore

$$\|\Sigma - \hat{\Sigma}\|_F \leq \sqrt{2}\|U(\Sigma) - U(\hat{\Sigma})\|_2 \tag{11}$$

which combined with the fact that $\|\cdot\|_2 \leq \|\cdot\|_F$ yields the desired result.

3 Bounding the Eigenvectors of the Covariance Matrix

Theorem 3 (Eigenvalue-eigenvector identity [5]). *Let* Σ *be a* $p \times p$ *Hermitian matrix. Denote by* V_{ij} *the* j*th component of the* i*th eigenvector of* Σ*, and* $\lambda_i(\Sigma)$ *be the* i*th eigenvalue. The following identity holds:*

$$|V_{ij}|^2 \prod_{k=1;k\neq i}^{p} \left(\lambda_i(\Sigma) - \lambda_k(\Sigma)\right) = \prod_{k=1}^{p-1} \left(\lambda_i(\Sigma) - \lambda_k(M_j)\right), \tag{12}$$

where M_j *is the* $p-1 \times p-1$ *minor formed from* Σ *by deleting its* j*th row and column.*

Assume that we have a perturbed observation $\hat{\Sigma}$ satisfying $\|\hat{\Sigma} - \Sigma\|_2 \leq \varepsilon$ with high probability for some known ε. We may construct such a ε, e.g. when estimating a covariance matrix from a finite sample.

Theorem 4 (Weyl's Theorem [17]). *For two positive definite matrices* $\hat{\Sigma}$ *and* Σ*, if*

$$|\lambda_k(\hat{\Sigma}) - \lambda_k(\Sigma)| \leq \|\hat{\Sigma} - \Sigma\|_2 \leq \varepsilon \tag{13}$$

where $0 < \varepsilon < \lambda_k(\Sigma)\ \forall k$*, then*

$$\lambda_k(\hat{\Sigma}) - \varepsilon \leq \lambda_k(\Sigma) \leq \lambda_k(\hat{\Sigma}) + \varepsilon\ \forall k. \tag{14}$$

Proposition 1. *Assuming that $\|\hat{\Sigma} - \Sigma\|_2 \leq \varepsilon$ and the fact that the eigenvectors are orthonormal, it follows that*

$$|V_{ij}|^2 \leq \begin{cases} 1 & \text{if } \exists k \neq i : |\lambda_i(\hat{\Sigma}) - \lambda_k(\hat{\Sigma})| \leq 2\varepsilon \\ \min(\alpha_{ij}, 1) & \text{otherwise,} \end{cases} \tag{15}$$

and

$$|V_{ij}|^2 \geq \begin{cases} 0 & \text{if } \exists k : |\lambda_i(\hat{\Sigma}) - \lambda_k(\hat{M}_j)| \leq 2\varepsilon \\ \beta_{ij} & \text{otherwise,} \end{cases} \tag{16}$$

where

$$\alpha_{ij} = \frac{\prod_{k=1}^{p-1} \left(|\lambda_i(\hat{\Sigma}) - \lambda_k(\hat{M}_j)| + 2\varepsilon \right)}{\prod_{k=1;k\neq i}^{p} \left(|\lambda_i(\hat{\Sigma}) - \lambda_k(\hat{\Sigma})| - 2\varepsilon \right)} \tag{17}$$

$$\beta_{ij} = \frac{\prod_{k=1}^{p-1} \left(|\lambda_i(\hat{\Sigma}) - \lambda_k(\hat{M}_j)| - 2\varepsilon \right)}{\prod_{k=1;k\neq i}^{p} \left(|\lambda_i(\Sigma) - \lambda_k(\Sigma)| + 2\varepsilon \right)} \tag{18}$$

Proof. We first note that $\|\hat{M}_j - M_j\|_2 \leq \|\hat{\Sigma} - \Sigma\|_2$, where $\hat{M}_j$ denotes the corresponding minor of $\hat{\Sigma}$, which follows directly from the subadditivity of the norm. This indicates we can also use ε to bound the perturbation on the eigenvalues of $\hat{M}_j$, though it would be possible to compute a slightly tighter bound on $|V_{ij}|^2$ at the expense of extra computation for a tighter bound on $\|\hat{M}_j - M_j\|_2$.

We subsequently note that the quantity to be bounded is non-negative, which allows us to replace all eigenvalue differences with the absolute value. The rest of the inequality follows by a simple application of interval arithmetic [12] based on eigenvalue bounds from Weyl's theorem.

Corollary 1. *If the lower bound on $|V_{ij}|^2$ is greater than zero, we have that the sign of V_{ij} is equal to the sign of $\hat{V}_{ij}$ and we can recover an upper and lower bound on V_{ij} directly. Otherwise, we conclude that the signed lower bound is the negative of the unsigned upper bound.*

4 Experiments

We show empirical results for producing confidence intervals of the eigendecomposition on real-world data from the medical and physics domain, described below.

4.1 Datasets

Knee osteoarthritis We used the data from baselines of two publicly available follow-up patient cohorts – Osteoarthritis Initiative (OAI, $n_{subjects} = 4796$), and Multicenter Osteoarthritis Study (MOST, $n_{subjects} = 3026$). We applied our method for conducting an analysis for assessing the dependence between the symptoms and radiographic progression of knee osteoarthritis in the medial side of the joint.

We leveraged the gradings done according to the Osteoarthritis Research Society International (OARSI) grading atlas [1]. The following variables from the radiographic part of the OAI and MOST datasets were used:

1. Osteophytes (bone spures) severity in the tibia bone (OSTM).
2. Ostophytes severity in the femur (OSFM).

We also included the symptomatic assessments done according to the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) [4]. WOMAC allows for quantification of the patient's pain, and we included the total WOMAC score, calculated as a sum of all the scores of its subsections.

