Please click here for the latest version.

INFERENCE ON NETWORK STATISTICS FROM ESTIMATED ADJACENCY MATRICES

JÉRÔME R. SIMONS

Abstract. PRELIMINARY, PLEASE USE ONLY FOR CONFERENCE REVIEW PURPOSES AND DO NOT CIRCULATE WITHOUT PERMISSION. We establish conditions under which a wide class of network statistics such as eigenvector centralities or clustering coefficients are consistent estimators of their population counterparts whenever we encounter an estimate of the adjacency matrix. Furthermore, we derive analytical expressions of standard errors of these statistics and characterise their asymptotic distributions. We shall use the delta method via the associated Jacobians of the aforementioned transformations. To arrive at this result, we prove that the maps assigning singular and invariant subspaces of general matrices are smooth. Our proof strategy is two-fold: building on perturbation expansions of invariant subspaces in Sun (1991) and singular subspaces in J. Liu, X. Liu, and Ma (2007), we derive methods for generic inference on normalised bases of these subspaces. Thereby, we extend the inference procedure of Tyler (1981) to generic invariant subspaces of non-symmetric and non-diagonalisable matrices. We establish distribution theory for a Wald and t-test for full-vector and individual coefficient hypotheses, for both invariant and singular subspaces respectively. We demonstrate the utility of our approach through finding confidence intervals for eigenvector centralities estimated from weighted, directed graph adjacency matrices and demonstrate the quality of the approximations through a Monte Carlo study.

1. Introduction

Network statistics are fundamental in quantifying structure and quality of a graph. For example, researchers may ask whether a given network has large clusters or cliques. Alternatively, they may want to know how important individual nodes are. If networks, and the associated adjacency matrices, are estimated or measured with error, it is natural that associated network statistics inherit this randomness and appear as stochastic objects themselves. This study establishes methods to derive the induced (asymptotic) distributions of network statistics from those of estimated adjacency matrices. We shall supply consistency proofs and expressions for Jacobians to convert standard errors of estimated network links into those for associated network statistics.

The underlying methodology for these results occurs via finding both invariant and singular subspaces of general matrices. Indeed, most network statistics are expressible as either singular or invariant vectors of adjacency matrices. We therefore establish results for inference on those vectors. For invariant vectors, we note that these inference procedures are new for the case of potentially non-diagonalisable, non-symmetric matrices and are an extension of Tyler (1981). For singular vectors, most of the literature has focussed on the case of matrices of

Date: 29 February ‘24.

growing dimension. In our case, however, the matrix is of fixed size, but measured with error or estimated via some other scheme with arbitrary covariance structure among the columns.

Our study proceeds via formally characterising subspaces of such matrices and deriving their Jacobians. In the presence of random noise, these subspaces to appear as stochastic objects for which we offer hypothesis tests. In particular, we are interested in testing for membership in subspaces and the construction of confidence regions for associated basis matrices.

In the special case of a diagonalisable, non-symmetric matrix, our invariant subspace methodology provides inference procedures for eigenvectors. If we specialise further to symmetric matrices, invariant and singular vectors coincide so that those otherwise distinct inference procedures coincide. Once we have characterised invariant and singular subspaces as induced maps on Banach spaces we fill in the link between these subspaces and subspace-dependent network statistics, with the eigenvector centrality being the most prominent example. For this chain of maps, we prove that an inverse mapping theorem applies to a large class of functions of matrices and compute their Jacobians.

Setup. For a $p \times p$ matrix M, the columns of the $p \times q$ matrix R span a right-invariant subspace $\mathcal{S}_I(M)$ of dimension q iff the relation

$$MR = R\Lambda \tag{1.1}$$

holds for some not necessarily full-rank $q \times q$ matrix Λ.[1] In the above example, the rows of matrices L^T span a left invariant subspace whereby $L^\mathsf{T} M = \Lambda L^\mathsf{T}$. If M is instead an $m \times l$ general matrix of rank $q \leq \min(m, l)$, the columns of the $l \times q$ and $m \times q$ matrices V and U span the left- and right singular subspaces of dimension q iff instead the relation

$$MV = U\Sigma$$

holds for a diagonal $q \times q$ matrix Σ.

Our study provides inference methods for right-invariant and left-singular subspaces for M.[2] If M is a symmetric, square matrix and hence diagonalisable, left- and right-invariant as well as singular subspaces coincide and $R = L = U = V$.

For either type of subspace, we assume that we observe a sample of n matrices $M_1, \ldots, M_n$, whence we estimate the mean $\hat{M}_n$ so that the columns of $a_n\left(\hat{M}_n - M\right)$ converge weakly to a multivariate normal distribution centered at zero for an increasing sequence a_n. Our approach is to construct test statistics for membership of υ of invariant and singular subspaces. The maps $\psi(M; \upsilon)$ and $\Psi(M; \upsilon)$ represent these statistics and have informative null distributions for invariant and singular subspaces of M, spanned by the columns of R or V. We derive these distributions using the delta method to find the distributions of $\sqrt{n}\left(\psi\left(\hat{M}_n\right) - \psi(M)\right)$ and

[1] If A is rank-deficient, then some columns in X belong to the kernel of the linear transformation M.

[2] Confusingly, an eigendecomposition of $M = R\Lambda L^H$ implies that R contains right eigenvectors, while for a singular value decomposition, $M = U\Sigma V^\mathsf{T}$, U contains left singular vectors. Our study focusses on those matrices appearing ‘on the left.’

$\sqrt{n}\left(\Psi\left(\hat{M}_n\right)-\Psi\left(M\right)\right)$ based on that of $\sqrt{n}\left(\hat{M}_n-M\right)$. Inverting coefficient-wise versions of $\psi\left(M;\upsilon\right)$ and $\Psi\left(M;\upsilon\right)$ yields confidence regions for υ for either subspace.

We are interested in testing the associated null hypothesis of υ belonging to a subspace of M potentially indexed by elements of a set $\mathcal{L}$, which in the case of eigenspaces contains characteristic roots and singular values otherwise. Formally, the null hypothesis to test is

$$H_0 : \upsilon \in \mathcal{S}_I\left(M\right) \tag{1.2}$$

against the one-sided alternative $H_1 : \upsilon \notin \mathcal{S}_I\left(M\right)$. To introduce inference on invariant subspaces, we will use the case of diagonalisable M and introduce eigenvalues and singular values into the discussion where necessary. Associating subspace vectors with a set of such values, we assume in either case that the elements of $\mathcal{L}$ split according to

$$\mathcal{L} := \mathcal{L}_I \cup \mathcal{L}_J,$$

where $\mathcal{L}_I$ denotes a set of interest and $\mathcal{L}_J$ the remainder. For eigenspaces, $\mathcal{L}_I$ contains the roots associated with the directions of interest.

Throughout, we assume that $\operatorname{rk}\upsilon = q \leq \left|\mathcal{L}_I\right|$, i.e. the number of linearly independent columns in υ is less than or equal to the dimension of the subspace spanned by the vectors associated with elements in $\mathcal{L}_I$. If $\operatorname{rk}\upsilon = r > \left|\mathcal{L}_I\right|$, i.e. the number of linearly independent columns in υ exceeds the number of elements in $\mathcal{L}_I$, we define the alternative null hypothesis

$$H_0^* : \mathcal{S}_I\left(M\right) \in \operatorname{sp}\, \upsilon,$$

which can be tested by constructing the $p \times q$ matrix $\upsilon_\perp$ such that

$$\upsilon_\perp^\intercal \upsilon = 0. \tag{1.3}$$

Consequently, we write

$$H_0^* : \upsilon_\perp \in \mathcal{S}_J\left(M\right). \tag{1.4}$$

The case of $\operatorname{rk}\upsilon = r \leq \left|\mathcal{L}_I\right|$ is intuitive because the columns of υ span a lower-dimensional subspace than $\mathcal{S}_I\left(M\right)$ while $\operatorname{rk}\upsilon = r > \left|\mathcal{L}_I\right|$ is equivalent to an overdetermination of the subspace of interest so that the estimand $\mathcal{S}_I\left(M\right)$, under the null hypothesis, is wholly contained in the candidate space $\operatorname{sp}\, \upsilon$. For eigenvectors, Tyler (1981) implements this scenario by inverting the hypothesis (1.2) to achieve (1.4) which reduces to H_0 again. The orthocomplement, $\upsilon_\perp$, is always easy to compute from υ using, e.g. an LU factorization or the Gram-Schmidt algorithm. Moreover, working with H_0^* and hence $\upsilon_\perp$ implies normalisation-invariant test statistics and offers numerical advantages as Monte Carlo experiments showed. We shall therefore write our arguments in the form of H_0^*, where inference on $\upsilon_\perp$ is equivalent to inference on υ without loss of generality.

The main limitation of the existing approaches for invariant subspaces is that per Tyler (1981), we have to assume that M is such that ΓM is symmetric for a positive definite symmetric matrix Γ. We illustrate this issue with

Example 1. The matrix M_1 is symmetric in the metric of Γ for

$$\Gamma = \begin{bmatrix} 2 & 1 \\ 1 & 6 \end{bmatrix} \qquad\qquad M_1 = \begin{bmatrix} 1 & 3 \\ 1 & 1 \end{bmatrix}$$

whereas M_2 is not symmetric in the metric of Γ,

$$M_2 = \begin{bmatrix} \lambda & a \\ 0 & \lambda \end{bmatrix}.$$

A 2×2 matrix Γ is not guaranteed to be positive-definite for $M_2\Gamma$ to be symmetric. Moreover, M_2 only has a single eigenvector $\left[\begin{smallmatrix} 1 & 0 \end{smallmatrix}\right]^\mathsf{T}$, and is therefore not diagonalisable although its single eigenvector obviously spans an invariant subspace. Therefore, inferential procedures for non-symmetric matrices apply to M_1 but not to M_2. In the following, we shall establish that no such requirement is necessary for inference on invariant subspaces of non-symmetric matrices.

1.1. **Related literature on subspace inference.** Statistical analysis of eigenvectors and eigenvalues finds use in diverse applications such as quasi-cointegration analysis, the study of centrality in social networks (Bonacich 1987), or in radio signal processing (Penna, Garello, and Spirito 2009).

The study of inference and estimation of eigenvectors of more general, dense square random matrices $M \in \mathbb{R}^{p \times p}$ with p fixed began with the seminal study of Davis (1977) in the non-normal case. James (1977) posed the problem of eigenvector inference for general eigenvectors without associating them with a particular set of eigenvalues. Tyler (1981) considers null hypotheses of eigenvectors associated with particular eigenvalues. His study served as the inspiration for our paper and requires quasi-symmetry of M, which we relax.

The first study to examine spectra of symmetric random matrices is the seminal paper by Anderson (1963), which provided joint density functions for eigenvectors and eigenvalues of covariance matrices and a testing procedure for equality of eigenvalues. Subsequently, James (1964) offered a general method to obtain the distribution of statistics derived from random matrices. Since then, the field of invariant subspace statistics based on random matrices has progressed along two strands.

The first branch of the literature considers the spectra of Wishart matrices. To see how they arise, consider a matrix $A \in \mathbb{R}^{p \times n}$ whose columns A_i are independent and normal according to $A_i \sim N_p(0, \Omega)$ where $\Omega \in \mathbb{R}^{p \times p}$ is a positive-definite covariance matrix. The variate $M := AA^\mathsf{T}$ will then have a central, p-variate Wishart distribution with n degrees of freedom, usually denoted by $M \sim W_p(\Omega, n)$. A representative study with a further overview of the development of density functions of such matrices is given in Zanella, Chiani, and Win (2009).[3] Further refinements include expanding or restricting the spectral structure such as the study by Takemura and Sheena (2005) who derive the asymptotic distribution of eigenvalues and eigenvectors of Wishart matrices in the case of infinitely dispersed population eigenvalues.

[3]Owing to their use as representations of physical systems, studies in eigenstatistics of Wishart matrices tend to prioritize those with complex entries. For the social sciences, this literature is therefore less useful because results derived using complex matrices do not, in general, apply to real matrices.

A second strand of inference on invariant subspaces are the spectra of symmetric square matrices $M \in \mathbb{R}^{p \times p}$ with $p \to \infty$. The often sparse nature of this type of matrix implies that one can truncate spectral data or singular value decomposition to estimate low-rank representations of such matrices. This literature began with the semi-circle law of Wigner (1958) for the joint eigenvalue distribution and Tracy and Widom (1994) for the distribution of the largest eigenvalue. For applied researchers, Chen, Cheng, and Fan (2020) develop a method to conduct inference on low-rank representations of random matrices that obtain from deterministic symmetric matrices that are (potentially asymmetrically) perturbed. As a further refinement of this case, Cheng, Wei, and Chen (2021) consider matrices with finely spaced eigenvalues while Iain M Johnstone, Onatski, et al. (2020) develop a testing framework for high-dimensional matrices with a single, dominant eigenvalue, or *spike*. That latter phenomenon arises for example in auto-covariance matrices of embedded time series if the process possesses a root of unity. In the class of low-rank matrices with $p \to \infty$, Cheng, Wei, and Chen (2021) offer an estimation and inference procedure for eigenvectors of such random matrices. Bura and Pfeiffer (2008) add theory for inference on singular vectors.

A subspecialty of the aforementioned literature studies estimation and inference procedures for eigenvectors and principal component vectors of covariance matrices, as they arise in in factor models. If the column dimension of A_n diverges as $n \to \infty$, central limit theorems imply that a suitably normalized $M_n := A_n A_n^{\intercal}$ converges to a central Wishart matrix as above. This result nests the limiting distributions of a large class of covariance matrix estimators. Tyler (1983) provides a procedure to compute asymptotic confidence sets for principal components which obtain from invariant subspaces of covariance matrices. Iain M. Johnstone (2001) and Lee and Schnelli (2016) contribute results on the largest eigenvalue statistics of principal component estimators and covariance matrices. Koltchinskii, Löffler, Nickl, et al. (2020) consider a representative functional extension and focus on efficiently estimating linear functionals of principal components, which have symmetric underlying covariance matrices.

A related literature studies the singular value decomposition of rectangular random matrices $M \in \mathbb{R}^{p \times d}$ where $p, d < \infty$ and p and d are fixed. Cai et al. (2019) offer a method to estimate the column space via a singular value decomposition and Girko (1999) provides conditions for and derive a convergence law for associated singular values.

Organisation. The rest of this article is organised as follows. Section 2 defines the set of matrices that forms the domain and distinguishes between invariant and singular subspaces. The currently proposed matrix estimator is a mere placeholder intended to induce a well-behaved distribution. Section 3 defines perturbation expansions and informally discusses their connections with Gateaux derivatives and provides Jacobian matrices. Section 4 introduces practically useful estimators of normalised basis matrices for invariant and singular subspaces and uses the results of the previous section to derive their asymptotic distributions. Section 5 derives hypothesis tests for invariant and singular subspaces. Section 6 details results of a Monte Carlo study of the inference procedures for generic subspaces. Section 7 extends the framework to general network statistics and displays how the methods can be used to derive asymptotically valid confidence intervals for centrality scores if networks are estimated with error. Section 8 concludes while Section 9 contains all proofs.

