Unperturbed: spectral analysis beyond Davis-Kahan

Justin Eldridge ELDRIDGE@CSE.OHIO-STATE.EDU
Department of Computer Science and Engineering
The Ohio State University
Columbus, Ohio 43210, USA

Mikhail Belkin MBELKIN@CSE.OHIO-STATE.EDU
Department of Computer Science and Engineering
The Ohio State University
Columbus, Ohio 43210, USA

Yusu Wang YUSU@CSE.OHIO-STATE.EDU
Department of Computer Science and Engineering
The Ohio State University
Columbus, Ohio 43210, USA

Editor: Mehryar Mohri and Karthik Sridharan

Abstract

Classical matrix perturbation results, such as Weyl's theorem for eigenvalues and the Davis-Kahan theorem for eigenvectors, are general purpose. These classical bounds are tight in the worst case, but in many settings sub-optimal in the typical case. In this paper, we present perturbation bounds which consider the nature of the perturbation and its interaction with the unperturbed structure in order to obtain significant improvements over the classical theory in many scenarios, such as when the perturbation is random. We demonstrate the utility of these new results by analyzing perturbations in the stochastic blockmodel where we derive much tighter bounds than provided by the classical theory.

1. Introduction

In many applications the interesting structure of information is encoded by the eigenvalues and eigenvectors of an appropriately defined matrix. For instance, the top eigenvectors of the covariance matrix reveal the principal directions of the distribution, and the bottom eigenvalues and eigenvectors of a graph's Laplacian capture important details about its cluster structure. When learning from data, however, we typically do not have access to the matrix itself but rather a version which has been contaminated by (oftentimes random) noise. In such cases the following problem is of great interest: let M and H be $n \times n$ symmetric matrices with real entries. Suppose we "perturb" the matrix M by adding H. How do the eigenvalues and eigenvectors of $M + H$ relate to those of M?

For eigenvalues, the classical answer to this question comes in the form of Weyl's theorem (Weyl, 1912). Let the eigenvalues of M be $\lambda_1 \geq \cdots \geq \lambda_n$ and the eigenvalues of $M + H$ be $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_n$. Denote by $\|H\|$ the *spectral norm* of H; that is, the largest eigenvalue of H in absolute value. We have:

Theorem 1 (Weyl's theorem)**.** *For any* $i \in [n]$, $|\lambda_i - \tilde{\lambda}_i| \leq \|H\|$.

For the perturbation of eigenvectors, the classical result is the Davis-Kahan theorem (Davis and Kahan, 1969). For any fixed $t \in [n]$, let $u^{(t)}$ be an eigenvector of M with eigenvalue λ_t, and let $\tilde{u}^{(t)}$ be an eigenvector of $M + H$ with eigenvalue $\tilde{\lambda}_t$. Assume that λ_t and $\tilde{\lambda}_t$ have unit multiplicity; this assumption can be removed at the cost of complicating the statement of the result. The Davis-Kahan theorem bounds the angle θ_t between $u^{(t)}$ and $\tilde{u}^{(t)}$:

Theorem 2 (The Davis-Kahan theorem)**.** *Define* $\delta_t = \min\{|\tilde{\lambda}_j - \lambda_t| : j \neq t\}$. *Then* $\sin\theta_t \leq \|H\|/\delta_t$.

These classical results bound matrix perturbations in general cases, and do not use information about the structure of the matrices M and H or the relation between them. In applications, however, we often make assumptions about the nature of M and H; for example, we might assume that H is random noise added to a low rank M. In such instances the generality of Weyl's theorem and Davis-Kahan may result in severely suboptimal bounds.

In this work we present perturbation bounds which incorporate knowledge of the interaction between H and the eigenvectors of M. We obtain significant improvements over the classical theory in settings where this interaction is weak, such as when the perturbation H is random. In Section 3, we present an eigenvalue perturbation bound in the following spirit:

"Theorem" 1. *In many settings,* $|\tilde{\lambda}_t - \lambda_t|$ *is on the order of* $|\langle u^{(t)}, Hu^{(t)}\rangle| \ll \|H\|$.

We will show that when H is random the perturbation of the top eigenvalues is on the order of $\sqrt{\log n}$, whereas Weyl's theorem gives a bound on the order of $\sqrt{n}$. Next, in Section 4, we develop a theory of eigenvector perturbations in ∞-norm which is informally stated as follows:

"Theorem" 2. *In many settings,* $\|\tilde{u}^{(t)} - u^{(t)}\|_\infty$ *is on the order of* $\left\|\sum_{p\geq 1}^{\infty} (H/\lambda_t)^p\, u^{(t)}\right\|_\infty$.

We will show that when H is random and the top eigenvectors of M have small ∞-norm (which, for example, is the case when M has block-constant structure), the perturbation is also small. In many natural settings, our perturbation result improves upon the classical theory by a factor of $1/\sqrt{n}$.

Among the techniques used to derive the above results, we highlight the importance of what we call the *Neumann trick* – a particular expansion of the perturbed eigenvector which diminishes the effect of components whose interaction with H is hard to bound. To see the utility of the trick, consider bounding some norm of the perturbation $u^{(1)} - \tilde{u}^{(1)}$: Begin by writing $\tilde{u}^{(1)}$ as $\alpha u^{(1)} + \beta u^{\perp}$, where $u^{\perp}$ is some unit vector orthogonal to $u^{(1)}$. In the usual approach, controlling the norms of $Hu^{(1)}$ and $Hu^{\perp}$ are crucial in bounding the size of $u^{(1)} - \tilde{u}^{(1)}$. In the worst case these norms are bounded by $\|H\|$. It turns out that $\|Hu^{(1)}\|_2$ is often close to this worst-case bound in practice, but that $\|Hu^{(1)}\|_\infty$ can be much smaller than $\|H\|$, particularly when H is random. As a result, analyzing the interaction between H and $u^{(1)}$ often leads to an improved perturbation bound in ∞-norm.

However, while obtaining a tighter bound on $\|Hu^{(1)}\|_\infty$ is often possible, it can be difficult to derive an improved bound on $\|Hu^{\perp}\|_\infty$. Specifically, note that $u^{(1)}$ is a fixed vector independent of the perturbation H, but $u^{\perp}$ depends on H. If H is random, for instance, then $u^{\perp}$ is a random vector depending on H and the statistical interaction between H and $u^{\perp}$ can be hard to analyze. As a result, we often cannot bound the norm of $Hu^{\perp}$ any better than by the spectral norm of H. The Neumann trick allows us to replace the hard-to-analyze norm of $Hu^{\perp}$ with λ_2; if λ_2 is smaller than $\|H\|$ the Neumann trick presents significant advantages over the classical approach, as we will see.

We believe that the Neumann trick has the potential to substantially improve eigenvector perturbation bounds in many settings. As an example, we use it to analyze perturbations in the stochastic blockmodel and obtain much finer bounds than provided by the classical eigenvalue/eigenvector perturbation theory. As an easy corollary of our perturbation bounds, we obtain a straightforward proof that a simple and natural graph clustering algorithm indeed recovers the correct clustering of even very sparse graphs

1.1 Related work

Improving classical perturbation bounds has been the subject of recent interest. The work of Fan et al. (2016) bounds the ∞-norm perturbation of singular vectors under the assumption that M is low rank and incoherent. Our theory does not place either of these assumptions on M. Moreover, we will obtain improved bounds in some settings where (Fan et al., 2016) does not apply, such as in the stochastic blockmodel. Both Vu (2010) and O'Rourke et al. (2013) consider the case where H is random and M is low rank and present bounds in 2-norm which improve upon Davis-Kahan in certain settings. In contrast, our results are for the ∞-norm, we do not assume that M is low rank, and H needs not be random. Furthermore, in certain settings where M is low rank – such as in the case of the blockmodel – the results of the aforementioned papers do not necessarily improve upon the classical theory, while ours will. We note that the eigenvalue perturbation analysis in (O'Rourke et al., 2013) bears resemblance to that presented herein, but ours will hold for full-rank M and non-random H.

Jain and Netrapalli (2015) bound the max-norm error of the rank-k approximation of a matrix with the help of the Neumann trick. The scope of their paper is limited to matrix completion, and they do not provide eigenvector perturbation bounds. On the other hand, we develop an entrywise perturbation bound for the eigenvectors of general matrices (not necessarily low rank). In the special case of rank-k matrices, our strong control of the eigenvectors can be used to derive a max-norm error bound for matrix approximations similar to that provided by Jain and Netrapalli (2015) – for an example in the setting of a stochastic blockmodel, see Corollary 1.

Several works have studied the similar setting of random perturbations of spiked covariance matrices – see, for instance, the work of Benaych-Georges and Nadakuditi (2009) and Baik et al. (2004). These works derive the distribution of the eigenvalues of the perturbed matrix in the large matrix limit and study the phase transition threshold below which the extreme eigenvalues are not separated from the bulk eigenvalues. In contrast, the current paper bounds the effect of perturbations of a finite size.

Also related to the present work are the theories of random graphs and matrices. Perhaps most relevant is the work of Erdős et al. (2011), which analyzes the spectral statistics of Erdő-Rényi graphs using the Neumann trick. In contrast, we will develop the Neumann trick into a tool for analyzing general perturbations. Another related work is that of Mitra (2009), which bounds the ∞-norm perturbation of the top eigenvector of an Erdős-Rényi graph and provides a simple algorithm for clustering a sparse stochastic blockmodel with two communities. However, it is not clear how to generalize this method beyond the first eigenvector and therefore to blockmodels with $K \geq 2$ communities. Our method will give useful bounds on the top K eigenvectors, and our algorithm will work on models with an arbitrary (but constant) number of communities.

The stochastic blockmodel has been well-studied; for a summary, see the survey of Abbe (2017). A problem of particular interest is that of exact recovery of the latent communities in a sparse blockmodel. In this direction, information-theoretic limitations have been discovered and efficient algorithms developed (Abbe and Sandon, 2015). It is known, for example, that exact recovery is possible in the balanced 2-block model if the expected node degrees are super-logarithmic; when they are logarithmic, recovery is possible for some choices of constant factors but not for others. Recently, Vu (2014) analyzed an algorithm based on the SVD which recovers clusters exactly all the way down to the $\log n$ degree barrier. We will use our perturbation results to analyze a related algorithm which exactly recovers the communities of graphs with polylogarithmic degree. The main advantage of our method is its simplicity; while our algorithm does not improve on that of Vu (2014) in terms of performance, it is very natural and simple, and the guarantee of its correctness is the byproduct of our general perturbation results. It is also easy to generalize our method to blockmodels with a super-constant number of communities, or in which the block sizes scale at different rates.

1.2 Conventions and notation

We write $[n]$ to denote the set $\{1, \dots, n\}$. If $X^{(n)}$ is a sequence of random variables indexed by n, we say $X = O(f(n))$ with high probability (w.h.p.) if there exists a constant C such that

$\mathbb{P}(|X^{(n)}| \leq Cf(n)) \to 1$ as $n \to \infty$. We adopt the analogous definitions for the other asymptotic notations, such as $\Theta(f(n))$. We assume that eigenvectors have unit 2-norm.

2. Example: the stochastic blockmodel

Our perturbation results are sometimes rather technical when stated in their full generality. Therefore, to aid in the exposition, we adopt in this section a setting in which our main results are simpler to state. In particular, we assume the *stochastic blockmodel* – a popular random graph model with community structure. We will show that, here, our results provide much finer control over the perturbation of eigenvalues and eigenvectors as compared to Weyl's theorem and the theorem of Davis and Kahan. Additionally, we show that the proof of a simple and intuitive graph clustering algorithm is an easy corollary of our stronger perturbation bounds. We stress that this section serves as an example of how our theory is applied in a special case; its purpose is to provide the reader with a flavor of our more technical general results. The main results of the paper are stated in their full generality in Sections 3 and 4.

To begin, we formally define the K-block model:

Definition 1. An (n, K)-*stochastic blockmodel* is a pair (z, P), where $z : [n] \to [K]$ is a surjective map and P is a $K \times K$ symmetric matrix of rank K, with $P_{ij} \in [0, 1]$. We call z the *assignment* and P the *inter-community edge probability matrix*. The *edge probability matrix* M is the $n \times n$ symmetric matrix with entries $M_{ij} = P_{z(i),z(j)}$.

To generate a graph G from a blockmodel we sample to obtain its symmetric adjacency matrix $A = A_G$, where the upper triangular entries ($j \geq i$) are such that $A_{ij} \sim \text{Bernoulli}(M_{ij})$ and the lower triangular entries ($j < i$) are constrained to $A_{ij} = A_{ji}$. We view the random matrix A as a perturbation of M by the symmetric random matrix $H = A - M$, so that $A = M + H$. In what follows, let the eigenvectors and eigenvalues of M be $u^{(1)}, \ldots, u^{(n)}$ and $\lambda_1 \geq \cdots \geq \lambda_n$; similarly, let the eigenvectors and eigenvalues of A be $\tilde{u}^{(1)}, \ldots, \tilde{u}^{(n)}$ and $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_n$.

We will study sequences of blockmodels in which the expected node degree is permitted to grow sublinearly in the size the network; this is the *sparse* régime. For simplicity, we assume that every community has the same number of nodes, and that P is shared by all blockmodels in the sequence up to a density scaling factor of ρ. More precisely, we will adopt the following setting:

Setting 1 (ρ-sparse balanced blockmodel)**.** Let $K \in \mathbb{N}^+$ be a constant and fix a $K \times K$ inter-community edge probability matrix $P^{(0)}$. Assume for simplicity that each of the eigenvalues of $P^{(0)}$ is positive and unique. Let $\rho : \mathbb{N}^+ \to (0, 1]$ be such that each entry of $\rho(i) \cdot P^{(0)}$ is in the interval $[0, 1]$ for all $i \in \mathbb{N}^+$, and $\rho = \Omega(1/n)$. For any $m \in \mathbb{N}^+$, let $n = mK$ and define $P^{(m)} = \rho(n) \cdot P^{(0)}$. Consider a sequence of blockmodels $((z^{(m)}, P^{(m)}))_{m=1}^{\infty}$ in which $z^{(m)} : [n] \to [K]$ is an assignment of n nodes into K communities such that each is of size m.

ρ	$\Omega(1/n), O(1)$
$\mathbb{E}\lvert H_{ij}\rvert^k$	$O(\rho), k \geq 1$
$\lVert H \rVert$	$O(\sqrt{\rho n})$
λ_t	$\Theta(\rho n)$
$\lVert u^{(t)} \rVert_\infty$	$\Theta(1/\sqrt{n})$

Table 1: $t \in [K]$

The sequence of blockmodels has associated sequences of edge probability matrices $M^{(m)}$, random adjacency $A^{(m)}$ matrices, and so forth. For conciseness, we often omit the sequence index. We also remark that the assumptions on the eigenvalues of $P^{(0)}$ are made to simplify the exposition; the following results will hold in general with minor modification.

The asymptotic behaviors of the important quantities of Setting 1 are collected in Table 1 for $t \in [K]$. The fact that $\mathbb{E}|H_{ij}|^k = O(\rho)$ for all $k \geq 1$ follows from a simple calculation (see Lemma 2 in Appendix A.1). The bound on $\|H\|$ follows from a result in the theory of random matrices (see Theorem 14 in Appendix D.1). The nonzero eigenvalues of M are the eigenvalues of P scaled by ρn, and hence $\lambda_t = \Theta(\rho n)$ for any $t \in [K]$. Since the eigenvalues of P are

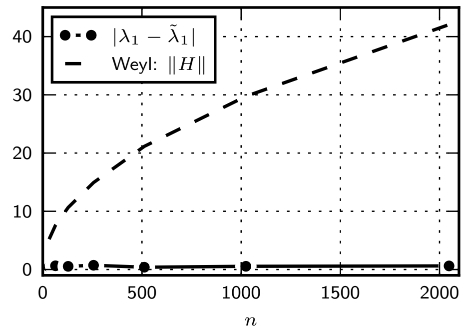

Figure 1: Empirical eigenvalue perturbations.

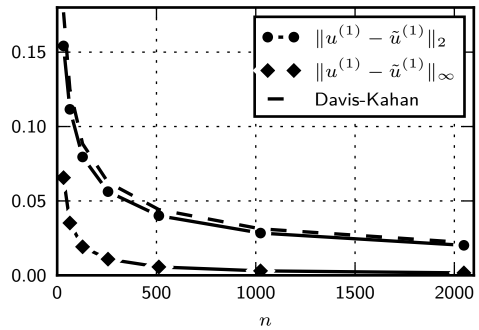

Figure 2: Empirical eigenvector perturbations.

assumed to be distinct in Setting 1, a gap of size $\Theta(\rho n)$ is ensured between the non-zero eigenvalues of M. It can also be shown that an eigenvector u of M which corresponds to a nonzero eigenvalue is constant on each block; i.e., $z(i) = z(j) \Rightarrow u_i = u_j$. Since each community has m members, it follows from the normalization constraint that $\|u^{(t)}\|_\infty = \Theta(1/\sqrt{m}) = \Theta(1/\sqrt{n})$.

	Classical	Ours (Thm. 3)
$\lvert\tilde{\lambda}_t - \lambda_t\rvert$	$O(\sqrt{\rho n})^\dagger$	$O(\sqrt{\log n})$
$\tilde{\lambda}_t$	$\Theta(\rho n)$	
$\sin\angle(\tilde{u}^{(t)}, u^{(t)})$	$O(1/\sqrt{\rho n})$	
$\|\tilde{u}^{(t)} - u^{(t)}\|_2$	$O(1/\sqrt{\rho n})$	
$\|\tilde{u}^{(t)} - u^{(t)}\|_\infty$	$O(1/\sqrt{\rho n})$	$O\left(\frac{\log^\xi n}{n\sqrt{\rho}}\right)$

Table 2: $t \in [K]$, $\quad \dagger : t \in [n]$

The predictions of the classical matrix perturbation theory as applied in this setting are collected in Table 2: Weyl's theorem bounds the eigenvalues and Davis-Kahan bounds the eigenvectors. To assess the quality of these bounds, the perturbation in the top eigenvalue and eigenvector of a sequence of growing blockmodels ($K = 1, \rho = 1, P = 1/2$) was measured; the results are shown in Figures 1 and 2. In the case of eigenvalues, we see that the actual perturbation is much smaller than Weyl's bound of $\|H\|$. For eigenvectors, the perturbation in 2-norm is close to the bound provided by the Davis-Kahan theorem, but the perturbation in ∞-norm is much smaller than predicted. Our general perturbation theory will explain both of these phenomena. In particular, our general results will imply the following in the current setting:

Theorem 3 (Special case: the stochastic blockmodel)**.** *Assume Setting 1; i.e., the ρ-sparse balanced stochastic blockmodel with $K \geq 1$. Suppose that $\rho = \Omega(n^{-1}\log^\epsilon n)$ for some $\epsilon > 2$. Let $1 < \xi < \epsilon/2$. Then there exist constants C_1, C_2 such that for any blockmodel in the sequence and all $t \in [K]$, with high probability as $n \to \infty$:*

$$|\lambda_t - \tilde{\lambda}_t| \leq C_1\sqrt{\log n} \qquad \textit{and} \qquad \|u^{(t)} - \tilde{u}^{(t)}\|_\infty \leq \frac{C_2(\log n)^\xi}{n\sqrt{\rho}}.$$

Our bounds are compared to their classical counterparts in Table 2. In particular, assume $\rho = \Theta(1)$: then the eigenvalue perturbation bound is improved from $O(\sqrt{n})$ to $O(\sqrt{\log n})$, and the eigenvector perturbation bound in ∞-norm is improved from $O(1/\sqrt{n})$ to $O(n^{-1}\log^\xi n)$; both bounds are substantially better. The proof of Theorem 3 will be given in two examples in later sections which serve to demonstrate how our more general perturbation results can be applied to specific settings. The eager reader can find the proof for eigenvalues in Section 3, Example 1 and the proof for eigenvectors in Section 4, Example 2.

