Biometrika (2019), **106**, 1, *pp.* 243–250
Printed in Great Britain
doi: 10.1093/biomet/asy070
Advance Access publication 7 January 2019

Miscellanea

Signal-plus-noise matrix models: eigenvector deviations and fluctuations

BY J. CAPE, M. TANG AND C. E. PRIEBE

Department of Applied Mathematics and Statistics, Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, U.S.A.

joshua.cape@jhu.edu minh@jhu.edu cep@jhu.edu

SUMMARY

Estimating eigenvectors and low-dimensional subspaces is of central importance for numerous problems in statistics, computer science and applied mathematics. In this paper we characterize the behaviour of perturbed eigenvectors for a range of signal-plus-noise matrix models encountered in statistical and random-matrix-theoretic settings. We establish both first-order approximation results, i.e., sharp deviations, and second-order distributional limit theory, i.e., fluctuations. The concise methodology presented in this paper synthesizes tools rooted in two core concepts, namely deterministic decompositions of matrix perturbations and probabilistic matrix concentration phenomena. We illustrate our theoretical results with simulation examples involving stochastic block model random graphs.

Some key words: Asymptotic normality; Eigenvector perturbation; Principal component analysis; Random matrix; Signal-plus-noise.

1. INTRODUCTION

In this paper we consider the setting where M and E are large $n \times n$ symmetric real-valued matrices, with $\hat{M} = M + E$ representing an additive perturbation of M by E. We investigate the relationship between the $n \times r$ matrices U and $\hat{U}$ whose columns are orthonormal eigenvectors corresponding to the $r \ll n$ leading eigenvalues of M and $\hat{M}$, respectively. Under quite general structural assumptions on U, M and E, our main results address this relationship both at the level of first-order deviations and at the level of second-order fluctuations. Theorems 1 and 2 quantify the entrywise closeness of $\hat{U}$ to U modulo a necessary orthogonal transformation W, which will subsequently be made precise. Theorem 3 states a multivariate distributional limit result for the rows of the matrix $\hat{U} - UW$.

Numerous problems in statistics deal with the eigenstructure of large symmetric matrices. Prominent examples of such problems include spike population and covariance matrix estimation (Silverstein, 1984, 1989; Johnstone, 2001; Yu et al., 2014), as well as principal component analysis (Jolliffe, 1986; Nadler, 2008; Paul, 2007), areas which have received additional attention as a result of advances in random-matrix theory (Bai & Silverstein, 2010; Benaych-Georges & Nadakuditi, 2011; Paul & Aue, 2014). Within the study of networks, the problem of community detection and the success of spectral clustering methods have also generated widespread interest in understanding spectral perturbations of large matrices, in particular graph Laplacian and adjacency matrices (Rohe et al., 2011; Lei & Rinaldo, 2015; Sarkar & Bickel, 2015; Le et al., 2017; Tang & Priebe, 2018). Towards these ends, recent ongoing and concurrent efforts in the statistics, computer science and mathematics communities have been devoted to obtaining precise entrywise bounds on eigenvector perturbations (Cape et al., 2019; Eldridge et al., 2018; Fan et al., 2018); see also Abbe et al. (2017), Mao et al. (2017) and Tang et al. (2017).

© 2019 Biometrika Trust

Downloaded from https://academic.oup.com/biomet/article/106/1/243/5280315 by University of Pittsburgh user on 20 October 2020

This paper distinguishes itself from the existing literature by presenting both deviation and fluctuation results within a concise yet flexible signal-plus-noise matrix model framework that is amenable to statistical applications. We extend the machinery and perturbation considerations introduced in Cape et al. (2019) to obtain strong first-order bounds. We then demonstrate how careful analysis within a unified framework leads to second-order multivariate distributional limit theory. Our characterization of eigenvector perturbations relies on a matrix perturbation series expansion together with an approximate commutativity argument for certain matrix products.

The results in this paper apply to principal component analysis in spike matrix models, including those of the form $Y = \lambda uu^{\mathrm{T}} + n^{-1/2}E$, where $u \in \mathbb{R}^n$ denotes a spike unit vector and $E \in \mathbb{R}^{n\times n}$ is a random symmetric matrix. We consider the supercritical regime, $\lambda > 1$, for which it is known, for example, that the leading eigenvector $\hat{u}$ of Y has nontrivial correlation with u when E is drawn from the Gaussian orthogonal ensemble, i.e., $|\langle \hat{u}, u\rangle|^2 \to 1 - 1/\lambda^2$ almost surely (Benaych-Georges & Nadakuditi, 2011). Here we obtain stronger local results for spike vector estimation in the presence of sufficient vector delocalization, provided the signal in $\lambda \gg 1$ is sufficiently informative with respect to E. Loosely speaking, we establish that $\|\hat{u}-u\|_\infty \leqslant C(\log n)^c \lambda^{-1}\|u\|_\infty$ with high probability for some positive constants C and c, and we prove that $n(\hat{u}_i - u_i)$ is asymptotically normally distributed. Our results hold more generally for r-dimensional spike models exhibiting eigenvalue multiplicity and for E exhibiting a heterogeneous variance profile.