Notation. We consider tests on the invariant subspaces of $M \in \mathbb{R}^{p\times p}$ with an associated sequence of estimates $\hat{M}_n$. In the context of tests on singular subspaces, M denotes a rectangular matrix $M \in \mathbb{R}^{m\times n}$. To simplify the subsequent exposition and to easily define test statistics, we vectorise M, so that the vec M represents its stacked columns. The notation B^H denotes conjugate transposition whereas $B^\intercal$ refers to transposition without conjugation even if B has complex entries as in vec $AXB = \left(B^\intercal \otimes A\right)$ vec X. Finally, λ^* denotes complex conjugation of the scalar λ. The expression A^+ refers to a generalised inverse of matrix A. We define generic null hypotheses via H_0 and H_0^* but may change H_0 depending on the context. We use $A \cdot B$ and $A \otimes B$ to denote the Hadamard and Kronecker products, respectively. Finally, $\mathbf{1}$ denotes a column vector of ones of appropriate dimension. Matrices with a tilde, $\tilde{M}(t)$ are perturbed versions of their exact counterparts M where $t \in \mathbb{C}$ is a parameter that helps us keep track of the expansion order and $\tilde{M}(0) = M$.

2. Setup

2.1. General framework.

Assumption 1. *Let $\Omega \in \mathbb{R}^{ml\times ml}$ denote a positive-definite covariance matrix and $M \in \mathbb{R}^{m\times l}$ be a general matrix. Then,*

1. *The model $M_t = M + \varepsilon_t$ generates data $\{M_t\}_{t=1}^n$ where ε_t are* i.i.d. *with general covariance matrix $\Omega = \mathbb{E} \text{vec}\, \varepsilon_i \text{ vec}\, \varepsilon_j^\intercal$ for all $i, j = 1, \dots, T$ and $m = l = p$ for the case of square M.*
2. *We can consistently estimate the covariance matrix by some covariance estimator $\hat{\Omega}$ so that $\hat{\Omega} \xrightarrow{p} \Omega$.*

Assumption 1.1 implies that for an estimator

$$\hat{M}_n := n^{-1} \sum_{t=1}^{n} M_t \tag{2.1}$$

we have

$$\sqrt{n}\,\text{vec}\,\left(\hat{M}_n - M\right) \rightsquigarrow N\left(0, \Omega\right).$$

This setup admits a general covariance structure within Ω and imposes a stationary error distribution over the index t. An important special case arises when the columns of the error matrix are homoskedastic themselves, which we define for $\varepsilon_t := \left[\mathbf{a}_1\, \mathbf{a}_2\, \dots\, \mathbf{a}_p\right]_t$ via

$$\mathbb{E}\mathbf{a}_i\mathbf{a}_j^\intercal = \Omega_M \delta_{ij} \tag{2.2}$$

so that $\Omega = I_p \otimes \Omega_M$. None of our results are sensitive to these restrictions and, if equipped with a suitable covariance matrix estimator, one could allow for autocorrelated errors and heteroskedasticity across time, as well, an application of which we offer in Section 7.2.

Finally, Assumption 1.2 ensures that we can estimate the covariance matrix of the residuals.

2.2. **Covariance Matrix Estimation.** A covariance matrix estimator associated with (2.1) is, letting $m := \operatorname{vec} M$,

$$\hat{\Omega} := T^{-1} \sum_{t=1}^{n} (\hat{m}_n - m_t)(\hat{m}_n - m_t)^\mathsf{T},$$

which we can make more efficient if we know that $\Omega = I_p \otimes \Omega_M$ by enforcing this structure during estimation. If we interpret the index t as referring to time, we can also accommodate autocorrelation in the residuals, see Newey and West (1987) and the empirical application in Section 7.2. To be able to construct Wald tests, we need to consistently estimate estimate Ω^{-1}. However, neither the rank nor a generalised inverse of Ω are continuous. To deal with this issue, we recommend the procedure outlined in Lütkepohl and Burda (1997) and invoke Yuchen Xu and Matteson (2023, Lemma 2.1) to establish that $\hat{\Omega}^{+} \xrightarrow{p} \Omega^{-1}$.

2.3. **Frameworks for invariant and singular subspace inference.** To characterise our setup for invariant subspace inference, we have

Assumption 2. *Let $M \in \mathbb{R}^{p\times p}$ have a non-trivial invariant subspace. Then,*

(1) There exists an invariant subspace of interest spanned by the columns of $R_I \in \mathbb{R}^{p\times q}$ such that $MR_I = R_I A_{11}$ for some matrix $A_{11} \in \mathbb{R}^{q\times q}$. In particular, there exists a non-singular matrix $R := \left[\begin{smallmatrix} R_I & R_J \end{smallmatrix}\right] \in \mathbb{R}^{p\times p}$ such that

$$R^{-1}MR = \begin{bmatrix} \Lambda_I & \Lambda_{IJ} \\ 0 & \Lambda_J \end{bmatrix} =: \Lambda$$

where $\Lambda_I \in \mathbb{R}^{q\times q}$ and $\Lambda_J \in \mathbb{R}^{r\times r}$ for $r = p - q$.

(2) The eigenvalues of Λ_I and Λ_J denoted by $\mathcal{L}_I$ and $\mathcal{L}_J$ obey $\mathcal{L}_I \cap \mathcal{L}_J = \emptyset$ and $\lambda \in \mathcal{L}_I$ implies that $\lambda^* \in \mathcal{L}_I$, i.e. $\mathcal{L}_I$ is closed under conjugation.

Assumption 2.1 defines our setup as that of estimating invariant subspaces. The most important special case of invariant subspaces are eigenspaces which obtain if $\Lambda_{IJ} = 0$ and Λ_I and Λ_J are diagonal. An intermediate case is that of a partially diagonalisable matrix, which we could achieve by requiring Λ_I to be diagonal with distinct eigenvalues, $\Lambda_{IJ} = 0$, and Λ_J to be in Jordan normal form with blocks of arbitrary size.

Assumption 2.2 ensures that the map from M to its invariant subspaces spanned by the columns of M is differentiable and that we can discriminate between vectors of interest in sets I and those in J.

For singular subspaces, our setup changes ever so slightly.

Assumption 3. *Let $M \in \mathbb{R}^{m\times l}$, with $m < l$, have a non-trivial singular subspace and be of rank $F \leq m$ subspace. Then,*

(1) The singular subspaces split according to $M = U_I \Sigma_I V_I^\mathsf{T} + U_J \Sigma_J V_J^\mathsf{T}$ where $\operatorname{rk} M = F$ implies that $\Sigma_I \in \mathbb{R}^{F\times F}$ is a diagonal, square matrix and $\Sigma_J = 0$.

(2) For a subspace of dimension q, we have $U_I \in \mathbb{R}^{m\times q}$ where $q \leq F$.

Assumption 3.1 is innocuous in the context of our leading application, graph models, where M is usually sparse so that $F << \min(m, l)$. Without loss of generality, we further assume that $m < l$, i.e. that we encounter a wide matrix.

3. Matrix subspaces

This section develops our approach to building inference methods of matrix functionals based on perturbation expansions. Readers who wish to advance straight to the network model application may proceed to Section 7 without loss of continuity.

3.1. **From perturbation expansions to derivatives.** To identify the correspondence between a perturbation expansion and the appearance of derivatives, it is instructive to consider the following:

Example 2. Consider a continuous function $f(x)$ and a small ϵ, that perturbs x, such that $\tilde{x} := x + \epsilon$. Recall the difference quotient

$$d(\tilde{x}; x) := \frac{f(\tilde{x}) - f(x)}{\tilde{x} - x}$$

and observe that $\lim_{x \to \tilde{x}} d(\tilde{x}; x) = f'(x)$. Now, let $f(x) = x^2$. Define a perturbation expansion for f with first-order perturbation term $\dot{f}$ like so

$$\begin{aligned} f(\tilde{x}) &= f(x) + \dot{f}(x) + O(|\tilde{x} - x|) \\ x^2 + 2\epsilon x + \epsilon^2 &= x^2 + 2x\epsilon + O(|\tilde{x} - x|) \end{aligned}$$

and recognise that the relationship, as $\tilde{x} \to x$ or $\epsilon \to 0$, between the first-order perturbation expansion term $\dot{f}$ and the derivative f' satisfies

$$\dot{f} = f'\epsilon, \tag{3.1}$$

where we can identify the derivative from a perturbation expansion without taking the limit. Magnus and Neudecker (2019, Thm. 11) extends this relationship to matrix-valued functions. In non-scalar cases, it is important that the perturbation ϵ in (3.1) appears at the end of the expression, which we achieve in practice by vectorising matrix expressions and applying the identity $\text{vec}\, AXB = \left(B^\intercal \otimes A\right) \text{vec}\, X$.

We recall the maps delivering invariant and singular subspaces, $\psi(M; v)$ and $\Psi(M; v)$, respectively, which carry information about invariant and singular subspaces. Further recall that, if v belongs to a subspace of interest, i.e. $v \in \mathcal{S}_I(M)$ iff $\psi(M; v) = 0$ and $\Psi(M; v) = 0$. Exploiting the information thus expressed, we can build inference methods for subspaces based on these maps if we can derive the asymptotic distribution of the sample analogues of ψ and Ψ from that of $\sqrt{n}\text{vec}\, \hat{M}_n$. The critical ingredient for the extraction of the distribution are the Jacobians of these maps and appear in our central results in Lemma 1 and Lemma 2. In short, we shall apply the delta method to transform the covariance matrix of $\sqrt{n}\, \text{vec}\, \hat{M}_n$ into those of ψ and Ψ. To derive the Jacobians, we use expansions of ψ and Ψ that originate from a perturbed version of M analogously to Example 2, denoted by $\tilde{M}$. A typical perturbation expansion of this type appears as $\psi\left(\tilde{M}\right)$ about $\psi\left(M\right)$, so that

$$\psi\left(\tilde{M}\right) = \psi\left(M\right) + \dot{\psi}\left(M - \tilde{M}\right) + O\left(\left\|M - \tilde{M}\right\|\right). \tag{3.2}$$

An exact analogue with $\dot{\Psi}$ in lieu of $\dot{\psi}$ holds for Ψ. Importantly, we shall argue that the first-order term indeed corresponds to the Gateaux derivative of $\psi\left(\Psi\right)$.

In the spirit of Example 2, we consider the perturbation explicitly, $E \coloneqq \tilde{M} - M$, as the deviation from the true M. Following Magnus and Neudecker (2019), we define the Jacobian matrix in the direction of E of a vectorized, matrix-valued function $F\left(X\right) : \mathbb{R}^{m\times l} \to \mathbb{R}^{r\times q}$ via first differentials as the matrix $J \in \mathbb{R}^{ml\times rq}$ satisfying

$$\mathrm{dvec}\, F\left(X; E\right) = J \mathrm{vec}\, E,$$

whence Magnus and Neudecker (2019, Thm. 11) establishes that J is indeed unique. In such case,

$\psi\left(\tilde{M}\right) - \psi\left(M\right) \to \mathrm{d}\psi$ as $E \to 0$ so that, by (Sun 1991, Remark 4.2), (3.2) turns into the definition of the Gateaux derivative. Therefore, vec $\dot{\psi}\left(M - \tilde{M}\right)$ will allow us to derive the Jacobian J of the transformation ψ. More details appear in Section 9.1 and Section 9.4, while Section 3.2 and Section 3.3 sketch the steps required to derive the Jacobians for invariant and singular subspace maps.

3.2. **Expansion of invariant subspace map ψ.** The object of interest that we wish to conduct inference on is the column span of R_I, implicitly defined via the invariant subspace decomposition

$$M = R_I\left(\Lambda_I L_I^H + \Lambda_{IJ} L_J^H\right) + R_J \Lambda_J L_J^H, \tag{3.3}$$

where we use conjugate transposition on the matrices L_I and L_J to allow for the fact that some eigenvectors may be complex. Subsequently, we partition the matrices according to their column spans to obtain

$$R = \begin{bmatrix} R_I \, R_J \end{bmatrix} \; L = \begin{bmatrix} L_I \, L_J \end{bmatrix} \; \Lambda = \begin{bmatrix} \Lambda_I & \Lambda_{IJ} \\ 0 & \Lambda_J \end{bmatrix}, \tag{3.4}$$

where $\Lambda_{IJ} = 0$ implements the familiar eigendecomposition.

Consider the invariant subspace $\operatorname{col} R_I$, whose invariance depends on Λ in (3.4) being block upper-triangular. If Λ fails this requirement, i.e. that $\Lambda_{JI} \neq 0$, we have

$$M R_I = R_I \Lambda_I + R_J \Lambda_{JI}. \tag{3.5}$$

We learn from the preceding that $\|\Lambda_{\mathrm{JI}}\|$ measures how far $\operatorname{col} R_I$ is from being an invariant subspace. Therefore, we introduce the off-diagonal element corresponding to Λ_{JI} as

$$\Delta_{\mathrm{JI}} = L_J^H\left(\tilde{M} - M\right) R_J, \tag{3.6}$$

which for $\tilde{M} = M$ implies $\Lambda_{JI} = 0$. In this case, based on (3.3), (3.5) becomes

$$MR_I = R_I \Lambda_I. \tag{3.7}$$

We conclude for now that perturbations to M affect the column span of R_I via Δ_{JI}.

To motivate a test statistic, we introduce the skew projection

$$P_I := R_I L_I^H. \tag{3.8}$$

If the roots in the set $\mathcal{L}_I$ are closed under conjugation, then we may take R_I and L_I to be real without loss of generality. However, even if M itself is a real matrix, we may encounter estimation error that may induce complex invariant vectors and, in the case of an eigendecomposition, complex eigenvalues. Naturally, if invariant vectors of interest are real, the imaginary parts of their estimators shall converge to zero in probability. Under the hypothesis $H_0 : \upsilon \in \mathcal{S}_I(M)$, we obtain $\upsilon_\perp^\intercal P_I = 0$. Let the ring of matrices satisfying Assumption 1 be $\mathscr{Q}$. Define $\psi : \mathscr{Q} \mapsto \mathbb{R}^{r \times q}$ by

$$\psi(M) := \upsilon_\perp^\intercal P_I(M). \tag{3.9}$$

To write out the first-order term $\dot{\psi}$, introduce the operator $\boldsymbol{S} : \mathbb{R}^{r\times q} \to \mathbb{R}^{r \times q}$ such that

$$\boldsymbol{S}(Q) = Q\Lambda_I - \Lambda_J Q, \tag{3.10}$$

where $Q \in \mathbb{R}^{r\times q}$.