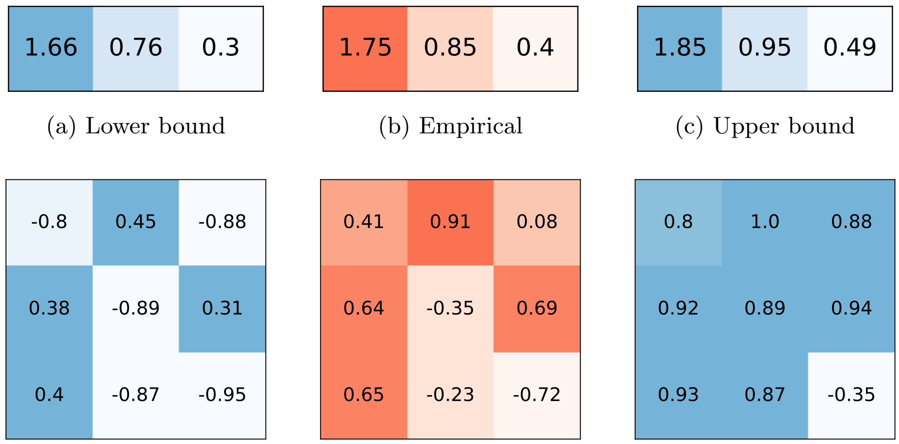
Higgs boson We consider the Higgs boson classification dataset from [3], containing 11 million observations. For this experiment we used the seven high-level features derived from the kinematic properties measured for the decay products after a particle collision occurs. Four of them, i.e., m_{jj}, m_{jjj}, m_{lv} and m_{jlv}, involve the observable decay products leptons l and jets, and the remaining three (m_{bb}, m_{wbb} and m_{wwbb}) are related to the generation of the Higgs bosons.

4.2 Results

We utilize the theoretical results presented above to bound the eigendecomposition of the covariance matrix for the two described experiments. The data is whitened as a pre-processing step. Figures 1 and 2 show that the obtained upper and lower bound on the empirical eigendecomposition are valid.

Application: Interpretation of the medical data In the case of the knee OA data (Figure 1), we observe the following relationships known in the medical literature using the eigenvectors and the bounds:

1. Osteoarthritis develops in both tibia and femur for most of the cases. This can be seen from the first eigenvector. Our method yields non-trivial bounds for the imaging features, and returns high uncertainty for the symptoms.
2. Symptoms (WOMAC) have very limited association with structural features of the disease (osteophytes denoted by OSTM and OSFM features). Here, we consider the second eigenvector: a non-trivial bound is obtained for the symptoms, and high uncertainty is seen for the imaging features.
3. There exist some small number of cases in the data, for which tibial and femoral medial OA do not develop simultaneously. We can see that the empirical value for the symptoms in the case of the third eigenvector is nearly zero, and it has high uncertainty. The imaging features, in contrast, have rather tight bounds of the same size.



(a) Lower bound (b) Empirical (c) Upper bound

(d) Lower bound (e) Empirical (f) Upper bound

Fig. 1: Bounds on the eigenvalues (first row) and eigenvectors (second row) for the Osteoarthritis data [WOMAC, OSTM, OSFM]

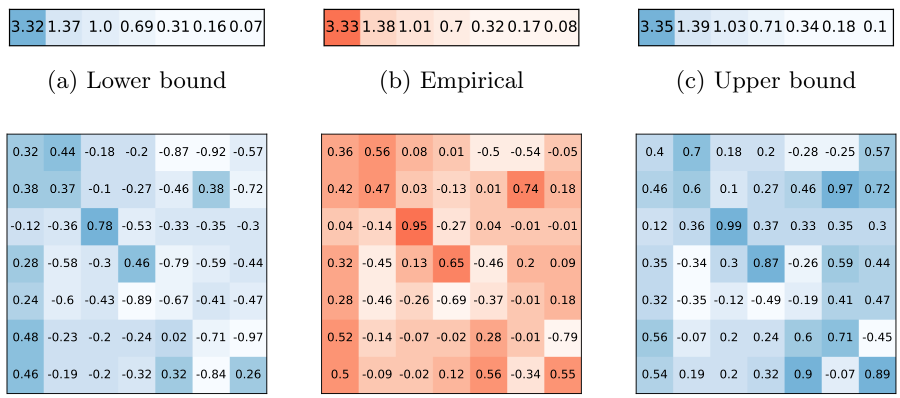
Scalability test The experiment with the Higgs boson dataset (Figure 2) shows that we can get non-trivial bounds on data with several million examples and multiple features. Furthermore, our unoptimized implementation of the method runs several orders of magnitude faster than a bootstrap procedure. We run all experiments on a CPU of a regular laptop.

5 Discussion and Conclusions

In this paper, we have presented a novel methodology for obtaining confidence intervals on the eigendecomposition of covariance matrices. This is done by first utilizing the theory of U-statistics to bound the perturbation of the empirical covariance matrix. We then use Weyl's theorem to obtain bounds on the eigenvalues and finally, we leverage the recently formalized eigenvalue-eigenvector identity to bound the eigenvectors.

The main strength of the proposed methodology is that it does not impose any restrictions on the data distribution and on the existence of finite higher-order moments. In fact, the only assumptions are that i) the covariance matrix exists and ii) an unbiased estimator converges to it. The conducted experiments verify the validity of the confidence intervals on the eigendecomposition.

Despite the benefits of the method, this paper has still some limitations. As such, the main downside of our approach is that, in practice, the number of samples has to be exponentially larger than the dimensionality of the data in order to obtain non-trivial bounds. Secondly, we did not fully explore the possibilities for tightening the bounds on the eigenvectors, however, we propose orthonormality



(a) Lower bound (b) Empirical (c) Upper bound

(d) Lower bound (e) Empirical (f) Upper bound

Fig. 2: Bounds on the eigenvalues (first row) and eigenvectors (second row) for the Higgs dataset [m_{jj}, m_{jjj}, m_{lv}, m_{jlv}, m_{bb}, m_{wbb} and m_{wwbb}].

constraints-based approach in Appendix B. We note that the method is consistent for all covariances with non-repeated eigenvalues, though possibly with poor convergence in the event that eigenvalues are closely spaced. This requirement is a direct consequence of the use of the eigenvalue-eigenvector identity, which is trivial in the case of repeated eigenvalues: Equation (12) degenerates to $0 = 0$ for the corresponding eigenvectors.

To conclude, we have proposed, to the best of our knowledge, the first application of the recently introduced eigenvalue-eigenvector identity to the estimation of confidence intervals over eigenvectors, resulting in a consistent estimator with computation linear in the number of samples. Furthermore, we have demonstrated its practical application in real applications, including particle physics and medicine. Source code of the method is available at `https://github.com/tpopordanoska/confidence-intervals`.

Acknowledgements

We acknowledge funding from the Flemish Government under the "Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen" programme.

