



Published in final edited form as:

*Ann Stat.* 2020 June ; 48(3): 1452–1474. doi:10.1214/19-aos1854.

## ENTRYWISE EIGENVECTOR ANALYSIS OF RANDOM MATRICES WITH LOW EXPECTED RANK

**Emmanuel Abbe,**

PACM and Department of EE, Princeton University, Princeton, NJ 08544, USA

**Jianqing Fan,**

Department of ORFE, Princeton University, Princeton, NJ 08544, USA

**Kaizheng Wang,**

Department of ORFE, Princeton University, Princeton, NJ 08544, USA

**Yiqiao Zhong**

Department of ORFE, Princeton University, Princeton, NJ 08544, USA

### Abstract

Recovering low-rank structures via eigenvector perturbation analysis is a common problem in statistical machine learning, such as in factor analysis, community detection, ranking, matrix completion, among others. While a large variety of bounds are available for average errors between empirical and population statistics of eigenvectors, few results are tight for entrywise analyses, which are critical for a number of problems such as community detection.

This paper investigates entrywise behaviors of eigenvectors for a large class of random matrices whose expectations are low-rank, which helps settle the conjecture in Abbe et al. (2014b) that the spectral algorithm achieves exact recovery in the stochastic block model without any trimming or cleaning steps. The key is a first-order approximation of eigenvectors under the  $\ell_\infty$  norm:

$$u_k \approx \frac{Au_k^*}{\lambda_k^*},$$

where  $\{u_k\}$  and  $\{u_k^*\}$  are eigenvectors of a random matrix  $A$  and its expectation  $\mathbb{E}A$ , respectively.

The fact that the approximation is both tight and linear in  $A$  facilitates sharp comparisons between  $u_k$  and  $u_k^*$ . In particular, it allows for comparing the signs of  $u_k$  and  $u_k^*$  even if  $\|u_k - u_k^*\|_\infty$  is large.

The results are further extended to perturbations of eigenspaces, yielding new  $\ell_\infty$ -type bounds for synchronization ( $\mathbb{Z}_2$ -spiked Wigner model) and noisy matrix completion.

---

eabbe@princeton.edu.

SUPPLEMENTARY MATERIAL

Supplementary A: proofs

(link). We provide detailed proofs of all stated results.

## Keywords

eigenvector perturbation; spectral analysis; synchronization; community detection; matrix completion; low-rank structures; random matrices; Primary 62H25; secondary 60B20; 62H12

## 1. Introduction

Many estimation problems in statistics involve low-rank matrix estimators that are NP-hard to compute, and many of these estimators are solutions to nonconvex programs. This is partly because of the widespread use of maximum likelihood estimation (MLE) which, while enjoying good statistical properties, often poses computational challenges due to nonconvex or discrete constraints inherent in the problems.

Fortunately, computationally efficient algorithms using eigenvectors often afford good performance. The eigenvectors either directly lead to final estimates (Shi and Malik, 2000; Ng et al., 2002), or serve as warm starts followed by further refinements (Keshavan et al., 2010a; Jain et al., 2013; Candès et al., 2015). Such algorithms mostly rely on computation of leading eigenvectors and matrix-vector multiplications, which are easily implemented.

While various heuristics abound, theoretical understanding remains scarce on the entrywise analysis, and on when refinements are needed or can be avoided. In particular, it remains open in various cases to determine whether a vanilla eigenvector-based method without preprocessing steps (e.g., trimming of outliers) or without refinement steps (e.g., cleaning with local improvements) enjoys the same optimality results as the MLE (or SDP) does. A crucial missing step is a sharp entrywise perturbation analysis of eigenvectors. This is partly because the  $\mathcal{L}_\infty$  distance between the eigenvectors of a random matrix and their expected counterparts may not be the correct quantity to look at; errors per entry can be asymmetrically distributed, as we shall see in this paper.

This paper investigates entrywise behaviors of eigenvectors and more generally, eigenspaces, for random matrices with low expected rank using the following approach. Let  $A$  be a random matrix,  $A^* = \mathbb{E}A$ , and  $E = A - A^*$  be the ‘error’ of  $A$ . In many cases,  $A^*$  is a symmetric matrix with low rank determined by the structure of a statistical problem, such as low-rank with blocks in community detection.

Consider for now the case of symmetric  $A$ , and let  $u_k$ , resp.  $u_k^*$ , be the eigenvector corresponding to the  $k$ -th largest eigenvalue of  $A$ , resp.  $A^*$ . Roughly speaking, if  $E$  is moderate, our first-order approximation reads

$$u_k = \frac{Au_k}{\lambda_k} \approx \frac{Au_k^*}{\lambda_k^*} = u_k^* + \frac{Eu_k^*}{\lambda_k^*}.$$

While  $u_k$  is a nonlinear function of  $A$  (or equivalently  $E$ ), the approximation is linear in  $A$ , which greatly facilitates the analysis. Under certain conditions, the maximum entrywise approximation error  $\|u_k - Au_k^*/\lambda_k^*\|_\infty$  can be much smaller than  $\|u_k^*\|_\infty$ , allowing us to study

$u_k$  through  $Au_k^*/\lambda_k^*$ . To obtain such results, a key part in our theory is to characterize concentration properties of  $A$  and structural assumptions on its expectation  $A^*$ .