Then, the first-order term

$$\dot{\psi}(M; \Delta_{\mathrm{JI}}) = \upsilon_\perp^\intercal R_J \boldsymbol{S}^{-1}(\Delta_{\mathrm{JI}}) L_J^H. \tag{3.11}$$

Per the reasoning in Section 3.1, we can now find the Jacobian of ψ based on vectorising the first-order perturbation expansion term $\dot{\psi}$ to obtain $\operatorname{vec}\dot{\psi} = B\operatorname{vec}\left(\tilde{M} - M\right)$, which we summarise formally in our result below.

Lemma 1. *Let the map $\psi(M)$ be as in* (3.9)*. Then, the Jacobian matrix with respect to M in the direction of $E \in \mathbb{R}^{p\times p}$ of ψ is the matrix $B \in \mathbb{R}^{p^2 \times rq}$ such that $\mathrm{d}$vec $\psi(M; E) = B$vec E where*

$$B = \left(L_I \otimes \upsilon_\perp^\intercal R_J\right)\left\{\left(\Lambda_I^H \otimes I_J\right) - \left(I_I \otimes \Lambda_J\right)\right\}^{-1}\left(R_I^H \otimes L_J^H\right). \tag{3.12}$$

3.3. **Expansion of singular subspace map Ψ.** This section focusses on the map Ψ that helps us conduct inference on the column span of U_I. To motivate our test statistic and sketch the argument of its asymptotic distribution, consider the singular value decomposition

$$M = U_I \Sigma_I V_I + U_J \Sigma_J V_J, \tag{3.13}$$

where $\operatorname{rk} U_I = F \leq \min(m, l)$ so that $\Sigma_J = 0$. To study the derivative of the implicit map that delivers U_I, we consider a perturbation to M, $\tilde{M}$, which in turn triggers perturbed versions of the singular factors in (3.13). The aim is to construct an expansion

$$\Psi\left(\tilde{M}\right) = \dot{\Psi}\left(M - \tilde{M}\right) + O\left(\left\|M - \tilde{M}\right\|\right). \tag{3.14}$$

Our focus is on left-singular vectors, which motivates the map

(3.15) $$\Psi\left(M\right) \coloneqq \upsilon_{\perp}^{\intercal} U_I\left(M\right),$$

Using the expansion in J. Liu, X. Liu, and Ma (2007), we obtain the first-order term

(3.16) $$\dot{\Psi}\left(M-\tilde{M}\right) = \upsilon_{\perp}^{\intercal}\left(U_I K - U_J U_J^H \left(M-\tilde{M}\right) V_I \Sigma_I^{-1}\right)$$

for $K \coloneqq D \cdot \left(U_I^{\intercal}\left(M-\tilde{M}\right) V_I \Sigma_I + \Sigma_I V_I^{\intercal}\left(M-\tilde{M}\right)^{\intercal} U_I\right)$ and $D_{ij} \coloneqq \left(\iota_i^2 - \iota_j^2\right)^{-1}$ where $i \neq j$, $D_{ii} = 0$, and $i, j = 1, \ldots, F$. From the first-order term (3.16), we see that it collapses to zero when $M = \tilde{M}$. We shall also apply Magnus and Neudecker (2019, Thm. 11) and (Sun 1991, Remark 4.2) to (3.14) to obtain the Jacobian of the right-singular subspace map, collected in

Lemma 2. *Let the map $\Psi\left(M\right)$ be as in* (3.15)*. Then, for a null hypothesis $\upsilon_{\perp}^{\intercal} \in \mathbb{R}^{m\times(m-h)}$ where $h \leqq F$ is the size of the hypothesised singular subspace, we obtain the Jacobian matrix with respect to M in the direction of $E \in \mathbb{R}^{m\times l}$ of Ψ as $B_{SVD} \in \mathbb{R}^{F(m-h)\times ml}$. In particular, $\mathrm{d}\mathit{vec}\,\Psi\left(M;E\right) = B_{SVD}\mathit{vec}\,E$ where*

(3.17) $$\begin{aligned} B_{SVD} = {} & \left(\Sigma_I^{-1\intercal} V_I^{\intercal} \otimes \upsilon_{\perp}^{\intercal} U_J U_J^{\intercal}\right) \\ & + \left(I_F \otimes \upsilon_{\perp}^{\intercal} U_I\right)\left[\left(\mathbf{1}^{\intercal} \otimes \mathit{vec}\,D\right) \cdot \left(\left(\Sigma_I^{\intercal} V_I^{\intercal} \otimes U_I^{\intercal}\right) + \left(U_I^{\intercal} \otimes \Sigma_I V_I^{\intercal}\right) K\right)\right]. \end{aligned}$$

3.4. **Projections and generalised inverses.** In this section we briefly review the theory of projections onto subspaces following Tyler (1981). To further motivate the introduction of the map (3.9) and aid the exposition of our method, we shall in the following assume that the eigenvalues of Λ_I are semi-simple so that each corresponds to a linearly independent column of R_I and that $\Lambda_{IJ} = 0$, which corresponds to a diagonalisable M. Define the eigenprojection onto the subspace spanned by eigenvectors associated with roots in $\mathcal{L}_I$ as

$$P_I \coloneqq \sum_{\lambda \in \mathcal{L}_I} \lambda r_\lambda l_\lambda^H,$$

where r_λ and l_λ belong to root λ. Now, consider

Example 3. Let

(3.18) $$M = \begin{bmatrix} 0.8 & 0.5 \\ 0 & 0.4 \end{bmatrix}.$$

Then, the right eigenvectors are $r_1^{\intercal} = \left[\begin{smallmatrix} 1 & 0 \end{smallmatrix}\right]$ and $r_2^{\intercal} = \left[\begin{smallmatrix} 1 & -0.8 \end{smallmatrix}\right]$ with eigenvalues are $\lambda_1 = 0.8$ and $\lambda_2 = 0.4$. Its left eigenvectors are $l_1^{\intercal} = \left[\begin{smallmatrix} 1 & 1.25 \end{smallmatrix}\right]$ and $l_2^{\intercal} = \left[\begin{smallmatrix} 0 & 1.6 \end{smallmatrix}\right]$. The associated eigenprojections are

(3.19) $$P_{\lambda_i} = r_i l_i^H$$

for $i \in \{1, 2\}$ the pair of matrices

$$P_{\lambda_1} = \begin{bmatrix} 1 & 1.25 \\ 0 & 0 \end{bmatrix} \qquad P_{\lambda_2} = \begin{bmatrix} 0 & -1.25 \\ 0 & 1 \end{bmatrix}$$

We see that P_{λ_i} are rank-1 projection matrices for which $P_{\lambda_i} P_{\lambda_j} = I_2 \delta_{ij}$ and if υ belongs to an eigenspace associated with root λ_i, we have $\upsilon P_{\lambda_i} = P_{\lambda_i} \upsilon$ and $P_{\lambda_1} + P_{\lambda_2} = I$. Finally, under the maintained null hypothesis that $\upsilon \in \mathcal{S}_I(M)$ for $I = \{1\}$,

$$P_{\lambda_1} \upsilon_{\perp} = 0.$$

Based on eigenprojections, it is easy to construct generalised inverses A^+ such that $AA^+A = A$. For a matrix of rank s, let the eigenprojection of A be as in Example 3. Then, a generalized inverse

$$A^+ = \sum_{i=1}^{s} \lambda_i^{-1} P_{\lambda_i}. \tag{3.20}$$

If we wish to relax the diagonalisability assumption, we may easily construct a generalised inverse via a singular value decomposition, which exists independently of diagonalisability. Proceed via

$$A^+ = \sum_{i=1}^{s} \iota_i^{-1} v_i u_i^{\mathsf{T}}, \tag{3.21}$$

where $\{\iota_i\}_{i=1}^{s}$ collects the non-zero singular values and u_i and v_i are left- and right-singular vectors of A. For the purpose of inverting estimated covariance matrices, either (3.20) or (3.21) would be suitable.

4. Asymptotic distribution of estimators

This section details estimators of basis vectors for singular and invariant subspaces. In general, neither invariant nor singular vectors are uniquely determined from the associated matrix decompositions. For example, $\operatorname{col} R_I$ is an invariant subspace iff $MR_I = R_I \Lambda_I$ for some general Λ_I and continues to hold also for $\operatorname{col} \bar{R}_I = \operatorname{col} R_I$ for $\bar{R}_I := R_I Q$ for some full-rank matrix $Q \in \mathbb{R}^{q \times q}$. Similarly, for $M = U_I \Sigma_I V_I^{\mathsf{T}}$, only Σ_I is unique if we order the singular values. The matrix U_I is in general not unique.

To circumvent the lack of uniqueness, we shall normalise our estimates of R_I and U_I in the following manner. For a candidate υ, we partition

$$\upsilon =: \begin{bmatrix} \upsilon_1 \\ \upsilon_2 \end{bmatrix},$$

where $\upsilon_1 \in \mathbb{R}^{r \times q}$ and $\upsilon_2 \in \mathbb{R}^{q \times q}$. Then, requiring $\upsilon^{\mathsf{T}} N \upsilon = I_q$ for

$$N = \begin{bmatrix} \left(\upsilon_2^{\mathsf{T}} \upsilon_2\right)^{-1} & 0 \\ 0 & \upsilon_2^{-1\mathsf{T}} \left(I_q - \upsilon_1^{\mathsf{T}} \left(\upsilon_2^{\mathsf{T}} \upsilon_2\right)^{-1} \upsilon_1\right) \upsilon_2^{-1} \end{bmatrix},$$

or, equivalently, letting $D^{\mathsf{T}} := \upsilon_1 \upsilon_2^{-1}$, we have

$$\upsilon := \begin{bmatrix} D^{\mathsf{T}} \\ I_q \end{bmatrix}, \tag{4.1}$$

and

(4.2) $$v_\perp := \begin{bmatrix} I_r \\ -D \end{bmatrix}$$

Unfortunately, uniqueness comes at the steep cost of making the second order moment potentially unbounded moments as Anderson (2010) discusses. Letting Ω_M be as in (2.2), we have in the case of homoskedastic columns of $\hat{M}$ the alternative normalization $v^\mathsf{T}\Omega_M v = I_q$, but shall not pursue these arguments further.[4]

To estimate basis vectors, we define for $r' \leq r$ the set estimator

(4.3) $$\widehat{v}_{\perp,n} := \left\{ v_\perp \in \mathbb{R}^{p\times r'} \ : \ \hat{\Phi}\left(v_\perp^\mathsf{T}\right) = 0 \right\}$$

for

$$\hat{\Phi}\left(v_\perp^\mathsf{T}\right) := \begin{cases} \psi\left(\hat{M}_n, v_\perp^\mathsf{T}\right) & \text{invariant subspace.} \\ \Psi\left(\hat{M}_n, v_\perp^\mathsf{T}\right) & \text{singular subspace.} \end{cases}$$

4.1. **Invariant subspaces.** We define the estimator of $\psi(M)$ by evaluating ψ at the estimate $\hat{M}$, so that $\psi\left(\hat{M}_n\right) = v_\perp^\mathsf{T} \hat{P}_I\left(\hat{M}_n\right)$. Using Lemma 1, we obtain the asymptotic distribution of the sample analogue of $\psi(M)$ in the following result.

Proposition 1. *Suppose Assumption 1 holds and let B as in* (3.12) *and let $v_\perp$ be as in* (1.3). *Then,*

$$\sqrt{n}\,\mathrm{vec}\, v_\perp^\mathsf{T} \hat{P}_I\left(\hat{M}_n\right) \rightsquigarrow N\left(0, B\Omega B^H\right).$$

Proposition 1 forms the basis for the Wald statistic. For general eigenvectors, i.e. if the set of eigenvalues $\mathcal{L}_I$ is not closed under conjugation, $\sqrt{n}\,\mathrm{vec}\, v_\perp^\mathsf{T} \hat{P}_I\left(\hat{M}_n\right)$ converges to a multivariate complex normal distribution, with covariance $B\Omega B^H$ and relation $B\Omega B^\mathsf{T}$. For now, we shall continue to assume that $\mathcal{L}_I$ is indeed closed under conjugation or, in the case of $|\mathcal{L}_I| = 1$ that the principal invariant vector is real so that we can avoid the complex normal distribution. However, we shall return to this issue in Section 7.1, when we deal with absolute values of the coefficients in v.

For $\hat{P}_I := \hat{R}_{I,n}\hat{L}_{I,n}^H$ using sample analogues of R_I and L_I, we have the estimator in (4.3). For a closed-form expression of $\hat{D}_n$ under the normalization (4.2), partition R_I conformably for $R_{I,1} \in \mathbb{R}^{r\times q}$ and $R_{I,2} \in \mathbb{R}^{q\times q}$ with $\mathrm{rk}\, R_{I,2} = q$.

(4.4) $$\begin{bmatrix} R_{I,1} \\ R_{I,2} \end{bmatrix} := R_I.$$

Consequently, define

(4.5) $$\hat{D}_{I,n}^\mathsf{T} := \hat{R}_{I,1,n}\hat{R}_{I,2,n}^{-1}.$$

[4]For an approach that avoids normalizations see Silin and Fan (2020) in the context of symmetric matrices.

and observe that (4.5) supplies a unique estimator. Our next result characterises the asymptotic distribution of $\hat{D}_{I,n}^{\intercal}$. Define

$$B_{D^{\intercal}} := \left(R_{I,2}^{\intercal -1} \otimes v_{\perp}^{\intercal} R_J\right) \left\{\left(\Lambda_I^{\intercal} \otimes I_J\right) - \left(I_I \otimes \Lambda_J\right)\right\}^{-1} \left(R_I^{\intercal} \otimes L_J^{\intercal}\right) \tag{4.6}$$

Theorem 1. *Let $\hat{D}_n$ be as in* (4.5) *and suppose Assumption 1 holds. Then,*

$$n^{1/2} vec \left\{\hat{D}_n^{\intercal} - D^{\intercal}\right\} \rightsquigarrow N\left(0, B_{D^{\intercal}} \Omega B_{D^{\intercal}}^H\right).$$

To obtain some intuition for how the distribution in Theorem 1 arises, it is instructive to consider the first term in the Jacobian matrix (4.6), $R_{I,2}^{\intercal -1}$, which is exactly the factor post-multiplying $R_{I,1}$ in (4.4), which demonstrates that we can derive the distribution of the estimator for the normalised basis coefficients collected in D from the same mapping ψ defined in (3.9). The above results rely on consistent covariance matrix estimators. To establish those, we have

Theorem 2. *Under Assumption 1, the following hold:*

(1) $\hat{M}_n \xrightarrow{p} M$.
(2) $\hat{\Omega}_n \xrightarrow{p} \Omega$.
(3) $\operatorname{col} \hat{R}_n \xrightarrow{p} \operatorname{col} R$.
(4) $\operatorname{col} \hat{L}_n \xrightarrow{p} \operatorname{col} L$.
(5) $\hat{\Lambda} \xrightarrow{p} \Lambda$.
(6) $\hat{B} \xrightarrow{p} B$.
(7) $\hat{B}_j \hat{\Omega} \hat{B}_j^{\intercal} \xrightarrow{p} B_j \Omega B_j$ *for* $j \in \left\{\emptyset, D^{\intercal}, W\right\}$.
(8) $\hat{D}_n \xrightarrow{p} D$.