References

1. Altman, R.D., Gold, G.: Atlas of individual radiographic features in osteoarthritis, revised. Osteoarthritis and cartilage **15**, A1–A56 (2007)
2. Bai, Z.D., Miao, B.Q., Pan, G.M.: On asymptotics of eigenvectors of large sample covariance matrix. The Annals of Probability **35**(4), 1532–1572 (2007), `http://www.jstor.org/stable/25450021`

3. Baldi, P., Sadowski, P., Whiteson, D.: Searching for exotic particles in high-energy physics with deep learning. Nature communications **5**(1), 1–9 (2014)
4. Bellamy, N.: Womac osteoarthritis index: a user's guide. London, Ontario **1995** (1995)
5. Denton, P.B., Parke, S.J., Tao, T., Zhang, X.: Eigenvectors from eigenvalues: a survey of a basic identity in linear algebra (2019)
6. Efron, B.: Bootstrap methods: another look at the jackknife. In: Breakthroughs in statistics, pp. 569–593. Springer (1992)
7. Hoeffding, W.: A class of statistics with asymptotically normal distribution. The Annals of Mathematical Statistics pp. 293–325 (1948)
8. Ledoit, O., Peche, S.: Eigenvectors of some large sample covariance matrix ensembles. Probability Theory and Related Fields **151** (11 2009). https://doi.org/10.1007/s00440-010-0298-3
9. Lee, A.J.: U-statistics: Theory and practice. CRC Press (1990)
10. Lehmann, E.L.: Elements of large-sample theory. Springer (1999)
11. Marčenko, V., Pastur, L.: Distribution of eigenvalues for some sets of random matrices. Math USSR Sb **1**, 457–483 (01 1967)
12. Moore, R.E., Kearfott, R.B., Cloud, M.J.: Introduction to Interval Analysis. Society for Industrial and Applied Mathematics (2009)
13. Serfling, R.J.: Approximation theorems of mathematical statistics. John Wiley & Sons (2009)
14. Silverstein, J.: Weak convergence of random functions defined by the eigenvectors of sample covariance matrices. The Annals of Probability **18** (02 2000). https://doi.org/10.1214/aop/1176990741
15. Silverstein, J.W.: Some limit theorems on the eigenvectors of large dimensional sample covariance matrices. Journal of Multivariate Analysis **15**(3), 295–324 (December 1984)
16. Silverstein, J.W.: On the eigenvectors of large dimensional sample covariance matrices. Journal of Multivariate Analysis **30**(1), 1–16 (July 1989)
17. Weyl, H.: Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung). Mathematische Annalen **71**, 441–479 (1912)
18. Xi, H., Yang, F., Yin, J.: Convergence of eigenvector empirical spectral distribution of sample covariance matrices. Annals of Statistics **48**(2), 953–982 (Apr 2020). https://doi.org/10.1214/19-aos1832, `http://dx.doi.org/10.1214/19-AOS1832`
19. Xia, N., Qin, Y., Bai, Z.: Convergence rates of eigenvector empirical spectral distribution of large dimensional sample covariance matrix. The Annals of Statistics **41**(5), 2572–2607 (2013)

A Proof of Theorem 2: Derivation of the covariance of the U-statistics for the covariance matrix

A.1 Overview

In this appendix, we show the details of the derivation of Theorem 2. We derive minimum-variance unbiased estimates of the covariance between two U-statistics estimates $\hat{\Sigma}_{ij}$ and $\hat{\Sigma}_{kl}$, where (i, j, k, l) range over each of the d variates in a covariance matrix $\hat{\Sigma}$. We note h and g the corresponding kernel of order 2 for $\hat{\Sigma}_{ij}$ and $\hat{\Sigma}_{kl}$, where

$$h(u_1, u_2) = \frac{1}{2}\left(X_{i_1} - X_{i_2}\right)\left(X_{j_1} - X_{j_2}\right), \text{with } u_r = (X_{i_r}, X_{j_r})^T \tag{19}$$

$$g(v_1, v_2) = \frac{1}{2}\left(X_{k_1} - X_{k_2}\right)\left(X_{l_1} - X_{l_2}\right), \text{with } v_r = (X_{k_r}, X_{l_r})^T. \tag{20}$$

Then, the covariance $\mathrm{Cov}(\hat{\Sigma}_{ij}, \hat{\Sigma}_{kl})$ for the two U-statistics $\hat{\Sigma}_{ij}$ and $\hat{\Sigma}_{kl}$ is

$$\begin{aligned} \mathrm{Cov}(\hat{\Sigma}_{ij}, \hat{\Sigma}_{kl}) &= \binom{n}{2}^{-1} (2(n-2)\zeta_1 + \zeta_2) \\ &= \binom{n}{2}^{-1} (2(n-2)\zeta_1) + \mathcal{O}(n^{-2}) \end{aligned} \tag{21}$$

where $\zeta_1 = \mathrm{Cov}\left(\mathrm{E}_{u_2}[h(u_1, u_2)], \mathrm{E}_{v_2}[g(v_1, v_2)]\right)$.

Depending on the equality and inequality of these four index variables, the empirical covariance estimate takes a different kernel form. We have employed a computer assisted proof to determine that there are seven different forms and that each of the unique $\binom{p^2 - \binom{p}{2}}{2}$ entries in $\mathrm{Cov}(\hat{\Sigma})$ (cf. Eq. (7) from the main text) can be mapped to one of these seven cases by a simple variable substitution.

In the sequel, we first describe the algorithm that determines the seven cases (Sec. A.2), we derive empirical estimators for each of these seven cases (Sec. A.3) and show that in all cases we have linear computation time in the number of samples (Sec. A.4).

A.2 Description of the algorithm providing the seven cases

We formally describe the algorithm that provided us 7 cases for the derivation of $\mathrm{Cov}(\hat{\Sigma}_{ij}, \hat{\Sigma}_{kl})$ of Theorem 2, where (i, j, k, l) vary over the set of d variables.

Enumeration First, we enumerate all configurations of $\mathrm{Cov}(\hat{\Sigma}_{ij}, \hat{\Sigma}_{kl})$, which can be encoded as a non-unique assignment matrix of variables i, j, k, l to instantiated variables (a, b, c, d). For a fixed assignment of i to variable a, we can list all possible assignments of the 3 remaining variables (j, k, l) to any (a, b, c, d). Naïvely, we have 4^3 possible assignments, but many of them will be equivalent by variable substitution. To test whether two forms are equivalent, it is sufficient to test a reduced form for equality.

Reduced Form We map a variable assignment to a reduced form by re-labeling variables sorted by the number of occurrences, which reduces the number of possible matches up-to non-uniqueness of the mapping due to equal numbers of variable occurrences. This ambiguity is then resolved by testing for symmetries.