This perturbation analysis leads to new and sharp theoretical guarantees. In particular, we find that for the exact recovery problem in stochastic block model, the vanilla spectral algorithm (without trimming or cleaning) achieves the information-theoretic limit, and it coincides with the MLE estimator *whenever* the latter succeeds. This settles in particular a conjecture left open in Abbe et al. (2014b, 2016). Therefore, MLE and SDP do not have advantage over the spectral method in terms of exact recovery, if the model is correct. SDP may be preferred in some applications for its robustness and optimality certificates, but that is beyond the scope of this paper.

### 1.1. A sample problem

Let us consider a network model that has received widespread interest in recent years: the stochastic block model (SBM). Suppose that we have a graph with vertex set  $\{1, 2, \dots, n\}$ , and assume for simplicity that  $n$  is even. There is an unknown index set  $J \subseteq \{1, 2, \dots, n\}$  with  $|J| = n/2$  such that the vertex set is partitioned into two groups  $J$  and  $J^c$ . Within groups, there is an edge between each pair of vertices with probability  $p$ , and between groups, there is an edge with probability  $q$ . Let  $x \in \mathbb{R}^n$  be the group membership vector with  $x_i = 1$  if  $i \in J$  and  $x_i = -1$  otherwise. The goal is to recover  $x$  from the observed edges of the graph.

This random-graph-based model was first proposed for social relationship networks (Holland et al., 1983), and many more realistic models have been developed based on the SBM since then. Given its fundamental importance, there are a plurality of papers addressing statistical properties and algorithmic efficiencies; see Abbe (2017) for a survey.

Under the regime  $p = \frac{a \log n}{n}$ ,  $q = \frac{b \log n}{n}$  where  $a > b > 0$  are constants, Abbe et al. (2016) and Mossel et al. (2014) proved that exact recovery is possible if and only if  $\sqrt{a} - \sqrt{b} > \sqrt{2}$ , and that the limit can be achieved by efficient algorithms. They used two-round procedures (with a clean-up phase) to achieve the threshold. Semidefinite relaxations are also known to achieve the threshold (Abbe et al., 2016; Hajek et al., 2016; Agarwal et al., 2015; Bandeira, 2015), as well as spectral methods with local refinements (Abbe and Sandon, 2015; Yun and Proutiere, 2016; Gao et al., 2015). We will discuss more in Sections 1.5 and 3.2.

While existing works tackle exact recovery rather successfully, some fundamental questions remain unsolved: how do the simple statistics—top eigenvectors of the adjacency matrix—behave? Are they informative enough to reveal the group structure under very challenging regimes?

To study these questions, we start with the eigenvectors of  $A^* = \mathbb{E}A$ . By definition,  $A_{ij}$  is a Bernoulli random variable, and  $\mathbb{P}(A_{ij} = 1)$  depends on whether  $i$  and  $j$  are from the same groups. The expectation  $\mathbb{E}A$  must be a block matrix of the following form:

$$\mathbb{E}A = \frac{\log n}{n} \begin{pmatrix} a\mathbf{1}_{\frac{n}{2}} \times \frac{n}{2} & b\mathbf{1}_{\frac{n}{2}} \times \frac{n}{2} \\ b\mathbf{1}_{\frac{n}{2}} \times \frac{n}{2} & a\mathbf{1}_{\frac{n}{2}} \times \frac{n}{2} \end{pmatrix},$$

where  $\mathbf{1}_{m \times m}$  is the  $m \times m$  all-one matrix. Here, for convenience, we represent  $\mathbb{E}A$  as if  $J = \{1, 2, \dots, n/2\}$ . But in general  $J$  is unknown, and there is a permutation of indices  $\{1, \dots, n\}$  in the matrix representation.

From the matrix representation it is clear that  $\mathbb{E}A$  has rank 2, with two nonzero eigenvalues  $\lambda_1^* = \frac{a+b}{2} \log n$  and  $\lambda_2^* = \frac{a-b}{2} \log n$ . Simple calculations give the corresponding (normalized) eigenvectors:  $u_1^* = \frac{1}{\sqrt{n}} \mathbf{1}_n$ , and  $(u_2^*)_i = 1/\sqrt{n}$  if  $i \in J$  and  $(u_2^*)_i = -1/\sqrt{n}$  if  $i \in \mathcal{J}$ . Since  $u_2^*$  perfectly aligns with the group assignment vector  $x$ , we hope to show its counterpart  $u_2$ , i.e., the second eigenvector of  $A$ , also has desirable properties.