The proof of Theorem 2 appears in Section 9.5. Note that Theorem 2.3 and 4 implies that we identify only the column spans R and L. The main utility of this result is the application of Slutsky's theorem to establish weak convergence of our test statistics.

4.2. **Singular subspaces.** We define the estimator of $\Psi(M)$ by evaluating Ψ at the estimate $\hat{M}$, so that $\Psi\left(\hat{M}_n\right) = v_{\perp}^{\intercal} \hat{U}_I\left(\hat{M}_n\right)$. Using Lemma 2, we obtain the asymptotic distribution of the sample analogue of $\Psi(M)$ in the following result.

Proposition 2. *Suppose Assumption 1 and Assumption 3 hold and let B_{SVD} be as in* (3.17) *and let $v_{\perp}$ be as in* (1.3). *Then,*

$$\sqrt{n}\, vec\, v_{\perp}^{\intercal} \hat{U}_I\left(\hat{M}_n\right) \rightsquigarrow N\left(0, B_{SVD} \Omega B_{SVD}^{\intercal}\right).$$

The result in Proposition 2 forms the basis for a Wald test for the orthocomplement of basis vectors of the right-singular subspace of M.

5. Hypothesis tests

5.1. **Invariant subspaces.**

5.1.1. *Wald test for matrix hypothesis.* To test $H_0^* : \upsilon_\perp^\intercal P_I = 0$ for a fully specified $\upsilon_\perp$ corresponding to some υ_0, we define a Wald test statistic based on Proposition 1. Letting $\hat{\Omega}_W^+$ denote a generalised inverse of $\hat{\Omega}_W$, defined in either (3.20) or (3.21), for $\hat{\Omega}_W := \hat{B}_W \hat{\Omega} \hat{B}_W$, we have

$$\hat{W}_n\left(\upsilon_\perp\right) := \text{vec}\left(\upsilon_\perp^\intercal \hat{P}_I\right)^H \hat{\Omega}_W^+ \text{vec}\left(\upsilon_\perp^\intercal \hat{P}_I\right) \tag{5.1}$$

and an associated Jacobian matrix estimator for

$$\hat{B}_W := \left(\hat{L}_I \otimes \upsilon_\perp^\intercal \hat{R}_J\right)\left\{\left(\hat{\Lambda}_I^\intercal \otimes I_J\right) - \left(I_I \otimes \hat{\Lambda}_J\right)\right\}^{-1}\left(\hat{R}_I^\intercal \otimes \hat{L}_J^\intercal\right), \tag{5.2}$$

which we obtain by simply replacing sample analogues into the expression for the Jacobian of $\psi\left(X\right)$ defined in (3.12). Consequently, we obtain the asymptotic distribution of the Wald test in

Theorem 3. *Suppose Assumption 1 holds. Then,*

$$\hat{W}_n\left(\upsilon_\perp\right) \rightsquigarrow \chi^2_{qm}. \tag{5.3}$$

The proof of Theorem 3 appears in Section 9.1 and builds on the result in Proposition 1, which establishes the asymptotic normality of the vectors associated with the Wald test statistic. Section 6.1 collects information on the quality of the offered asymptotic approximation. The advantage of this test is that it is relatively straightforward to compute although a researcher has to specify a full vector or matrix hypothesis. We imagine that this scenario is most likely of interest when testing whether localised basis vectors belong to an invariant subspace of M.

5.1.2. *t-test for individual coefficients.* If researchers wish to conduct inference on individual coefficients of invariant vectors or matrices, it may be more useful to consider a scalar-valued t-test. This scenario is useful when We therefore consider a null hypothesis about individual entries of D, $d_{ij} := e_i^\intercal D e_j$ for $e_i \in \mathbb{R}^{q\times 1}$ and $e_j \in \mathbb{R}^{r\times 1}$ where the vectors e_i have unit entries at i and zero elsewhere.

Let $\hat{R}_{I,2}^{-1}$ be the empirical analogue of $R_{I,2}^{-1}$ defined in (4.4) and define analogously to (5.2),

$$B_{ij} = \left(e_i^\intercal R_{2,I}^{\intercal -1} \otimes e_j^\intercal \upsilon_\perp^\intercal R_J\right)\left[\left(\Lambda_I^\intercal \otimes I_J\right) - \left(I_I \otimes \Lambda_J\right)\right]^{-1}\left(R_I^\intercal \otimes L_J^\intercal\right),$$

to be able to write the scalar variance $\sigma_{ij}^2 := B_{ij}\Omega B_{ij}^H$ of $\sqrt{n}\left(\hat{d}_{ij} - d\right)$ Then, we define the t-test statistic for inference on the ijth coefficient of D,

$$t_{ij,n}\left(d_0\right) := \frac{\hat{d}_{n,ij} - d_0}{\sqrt{\hat{\sigma}_{ij}^2/n}}. \tag{5.4}$$

We construct the estimator of the variance, $\hat{\sigma}_{ij}$ by replacing B_{ij} with $\hat{B}_{ij}$ which in turn contains sample analogues of $\hat{R}_i$ and $\hat{L}_i$ for $i \in \{I, J\}$ which are similar to the replacements appearing in (5.2). Our next result collects the weak limit of $t_{ij,n}\left(d_0\right)$.

Theorem 4. *Suppose Assumption 1 holds, then*

$$t_{ij,n}\left(d_0\right) \rightsquigarrow N\left(0,1\right).$$

The result Theorem 4 is particularly convenient for constructing confidence intervals on individual coefficients. Furthermore, we consider an important special case that arises in the study of eigenvector centralities that graph adjacency matrices induce. In such cases, we are interested in the distribution of $\left|\hat{d}_{ij}\right|$, which will be a folded normal distribution with cumulative distribution function

$$F_G\left(x; d_{ij}, \sigma_{ij}\right) \coloneqq \Phi\left(\frac{x - d_{ij}}{\sigma_{ij}}\right) - \Phi\left(\frac{-x - d_{ij}}{\sigma_{ij}}\right) \tag{5.5}$$

where $\Phi\left(\frac{x-\mu}{\sigma}\right)$ denotes the normal cumulative distribution function with mean μ and variance σ^2. For the asymptotic distribution of the absolute values of the normalised eigenvector entries, we record

Corollary 1. *Suppose Assumption 1 holds, then* $\sqrt{n}\left|\hat{d}_{ij}\right| \rightsquigarrow G$ *where* G *is a random variable that has a folded normal distribution with c.d.f.* $F_G\left(x\right)$.

Proof. See Tsagris, Beneki, and Hassani (2014). □

The result in Corollary 1 allows one to construct one-sided confidence intervals, which appear in Section 7.1.

5.2. **Inference on clustering coefficients and singular subspaces.** In this section, we develop methods to conduct inference on singular subspaces in the case of M being an $m \times n$ rectangular matrix. We begin by considering a Wald test statistic which is amenable for inference on subspaces or individual vectors and then move on to consider t-tests for individual coefficients. We specialise the latter to clustering coefficients as they arise in network models.

5.2.1. *Wald test for matrix hypothesis on singular subspaces.*

5.2.2. *t-test for individual coefficients of singular vectors.*

6. Simulation study

We ran simulations to study the performance of the t- and Wald tests using quantile-quantile plots as well as the empirical cumulative distribution functions compared with their theoretical counterparts. These visual aids demonstrate the quality of the asymptotic approximations found in Section 5.2 and Section 5.2.[5]

6.1. **Monte Carlo evidence for invariant subspace inference.**

6.1.1. *Data-generating processes.* We used a simplified covariance matrix $\Omega = I_p \otimes \Omega_M$ for a matrix with $p = 2$ and detail the data-generating process in Algorithm 6.1. The online supplement to this paper contains further DGPs as well as examples on how to include heteroskedasticity and autocorrelation-robust covariance matrix estimators.

[5]Complementary to the results presented in this section, we refer the reader to the supplementary material hosted at https://github.com/jsimons8/networkmodelssubspaces.

Algorithm 6.1 DGP underlying results in Figure 6.1.

Input: Dimension of the matrix p, sample size n, Monte Carlo repetitions K, Dimension of invariant subspace of interest q, a candidate matrix M, and a null hypothesis $\upsilon_\perp = \begin{pmatrix} I_r & -D^\intercal \end{pmatrix}$

Output: K samples of t and Wald statistic.

Generate n samples of $M_t = M + E_t \in \mathbb{R}^{p\times p}$, K times, where the columns of E are drawn from a mean zero multivariate normal distribution with covariance matrix Ω_W. Store the result.

for $i = 1$ **to** K **do**

  Estimate $\hat{M}_n$ via $n^{-1}\sum_{t=1}^{n} M_t$.

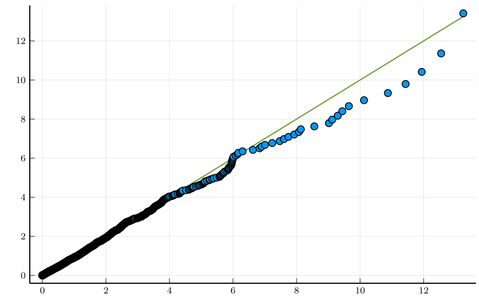
  Estimate $\hat{R}_n$ and $\hat{L}_n$ and partition them like in (3.4), $R_I = R[:, 1 : q]$ and $R_J = R[:, q + 1 : p]$. Find $L = \begin{pmatrix} L_I & L_J \end{pmatrix}$ via (generalised) inversion.

  Construct $\hat{D}$ based on (4.5), take the (i, j) element for $d_{i,j}$.

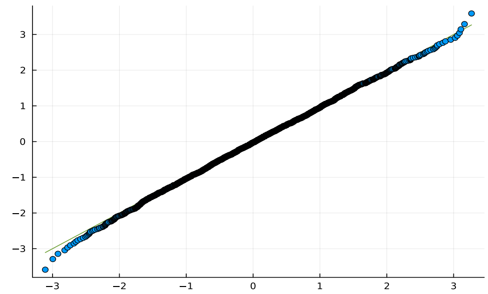
  Construct Wald or t-statistic for null hypothesis υ based on (5.1) or (5.4), respectively.

  For the t-statistic, take the real part only. The Wald statistic will necessarily be real.

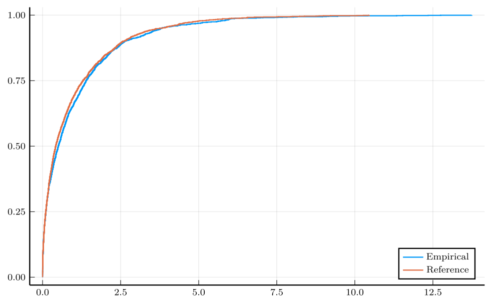
end for



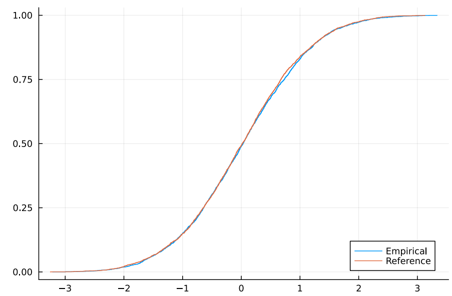
(a) Wald test quantiles.



(b) t-test quantiles.



(c) CDF comparisons of Wald test statistic.



(d) CDF comparisons of t-test statistic.

FIGURE 6.1. Quality of asymptotic approximation. Left column: Wald test, right column: t-test. 5000 MC repetitions were used for a sample size of 100. : QQ plots are theoretical (y) vs. empirical (x).

6.1.2. *Results.* Figure 6.1 shows that both the Wald and t-test statistics perform well in simulation exercises. The left panel lets us judge the approximation made in Theorem 3 while the right panel displays the quality of the t-test approximation of Theorem 4. The overall performance is very good with only few outliers. Panel (b) shows that the t-test has slightly weaker tails compared with the standard normal distribution, although they are in the acceptable range. Histograms overlain with densities displayed the same tendency and appear in the online supplement. The bottom panels show the same tendency where the cumulative distribution functions track their empirical counterparts fairly well.

Algorithm 6.2 DGP underlying results in Figure 6.2.

Input: Dimension of the matrix $m \times l$, sample size n, Monte Carlo repetitions K, Dimension of singular subspace of interest F, (if applicable) a singular vector of interested indexed by f, a candidate matrix M, and a null hypothesis $v_\perp = \begin{pmatrix} I_r & -D^\intercal \end{pmatrix}$

Output: K samples of t and Wald statistic.

Generate n samples of $M_t = M + E_t \in \mathbb{R}^{m \times l}$, K times, where the columns of E are drawn from a mean zero multivariate normal distribution with covariance matrix Ω_W. Store the result.

for $i = 1$ **to** K **do**

  Estimate $\hat{M}_n$ via $n^{-1} \sum_{t=1}^{n} M_t$.

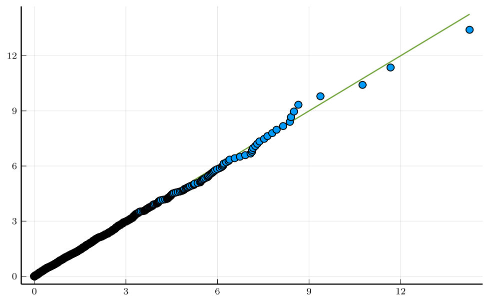
  Estimate $\hat{U}_n$ and $\hat{V}_n$ and partition them like in (3.4), $R_I = R[:, 1:q]$ and $R_J = R[:, q+1:p]$. Find $L = \begin{pmatrix} L_I & L_J \end{pmatrix}$ via (generalised) inversion.

  Construct $\hat{D}$ based on (4.5), take the (i, j) element for $d_{i,j}$.

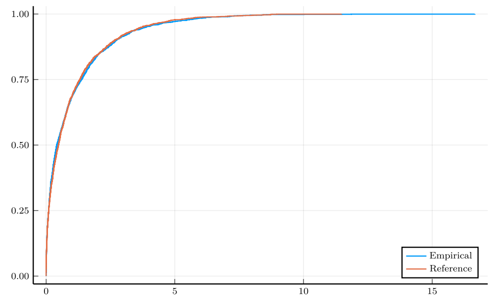
  Construct Wald or t-statistic for null hypothesis v based on (5.1) or (5.4), respectively.