Symmetry Symmetry of the covariance operator brings the following equally that we take into consideration in testing for equivalence:

$$\begin{aligned} \mathrm{Cov}(\hat{\Sigma}_{ij}, \hat{\Sigma}_{kl}) &= \mathrm{Cov}(\hat{\Sigma}_{kl}, \hat{\Sigma}_{ij}) = \mathrm{Cov}(\hat{\Sigma}_{ij}, \hat{\Sigma}_{lk}) = \mathrm{Cov}(\hat{\Sigma}_{lk}, \hat{\Sigma}_{ij}) \\ &= \mathrm{Cov}(\hat{\Sigma}_{lk}, \hat{\Sigma}_{ji}) = \mathrm{Cov}(\hat{\Sigma}_{ji}, \hat{\Sigma}_{kl}) = \mathrm{Cov}(\hat{\Sigma}_{ji}, \hat{\Sigma}_{lk}) \end{aligned} \tag{22}$$

The algorithm outputs each variable assignment that is not equivalent by variable substitution to any previously enumerated assignment. The seven different cases are enumerated in Table 1.

Cases	Indices	Correspondence
1	$i \neq j, k, l;\ j \neq k, l;\ k \neq l$	$\mathrm{Cov}(\hat{\Sigma}_{ij}, \hat{\Sigma}_{kl})$
2	$i = j;\ j \neq k, l;\ k = l$	$\mathrm{Cov}(\hat{\Sigma}_{ii}, \hat{\Sigma}_{kk})$
3	$i = j;\ j \neq k, l;\ k \neq l$	$\mathrm{Cov}(\hat{\Sigma}_{ii}, \hat{\Sigma}_{kl})$
4	$i = k;\ j \neq i, k, l;\ k \neq l$	$\mathrm{Cov}(\hat{\Sigma}_{ij}, \hat{\Sigma}_{il})$
5	$i = k;\ i \neq j;\ j = l;$	$\mathrm{Var}(\hat{\Sigma}_{ij})$
6	$i = j = k;\ i \neq l$	$\mathrm{Cov}(\hat{\Sigma}_{ii}, \hat{\Sigma}_{il})$
7	$i = j, k, l$	$\mathrm{Var}(\hat{\Sigma}_{ii})$

Table 1: Enumeration and correspondence of the seven cases.

A.3 The seven exhaustive cases

We now derive linear-time finite-sample estimates of the covariance for each of the seven cases.

Notation We denote the p-dimensional data matrix with n i.i.d samples as $X \in \mathbb{R}^{p \times n}$, data distribution as P_X, X_i – i^{th} row of the data matrix, $\overline{X_i X_j} = \mathbb{E}_X[X_i X_j]$, $\overline{X_i}\ \overline{X_j} = \mathbb{E}_X[X_i]\mathbb{E}_X[X_j]$.

Case 1: $\boldsymbol{i \neq j, k, l;\ j \neq k, l;\ k \neq l}$ The kernels are

$$h(u_1, u_2) = \frac{1}{2}\left(X_{i_1} - X_{i_2}\right)\left(X_{j_1} - X_{j_2}\right); \qquad g(v_1, v_2) = \frac{1}{2}\left(X_{k_1} - X_{k_2}\right)\left(X_{l_1} - X_{l_2}\right)$$

$$\mathrm{E}_{u_2}[h(u_1, u_2)] = \frac{1}{2}\left(X_{i_1} - \overline{X_i}\right)\left(X_{j_1} - \overline{X_j}\right); \mathrm{E}_{u_2}[g(v_1, v_2)] = \frac{1}{2}\left(X_{k_1} - \overline{X_k}\right)\left(X_{l_1} - \overline{X_l}\right)$$

$$
\begin{aligned}
\zeta_1 &= \mathrm{Cov}\left[\frac{1}{2}\left(X_{i_1}-\overline{X_i}\right)\left(X_{j_1}-\overline{X_j}\right), \frac{1}{2}\left(X_{k_1}-\overline{X_k}\right)\left(X_{l_1}-\overline{X_l}\right)\right] \qquad (23)\\
&= \frac{1}{4}\left\{\mathrm{Cov}\left[X_{i_1}X_{j_1}-\overline{X_i}X_{j_1}-X_{i_1}\overline{X_j}; X_{k_1}X_{l_1}-\overline{X_k}X_{l_1}-X_{k_1}\overline{X_l}\right]\right\}\\
&= \frac{1}{4}\left\{\mathrm{E}_{u_1}\left[X_{i_1}X_{j_1}X_{k_1}X_{l_1}-\overline{X_i}X_{j_1}X_{k_1}X_{l_1}-X_{i_1}\overline{X_j}X_{k_1}X_{l_1}\right.\right.\\
&\qquad -X_{i_1}X_{j_1}\overline{X_k}X_{l_1}+\overline{X_i}X_{j_1}\overline{X_k}X_{l_1}+X_{i_1}\overline{X_j}\;\overline{X_k}X_{l_1}\\
&\qquad \left.-X_{i_1}X_{j_1}X_{k_1}\overline{X_l}+\overline{X_i}X_{j_1}X_{k_1}\overline{X_l}+X_{i_1}\overline{X_j}X_{k_1}\overline{X_l}\right]\\
&\qquad \left.-\mathrm{E}_{u_1}\left[X_{i_1}X_{j_1}-\overline{X_i}X_{j_1}-X_{i_1}\overline{X_j}\right]\mathrm{E}_{u_1}\left[X_{k_1}X_{l_1}-\overline{X_k}X_{l_1}-X_{k_1}\overline{X_l}\right]\right\}\\
&= \frac{1}{4}\left\{\overline{X_iX_jX_kX_l}-\overline{X_i}\;\overline{X_jX_kX_l}-\overline{X_j}\;\overline{X_iX_kX_l}\right.\\
&\qquad -\overline{X_k}\;\overline{X_iX_jX_l}+\overline{X_i}\;\overline{X_k}\;\overline{X_jX_l}+\overline{X_j}\;\overline{X_k}\;\overline{X_iX_l}\\
&\qquad -\overline{X_iX_jX_k}\;\overline{X_l}+\overline{X_i}\;\overline{X_l}\;\overline{X_jX_k}+\overline{X_j}\;\overline{X_l}\;\overline{X_iX_k}\\
&\qquad \left.-\left(\overline{X_iX_j}-2\;\overline{X_i}\;\overline{X_j}\right)\left(\overline{X_kX_l}-2\;\overline{X_k}\;\overline{X_l}\right)\right\}
\end{aligned}
$$