2.1 A simple, consistent clustering algorithm

We will now show that the consistency of a simple graph clustering algorithm is a simple corollary of our Theorem 3. The fact that the eigenvectors of the blockmodel can be recovered to such precision motivates Algorithm 1. The algorithm first computes a rank-K approximation $\hat{M}$ of M using the top K eigenvectors of A ordered by the magnitude of their eigenvalues. It then clusters together all columns which are within a threshold τ in ∞-norm.

Algorithm 1 Blockmodel clustering

Require: Adjacency matrix A, $\tau \in \mathbb{R}^+$, $K \in \mathbb{N}^+$
  $\tilde{\lambda}_{s_1}, \ldots, \tilde{\lambda}_{s_K} \leftarrow$ top K eigvals of A by magnitude
  $\tilde{u}^{(s_1)}, \ldots, \tilde{u}^{(s_K)} \leftarrow$ corresponding eigvecs of A
  $\hat{M} \leftarrow \sum_{i=1}^{K} \tilde{\lambda}_{s_i} \tilde{u}^{(s_i)} \otimes \tilde{u}^{(s_i)}$
  $E \leftarrow \{(i,j) : \|\hat{M}_i - \hat{M}_j\|_\infty < \tau\}$
  $G \leftarrow$ graph with node set $[n]$, edge set E
  return connected components of G

Intuitively, the correctness of this algorithm relies on $\hat{M}$ being close to M entrywise. A sufficiently-tight bound on $\|\hat{M} - M\|_{\max}$ can indeed be obtained by applying a recent result from the low-rank matrix completion literature to the current setting; in particular, see Lemma 2 of Jain and Netrapalli (2015). In fact, roughly the same bound can be recovered as an easy corollary of our Theorem 3. In particular, we obtain the following result, whose proof is located in Appendix A.2:

Corollary 1. *Suppose that the assumptions of Theorem 3 hold. Define $\hat{M}$ as in Algorithm 1. Then $\|\hat{M} - M\|_{max} = O(\sqrt{\rho/n} \cdot \log^\xi n)$ with high probability.*

The consistency of the algorithm follows easily; the formal statement of the result and its proof are located in Appendix A.3.

It was remarked by Vu (2014) that Algorithm 1 is very natural, but difficult to analyze. Indeed, while the Davis-Kahan theorem provides a useful bound on $\|\hat{M}-M\|_F$, it implies only a trivial bound on $\|\hat{M} - M\|_{\max}$; for details, see Appendix A.4. In contrast, we are able to obtain a sufficiently-strong bound on $\|\hat{M} - M\|_{\max}$ by controlling the entrywise perturbation of eigenvectors much more tightly than what is implied by the Davis-Kahan theorem.

3. Eigenvalue perturbation

In this section we derive an eigenvalue perturbation bound that is stated in terms of the interaction between the perturbation matrix H and the eigenvectors of the base matrix M. We will see that in many cases, particularly when H is random, this bound is much tighter than Weyl's. The perturbation for eigenvectors is much more sophisticated to analyze, and will be given in Section 4.

To see how incorporating the interaction between H and the eigenvectors of M may lead to improved bounds, consider the following informal analysis of the perturbation in the first eigenvalue. As usual, let M and H be $n \times n$ and symmetric. The eigenvalues and eigenvectors of M are $\lambda_1 \geq \cdots \geq \lambda_n$ and $u^{(1)}, \ldots, u^{(n)}$, and the eigenvalues/vectors of $M + H$ are $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_n$ and $\tilde{u}^{(1)}, \ldots, \tilde{u}^{(n)}$. We have $\lambda_1 = \langle u^{(1)}, M u^{(1)} \rangle$ and $\tilde{\lambda}_1 = \langle \tilde{u}^{(1)}, (M + H)\, \tilde{u}^{(1)} \rangle$. Intuitively, if $\tilde{u}^{(1)}$ is close to $u^{(1)}$ then $\tilde{\lambda}_1 \approx \langle u^{(1)}, (M+H)\, u^{(1)} \rangle$; hence $\tilde{\lambda}_1 - \lambda_1 \approx \langle u^{(1)}, H u^{(1)} \rangle$. In the worst case $|\langle u^{(1)}, H u^{(1)} \rangle|$ can be as large as $\|H\|$ and we recover Weyl's bound. However, $|\langle u^{(1)}, H u^{(1)} \rangle|$ could be much smaller than $\|H\|$. For example, suppose that the entries of H are independent random variables with standard Gaussian distribution. Then $\langle u^{(1)}, H u^{(1)} \rangle$ is the sum of centered and independent random variables and therefore concentrates around zero. In this case the spectral norm of H is $O(\sqrt{n})$ while $|\langle u^{(1)}, H u^{(1)} \rangle|$ is much smaller at $O(\sqrt{\log n})$; this leads to an $O(\sqrt{\log n})$ bound on the eigenvalue perturbation instead of Weyl's bound of $O(\sqrt{n})$.

We now formalize this argument. We use the following well-known characterization of eigenvalues.

Theorem 4 (Courant-Fischer-Weyl min-max/max-min principles Horn and Johnson (2012))**.** *Let B be an $n \times n$ symmetric matrix with eigenvalues $\mu_1 \geq \ldots \geq \mu_t \geq \ldots \mu_n$. For any $d \in \{1, \ldots, n\}$,* *write $\mathcal{V}_d$ for the set of d-dimensional subspaces of $\mathbb{R}^n$. Then*

$$\mu_t \quad = \quad \min_{V \in \mathcal{V}_{n-t+1}} \max_{\substack{x \in V \\ \|x\|=1}} \langle x, Bx \rangle \quad = \quad \max_{V \in \mathcal{V}_t} \min_{\substack{x \in V \\ \|x\|=1}} \langle x, Bx \rangle \, .$$

We will use the max-min principle to get a lower bound on the perturbed eigenvalue and the min-max principle to obtain an upper bound. We prove the lower bound here to provide intuition:

Theorem 5 (Eigenvalue lower bound)**.** *Let $T \in [n]$ and h be such that $|\langle x, Hx \rangle| \leq h$ for all $x \in \operatorname{Span}(\{u^{(1)}, \ldots, u^{(T)}\})$. Then $\tilde{\lambda}_t \geq \lambda_t - h$ for all $t \leq T$.*

Proof. The max-min principle tells us that

$$\tilde{\lambda}_t = \max_{V \in \mathcal{V}_t} \min_{\substack{x \in V \\ \|x\|=1}} \langle x, (M+H)x \rangle \, .$$

Let $V^* = \operatorname{Span}(\{u^{(1)}, \ldots, u^{(T)}\})$. Then the above is lower-bounded by:

$$\min_{\substack{x \in V^* \\ \|x\|=1}} \langle x, (M+H)x \rangle \geq \min_{\substack{x \in V^* \\ \|x\|=1}} \langle x, Mx \rangle - \max_{\substack{x \in V^* \\ \|x\|=1}} \langle x, Hx \rangle \, .$$

The first term is minimized by taking $x = u^{(t)}$, such that $\langle x, Mx \rangle = \left\langle u^{(t)}, Mu^{(t)} \right\rangle = \lambda_t$. The magnitude of the second term is bounded by h. ☐

The proof of the following upper bound is more involved and is therefore located in Appendix B.1.

Theorem 6 (Eigenvalue upper bound)**.** *Let $T \in [n]$ and h be such that $|\langle x, Hx \rangle| \leq h$ for all $x \in \operatorname{Span}(\{u^{(1)}, \ldots, u^{(T)}\})$. Let $t \leq T$ and suppose that $\lambda_t - \lambda_{T+1} > 2\|H\| - h$. Then:*

$$\tilde{\lambda}_t \leq \lambda_t + h + \frac{\|H\|^2}{\lambda_t - \lambda_{T+1} + h - \|H\|}.$$

Similar lower and upper bounds can be obtained for eigenvalues at the bottom of the spectrum by negating M and H. For ease of reference, the statement of that result is located in Appendix B.2.

3.1 Interactions with random perturbations

Theorems 5 and 6 show that a tighter bound on eigenvalue perturbations can be obtained when $|\langle x, Hx \rangle| \ll \|H\|$ for any x in a subspace spanned by the top (or bottom) eigenvectors of M. We now show that this is often the case when H is random. The following is an application of the usual Hoeffding inequality; the proof is located in Appendix D.2.

Lemma 1. *Let u, v be any two fixed unit vectors in $\mathbb{R}^n$. Let H be an $n \times n$ symmetric random matrix with independent entries along the upper-triangle such that for all $j \geq i$, $\mathbb{E}H_{ij} = 0$ and H_{ij} is sub-Gaussian with parameter $\sigma_{ij} \leq \sigma$. Then $\mathbb{P}(|\langle u, Hv \rangle| \geq \gamma) \leq 2\exp\{-\gamma^2/(8\sigma^2)\}$.*

Lemma 1 applies generally to many types of random perturbation, including Gaussian noise and Bernoulli noise, as well as the random graph noise encountered in the stochastic blockmodel example in Section 2. We typically integrate the lemma with Theorems 5 and 6 in the following way: We first partition the spectrum into a top (large positive) and the remainder (small positive and negative) by choosing $T \in [n]$ such that $\lambda_T \gg \lambda_{T+1}$. We then apply Lemma 1 to argue that $|\langle u^{(i)}, Hu^{(j)} \rangle|$ is small ($\leq h$) with high probability for any indices $i, j \leq T$. It follows that $|\langle x, Hx \rangle| \leq Th$ for any unit vector x lying within the span of the top T eigenvectors of M; see Lemma 5 in Appendix D.3 for a proof. To bound the negative eigenvalues we negate M and H and repeat the above process.

Example 1: Proof of eigenvalue perturbation bound stated in Theorem 3. To demonstrate the application of our eigenvalue perturbation results, we will prove that in the blockmodel setting assumed in Theorem 3, $|\tilde{\lambda}_t - \lambda_t| \leq C\sqrt{\log n}$ for $t \in [K]$. We begin by applying Lemma 1. Since $\lambda_{K+1}, \ldots, \lambda_n$ are zero, we naturally choose $T = K$ such that $\lambda_T - \lambda_{T+1} = \lambda_T = \Theta(\rho n)$. Each entry along the diagonal and in the upper triangle of H is bounded and hence sub-Gaussian with a variance parameter upper-bounded by some constant σ. Choosing $\gamma = \sqrt{C \log n}$ in Lemma 1, we find that $|\langle u^{(i)}, Hu^{(j)} \rangle| \leq \sqrt{C \log n}$ for all $i, j \leq T$ w.h.p. Thus $|\langle x, Hx \rangle| \leq T\sqrt{C \log n} = O(\sqrt{\log n})$ for all $x \in \text{Span}(\{u^{(s)} : s \leq T\})$. We therefore bound h by $O(\sqrt{\log n})$ w.h.p. in Theorems 5 and 6.

In Theorem 3 it was assumed that $\rho = \Omega(n^{-1} \log^2 n) = \omega(n^{-1} \log n)$. It follows from this and the results in Table 1 that $\lambda_t + h - \|H\|$ is dominated by λ_t, and is therefore $\Theta(\rho n)$. Hence the second term in Theorem 6 is $O(\|H\|^2/\lambda_t) = O(1)$, and both the upper and lower bounds are dominated by $h = O(\sqrt{\log n})$. □

4. Eigenvector perturbation

We now study how a tigher bound on eigenvector perturbations might be achieved by analyzing the interaction between H and eigenvectors of M. Proofs of results in this section are rather technical and mostly in appendices. To build intuition, we make a series of simplifying assumptions; our formal theory will be much more general. First suppose that all eigenvalues of M are non-negative and that $\lambda_1 \gg \lambda_2$. By writing $\tilde{u}^{(1)}$ as $\alpha u^{(1)} + \beta u^{\perp}$ for some unit vector $u^{\perp}$ orthogonal to $u^{(1)}$ and using the definition of an eigenvector, we obtain: $\tilde{u}^{(1)} = \tilde{\lambda}_1^{-1} (M + H)\, \tilde{u}^{(1)} = \tilde{\lambda}_1^{-1} \left(\alpha \lambda_1 u^{(1)} + \beta M u^{\perp} + \alpha H u^{(1)} + \beta H u^{\perp} \right)$. Note that $\|Mu^{\perp}\|_2 \leq \|Mu^{(2)}\|_2 = \lambda_2 \ll \lambda_1$. If λ_2 is sufficiently small, the contribution of $\beta M u^{\perp}$ to $\tilde{u}^{(1)}$ is negligible. Assume that this is so, that $\tilde{\lambda}_1 \approx \lambda_1$, and that $\alpha \approx 1$ such that $\beta \ll 1$. Then $u^{(1)} - \tilde{u}^{(1)} \approx \lambda_1^{-1}(Hu^{(1)} + \beta H u^{\perp})$. Therefore we see that to bound the norm of the perturbation it suffices to control the norms of $Hu^{(1)}$ and $Hu^{\perp}$.

The classical approach is to bound these quantities by the spectral norm of H. For instance, to derive a bound in 2-norm we observe that $\|Hu^{(1)}\|_2 \leq \|H\|$ and that $\|\beta H u^{\perp}\|_2 \leq \|\beta H\|$, and therefore $\|\tilde{u}^{(1)} - u^{(1)}\|_2 \lesssim \lambda_1^{-1}\|H\|$. Furthermore, since the 2-norm upper-bounds the ∞-norm, we get a bound of $\|\tilde{u}^{(1)} - u^{(1)}\|_\infty \lesssim \lambda_1^{-1}\|H\|$ "for free". However, the spectral norm does not utilize information about the interaction between H and M. Our hope is that by analyzing this interaction, tighter bounds on the norms of $Hu^{(1)}$ and $Hu^{\perp}$ might be obtained.

In particular, consider a random, centered H and $u^{(1)}$ (which is independent of H). Unfortunately, $\|Hu^{(1)}\|_2$ is typically on the same order as $\|H\|$ and analyzing the interaction does not improve the bound. On the other hand, $\|Hu^{(1)}\|_\infty$ is often much smaller than $\|H\|$ and analyzing the interaction leads to much tighter bounds. To see why, note that $\|Hu^{(1)}\|_2^2 = \sum_{i=1}^n (\sum_{j=1}^n H_{ij} u_j^{(1)})^2$. As the summand of the outer sum is squared and thus non-negative, it does not concentrate around zero. In contrast, the sum in $|[Hu^{(1)}]_i| = |\sum_{j=1}^n H_{ij} u_j^{(1)}|$ *does* concentrate around zero, and is often much less than the worst-case bound of $\|H\|$. For example, if H is the random Gaussian matrix described above then $[Hu^{(1)}]_i$ is on the order of one, and a union bound over the n entries results in a high-probability bound of $\|Hu^{(1)}\|_\infty \leq \sqrt{\log n}$. On the other hand, $\|H\| = O(\sqrt{n})$.

In this case and in others, $\|Hu^{(1)}\|_\infty$ can be bounded to be much smaller than $\|H\|$. Can a similar analysis be used to show that $\|Hu^{\perp}\|_\infty$ is much smaller than $\|H\|$? It turns out that this is difficult for a subtle reason: while $u^{(1)}$ is fixed, $u^{\perp}$ depends on the perturbation. When H is random, $u^{\perp}$ is also random and statistically dependent on H. As such, the interaction between H and $u^{\perp}$ is often difficult to analyze, and we must resort to using the worst-case bound of $\|Hu^{\perp}\|_\infty \leq \|H\|$, giving:

$$\|u^{(1)} - \tilde{u}^{(1)}\|_\infty \lesssim \tilde{\lambda}^{-1} \left(\|Hu^{(1)}\|_\infty + \|\beta H\| \right). \tag{1}$$

In many cases $\|\beta H\|$ is small enough that it is dominated by our bound on $\|Hu^{(1)}\|_\infty$ and we have $\|u^{(1)} - \tilde{u}^{(1)}\|_\infty \lesssim \tilde{\lambda}^{-1}\|Hu^{(1)}\|_\infty$. For example, it can be shown that in the sparse stochastic blockmodel described in Setting 1, $\|Hu^{(1)}\|_\infty = O(\sqrt{\rho \log n})$ w.h.p., while $\|\beta H\| = O(1)$. Therefore, if $\rho = \Omega(1/\log n)$, the bound on $\|Hu^{(1)}\|_\infty$ dominates and we have $\|\tilde{u}^{(1)} - u^{(1)}\|_\infty = O(\rho^{-1/2}\, n^{-1} \sqrt{\log n})$. Comparing this to the trivial bound of $O(1/\sqrt{\rho n})$ implied by Davis-Kahan, we see that analyzing the interaction leads to a $\tilde{O}(1/\sqrt{n})$ improvement over the classical theory.

4.1 The Neumann trick

There are important settings, however, in which using the spectral norm to bound $Hu^\perp$ is sub-optimal; for instance, in the blockmodel described above when $\rho = o(1/\log n)$ In this sparser régime, $\|\beta H\| = O(1)$ dominates our bound on $\|Hu^{(1)}\|_\infty$ and we find that $\|\tilde{u}^{(1)} - u^{(1)}\|_\infty = O(1/\tilde{\lambda}_1) = O(1/\rho n)$, which is not tight. In general, if $\|Hu^{(1)}\|_\infty$ can be bounded to be much smaller than $\|\beta H\|$, the latter term dominates Equation (1). Therefore, while the simple approach described in the previous section improves upon the classical bound, the presence of the hard-to-control $Hu^\perp$ limits its effectiveness.

It turns out that we can often obtain a better bound by applying what we call the *Neumann trick*, which we now describe for $\tilde{u}^{(1)}$. From the definition of an eigenvector, we have $(M+H)\tilde{u}^{(1)} = \tilde{\lambda}_1 \tilde{u}^{(1)}$, which implies $(\tilde{\lambda}_1 - H)\tilde{u}^{(1)} = M\tilde{u}^{(1)}$. If $\tilde{\lambda}_1$ is not an eigenvalue of H we may invert $(\tilde{\lambda}_1 - H)$ to obtain $\tilde{u}^{(1)} = \tilde{\lambda}_1^{-1}(I - H/\tilde{\lambda}_1)^{-1} M\tilde{u}^{(1)}$. Expanding the inverse in a Neumann series and decomposing $\tilde{u}^{(1)}$ as above, we find: $\tilde{u}^{(1)} = \tilde{\lambda}_1^{-1} \sum_{p \geq 0} (H/\tilde{\lambda}_1)^p \left[\alpha \lambda_1 u^{(1)} + \beta M u^\perp\right]$. Assuming that $\alpha \approx 1$ and $\lambda_1 \approx \tilde{\lambda}_1$, we have:

$$u^{(1)} - \tilde{u}^{(1)} \approx \left[\sum_{p \geq 1} (H/\tilde{\lambda}_1)^p\, u^{(1)}\right] + \left[\frac{\beta}{\tilde{\lambda}_1} \sum_{p \geq 0} (H/\tilde{\lambda}_1)^p\, M u^\perp\right]. \tag{2}$$

If the series involving $u^\perp$ converges, it is dominated by its first term: $Mu^\perp$. Since $u^\perp$ lies in the subspace orthogonal to $u^{(1)}$, $\|Mu^\perp\|_2$ is upper-bounded by λ_2, and hence so is $\|Mu^\perp\|_\infty$. Hence:

$$\|u^{(1)} - \tilde{u}^{(1)}\|_\infty \lesssim \left\| \sum_{p \geq 1} (H/\tilde{\lambda}_1)^p\, u^{(1)} \right\|_\infty + \frac{|\beta| \lambda_2}{\tilde{\lambda}_1}. \tag{3}$$

Thus the contribution of $u^\perp$ is bounded here by $\tilde{\lambda}_1^{-1}|\beta|\lambda_2$. Comparing this to the previous result of Equation (1) in which the contribution of $u^\perp$ was bounded by $\tilde{\lambda}_1^{-1}|\beta| \cdot \|H\|$, we see that the Neumann trick permits us to replace $\|H\|$ with the top eigenvalue corresponding to the subspace orthogonal to $u^{(1)}$. The tradeoff is that we must now analyze the interaction between all powers of H and $u^{(1)}$ in order to bound the first term in Equation (3).