  For the t-statistic, take the real part only. The Wald statistic will necessarily be real.

end for



(a) Wald test quantiles.



(b) CDF comparisons of Wald test statistic.

FIGURE 6.2. Quality of asymptotic approximation for Wald test-based inference on singular vectors. 2000 MC repetitions were used for a sample size of 500. : QQ plots are theoretical (y) vs. empirical (x).

6.2. Monte Carlo evidence for inference on singular vectors.

6.2.1. *Data-generating processes.* We considered a data-generating process using a covariance matrix $\Omega = I_m \otimes \Omega_W$ for a positive-definite $\Omega_W \in \mathbb{R}^{m \times m}$. Algorithm 6.2 details the steps of how we generate samples for the t- and Wald statistics.

6.2.2. *Results.* The quantile-quantile plots in the top row of Figure 6.2 show that the overall performance of the asymptotic approximations is very good with few outlier. Similarly, the bottom row details that the cumulative distribution functions track their empirical counterparts well. These results are encouraging that the perturbation theory delivers good approximations to the derivatives of the singular subspace maps.

7. Network statistics

In this section, we apply the methods of invariant and singular subspace inference to the empirical questions of estimating adjacency matrices and eigenvector centrality measures. We focus on directed graphs with potential self-loops, which trigger non-symmetric adjacency matrices and whose principal invariant subspaces collect information on the popularity of nodes.

We begin by connecting the theory of invariant subspaces to centrality measures as they arise from a network adjacency matrix. To study the performance of our inference methods in this setting, we therefore being by explicitly simulating adjacency matrices from a directed graph model. Subsequently, we consider two empirical examples and estimate trade and input-output networks from noisy data for which we estimate eigenvector centralities with associated confidence intervals. We find that in the case of input-output networks, the inclusion of confidence intervals reorders the nodes according to their importance in the network.

7.1. **Network centralities.**

We interpret M as the adjacency matrix representing an unweighted, directed graph with pairs of vertices indexed by i and j. The importance score s_i,

$$s_i := \frac{1}{\lambda} \sum_{j \in N_i} s_j, \tag{7.1}$$

measures the relative importance of firm i in the network. The sum runs over all j that are direct neighbours of firm i, denoted by N_i and λ is a constant. The measure s_i is known as the eigenvector centrality. For the full vector of scores s, making the substitution $\sum_{j \in N_i} s_j = \sum_{j=1}^{N} m_{ij} s_j$, we obtain the relationship

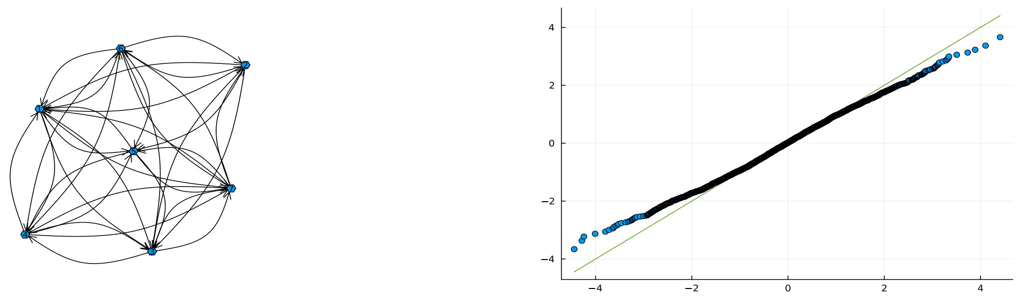
$$Ms = \lambda s, \tag{7.2}$$

which defines an eigenvector problem. We consider an estimate of M, $\hat{M}_n$, estimated with uncertainty which arises for example in the setup of de Paula, Rasul, and Souza (2023). We aim to test hypotheses of the form

$$H_0 \ : \ s_i \text{ is the relative importance of node } i.$$

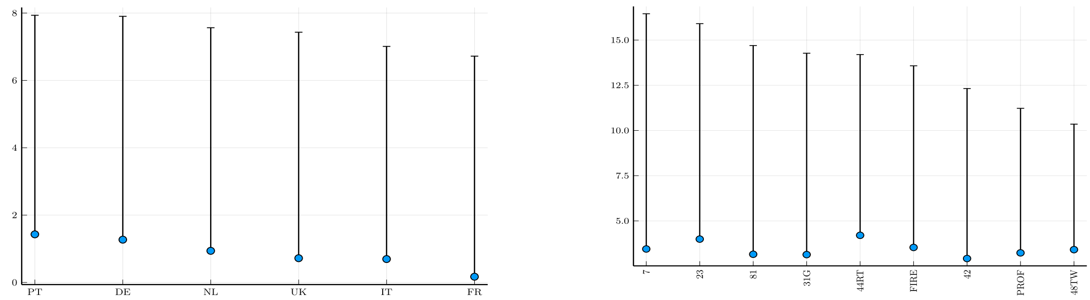
against the alternative that it is not. We use the absolute value version of the coefficient-wise t-test based on Corollary 1 to construct one-sided confidence intervals. We define a one-sided confidence interval with level α for $|s_i|$ via

$$\mathcal{C}_i := \left[|\hat{s}_i| , x_0 \right],$$



(a) Directed graph belonging to the adjacency matrix M. (b) Quality of approximation

FIGURE 7.1. Example graph measured with noise and quality of the asymptotic approximation for inference on eigenvector centralities. 1000 MC repetitions were used for a sample size of 500. : QQ plots are theoretical (y) vs. empirical (x).



(a) Centralities for trade network (logarithms). (b) Centralities for input-output network.

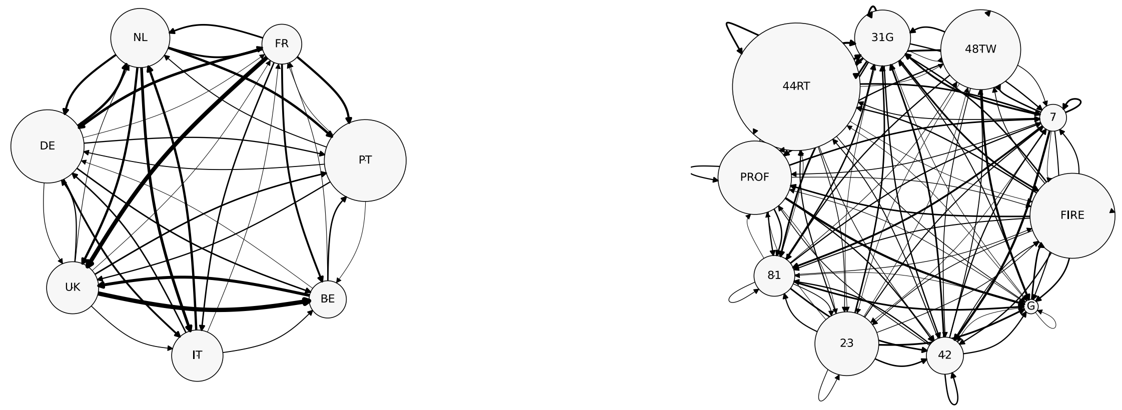
FIGURE 7.2. Left column: One-sided confidence intervals (95%) for eigenvector centralities.

Code	Description
23	Construction
31G	Manufacturing
42	Wholesale trade
44RT	Retail trade
48TW	Transportation and warehousing
FIRE	Finance, insurance, real estate, rental, and leasing
PROF	Professional and business services
7	Arts, entertainment, recreation, accommodation, and food services
81	Other services, except government
G	Government

TABLE 7.1. Sector codes for the input-output network model.

where the upper limit $x_0 > 0$ solves $F_G\left(x_0; |s_i|, \sigma_i\right) = 1 - \alpha$ for $x_0 > 0$ and (5.5) defines F_G and with consistent point estimates $|\hat{s}_i|$and $\hat{\sigma}_i$ substituted for $|s_i|$ and σ_i for $i = 1, \ldots, p-1$. Consequently, $\forall s \in \mathcal{C}$, we cannot reject $H_0 \ : \ |d_{ij}| = d$.

7.2. **Network model of international trade.** We next consider a setting where network ties are observed with error repeatedly, resulting in a noisy adjacency matrix. In particular, we are taking the



(a) Estimated trade network. (b) Estimated input-output network.

FIGURE 7.3. Estimated networks.

8. Conclusion

This paper has extended the theory of Tyler (1981) to cover non-symmetric matrices. In addition to testing full vector hypotheses, we also provide a t-test on individual coefficients of the vector and derive the asymptotic distribution of the estimator for normalised eigenvectors. We caution the reader that normalised estimates may suffer from a potentially unbounded second moment. However, we believe that the benefits of being able to normalise coefficients outweigh its drawbacks.

We have also seen how to apply the analytic perturbation theory of Sun (1991) to the problem of inference on eigenvectors and hope that our exposition can aid researchers in similar settings. The perturbation theory of Kato (1995) delivers the same symbolic results but lacks the discussion on the smoothness of maps. For readers who are new to the literature on perturbation theory, Greenbaum, Li, and Overton (2019) contains a pedagogic and wonderfully detailed treatment of the subject.

The leading application of our results is to graph Laplacians or adjacency matrices that directed graphs induce. We leave further development of a full model for inference on noisy graph Laplacians for future work.

9. Appendix

This section contains proofs of the main results of the paper. In particular, we will see how the χ^2 and normal approximations to the limiting distributions to the Wald and t-tests statistics in Theorem 3 and Theorem 4 arise by considering asymptotic perturbation expansions that depend on the spectral data of M. Following on, we shall study how these expansions arise from perturbation theory and establish that the implicit function theorem in our setup implies that the unnormalised basis vectors are analytic functions of the underlying matrices. This discussion then leads us to furnish a consistency proof that takes advantage of the implied continuity of eigenvectors and eigenvalues. We note that this result is of independent interest because eigenvectors are not in general continuous functions of the underlying matrix. They admit continuous representations if the eigenvalues split according to Assumption 1.2.

9.1. **Asymptotic distribution of Wald test statistic.** To prove Theorem 3, recall that

$$v_\perp^\top \hat{P}_{n,I} = v_\perp^\top \left(\hat{P}_{n,I} - P_I\right) \tag{9.1}$$

because $v_\perp^\top P_I = 0$ under the maintained hypothesis. We introduce the notation

$$\begin{bmatrix} \Delta_I & \\ \Delta_{IJ} & \Delta_J \end{bmatrix} := \begin{bmatrix} L_I^H E R_I & \\ L_J^H E R_I & L_J^H E R_J \end{bmatrix}, \tag{9.2}$$

which introduces $\Delta_\cdot$ to denote the diagonalisation of a perturbing matrix $E \in \mathbb{R}^{p\times p}$.

To expand the right-hand side of (9.1), we introduce the derivatives in

Lemma 3. *Suppose Assumption 1 holds. Then, the first and second order Gateaux derivatives of $\psi(M,E)$ at M are*

$$\dot{\psi}(X,E)\,|_{X=M} = v_\perp^\top R_J \boldsymbol{S}^{-1}(\Delta_{SL}) L_I^H. \tag{9.3}$$

$$\ddot{\psi}(X,E)\,|_{X=M} = 2 v_\perp^\top R_J \boldsymbol{S}^{-1}\left(\left(\Delta_J \boldsymbol{S}^{-1}(\Delta_{IJ}) - \boldsymbol{S}^{-1}(\Delta_{IJ})\Delta_I\right)\right) L_I^H. \tag{9.4}$$

The Gateaux derivatives in Lemma 3 crucially depend on the perturbation E. To operationalise these results for a first order expansion of $\psi\left(\hat{M}_n\right)$ about $\psi(M)$, we invoke Magnus and Neudecker (2019, Thm. 11) to write them as

Corollary 2. *Suppose Assumption 1 holds. Then, the first order differential form of $\psi(M,E)$ at M is*

$$\mathit{vec}\ \mathrm{d}\psi(M) = \mathit{vec}\, v_\perp^\top R_J \boldsymbol{S}^{-1}(\Delta_{SL}) L_I^H,$$

which lets us identify the Jacobian matrix B via

$$\mathit{vec}\ \mathrm{d}\psi(M;E) = \underbrace{\left(L_J \otimes v_\perp^\top R_I\right)\left\{\left(\Lambda_I^\top \otimes I_J\right) - \left(I_I \otimes \Lambda_J\right)\right\}^{-1}\left(R_I^\top \otimes L_J^\top\right)}_{B} \mathit{vec}\, E$$

The first derivative (9.3) thus identified allows assembly of a first-order expansion, which we pursue in

Lemma 4.

(1) *The standardized term admits the expansion centered around the truth*

$$\sqrt{n} v_\perp^\top \left(\hat{P}_{n,I} - P_I\right) = \sqrt{n} v_\perp^\top \hat{R}_{n,J} \boldsymbol{S}^{-1}(\Delta_{n,IJ}) \hat{L}_{n,I}^H + \sqrt{n} r\left(\hat{M}_n - M\right).$$

(2) *The remainder term obeys $r\left(\hat{M}_n - M\right) = O\left(n^{-1}\right)$.*

(3) *Let $\hat{\Delta}_{n,i}$ with $i \in \{I, IJ, J\}$ be as in* (9.2). *Then*

$$\begin{bmatrix} \hat{\Delta}_{n,I} & \\ \hat{\Delta}_{n,IJ} & \hat{\Delta}_{n,J} \end{bmatrix} = \begin{bmatrix} O_p\left(n^{-1/2}\right) & \\ O_p\left(n^{-1/2}\right) & O_p\left(n^{-1/2}\right) \end{bmatrix}.$$

Lemma 4.1 lets us derive an estimator of the covariance matrix while Lemma 4.2 ensures that the approximation error is $O\left(n^{-1/2}\right)$ and therefore converges in probability to zero. Our next result ensures the desired convergence. Define $B_{\mathrm{IJ}}^{\top} := R_I \otimes L_J$. Then, we have

Lemma 5. *Let $\Delta_{n,IJ}$ be as in Lemma 3. Then, for $\upsilon_{\perp} \in \mathbb{R}^{p\times(p-m)}$,*

(1) $\sqrt{n}\mathit{vec}\,\Delta_{n,IJ} \rightsquigarrow N\left(0, B_{IJ}\Omega B_{IJ}^{H}\right).$
(2) $\sqrt{n}\mathit{vec}\left\{\upsilon_{\perp}^{\top}\hat{R}_{n,J}\boldsymbol{S}^{-1}\left(\Delta_{n,IJ}\right)\hat{L}_{n,I}^{H}\right\} \rightsquigarrow N\left(0,\Omega_W\right).$
(3) $\operatorname{rk}\Omega_W = qm.$

Proposition 1 follows from application of Lemma 5.2 to Lemma 4.1. Similarly, Theorem 3 follows from Lemma 5.2 where the degrees of freedom follow from Lemma 5.3. The consistency of the covariance matrix and right and left eigenvector column space estimates that is necessary for the convergences to hold jointly follows from Theorem 2.