Case 2: $\boldsymbol{i=j;\ j\neq k,l;\ k=l}$ The kernels are

$$h(u_1,u_2)=\frac{1}{2}\left(X_{i_1}-X_{i_2}\right)^2; \qquad g(v_1,v_2)=\frac{1}{2}\left(X_{k_1}-X_{k_2}\right)^2$$

$$\mathrm{E}_{u_2}[h(u_1,u_2)]=\frac{1}{2}\left(X_{i_1}-\overline{X_i}\right)^2; \qquad \mathrm{E}_{u_2}[g(v_1,v_2)]=\frac{1}{2}\left(X_{k_1}-\overline{X_k}\right)^2$$

Then, we have

$$
\begin{aligned}
\zeta_1 &= \mathrm{Cov}\left[\frac{1}{2}\left(X_{i_1}-\overline{X_i}\right)^2; \frac{1}{2}\left(X_{k_1}-\overline{X_k}\right)^2\right] \qquad (24)\\
&= \frac{1}{4}\left\{\mathrm{Cov}\left[X_{i_1}^2-2X_{i_1}\overline{X_i}; X_{k_1}^2-2X_{k_1}\overline{X_k}\right]\right\}\\
&= \frac{1}{4}\left\{\mathrm{E}_{u_1}\left[X_{i_1}^2X_{k_1}^2-2X_{i_1}\overline{X_i}X_{k_1}^2-2X_{i_1}^2X_{k_1}\overline{X_k}+4X_{i_1}\overline{X_i}X_{k_1}\overline{X_k}\right]\right.\\
&\qquad \left.-\mathrm{E}_{u_1}\left[X_{i_1}^2-2X_{i_1}\overline{X_i}\right]\mathrm{E}_{u_1}\left[X_{k_1}^2-2X_{k_1}\overline{X_k}\right]\right\}\\
&= \frac{1}{4}\left\{\overline{X_i^2X_k^2}-2\;\overline{X_i}\;\overline{X_iX_k^2}-2\;\overline{X_i^2X_{k_1}}\;\overline{X_k}+4\overline{X_iX_k}\;\overline{X_i}\;\overline{X_k}\right.\\
&\qquad \left.-\left(\overline{X_i^2}-2\;\overline{X_i}^2\right)\left(\overline{X_k^2}-2\;\overline{X_k}^2\right)\right\}
\end{aligned}
$$

Case 3: $\boldsymbol{i=j;\ j\neq k,l;\ k\neq l}$ The kernels are

$$h(u_1,u_2)=\frac{1}{2}\left(X_{i_1}-X_{i_2}\right)^2; \qquad g(v_1,v_2)=\frac{1}{2}\left(X_{k_1}-X_{k_2}\right)\left(X_{l_1}-X_{l_2}\right)$$
$$\mathrm{E}_{u_2}[h(u_1,u_2)]=\frac{1}{2}\left(X_{i_1}-c\right)^2; \qquad \mathrm{E}_{u_2}[g(v_1,v_2)]=\frac{1}{2}\left(X_{k_1}-\overline{X_k}\right)\left(X_{l_1}-\overline{X_l}\right)$$

Then, we have

$$\begin{aligned}
\zeta_1 &= \mathrm{Cov}\left[\frac{1}{2}\left(X_{i_1}-\overline{X_i}\right)^2; \frac{1}{2}\left(X_{k_1}-\overline{X_k}\right)\left(X_{l_1}-\overline{X_l}\right)\right] \\
&= \frac{1}{4}\Big\{\mathrm{Cov}\left[X_{i_1}^2-2X_{i_1}\overline{X_i}; X_{k_1}X_{l_1}-\overline{X_k}X_{l_1}-X_{k_1}\overline{X_l}\right]\Big\} \\
&= \frac{1}{4}\Big\{\mathrm{E}_{u_1}\left[X_{i_1}^2X_{k_1}X_{l_1}-2X_{i_1}\overline{X_i}X_{k_1}X_{l_1}-X_{i_1}^2\overline{X_k}X_{l_1}\right. \\
&\qquad\qquad \left.+2X_{i_1}\overline{X_i}\ \overline{X_k}X_{l_1}-X_{i_1}^2X_{k_1}\overline{X_l}+2X_{i_1}\overline{X_i}X_{k_1}\overline{X_l}\right] \\
&\qquad -\mathrm{E}_{u_1}\left[X_{i_1}^2-2X_{i_1}\overline{X_i}\right]\mathrm{E}_{u_1}\left[X_{k_1}X_{l_1}-\overline{X_k}X_{l_1}-X_{k_1}\overline{X_l}\right]\Big\} \\
&= \frac{1}{4}\Big\{\overline{X_i^2X_kX_l}-2\ \overline{X_iX_kX_l}\ \overline{X_i}-\overline{X_i^2X_l}\ \overline{X_k} \\
&\qquad\qquad +2\ \overline{X_iX_l}\ \overline{X_i}\ \overline{X_k}-\overline{X_i^2X_{k_1}}\ \overline{X_l}+2\ \overline{X_iX_k}\ \overline{X_i}\ \overline{X_l} \\
&\qquad -\left(\overline{X_i^2}-2\ \overline{X_i}^2\right)\left(\overline{X_kX_l}-2\ \overline{X_k}\ \overline{X_l}\right)\Big\}
\end{aligned} \tag{25}$$

Case 4: $i=k$; $j\neq i,k,l$; $k\neq l$ The kernels are

$$h(u_1,u_2)=\frac{1}{2}\left(X_{i_1}-X_{i_2}\right)\left(X_{j_1}-X_{j_2}\right); \qquad g(v_1,v_2)=\frac{1}{2}\left(X_{i_1}-X_{i_2}\right)\left(X_{l_1}-X_{l_2}\right)$$
$$\mathrm{E}_{u_2}[h(u_1,u_2)]=\frac{1}{2}\left(X_{i_1}-\overline{X_i}\right)\left(X_{j_1}-\overline{X_j}\right); \mathrm{E}_{u_2}[g(v_1,v_2)]=\frac{1}{2}\left(X_{i_1}-\overline{X_i}\right)\left(X_{l_1}-\overline{X_l}\right)$$