The Neumann trick allows us to tighten the eigenvector perturbation bound in the sparse stochastic blockmodel discussed above. We have seen that the first approach of Equation (1) leads to a bound of $\|\tilde{u}^{(1)} - u^{(1)}\|_\infty = O(1/\rho n)$ when $\rho = o(1/\log n)$. Now if we use Neumann trick, we can show that the norm of the series in Equation (3) is $O(\log^\xi n / (\sqrt{\rho} n))$, where $\xi > 1$. Assume the blockmodel has only one block (for multiple blocks we will use the more general results in Theorem 7). Then $\lambda_2 = 0$ and the second term in Equation (3) disappears. We thus have $\|\tilde{u}^{(1)} - u^{(1)}\|_\infty = O(\log^\xi n/(\sqrt{\rho} n))$, which significantly outperforms $O(1/\rho n)$ in this sparse régime (where $\rho = o(1/\log n)$).

We now formally state the general *Neumann trick*. See Appendix C.1 for the proof.

Theorem 7 (Neumann trick)**.** *Fix a $t \in [n]$. Suppose that $\|H\| < |\tilde{\lambda}_t|$. Then:*

$$\tilde{u}^{(t)} = \sum_{s=1}^{n} \frac{\lambda_s}{\tilde{\lambda}_t} \cdot \left\langle \tilde{u}^{(t)}, u^{(s)} \right\rangle \sum_{p \geq 0} \left(\frac{H}{\tilde{\lambda}_t}\right)^p u^{(s)}.$$

Observe that the contribution of $u^{(s)}$ is filtered by its eigenvalue, λ_s. In the special case when M is rank-K, $\tilde{u}^{(t)}$ is expressed totally in terms of $u^{(1)}, \ldots, u^{(K)}$. The Neumann trick can be used in combination with Weyl's theorem and the Davis-Kahan theorem to obtain a tighter bound on the elementwise perturbation of eigenvectors.

The following theorem states the result in its full generality, where M may be full-rank with non-distinct eigenvalues. Its proof in Appendix C.4 is a corollary of Theorem 12 in Appendix C.2. Let u_α denote the α-th entry of vector u.

Theorem 8. *For any $s \in [n]$, let $\Lambda_s = \{i : \lambda_i = \lambda_s\}$. Define $d_s = |\Lambda_s|$, and let the gap be defined as $\delta_s = \min_{i \notin \Lambda_s} |\lambda_s - \lambda_i|$. For any $s, t \in [n]$, let $\Delta_{s,t}^{-1} = \min\{d_i/\delta_i\}_{i\in\{s,t\}}$. Define $\lambda_s^* = |\lambda_s| - \|H\|$. There exists an orthonormal set of eigenvectors $u^{(1)}, \ldots, u^{(n)}$ satisfying $Mu^{(s)} = \lambda_s u^{(s)}$ such that for all $t \in [n]$:*

$$\left|\tilde{u}_\alpha^{(t)} - u_\alpha^{(t)}\right| \leq \left|u_\alpha^{(t)}\right| \cdot \left(8d_t \left[\frac{\|H\|}{\delta_t}\right]^2 + \frac{\|H\|}{\lambda_t^*}\right) + \left(\frac{|\lambda_t|}{\lambda_t^*}\right)^2 \cdot \zeta_\alpha(u^{(t)}; H, \lambda_t) + \frac{2\sqrt{2}\cdot\|H\|}{\lambda_t^*} \sum_{s\neq t} \frac{|\lambda_s|}{\Delta_{s,t}} \left[|u_\alpha^{(s)}| + \frac{|\lambda_t|}{\lambda_t^*} \cdot \zeta_\alpha(u^{(s)}; H, \lambda_t)\right], \tag{4}$$

where $\zeta(u; H, \lambda)$ is the n-vector whose αth entry is defined to be $\zeta_\alpha(u; H, \lambda) = \left|\left[\sum_{p\geq 1} \left(\frac{H}{\lambda}\right)^p u\right]_\alpha\right|$.

4.2 Interactions with random perturbations

The interaction between the eigenvectors of M and the perturbation H appears in Theorem 8 through ζ; in many applications ζ will dominate the bound. It turns out that when H is random and the eigenvectors of M have small ∞-norm, ζ is also small. The following result makes this precise. See Appendix E.3 for the proof.

Theorem 9. *Let H be an $n \times n$ symmetric random matrix with independent entries along the diagonal and upper triangle satisfying $\mathbb{E}H_{ij} = 0$. Suppose γ is such that $\mathbb{E}|H_{ij}/\gamma|^p \leq 1/n$ for all $p \geq 2$. Choose $\xi > 1$ and $\kappa \in (0,1)$. Let $\lambda \in \mathbb{R}$ and suppose that $\gamma < \lambda(\log n)^\xi$ and $\lambda > \|H\|$. Fix $u \in \mathbb{R}^n$. Then: with probability $1 - n^{-\frac{1}{4}(\log_b n)^{\xi-1}(\log_b e)^{-\xi}+1}$, where $b = \left(\frac{\kappa+1}{2}\right)^{-1}$.*

$$\left\|\sum_{p\geq 1} \left(\frac{H}{\lambda}\right)^p u\right\|_\infty \leq \frac{\gamma(\log n)^\xi}{\lambda - \gamma(\log n)^\xi} \cdot \|u\|_\infty + \frac{\|H/\lambda\|^{\lfloor \frac{\kappa}{8}(\log n)^\xi + 1\rfloor}}{1 - \|H/\lambda\|} \cdot \|u\|_2. \tag{5}$$

As an example, we apply Theorem 9 to bound the perturbation of eigenvectors in the stochastic blockmodel.

Example 2: Proof of eigenvector perturbation bound stated in Theorem 3. Consider again the setting of Theorem 3. We will use Theorems 8 and 9 to derive the bound of $\|u^{(t)} - \tilde{u}^{(t)}\|_\infty = O(\rho^{-1/2}\, n^{-1} \log^\xi n)$ w.h.p. for all $t \in [K]$.

First note that all but $K-1$ terms of the sum in Equation (4) vanish due to λ_s being zero; only the terms corresponding to $s \in [K]$ remain. Fix a $t \in [n]$. Referring to Table 1, we find that: $\|H\| = O(\sqrt{\rho n})$, $\lambda_t^* = \Theta(\rho n)$, $\delta_t = \Theta(\rho n)$, and $\|u^{(s)}\|_\infty = \Theta(1/\sqrt{n})$ for any $s \in [K]$. Substituting these bounds into Equation (4) and assuming that Z is an upper bound for $\|\zeta(u^{(s)}; H, \lambda_t)\|_\infty$ for all $s \in [K]$, we see that the first term in Theorem 8 is $O(n^{-1}\rho^{-1/2})$, the second term is $O(Z)$ and the third term is $O(n^{-1}\rho^{-1/2} + (\rho n)^{-1/2} Z)$. Therefore $\|u^{(t)} - \tilde{u}^{(t)}\|_\infty = O(n^{-1}\rho^{-1/2} + Z)$ with high probability.

We now bound Z. Consulting Table 1, we see that $\mathbb{E}|H_{ij}|^k = O(\rho)$ for all $k \geq 1$. Therefore, for a constant C sufficiently large, setting $\gamma = C\sqrt{\rho n}$ results in $\mathbb{E}|H_{ij}/\gamma|^k \leq 1/n$ for all $k \geq 2$ w.h.p. Since $\rho = \omega(n^{-1}\log^\epsilon n)$ and $\epsilon > 2\xi$ by the assumptions of Theorem 3, $\lambda_t - \gamma(\log n)^\xi$ is dominated by λ_t and so the first term in Equation (5) is $O(\lambda_t^{-1}\,\gamma \cdot \|u^{(s)}\|_\infty \cdot \log^\xi n) = O(\log^\xi n/(\sqrt{\rho} n))$. Next, we have

$\|H/\lambda_t\| = O(1/\sqrt{\rho n})$ w.h.p. Since κ and ξ are fixed constants, the exponent $\frac{\kappa}{8}\log^{\xi} n$ is unbounded as $n \to \infty$ and hence the second term is dominated by the first. Using this result as Z, we find that $\|u^{(t)} - \tilde{u}^{(t)}\|_\infty = O(\log^{\xi} n/(\sqrt{\rho}n))$ w.h.p. □

5. Conclusion

In this paper, we have seen that the classical perturbation theories of Weyl (1912) and Davis and Kahan (1969) are substantially improved by incorporating information about the interaction between the perturbation and the structure. Considering such interactions has been a fruitful line of recent research into spectral perturbation theory (Fan et al., 2016; Vu, 2010; O'Rourke et al., 2013; Jain and Netrapalli, 2015), however, it is typically assumed that the structure matrix is of low-rank and the noise is random. In contrast, our results hold for structure matrices of full rank and for arbitrary perturbations.

Still, we feel that the story of spectral perturbation theory is incomplete. First, it is not clear that improved eigenvector bounds should only be achieved when the structure is low-rank (or approximately so); numerical experiments indeed show that random perturbations change the eigenvectors of a full-rank matrix by much less than the bound provided by Davis-Kahan would suggest. We posit that the low-rank setting commonly assumed in the literature is merely a tractable special case of a more general phenomenon which is yet to be explained.

Second, we seek to better understand the connection between the perturbation of eigenvalues and the perturbation of eigenvectors; this connection plays a central role in the classical theory. For instance, a useful application of the Davis-Kahan theorem requires that the eigengap is bounded away from zero – this means that the perturbation in eigenvalues must not be too large compared to the gap between them. Given the ubiquity of the Davis-Kahan theorem, it is easy, perhaps, to conclude that recovery of the eigenvectors is impossible if the eigenvalue perturbation is too large. But this is clearly not true, as the spectral theorem tells us that any polynomial of a diagonalizable matrix leaves eigenvectors unchanged, while potentially greatly perturbing eigenvalues. In general, the connection between eigenvalue perturbations and eigenvector perturbations depends upon the nature of the perturbation itself. We conjecture that random perturbations are one instance in which eigenvector perturbations are less sensitive to eigenvalue perturbations than the classical theory suggests. If this is indeed the case, it would seem that the eigenvectors of the unperturbed matrix are possible to recover even when the eigengap is much smaller than that required by Davis-Kahan. This would potentially make spectral analysis feasible for problems in statistics and machine learning for which Davis-Kahan provides no meaningful bound.

Acknowledgements

This work was supported by NSF grants CCF-1422830, DMS-1547357, & RI-1550757. It was done in part while the authors were visiting the Simons Institute for the Theory of Computing.

References

Emmanuel Abbe. Community detection and stochastic block models: recent developments. 29 March 2017. arXiv preprint.

Emmanuel Abbe and Colin Sandon. Community detection in general stochastic block models: fundamental limits and efficient recovery algorithms. March 2015.

Jinho Baik, Gerard Ben Arous, and Sandrine Peche. Phase transition of the largest eigenvalue for non-null complex sample covariance matrices. March 2004.

Florent Benaych-Georges and Raj Rao Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. October 2009.

Chandler Davis and W M Kahan. Some new bounds on perturbation of subspaces. *Bull. Am. Math. Soc.*, 75(4):863–868, July 1969.

László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Spectral statistics of Erdős-Rényi graphs i: Local semicircle law. 9 March 2011. arXiv preprint.

Jianqing Fan, Weichen Wang, and Yiqiao Zhong. An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. 11 March 2016. arXiv preprint.

Roger A Horn and Charles R Johnson. *Matrix Analysis*. Cambridge University Press, 22 October 2012.

Prateek Jain and Praneeth Netrapalli. Fast exact matrix completion with finite samples. In *Conference on Learning Theory*, pages 1007–1034, 26 June 2015.

Pradipta Mitra. Entrywise bounds for eigenvectors of random graphs. *Electron. J. Comb.*, 16(1): R131, 31 October 2009.

Sean O'Rourke, Van Vu, and Ke Wang. Random perturbation of low rank matrices: Improving classical bounds. 12 November 2013. arXiv preprint.

Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. 12 November 2010. arXiv preprint.

Van Vu. Spectral norm of random matrices. *Combinatorica*, 27(6):721–736, 1 November 2007.

Van Vu. Singular vectors under random perturbation. 12 April 2010. arXiv preprint.

Van Vu. A simple SVD algorithm for finding hidden partitions. 15 April 2014. arXiv preprint.

Hermann Weyl. Das asymptotische verteilungsgesetz der eigenwerte linearer partieller differentialgleichungen (mit einer anwendung auf die theorie der hohlraumstrahlung). *Math. Ann.*, 71(4): 441–479, 1 December 1912.

Y Yu, T Wang, and R J Samworth. A useful variant of the Davis–Kahan theorem for statisticians. *Biometrika*, 102(2):315–323, 1 June 2015.

P Zhu and A V Knyazev. Angles between subspaces and their tangents. *J. Numer. Math.*, 21(4), January 2013.

Appendix A. Regarding the stochastic blockmodel

A.1 On the entries of H

Lemma 2. *For any $k \geq 1$, $\mathbb{E}|H_{ij}|^k = O(\rho)$.*

Proof. Recall that we define $H = A - M$, and hence $H_{ij} = A_{ij} - M_{ij}$. A_{ij} is 1 with probability M_{ij}, and 0 with probability $1 - M_{ij}$. Therefore:

$$H_{ij} = \begin{cases} 1 - M_{ij}, & \text{with probability } M_{ij}, \\ -M_{ij}, & \text{with probability } 1 - M_{ij}. \end{cases}$$

Therefore:

$$\mathbb{E}|H_{ij}|^k = (1 - M_{ij})^k \cdot M_{ij} + (M_{ij})^k \cdot (1 - M_{ij}).$$

Recall that $M_{ij} = \rho \cdot P^{(0)}_{z(i),z(j)}$, where z is the assignment map from nodes to communities. In Setting 1, is assumed that each entry of $\rho \cdot P^{(0)}$ is in the unit interval; therefore, $0 \leq M_{ij} \leq 1$. Hence $0 \leq (1 - M_{ij})^k \leq 1$ for all $k \geq 0$, and we have:

$$\leq M_{ij} + (M_{ij})^k.$$

Suppose $k \geq 1$. Then since $0 \leq M_{ij} \leq 1$, we have $(M_{ij})^k \leq M_{ij}$. Hence:

$$\leq 2M_{ij}.$$

Since $M_{ij} = \rho \cdot P^{(0)}_{z(i),z(j)}$, and $P^{(0)}$ is a fixed matrix with constant entries, $M_{ij} = \Theta(\rho)$. Therefore:

$$= O(\rho).$$

□

A.2 Proof of Corollary 1

We now prove the following result which was originally stated in Section 2:

Corollary 1. *Suppose that the assumptions of Theorem 3 hold. Define $\hat{M}$ as in Algorithm 1. Then $\|\hat{M} - M\|_{max} = O(\sqrt{\rho/n} \cdot \log^\xi n)$ with high probability.*

Proof. Recall that we define $\hat{M}$ to be the rank-K approximation of M using the top K eigenvectors of A in magnitude. Let $s_1, \ldots, s_K$ be such that $|\lambda_{s_1}| \geq |\lambda_{s_2}| \geq \cdots \geq |\lambda_{s_K}|$ are the top K eigenvalues of M in absolute value. We first argue that $|\tilde{\lambda}_{s_1}| \geq |\tilde{\lambda}_{s_2}| \geq \cdots \geq |\tilde{\lambda}_{s_K}|$ are the top eigenvalues of A in absolute value with high probability as $n \to \infty$. This follows from a simple eigenvalue perturbation argument: By Weyl's theorem, for any $t \in [n]$, $|\tilde{\lambda}_t - \lambda_t| \leq \|H\| = O(\sqrt{\rho n})$. As a result, if $\lambda_t = 0$ then $\tilde{\lambda}_t = O(\sqrt{\rho n})$. Since $\tilde{\lambda}_{s_1}, \ldots, \tilde{\lambda}_{s_K}$ are $\Theta(\rho n)$, there is a gap of size $\Theta(\rho n)$ w.h.p., between them and the remaining eigenvalues of A, and therefore the top K eigenvalues of A are as claimed.

Therefore, we assume that the top K eigenvalues of A in absolute value are $\tilde{\lambda}_{s_1}, \ldots, \tilde{\lambda}_{s_K}$. Then:

$$\hat{M} = \sum_{k=1}^{K} \tilde{\lambda}_{s_k} \tilde{u}^{(s_k)} \otimes \tilde{u}^{(s_k)},$$

where $\tilde{u}^{(s_k)} \otimes \tilde{u}^{(s_k)}$ is the outer product of these two vectors. Since M is rank K, we have

$$M = \sum_{k=1}^{K} \lambda_{s_k} u^{(s_k)} \otimes u^{(s_k)}.$$

As a result, we have

$$M_{ij} = \sum_{k=1}^{K} \lambda_{s_k} u_i^{(s_k)} u_j^{(s_k)}, \qquad \hat{M}_{ij} = \sum_{k=1}^{K} \tilde{\lambda}_{s_k} \tilde{u}_i^{(s_k)} \tilde{u}_j^{(s_k)}.$$

For any $t \in \{s_1, \ldots, s_K\}$, define $\Delta^{(t)} = \tilde{u}^{(t)} - u^{(t)}$ and let $\epsilon_t = \tilde{\lambda}_t - \lambda_t$. Then $\tilde{u}^{(t)} = u^{(t)} + \Delta^{(t)}$ and $\tilde{\lambda}_t = \lambda_t + \epsilon_t$. Hence:

$$\hat{M}_{ij} = \sum_{k=1}^{K} (\lambda_{s_k} + \epsilon_{s_k})(u_i^{(s_k)} + \Delta_i^{(s_k)})(u_j^{(s_k)} + \Delta_j^{(s_k)}).$$

From Theorem 3, we have that $|\Delta_i^{(t)}| \leq C\rho^{-1/2} n^{-1} \log^\xi n$ simultaneously for all $t \in \{s_1, \ldots, s_K\}$ and $i \in [n]$ with high probability. Furthermore, consulting Table 1 shows that $|u_i^{(t)}| = \Theta(1/\sqrt{n})$. Combining this with Weyl's bound of $\epsilon_t \leq \|H\| = O(\sqrt{\rho n})$, it is easy to see that:

$$\begin{aligned}
\hat{M}_{ij} &= M_{ij} + O\left(\sum_{k=1}^{K} \lambda_{s_k} u_i^{(s_k)} \Delta_j^{(s_k)} \right), \\
&= M_{ij} + O\left(K \cdot \rho n \cdot \frac{1}{\sqrt{n}} \cdot \frac{\log^\xi n}{n\sqrt{\rho}} \right), \\
&= M_{ij} + O\left(\sqrt{\frac{\rho}{n}} \log^\xi n \right).
\end{aligned}$$

□

A.3 Proof of Algorithm 1's consistency

We now prove the consistency of Algorithm 1.