The first reassuring fact is that, the top eigenvalues preserve proper ordering: by Weyl's inequality, the deviation of any eigenvalue  $\lambda_i$  ( $i \in [n]$ ) from  $\lambda_i^*$  is bounded by  $\|A - \mathbb{E}A\|_2$ , which is  $O(\sqrt{\log n})$  with high probability; see supplementary materials (Abbe et al., 2018). The Davis-Kahan sin  $\Theta$  theorem asserts that  $u_1$  and  $u_2$  are weakly consistent estimators for  $u_1^*$  and  $u_2^*$  respectively, in the sense that  $|\langle u_k, u_k^* \rangle| \xrightarrow{\mathbb{P}} 1$  for  $k = 1, 2$ . However, this is not helpful for understanding their entrywise behaviors in the uniform sense, which is crucial for exact recovery. Nor can it explain the sharp phase transition phenomenon. This makes entrywise analysis both interesting and challenging.

This problem motivates some simulations about the coordinates of top eigenvectors of  $A$ . In Figure 1, we calculate the rescaled second eigenvector  $\sqrt{n}u_2$  of one typical realization  $A$ , and make a histogram plot of its coordinates. (Note the first eigenvector is aligned with the all-one vector  $\mathbf{1}_n$ , which is uninformative.) The parameters we choose are  $n = 5000$ ,  $a = 4.5$  and  $b = 0.25$ , for which exact recovery is possible with high probability. Visibly, the coordinates of  $\sqrt{n}u_2$  form two clusters around  $\pm 1$  which, marked by red dashed lines, are coordinates of  $\sqrt{n}u_2^*$ . Intuitively, the signs of the former should suffice to reveal the group structure.

To probe into the second eigenvector  $u_2$ , we expand the perturbation  $u_2 - u_2^*$  as follows:

$$u_2 - u_2^* = \left( \frac{Au_2^*}{\lambda_2^*} - u_2^* \right) + \left( u_2 - \frac{Au_2^*}{\lambda_2^*} \right). \quad (1.1)$$

The first term is exactly  $Eu_2^*/\lambda_2^*$ , which is linear in  $E$  and can be viewed as the first-order perturbation. The second term is nonlinear in general, representing the error of higher order. Figure 1 shows boxplots of the infinity norm of rescaled perturbation errors over 100 realizations (see (i)-(iii)), which illustrates that  $\|u_2 - Au_2^*/\lambda_2^*\|_\infty$  is much smaller than  $\|u_2 - u_2^*\|_\infty$  and  $\|Au_2^*/\lambda_2^* - u_2^*\|_\infty$ . Indeed, we will see in Theorem 1.1 that

$$\|u_2 - Au_2^*/\lambda_2^*\|_\infty = o_{\mathbb{P}}\left(\min_i |(u_2^*)_i|\right) = o_{\mathbb{P}}(1/\sqrt{n}). \tag{1.2}$$

The result holds ‘up to sign’, i.e. can choose an appropriate sign for the eigenvector  $u_2$  as it is not uniquely defined; see Theorem 1.1 for its precise meaning. Therefore, the entrywise behavior of  $u_2 - u_2^*$  is captured by its first-order term, which is much more amenable to analysis. This observation will finally lead to sharp eigenvector results in Section 3.2.

We remark that it is also possible to study the top eigenvector (denoted as  $\bar{u}$ ) of the centered adjacency matrix  $\bar{A} = A - \frac{\hat{d}}{n}\mathbf{1}_n\mathbf{1}_n^T$ , where  $\hat{d} = \sum_{i,j} A_{ij}/n$  is the average degree of all vertices. The top eigenvector of  $\mathbb{E}\bar{A}$  is exactly  $u_2^*$ , and its empirical counterpart  $\bar{u}$  is very similar to  $u_2$ . In fact, the same reasoning and analysis applies to  $\bar{u}$ , and one obtains similar plots as Figure 1 (omitted here).

### 1.2. First-order approximation of eigenvectors

Now we present a simpler version of our result that justifies the intuitions above. Consider a general symmetric random matrix (more precisely, this should be a sequence of random matrices with growing dimensions)  $A \in \mathbb{R}^{n \times n}$  with independent entries on and above its diagonal. Suppose its expectation  $A^* = \mathbb{E}A \in \mathbb{R}^{n \times n}$  is low-rank and has  $r$  nonzero eigenvalues. Let us assume that

- (a)  $r = O(1)$ , these  $r$  eigenvalues are positive and in descending order ( $\lambda_1^* \geq \lambda_2^* \geq \dots \geq \lambda_r^* > 0$ ), and  $\lambda_1^* \asymp \lambda_r^*$ .

Their corresponding eigenvectors are denoted by  $u_1^*, \dots, u_r^* \in \mathbb{R}^n$ . In other words, we have spectral decomposition  $A^* = \sum_{j=1}^r \lambda_j^* u_j^* (u_j^*)^T$ .