2. Preliminaries

For $n \times r$ real matrices with orthonormal columns, denoted by $\hat{U}, U \in \mathbb{O}_{n,r}$, the columns of $\hat{U}$ and U each form orthonormal bases for r-dimensional subspaces of $\mathbb{R}^n$. The distance between subspaces is commonly defined via the notion of canonical angles and the cosine-sine matrix decomposition, which crucially involve the singular values of the matrix $U^{\mathrm{T}}\hat{U}$. Specifically, by writing the singular values of $U^{\mathrm{T}}\hat{U}$ as $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_r$, the $r \times r$ diagonal matrix of canonical angles is defined on the main diagonal as $\Theta(\hat{U}, U)_{ii} = \arccos(\sigma_i)$ for $i \in [r] = \{1, 2, \ldots, r\}$ (Bhatia, 1997, § 7.1).

One frequently encounters the entrywise-defined matrix $\sin\Theta(\hat{U}, U) \in \mathbb{R}^{r\times r}$, since for the commonly considered spectral and Frobenius matrix norms, small values of $\|\sin\Theta(\hat{U}, U)\|_\eta$ indicate small angular separation, or distance, between the subspaces corresponding to $\hat{U}$ and U. Importantly, the canonical angle notion of distance between r-dimensional subspaces takes into account basis alignment in the form of right-multiplication by an $r \times r$ orthogonal matrix $W \in \mathbb{O}_{r,r} \equiv \mathbb{O}_r$, and each choice of norm $\eta \in \{\cdot, \mathrm{F}\}$ satisfies (Cai & Zhang, 2018, Lemma 1)

$$\|\sin\Theta(\hat{U}, U)\|_\eta \leqslant \inf_{W\in\mathbb{O}_r} \|\hat{U} - UW\|_\eta \leqslant \surd 2\, \|\sin\Theta(\hat{U}, U)\|_\eta.$$

In this paper we focus on matrices of the form $\hat{U} - UW \in \mathbb{R}^{n\times r}$, but instead consider the two-to-infinity matrix norm, which is defined via the ℓ_2 and ℓ_∞ vector norms as $\|T\|_{2\to\infty} = \sup_{\|x\|=1}\|Tx\|_\infty$ for any matrix T. The quantity $\|T\|_{2\to\infty}$ has the convenient interpretation of being the maximum Euclidean norm of the rows of T and therefore affords the advantage of being invariant with respect to right-multiplication by orthogonal matrices. Our subsequent analysis will be shown to be particularly meaningful when U exhibits low/bounded coherence (Candès & Recht, 2009), i.e., when U is suitably delocalized (Rudelson & Vershynin, 2015) in the sense that $\|U\|_{2\to\infty}$ decays sufficiently quickly in n.

For tall, thin matrices $T \in \mathbb{R}^{n\times r}$ with $n \gg r$, such as $\hat{U} - UW$, standard norm relations reveal that $\|T\|_{\max} = \max_{i,j}|T_{ij}|$ and $\|T\|_{2\to\infty}$ differ by at most a factor depending on r. The same well-known relationship holds for the spectral and Frobenius norms, $\|T\| = \sigma_1(T)$ and $\|T\|_{\mathrm{F}} = (\sum_i \sigma_i^2(T))^{1/2}$, since necessarily $\mathrm{rank}(T) \leqslant r$. In contrast, $\|T\|_{2\to\infty}$ may in certain cases be much smaller than $\|T\|$ by a factor depending on n, summarized as

$$\|T\|_{\max} \overset{r}{\asymp} \|T\|_{2\to\infty} \overset{n}{\ll} \|T\| \overset{r}{\asymp} \|T\|_{\mathrm{F}}.$$

Taken together, these properties suggest the appropriateness of the two-to-infinity norm when viewing the rows of T as a point cloud of residuals in low-dimensional Euclidean space. We refer the reader to

Downloaded from https://academic.oup.com/biomet/article/106/1/243/5280315 by University of Pittsburgh user on 20 October 2020

Cape et al. (2019) for a more general discussion of the two-to-infinity norm and for further details on statistical applications, including community detection and principal subspace estimation, which are of interest here. In the current paper, additional model assumptions and more refined technical analysis yield stronger results for these applications.