Theorem 10 (Consistency of Algorithm 1)**.** *Suppose that the assumptions of Theorem 3 hold. Let* $\tau = \omega(\sqrt{\rho/n} \cdot \log^\xi n)$ *and* $\tau = o(\rho)$. *Define* $\Gamma = \{z^{-1}(k)\}_{k=1}^K$ *to be the partition of* $[n]$ *into the ground-truth communities, and let* $\hat{\Gamma}$ *be the clustering returned by Algorithm 1 with inputs* A, $\tau = \tau(n)$, *and* K. *Then* $\mathbb{P}(\text{communities recovered exactly}) = \mathbb{P}(\Gamma = \hat{\Gamma}) \to 1$ *as* $n \to \infty$.

Proof. We will use Corollary 1 to show that, with high probability as $n \to \infty$, for all pairs of graph nodes i and j simultaneously, i and j belong to the same latent community if and only if $\|M_i - M_j\|_\infty < \tau$.

Recall that we write $z(i)$ to denote the latent community label of node i. Define:

$$\Delta = \min_{\substack{i,j \\ z(i) \neq z(j)}} \|M_i - M_j\|_\infty.$$

Since $M_{ij} = \rho \cdot P^{(0)}_{z(i),z(j)}$ we have:

$$\Delta = \rho \cdot \min_{k \neq k'} \|P^{(0)}_k - P^{(0)}_{k'}\|_\infty = \Theta(\rho).$$

Thus there exists a constant C (depending on $P^{(0)}$) such that for all blockmodels in the sequence, if i and j belong to different communities, then $\|M_i - M_j\|_\infty \geq C\rho$. Therefore we are able to recover the communities exactly if M is known.

Observe that:

$$\|\hat{M}_i - \hat{M}_j\|_\infty = \|M_i + (\hat{M}_i - M_i) - M_j - (\hat{M}_j - M_j)\|_\infty,$$
$$= \|(M_i - M_j) + (\hat{M}_i - M_i) - (\hat{M}_j - M_j)\|_\infty.$$

As a result,

$$\Big| \|M_i - M_j\|_\infty - \|\hat{M}_i - \hat{M}_j\|_\infty \Big| \le \|\hat{M}_i - M_i\|_\infty + \|\hat{M}_j - M_j\|_\infty,$$
$$= O\left(\sqrt{\frac{\rho}{n}} \cdot \log^\xi n \right),$$

where we have substituted the result of Corollary 1. Since $\xi < \epsilon/2$ by assumption, we have that

$$\frac{\log^\xi n}{\sqrt{n}} = o\left(\sqrt{\frac{\log^\epsilon n}{n}} \right) = o(\sqrt{\rho}),$$

where in the last step we used the assumption that $\rho = \omega(n^{-1} \log^\epsilon n)$. Therefore $\sqrt{\rho/n} \cdot \log^\xi n = o(\rho)$. In particular, if i and j belong to different communities then

$$\|\hat{M}_i - \hat{M}_j\|_\infty \ge C\rho - O\left(\sqrt{\frac{\rho}{n}} \cdot \log^\xi n \right) = \Omega(\rho).$$

Hence if $\tau = o(\rho)$, $\|\hat{M}_i - \hat{M}_j\|_\infty > \tau$ w.h.p. and thus i and j will be clustered into different communities by Algorithm 1 with high probability as $n \to \infty$.

On the other hand, suppose that i and j belong to the same community. Then, as shown above, $\|\hat{M}_i - \hat{M}_j\|_\infty = O(\sqrt{\rho/n} \cdot \log^\xi n)$. Therefore, if $\tau = \omega(\sqrt{\rho/n} \cdot \log^\xi n)$, $\|\hat{M}_i - \hat{M}_j\|_\infty \le \tau$ with high probability as $n \to \infty$, and therefore i and j are clustered together. □

A.4 A remark on the classical theory

It was remarked above that the Davis-Kahan theorem provides only a trivial bound on $\|\hat{M} - M\|_{\max}$. We now expand on this remark.

We have seen that in the context of the sparse stochastic blockmodel (i.e., Setting 1) the classical bound on the perturbation of the top K eigenvectors in 2-norm is $\Theta(1/\sqrt{\rho n})$; see Table 2 and the discussion in Theorem 3 for reference. We now argue that this implies a bound of

$$\|\hat{M} - M\|_F = \sqrt{\sum_{i,j} (\hat{M}_{ij} - M_{ij})^2} = O(\sqrt{\rho n}).$$

Recall that we have assumed for simplicity that the eigenvalues of M are non-negative. Then the top K eigenvalues of M in absolute value are simple $\lambda_1, \ldots, \lambda_K$, and:

$$M = \sum_{k=1}^{K} \lambda_k u^{(k)} \otimes u^{(k)}.$$

Assume that the top K eigenvalues of A are the largest in magnitude – as argued above, this will be true with high probability as $n \to \infty$. Then the rank K approximation of M is:

$$\hat{M} = \sum_{k=1}^{K} \tilde{\lambda}_k \tilde{u}^{(k)} \otimes \tilde{u}^{(k)}.$$

Consider the tth eigenvalue and eigenvector for $t \in [K]$; the following argument will hold for the remaining of the top K eigenvalues since they are of the same order. Write $\tilde{\lambda}_t = \lambda_t + \epsilon_t$. We have:

$$\begin{aligned}\|\lambda_t u^{(t)} \otimes u^{(t)} - \tilde{\lambda}_t \tilde{u}^{(t)} \otimes \tilde{u}^{(t)}\|_F &= \|\lambda_t u^{(t)} \otimes u^{(t)} - (\lambda_t + \epsilon_t)\tilde{u}^{(t)} \otimes \tilde{u}^{(t)}\|_F, \\ &\leq \underbrace{\lambda_t \|u^{(t)} \otimes u^{(t)} - \tilde{u}^{(t)} \otimes \tilde{u}^{(t)}\|_F}_{A} + \underbrace{|\epsilon_t| \cdot \|\tilde{u}^{(t)} \otimes \tilde{u}^{(t)}\|_F}_{B},\end{aligned}$$

Weyl's theorem gives a bound of $|\epsilon_t| \leq \|H\| = O(\sqrt{\rho n})$. Since $\tilde{u}^{(t)}$ is a unit vector, $\|\tilde{u}^{(t)} \otimes \tilde{u}^{(t)}\|_F \leq 1$, and so $B = O(\sqrt{\rho n})$.

We now bound A. Let $\Delta = \tilde{u}^{(t)} - u^{(t)}$. We have:

$$\begin{aligned}\|u^{(t)} \otimes u^{(t)} - \tilde{u}^{(t)} \otimes \tilde{u}^{(t)}\|_F &= \|u^{(t)} \otimes u^{(t)} - (u^{(t)} + \Delta) \otimes (u^{(t)} + \Delta)\|_F, \\ &\leq \|u^{(t)} \otimes \Delta\|_F + \|\Delta \otimes u^{(t)}\|_F + \|\Delta \otimes \Delta\|_F.\end{aligned}$$

Using the submultiplicative property of the Frobenius norm, we bound each of these terms by $\|\Delta\|_F = \|\Delta\|_2 = O(1/\sqrt{\rho n})$. Then, since $\lambda_t = \Theta(\rho n)$, we have a bound on A and also $\|\hat{M} - M\|_F$ of $O(\sqrt{\rho n})$.

Such a bound is not sufficient to cluster the columns of $\hat{M}$ in a way that recovers the correct clustering exactly with high probability. For instance, suppose that i and j belong to different clusters. Let $\hat{M}$ be the matrix which is identical to M, except that column and row i is made to look exactly like column j. It is easy to see that $\hat{M}$ differs from M in $O(n)$ entries, and each difference has magnitude ρ. Therefore, $\|\hat{M} - M\|_F = O(\sqrt{\rho n})$. But by construction it is impossible to distinguish i from j using $\hat{M}$. On the other hand, our bound on $\|\hat{M} - M\|_{\max}$ is sufficient, as shown in the proof of Theorem 10 above.

Appendix B. Eigenvalue perturbation proofs

B.1 Proof of Theorem 6

Theorem 6 (Eigenvalue upper bound)**.** *Let $T \in [n]$ and h be such that $|\langle x, Hx\rangle| \leq h$ for all $x \in \operatorname{Span}(\{u^{(1)}, \ldots, u^{(T)}\})$. Let $t \leq T$ and suppose that $\lambda_t - \lambda_{T+1} > 2\|H\| - h$. Then:*

$$\tilde{\lambda}_t \leq \lambda_t + h + \frac{\|H\|^2}{\lambda_t - \lambda_{T+1} + h - \|H\|}.$$

Proof. The min-max priciple says

$$\tilde{\lambda}_t = \min_{S \in \mathcal{S}_{n-t+1}} \max_{\substack{x \in S \\ \|x\|=1}} x^\intercal (M + H)x,$$

where $\mathcal{S}_{n-t+1}$ is the set of all subspaces of $\mathbb{R}^n$ of dimension $n - t + 1$. In particular, fix the subspace to be $S_{t:n} = \operatorname{Span}(\{u^{(t)}, \ldots, u^{(n)}\})$ such that

$$\leq \max_{x \in S_{t:n}} x^\intercal (M + H)x.$$

We may write any unit vector $x \in S_{t:n}$ as $\alpha u + \beta u_\perp$ for some unit vector $u \in S_{t:T}$ and some unit vector $u_\perp \in S_{T+1:n}$, with the constraint $\alpha^2 + \beta^2 = 1$. As such, the above maximization is equivalent to:

$$= \max_{\substack{\alpha, \beta \\ \alpha^2 + \beta^2 = 1}} \max_{u \in S_{t:T}} \max_{u_\perp \in S_{T+1:n}} (\alpha u + \beta u_\perp)^\intercal (M + H)(\alpha u + \beta u_\perp).$$

Expanding the quadratic form:

$$= \max_{\substack{\alpha,\beta \\ \alpha^2+\beta^2=1}} \max_{u\in S_{t:T}} \max_{u_\perp \in S_{T+1:n}} \Big\{ \alpha^2 u^\intercal M u + \alpha^2 u^\intercal H u + \cancel{2\alpha\beta u^\intercal M u_\perp} + 2\alpha\beta u^\intercal H u_\perp + \beta^2 u_\perp^\intercal M u_\perp + \beta^2 u_\perp^\intercal H u_\perp \Big\}.$$

The $u^\intercal M u_\perp$ term drops, since $M u_\perp \in S_{T+1:n}$, and this subspace is orthogonal to $S_{t:T}$, of which u is a member. We bound the remaining terms individually. First, $u^\intercal M u$ is at most λ_t, since u is restricted to $S_{t:T}$. We then bound $u^\intercal H u \leq h$ using the assumption. Both $u^\intercal H u_\perp$ and $u_\perp^\intercal H u_\perp$ can be at most $\|H\|$. Lastly, $u_\perp^\intercal M u_\perp$ can be at most λ_{T+1}, since $u_\perp \in S_{T+1:n}$. Collecting these upper bounds, we have:

$$\leq \max_{\substack{\alpha,\beta \\ \alpha^2+\beta^2=1}} \left\{\alpha^2\lambda_t + \alpha^2 h + 2\alpha\beta\|H\| + \beta^2\lambda_{T+1} + \beta^2\|H\|\right\}.$$

Now, $\alpha\beta\|H\| \leq |\beta|\|H\|$ due to the constraint $\alpha^2+\beta^2=1$. As such, the above is bounded by:

$$\begin{aligned}
&\leq \max_{0\leq\beta\leq1} \left\{(1-\beta^2)\lambda_t + (1-\beta^2)h + 2\beta\|H\| + \beta^2\lambda_{T+1} + \beta^2\|H\|\right\}, \\
&= \lambda_t + h + \max_{0\leq\beta\leq1} \Bigg\{ \underbrace{\beta^2\left(\lambda_{T+1} - \lambda_t - h + \|H\|\right) + 2\beta\|H\|}_{g(\beta)} \Bigg\}, \\
&= \lambda_t + h + \max_{0\leq\beta\leq1} g(\beta).
\end{aligned}$$

Thus we bound $\tilde{\lambda}_t$ by maximizing $g(\beta)$ subject to $\beta \in [0,1]$. The derivative is:

$$g'(\beta) = 2\beta\left(\lambda_{T+1} - \lambda_t - h + \|H\|\right) + 2\|H\|.$$

Solving $g'(\beta^*) = 0$ for β^*, we have:

$$\beta^* = \frac{\|H\|}{\lambda_t - \lambda_{T+1} + h - \|H\|}.$$

Note that $\beta^* \in [0,1]$ as a consequence of the assumption $\lambda_t - \lambda_{T+1} > 2\|H\| - h$. Lastly, substituting this maximizing value into $g(\beta)$, we obtain:

$$\begin{aligned}
\tilde{\lambda}_t &\leq \lambda_t + h + g(\beta^*), \\
&\leq \lambda_t + h - \frac{\|H\|^2}{\lambda_t - \lambda_{T+1} + h - \|H\|} + \frac{2\|H\|^2}{\lambda_t - \lambda_{T+1} + h - \|H\|}, \\
&= \lambda_t + h + \frac{\|H\|^2}{\lambda_t - \lambda_{T+1} + h - \|H\|}.
\end{aligned}$$

□

B.2 Bounding perturbations at both ends of the spectrum

We now give the general result which bounds the perturbation of eigenvalues at both ends of the spectrum.

Theorem 11 (Eigenvalue perturbation)**.** *Let $s^\uparrow, s^\downarrow \in \{0, \ldots, n+1\}$ be such that $s^\uparrow < s^\downarrow$. Let h be such that $|\langle x, Hx\rangle| \leq h$ for all $x \in \mathrm{Span}\left(\{u^{(1)}, \ldots, u^{(s^\uparrow)}\}\right)$ and for all $x \in \mathrm{Span}\left(\{u^{(s^\downarrow)}, \ldots, u^{(n)}\}\right)$. Then for any $t \leq s^\uparrow$, if $\lambda_t - \lambda_{s^\uparrow+1} > 2\|H\| - h$:*

$$\lambda_t - h \leq \tilde{\lambda}_t \leq \lambda_t + h + \frac{\|H\|^2}{\lambda_t - \lambda_{s^\uparrow+1} + h - \|H\|},$$

and for any $t \geq s^\downarrow$, if $\lambda_{s^\downarrow} - \lambda_t > 2\|H\| - h$:

$$\lambda_t - h - \frac{\|H\|^2}{\lambda_{s^\downarrow+1} - \lambda_t + h - \|H\|} \leq \tilde{\lambda}_t \leq \lambda_t + h.$$

Proof. The statement for $t \leq s^\uparrow$ has already been proven in Theorems 5 and 6. The statement for $t \geq s^\downarrow$ follows from a symmetric argument. Let $\hat{M} = -M$ and $\hat{H} = -H$. Let $\mu_1 \geq \cdots \geq \mu_n$ be the eigenvalues of $\hat{M}$. Then $\mu_i = -\lambda_{n-i+1}$ for any $1 \leq i \leq n$. Similarly, $\lambda_i = -\mu_{n-i+1}$. Furthermore, define $v^{(i)} = u^{(n-i+1)}$. Then $v^{(i)}$ is an eigenvector of $\hat{M}$ for the eigenvalue μ_i. It follows that for any $x \in \mathrm{Span}(\{v^{(1)}, \ldots, v^{(n-s^\downarrow+1)}\})$, we have $|x^\intercal \hat{M} x| \leq h$. In addition, we have $\mu_{n-s^\downarrow+1} - \mu_{n-s^\downarrow+2} > 2\|H\| - h$. Therefore, applying Theorems 5 and 6 to $\hat{M} + \hat{H}$, we have, for any $t \leq n - s^\downarrow + 1$:

$$\mu_t - h \leq \tilde{\mu}_t \leq \mu_t + h + \frac{\|H\|^2}{\mu_t - \mu_{n-s^\downarrow+2} + h - \|H\|}.$$

Now, $\tilde{\mu}_t = -\tilde{\lambda}_{n-t+1}$, such that:

$$-\mu_t - h - \frac{\|H\|^2}{\mu_t - \mu_{n-s^\downarrow+2} + h - \|H\|} \leq \tilde{\lambda}_{n-t+1} \leq -\mu_t + h.$$

And recall that $-\mu_t = \lambda_{n-t+1}$. Hence, for any $t \leq n - s^\downarrow + 1$:

$$\lambda_{n-t+1} - h - \frac{\|H\|^2}{\lambda_{s^\downarrow-1} - \lambda_{n-t+1} + h - \|H\|} \leq \tilde{\lambda}_{n-t+1} \leq \lambda_{n-t+1} + h.$$

Finally, we make a change of index such that $t \mapsto n - t + 1$. Then for any $t \geq s^\downarrow$:

$$\lambda_t - h - \frac{\|H\|^2}{\lambda_{s^\downarrow-1} - \lambda_t + h - \|H\|} \leq \tilde{\lambda}_t \leq \lambda_t + h.$$

□

Appendix C. Eigenvector perturbation proofs

C.1 Proof of Theorem 7: the Neumann trick

Theorem 7 (Neumann trick)**.** *Fix a $t \in [n]$. Suppose that $\|H\| < |\tilde{\lambda}_t|$. Then:*

$$\tilde{u}^{(t)} = \sum_{s=1}^{n} \frac{\lambda_s}{\tilde{\lambda}_t} \cdot \left\langle \tilde{u}^{(t)}, u^{(s)} \right\rangle \sum_{p \geq 0} \left(\frac{H}{\tilde{\lambda}_t}\right)^p u^{(s)}.$$

Proof. Since $\tilde{u}^{(t)}$ is an eigenvector of $M + H$ with eigenvalue $\tilde{\lambda}_t$, we have $(M + H)\tilde{u}^{(t)} = \tilde{\lambda}_t \tilde{u}^{(t)}$. Rearranging, we obtain $M\tilde{u}^{(t)} = (\tilde{\lambda}_t I - H)\tilde{u}^{(t)}$. By the assumption that $\|H\| < |\tilde{\lambda}_t|$ it follows that $\tilde{\lambda}_t$ is not an eigenvalue of H, and and so $(\tilde{\lambda}_t I - H)$ is invertible. Therefore:

$$\tilde{u}^{(t)} = \frac{1}{\tilde{\lambda}_t} \left(I - \frac{H}{\tilde{\lambda}_t}\right)^{-1} M \tilde{u}^{(t)}.$$

Since $\|H\| < \tilde{\lambda}_t$, we may expand $(I - H/\tilde{\lambda}_t)$ in a Neumann series:

$$= \frac{1}{\tilde{\lambda}_t} \sum_{k \geq 0} \left(\frac{H}{\tilde{\lambda}_t} \right)^k M \tilde{u}^{(t)}.$$

The eigenvectors of M form an orthonormal basis for $\mathbb{R}^n$. We may therefore write $\tilde{u}^{(t)} = \sum_{s=1}^n \langle \tilde{u}^{(t)}, u^{(s)} \rangle u^{(s)}$. Using this in the above, we find:

$$\begin{aligned}
&= \frac{1}{\tilde{\lambda}_t} \sum_{k \geq 0} \left(\frac{H}{\tilde{\lambda}_t} \right)^k \sum_{s=1}^n \left\langle \tilde{u}^{(t)}, u^{(s)} \right\rangle M u^{(s)}, \\
&= \frac{1}{\tilde{\lambda}_t} \sum_{k \geq 0} \left(\frac{H}{\tilde{\lambda}_t} \right)^k \sum_{s=1}^n \lambda_s \left\langle \tilde{u}^{(t)}, u^{(s)} \right\rangle u^{(s)}, \\
&= \sum_{s=1}^n \frac{\lambda_s}{\tilde{\lambda}_t} \left\langle \tilde{u}^{(t)}, u^{(s)} \right\rangle \sum_{k \geq 0} \left(\frac{H}{\tilde{\lambda}_t} \right)^k u^{(s)}.
\end{aligned}$$

□

C.2 A general perturbation bound based on the Neumann trick

The result stated in Theorem 8 is a corollary of a more general perturbation result, which we state below. The theorem takes as input bounds on the perturbation of eigenvalues and the angle of the perturbation in eigenvectors. Theorem 8 uses Weyl's theorem and the Davis-Kahan to provide these bounds, however if better bounds are available the following result will take advantage of them.