We fix  $k \in [r]$  and study the  $k$ -th eigenvector  $u_k$ . Define the eigen-gap (or spectral gap) as  $\Delta^* = \min\{\lambda_{k-1}^* - \lambda_k^*, \lambda_k^* - \lambda_{k+1}^*\}$ , where we adopt the convention  $\lambda_0^* = +\infty$  and  $\lambda_{n+1}^* = -\infty$ . Assume that

- (b)  $A$  *concentrates* under the spectral norm, i.e., there is a suitable  $\gamma = \gamma_n = o(1)$  such that  $\|A - A^*\|_2 \leq \gamma \Delta^*$  holds with probability  $1 - o(1)$ .

A direct yet important implication is that, the fluctuation of  $\lambda_k$  is much smaller than the gap  $\Delta^*$ , since Weyl’s inequality forces  $|\lambda_k - \lambda_k^*| \leq \|A - A^*\|_2$ . Thus,  $\lambda_k$  is well separated from other eigenvalues, including the ‘bulk’  $n-r$  eigenvalues whose magnitudes are at most  $\|E\|_2$ .

In addition, we assume that  $A$  *concentrates in a row-wise sense*:

- (c) there exists a continuous non-decreasing function  $\varphi: \mathbb{R}_+ \rightarrow \mathbb{R}_+$  that possibly depends on  $n$ , such that  $\varphi(0) = 0$ ,  $\varphi(x)/x$  is non-increasing, and that for any  $m \in [n]$ ,  $w \in \mathbb{R}^n$ , with probability  $1 - o(n^{-1})$ ,

$$|(A - A^*)_{m \cdot} w| \leq \Delta^* \|w\|_\infty \varphi\left(\frac{\|w\|_2}{\sqrt{n} \|w\|_\infty}\right).$$

Here, the notation  $(A - A^*)_{m \cdot}$  means the  $m$ -th row vector of  $A - A^*$ .

For the Gaussian case where  $A_{ij} \sim N(A_{ij}^*, \sigma^2)$ , we can simply choose a linear function  $\varphi(x) = c(\Delta^*)^{-1} \sigma \sqrt{n \log n} x$  where  $c > 0$  is some proper constant. The condition then reads

$$\mathbb{P}(|(A - A^*)_{m \cdot} w| \leq c \sigma \sqrt{\log n} \|w\|_2) = 1 - o(n^{-1}),$$

which directly follows from Gaussian tail bound since  $(A - A^*)_{m \cdot} w \sim N(0, \sigma^2 \|w\|_2^2)$ . The tail of  $(A - A^*)_{m \cdot} w$  is completely determined by  $\|w\|_2$ . For Bernoulli variables, we will use Bernstein-type inequalities to study  $(A - A^*)_{m \cdot} w$ , which will inevitably involve both  $\|w\|_2$  and  $\|w\|_\infty$ . Hence the function  $\varphi(x)$  can no longer be linear. It turns out that  $\varphi(x) \propto (1 \vee \log(1/x))^{-1}$ , shown in Figure 2, is a suitable choice. More details can be found in Section 2.1 and the supplementary material Abbe et al. (2018). In both cases we have  $\varphi(1) = O(1)$  under suitable signal-to-noise conditions.

**Theorem 1.1**—(Simpler form of Theorem 2.1). *Let  $k \in [r] = \{1, 2, \dots, r\}$  be fixed. Suppose that Assumptions (a), (b) and (c) hold, and  $\|u_k^*\|_\infty \leq \gamma$ . Then, with probability  $1 - \alpha(1)$ ,*

$$\min_{s \in \{\pm 1\}} \|u_k - s A u_k^* / \lambda_k^*\|_\infty = O((\gamma + \varphi(\gamma)) \|u_k^*\|_\infty) = o(\|u_k^*\|_\infty), \quad (1.3)$$

where the notations  $O(\cdot)$  and  $o(\cdot)$  hide dependencies on  $\varphi(1)$ .

On the left-hand side, we are allowed to choose a suitable sign  $s$  as eigenvectors are not uniquely defined. The second bound is a consequence of the first one, since  $\gamma = \alpha(1)$  and  $\lim_{\gamma \rightarrow 0} \varphi(\gamma) = 0$  by continuity. We hide dependency on  $\varphi(1)$  in the above bound, since  $\varphi(1)$  is bounded by a constant under suitable signal-to-noise ratio. More details can be found in Theorem 2.1. Therefore, the approximation error  $\|u_k - A u_k^* / \lambda_k^*\|_\infty$  is much smaller than  $\|u_k^*\|_\infty$ . This rigorously confirms the intuitions in Section 1.1.

Here are some remarks. (1) This theorem enables us to study  $u_k$  via its linearization  $A u_k^* / \lambda_k^*$ , since the approximation error is usually small order-wise. (2) The conditions of the theorem are fairly mild. For SBM, the theorem is applicable as long as we are in the  $\frac{\log n}{n}$  regime ( $p = a \frac{\log n}{n}$  and  $q = b \frac{\log n}{n}$ ), regardless of the relative sizes of  $a$  and  $b$ .