3. Main results

3.1. *Setting*

Let $M \equiv M_n \in \mathbb{R}^{n\times n}$ be a symmetric matrix with block spectral decomposition

$$M \equiv [U|U_\perp][\Lambda \oplus \Lambda_\perp][U|U_\perp]^{\mathrm{T}} = U\Lambda U^{\mathrm{T}} + U_\perp \Lambda_\perp U_\perp^{\mathrm{T}},$$

where the diagonal matrix $\Lambda \in \mathbb{R}^{r\times r}$ contains the r largest-in-magnitude nonzero eigenvalues of M, with $|\Lambda_{11}| \geqslant \cdots \geqslant |\Lambda_{rr}| > 0$, and $U \in \mathbb{O}_{n,r}$ is an $n\times r$ matrix whose orthonormal columns are the corresponding eigenvectors of M. The diagonal matrix $\Lambda_\perp \in \mathbb{R}^{(n-r)\times(n-r)}$ contains the remaining $n-r$ eigenvalues of M, with the associated matrix of orthonormal eigenvectors $U_\perp \in \mathbb{O}_{n,(n-r)}$. Let $E \in \mathbb{R}^{n\times n}$ be a symmetric matrix, and write the perturbation of M by E as $\hat{M} \equiv M + E = \hat{U}\hat{\Lambda}\hat{U}^{\mathrm{T}} + \hat{U}_\perp\hat{\Lambda}_\perp\hat{U}_\perp^{\mathrm{T}}$.

Assumption 1. Let ρ_n denote a possibly n-dependent scaling parameter such that $(0,1] \ni \rho_n \to c_\rho \in [0,1]$ as $n\to\infty$, with $n\rho_n \geqslant c_1(\log n)^{c_2}$ for some constants $c_1, c_2 \geqslant 1$.

Assumption 2. There exist constants $C, c > 0$ such that for all $n \geqslant n_0(C,c)$, $|\Lambda_{rr}| \geqslant c(n\rho_n)$ and $|\Lambda_{11}||\Lambda_{rr}|^{-1} \leqslant C$, while $\Lambda_\perp \equiv 0$.

Assumption 3. There exist constants $C, c > 0$ such that $\|E\| \leqslant C(n\rho_n)^{1/2}$ with probability at least $1 - n^{-c}$ for all $n \geqslant n_0(C,c)$, written succinctly as $\|E\| = O_{\mathbb{P}}\{(n\rho_n)^{1/2}\}$.

Assumption 1 introduces a sparsity scaling factor ρ_n for added flexibility. This paper focuses on the large-n regime and so we often suppress the dependence of sequences of matrices on n for notational convenience.

Assumption 2 specifies the magnitude of the leading eigenvalues corresponding to the leading eigenvectors of interest. For simplicity and specificity, all leading eigenvalues are taken to be of the same prescribed order, and the remaining eigenvalues are assumed to vanish. Remark 2 briefly addresses the situation where the leading eigenvalues differ in order of magnitude, where $\Lambda_\perp \neq 0$, and when the spike dimension r is unknown.

Assumption 3 specifies that the random matrix E is concentrated in spectral norm in the classical probabilistic sense. Such concentration holds widely for random matrix models where E is centred, in which case $\hat{M}$ has low-rank expectation equal to M. The advantage of Assumption 3 when coupled with Assumption 2 is that, together with an application of Weyl's inequality (Bhatia, 1997, Corollary 3.2.6), the implicit signal-to-noise ratio terms behave as $\|E\||\Lambda_{rr}|^{-1}, \|E\||\hat{\Lambda}_{rr}|^{-1} = O_{\mathbb{P}}\{(n\rho_n)^{-1/2}\}$. It is straightforward to adapt our analysis and results under less explicit assumptions, albeit at the expense of succinctness and clarity.

Assumption 4 below specifies an additional probabilistic concentration requirement that arises in conjunction with the model flexibility introduced via the sparsity scaling factor ρ_n in Assumption 1. The notation $\lceil\cdot\rceil$ is used for the ceiling function.

Assumption 4. There exist constants $C_E, \nu > 0$ and $\xi > 1$ such that for all $1 \leqslant k \leqslant k(n) = \lceil \log n / \log(n\rho_n)\rceil$, for each standard basis vector e_i, and for each column vector u of U,

$$\left|\langle E^k u, e_i\rangle\right| \leqslant (C_E n\rho_n)^{k/2}(\log n)^{k\xi}\|u\|_\infty \tag{1}$$

with probability at least $1 - \exp\{-\nu(\log n)^{\xi}\}$, provided that $n \geqslant n_0(C_E, \nu, \xi)$.