9.2. **Asymptotic distributions of estimator for D and t-test statistic.** To prove Theorem 1, apply to $\sqrt{n}\left(\hat{D}_n^{\top} - D^{\top}\right)$ the expression given in

Lemma 6. *Let the estimator $\hat{D}_n^{\top} = \hat{R}_{n,1,I}\hat{R}_{n,2,I}^{-1}$. Then, the centered estimator*

$$\left(\hat{D}_n - D\right)^{\top} = \upsilon_{\perp}^{\top}\hat{R}_{n,1,I}\hat{R}_{n,2,I}^{-1} + o_p\left(1\right). \tag{9.5}$$

To study the standardised and centred estimator, we can work with the right-hand side of (9.5). For the asymptotic limit, we have in analogy to Lemma 4,

Corollary 3. *We have the expansion for the centered estimator*

$$\sqrt{n}\left(\hat{D}_n^{\top} - D^{\top}\right) = \sqrt{n}\upsilon_{\perp}^{\top}\hat{R}_{n,J}\boldsymbol{S}^{-1}\left(\Delta_{n,\mathrm{IJ}}\right)\hat{R}_{n,2,I}^{-1} + \sqrt{n}r\left(\hat{M}_n - M\right)$$

We note that Lemma 4.2 and 3 remain unchanged in the case of expansions for $\hat{D}_n$ and the associated t-test statistic.

Theorem 1 follows from applying Lemma 5 to the expansion in Corollary 3 by substituting $R_{n,2,I}^{-1}$ for $L_{n,I}^{H}$.

We now turn to proving Theorem 4. By Theorem 1 and Assumption 1,

$$\sqrt{n}\left(\hat{B}_D\hat{\Omega}_M\hat{B}_D^{\top}\right)^{-1/2}\left(\text{vec}\,\hat{D}_n^{\top} - \text{vec}\,D^{\top}\right) \rightsquigarrow N(0, I_{rq\times rq}).$$

Let $e_{s,i} \in \mathbb{R}^s$ denote a vector with zero everywhere except for a 1 in the sth position so that $d_{ij} := e_{i,q}^{\top}D^{\top}e_{j,r}$ is the entry of $D^{\top}$ in row i and column j, which corresponds to the ijth element of vec $\hat{D}_n^{\top}$ − vec $D_0^{\top}$. To find its associated standard error, we let σ_{ij}^2 denote the ijth diagonal entry of $\left(\hat{B}_D\hat{\Omega}_M\hat{B}_D^{H}\right)^{-1}$. Then, we define as the standard error for d_{ij}, $\hat{\sigma}_{ij} := \sqrt{\frac{\hat{\sigma}_{ij}^2}{n}}$ and find that $\frac{\hat{d}_{ij}-d_{ij}}{\hat{\sigma}_{ij}} \rightsquigarrow N\left(0,1\right)$. Note that $\hat{B}_D\hat{\Omega}_M\hat{B}_D^{H}$ is always real by construction in analogy to the covariance matrix of the Wald test statistic.

9.3. **Perturbation theory for left singular vectors.** We write for the mapping that represents the test statistic

$$\begin{aligned}&\Psi\left(M\left(t\right)\right)\\=&\upsilon_{\perp}^{\intercal}U_{I}\left(t\right)\\=&\upsilon_{\perp}^{\intercal}\left(\tilde{U}_{I}-U_{I}\right)\left(t\right).\end{aligned}$$

To first-order, we expand $\Psi\left(M\left(t\right)\right)$, using the expansion in Liu et al.:

$$\begin{aligned}&\Psi\left(M\left(t\right)\right)\\=&\upsilon_{\perp}^{\intercal}U_{I}\left(D\cdot\left(U_{I}^{\intercal}E\left(t\right)V_{I}\Sigma_{I}+\Sigma_{I}V_{I}^{\intercal}E^{\intercal}\left(t\right)U_{I}\right)\right)\\&+\upsilon_{\perp}^{\intercal}U_{J}U_{J}^{\intercal}E\left(t\right)V_{I}\Sigma_{I}^{-1}.\end{aligned}$$

In order to identify the Jacobian of the perturbation expansion of $\Psi\left(M\left(t\right)\right)$, we need to find vec $\Psi\left(M\left(t\right)\right)$, for which we write

$$\begin{aligned}&\text{vec }\Psi\left(M\left(t\right)\right)\\=&\text{vec }\upsilon_{\perp}^{\intercal}U_{I}\left(D\cdot\left(U_{I}^{\intercal}E\left(t\right)V_{I}\Sigma_{I}+\Sigma_{I}V_{I}^{\intercal}E^{\intercal}\left(t\right)U_{I}\right)\right)\\&+\text{vec }\upsilon_{\perp}^{\intercal}U_{J}U_{J}^{\intercal}E\left(t\right)V_{I}\Sigma_{I}^{-1}.\end{aligned}$$

We focus on the second term first:

$$\begin{aligned}&\text{vec }\upsilon_{\perp}^{\intercal}U_{J}U_{J}^{\intercal}E\left(t\right)V_{I}\Sigma_{I}^{-1}\\=&\left(\Sigma_{I}^{-1\intercal}V_{I}^{\intercal}\otimes\upsilon_{\perp}^{\intercal}U_{J}U_{J}^{\intercal}\right)\text{vec }E\left(t\right)\end{aligned}$$

Regarding the first term,

$$\begin{aligned}&\text{vec }\upsilon_{\perp}^{\intercal}U_{I}\left(D\cdot\left(U_{I}^{\intercal}E\left(t\right)V_{I}\Sigma_{I}+\Sigma_{I}V_{I}^{\intercal}E^{\intercal}\left(t\right)U_{I}\right)\right)I_{F}\\=&\left(I_{F}\otimes\upsilon_{\perp}^{\intercal}U_{I}\right)\left[\left(\text{vec }D\right)\cdot\text{vec }\left(U_{I}^{\intercal}E\left(t\right)V_{I}\Sigma_{I}+\Sigma_{I}V_{I}^{\intercal}E^{\intercal}\left(t\right)U_{I}\right)\right]\\=&\left(I_{F}\otimes\upsilon_{\perp}^{\intercal}U_{I}\right)\left[\left(\text{vec }D\right)\cdot\left(\left(\Sigma_{I}^{\intercal}V_{I}^{\intercal}\otimes U_{I}^{\intercal}\right)\text{vec }\left(E\left(t\right)\right)+\left(U_{I}^{\intercal}\otimes\Sigma_{I}V_{I}^{\intercal}\right)K\text{vec }E\left(t\right)\right)\right]\\=&\left(I_{F}\otimes\upsilon_{\perp}^{\intercal}U_{I}\right)\left[\left(\text{vec }D\right)\cdot\left(\left(\left(\Sigma_{I}^{\intercal}V_{I}^{\intercal}\otimes U_{I}^{\intercal}\right)+\left(U_{I}^{\intercal}\otimes\Sigma_{I}V_{I}^{\intercal}\right)K\right)\text{vec }E\left(t\right)\right)\right]\end{aligned}$$

where I_F is an $F\times F$ identity matrix and F is the true rank of M. We now apply Lemma 7 to $\left(\text{vec }D\right)\cdot\left[\left(\Sigma_{I}^{\intercal}V_{I}^{\intercal}\otimes U_{I}^{\intercal}\right)\text{vec }E\left(t\right)\right]$ to obtain $\left[\left(\mathbf{1}^{\intercal}\otimes\text{vec }D\right)\cdot\left(\Sigma_{I}^{\intercal}V_{I}^{\intercal}\otimes U_{I}^{\intercal}\right)\right]\text{vec }E\left(t\right)$ and similarly, $\left(\text{vec }D\right)\cdot\left(U_{I}^{\intercal}\otimes\Sigma_{I}V_{I}^{\intercal}\right)K\text{vec }E\left(t\right)=\left[\left(\mathbf{1}^{\intercal}\otimes\text{vec }D\right)\cdot\left(U_{I}^{\intercal}\otimes\Sigma_{I}V_{I}^{\intercal}\right)K\right]\text{vec }E\left(t\right)$. Therefore, we obtain

$$\begin{aligned}&\text{vec }\Psi\left(M\left(t\right)\right)\\=&\left(\Sigma_{I}^{-1\intercal}V_{I}^{\intercal}\otimes\upsilon_{\perp}^{\intercal}U_{J}U_{J}^{\intercal}\right)\text{vec }E\left(t\right)\\&+\left(I_{F}\otimes\upsilon_{\perp}^{\intercal}U_{I}\right)\left[\left(\mathbf{1}^{\intercal}\otimes\text{vec }D\right)\cdot\left(\left(\Sigma_{I}^{\intercal}V_{I}^{\intercal}\otimes U_{I}^{\intercal}\right)+\left(U_{I}^{\intercal}\otimes\Sigma_{I}V_{I}^{\intercal}\right)K\right)\right]\text{vec }E\left(t\right).\end{aligned}$$

Hence, we can write the map $\Psi\left(M\left(t\right)\right)$ in differential form as $\mathrm{d}\Psi\left(M;E\right)=:B_{\text{SVD}}\text{vec }E$ for a perturbation E and identify the Jacobian matrix invoking Magnus and Neudecker (2019, Thm.

11) as

$$B_{\text{SVD}} = \left(\Sigma_I^{-1\mathsf{T}} V_I^\mathsf{T} \otimes \upsilon_\perp^\mathsf{T} U_J U_J^\mathsf{T}\right) \\ + \left(I_F \otimes \upsilon_\perp^\mathsf{T} U_I\right) \left[\left(\mathbf{1}^\mathsf{T} \otimes \text{vec } D\right) \cdot \left(\left(\Sigma_I^\mathsf{T} V_I^\mathsf{T} \otimes U_I^\mathsf{T}\right) + \left(U_I^\mathsf{T} \otimes \Sigma_I V_I^\mathsf{T}\right) K\right)\right].$$

Lemma 7. *Let A, B, and C be matrices of dimensions $a \times b$, $ab \times cd$, and $c \times d$, respectively. Then, for a matrix of ones of dimension $cd \times 1$, $\mathbf{1}$, we have the identities*

$$\textit{vec } A \cdot (B \textit{vec } C) = \left[\left(\mathbf{1}^\mathsf{T} \otimes \textit{vec } A\right) \cdot B\right] \textit{vec } C$$

and

$$\textit{vec } (A \cdot B) = \textit{vec } A \cdot \textit{vec } B.$$

Writing out individual entries establishes the result. □

For a single vector of interest, the Jacobian simplifies considerably, for a single vector. Write for the statistic

$$\begin{aligned} &\Psi_s(M(t)) \\ =&\upsilon_\perp^\mathsf{T} u_i(t) \\ =&\upsilon_\perp^\mathsf{T} (\tilde{u}_i - u_i). \\ =&\upsilon_\perp^\mathsf{T} U_I D_i U_I^\mathsf{T} E(t) v_i s_i \\ +&\upsilon_\perp^\mathsf{T} U_I D_i \Sigma_I V_I^\mathsf{T} E^\mathsf{T}(t) u_i \\ +&\upsilon_\perp^\mathsf{T} U_J U_J^\mathsf{T} E(t) v_i s_i^{-1}. \end{aligned}$$

Again, we vectorize term by term, to obtain vec $\upsilon_\perp^\mathsf{T} U_I D_i U_I^\mathsf{T} E v_i s_i = \left(s_i v_i^\mathsf{T} \otimes \upsilon_\perp^\mathsf{T} U_I D_i U_I^\mathsf{T}\right) \text{vec } E(t)$, to obtain vec $\upsilon_\perp^\mathsf{T} U_I D_i \Sigma_I V_I^\mathsf{T} E^\mathsf{T}(t) u_i = \left(u_i^\mathsf{T} \otimes \upsilon_\perp^\mathsf{T} U_I D_i \Sigma_I V_I^\mathsf{T}\right) K \text{vec } E(t)$, and $\left(s_i^{-1} v_i^\mathsf{T} \otimes \upsilon_\perp^\mathsf{T} U_J U_J^\mathsf{T}\right) \text{vec } E(t)$

$$\begin{aligned} &\text{vec } \Psi_s(M(t)) \\ &= \left(s_i v_i^\mathsf{T} \otimes \upsilon_\perp^\mathsf{T} U_I D_i U_I^\mathsf{T}\right) \text{vec } E(t) \\ &+ \left(u_i^\mathsf{T} \otimes \upsilon_\perp^\mathsf{T} U_I D_i \Sigma_I V_I^\mathsf{T}\right) K \text{vec } E(t) \\ &+ \left(s_i^{-1} v_i^\mathsf{T} \otimes \upsilon_\perp^\mathsf{T} U_J U_J^\mathsf{T}\right) \text{vec } E(t), \end{aligned}$$

so that we can write the map $\Psi_s(M(t))$ in differential form as $\mathrm{d}\Psi_s(M; E) =: B_s \text{vec } E$, again invoking Magnus and Neudecker (2019, Thm. 11) where

$$\begin{aligned} &B_s \\ &= \left(s_i v_i^\mathsf{T} \otimes \upsilon_\perp^\mathsf{T} U_I D_i U_I^\mathsf{T}\right) \\ &+ \left(u_i^\mathsf{T} \otimes \upsilon_\perp^\mathsf{T} U_I D_i \Sigma_I V_I^\mathsf{T}\right) K \\ &+ \left(s_i^{-1} v_i^\mathsf{T} \otimes \upsilon_\perp^\mathsf{T} U_J U_J^\mathsf{T}\right). \end{aligned}$$

9.4. **Proofs for perturbation expansions.** In this section, we derive the perturbation expansions of our test statistics explicitly and begin with