### 1.3. MLE, spectral algorithm, and strong consistency

Once we obtain the approximation result (1.3), the analysis of entrywise behavior of eigenvectors boils down to that of  $A u_k^* / \lambda_k^*$ . In the SBM example, suppose we have (1.2) and with probability  $1 - \alpha(1)$ ,  $\text{sgn}(A u_2^* / \lambda_2^*) = \text{sgn}(u_2^*)$  and all the entries of  $A u_2^* / \lambda_2^*$  are bounded

away from zero by an order of  $1/\sqrt{n}$ . Then  $\text{sgn}(u_2) = \text{sgn}(Au_2^*/\lambda_2^*)$  holds with probability  $1 - o(1)$ . Here  $\text{sgn}(\cdot)$  denotes the entrywise sign function. The eigenvector-based estimator  $\text{sgn}(u_2)$  for block membership can be conveniently analyzed through  $Au_2^*/\lambda_2^*$ , whose entries are just linear combinations of Bernoulli variables.

We remark on a subtlety of our result: our central analysis is a good control of  $\|u_k - Au_k^*/\lambda_k^*\|_\infty$ , not necessarily of  $\|u_k - u_k^*\|_\infty$ . For example, in SBM, an inequality such as  $\|u_2 - u_2^*\|_\infty < \|u_2^*\|_\infty$  is not true in general. In Figure 1, the second boxplot shows that  $\sqrt{n}\|u_2 - u_2^*\|_\infty$  may well exceed 1 even if  $\text{sgn}(u_2) = \text{sgn}(u_2^*)$ . This suggests that the distributions of the coordinates of the two clusters, though well separated, have asymmetric tails. Our Theorem 3.3 asserts that it is in vain to seek a good bound for  $\|u_2 - u_2^*\|_\infty$ . Instead, one should resort to the central quantity  $Au_2^*/\lambda_2^*$ . This may partly explain why the conjecture has remained open for long.

The vector  $Au_k^*/\lambda_k^*$  also plays a pivotal role in the information-theoretic lower bound for exact recovery in SBM, established in Abbe et al. (2016). It is necessary to ensure  $(Au_2^*/\lambda_2^*)_i > 0, \forall i \in J$  to hold with probability at least  $1/3$ . Otherwise, by symmetry and the union bound, with probability at least  $1/3$  we can find some  $i \in J$  and  $i' \in \mathcal{J}$  with  $(Au_2^*/\lambda_2^*)_i < 0$  and  $(Au_2^*/\lambda_2^*)_{i'} > 0$ . Elementary calculation shows that in that case, a swap of group assignments of  $i$  and  $i'$  increases the likelihood. Thus the MLE  $\hat{x}_{MLE}$  fails to exactly recover  $J$ . With a uniform prior on group assignments, the MLE is equivalent to the maximum *a posteriori* estimator, which is optimal for exact recovery. Therefore, we must eliminate such local refinements to make exact recovery possible. This forms the core argument in Abbe et al. (2016). The analysis above suggests an interesting property about the eigenvector-based estimator  $\hat{x}_{eig}(A) = \text{sgn}(u_2)$ :

**Corollary 1.1**—*Suppose we are given  $a > b > 0$  such that  $\sqrt{a} \neq \sqrt{b} + \sqrt{2}$ , i.e., we exclude the regime where  $(a, b)$  is at the boundary of the phase transition. Then, whenever the MLE is successful, in the sense that  $\hat{x}_{MLE} = x$  (up to sign) with probability  $1 - o(1)$ , we have*

$$\hat{x}_{eig}(A) = \hat{x}_{MLE}(A) = x$$

*with probability  $1 - o(1)$ . Here  $x$  is the signed indicator of true communities.*

This is because the success of  $\hat{x}_{MLE}$  hinges on  $\text{sgn}(Au_2^*/\lambda_2^*) = \text{sgn}(u_2^*)$ , which also guarantees  $\hat{x}_{eig}$  to work. See Section 3.2 for details. Such phenomenon appears in two applications considered in this paper.



#### 1.4. An iterative perspective: power iterations

In the SBM, a key observation is that  $\|u_2 - Au_2^*/\lambda_2^*\|_\infty$  is small. Here we give some intuitions from an iterative (or algorithmic) perspective. For simplicity, we will focus on the top eigenvector  $\bar{u}$  of the centered adjacency matrix  $\bar{A} = A - \frac{\hat{d}}{n}\mathbf{1}_n\mathbf{1}_n^T$ .

It is well known that the top eigenvector of a symmetric matrix can be computed via the power method. For almost any possible initialization  $u^0$ , the iterations  $u^{t+1} = \bar{A}u^t/\|\bar{A}u^t\|_2$  converge to  $\bar{u}$ . Suppose we set  $u^0 = u_2^*$ , the top eigenvector of  $\mathbb{E}\bar{A}$ . Although this is not a real algorithm due to the initialization, it helps us gain theoretical insights.