Downloaded from https://academic.oup.com/biomet/article/106/1/243/5280315 by University of Pittsburgh user on 20 October 2020

Assumption 4 states a higher-order concentration estimate that reflects behaviour exhibited by a broad class of random symmetric matrices, including Wigner matrices whose entries exhibit subexponential decay and nonidentical variances (Erdős et al., 2013, modification of Lemma 7.10; Remark 2.4); see also Mao et al. (2017). For example, using our notation, the proof of Lemma 7.10 in Erdős et al. (2013) establishes that $|\langle (C_E n\rho_n)^{-k/2} E^k e, e_i\rangle| \leqslant (\log n)^{k\xi}$ with high probability, where e is the vector of all ones and the symmetric matrix E has independent zero-mean entries with bounded variances. Taking a union bound collectively over $1 \leqslant k \leqslant k(n)$, the standard basis vectors in $\mathbb{R}^n$ and the columns of U yields an event that holds with probability at least $1 - n^{-c}$ for some constant $c > 0$ and sufficiently large n.

The function $k(n)$ is fundamentally model-dependent through its connection with the sparsity factor ρ_n and satisfies $(n\rho_n)^{-k(n)/2} \leqslant n^{-1/2}$ for n sufficiently large. In the case where $\rho_n \equiv 1$, we have $k(n) \equiv 1$, and the behaviour reflected in (1) reduces to commonly encountered Bernstein-type probabilistic concentration. In contrast, if $\rho_n \to 0$ and, for example, $(n\rho_n) = n^{\epsilon}$ for some $\epsilon \in (0, 1)$, then $k(n) \equiv \epsilon^{-1}$. If instead $(n\rho_n) = (\log n)^{c_2}$ for some $c_2 \geqslant 1$, then $k(n) = \lceil \log n/(c_2 \log\log n)\rceil$. All regimes in which $\rho_n \to c_\rho > 0$ functionally correspond to the regime where $\rho_n \equiv 1$ by appropriate rescaling.

3.2. *First-order approximation*

Under Assumptions 2 and 3, spectral norm analysis via the Davis–Kahan sin Θ theorem (Bhatia, 1997, § 7.3) yields that for large n there exists $W \equiv W_n \in \mathbb{O}_r$ such that

$$\|\hat{U} - UW\| = O_{\mathbb{P}}\{(n\rho_n)^{-1/2}\}. \tag{2}$$

Equation (2) provides a coarse benchmark bound for the quantity $\|\hat{U} - UW\|_{2\to\infty}$, a quantity which is shown below to be much smaller at times.

Theorem 1. *Suppose that Assumptions 1–4 hold and that $n\rho_n = \omega\{(\log n)^{2\xi}\}$ with $r^{1/2} \leqslant (\log n)^{\xi}$. Then there exists $W \equiv W_n \in \mathbb{O}_r$ such that*

$$\|\hat{U} - UW\|_{2\to\infty} = O_{\mathbb{P}}\big[(n\rho_n)^{-1/2} \times \min\{r^{1/2}(\log n)^{\xi}\|U\|_{2\to\infty}, 1\}\big]. \tag{3}$$

The bound obtained by two-to-infinity norm methods in (3) is demonstrably superior to the bound implied by (2) when $r^{1/2}(\log n)^{\xi}\|U\|_{2\to\infty} \to 0$ as $n \to \infty$, i.e., when $\|U\|_{2\to\infty} \to 0$ sufficiently quickly. Such behaviour arises both in theory and in applications, including under the guise of eigenvector delocalization (Erdős et al., 2013; Rudelson & Vershynin, 2015) and of subspace basis coherence (Candès & Recht, 2009).

The proof of Theorem 1 first proceeds by way of refined deterministic matrix decompositions and then leverages the aforementioned probabilistic concentration assumptions. Our proof framework further permits second-order analysis, culminating in Theorem 3 in § 3.3. In the process of proving Theorem 3, we also prove Theorem 2, an extension and refinement of Theorem 1. Details of the proofs are provided in the Supplementary Material.