Theorem 12. *Fix $t \in [n]$. Define $\epsilon = |\lambda_t - \tilde{\lambda}_t| / |\lambda_t|$ and let θ_s be the angle between $\tilde{u}^{(t)}$ and $u^{(s)}$. Then:*

$$\begin{aligned}
\left| u_\alpha^{(t)} - \tilde{u}_\alpha^{(t)} \right| \leq \left| u_\alpha^{(t)} \right| \cdot \left(\sin^2 \theta_t + \frac{\epsilon}{|\lambda_t| - \epsilon} \right) + \left(\frac{|\lambda_t|}{|\lambda_t| - \epsilon} \right)^2 \cdot \zeta_\alpha^{(s)} \\
+ \sum_{s \neq t} \frac{|\lambda_s| \cdot |\cos \theta_s|}{|\lambda_t| - \epsilon} \cdot \left(\left| u_\alpha^{(s)} \right| + \left[\frac{|\lambda_t|}{|\lambda_t| - \epsilon} \right] \cdot \zeta_\alpha^{(s)} \right).
\end{aligned}$$

where $\zeta^{(s)}$ is the n-vector whose αth entry is defined to be

$$\zeta_\alpha^{(s)} = \left| \left[\sum_{k \geq 1} \left(\frac{H}{\lambda_t} \right)^k u^{(s)} \right]_\alpha \right|.$$

Proof. Define

$$\psi^{(s)} = \frac{\lambda_s}{\tilde{\lambda}_t} \left\langle \tilde{u}^{(t)}, u^{(s)} \right\rangle \sum_{k \geq 0} \left(\frac{H}{\tilde{\lambda}_t} \right)^k u^{(s)}.$$

Note that $\psi^{(s)}$ is a vector, and we write $\psi_\alpha^{(s)}$ to denote its αth element. Using this notation, Theorem 7 is simply restated as: $\tilde{u}^{(t)} = \sum_{s=1}^n \psi^{(s)}$. In particular we have equality for every entry, such that:

$$\tilde{u}_\alpha^{(t)} = \sum_{s=1}^n \psi_\alpha^{(s)}.$$

Our goal is to bound $|u_\alpha^{(t)} - \tilde{u}_\alpha^{(t)}|$. Using the above expression for $\tilde{u}_\alpha^{(t)}$, we obtain:

$$\left| u_\alpha^{(t)} - \tilde{u}_\alpha^{(t)} \right| = \left| u_\alpha^{(t)} - \sum_{s=1}^n \psi_\alpha^{(s)} \right|.$$

We extract the $s = t$ term from the sum and use the triangle inequality to obtain:

$$
\begin{aligned}
&= \left| u_\alpha^{(t)} - \psi_\alpha^{(t)} - \sum_{s \neq t} \psi_\alpha^{(s)} \right|, \\
&\leq \left| u_\alpha^{(t)} - \psi_\alpha^{(t)} \right| + \sum_{s \neq t} \left| \psi_\alpha^{(s)} \right|.
\end{aligned} \tag{6}
$$

We begin by bounding the first term. We have:

$$
\begin{aligned}
\left| u_\alpha^{(t)} - \psi_\alpha^{(t)} \right| &= \left| u_\alpha^{(t)} - \frac{\lambda_t}{\tilde{\lambda}_t} \left\langle \tilde{u}^{(t)}, u^{(t)} \right\rangle \left[\sum_{k \geq 0} \left(\frac{H}{\tilde{\lambda}_t} \right)^k u^{(t)} \right]_\alpha \right|, \\
&= \left| u_\alpha^{(t)} - \frac{\lambda_t}{\tilde{\lambda}_t} \cdot \cos\theta_t \cdot \left[\sum_{k \geq 0} \left(\frac{H}{\tilde{\lambda}_t} \right)^k u^{(t)} \right]_\alpha \right|.
\end{aligned}
$$

Here we used the assumption that the angle between $\tilde{u}^{(t)}$ and $u^{(t)}$ is acute. We extract the $k = 0$ term from the series and use the triangle inequality again:

$$
= \underbrace{\left| u_\alpha^{(t)} - \frac{\lambda_t}{\tilde{\lambda}_t} \cdot \cos\theta_t \cdot u_\alpha^{(t)} \right|}_{\text{A}} + \underbrace{\left| \frac{\lambda_t}{\tilde{\lambda}_t} \cdot \cos\theta_t \cdot \left[\sum_{k \geq 1} \left(\frac{H}{\tilde{\lambda}_t} \right)^k u^{(t)} \right]_\alpha \right|}_{\text{B}}. \tag{7}
$$

We now bound A. We have

$$
\begin{aligned}
\left| u_\alpha^{(t)} - \frac{\lambda_t}{\tilde{\lambda}_t} \cos\theta_t u_\alpha^{(t)} \right| &= \left| u_\alpha^{(t)} \right| \cdot \left| 1 - \frac{\lambda_t}{\tilde{\lambda}_t} \cos\theta_t \right|, \\
&= \left| u_\alpha^{(t)} \right| \cdot \left| 1 - \frac{\tilde{\lambda}_t + (\lambda_t - \tilde{\lambda}_t)}{\tilde{\lambda}_t} \cos\theta_t \right|, \\
&= \left| u_\alpha^{(t)} \right| \cdot \left| 1 - \left(1 - \frac{\lambda_t - \tilde{\lambda}_t}{\tilde{\lambda}_t} \right) \cos\theta_t \right|, \\
&\leq \left| u_\alpha^{(t)} \right| \cdot \left(|1 - \cos\theta_t| + \left| \frac{\lambda_t - \tilde{\lambda}_t}{\tilde{\lambda}_t} \cdot \cos\theta_t \right| \right).
\end{aligned}
$$

Since θ_t is an acute angle, we have $0 \leq \cos\theta_t \leq 1$, and so $|1 - \cos\theta_t| = 1 - \cos\theta_t$. But $\cos\theta_t = \sqrt{1 - \sin^2\theta_t} \leq 1 - \sin^2\theta_t$, such that:

$$
\leq \left| u_\alpha^{(t)} \right| \cdot \left(\sin^2\theta_t + \left| \frac{\lambda_t - \tilde{\lambda}_t}{\tilde{\lambda}_t} \right| \right). \tag{8}
$$

Because we view $\tilde{\lambda}_t$ as a perturbation of λ_t, it is natural to assume that λ_t is known and that we have a bound on $|\lambda_t - \tilde{\lambda}_t|$, and that we do not know $\tilde{\lambda}_t$. It is therefore desirable to upper bound $1/|\tilde{\lambda}_t|$ in terms of $\epsilon = |\lambda_t - \tilde{\lambda}_t|$ and λ_t. We have:

$$
\frac{1}{\left| \tilde{\lambda}_t \right|} = \frac{1}{\left| \lambda_t + \tilde{\lambda}_t - \lambda_t \right|} \leq \frac{1}{|\lambda_t| - \left| \tilde{\lambda}_t - \lambda_t \right|} = \frac{1}{|\lambda_t| - \epsilon}. \tag{9}
$$

Therefore we may write Equation (8) as:

$$\left|u_\alpha^{(t)} - \frac{\lambda_t}{\tilde{\lambda}_t}\left\langle \tilde{u}^{(t)}, u^{(t)}\right\rangle u_\alpha^{(t)}\right| \leq \left|u_\alpha^{(t)}\right| \cdot \left(\sin^2\theta_t + \frac{\epsilon}{|\lambda_t| - \epsilon}\right).$$

We now turn to bounding part B of Equation (7). We have:

$$\begin{aligned}
\left|\left[\sum_{k\geq 1}\left(\frac{H}{\tilde{\lambda}_t}\right)^k u^{(t)}\right]_\alpha\right| &\leq \sum_{k\geq 1}\left|\left[\left(\frac{H}{\tilde{\lambda}_t}\right)^k u^{(t)}\right]_\alpha\right|, \\
&\leq \sum_{k\geq 1}\left|\left[\left(\frac{\lambda_t}{\tilde{\lambda}_t}\right)^k \left(\frac{H}{\lambda_t}\right)^k u^{(t)}\right]_\alpha\right|, \\
&= \sum_{k\geq 1}\left|\frac{\lambda_t}{\tilde{\lambda}_t}\right|^k \cdot \left|\left[\left(\frac{H}{\lambda_t}\right)^k u^{(t)}\right]_\alpha\right|,
\end{aligned}$$

From Equation (9), we have:

$$\begin{aligned}
&\leq \sum_{k\geq 1}\left(\frac{|\lambda_t|}{|\lambda_t| - \epsilon}\right)^k \cdot \left|\left[\left(\frac{H}{\lambda_t}\right)^k u^{(t)}\right]_\alpha\right|, \\
&\leq \frac{|\lambda_t|}{|\lambda_t| - \epsilon} \cdot \sum_{k\geq 1}\left|\left[\left(\frac{H}{\lambda_t}\right)^k u^{(t)}\right]_\alpha\right|, \\
&= \frac{|\lambda_t|}{|\lambda_t| - \epsilon} \cdot \zeta_\alpha^{(t)} \qquad (10)
\end{aligned}$$

As such, part B is bounded as:

$$\begin{aligned}
\left|\frac{\lambda_t}{\tilde{\lambda}_t} \cdot \cos\theta_t \cdot \left[\sum_{k\geq 1}\left(\frac{H}{\tilde{\lambda}_t}\right)^k u^{(t)}\right]_\alpha\right| &\leq \left|\frac{\lambda_t}{\tilde{\lambda}_t}\right| \cdot \cos\theta_t \cdot \left|\left[\sum_{k\geq 1}\left(\frac{H}{\tilde{\lambda}_t}\right)^k u^{(t)}\right]_\alpha\right|, \\
&\leq \left(\frac{|\lambda_t|}{|\lambda_t| - \epsilon}\right)^2 \cdot \zeta_\alpha^{(t)}.
\end{aligned}$$

Where we used the fact that $\cos\theta_t \leq 1$ in the last line. We have therefore bounded the first term in Equation (6) by:

$$\left|u_\alpha^{(t)} - \psi_\alpha^{(t)}\right| \leq \left|u_\alpha^{(t)}\right| \cdot \left(\sin^2\theta_t + \frac{\epsilon}{|\lambda_t| - \epsilon}\right) + \left(\frac{|\lambda_t|}{|\lambda_t| - \epsilon}\right)^2 \cdot \zeta_\alpha^{(t)}. \qquad (11)$$

We now bound the second term in Equation (6):

$$\left|\psi_\alpha^{(s)}\right| = \left|\frac{\lambda_s}{\tilde{\lambda}_t}\left\langle \tilde{u}^{(t)}, u^{(s)}\right\rangle \left[\sum_{k\geq 0}\left(\frac{H}{\tilde{\lambda}_t}\right)^k u^{(s)}\right]_\alpha\right|,$$

First, the magnitude of the dot product is $|\cos\theta_s|$ by definition, hence:

$$= \left|\frac{\lambda_s}{\tilde{\lambda}_t}\right| \cdot |\cos\theta_s| \cdot \left|\left[\sum_{k\geq 0}\left(\frac{H}{\tilde{\lambda}_t}\right)^k u^{(s)}\right]_\alpha\right|.$$

Extracting the $k = 0$ term from the sum, we have:

$$= \left|\frac{\lambda_s}{\tilde{\lambda}_t}\right| \cdot |\cos\theta_s| \cdot \left| u_\alpha^{(s)} + \left[\sum_{k\geq 1}\left(\frac{H}{\tilde{\lambda}_t}\right)^k u^{(s)}\right]_\alpha \right|,$$

$$\leq \left|\frac{\lambda_s}{\tilde{\lambda}_t}\right| \cdot |\cos\theta_s| \cdot \left(\left|u_\alpha^{(s)}\right| + \left| \left[\sum_{k\geq 1}\left(\frac{H}{\tilde{\lambda}_t}\right)^k u^{(s)}\right]_\alpha \right| \right).$$

We can bound the sum as we did in Equation (10). We obtain:

$$\leq \left|\frac{\lambda_s}{\tilde{\lambda}_t}\right| \cdot |\cos\theta_s| \cdot \left(\left|u_\alpha^{(s)}\right| + \left[\frac{|\lambda_t|}{|\lambda_t| - \epsilon}\right] \cdot \zeta_\alpha^{(s)} \right).$$

Using the bound for $1/|\tilde{\lambda}_t|$ derived in Equation (9), we have:

$$\leq \frac{|\lambda_s|}{|\lambda_t| - \epsilon} \cdot |\cos\theta_s| \cdot \left(\left|u_\alpha^{(s)}\right| + \left[\frac{|\lambda_t|}{|\lambda_t| - \epsilon}\right] \cdot \zeta_\alpha^{(s)} \right).$$

Substituting this result and Equation (11) into Equation (6), we arrive at:

$$\begin{aligned} \left|u_\alpha^{(t)} - \tilde{u}_\alpha^{(t)}\right| \leq \left|u_\alpha^{(t)}\right| \cdot \left(\sin^2\theta_t + \frac{\epsilon}{|\lambda_t| - \epsilon}\right) + \left(\frac{|\lambda_t|}{|\lambda_t| - \epsilon}\right)^2 \cdot \zeta_\alpha^{(s)} \\ + \sum_{s\neq t} \frac{|\lambda_s| \cdot |\cos\theta_s|}{|\lambda_t| - \epsilon} \cdot \left(\left|u_\alpha^{(s)}\right| + \left[\frac{|\lambda_t|}{|\lambda_t| - \epsilon}\right] \cdot \zeta_\alpha^{(s)} \right). \end{aligned}$$

□

C.3 Results concerning the perturbation of subspaces

In this section, we state results on the perturbation of subspaces which are used in various proofs; in particular, the proof of Theorem 8. The purpose of these results is to handle the case when an eigenspace $\mathcal{U}$ of M has dimensionality larger than one. In this case, the basis of $\mathcal{U}$ is determined only up to an orthogonal transformation. In most practical applications, however, we assume that the corresponding subspace of the perturbed matrix $M + H$ has a fixed basis. Therefore we wish to find a basis of $\mathcal{U}$ and a bijection between its basis vectors and the basis of $\tilde{\mathcal{U}}$ such that each vector is close to its counterpart in angle.

To begin, recall the definition of the principal angles between subspaces:

Definition 2 (Principal angles between subspaces Zhu and Knyazev (2013))**.** Let $\mathcal{U}$ and $\tilde{\mathcal{U}}$ be two d-dimensional subspaces of $\mathbb{R}^n$, and let U and $\tilde{U}$ be any orthogonal matrices whose columns form orthonormal bases for $\mathcal{U}$ and $\tilde{\mathcal{U}}$ respectively. Let $\sigma_1 \geq \cdots \geq \sigma_d$ be the singular values of $U^\intercal \tilde{U}$. The ith principal angle between $\mathcal{U}$ and $\tilde{\mathcal{U}}$ is defined to be $\cos^{-1}\sigma_i$. We write

$$\Theta(\mathcal{U}, \tilde{\mathcal{U}}) = \Theta(U, \tilde{U}) = \operatorname{diag}(\cos^{-1}\sigma_1, \ldots, \cos^{-1}\sigma_d),$$

for the $d \times d$ diagonal matrix of principal angles, and $\sin\Theta(\mathcal{U}, \tilde{\mathcal{U}}) = \Theta(U, \tilde{U})$ for the diagonal matrix obtained by applying sine to every principal angle.

The Davis-Kahan theorem in its full generality bounds the principal angles between the subspaces of M and the perturbation $M + H$:

Theorem 13 (Davis-Kahan for statisticians; Yu et al. (2015))**.** *Let M and H be $n \times n$ symmetric matrices. Let the eigenvalues of M and $M + H$ be $\lambda_1 \geq \cdots \geq \lambda_n$ and $\tilde{\lambda}_1 \cdots \tilde{\lambda}_n$ respectively. Fix $1 \leq r \leq s \leq n$ and define $\delta = \min(\lambda_{r-1} - \lambda_r, \lambda_s - \lambda_{s+1})$, where we have defined $\lambda_0 = \infty$ and $\lambda_{n+1} = -\infty$ for convenience. Assume that $\delta > 0$. Let $d = s-r+1$, and let $U = (u^{(r)}, u^{(r+1)}, \ldots, u^{(s)})$ and $\tilde{U} = (\tilde{u}^{(r)}, \tilde{u}^{(r+1)}, \ldots, \tilde{u}^{(s)})$ be orthonormal $n \times d$ matrices such that $Mu^{(i)} = \lambda_i u^{(i)}$ and $(M + H)\tilde{u}^{(i)} = \tilde{\lambda}_i \tilde{u}^{(i)}$ for all $i \in \{r, \ldots, s\}$. Then:*

$$\|\sin\Theta(U, \tilde{U})\|_F \leq 2\sqrt{d} \cdot \frac{\|H\|}{\delta}.$$

The next result shows that if the basis of $\mathcal{Y}$ is fixed and we know that the maximum principal angle between $\mathcal{Y}$ and another subspace $\mathcal{X}$ is small, then we can find a suitable orthonormal basis for $\mathcal{X}$ such that the basis vectors of both subspaces are roughly aligned.

Lemma 3. *Let $\mathcal{X}$ and $\mathcal{Y}$ be d-dimensional subspaces of $\mathbb{R}^n$. Suppose that the maximum principal angle[1] between $\mathcal{X}$ and $\mathcal{Y}$ is θ, and define $\delta = \sin\theta$. Then for any orthonormal basis $y_1, \ldots, y_d$ for $\mathcal{Y}$, there exists an orthonormal basis $\hat{x}_1, \ldots, \hat{x}_d$ for $\mathcal{X}$ such that*

$$\begin{aligned} \langle \hat{x}_i, y_i \rangle &\geq 1 - \delta^2, \qquad \forall i, \\ |\langle \hat{x}_i, y_j \rangle| &\leq \delta^2, \qquad \text{when } i \neq j. \end{aligned}$$

Proof. Let $Y = (y_1, \ldots, y_d)$ be the $n \times d$ matrix of basis vectors of $\mathcal{Y}$. Let $X = (x_1, \ldots, x_d)$ be an $n \times d$ matrix whose orthonormal columns form a basis for $\mathcal{X}$; the choice of basis is arbitrary. It known that the principal angles between subspaces can be calculated by a singular value decomposition. In particular, let $U\Sigma V^\intercal$ be the SVD of $X^\intercal Y$. Assume that the singular values σ_i are placed in decreasing order along the diagonal of Σ. Let θ_i be the ith smallest principal angle. Then $\sigma_i = \cos\theta_i$. Note that

$$\cos\theta_i = \sqrt{1 - \sin^2\theta_i} \geq \sqrt{1 - \sin^2\theta} \geq \sqrt{1 - \delta^2} \geq 1 - \delta^2,$$

and therefore every singular value is bounded as $1 - \delta^2 \leq \sigma_i \leq 1$.