The first, iterate after initialization is  $u^1 = \bar{A}u_2^*/\|\bar{A}u_2^*\|_2$ . Standard concentration inequalities show that  $\|\bar{A}u_2^*\|_2 \approx \bar{\lambda}^*$ , the top eigenvalue of  $\mathbb{E}\bar{A}$ . Therefore,  $u^1$  is approximately  $\bar{A}u_2^*/\bar{\lambda}^*$ , which coincides with our first-order approximation. If  $u^t$  converges to  $\bar{u}$  sufficiently fast,  $u^1$  can already be good enough. This is similar to the rationale of one-step estimator (Bickel, 1975): a single, carefully designed iterate may improve the precision of a good initialization to the desired level. Figure 3 helps illustrate this idea.

The iterative perspective has been explored in recent works (Zhong, 2017; Zhong and Boumal, 2018), where the latter studied both the eigenvector estimator and the MLE of a nonconvex problem. We are not going to show any proof with iterations or induction. Instead, we resort to the Davis-Kahan sin  $\Theta$  theorem, combined with a “leave-one-out” technique. Nevertheless, we believe the iterative perspective is helpful to many other (nonconvex) problems where a counterpart of Davis-Kahan theorem is absent.

#### 1.5. Related works

The study of eigenvector perturbation dates back to Rayleigh (Rayleigh, 1896) and Schrödinger (Schrödinger, 1926), in which asymptotic expansions were obtained. Later, Davis and Kahan (1970) developed elegant nonasymptotic perturbation bounds for eigenspaces gauged by unitary-invariant norms. These were extended to general rectangular matrices in Wedin (1972). See Stewart and Sun (1990) for a comprehensive investigation. Recently, O’Rourke et al. (2018) showed significant improvements of classical, deterministic bounds when the perturbation is random. Norms that depend on the choice of basis, such as the  $\mathcal{L}_\infty$  norm, are not addressed in these works but are of great interest in statistics.

There are several recent papers related to the study of entrywise perturbation. Fan et al. (2016) obtained  $\mathcal{L}_\infty$  eigenvector perturbation bounds. Their results were improved by Cape et al. (2017), in which the authors focused on  $2 \rightarrow \infty$  norm bounds for eigenspaces. Eldridge et al. (2017) developed an  $\mathcal{L}_\infty$  perturbation bound by expanding the eigenvector perturbation into infinite series. These results are deterministic by nature, and thus yield suboptimal bounds under challenging stochastic regimes with small signal-to-noise ratio. By taking advantage of randomness, Koltchinskii et al. (2016) and Koltchinskii and Xia (2016) studied bilinear forms of singular vectors, leading to a sharp bound on  $\mathcal{L}_\infty$  error that was later extended to tensors (Xia and Zhou, 2017). Zhong (2017) characterized entrywise behaviors



of eigenvectors and explored their connections with Rayleigh-Schrödinger perturbation theory. Zhong and Boumal (2018) worked on a related but slightly more complicated problem named “phase synchronization”, and analyzed entrywise behaviors of both the spectral estimator and MLE under a near-optimal regime. Chen et al. (2017) used similar ideas to derive the optimality of both the spectral estimator and MLE in top- $K$  ranking problem.

There is a rich literature on the three applications in this paper. The synchronization problems (Singer, 2011; Cucuringu et al., 2012) aim at estimating unknown signals (usually group elements) from their noisy pairwise measurements, and have attracted much attention in optimization and statistics community recently (Bandeira et al., 2016; Javanmard et al., 2016). They are very relevant models for cryo-EM, robotics (Singer, 2011; Rosen et al., 2016) and more.

The stochastic block model has been studied extensively in the past decades, with renewed activity in the recent years (Coja-Oghlan, 2006; Decelle et al., 2011; Massoulié, 2014; Mossel et al., 2013; Krzakala et al., 2013; Abbe et al., 2016; Guédon and Vershynin, 2016; Amini and Levina, 2014; Abbe and Sandon, 2015; Montanari and Sen, 2016; Bordenave et al., 2015; Abbe and Sandon, 2017; Banks et al., 2016), see Abbe (2017) for further references, and in particular McSherry (2001), Vu (2014), Yun and Proutiere (2014), Lelarge et al. (2015), Chin et al. (2015) and Yun and Proutiere (2016), which are closest to this paper in terms of regimes and algorithms. The matrix completion problems (Candès and Recht, 2009; Candès and Plan, 2010; Keshavan et al., 2010b) have seen great impacts in many areas, and new insights and ideas keep flourishing in recent works (Ge et al., 2016; Sun and Luo, 2016). These lists are only a small fraction of the literature and are far from complete.

We organize our paper as follows: we present our main theorems of eigenvector and eigenspace perturbation in Section 2, which are rigorous statements of the intuitions introduced in Section 1. In Section 3, we apply the theorems to three problems:  $\mathbb{Z}_2$ -synchronization, SBM, and matrix completion from noisy entries. In Section 4, we present simulation results to verify our theories. Finally, we conclude and discuss future works in Section 5.