Theorem 2. *Suppose that Assumptions 1–4 hold and that (1) holds for k up to $k(n) + 1$. Suppose $n\rho_n = \omega\{(\log n)^{2\xi}\}$ and $r^{1/2} \leqslant (\log n)^{\xi}$. Then there exists $W \equiv W_n \in \mathbb{O}_r$ such that*

$$\hat{U} - UW = EU\Lambda^{-1}W + R$$

for some matrix $R \in \mathbb{R}^{n\times r}$ satisfying

$$\|R\|_{2\to\infty} = O_{\mathbb{P}}\big[(n\rho_n)^{-1} \times r \times \max\{(\log n)^{2\xi}, \|U^{\mathrm{T}}EU\| + 1\} \times \|U\|_{2\to\infty}\big].$$

Moreover,

$$\|EU\Lambda^{-1}W\|_{2\to\infty} = O_{\mathbb{P}}\{(n\rho_n)^{-1/2} \times r^{1/2}(\log n)^{\xi}\|U\|_{2\to\infty}\}.$$

Downloaded from https://academic.oup.com/biomet/article/106/1/243/5280315 by University of Pittsburgh user on 20 October 2020

Theorem 2 provides a collective eigenvector, i.e., subspace, characterization of the relationship between the leading eigenvectors of M and $\hat{M}$ via the perturbation E, summarized as

$$\hat{U} \approx \hat{M} U \Lambda^{-1} W = UW + EU\Lambda^{-1} W.$$

The unperturbed eigenvectors satisfy $UW \equiv MU\Lambda^{-1}W$, leading to the striking observation that the eigenvector perturbation characterization is approximately linear in the perturbation E.

Remark 1. It is always true that $\|U^{\mathrm{T}} EU\| \leqslant \|E\|$, where $\leqslant$ can be replaced by $\ll$ upon invoking Hoeffding-type concentration or, more generally, (C, c, γ)-concentration (O'Rourke et al., 2018) for suitable choices of E. Moreover, $\|R\|_{2\to\infty} \ll \|EU\Lambda^{-1}W\|_{2\to\infty}$ holds with high probability in Theorem 2 for numerous regimes in which $n\rho_n \to \infty$ and $\|U\|_{2\to\infty} \to 0$.

Remark 2. Strictly speaking, (2) holds even when the leading eigenvalues of M are not of the same order of magnitude, for the bound is fundamentally given by $C\|E\|(|\Lambda_{rr}| - \|\Lambda_{\perp}\|)^{-1}$. Similarly, the first-order bounds in this paper still hold for $\Lambda_{\perp} \neq 0$ provided $\|\Lambda_{\perp}\|$ is sufficiently small, in which case naïve analysis introduces additional terms of the form $\|\Lambda_{\perp}\| \|\Lambda^{-1}\| \| \sin \Theta(\hat{U}, U)\|$. In practice the exact spike dimension may be unknown, though it can often be consistently estimated via the elbow in the scree plot approach (Zhu & Ghodsi, 2006) provided that $\|E\|$ is sufficiently small relative to the leading nonzero eigenvalues of M.

3.3. *Second-order limit theory*

In this subsection additional structure is specified on M and E for the purpose of establishing second-order limit theory, i.e., fluctuations. Here, M is assumed to have strictly positive leading eigenvalues, reminiscent of a spike covariance or kernel population matrix setting. It is possible, though more involved, to obtain similar second-order results when M is allowed to have both strictly positive and strictly negative leading eigenvalues of the same order. Specifically, such modifications would give rise to considerations involving structured orthogonal matrices and the indefinite orthogonal group.

Assumption 5. The matrix M can be written as $M \equiv \rho_n XX^{\mathrm{T}} \equiv U\Lambda U^{\mathrm{T}}$ with $X = [X_1 | \dots | X_n]^{\mathrm{T}} \in \mathbb{R}^{n\times r}$ and $(n^{-1}X^{\mathrm{T}}X) \to \Xi \in \mathbb{R}^{r\times r}$ as $n \to \infty$ for some symmetric invertible matrix Ξ. Moreover, for a fixed index i, the scaled ith row of EX, written as $(n\rho_n)^{-1/2}(EX)_i = (n\rho_n)^{-1/2}(\sum_{j=1}^{n} E_{ij}X_j)$, converges in distribution to a centred multivariate normal random vector $Y_i \in \mathbb{R}^r$ with second moment matrix $\Gamma_i \in \mathbb{R}^{r\times r}$.