Proof of Lemma 3. We wish to construct the first and second derivatives of the map

$$\psi\left(M\left(t\right)\right)=v_{\perp}^{\intercal}R_{I}\left(t\right)L_{I}^{\intercal}\left(t\right)=v_{\perp}^{\intercal}P_{I}\left(t\right),$$

where we have made the dependence on t explicit and $M\left(0\right)=M$. Then based on Sun (1991, Remark 4.2), the Fréchet and Gateaux derivatives coincide and we can write

$$\dot{\psi}_{R}\left(\hat{M}_{n}-M\right)=v_{\perp}^{\intercal}\dot{P}_{I}\left(t\right)|_{t=0},$$

which by the product rule implies

$$v_{\perp}^{\intercal}\dot{P}\left(0\right)=v_{\perp}^{\intercal}\dot{R}_{I}\left(0\right)L_{I}^{\intercal}+v_{\perp}^{\intercal}R_{I}\dot{L}_{I}^{\intercal}\left(0\right)=v_{\perp}^{\intercal}\dot{R}_{I}\left(0\right)L_{I}^{\intercal}$$

because $v_{\perp}^{\intercal}R_{I}(0)=v_{\perp}^{\intercal}R_{I}=0$. □

Proof of Lemma 1. By Lemma 3,

$$\text{vec }\Delta_{\text{JI}}=\left(R_{I}^{\intercal}\otimes L_{J}^{\intercal}\right)\text{vec }\left(E\right), \tag{9.6}$$

and the derivative in (9.3), we take the first differential and vectorise the entire expression to obtain dvec $v_{\perp}^{\intercal}R_{J}L^{H}=B\,\text{dvec }\left(M\right)$. In particular vectorising $\boldsymbol{S}^{-1}\left(\hat{\Delta}_{n,\text{IJ}}\right)$ obtains from vectorising the forward problem $\boldsymbol{S}\left(X\right)=X\Lambda_{I}-\Lambda_{J}X=Y$ to obtain vec $Y=\left\{\left(\Lambda_{I}^{\intercal}\otimes I_{J}\right)-\left(I_{I}\otimes\Lambda_{J}\right)\right\}$ vec X so that vec $\boldsymbol{S}^{-1}\left(\hat{\Delta}_{n,\text{IJ}}\right)=\left\{\left(\Lambda_{I}^{\intercal}\otimes I_{J}\right)-\left(I_{I}\otimes\Lambda_{J}\right)\right\}$ vec $\hat{\Delta}_{n,\text{IJ}}$ which we insert into (9.6) to obtain

$$\text{vec }\boldsymbol{S}^{-1}\left(\Delta_{\text{JI}}\right)=\left\{\left(\Lambda_{I}^{\intercal}\otimes I_{J}\right)-\left(I_{I}\otimes\Lambda_{J}\right)\right\}^{-1}\left(R_{I}^{\intercal}\otimes L_{J}^{\intercal}\right)\text{vec }\left(M\right).$$

Combining the above display with (9.3), we obtain

$$\begin{aligned}&\text{vec }v_{\perp}^{\intercal}R_{I}\boldsymbol{S}^{-1}\left(\Delta_{\text{JI}}\right)L_{J}^{H}\\=&\left(L_{J}\otimes v_{\perp}^{\intercal}R_{I}\right)\left\{\left(\Lambda_{I}^{\intercal}\otimes I_{J}\right)-\left(I_{I}\otimes\Lambda_{J}\right)\right\}^{-1}\left(R_{I}^{\intercal}\otimes L_{J}^{\intercal}\right)\text{vec }\left(M\right).\end{aligned}$$

□

Proof of Lemma 5.

(1) By Lemma 3,

$$\text{vec }\hat{\Delta}_{n,\text{IJ}}=\left(R_{I}^{\intercal}\otimes L_{J}^{\intercal}\right)\text{vec }\left(\hat{M}_{n}-M\right).$$

The result follows by the continuous mapping theorem and Assumption 1.

(2) By the preceding argument, and the definition of $\boldsymbol{S}^{-1}$ in Lemma 3, we see that

$$v_{\perp}^{\intercal}R_{n,J}\boldsymbol{S}^{-1}\left(\hat{\Delta}_{n,\text{IJ}}\right)L_{n,I}^{\intercal}$$

is a linear and continuous transformation of $\hat{\Delta}_{n,\text{IJ}}$. Vectorizing $\boldsymbol{S}^{-1}\left(\hat{\Delta}_{n,\text{IJ}}\right)$ and taking the expectation of the outer product provides the result.

(3) To establish the rank of the covariance matrix, we shall take the constituent elements apart. For $\Omega_{W}=B_{W}\Omega B_{W}$ where we reproduce (5.2) for convenience as

$$B_{W}=\left(L_{I}\otimes v_{\perp}^{\intercal}R_{J}\right)\left\{\left(\Lambda_{I}^{\intercal}\otimes I_{J}\right)-\left(I_{I}\otimes\Lambda_{J}\right)\right\}^{-1}\left(R_{I}^{\intercal}\otimes L_{J}^{\intercal}\right).$$

First, we establish that $\operatorname{rk} \upsilon_{\perp}^{\intercal} R_J = m$. By Assumption 1.1, $R_J \in \mathbb{R}^{p \times r}$ has r linearly independent columns so that $\operatorname{rk} R_J = r$. By the properties of ranks as linearly independent columns we obtain the inequality $\operatorname{rk} \upsilon_{\perp}^{\intercal} R_J \leq \min(m, r) = m$ where the last inequality follows from $r \leq m$. By Sylvester's rank inequality, we obtain the lower bound $m + r - p \leq \operatorname{rk} \upsilon_{\perp}^{\intercal} R_J \leq m$, which implies $m - q \leq \operatorname{rk} \upsilon_{\perp}^{\intercal} R_J \leq m$. To see that the lower bound is slack and the upper bound is tight, we induct on q: put $q = 1$ and assume without loss of generality that $m = r$. Then, $r - 1 \leq \operatorname{rk} \upsilon_{\perp}^{\intercal} R_J \leq r$, so premultiplication of $\upsilon_{\perp}^{\intercal}$ causes the rank of R_J to drop by one, which implies that one column of $\upsilon_{\perp}^{\intercal}$ is perpendicular to one column of R_J. However, by Assumption 1.2, we have that no columns parallel to any in R_I must reside in R_J, which means that $\operatorname{rk} \upsilon_{\perp}^{\intercal} R_J = r$. Therefore, per the induction hypothesis the claim holds for $q = 1$. Now, we show that it holds for $q+1$. Suppose that $\operatorname{rk} \upsilon_{\perp}^{\intercal} R_J = r - q - 1$, which implies that we have again one column of $\upsilon_{\perp}^{\intercal}$ that is perpendicular to a column of R_J, leading to the same contradiction as before. Therefore, for any q and $q + 1$ we must have that the lower bound has to be maximally slack while the upper bound binds, thus establishing the claim for $m = r$. As there was nothing special about this choice of m, the claim also holds for all $m \leq r$. Next, we use the property of the Kronecker product $\operatorname{rk} L_I \otimes \upsilon_{\perp}^{\intercal} R_J = \operatorname{rk} L_I \operatorname{rk} \upsilon_{\perp}^{\intercal} R_J = qm$. The Jacobian $\left\{ \left(\Lambda_I^{\intercal} \otimes I_J \right) - \left(I_I \otimes \Lambda_J \right) \right\}^{-1}$ is a $qr \times qr$ square matrix of full rank. Then, $\operatorname{rk} \left(R_I^{\intercal} \otimes L_J^{\intercal} \right) = qr$ so that $\operatorname{rk} \left\{ \left(\Lambda_I^{\intercal} \otimes I_J \right) - \left(I_I \otimes \Lambda_J \right) \right\}^{-1} \left(R_I^{\intercal} \otimes L_J^{\intercal} \right) = qr$. Putting $A_1 := \left(R_I^{\intercal} \otimes L_J^{\intercal} \right)$ and $A_2 := \left\{ \left(\Lambda_I^{\intercal} \otimes I_J \right) - \left(I_I \otimes \Lambda_J \right) \right\}^{-1} \left(R_I^{\intercal} \otimes L_J^{\intercal} \right)$ we have $qm \pm qr \leq \operatorname{rk} AB \leq \min(qm, qr) = qm$ where the first relation follows from Sylvester's rank inequality and the last equality follows from $m \leq r$. Hence, $qm \leq \operatorname{rk} A_1 A_2 \leq qm$ or $\operatorname{rk} A_1 A_2 = \operatorname{rk} B_W = qm$. The same argument then establishes that $\operatorname{rk} B_W \Omega B_W^{\intercal} = qm$: in detail, $qm \pm p^2 \leq \operatorname{rk} B_W \Omega^{1/2} \leq \min(qm, p^2) = qm$. Then, $qm \pm p^2 \leq \operatorname{rk} B_W \Omega B_W^{\intercal} \leq \min(qm, p^2) = p^2$ and the claim follows.

□

Proof of Lemma 4.

(1) We Taylor-expand

$$
\begin{aligned}
\sqrt{n} \upsilon_{\perp}^{\intercal} \hat{P}_{n,I} &= \sqrt{n} \psi \left(\hat{M}_n \right) \\
&=_{(2)} \sqrt{n} \left(\psi \left(\hat{M}_n \right) - \psi(M) \right) \\
&= \sqrt{n} \dot{\psi} \left(\hat{M}_n - M \right) + \sqrt{n} r \left(\hat{M}_n - M \right) \\
&=_{(3)} \sqrt{n} \upsilon_{\perp}^{\intercal} R_{n,J} \boldsymbol{S}^{-1} \left(\hat{\Delta}_{n,\mathrm{IJ}} \right) L_{n,I}^{\intercal} + \sqrt{n} r \left(\hat{M}_n - M \right)
\end{aligned}
$$

where $=_{(2)}$ follows from application of the null hypothesis and $=_{(3)}$ follows from Lemma 3.

(2) We obtain a remainder term

$$
r \left(\hat{M} - M \right) := \psi \left(\hat{M} \right) - \psi(M) - \dot{\psi} \left(\hat{M} - M \right).
$$

To ensure that the first order term dominates the remainder, we need to verify that standardisation by $\sqrt{n}$ does not cause dominant second order terms. By Assumption 1, $\left\|\hat{M} - M\right\| = O_p\left(n^{-1/2}\right)$. By Taylor's theorem, $r\left(\hat{M} - M\right) = O\left(n^{-1/2}\right)$. For simplicity, we can treat $\left(\hat{M}_n - M\right)$ as a deterministic sequence and then apply van der Vaart (2000, Lemma 2.12, Ch. 2) to obtain the relevant statistical result. First, recognize that $\boldsymbol{S}^{-1}\left(a_n M\right) = a_n \boldsymbol{S}^{-1}\left(M\right)$ for any scalar sequence a_n and admissible argument M. Then, rewrite r as the sum of the next higher order term and another unspecified remainder term of known order via application of Lemma 3:

$$r\left(\hat{M}_n - M\right) = 2v_{\perp}^{\intercal} R_J \boldsymbol{S}^{-1}\left(\left(\hat{\Delta}_{n,J} V - V\hat{\Delta}_{n,I}\right)\right) L_I^{\intercal} + r_2\left(\hat{M}_n - M\right),$$

where r_2 collects higher order terms. By Lemma 4.3 and Lemma 3, we obtain the tighter bound $n^{1/2} r\left(\hat{M}_n - M\right) = O\left(n^{-1/2}\right)$, which implies that a normalization by $\sqrt{n}$ is innocuous and does not affect convergence.

(3) For a consistent and asymptotically normal least-squares estimator Assumption 1.1 and the definition in (9.2), the result follows by diagonalising $\hat{M}_n - M$ via $\hat{L}_I^H\left(\hat{M}_n - M\right)\hat{R}_I$

□

Proof of Lemma 6. Add and subtract $v_{\perp}^{\intercal}$ to obtain $\left(\left[\,\hat{D}_n^{\intercal}\; I_r\,\right] \pm \left[\,D^{\intercal}\; I_r\,\right]\right)\hat{R}_{n,I} = 0$. Therefore,

$$\left(\hat{D}_n^{\intercal} - D^{\intercal}\right)\hat{R}_{n,2,I} = \left[\hat{D}_n^{\intercal} \quad I_r\right]\left(\hat{R}_{n,I} - R_I R_{2,I}^{-1}\hat{R}_{n,2,I}\right),$$

whence the result follows. □

Proof of Corollary 3. Apply Corollary 4 to the expression in Lemma 6 and the result follows.

□

9.5. **Invariant subspace maps and consistency.** To complete the argument that renders the invariant subspaces of interest well-defined, we use the following result to argue that the map $\psi\left(M\left(t\right)\right) = v_{\perp}^{\intercal} P_I\left(M\left(t\right)\right)$ is smooth for a perturbation parameter $t \in \mathbb{C}$. Letting $t \in \mathbb{C}$ we perturb M and obtain the perturbed map $M\left(t\right) \coloneqq M + tC$. In the following, we shall make dependence on t explicit.

Lemma 8. *Let $\psi\left(t\right)$ be as in* (3.9)*. Then, both $\psi\left(t\right)$ and the perturbation expansion terms given in Lemma 3 are analytic functions of t in a neighbourhood around zero.*

Proof. To achieve analyticity, we observe that $\left|\left(\Lambda_I^{\intercal} \otimes I_J\right) - \left(I_I \otimes \Lambda_J\right)\right| \neq 0$ by Assumption 1.2 so that the condition in the second display at the top of Sun (1991, p. 90) is satisfied. For the case of real matrices, we appeal to Lang (1993, Thm. XIV.2.1). □

For the map induced by the normalised estimator $\hat{D}_n^{\intercal}$, we have in analogy to $\psi\left(M\right)$ defined in (3.9), the map $\psi_D : \mathscr{Q} \mapsto \mathbb{R}^{r \times q}$ with $\psi_D\left(M\right) \coloneqq v_{\perp}^{\intercal} R_I\left(M\right) R_{2,I}^{-1}$. Then, we have

Corollary 4. *The Fréchet derivative of ψ_D,*

$$\dot{\psi}_D\left(\hat{M}_n - M\right) = v_{\perp}^{\intercal} R_{n,J} \boldsymbol{S}^{-1}\left(\hat{\Delta}_{n,SL}\right) L_{n,I}^H R_{n,2,I}^{-1}.$$

Proof of Corollary 4. In Lemma 3, replace $L_I^\intercal$ by $R_{2,I}^{-1}$ and the result follows. □

Proof of Theorem 2. Theorem 2.1 follows from a law of large numbers for $\hat{M}_n = M + n^{-1}\sum_{i=1}^n \varepsilon_i$ and $\mathbb{E}\varepsilon_i = O$. Similarly, Theorem 2.2 follows from a standard argument: let $\hat{\varepsilon}_{i(j)}$ denote the jth column of the ith residual matrix $\hat{\varepsilon}_i \in \mathbb{R}^{p\times p}$. Then, $\hat{\Omega} = \frac{1}{np}\sum_{j=1}^p \sum_{i=1}^n \hat{\varepsilon}_{i(j)}\hat{\varepsilon}_{i(j)}^\intercal \xrightarrow{p} \Omega$ where we have exploited the homoskedasticity across columns of $\hat{M} - M$. Then, Lemma 8 implies that perturbation expansions of $\psi(M)$ in terms of basis vectors of right eigenvectors are equal to their map as defined implicitly by (3.7). Therefore, $\psi(M)$ is smooth in M and thus an immediate consequence of Lemma 8 is that $\hat{M}_n \xrightarrow{p} M$ implies that $\hat{R}_n \xrightarrow{p} R$ because the implicit map for R defined by $MR = R\Lambda$ is analytic. Theorem 2.4 follows from the continuous mapping theorem and $\hat{L}_n^H = \hat{R}_n^{-1}$ where we note that matrix inversion is continuous. For Theorem 2.5, $\hat{\Lambda}_n \xrightarrow{p} \Lambda$, we appeal to Tyrtyshnikov (1997, Thm. 3.9.1). Theorem 2.6-Theorem 2.7 follow from the continuous mapping theorem. Finally, the map underlying the estimator $\hat{D}_n$ defined in (4.5) inherits the analyticity property by Corollary 4. Observing that $\psi_D(M)$ is smooth in M, we can apply the continuous mapping theorem whence Theorem 2.8 follows. □

To prove Lemma 3, we find an expansion of basis vectors of invariant subspaces of a matrix in response to a small perturbation of that matrix. To aid with the general exposition of the idea behind invariant subspaces and associated perturbations, we harmonise notations of Sun (1991) with those employed here. Similar results to those considered here are available in Kato (1995, Ch. 2). Importantly, matrix perturbation theory seldomly considers real matrices. Lemma 8 ensures that our results apply to both the complex and real fields. Moreover, our test statistics remain practically useful even when $R, \Lambda \in \mathbb{C}^{p\times p}$, because they are always real by construction.