Then, we have

$$\begin{aligned}
\zeta_1 &= \mathrm{Cov}\left[\frac{1}{2}\left(X_{i_1}-\overline{X_i}\right)\left(X_{j_1}-\overline{X_j}\right);\frac{1}{2}\left(X_{i_1}-\overline{X_i}\right)\left(X_{l_1}-\overline{X_l}\right)\right] \qquad (26)\\
&= \frac{1}{4}\left\{\mathrm{Cov}\left[X_{i_1}X_{j_1}-\overline{X_i}X_{j_1}-X_{i_1}\overline{X_j};X_{i_1}X_{l_1}-\overline{X_i}X_{l_1}-X_{i_1}\overline{X_l}\right]\right\}\\
&= \frac{1}{4}\left\{\mathrm{E}_{u_1}\left[X_{i_1}^2X_{j_1}X_{l_1}-\overline{X_i}X_{j_1}X_{i_1}X_{l_1}-X_{i_1}^2\overline{X_j}X_{l_1}\right.\right.\\
&\qquad -X_{i_1}X_{j_1}\overline{X_i}X_{l_1}+\overline{X_i}^2\ X_{j_1}X_{l_1}+X_{i_1}\overline{X_j}\ \overline{X_i}X_{l_1}\\
&\qquad \left.-X_{i_1}^2X_{j_1}\overline{X_l}+\overline{X_i}X_{j_1}X_{i_1}\overline{X_l}+X_{i_1}^2\overline{X_j X_l}\right]\\
&\quad \left.-\mathrm{E}_{u_1}\left[X_{i_1}X_{j_1}-\overline{X_i}X_{j_1}-X_{i_1}\overline{X_j}\right]\mathrm{E}_{u_1}\left[X_{i_1}X_{l_1}-\overline{X_i}X_{l_1}-X_{i_1}\overline{X_l}\right]\right\}\\
&= \frac{1}{4}\left\{\overline{X_{i_1}^2X_{j_1}X_{l_1}}-\overline{X_i}\ \overline{X_{j_1}X_{i_1}X_{l_1}}-\overline{X_{i_1}^2X_{l_1}}\ \overline{X_j}\right.\\
&\qquad -\overline{X_{i_1}X_{j_1}X_{l_1}}\ \overline{X_i}+\overline{X_i}^2\ \overline{X_{j_1}X_{l_1}}+\overline{X_{i_1}X_{l_1}}\ \overline{X_j}\ \overline{X_i}\\
&\qquad -\overline{X_{i_1}^2X_{j_1}}\ \overline{X_l}+\overline{X_i}\ \overline{X_{j_1}X_{i_1}}\ \overline{X_l}+\overline{X_{i_1}^2}\ \overline{X_j}\ \overline{X_l}]\\
&\quad \left.-\left(\overline{X_iX_j}-2\ \overline{X_i}\ \overline{X_j}\right)\left(\overline{X_iX_l}-2\ \overline{X_i}\ \overline{X_l}\right)\right\}
\end{aligned}$$

Case 5: $i = k$; $i \neq j$; $j = l$;

$$h(u_1,u_2)=\frac{1}{2}\left(X_{i_1}-X_{i_2}\right)\left(X_{j_1}-X_{j_2}\right);$$

$$\mathrm{E}_{u_2}[h(u_1,u_2)]=\frac{1}{2}\left(X_{i_1}-\overline{X_i}\right)\left(X_{j_1}-\overline{X_j}\right);$$

$$g(v_1,v_2)=h(u_1,u_2)$$

$$\mathrm{E}_{u_2}[g(v_1,v_2)]=\mathrm{E}_{u_2}[h(u_1,u_2)]$$

Then, we have

$$\begin{aligned}
\zeta_1 &= \mathrm{Var}\left[\frac{1}{2}\left(X_{i_1}-\overline{X_i}\right)\left(X_{j_1}-\overline{X_j}\right)\right] \qquad (27)\\
&= \frac{1}{4}\left\{\mathrm{Var}\left[X_{i_1}X_{j_1}-\overline{X_i}X_{j_1}-X_{i_1}\overline{X_j}\right]\right\}\\
&= \frac{1}{4}\left\{\mathrm{E}_{u_1}\left[(X_{i_1}X_{j_1}-\overline{X_i}X_{j_1}-X_{i_1}\overline{X_j})^2\right]-\mathrm{E}_{u_1}\left[X_{i_1}X_{j_1}-\overline{X_i}X_{j_1}-X_{i_1}\overline{X_j}\right]^2\right\}\\
&= \frac{1}{4}\left\{\mathrm{E}_{u_1}\left[X_{i_1}^2X_{j_1}^2-2X_{i_1}X_{j_1}^2\overline{X_i}+\overline{X_i}^2X_{j_1}^2-2X_{i_1}^2X_{j_1}\overline{X_j}+2\overline{X_i}X_{j_1}X_{i_1}\overline{X_j}+X_{i_1}^2\overline{X_j}^2\right]\right.\\
&\qquad\left.-\left(\overline{X_iX_j}-2(\overline{X_i}\ \overline{X_j})\right)^2\right\}\\
&= \frac{1}{4}\left\{\overline{X_i^2X_j^2}-2\overline{X_iX_j^2}\ \overline{X_i}+\overline{X_i}^2\ \overline{X_j^2}-2\overline{X_i^2X_j}\ \overline{X_j}+2\overline{X_i}\ \overline{X_j}\ \overline{X_jX_i}+\overline{X_i^2}\ \overline{X_j}^2\right.\\
&\qquad\left.-\left(\overline{X_iX_j}-2(\overline{X_i}\ \overline{X_j})\right)^2\right\}
\end{aligned}$$

Case 6: $i = j = k$; $i \neq l$ The kernels are

$$h(u_1,u_2)=\frac{1}{2}\left(X_{i_1}-X_{i_2}\right)^2; \qquad g(v_1,v_2)=\frac{1}{2}\left(X_{i_1}-X_{i_2}\right)\left(X_{l_1}-X_{l_2}\right)$$

$$\mathrm{E}_{u_2}[h(u_1,u_2)]=\frac{1}{2}\left(X_{i_1}-\overline{X_i}\right)^2; \qquad \mathrm{E}_{u_2}[g(v_1,v_2)]=\frac{1}{2}\left(X_{i_1}-\overline{X_i}\right)\left(X_{l_1}-\overline{X_l}\right)$$