THEOREM 3. *Suppose that Assumptions* 1–5 *hold and that* (1) *holds for k up to $k(n) + 1$. Suppose in addition that $n\rho_n = \omega\{(\log n)^{2\xi}\}$, $r^{1/2} \leqslant (\log n)^{\xi}$, and*

$$\rho_n^{-1/2} \times r \times \max\left\{(\log n)^{2\xi}, \|U^{\mathrm{T}}EU\| + 1\right\} \times \|U\|_{2\to\infty} \to 0 \tag{4}$$

in probability as $n \to \infty$. Let $\hat{U}_i$ and U_i be column vectors corresponding to the ith rows of $\hat{U}$ and U, respectively. Then there exist sequences of orthogonal matrices (W) and (W_X) depending on n such that the random vector $n\rho_n^{1/2} W_X^{\mathrm{T}}(W\hat{U}_i - U_i)$ converges in distribution to a centred multivariate normal random vector with covariance matrix $\Sigma_i = \Xi^{-3/2}\Gamma_i\Xi^{-3/2}$, i.e.,

$$n\rho_n^{1/2} W_X^{\mathrm{T}}\left(W\hat{U}_i - U_i\right) \Rightarrow N_r(0, \Sigma_i).$$

Equation (4) amounts to a mild regularity condition that ensures $n\rho_n^{1/2}\|R\|_{2\to\infty} \to 0$ in probability for $R \equiv R_n \in \mathbb{R}^{n\times r}$ as in Theorem 2. This condition holds, for example, when $\|U\|_{2\to\infty} = O\{(\log n)^{c_3} n^{-1/2}\}$, in which case the left-hand side of (4) can often be shown to behave as $O_{\mathbb{P}}\{(\log n)^{c_4}(n\rho_n)^{-1/2}\}$ where $(\log n)^{c_4}(n\rho_n)^{-1/2} \to 0$ as $n \to \infty$. Such bounds on $\|U\|_{2\to\infty}$ provably arise when the ratio $(\max_i \|X_i\|)/(\min_i \|X_i\|)$ is at most polylogarithmic in n.

Downloaded from https://academic.oup.com/biomet/article/106/1/243/5280315 by University of Pittsburgh user on 20 October 2020

Remark 3 (Example: matrix M with kernel-type structure). Let F be a probability distribution defined on $\mathcal{X} \subseteq \mathbb{R}^r$, and let $X_1, \ldots, X_n \sim F$ be independent random vectors with invertible second moment matrix $\Xi \in \mathbb{R}^{r\times r}$. For $X = [X_1 | \ldots | X_n]^{\mathrm{T}} \in \mathbb{R}^{n\times r}$, let $M = \rho_n XX^{\mathrm{T}} \equiv U\Lambda U^{\mathrm{T}}$, so for each n there exists an $r \times r$ orthogonal matrix W_X such that $\rho_n^{1/2}X = U\Lambda^{1/2}W_X$. The strong law of large numbers guarantees that $(n^{-1}X^{\mathrm{T}}X) \to \Xi$ almost surely as $n \to \infty$, and so M has r eigenvalues of order $\Theta(n\rho_n)$ asymptotically almost surely. Moreover, $\|U\|_{2\to\infty} \leqslant Cn^{-1/2}\|X\|_{2\to\infty}$ asymptotically almost surely for some constant $C > 0$, where $\|X\|_{2\to\infty}$ can be suitably controlled by imposing additional assumptions, such as taking $\mathcal{X}$ to be bounded or imposing moment conditions on $\|X_1\|$. Conditioning on X yields a deterministic choice of M for the purposes of Assumption 5.

Remark 4 (Example: matrix E and multivariate normality). To continue the discussion from Remark 3, let all the entries of E be centred and independent up to symmetry with common variance $\sigma_E^2 > 0$. Then, by the classical multivariate central limit theorem, the asymptotic normality condition in Assumption 5 holds and $n\rho_n^{1/2}W_X^{\mathrm{T}}(W\hat{U}_i - U_i) \Rightarrow N_r(0, \sigma_E^2\Xi^{-2})$ by Theorem 3. There are a variety of other regimes in which the multivariate central limit theorem can be invoked for $(n\rho_n)^{-1/2}(\sum_{j=1}^n E_{ij}X_j)$ in order to satisfy the normality condition in Assumption 5, including cases where the entries of E have heterogeneous variances. In practice, we remark that Assumption 5 is probabilistically milder than Assumption 4 with respect to E.

3.4. *Simulations*

The K-block stochastic block model (Holland et al., 1983) is a simple yet ubiquitous random graph model in which vertices are assigned to one of K possible communities, or blocks, and where the adjacency of any two vertices is conditionally independent given the two vertices' community memberships. For stochastic block model graphs on n vertices, the binary symmetric adjacency matrix $A \in \{0, 1\}^{n\times n}$ can be viewed as an additive perturbation of a low-rank population edge probability matrix $P \in [0, 1]^{n\times n}$, $A = P + E$, where for K-block model graphs the matrix P corresponds to an appropriate dilation of the block edge probability matrix $B \in [0, 1]^{K\times K}$. In the language of this paper, $\hat{M} = A$ and $M = P$. It can be verified that versions of the aforementioned assumptions and hypotheses hold for the following examples. Here we set $\rho_n \equiv 1$.