## 1.6. Notations

We use the notation  $[n]$  to refer to  $\{1, 2, \dots, n\}$  for  $n \in \mathbb{Z}_+$ , and let  $\mathbb{R}_+ = [0, +\infty)$ . For any real numbers  $a, b \in \mathbb{R}$ , we denote  $a \vee b = \max\{a, b\}$  and  $a \wedge b = \min\{a, b\}$ . For nonnegative  $a_n$  and  $b_n$  that depend on  $n$  (e.g., problem size), we write  $a_n \lesssim b_n$  to mean  $a_n \leq Cb_n$  for some constant  $C > 0$ . The notation  $\asymp$  is similar, hiding two constants in upper and lower bounds.

For any vector  $x \in \mathbb{R}^n$ , we define  $\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}$  and  $\|x\|_\infty = \max_i |x_i|$ . For any matrix  $M \in \mathbb{R}^{n \times d}$ ,  $M_i$  refers to its  $i$ -th row, which is a row vector, and  $M_{\cdot j}$  refers to its  $j$ -th column, which is a column vector. The matrix spectral norm is  $\|M\|_2 = \max_{\|x\|_2=1} \|Mx\|_2$ , the matrix max-norm is  $\|M\|_{\max} = \max_{i,j} |M_{ij}|$ , and the matrix  $2 \rightarrow \infty$  norm is

$\|M\|_{2 \rightarrow \infty} = \max_{\|x\|_2=1} \|Mx\|_\infty = \max_i \|M_i\|_2$ . The set of  $n \times r$  matrices with orthonormal columns is denoted by  $\mathcal{O}_{n \times r}$ .

## 2. Main results

### 2.1. Random matrix ensembles

Suppose  $A \in \mathbb{R}^{n \times n}$  is a symmetric random matrix and  $A^* = \mathbb{E}A$ . Denote the eigenvalues of  $A$  by  $\lambda_1 \leq \dots \leq \lambda_n$ , and their associated eigenvectors by  $\{u_j\}_{j=1}^n$ . Analogously for  $A^*$ , the eigenvalues and eigenvectors are  $\lambda_1^* \geq \dots \geq \lambda_n^*$  and  $\{u_j^*\}_{j=1}^n$ , respectively. We also adopt the convention  $\lambda_0 = \lambda_0^* = +\infty$  and  $\lambda_{n+1} = \lambda_{n+1}^* = -\infty$ . We allow some eigenvalues to be identical. Thus, some eigenvectors may be defined up to rotations.

Suppose  $r$  and  $s$  are two integers satisfying  $1 \leq r \leq n$  and  $0 \leq s \leq n - r$ . Let

$$U = (u_{s+1}, \dots, u_{s+r}) \in \mathbb{R}^{n \times r}, U^* = (u_{s+1}^*, \dots, u_{s+r}^*) \in \mathbb{R}^{n \times r} \text{ and}$$

$\Lambda^* = \text{diag}(\lambda_{s+1}^*, \dots, \lambda_{s+r}^*) \in \mathbb{R}^{r \times r}$ . We are interested in the eigenspace  $\text{span}(U)$ . To this end, we assume there is an eigen-gap  $\delta^*$  separating  $\{\lambda_{s+j}^*\}_{j=1}^r$  from 0 and other eigenvalues (see Figure 4), i.e.,

$$\Delta^* = (\lambda_s^* - \lambda_{s+1}^*) \wedge (\lambda_{s+r}^* - \lambda_{s+r+1}^*) \wedge \min_{i \in [r]} |\lambda_{s+i}^*|. \tag{2.1}$$

Compared with the usual eigen-gap (Davis and Kahan, 1970), our definition also takes the distances between eigenvalues and 0 into consideration. When  $A^*$  is rank-deficient, 0 is itself an eigenvalue.

We define  $\kappa := \max_{i \in [r]} |\lambda_{s+i}^*| / \Delta^*$ , which is always bounded from below by 1. In our applications,  $\kappa$  is usually bounded from above by a constant, i.e.,  $\delta^*$  is comparable to  $\{\lambda_{s+j}^*\}_{j=1}^r$  in terms of magnitude.

The concentration property is characterized by a parameter  $\gamma \geq 0$ , and a function  $\varphi(x): \mathbb{R}_+ \rightarrow \mathbb{R}_+$ . Roughly speaking,  $\gamma^{-1}$  resembles the signal-to-noise ratio, and  $\gamma$  typically vanishes as  $n$  tends to infinity.  $\varphi(x)$  is chosen according to the distribution of  $A$ , and is typically bounded by a constant for  $x \in [0, 1]$ . In particular, we take  $\varphi(x) \propto x$  for Gaussian matrices and  $\varphi(x) \propto (1 \vee \log(1/x))^{-1}$  for Bernoulli matrices —see Figure 2. In addition, we will also make a mild structural assumption:  $\|A^*\|_{2 \rightarrow \infty} \leq \gamma \delta^*$ . In many applications involving low-rank structure, the eigenvalues of interest (and thus  $\delta^*$ ) typically scale with  $n$ , whereas  $\|A^*\|_{2 \rightarrow \infty}$  scales with  $\sqrt{n}$ .

Based on the quantities above, we make the following assumptions.