Then, we have

$$\begin{aligned}
\zeta_1 &= \mathrm{Cov}\left[\frac{1}{2}\left(X_{i_1}-\overline{X_i}\right)^2;\frac{1}{2}\left(X_{i_1}-\overline{X_i}\right)\left(X_{l_1}-\overline{X_l}\right)\right] \qquad (28)\\
&= \frac{1}{4}\left\{\mathrm{Cov}\left[X_{i_1}^2-2X_{i_1}\overline{X_i};X_{i_1}X_{l_1}-\overline{X_i}X_{l_1}-X_{i_1}\overline{X_l}\right]\right\}\\
&= \frac{1}{4}\left\{\mathrm{E}_{u_1}\left[X_{i_1}^2X_{i_1}X_{l_1}-2X_{i_1}\overline{X_i}X_{i_1}X_{l_1}-X_{i_1}^2\overline{X_i}X_{l_1}\right.\right.\\
&\qquad\left.+2X_{i_1}\overline{X_i}\ \overline{X_i}X_{l_1}-X_{i_1}^2X_{i_1}\overline{X_l}+2X_{i_1}\overline{X_i}X_{i_1}\overline{X_l}\right]\\
&\qquad\left.-\mathrm{E}_{u_1}\left[X_{i_1}^2-2X_{i_1}\overline{X_i}\right]\mathrm{E}_{u_1}\left[X_{i_1}X_{l_1}-\overline{X_i}X_{l_1}-X_{i_1}\overline{X_l}\right]\right\}\\
&= \frac{1}{4}\left\{\overline{X_i^3X_l}-3\ \overline{X_i^2X_l}\ \overline{X_i}+2\ \overline{X_iX_l}\ \overline{X_i}^2-\overline{X_i^3}\ \overline{X_l}+2\ \overline{X_i^2}\ \overline{X_i}\ \overline{X_l}\right.\\
&\qquad\left.-\left(\overline{X_i^2}-2\ \overline{X_i}^2\right)\left(\overline{X_iX_l}-2\ \overline{X_i}\ \overline{X_l}\right)\right\}
\end{aligned}$$

Case 7: $i = j, k, l$ The kernels are

$h(u_1,u_2) = \frac{1}{2}\left(X_{i_1} - X_{i_2}\right)^2;$ $\qquad g(v_1,v_2) = h(u_1,u_2)$

$\mathrm{E}_{u_2}[h(u_1,u_2)] = \frac{1}{2}\left(X_{i_1} - \overline{X_i}\right)^2;$ $\qquad \mathrm{E}_{u_2}[g(v_1,v_2)] = \mathrm{E}_{u_2}[h(u_1,u_2)]$

Then, we have

$$\begin{aligned}
\zeta_1 &= \mathrm{Var}\left[\frac{1}{2}\left(X_{i_1} - \overline{X_i}\right)^2\right] \qquad (29)\\
&= \frac{1}{4}\mathrm{Var}\left[X_{i_1}^2 - 2X_{i_1}\overline{X_i}\right]\\
&= \frac{1}{4}\left\{\mathrm{E}_{u_1}\left[\left(X_{i_1}^2 - 2X_{i_1}\overline{X_i}\right)^2\right] - \mathrm{E}_{u_1}\left[X_{i_1}^2 - 2X_{i_1}\overline{X_i}\right]^2\right\}\\
&= \frac{1}{4}\left\{\overline{X_i^4} - 4\overline{X_i^3}\ \overline{X_i} + 4\overline{X_i^2}\ \overline{X_i}^2 - \left(\overline{X_i^2} - 2\overline{X_i}^2\right)^2\right\}
\end{aligned}$$

A.4 Derivation in $\mathcal{O}(n)$ time for all terms

In section A.3, all terms are in the form of $\mathrm{E}[X]$,$\mathrm{E}[XY]$,$\mathrm{E}[XYZ]$ and $\mathrm{E}[XYUV]$ and can be computed in $\mathcal{O}(n)$ as follows

$$\mathrm{E}[X] \approx \frac{1}{m}\sum_{q=1}^{n} X_q \qquad (30)$$

$$\mathrm{E}[XY] \approx \frac{1}{m}\sum_{q=1}^{n} X_q \odot Y_q \qquad (31)$$

$$\mathrm{E}[XYZ] \approx \frac{1}{m}\sum_{q=1}^{n} X_q \odot Y_q \odot Z_q \qquad (32)$$

$$\mathrm{E}[XYUV] \approx \frac{1}{m}\sum_{q=1}^{n} X_q \odot Y_q \odot U_q \odot V_q \qquad (33)$$

B Orthonormality constraints

The bounds on the eigenvectors described in the main text are valid, but do not exploit the orthonormality property of the eigenvectors. Here we show how one can make use of it, and tighten the bounds.

Let $\hat{\beta}$ (respectively $\hat{\alpha}$) denote a previously obtained lower bound (respectively upper bound) on $V \odot V$, where $\odot$ denotes the Hadamard product. From the fact that the norm of each eigenvector is equal to one, we additionally obtain $0 \leq |V_{ij}| \leq 1$, and

$$1 - \sum_{k\neq i}\hat{\alpha}_{kj} \leq |V_{ij}|^2 \leq 1 - \sum_{k\neq i}\hat{\beta}_{kj}. \qquad (34)$$

Orthogonality of the eigenvectors,

$$\sum_k V_{ki} V_{kj} = 0, \tag{35}$$

implies additional constraints on V_{ij}. Let $\alpha \geq V \geq \beta$ be previously obtained (probabilistic) signed bounds on V (here the inequalities are taken to be element wise). Using interval arithmetic [12] we can compute bounds

$$\mu_{ijl} \leq \sum_{k \neq l} V_{ki} V_{kj} \leq \nu_{ijl}, \tag{36}$$

and subsequently infer

$$\min(V_{li}\alpha_{lj}, V_{li}\beta_{lj}) + \mu_{ijl} \leq 0 \leq \max(V_{li}\alpha_{lj}, V_{li}\beta_{lj}) + \nu_{ijl}, \tag{37}$$

which in turn leads to additional upper and lower bounds on V_{li}, the forms of which depend on the signs of the coefficients. We enumerate the cases here:

1. $\alpha_{lj} < 0$: We start with the constraint

   $$\min(V_{li}\alpha_{lj}, V_{li}\beta_{lj}) \leq -\mu_{ijl}. \tag{38}$$

   Assuming $V_{li} \geq 0$ yields

   $$V_{li} \geq -\frac{\mu_{ijl}}{\beta_{lj}} \tag{39}$$

   Assuming $V_{li} < 0$ yields

   $$V_{li} \geq -\frac{\mu_{ijl}}{\alpha_{lj}} \tag{40}$$

   If we have already constrained the sign of V_{li} we can now directly use one of these two inequalities, but we know in any case that

   $$V_{li} \geq \min\left(-\frac{\mu_{ijl}}{\alpha_{lj}}, -\frac{\mu_{ijl}}{\beta_{lj}}\right). \tag{41}$$