To delve deeper into the underlying perturbation theory,. let $X_1 \in \mathbb{R}^{p\times q}$ with $\operatorname{rk} X_1 = q$, $X_1^\intercal X_1 = I_q$ and $MX_1 = X_1 M_1$ for some matrix $M_1 \in \mathbb{R}^{q\times q}$. Then $\operatorname{sp} X_1$ is an invariant subspace of $M \in \mathbb{R}^{p\times p}$ if and only if there exists a non-singular matrix $X = \left[\begin{smallmatrix} X_1 & X_2 \end{smallmatrix}\right] \in \mathbb{R}^{p\times p}$ with $X_2^\intercal X_2 = I_{p-q}$ such that

$$X^{-1}MX = \begin{bmatrix} M_{11} & M_{12} \\ 0 & M_{22} \end{bmatrix},\ M_{11} \in \mathbb{R}^{q\times q}.$$

We now introduce the perturbation matrix of A,

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}.$$

The block C_{21} measures the departure from $M(t)$ being upper-triangular. For the purposes of this study, we are interested in the span of the vectors that are orthogonal to the columns of X_1. Recall the operator (3.10),

$$\boldsymbol{S}(Q) = QM_{11} - M_{22}Q$$

and $Q \in \mathbb{R}^{(p-q)\times q}$. Then, a necessary condition for Theorem 2.1 in Sun (1991) to apply is that A_{11} and A_{22} do not share any eigenvalues as stipulated in Assumption 1.2. Introducing the parameter $t \in \mathbb{C}$ to help us keep track of the order of the expansion. We state Sun (1991, Thm. 1):

Theorem 5. *There exists a unique* q*-dimensional invariant subspace* $\operatorname{sp} X_1(t)$ *of* $M(t)$ $(t \in \mathbb{C})$ *such that* $\operatorname{sp} X_1(0) = \operatorname{sp} X_1$ *and the basis vectors (columns of* X_1*) may be defined to be analytic functions of* t *in some neighbourhood of the origin of* $\mathbb{C}$*. Further, the analytic matrix-valued function* $X_1(t)$ *has the second order perturbation expansion*

$$
\begin{aligned}
X_1(t) = {} & X_1 \\
& + X_2 \boldsymbol{S}^{-1}(C_{21})\, t \\
& + X_2 \boldsymbol{S}^{-1}\left(C_{22}\boldsymbol{S}^{-1}(C_{21}) - \boldsymbol{S}^{-1}(C_{21})\, C_{11} - \boldsymbol{S}^{-1}(C_{21})\, M_{12}\boldsymbol{S}^{-1}(C_{21})\right) t^2 \\
& + O\left(t^3\right).
\end{aligned} \tag{9.7}
$$

We interpret Theorem 5 to make it amenable for statistical analysis. Expression (9.7) shows how small perturbations in the matrix A affect the column space of X_1. which correspond to the first and second-order derivatives. The equivalents of $X^{-1}MX$ and X_1 are Λ and R_I respectively. Furthermore, $M_{11} := \Lambda_I$ and $M_{22} := \Lambda_J$, so that by Assumption 1.2, M_{11} and M_{22} have no eigenvalues in common.

9.6. **Sufficiency for perturbation theory.** We mentioned that Assumption 1.2 was merely necessary for the perturbative arguments to apply. To furnish sufficiency we show that we can assume $R_I^H R_I = I_q$ without loss of generality. A corollary of this result is that the eigenvectors do not need to be normalised for the maps ψ_R and ψ_D to be smooth.

Importantly, the basis vectors spanning our invariant subspace of interest, $\operatorname{col} R_I$, are not in general biorthogonal, so a relation that left and right eigenvectors fulfill, $l_i^\top r_i = \delta_{ij}$ does not in general hold for right or left eigenvectors alone, i.e. $r_i^\top r_j = \delta_{ij}$ is true only for symmetric matrices when left and right eigenvectors coincide. Therefore, to operationalise the result Sun (1991, Thm 2.1), we need the invariant subspace, $\operatorname{col} R_I$, to come with orthogonal basis vectors whence we can recover the original eigenvectors. The following result ensures that the perturbation arguments apply:

Lemma 9. *The following are equivalent:*

1. There exists an orthonormal basis of vectors $s_1, \ldots, s_q$ for $\operatorname{col} R_I$ such that $\operatorname{col} R_I = \operatorname{sp}\{s_1, \ldots, s_q\}$.
2. There exists a bijection to recover right eigenvectors from the orthonormal basis vectors $\{s_1, \ldots, s_q\} \rightleftarrows \{r_1, \ldots, r_q\}$.

Proof. Employ the QR factorisation which exists for every real and complex matrix whereby $R_I = SU$ with $S^\top S = I_q$ and U is an upper triangular matrix. Because $\operatorname{rk} U = q$, $\operatorname{col} R_I = \operatorname{col} S$ so that the invariant subspace of interest that we wish to conduct inference on can always be identified. Finally, $MR_I = R_I \Lambda_I$ implies

$$
MS = S \underbrace{U \Lambda_I U^{-1}}_{\text{upper triangular}},
$$

so that $\operatorname{col} S$ is indeed an invariant subspace of M. To see why $U\Lambda_I U^{-1}$ is upper triangular, we need only show that U^{-1} is upper triangular whenever U is. Represent $U = D(I_q + K)$ where D is diagonal and K is strictly upper triangular. Then, $(I_q + K)^{-1} = I_q - K + K^2 - \cdots + (-1)^{q-1} K^{q-1}$ which has only strictly upper triangular summands. Moreover, K^q and higher powers are zero by the Cayley-Hamilton theorem. Then, the columns of S span the invariant subspace of M associated with roots in R_I. Finally, we have the relation for the columns of R_I, $r_i = \sum_{k=1}^{i} u_{ki} s_k$, where s_k are the columns of S and u_{ki} is the (k, i)th entry of U. This mapping is bijective because U is of full rank so that a reverse mapping can be achieved by swapping r and s and replacing u_{ki} with the entries of U^{-1}. □

References

Anderson, Theodore W. (1963). "Asymptotic theory for principal component analysis". In: *The Annals of Mathematical Statistics* 34.1, pp. 122–148.

– (2010). "The LIML estimator has finite moments!" In: *Journal of Econometrics* 157.2, pp. 359–361. ISSN: 0304-4076. DOI: `10.1016/j.jeconom.2010.02.004`. URL: `http://www.sciencedirect.com/science/article/pii/S0304407610000801`.

Bonacich, Phillip (1987). "Power and centrality: A family of measures". In: *American Journal of Sociology* 92.5, pp. 1170–1182.

Bura, E and R Pfeiffer (2008). "On the distribution of the left singular vectors of a random matrix and its applications". In: *Statistics & Probability Letters* 78.15, pp. 2275–2280.

Cai, Changxiao et al. (2019). "Subspace Estimation from Unbalanced and Incomplete Data Matrices: Statistical Guarantees". In: *arXiv preprint arXiv:1910.04267*.

Chen, Yuxin, Chen Cheng, and Jianqing Fan (2020). *Asymmetry Helps: Eigenvalue and Eigenvector Analyses of Asymmetrically Perturbed Low-Rank Matrices.* arXiv: `1811.12804 [math.ST]`.

Cheng, Chen, Yuting Wei, and Yuxin Chen (2021). *Tackling small eigen-gaps: Fine-grained eigenvector estimation and inference under heteroscedastic noise.* arXiv: `2001.04620 [math.ST]`.

Davis, AW (1977). "Asymptotic Theory for Principal Component Analysis: Non-normal Case". In: *Australian Journal of Statistics* 19.3, pp. 206–212.

Girko, V. L. (1999). "Strong Law for the singular values and eigenvectors of random matrices III. Inequalities for the spectral radius of large random matrices". In: *Random Operators and Stochastic Equations* 7.2, pp. 177–200. DOI: `https://doi.org/10.1515/rose.1999.7.2.177`. URL: `https://www.degruyter.com/view/journals/rose/7/2/article-p177.xml`.

Greenbaum, Anne, Ren-cang Li, and Michael L. Overton (Mar. 2, 2019). "First-order Perturbation Theory for Eigenvalues and Eigenvectors". Version 1. In: arXiv: `http://arxiv.org/abs/1903.00785v2 [math.NA, cs.NA, 47A55, 65F15]`. URL: `http://arxiv.org/abs/1903.00785v2`.

James, Alan T. (1964). "Distributions of Matrix Variates and Latent Roots Derived from Normal Samples". In: *The Annals of Mathematical Statistics*, pp. 475–501.

James, Alan T. (1977). "Tests for a prescribed subspace of principal components". In: *Multivariate Analysis IV, Ed. PR Krishnaiah, North-Holland: Amsterdam*, pp. 73–77.

Johnstone, Iain M, Alexei Onatski, et al. (2020). "Testing in high-dimensional spiked models". In: *Annals of Statistics* 48.3, pp. 1231–1254.

Johnstone, Iain M. (2001). "On the Distribution of the Largest Eigenvalue in Principal Components Analysis". In: *The Annals of Statistics* 29.2, pp. 295–327. ISSN: 00905364. URL: http://www.jstor.org/stable/2674106.

Kato, Tosio (1995). *Perturbation Theory for Linear Operators.* 2nd. Springer-Verlag Berlin and Heidelberg.

Koltchinskii, Vladimir, Matthias Löffler, Richard Nickl, et al. (2020). "Efficient estimation of linear functionals of principal components". In: *The Annals of Statistics* 48.1, pp. 464–490.

Lang, Serge (1993). *Real and Functional Analysis.* 3rd. New York (USA): Springer.

Lee, Ji Oon and Kevin Schnelli (Dec. 2016). "Tracy–Widom distribution for the largest eigenvalue of real sample covariance matrices with general population". In: *Ann. Appl. Probab.* 26.6, pp. 3786–3839. DOI: 10.1214/16-AAP1193. URL: https://doi.org/10.1214/16-AAP1193.

Liu, Jun, Xiangqian Liu, and Xiaoli Ma (2007). "First-Order Perturbation Analysis of Singular Vectors in Singular Value Decomposition". In: *2007 IEEE/SP 14th Workshop on Statistical Signal Processing*, pp. 532–536. DOI: 10.1109/SSP.2007.4301315.

Lütkepohl, Helmut and Maike M. Burda (1997). "Modified wald tests under nonregular conditions". In: *Journal of Econometrics* 78.1, pp. 315–332. ISSN: 0304-4076. DOI: https://doi.org/10.1016/S0304-4076(97)80015-2. URL: https://www.sciencedirect.com/science/article/pii/S0304407697800152.

Magnus, Jan R and Heinz Neudecker (2019). *Matrix differential calculus with applications in statistics and econometrics.* John Wiley & Sons.

Newey, Whitney K. and Kenneth D. West (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". In: *Econometrica* 55.3, pp. 703–708. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1913610 (visited on 02/06/2024).

Paula, Aureo de, Imran Rasul, and Pedro Souza (2023). *Identifying Network Ties from Panel Data: Theory and an Application to Tax Competition.* arXiv: 1910.07452 [econ.EM].

Penna, F., R. Garello, and M. A. Spirito (2009). "Cooperative spectrum sensing based on the limiting eigenvalue ratio distribution in wishart matrices". In: *IEEE Communications Letters* 13.7, pp. 507–509.

Silin, Igor and Jianqing Fan (2020). "Hypothesis testing for eigenspaces of covariance matrix". In: *arXiv preprint arXiv:2002.09810.*

Sun, Ji-guang (1991). "Perturbation expansions for invariant subspaces". In: *Linear Algebra And Its Applications* 153, pp. 85–97.

Takemura, Akimichi and Yo Sheena (2005). "Distribution of eigenvalues and eigenvectors of Wishart matrix when the population eigenvalues are infinitely dispersed and its application to minimax estimation of covariance matrix". In: *Journal of Multivariate Analysis* 94.2, pp. 271–299.

Tracy, Craig A. and Harold Widom (1994). "Level-spacing distributions and the Airy kernel". In: *Comm. Math. Phys.* 159.1, pp. 151–174. URL: https://projecteuclid.org:443/euclid.cmp/1104254495.

Tsagris, Michail, Christina Beneki, and Hossein Hassani (Feb. 2014). "On the Folded Normal Distribution". In: *Mathematics* 2.1, pp. 12–28. ISSN: 2227-7390. DOI: 10.3390/math2010012. URL: http://dx.doi.org/10.3390/math2010012.

Tyler, David E. (1981). "Asymptotic Inference for Eigenvectors". In: *The Annals of Statistics* 9, pp. 725–736. ISSN: 0090-5364. DOI: 10.1214/aos/1176345514.

– (1983). "A Class of Asymptotic Tests for Principal Component Vectors". In: *The Annals of Statistics* 11.4, pp. 1243–1250. DOI: 10.1214/aos/1176346337. URL: http://dx.doi.org/10.1214/aos/1176346337.

Tyrtyshnikov, Eugene E (1997). *A brief introduction to numerical analysis*. Springer Science & Business Media.

Vaart, Aad W. van der (2000). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Wigner, Eugene P (1958). "On the distribution of the roots of certain symmetric matrices". In: *Annals of Mathematics*, pp. 325–327.

Yuchen Xu, Marie-Christine Düker and David S. Matteson (2023). "Testing Simultaneous Diagonalizability". In: *Journal of the American Statistical Association* 0.0, pp. 1–13. DOI: 10.1080/01621459.2023.2202435. eprint: https://doi.org/10.1080/01621459.2023.2202435. URL: https://doi.org/10.1080/01621459.2023.2202435.

Zanella, Alberto, Marco Chiani, and Moe Z Win (2009). "On the marginal distribution of the eigenvalues of Wishart matrices". In: *IEEE Transactions on Communications* 57.4, pp. 1050–1060.