Let $\tilde{X} = XU$ and $\tilde{Y} = YV$. Then

$$\tilde{X}^\intercal \tilde{Y} = U^\intercal X^\intercal Y V = U^\intercal U \Sigma V^\intercal V = \Sigma,$$

where we used the fact that U and V are orthonormal $d \times d$ matrices. Next, note that $Y = \tilde{Y}V^\intercal$, and define $\hat{X} = \tilde{X}V^\intercal$. We claim that the columns of $\hat{X}$ form an orthonormal basis for $\mathcal{X}$. To see this, we first show orthonormality of the columns. We have

$$\hat{X}^\intercal \hat{X} = V\tilde{X}^\intercal \tilde{X} V^\intercal = V(XU)^\intercal (XU)V^\intercal = VU^\intercal X^\intercal XUV^\intercal = I,$$

where in the last step we use the fact that the columns of X are orthonormal, and that U and V are orthonormal matrices. Next we show that the columns of $\hat{X}$ form a basis for $\mathcal{X}$. We do so by proving that the projection operator $\hat{X}\hat{X}^\intercal$ is in fact equal to $XX^\intercal$. We have

$$\begin{aligned} \hat{X}\hat{X}^\intercal &= (\tilde{X}V^\intercal)(\tilde{X}V^\intercal)^\intercal, \\ &= \tilde{X}V^\intercal V \tilde{X}, \\ &= \tilde{X}\tilde{X}^\intercal, \\ &= (XU)(XU)^\intercal, \\ &= XUU^\intercal X^\intercal, \\ &= XX^\intercal. \end{aligned}$$

1. A principal angle θ_i is such that $0 \leq \theta_i \leq \pi/2$ by definition.

And so our claim is proven.

Now we wish to show that the basis given by $\hat{X}$ is "aligned" with the basis given by Y in the sense that the angle between corresponding basis elements is small. See that

$$\hat{X}^\intercal Y = V\tilde{X}^\intercal \tilde{Y} V^\intercal = V\Sigma V^\intercal.$$

Defining $\hat{x}_i$ as the ith column of $\hat{X}$, we have that $\langle \hat{x}_i, y_j \rangle$ is the ij element of $V\Sigma V^\intercal$. Therefore:

$$\langle \hat{x}_i, y_j \rangle = \sum_{k=1}^{d} V_{ik}\sigma_k V_{jk}.$$

Write $\sigma_k = 1 - r_k$, where $0 \le r_k \le \delta^2$. Then:

$$\begin{aligned} &= \sum_{k=1}^{d} V_{ik}V_{jk}(1 - r_k), \\ &= \sum_{k=1}^{d} V_{ik}V_{jk} - \sum_{k=1}^{d} r_k V_{ik}V_{jk}. \end{aligned}$$

The first sum is simply the dot product between the ith and jth column of V. Since V is orthogonal, this is 1 if $i = j$, and 0 otherwise. Using the notation $\delta_{i,j}$ for the Kronecker function, we have:

$$= \delta_{i,j} - \sum_{k=1}^{d} r_k V_{ik}V_{jk}.$$

We can easily bound the magnitude of the remaining sum:

$$\begin{aligned} \left| \sum_{k=1}^{d} r_k V_{ik}V_{jk} \right| &\le \sum_{k=1}^{d} r_k \left| V_{ik}V_{jk} \right|, \\ &= r_k \sum_{k=1}^{d} |V_{ik}||V_{jk}|, \\ &\le \delta^2 \sum_{k=1}^{d} |V_{ik}||V_{jk}|. \end{aligned}$$

Define the d-vector $\tilde{v}^{(\ell)}$ to be the entrywise absolute value of the ℓ-th row of V; i.e., $\tilde{v}_k^{(\ell)} = |V_{\ell k}|$. Then the above is:

$$= \delta^2 \langle \tilde{v}_k^{(i)}, \tilde{v}_k^{(j)} \rangle.$$

Applying the Cauchy-Schwarz inequality, we find:

$$\le \delta^2 \|\tilde{v}_k^{(i)}\| \|\tilde{v}_k^{(j)}\|.$$

It is easily seen that $\|\tilde{v}_k^{(\ell)}\|$ is the norm of the ℓ-th row of V. Since V is orthonormal, this is simply one. Therefore:

$$\le \delta^2.$$

As such, $\langle \hat{x}_i, y_j \rangle$ is not more than δ^2 away from $\delta_{i,j}$, proving the result. □

The following result combines the previous lemma with the Davis-Kahan theorem.

Lemma 4. *Let M and H be $n \times n$ symmetric matrices. Let the eigenvalues of M be $\lambda_1, \ldots, \lambda_n$, and the eigenvalues of $M+H$ be $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_n$. Let $\tilde{u}^{(1)}, \ldots, \tilde{u}^{(n)}$ be an orthonormal set of eigenvectors of $M+H$ such that $(M+H)\tilde{u}^{(t)} = \tilde{\lambda}_t \tilde{u}^{(t)}$. For any $s \in [n]$, let $\Lambda_s = \{i : \lambda_i = \lambda_s\}$. Define $d_s = |\Lambda_s|$, and let the gap be defined as $\delta_s = \min_{i \notin \Lambda_s} |\lambda_s - \lambda_i|$. Denote by θ_s the angle between $\tilde{u}^{(t)}$ and $u^{(s)}$. There exists an orthonormal set of eigenvectors $u^{(1)}, \ldots, u^{(n)}$ satisfying $Mu^{(s)} = \lambda_s u^{(s)}$ such that for $t \in [n]$:*

$$\sin\theta_t \leq 2\sqrt{2d_t} \cdot \frac{\|H\|}{\delta_t}, \qquad |\cos\theta_s| \leq 2\sqrt{2} \cdot \|H\| \cdot \min\left\{\frac{\sqrt{d_i}}{\delta_i}\right\}_{i \in \{s,t\}}.$$

Proof. We first show that there exists an orthonormal basis $u^{(1)}, \ldots, u^{(n)}$ of eigenvectors of M such that $u^{(i)}$ is close in angle to $\tilde{u}^{(i)}$ for all $i \in [n]$, provided that the perturbation is too large. Choose any $s \in [n]$. Define $\mathcal{X}_s$ to be the subspace of the range of M corresponding to Λ_s. That is:

$$\mathcal{X}_s = \text{Span}\left(\{x : Mx = \lambda_s x\}\right).$$

Similarly:

$$\mathcal{Y}_s = \text{Span}(\{\tilde{u}^{(i)} : i \in \Lambda_s\}).$$

Let θ be the maximum principal angle between $\mathcal{X}_s$ and $\mathcal{Y}_s$. In particular, $|\sin\theta| \leq \|\Theta(\mathcal{X}_s, \mathcal{Y}_s)\|_F$. Therefore, applying the Davis-Kahan theorem, we have that $|\sin\theta| \leq 2\sqrt{d_s} \cdot \|H\|/\delta_s$. Lemma 3 states that there exists an orthonormal basis $\{u^{(i)}\}_{i \in \Lambda_s}$ for $\mathcal{X}_s$ such that for every $i \in \Lambda_s$:

$$\left\langle \tilde{u}^{(i)}, u^{(i)} \right\rangle \geq 1 - \sin^2\theta = 1 - 4d_s \left(\frac{\|H\|}{\delta_s}\right)^2.$$

Since $\mathcal{X}_s$ is a d_s-dimensional subspace spanned by eigenvectors with the same eigenvalue, any vector in the subspace is an eigenvector. Namely, $\{u^{(i)}\}_{i \in \Lambda_s}$ is an orthonormal set of eigenvectors spanning $\mathcal{X}_s$. We can repeat this process for each eigenspace of M, resulting in the desired orthonormal basis.

Assume this basis, and consider t as fixed. Note that for any $r \in [n]$ we have:

$$\sqrt{1 - \left\langle \tilde{u}^{(r)}, u^{(r)} \right\rangle^2} \leq \sqrt{1 - \left[1 - 4d_r \cdot \left(\frac{\|H\|}{\delta_r}\right)^2\right]^2},$$

Expanding the square:

$$\begin{aligned}
&= \sqrt{1 - \left[1 - 8d_r \cdot \left(\frac{\|H\|}{\delta_r}\right)^2 + 16d_r^2 \cdot \left(\frac{\|H\|}{\delta_r}\right)^4\right]}, \\
&= \sqrt{8d_r \cdot \left(\frac{\|H\|}{\delta_r}\right)^2 - 16d_r^2 \cdot \left(\frac{\|H\|}{\delta_r}\right)^4}, \\
&\leq \sqrt{8d_r \cdot \left(\frac{\|H\|}{\delta_r}\right)^2}, \\
&= 2\sqrt{2d_r} \cdot \frac{\|H\|}{\delta_r}.
\end{aligned}$$

Define θ_s to be the angle between $\tilde{u}^{(t)}$ and $u^{(s)}$. Namely, we have

$$\sin^2\theta_t = 1 - \cos^2\theta_t = 1 - \left\langle \tilde{u}^{(t)}, u^{(t)} \right\rangle^2 \leq 8d_t \cdot \left(\frac{\|H\|}{\delta_t}\right)^2.$$

By the same token:

$$\begin{aligned}|\cos\theta_s| &= \left|\left\langle \tilde{u}^{(t)}, u^{(s)}\right\rangle\right|, \\ &\le \sqrt{1-\left\langle \tilde{u}^{(s)}, u^{(s)}\right\rangle^2}, \\ &\le 2\sqrt{2d_s}\cdot\frac{\|H\|}{\delta_s}.\end{aligned}$$

But we also have

$$\begin{aligned}|\cos\theta_s| &= \left|\left\langle \tilde{u}^{(t)}, u^{(t)}\right\rangle\right|, \\ &\le \sqrt{1-\left\langle \tilde{u}^{(s)}, u^{(t)}\right\rangle^2}, \\ &\le 2\sqrt{2d_t}\cdot\frac{\|H\|}{\delta_t}.\end{aligned}$$

Therefore:

$$\left|\left\langle \tilde{u}^{(t)}, u^{(s)}\right\rangle\right| \le 2\sqrt{2}\cdot\|H\|\cdot\min\left\{\frac{\sqrt{d_i}}{\delta_i}\right\}_{i\in\{s,t\}}.$$

□

C.4 Proof of Theorem 8

We will prove the following theorem which was originally stated in Section 4.

Theorem 8. *For any $s\in[n]$, let $\Lambda_s=\{i:\lambda_i=\lambda_s\}$. Define $d_s=|\Lambda_s|$, and let the gap be defined as $\delta_s=\min_{i\notin\Lambda_s}|\lambda_s-\lambda_i|$. For any $s,t\in[n]$, let $\Delta_{s,t}^{-1}=\min\{d_i/\delta_i\}_{i\in\{s,t\}}$. Define $\lambda_s^*=|\lambda_s|-\|H\|$. There exists an orthonormal set of eigenvectors $u^{(1)},\ldots,u^{(n)}$ satisfying $Mu^{(s)}=\lambda_s u^{(s)}$ such that for all $t\in[n]$:*

$$\begin{aligned}\left|\tilde{u}_\alpha^{(t)}-u_\alpha^{(t)}\right| \le \left|u_\alpha^{(t)}\right|\cdot\left(8d_t\left[\frac{\|H\|}{\delta_t}\right]^2+\frac{\|H\|}{\lambda_t^*}\right)+\left(\frac{|\lambda_t|}{\lambda_t^*}\right)^2\cdot\zeta_\alpha(u^{(t)};H,\lambda_t) \\ +\frac{2\sqrt{2}\cdot\|H\|}{\lambda_t^*}\sum_{s\neq t}\frac{|\lambda_s|}{\Delta_{s,t}}\left[|u_\alpha^{(s)}|+\frac{|\lambda_t|}{\lambda_t^*}\cdot\zeta_\alpha(u^{(s)};H,\lambda_t)\right],\end{aligned}\tag{4}$$

where $\zeta(u;H,\lambda)$ is the n-vector whose αth entry is defined to be $\zeta_\alpha(u;H,\lambda)=\left|\left[\sum_{p\ge1}\left(\frac{H}{\lambda}\right)^p u\right]_\alpha\right|$.

Proof. The proof is an immediate corollary of combining Theorem 12 (given in Appendix C.2) with Lemma 4 (given in Appendix C.3), and using Weyl's bound of $\|H\|$ for the perturbation of eigenvalues. □

Appendix D. Results concerning random perturbations

In this section we collect the proof of various results on random perturbations which are used throughout the paper. In what follows, the term *symmetric random matrix* will have a technical meaning.

Definition 3. A *symmetric random matrix* H is an $n\times n$ matrix whose entries are random variables satisfying $\mathbb{E}H_{ij}=0$. Furthermore, we assume that the entries along the diagonal and in the upper-triangle ($j\ge i$) are statistically independent, while the entries in the lower-triangle ($j<i$) are constrained to be equal to their transposes: $H_{ij}=H_{ji}$.

D.1 The spectral norm of random matrices

Throughout this paper we have used the following standard result from random matrix theory:

Theorem 14 (Spectral norm of random matrices, Vu (2007))**.** *There are constants C and C' such that the following holds. Let H be an $n \times n$ symmetric random matrix whose entries satisfy,*

$$\mathbb{E}H_{ij} = 0, \qquad \mathbb{E}\left(H_{ij}\right)^2 \leq \sigma^2, \qquad |H_{ij}| \leq B.$$

where $\sigma \geq C' n^{-1/2} B \log^2 n$. Then, almost surely:

$$\|H\| \leq 2\sigma\sqrt{n} + C\sqrt{B\sigma} \cdot n^{1/4} \log n.$$

It can be shown that a similar lower bound holds in many cases. For instance, when the entries of H have the Gaussian distribution with unit variance, the spectral norm of H is not only $O(\sqrt{n})$, but $\Theta(\sqrt{n})$ with high probability. Since we typically use $\|H\|$ to obtain an upper-bound on the size of the perturbation, we will not need this result.

D.2 Proof of Lemma 1

Lemma 1. *Let u, v be any two fixed unit vectors in $\mathbb{R}^n$. Let H be an $n \times n$ symmetric random matrix with independent entries along the upper-triangle such that for all $j \geq i$, $\mathbb{E}H_{ij} = 0$ and H_{ij} is sub-Gaussian with parameter $\sigma_{ij} \leq \sigma$. Then $\mathbb{P}(|\langle u, Hv\rangle| \geq \gamma) \leq 2\exp\{-\gamma^2/(8\sigma^2)\}$.*

Proof. We have

$$\langle u, Hv\rangle = \sum_{i=1}^n \sum_{j=1}^n u_i H_{ij} v_j = \sum_{i=1}^n u_i v_i H_{ii} + \sum_{j>i} (u_i v_j + u_j v_i) H_{ij}.$$

The right hand side is a sum of independent random variables. We therefore apply the Hoeffding inequality in its general form for sub-Gaussian random variables to obtain an upper bound (see Proposition 5.10 in Vershynin (2010)). We find:

$$\begin{aligned} \mathbb{P}(|\langle u, Hv\rangle| \geq \gamma) &\leq 2\exp\left\{ -\frac{\frac{1}{2}\gamma^2}{\sum_{i=1}^n (u_i v_i \sigma_{ii})^2 + \sum_{j>i} \left[(u_i v_j + u_j v_i)\sigma_{ij}\right]^2} \right\}, \\ &\leq 2\exp\left\{ -\frac{\frac{1}{2}\gamma^2}{\sigma^2 \left[\sum_{i=1}^n (u_i v_i)^2 + \sum_{j>i} (u_i v_j + u_j v_i)^2\right]} \right\}. \end{aligned} \tag{12}$$

We have

$$\sum_{i=1}^n (u_i v_i)^2 \leq \sum_{i=1}^n \sum_{j=1}^n (u_i v_j)^2 = \sum_{i=1}^n u_i^2 \sum_{j=1}^n v_j^2 = \|u\|_2^2 \cdot \|v\|_2^2 = 1. \tag{13}$$

Similarly,

$$\begin{aligned} \sum_{j>i} (u_i v_j + u_j v_i)^2 &\leq \sum_{j>i} \left[(u_i v_j)^2 + (u_j v_i)^2 + 2|u_i u_j v_i v_j|\right], \\ &= \sum_{j>i} (u_i v_j)^2 + \sum_{j>i} (u_j v_i)^2 + \sum_{j>i} +2|u_i u_j v_i v_j|, \\ &\leq \sum_{i=1}^n \sum_{j=1}^n (u_i v_j)^2 + \sum_{i=1}^n \sum_{j=1}^n (u_j v_i)^2 + \sum_{j>i} +2|u_i u_j v_i v_j|. \end{aligned}$$

The first two sums are each bounded by 1, as before:

$$
\begin{aligned}
&\le 2 + \sum_{j>i} +2|u_i u_j v_i v_j|, \\
&\le 2 + \sum_{i=1}^{n} \sum_{j=1}^{n} |u_i u_j v_i v_j|, \\
&= 2 + \sum_{i=1}^{n} |u_i v_i| \sum_{j=1}^{n} |u_j v_j|.
\end{aligned}
$$

Each sum is bounded by 1 by an application of Cauchy-Schwarz. Therefore we find that the total sum is bounded by 3. Substituting this and Equation (13) into Equation (12) we see that

$$
\mathbb{P}(|\langle u, Hv \rangle| \ge \gamma) \le 2 \exp\left\{ -\frac{\gamma^2}{8\sigma^2} \right\}.
$$

□

D.3 Proof of Lemma 5

Lemma 5. *Let $\{u^{(1)}, \ldots, u^{(d)}\}$ be an orthonormal set of d vectors, and suppose that $|\langle u^{(i)}, Hu^{(j)} \rangle| \le h$ for all $i, j \in [d]$. Then $|\langle x, Hx \rangle| \le dh$ for any unit vector $x \in \mathrm{Span}(u^{(1)}, \ldots, u^{(d)})$.*

Proof. Since $u^{(1)}, \ldots, u^{(d)}$ form an orthonormal basis for the space in which x lies, we can expand x as

$$
x = \sum_{i=1}^{d} \alpha_i u^{(i)},
$$

where $\alpha_i = \langle x, u^{(i)} \rangle$. Therefore:

$$
\begin{aligned}
\langle x, Hx \rangle &= \left\langle \sum_{i=1}^{d} \alpha_i u^{(i)},\ H \sum_{j=1}^{d} \alpha_j u^{(j)} \right\rangle, \\
&= \sum_{i=1}^{d} \sum_{j=1}^{d} \alpha_i \alpha_j \left\langle u^{(i)}, Hu^{(j)} \right\rangle, \\
&\le h \sum_{i=1}^{d} \sum_{j=1}^{d} |\alpha_i \alpha_j|, \\
&= h \left(\sum_{i=1}^{d} |\alpha_i| \right) \left(\sum_{j=1}^{d} |\alpha_j| \right).
\end{aligned}
$$

Let α be the vector $(\alpha_1, \ldots, \alpha_d)^\intercal$. Then:

$$
= h \|\alpha\|_1^2.
$$

We know that $\|\alpha\|_2 = 1$ since x is a unit vector. The 1-norm is bounded by $\sqrt{d}$ times the 2-norm. Hence $\|\alpha\|_1 \le \sqrt{d} \cdot \|\alpha\|_2 = \sqrt{d}$. Hence $|\langle x, Hx \rangle| \le hd$.

□

Appendix E. Powers of random matrices and their interaction with delocalized vectors

We have seen that using the Neumann trick to bound the perturbation in eigenvectors requires bounding series expansions of the form

$$\zeta(u; H, \lambda) = \sum_{p \geq 0} \left(\frac{H}{\lambda} \right)^p u,$$

where H is a random matrix. We have given one result in Theorem 9 which shows that the ∞-norm of this series is small when u has small ∞-norm. In this section we will prove this result. In this and what follows, *symmetric random matrix* has the precise meaning as given in Definition 3 above.

E.1 Proof of the main interaction

The proof of Theorem 9 depends heavily on the following Theorem 15. Much of the proof of Theorem 15 is due to Erdős et al. (2011). We have amended this proof to provide precise bounds on the probability of the event.