   A similar line of reasoning yields

   $$V_{li} \leq \max\left(-\frac{\nu_{ijl}}{\alpha_{lj}}, -\frac{\nu_{ijl}}{\beta_{lj}}\right), \tag{42}$$

   which can be similarly sharpened if the sign of V_{li} is known.
2. $\beta_{lj} > 0$: Analogous to the previous case, we obtain

   $$\min\left(-\frac{\nu_{ijl}}{\alpha_{lj}}, -\frac{\nu_{ijl}}{\beta_{lj}}\right) \leq V_{li} \leq \max\left(-\frac{\mu_{ijl}}{\alpha_{lj}}, -\frac{\mu_{ijl}}{\beta_{lj}}\right). \tag{43}$$

3. $\alpha_{lj} > 0 \land \beta_{lj} < 0$: Considering the first inequality

$$\min(V_{li}\alpha_{lj}, V_{li}\beta_{lj}) \leq -\mu_{ijl}. \tag{44}$$

Assume $V_{li} \geq 0$:

$$V_{li} \geq -\frac{\mu_{ijl}}{\beta_{lj}} \tag{45}$$

Assume $V_{li} < 0$:

$$V_{li} \leq -\frac{\mu_{ijl}}{\alpha_{lj}} \tag{46}$$

Considering the second inequality

$$\max(V_{li}\alpha_{lj}, V_{li}\beta_{lj}) \geq -\nu_{ijl}. \tag{47}$$

Assuming $V_{li} \geq 0$ yields

$$V_{li} \geq -\frac{\nu_{ijl}}{\alpha_{lj}}. \tag{48}$$

Assuming $V_{li} < 0$ yields

$$V_{li} \leq -\frac{\nu_{ijl}}{\beta_{lj}}, \tag{49}$$

and we can only add an additional constraint if the sign of V_{li} is known.

4. $\beta_{lj} = 0$ (the case of $\alpha_{lj} = 0$ is symmetric): The first constraint simplifies to

$$\min(V_{li}\alpha_{lj}, 0) \leq -\mu_{ijl}. \tag{50}$$

If μ_{ijl} is negative, this is satisfied trivially, so under the assumption that $\mu_{ijl} > 0$, this reduces to

$$V_{li} \leq -\frac{\mu_{ijl}}{\alpha_{lj}}. \tag{51}$$

The second constraint simplifies to

$$\max(V_{li}\alpha_{lj}, 0) \geq -\nu_{ijl}, \tag{52}$$

and implies (under the assumption that $\nu_{ijl} < 0$)

$$V_{li} \geq -\frac{\nu_{ijl}}{\alpha_{lj}}. \tag{53}$$

We note that Equations (34) and (37) may compliment each other, so we may iterate their application until there is no more improvement in the confidence intervals, or until improvement falls below a given tolerance.

The steps for implementing the orthonormality constraints are summarized in Algorithm 1.

Algorithm 1 Orthonormality constraints.

Require: Previously obtained lower and upper bound on $V \odot V$, $\hat{\beta}$ and $\hat{\alpha}$ respectively; N - number of eigenvectors

Update $\hat{\beta}$ and $\hat{\alpha}$ using Equation (34)
Obtain β and α, the signed lower and upper bound respectively, using Corollary 1
$\mu, \nu \leftarrow$ GetBoundsOnSum(α, β)
for $l = 1, 2, \ldots, N$ **do**
  for $i = 1, 2, \ldots, N$ **do**
    for $j = 1, 2, \ldots, N, j \neq i$ **do**
      if $\alpha_{lj} < 0$ **then**
        $\beta_{li} = \max\left(\beta_{li}, \min(-\frac{\mu_{ijl}}{\alpha_{lj}}, -\frac{\mu_{ijl}}{\beta_{lj}})\right)$
        $\alpha_{li} = \min\left(\alpha_{li}, \min(-\frac{\nu_{ijl}}{\alpha_{lj}}, -\frac{\nu_{ijl}}{\beta_{lj}})\right)$
      if $\beta_{lj} > 0$ **then**
        $\beta_{li} = \max\left(\beta_{li}, \min(-\frac{\nu_{ijl}}{\alpha_{lj}}, -\frac{\nu_{ijl}}{\beta_{lj}})\right)$
        $\alpha_{li} = \min\left(\alpha_{li}, \min(-\frac{\mu_{ijl}}{\alpha_{lj}}, -\frac{\mu_{ijl}}{\beta_{lj}})\right)$
      if $\alpha_{lj} > 0 \wedge \beta_{lj} < 0$ **then**
        if $\beta_{li} \geq 0$ **then** $\beta_{li} = \max\left(\beta_{li}, -\frac{\mu_{ijl}}{\beta_{lj}}, -\frac{\nu_{ijl}}{\alpha_{lj}}\right)$
        if $\alpha_{li} < 0$ **then** $\alpha_{li} = \min\left(\alpha_{li}, -\frac{\mu_{ijl}}{\alpha_{lj}}, -\frac{\nu_{ijl}}{\beta_{lj}}\right)$
      if $\beta_{lj} = 0$ **then**
        if $\nu_{ijl} < 0$ **then** $\beta_{li} = \max\left(\beta_{li}, -\frac{\nu_{ijl}}{\alpha_{lj}}\right)$
        if $\mu_{ijl} > 0$ **then** $\alpha_{li} = \min\left(\alpha_{li}, -\frac{\mu_{ijl}}{\alpha_{lj}}\right)$
      if $\alpha_{lj} = 0$ **then**
        if $\mu_{ijl} > 0$ **then** $\beta_{li} = \max\left(\beta_{li}, -\frac{\mu_{ijl}}{\beta_{lj}}\right)$
        if $\nu_{ijl} < 0$ **then** $\alpha_{li} = \min\left(\alpha_{li}, -\frac{\nu_{ijl}}{\beta_{lj}}\right)$

---

function GetBoundsOnSum(α, β)
  for $l = 1, 2, \ldots, N$ **do**
    for $i = 1, 2, \ldots, N-1$ **do**
      for $j = i+1 \ldots, N$ **do**
        $\mu_{ijl} = \sum_{k \neq l} \min(\beta_{ki}\beta_{kj}, \beta_{ki}\alpha_{kj}, \alpha_{ki}\beta_{kj}, \alpha_{ki}\alpha_{kj})$
        $\nu_{ijl} = \sum_{k \neq l} \max(\beta_{ki}\beta_{kj}, \beta_{ki}\alpha_{kj}, \alpha_{ki}\beta_{kj}, \alpha_{ki}\alpha_{kj})$
  return μ, ν