Consider n-vertex graphs arising from the three-block stochastic block model with equal block sizes where the within-block and between-block Bernoulli edge probabilities are given by $B_{ii} = 0.5$ for $i = 1, 2, 3$ and $B_{ij} = 0.3$ for $i \neq j$, respectively. Here $\mathrm{rank}(M) = 3$, and the second-largest eigenvalue of M has multiplicity two. Figure 1(a) plots the empirical mean and 95% empirical confidence interval for $\|\hat{U} - UW\|_{2\to\infty}$ computed from 100 independent simulated adjacency matrices for each value of n and the function $\phi(n) = \{\lambda_3^{-1/2}(M)\}(\log n)n^{-1/2}$, which for large n captures the behaviour of the leading-order term in Theorem 2. This illustration does not pursue optimal constants or logarithmic factors. Here, using Θ notation, $\lambda_3(M) = \Theta(n\rho_n) = \Theta\{(n\rho_n)^{1/2}\lambda\}$ with respect to λ at the end of § 1.

Figure 1(b) shows a scatterplot of the uncentred, block-conditional scaled leading eigenvector components for an $n = 200$ vertex graph arising from a two-block model with 40% of the vertices belonging to the first block and where the block edge probability matrix B has entries $B_{11} = 0.5$, $B_{12} = B_{21} = 0.3$ and $B_{22} = 0.3$. This small-n example is complemented by additional simulation results provided in the Supplementary Material. We remark that the normalized random row vectors are jointly dependent but have decaying pairwise correlations; rows within any fixed finite collection are provably asymptotically independent as $n \to \infty$.

Acknowledgement

We are grateful to the editor, associate editor, and reviewers for their consideration of our paper and for their suggestions. This research was supported by the D3M programme of the U.S. Defense Advanced Research Projects Agency and by the Acheson J. Duncan Fund for the Advancement of Research in Statistics at Johns Hopkins University.

Downloaded from https://academic.oup.com/biomet/article/106/1/243/5280315 by University of Pittsburgh user on 20 October 2020

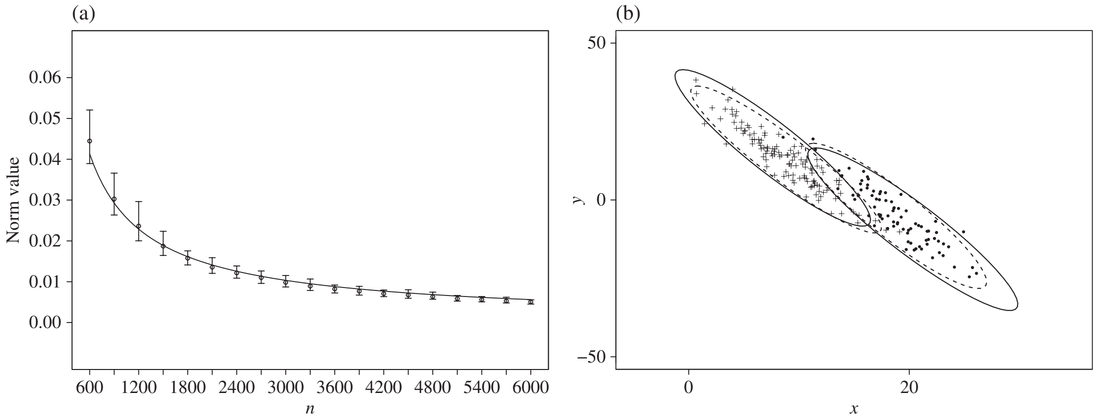

Fig. 1. (a) First-order simulations for the three-block model with number of vertices n on the x-axis and values of $\|\hat{U} - UW\|_{2\to\infty}$ on the y-axis; vertical bars depict 95% empirical confidence intervals, and the solid line reflects Theorem 2. (b) Second-order simulations for the two-block model with $n = 200$ where point shape reflects the block membership of the corresponding vertices; dashed ellipses give the 95% level curves for the empirical distributions, and solid ellipses represent the 95% level curves for the theoretical distributions according to Theorem 3.

Supplementary material

Supplementary material available at *Biometrika* online includes additional simulation examples as well as a joint proof of Theorems 1, 2 and 3.