Theorem 15. *Let X be a symmetric and centered random matrix of size $n \times n$. Let u be an n-vector with $\|u\|_\infty = 1$. Choose $\xi > 1$ and $0 < \kappa < 1$. Define $\mu = \left(\frac{\kappa+1}{2}\right)^{-1}$. Then with probability $1 - n^{-\frac{1}{4}(\log_\mu n)^{\xi-1}(\log_\mu e)^{-\xi}}$, for any $k \leq \frac{\kappa}{8}(\log n)^\xi$, if $\mathbb{E}|X_{ij}|^p \leq \frac{1}{n}$ for all $p \geq 2$, we have*

$$\left| \left(X^k u\right)_\alpha \right| < (\log n)^{k\xi}.$$

Proof. We will bound $\left| \left(X^k u\right)_\alpha \right|$ with a high-moment Markov inequality. Let p be a positive even integer. Then

$$\mathbb{P}\left(\left| \left(X^k u\right)_\alpha \right|\right) \leq \frac{\mathbb{E}\left[\left(X^k u\right)_\alpha^p\right]}{t^p}. \tag{14}$$

Bounding the expectation is non-trivial. We will utilize the following proof is to be found in the next subsection.

Lemma 6. *If $\mathbb{E}\left[|X_{ij}|^s\right] \leq 1/n$ for all $s \geq 2$, then*

$$\mathbb{E}\left[\left(X^k u\right)_\alpha^p\right] \leq (2pk)^{pk}.$$

Returning to the Markov inequality in Equation (14), we will choose $t = (\log n)^{k\xi}$, giving:

$$\mathbb{P}\left(\left| \left(X^k u\right)_\alpha \right| \geq (\log n)^{k\xi}\right) \leq \frac{\mathbb{E}\left[\left(X^k u\right)_\alpha^p\right]}{\left[(\log n)^{k\xi}\right]^p},$$

We now apply Lemma 6 to bound the expectation:

$$\begin{aligned} &\leq \frac{(2pk)^{pk}}{(\log n)^{pk\xi}}, \\ &= \left[\frac{2pk}{(\log n)^\xi}\right]^{pk}. \end{aligned}$$

The bound above holds for any positive even integer p. We will choose $p = \hat{p}$, where $\hat{p}$ is the smallest even integer greater than or equal to $\tilde{p} = \frac{1}{4k}(\log n)^\xi$. Since $k < \frac{1}{8}(\log n)^\xi$, we have $\tilde{p} \geq 2$, and so

$\hat{p} \geq 2$. Furthermore, we have $\hat{p} = \tilde{p} + \delta$, where $0 \leq \delta < 2$. Hence:

$$\begin{aligned}\left[\frac{2\hat{p}k}{(\log n)^\xi}\right]^{\hat{p}k} &= \left[\frac{2(\tilde{p}+\delta)k}{(\log n)^\xi}\right]^{(\tilde{p}+\delta)k}, \\ &= \left[\frac{2(\tilde{p}+\delta)k}{(\log n)^\xi}\right]^{\tilde{p}k} \cdot \left[\frac{2(\tilde{p}+\delta)k}{(\log n)^\xi}\right]^{\delta k}.\end{aligned}$$

We see that $2\tilde{p}k/(\log n)^\xi = 1/2$, hence:

$$= \left[\frac{1}{2} + \frac{2\delta k}{(\log n)^\xi}\right]^{\tilde{p}k} \cdot \left[\frac{1}{2} + \frac{2\delta k}{(\log n)^\xi}\right]^{\delta k}.$$

Because $0 \leq \delta < 2$, we have $\frac{1}{2} < \frac{1}{2} + \frac{2\delta k}{(\log n)^\xi} < 1$. And since $\delta k > 0$, the second term in the above is at most 1. Therefore:

$$\leq \left[\frac{1}{2} + \frac{2\delta k}{(\log n)^\xi}\right]^{\tilde{p}k}.$$

Using $\delta < 2$ and substituting the definitions of $\tilde{p}$ and k, we arrive at:

$$\begin{aligned}&\leq \left[\frac{1}{2} + \frac{4k}{(\log n)^\xi}\right]^{\frac{1}{4}(\log n)^\xi}, \\ &< \left[\frac{\kappa+1}{2}\right]^{\frac{1}{4}(\log n)^\xi}.\end{aligned}$$

We recognize the base of the exponent as μ^{-1}, therefore:

$$\begin{aligned}&= \mu^{-\frac{1}{4}(\log n)^\xi}, \\ &= \mu^{-\frac{1}{4}(\log_\mu n)^\xi (\log_\mu e)^{-\xi}}, \\ &= \mu^{-\frac{1}{4}(\log_\mu n)(\log_\mu n)^{\xi-1}(\log_\mu e)^{-\xi}}, \\ &= n^{-\frac{1}{4}(\log_\mu n)^{\xi-1}(\log_\mu e)^{-\xi}}.\end{aligned}$$

Therefore:

$$\begin{aligned}\mathbb{P}\left(\left|\left(X^k u\right)_\alpha\right| \geq (\log n)^{k\xi}\right) &\leq \left[\frac{2\hat{p}k}{(\log n)^\xi}\right]^{\hat{p}k}, \\ &\leq n^{-\frac{1}{4}(\log_\mu n)^{\xi-1}(\log_\mu e)^{-\xi}}.\end{aligned}$$

□

E.2 Proofs of moment bound, Lemma 6

In this subsection we derive a bound on $\mathbb{E}\left[\left(X^k u\right)_\alpha^p\right]$ under an assumption on the variance of the entries of X. In particular, we will prove Lemma 6 which is a critical component of Theorem 15.

E.2.1 Some useful results

First we derive a formalism for working with moments of random matrix products. It follows from the definition of matrix multiplication that the αth element of the vector $X^k u$ has the expansion:

$$\left(X^k u\right)_\alpha = \sum_{i_1,\ldots,i_k} X_{\alpha i_1} X_{i_1 i_2} \cdots X_{i_{k-1} i_k} u_{i_k}.$$

As a result, we have:

$$\begin{aligned}\mathbb{E}\left[(X^k u)^p_\alpha\right] &= \mathbb{E}\left[\left(\sum_{i_1,\dots,i_k} X_{\alpha i_1} X_{i_1 i_2}\cdots X_{i_{k-1} i_k} u_{i_k}\right)^p\right],\\ &= \mathbb{E}\left[\sum_{i_1^{(1)},\dots,i_k^{(1)}} \cdots \sum_{i_1^{(p)},\dots,i_k^{(p)}} \prod_{r=1}^{p} X_{\alpha i_1^{(r)}} X_{i_1^{(r)} i_2^{(r)}} \cdots X_{i_{k-1}^{(r)} i_k^{(r)}} u_{i_k^{(r)}}\right].\end{aligned}$$

Here there are p summations, each over an independently-varying set of k variables $i_1^{(r)}, \dots, i_k^{(r)}$ which range from 1 to n. We replace the variables of summation with indexing functions, defined as follows.

Definition 4. For positive integers p and k and an index $\alpha \in [n]$, a (p,k,α)-*indexing function* is a discrete map $\tau : [p] \times \{0,\dots,k\} \to [n]$ satisfying $\tau(r,0) = \alpha$ for all $r \in [p]$.

An indexing function τ corresponds to a single configuration of the variables of summation in the expectation above. That is, we may interpret $\tau(r,\ell)$ as the value of the variable $i_\ell^{(r)}$ in a particular configuration. As such, we will use the shorthand notation $\tau_\ell^{(r)} = \tau(r,\ell)$ so that $X_{i_{\ell-1}^{(r)} i_\ell^{(r)}}$ is replaced by $X_{\tau_{\ell-1}^{(r)} \tau_\ell^{(r)}}$.

Let $\mathcal{Z}_{p,k,\alpha}$ be the set of all (p,k,α)-index functions. The above expectation can be written as:

$$\begin{aligned}\mathbb{E}\left[(X^k u)^p_\alpha\right] &= \mathbb{E}\left[\sum_{\tau \in \mathcal{Z}_{p,k,\alpha}} \prod_{r=1}^{p} X_{\tau_0^{(r)}\tau_1^{(r)}} X_{\tau_1^{(r)}\tau_2^{(r)}} \cdots X_{\tau_{k-1}^{(r)}\tau_k^{(r)}} u_{\tau_k^{(r)}}\right],\\ &= \sum_{\tau \in \mathcal{Z}_{p,k,\alpha}} \mathbb{E}\left[\prod_{r=1}^{p} X_{\tau_0^{(r)}\tau_1^{(r)}} X_{\tau_1^{(r)}\tau_2^{(r)}} \cdots X_{\tau_{k-1}^{(r)}\tau_k^{(r)}} u_{\tau_k^{(r)}}\right],\\ &\le \sum_{\tau \in \mathcal{Z}_{p,k,\alpha}} \left|\mathbb{E}\left[\prod_{r=1}^{p} X_{\tau_0^{(r)}\tau_1^{(r)}} X_{\tau_1^{(r)}\tau_2^{(r)}} \cdots X_{\tau_{k-1}^{(r)}\tau_k^{(r)}} u_{\tau_k^{(r)}}\right]\right|,\\ &= \sum_{\tau \in \mathcal{Z}_{p,k,\alpha}} \underbrace{\left|\left(\prod_{r=1}^{p} u_{\tau_k^{(r)}}\right)\right|}_{\omega_u(\tau)} \cdot \underbrace{\left|\mathbb{E}\left[\prod_{r=1}^{p} X_{\tau_0^{(r)}\tau_1^{(r)}} X_{\tau_1^{(r)}\tau_2^{(r)}} \cdots X_{\tau_{k-1}^{(r)}\tau_k^{(r)}}\right]\right|}_{\varphi(\tau)},\\ &= \sum_{\tau \in \mathcal{Z}_{p,k,\alpha}} \omega_u(\tau) \cdot \varphi(\tau). \qquad (15)\end{aligned}$$

Here we write ω_u to show that ω_u is parametrized by the vector u. On the other hand, φ does not depend on u. In the following two parts, we derive bounds on this quantity under assumptions on the magnitude or variance of X_{ij}. In each case the core approach is the same: we bound the size of $\varphi(\tau)$ for any τ by using the assumptions on X, and then bound the number of τ for which φ and ω_u are non-zero.

The entries in the upper-triangle of the random matrix X are independent, but not necessarily identically distributed. Rather, the assumptions that we will place on the entries of X will not depend on the indices. As a result, it is not important to use the precise knowledge of which entries of X are selected by an indexing function τ in order to bound $\varphi(\tau)$. We will therefore partition the set of indexing functions into equivalence classes which characterize the important structure of the indexing, and then derive a bound for each equivalence class independently.

First, some notation: For a set of sets A, we write $[A]$ to denote the union of all elements of A; i.e., $[A] = \bigcup_{\gamma \in A} \gamma$. We introduce the following notion:

Definition 5. A (p,k)-*index partition* Γ is a partition of a subset of $\{1,\ldots,p\} \times \{0,\ldots,k\}$ with the property that there exists a block $\tilde{\gamma} \in \Gamma$ such that every pair of the form $(r,0)$ is in $\tilde{\gamma}$; that is:

$$\exists \tilde{\gamma} \in \Gamma \text{ s.t. } \tilde{\gamma} \supset \{(r,0) : r \in \{1,\ldots,p\}\}.$$

We call $\tilde{\gamma}$ the *root block* of Γ.

Note that a (p,k)-index partition is a partition of a *subset* of $[p] \times \{0,\ldots,k\}$; i.e., it is not necessarily the case that $[\Gamma]$ is the full set $[p] \times \{0,\ldots,k\}$. For example, any $(p-1,k-1)$-index partition is also a (p,k)-index partition by definition. We will later find it useful to make use of such "subpartitions", but for the time being we will only consider index partitions which in fact partition the full set. Let $\mathcal{P}_{p,k}$ be the set of all "full" (p,k)-index partitions Γ such that $[\Gamma] = [p] \times \{0,\ldots,k\}$.

Next, note that an index partition $\Gamma \in \mathcal{P}_{p,k}$ defines an equivalence relation on $[\Gamma]$. We use the following notation to denote this relation:

Notation. For pairs $(r,\ell), (\tilde{r},\tilde{\ell}) \in [\Gamma]$ we write $(r,\ell) \overset{\Gamma}{\sim} (\tilde{r},\tilde{\ell})$ if and only if there exists a block $\gamma \in \Gamma$ such that γ contains both (r,ℓ) and $(\tilde{r},\tilde{\ell})$.

We relate indexing functions and index partitions in the following way:

Definition 6. We say that an indexing function τ *respects* the partition $\Gamma \in \mathcal{P}_{p,k}$ when $\tau_\ell^{(r)} = \tau_{\ell'}^{(r')}$ if and only if $(r,\ell) \overset{\Gamma}{\sim} (r',\ell')$.

It is clear that for any indexing function τ, there is exactly one partition $\Gamma \in \mathcal{P}_{p,k}$ such that τ respects Γ. As such, we have implicitly established an equivalence relation between indexing functions: τ and τ' are equivalent if and only if they respect the same index partition. For an index partition $\Gamma \in \mathcal{P}_{p,k}$, write $\mathcal{Z}_{p,k,\alpha}\{\Gamma\}$ to denote the set of all indexing functions which respect Γ. Then Equation (15) can be re-written as:

$$\mathbb{E}\left[\left(X^k u\right)_\alpha^p\right] \le \sum_{\Gamma \in \mathcal{P}_{p,k}} \sum_{\tau \in \mathcal{Z}_{p,k,\alpha}\{\Gamma\}} \omega_u(\tau) \cdot \varphi(\tau). \tag{16}$$

Definition 7 (Twin property)**.** Let $\Gamma \in \mathcal{P}_{p,k}$. Let $(r,\ell) \in [\Gamma]$ and $(\tilde{r},\tilde{\ell}) \in [\Gamma]$ be distinct and such that $\ell, \tilde{\ell} > 0$. We say that (r,ℓ) and $(\tilde{r},\tilde{\ell})$ are *twins in* Γ if either:

1. $(r,\ell) \overset{\Gamma}{\sim} (\tilde{r},\tilde{\ell})$ and $(r,\ell-1) \overset{\Gamma}{\sim} (\tilde{r},\tilde{\ell}-1))$; or
2. $(r,\ell) \overset{\Gamma}{\sim} (\tilde{r},\tilde{\ell}-1)$ and $(r,\ell-1) \overset{\Gamma}{\sim} (\tilde{r},\tilde{\ell})$.

We say that a (p,k)-index partition Γ satisfies the *twin property* if for any pair $(r,\ell) \in [\Gamma]$ with $\ell > 0$ there exists a distinct $(\tilde{r},\tilde{\ell}) \in [\Gamma]$ with $\tilde{\ell} > 0$ such that (r,ℓ) and $(\tilde{r},\tilde{\ell})$ are twins in Γ.

Lemma 7. *Let τ be an indexing function respecting the partition Γ. Then (r,ℓ) and $(\tilde{r},\tilde{\ell})$ are twins in Γ if and only if $X_{\tau_{\tilde{\ell}-1}^{(\tilde{r})}\tau_{\tilde{\ell}}^{(\tilde{r})}} = X_{\tau_{\ell-1}^{(r)}\tau_\ell^{(r)}}$.*

Proof. Due to the symmetry of X and the independence of its entries along the upper triangle, we have that for any indices i,j,i',j', $X_{ij} = X_{i'j'}$ if and only if either 1) $(i,j) = (i',j')$ or 2) $(i,j) = (j',i')$. This is the case if and only if $\tau_{\tilde{\ell}-1}^{(\tilde{r})} = \tau_{\ell-1}^{(r)}$ and $\tau_{\tilde{\ell}}^{(\tilde{r})} = \tau_\ell^{(r)}$ or 2) $\tau_{\tilde{\ell}-1}^{(\tilde{r})} = \tau_\ell^{(r)}$ and $\tau_{\tilde{\ell}}^{(\tilde{r})} = \tau_{\ell-1}^{(r)}$. The fact that that this holds if and only if (r,ℓ) and $(\tilde{r},\tilde{\ell})$ are twins follows from the definition of twins and the notion of τ respecting the partition Γ. ∎

Definition 8. For any index partition $\Gamma \in \mathcal{P}_{p,k}$, denote by $T\{\Gamma\}$ the set of equivalence classes of the *twin relation*, defined on $[p] \times [k]$ by $(r, \ell) \sim (\tilde{r}, \tilde{\ell})$ if and only if (r, ℓ) and $(\tilde{r}, \tilde{\ell})$ are twins.

Notation. If τ is an indexing function which respects Γ and $\rho \in T\{\Gamma\}$, we write X_ρ to denote the random variable X_{ij} such that $X_{ij} = X_{\tau^{(r)}_{\ell-1}\tau^{(r)}_{\ell}}$ for every $(r, \ell) \in \rho$; this is well-defined as a result of Lemma 7.

Lemma 8. *Let* $\Gamma \in \mathcal{P}_{p,k}$ *and suppose* τ *is an indexing function which respects* Γ. *Then:*

$$\varphi(\tau) = \prod_{\rho \in T\{\Gamma\}} \left| \mathbb{E}\left[X_\rho^{|\rho|} \right] \right|.$$

Proof. Lemma 7 implies that the equivalence classes of the twin relation partition the pk terms of the product in φ into sets of random variables which are equal. Since the entries of X are independent random variables, the expectation factors. □

Since $\mathbb{E}X = 0$, we have the following corollary:

Corollary 2. *Suppose that* $\Gamma \in \mathcal{P}_{p,k}$ *does not satisfy the twin property; i.e., there exists a pair* $(\hat{r}, \hat{\ell}) \in [\Gamma]$ *that does not have a twin in* Γ. *Then* $\varphi(\tau) = 0$ *for every* τ *respecting* Γ.

Corollary 2 implies that only partitions satisfying the twin property contribute to the sum in Equation (16).

Lemma 9. *Let* $F = \{i : u_i \neq 0\}$. *Fix* $\alpha \in [n]$. *Suppose that* $\Gamma \in \mathcal{P}_{p,k}$ *is such that the root block* $\tilde{\gamma}$ *contains an element of the form* (r, k) *for some* $r \in [p]$. *Then if* $\alpha \notin F$ *we have* $\omega_u(\tau) = 0$ *for every* τ *which respects* Γ.

Proof. By the definition of an indexing function, $\tau_0^{(r)} = \alpha$ for every $r \in [p]$. Let r^* be such that $(r^*, k) \in \tilde{\gamma}$. If τ respects Γ, then it is necessarily the case that $\tau_k^{(r^*)} = \tau_0^{(r^*)} = \alpha$. Then $u_{\tau_k^{(r^*)}} = u_\alpha$. If $\alpha \notin F$, then $u_\alpha = 0$ and hence $\omega_u(\tau) = 0$. □

Definition 9. Fix a set $F \subset [n]$ and an index $\alpha \in [n]$. We write $\mathcal{P}_{p,k}^{+(F,\alpha)}$ to denote the set of all $\Gamma \in \mathcal{P}_{p,k}$ such that

1. Γ satisfies the twin property; and
2. if $\alpha \notin F$, the root block $\tilde{\gamma} \in \Gamma$ contains no elements of the form (r, k).

The partitions in $\mathcal{P}_{p,k}^{+(F,\alpha)}$ do not contribute to Equation (16). Hence:

$$\mathbb{E}\left[\left(X^k u \right)_\alpha^p \right] \leq \sum_{\Gamma \in \mathcal{P}_{p,k}^{+(F,\alpha)}} \sum_{\tau \in \mathcal{Z}_{p,k,\alpha}\{\Gamma\}} \omega_u(\tau) \cdot \varphi(\tau).$$

It is necessary for a partition Γ to be an element of $\mathcal{P}_{p,k}^{+(F,\alpha)}$ in order for a τ respecting it to be such that $\varphi(\tau) \neq 0$, however this is not a sufficient condition. Suppose that $\tau_k^{(r)} \notin F$ for some r. Then $u_{\tau_k^{(r)}} = 0$ and hence $\omega_u(\tau) = 0$. Therefore, we can restrict ourselves to considering τ which map (r, k) to F. Define:

$$\mathcal{Z}_{p,k,\alpha}^{+(F)}\{\Gamma\} = \{\tau \in \mathcal{Z}_{p,k,\alpha}\{\Gamma\} : \tau_k^{(r)} \in F \quad \forall r \in [p]\}.$$

Then:

$$\mathbb{E}\left[\left(X^k u \right)_\alpha^p \right] \leq \sum_{\Gamma \in \mathcal{P}_{p,k}^{+(F,\alpha)}} \sum_{\tau \in \mathcal{Z}_{p,k,\alpha}^{+(F)}\{\Gamma\}} \omega_u(\tau) \cdot \varphi(\tau).$$

Fix $\Gamma \in \mathcal{P}_{p,k}$. Suppose that $\varphi(\tau) \le \Phi_\Gamma$ for any $\tau \in \mathcal{Z}^{+(F)}_{p,k,\alpha}\{\Gamma\}$. Furthermore, suppose that $|\mathcal{Z}^{+(F)}_{p,k,\alpha}\{\Gamma\}| \le Z_\Gamma$. Note that $\omega_u(\tau) \in [0,1]$, since it is the product of magnitudes of entries of u and $\|u\|_\infty = 1$. Therefore:

$$\mathbb{E}\left[\left(X^k u\right)^p_\alpha\right] \le \sum_{\Gamma \in \mathcal{P}^{+(F,\alpha)}_{p,k}} Z_\Gamma \cdot \Phi_\Gamma.$$

If $Z_\Gamma \cdot \Phi_\Gamma \le B$ for all $\Gamma \in \mathcal{P}^{+(F,\alpha)}_{p,k}$, then:

$$\begin{aligned} &\le \sum_{\Gamma \in \mathcal{P}^{+(F,\alpha)}_{p,k}} B, \\ &= \left|\mathcal{P}^{+(F,\alpha)}_{p,k}\right| \cdot B. \end{aligned}$$

We can bound the number of partitions loosely using the following lemma:

Lemma 10. $|\mathcal{P}_{p,k}| \le (2pk)^{pk}$.

Proof. Let $\mathcal{P}'_{p,k}$ be the set of all partitions of $\{1,\dots,p\} \times \{1,\dots,k\}$. The number of such partitions is the pk-th Bell number; a well-known bound gives $|\mathcal{P}'_{p,k}| \le (pk)^{pk}$. We generate $\mathcal{P}_{p,k}$ from $\mathcal{P}'_{p,k}$ in the following way: For every $\Gamma \in \mathcal{P}'_{p,k}$, we

1. Create a new block $\tilde{\gamma} = \{(r,0) : r \in \{1,\dots,p\}\}$.
2. For every element (r,ℓ) in $\{1,\dots,p\} \times \{1,\dots,k\}$, make an independent decision about whether to move (r,ℓ) from the block of Γ containing it to the new block $\tilde{\gamma}$. There are 2^{pk} possible ways of deciding which elements to move, and so there are 2^{pk} partitions of $\{1,\dots,p\} \times \{0,\dots,k\}$ generated from Γ.

For each partition $\Gamma \in \mathcal{P}'_{p,k}$ we generate 2^{pk} partitions; in total, we generate $2^{pk} \cdot |\mathcal{P}'_{p,k}| = (2pk)^{pk}$. It is clear that $\mathcal{P}_{p,k}$ is a subset of the generated partitions. Since some of the partitions generated from Γ and a distinct partition Γ' will be identical, $(2pk)^{pk}$ is only an upper-bound on $|\mathcal{P}_{p,k}|$. □

Since $\mathcal{P}^{+(F,\alpha)}_{p,k} \subset \mathcal{P}_{p,k}$, we have $\left|\mathcal{P}^{+(F,\alpha)}_{p,k}\right| \le (2pk)^{pk}$. We have therefore derived the following result:

Lemma 11. *Fix a vector u and let $F = \{i : u_i \neq 0\}$. Fix an index $\alpha \in [n]$. For an index partition $\Gamma \in \mathcal{P}^{+(F,\alpha)}_{p,k}$, suppose that $\varphi(\tau) \le \Phi_\Gamma$ for any $\tau \in \mathcal{Z}^{+(F)}_{p,k,\alpha}\{\Gamma\}$, and that $|\mathcal{Z}^{+(F)}_{p,k,\alpha}\{\Gamma\}| \le Z_\Gamma$. If $Z_\Gamma \cdot \Phi_\Gamma \le B$ for all $\Gamma \in \mathcal{P}^{+(F,\alpha)}_{p,k}$, then:*

$$\mathbb{E}\left[\left(X^k u\right)^p_\alpha\right] \le (2pk)^{pk} \cdot B.$$

We will use this result as a starting point for proving Lemma 6. In the next two parts, we will derive B under different assumptions on the entries of X.

Lemma 12. *Fix a vector u and let $F = \{i : u_i \neq 0\}$. Let $\Gamma \in \mathcal{P}^{+(F,\alpha)}_{p,k}$. Then*

$$\left|\mathcal{Z}^{+(F)}_{p,k,\alpha}\{\Gamma\}\right| \le n^{|\Gamma|-1}.$$

Moreover, let $Q \subset \Gamma$ be the set of blocks in Γ which contain an element of the form (r,k) for some $r \in [p]$. Suppose that $\alpha \notin F$. Then:

$$\left|\mathcal{Z}^{+(F)}_{p,k,\alpha}\{\Gamma\}\right| \le n^{|\Gamma|-|Q|-1} \cdot |F|^{|Q|}.$$

Proof. By definition, $\tau^{(r)}_{\ell} = \tau^{(\tilde{r})}_{\tilde{\ell}}$ if and only if $(r, \ell) \overset{\Gamma}{\sim} (\tilde{r}, \tilde{\ell})$. Hence an indexing function τ respecting Γ takes a distinct value on each $\gamma \in \Gamma$. Exactly one block of the partition contains the pairs of the form $(r, 0)$, and on this block τ must take the value α. On the remaining $|\Gamma - 1|$ blocks τ takes a value in $[n]$. Ignoring the constraint that these values be distinct between blocks to obtain an upper bound, there are $n^{|\Gamma|-1}$ possible choices for the values of τ on these blocks; this gives the desired upper bound.

For the second part, recognize that since $\tau \in \mathcal{Z}^{+(F)}_{p,k,\alpha}\{\Gamma\}$ we have $\tau^{(r)}_{k} \in F$ by assumption. Hence the number of possible values which τ may take on a block in Q is bounded above by $|F|$. Furthermore, it is true that Q does not contain the root block of the partition – this follows from the definition of $\mathcal{P}^{+(F,\alpha)}_{p,k}$ and the assumption that $\alpha \notin F$. The result then follows immediately. □

E.2.2 Proof of Lemma 6

In this part, we will bound $\mathbb{E}\left[\left(X^k u\right)^p_\alpha\right]$ under the assumption that $\mathbb{E}\left[|X_{ij}|^s\right] \le 1/n$ for all $s \ge 2$. As per Lemma 11, it is sufficient to bound $Z_\Gamma \cdot \Phi_\Gamma$ for all partitions Γ satisying the twin property. In the following two lemmas, let $\Gamma \in \mathcal{P}^{+(F,\alpha)}_{p,k}$ and suppose that $\mathbb{E}\left[|X_{ij}|^s\right] \le 1/n$ for all $s \ge 2$.

Lemma 13. *For any $\tau \in \mathcal{Z}^{+(F)}_{p,k,\alpha}\{\Gamma\}$ we have $\varphi(\tau) \le \Phi_\Gamma$, where $\Phi_\Gamma = n^{-|T\{\Gamma\}|}$.*

Proof. As a result of Lemma 8:

$$\varphi(\tau) = \prod_{\rho \in T\{\Gamma\}} \left|\mathbb{E}\left[X^{|\rho|}_\rho\right]\right|.$$

We upper bound this by:

$$\le \prod_{\rho \in T\{\Gamma\}} \mathbb{E}\left[|X_\rho|^{|\rho|}\right].$$

Since Γ satisfies the twin property we have $|\rho| \ge 2$. Then $\mathbb{E}\left[|X_\rho|^{|\rho|}\right] \le 1/n$ by assumption, and so:

$$\le \prod_{\rho \in T\{\Gamma\}} n^{-1},$$
$$= n^{-|T\{\Gamma\}|}.$$

□

Lemma 14. *We have $|\mathcal{Z}^{+(F)}_{p,k,\alpha}\{\Gamma\}| \le Z_\Gamma$, where $Z_\Gamma = n^{|T\{\Gamma\}|}$.*

Proof. From Lemma 12 we have $|\mathcal{Z}^{+(F)}_{p,k,\alpha}\{\Gamma\}| \le n^{|\Gamma|-1}$. We now show that $|\Gamma| - 1 \le |T\{\Gamma\}|$. It is sufficient to find an injection from the set $V \subset \Gamma$ of non-root blocks of Γ to $T\{\Gamma\}$; The existence of an injection proves that $|V| \le |T\{\Gamma\}|$, and since Γ has exactly one root block it follows that $|\Gamma| - 1 \le |T\{\Gamma\}|$. We construct an injection $g : V \to T\{\Gamma\}$ as follows. For any block $\gamma \in \Gamma$, let $\min \gamma$ be the pair $(r^*, \ell^*) \in \gamma$ which is the minimum element with respect to the natural lexicographical order. That is, $(r^*, \ell^*) \in \gamma$ is the pair such that for any other $(r, \ell) \in \gamma$, either $r > r^*$ or it is the case that both $r = r^*$ and $\ell > \ell^*$. The injection g is defined by:

$$g : \gamma \mapsto \text{the equivalence class } \rho \in T\{\Gamma\} \text{ containing } \min \gamma.$$

First note that this is a function since $T\{\Gamma\}$ partitions the set $[p] \times [k]$ such that $g(\gamma)$ is uniquely defined. Next we show that it is indeed an injection. Suppose for a contradiction that γ and γ' are distinct members of Γ and that $g(\gamma) = g(\gamma')$. Let $(r, \ell) = \min \gamma$ and $(r', \ell') = \min \gamma'$, and assume (without loss of generality) that $(r, \ell) < (r', \ell')$ with respect to the lexicographical order on pairs.

The fact that $g(\gamma) = g(\gamma')$ implies that (r,ℓ) and (r',ℓ') are twins. Therefore one of two cases hold: In the first case, $(r,\ell) \overset{\Gamma}{\sim} (r',\ell')$ and $(r,\ell-1) \overset{\Gamma}{\sim} (r',\ell')$. This results in a contradiction, because then $\gamma = \gamma'$; i.e., they are not distinct. In the second case, $(r,\ell) \overset{\Gamma}{\sim} (r',\ell'-1)$ and $(r,\ell-1) \overset{\Gamma}{\sim} (r',\ell')$. In particular, $(r,\ell-1)$ and (r',ℓ') are both in the same block γ' of Γ. Note that $(r',\ell') = \min \gamma'$. But $(r,\ell-1) < (r,\ell) < (r',\ell')$. This is a contradiction. Since both cases lead to contradictions, the assumption cannot hold. Therefore $g(\gamma) \neq g(\gamma')$ when $\gamma \neq \gamma'$, and g is an injection. □

With these results it is easy to prove Lemma 6, restated below:

Lemma 6. *If* $\mathbb{E}\left[|X_{ij}|^s\right] \leq 1/n$ *for all* $s \geq 2$*, then*

$$\mathbb{E}\left[\left(X^k u\right)_\alpha^p\right] \leq (2pk)^{pk}.$$

Proof. Let Z_Γ and Φ_Γ be as defined in Lemma 11. Using the bounds derived in Lemmas 13 and 14, we have for any $\Gamma \in \mathcal{P}_{p,k}^{+(F,\alpha)}$:

$$Z_\Gamma \cdot \Phi_\Gamma \leq n^{|T\{\Gamma\}|} \cdot n^{-|T\{\Gamma\}|} = 1.$$

The result then follows immediately from Lemma 11. □

E.3 Proof of Theorem 9

We will now prove Theorem 9. First, we will state a minor technical lemma which will be used in the proof.

Lemma 15. *Suppose that* $\left|\left(X^k u\right)_\alpha\right| \leq Q^k$ *for all* $k \leq K$*. Let* η *be a positive number, and suppose* $\eta < \min\{Q^{-1}, \|X\|^{-1}\}$*. Then:*

$$\sum_{k\geq 1} \left|\left[(\eta X)^k u\right]_\alpha\right| \leq \frac{\eta Q}{1-\eta Q} + \frac{\|u\|_2 \cdot \|\eta X\|^{K+1}}{1-\|\eta X\|}.$$

Proof. We have:

$$\sum_{k\geq 1} \left|\left[(\eta X)^k u\right]_\alpha\right| = \underbrace{\sum_{k=1}^{K} \left|\left[(\eta X)^k u\right]_\alpha\right|}_{\#1} + \underbrace{\sum_{k>K} \left|\left[(\eta X)^k u\right]_\alpha\right|}_{\#2}.$$

We begin by bounding #1. For each $1 \leq k \leq K$, we have

$$\left|\left[(\eta X)^k u\right]_\alpha\right| = \eta^k \left|\left(X^k u\right)_\alpha\right| \leq (\eta Q)^k.$$

The last step follows from the assumption that $\eta Q < 1$. As a result:

$$\begin{aligned}
\sum_{k=1}^{K} \left|\left[(\eta X)^k u\right]_\alpha\right| &\leq \sum_{k=1}^{K} (\eta Q)^k, \\
&\leq \sum_{k=1}^{\infty} (\eta Q)^k, \\
&= \eta Q \sum_{k=0}^{\infty} (\eta Q)^k, \\
&= \frac{\eta Q}{1-\eta Q}.
\end{aligned}$$

We next bound #2. Here we will use the assumption that $\|\eta X\| < 1$ combined with the fact that the ∞-norm of a vector is bounded above by the 2-norm. We have:

$$\begin{aligned}
\sum_{k>K} \left| \left[(\eta X)^k u \right]_\alpha \right| &\leq \sum_{k>K} \left\| (\eta X)^k u \right\|_\infty, \\
&\leq \sum_{k>K} \left\| (\eta X)^k u \right\|_2, \\
&\leq \sum_{k>K} \left\| (\eta X)^k \right\| \cdot \|u\|_2, \\
&= \sum_{k>K} \|\eta X\|^k \cdot \|u\|_2, \\
&= \|u\|_2 \cdot \|\eta X\|^{K+1} \sum_{k\geq 0} \|\eta X\|^k, \\
&= \frac{\|u\|_2 \cdot \|\eta X\|^{K+1}}{1 - \|\eta X\|}.
\end{aligned}$$

□

We are now able to prove the main result of this section, restated below:

Theorem 9. *Let H be an $n \times n$ symmetric random matrix with independent entries along the diagonal and upper triangle satisfying $\mathbb{E}H_{ij} = 0$. Suppose γ is such that $\mathbb{E}|H_{ij}/\gamma|^p \leq 1/n$ for all $p \geq 2$. Choose $\xi > 1$ and $\kappa \in (0,1)$. Let $\lambda \in \mathbb{R}$ and suppose that $\gamma < \lambda(\log n)^\xi$ and $\lambda > \|H\|$. Fix $u \in \mathbb{R}^n$. Then: with probability $1 - n^{-\frac{1}{4}(\log_b n)^{\xi-1}(\log_b e)^{-\xi}+1}$, where $b = \left(\frac{\kappa+1}{2}\right)^{-1}$.*

$$\left\| \sum_{p\geq 1} \left(\frac{H}{\lambda}\right)^p u \right\|_\infty \leq \frac{\gamma(\log n)^\xi}{\lambda - \gamma(\log n)^\xi} \cdot \|u\|_\infty + \frac{\|H/\lambda\|^{\lfloor \frac{\kappa}{8}(\log n)^\xi + 1 \rfloor}}{1 - \|H/\lambda\|} \cdot \|u\|_2. \tag{5}$$

Proof. We have

$$\begin{aligned}
\zeta_\alpha &= \sum_{p\geq 1} \left| \left[\left(\frac{H}{\lambda}\right)^p u \right]_\alpha \right|, \\
&= \|u\|_\infty \sum_{p\geq 1} \left| \left[\left(\frac{H}{\lambda}\right)^p \cdot \frac{u}{\|u\|_\infty} \right]_\alpha \right|, \\
&= \|u\|_\infty \sum_{p\geq 1} \left| \left[\left(\frac{\gamma}{\lambda} \cdot \frac{H}{\gamma}\right)^p \cdot \frac{u}{\|u\|_\infty} \right]_\alpha \right|,
\end{aligned}$$

Define $X = H/\gamma$, $\eta = \frac{\gamma}{\lambda}$, and $v = u/\|u\|_\infty$. Then:

$$= \|u\|_\infty \sum_{p\geq 1} |[(\eta X)^p v]_\alpha|. \tag{17}$$

Note that $\mathbb{E}|X_{ij}|^p = \mathbb{E}|H_{ij}/\gamma|^p$. Thus for all $p \geq 2$ we have $\mathbb{E}|X_{ij}|^p \leq 1/n$. We may therefore invoke the first result in Theorem 15 to derive, for all $p \leq \frac{\kappa}{8}(\log n)^\xi$,

$$\mathbb{P}\left(|(X^p v)_\alpha|^k \geq (\log n)^{k\xi} \right) \leq 1 - n^{-\frac{1}{4}(\log_\mu n)^{\xi-1}(\log_\mu e)^{-\xi}}. \tag{18}$$

We now bound $\sum_{p\geq 1} |[(\eta X)^p v]_\alpha|$ by applying Lemma 15 with $X = H/\gamma$, $\eta = \gamma/\lambda$, $Q = (\log n)^\xi$ and $K = \left\lfloor \frac{\kappa}{8}(\log n)^\xi \right\rfloor$. One of the requirements of Lemma 15 is that $\eta = \gamma/\lambda$ must satisfy:

$$\frac{\gamma}{\lambda} < \min\left\{Q^{-1},\ \|X\|^{-1}\right\} = \min\left\{(\log n)^{-\xi},\ \gamma\|H\|^{-1}\right\}.$$

Hence we must have $\gamma < \lambda(\log n)^{-\xi}$ and $\lambda > \|H\|$, as assumed. Then, applying the result of Lemma 15, we have:

$$\begin{aligned}
\zeta_\alpha(H,\lambda,u) &= \|u\|_\infty \sum_{p\geq 1} |[(\eta X)^p\, v]_\alpha|, \\
&\leq \|u\|_\infty \left(\frac{\gamma(\log n)^\xi}{\lambda - \gamma(\log n)^\xi} + \frac{\|H/\lambda\|^{\lfloor \frac{\kappa}{8}(\log n)^\xi + 1 \rfloor}}{1 - \|H/\lambda\|} \cdot \frac{\|u\|_2}{\|u\|_\infty} \right), \\
&= \frac{\gamma(\log n)^\xi}{\lambda - \gamma(\log n)^\xi} \cdot \|u\|_\infty + \frac{\|H/\lambda\|^{\lfloor \frac{\kappa}{8}(\log n)^\xi + 1 \rfloor}}{1 - \|H/\lambda\|} \cdot \|u\|_2.
\end{aligned}$$

□