References

Abbe, E., Fan, J., Wang, K. & Zhong, Y. (2017). Entrywise eigenvector analysis of random matrices with low expected rank. *arXiv:* 1709.09565.

Bai, Z. & Silverstein, J. W. (2010). *Spectral Analysis of Large Dimensional Random Matrices*. New York: Springer.

Benaych-Georges, F. & Nadakuditi, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227**, 494–521.

Bhatia, R. (1997). *Matrix Analysis*, vol. 169 of *Graduate Texts in Mathematics*. New York: Springer.

Cai, T. T. & Zhang, A. (2018). Rate-optimal perturbation bounds for singular subspaces with applications to high-dimensional statistics. *Ann. Statist.* **46**, 60–89.

Candès, E. J. & Recht, B. (2009). Exact matrix completion via convex optimization. *Foundat. Comp. Math.* **9**, 717.

Cape, J., Tang, M. & Priebe, C. E. (2019). The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *Ann. Statist.* to appear.

Eldridge, J., Belkin, M. & Wang, Y. (2018). Unperturbed: Spectral analysis beyond Davis-Kahan. In *Proceedings of Algorithmic Learning Theory*, vol. 83 of *Proceedings of Machine Learning Research*. PMLR, pp. 321–58.

Erdős, L., Knowles, A., Yau, H.-T. & Yin, J. (2013). Spectral statistics of Erdős–Rényi graphs I: Local semicircle law. *Ann. Prob.* **41**, 2279–375.

Fan, J., Wang, W. & Zhong, Y. (2018). An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *J. Mach. Learn. Res.* **18**, 1–42.

Holland, P. W., Laskey, K. B. & Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social Networks* **5**, 109–37.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29**, 295–327.

Jolliffe, I. T. (1986). *Principal Component Analysis*. New York: Springer.

Le, C. M., Levina, E. & Vershynin, R. (2017). Concentration and regularization of random graphs. *Random Struct. Algor.* **51**, 538–61.

Lei, J. & Rinaldo, A. (2015). Consistency of spectral clustering in stochastic block models. *Ann. Statist.* **43**, 215–37.

Mao, X., Sarkar, P. & Chakrabarti, D. (2017). Estimating mixed memberships with sharp eigenvector deviations. *arXiv:* 1709.00407.

Nadler, B. (2008). Finite sample approximation results for principal component analysis: A matrix perturbation approach. *Ann. Statist.* **36**, 2791–817.

Downloaded from https://academic.oup.com/biomet/article/106/1/243/5280315 by University of Pittsburgh user on 20 October 2020

O'ROURKE, S., VU, V. & WANG, K. (2018). Random perturbation of low rank matrices: Improving classical bounds. *Lin. Algeb. Applic.* **540**, 26–59.

PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17**, 1617–42.

PAUL, D. & AUE, A. (2014). Random matrix theory in statistics: A review. *J. Statist. Plan. Infer.* **150**, 1–29.

ROHE, K., CHATTERJEE, S. & YU, B. (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *Ann. Statist.* **39**, 1878–915.

RUDELSON, M. & VERSHYNIN, R. (2015). Delocalization of eigenvectors of random matrices with independent entries. *Duke Math. J.* **164**, 2507–38.

SARKAR, P. & BICKEL, P. J. (2015). Role of normalization in spectral clustering for stochastic blockmodels. *Ann. Statist.* **43**, 962–90.

SILVERSTEIN, J. W. (1984). Some limit theorems on the eigenvectors of large dimensional sample covariance matrices. *J. Mult. Anal.* **15**, 295–324.

SILVERSTEIN, J. W. (1989). On the eigenvectors of large dimensional sample covariance matrices. *J. Mult. Anal.* **30**, 1–16.

TANG, M., CAPE, J. & PRIEBE, C. E. (2017). Asymptotically efficient estimators for stochastic blockmodels: The naive MLE, the rank-constrained MLE, and the spectral. *arXiv:* 1710.10936.

TANG, M. & PRIEBE, C. E. (2018). Limit theorems for eigenvectors of the normalized Laplacian for random graphs. *Ann. Statist.* **46**, 2360–415.

YU, Y., WANG, T. & SAMWORTH, R. J. (2014). A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika* **102**, 315–23.

ZHU, M. & GHODSI, A. (2006). Automatic dimensionality selection from the scree plot via the use of profile likelihood. *Comp. Statist. Data Anal.* **51**, 918–30.

[*Received on* 14 *February* 2018. *Editorial decision on* 10 *August* 2018]

Downloaded from https://academic.oup.com/biomet/article/106/1/243/5280315 by University of Pittsburgh user on 20 October 2020