**A1 (Incoherence)**  $\|A^*\|_{2 \rightarrow \infty} \leq \gamma \delta^*$ .

**A2 (Row- and column-wise independence)** For any  $m \in [n]$ , the entries in the  $m$ th row and column of  $A$  are independent with others, i.e.  $\{A_{ij}: i = m \text{ or } j = m\}$  are independent of  $\{A_{ij}: i \neq m, j \neq m\}$ .

**A3 (Spectral norm concentration)**  $\|A\|_{2 \rightarrow \infty} \leq 32 \kappa \max\{\gamma, \varphi(\gamma)\} \delta^*$  and for some  $\delta_0 \in (0, 1)$ ,

$$\mathbb{P}(\|A - A^*\|_2 \leq \gamma \Delta^*) \geq 1 - \delta_0. \tag{2.2}$$

**A4 (Row concentration)** Suppose  $\varphi(x)$  is continuous and non-decreasing in  $\mathbb{R}_+$  with  $\varphi(0) = 0$ ,  $\varphi(x)/x$  is non-increasing in  $\mathbb{R}_+$ , and  $\delta_1 \in (0, 1)$ . For any  $m \in [n]$  and  $W \in \mathbb{R}^{n \times r}$ ,

$$\mathbb{P}\left(\|(A - A^*)_{\cdot m} \cdot W\|_2 \leq \Delta^* \|W\|_2 \rightarrow \infty \varphi\left(\frac{\|W\|_F}{\sqrt{n} \|W\|_2 \rightarrow \infty}\right)\right) \geq 1 - \frac{\delta_1}{n}. \tag{2.3}$$

Here are some remarks and intuitions. Assumption 1 requires that no row of  $A^*$  is dominant. To relate it to the usual concept of incoherence (Candès and Recht, 2009; Candès et al., 2011), we consider the case  $A^* = U^* \Lambda^* (U^*)^T$  and let

$$\mu(U^*) = \frac{n}{r} \max_{i \in [n]} \sum_k (U_{ik}^*)^2 = \frac{n}{r} \|U^*\|_2^2 \rightarrow \infty. \text{ Note that}$$

$$\|U^* \Lambda^* (U^*)^T\|_2 \rightarrow \infty \leq \|U^*\|_2 \rightarrow \infty \|\Lambda^* (U^*)^T\|_2 = \|U^*\|_2 \rightarrow \infty \|\Lambda^*\|_2 \tag{2.4}$$

and  $\kappa = \|\Lambda^*\|_2 / \Delta^*$ . Then Assumption 1 is satisfied as long as  $\mu(U^*) \leq \frac{n\gamma^2}{r\kappa^2}$ , which is very mild.

Assumption 2 is a mild independence assumption, and it encompasses common i.i.d. noise assumptions.

Assumption 3 requires the spectral norm of the noise matrix  $A - A^*$  to be dominated by  $\Delta^*$ , which can be interpreted as signal strength. In our example of  $\mathbb{Z}_2$  synchronization (see Section 3.1), we have  $\Delta^* = n$ , and  $A - A^*$  have i.i.d.  $\mathcal{N}(0, \sigma^2)$  entries above the diagonal. Since  $\|A - A^*\|_2 \lesssim \sigma\sqrt{n}$  by standard concentration results, we need to require  $\sigma = O(\gamma\sqrt{n})$ .

Assumption 4 is a generalization of the row concentration assumption in Section 1.2, and the function  $\varphi$  is problem-dependent. Here we explain the role of  $\varphi$  using a special case where  $r = 1$  and  $A \in \{0, 1\}^{n \times n}$  has i.i.d. Bernoulli entries with parameter  $p = p_n$  on and above its diagonal. Then  $\Delta^* = np$ ,  $\sum_{i=1}^n A_{mi}^* = np$ . If  $p$  is not too small, with high probability we have  $\sum_{i=1}^n A_{mi} \lesssim np$  and thus

$$|(A - A^*)_{\cdot m} \cdot W| \leq \left\| W \right\|_{\infty} \sum_{i=1}^n |(A - A^*)_{mi}| \lesssim \left\| W \right\|_{\infty} np = \Delta^* \left\| W \right\|_{\infty}.$$

If many entries in  $W$  have magnitudes much less than  $\|W\|_{\infty}$ , there should be less fluctuation and better concentration. Indeed, Assumption 4 stipulates a tighter bound by a factor of  $\varphi\left(\frac{\|W\|_2}{\sqrt{n} \|W\|_{\infty}}\right)$ , where  $\frac{\|W\|_2}{\sqrt{n} \|W\|_{\infty}}$  is typically much smaller than 1 in this case. This delicate concentration bound turns out to be crucial in the analysis of SBM, where  $A$  is a sparse binary matrix.

## 2.2. Entrywise perturbation of general eigenspaces

In this section, we generalize Theorem 1.1 from individual eigenvectors to eigenspaces under milder conditions that are characterized by additional parameters. Note that neither  $U$  nor  $U^*$  is uniquely defined, and they can only be determined up to a rotation if the eigenvalues are identical. For this reason, our result has to involve an  $r \times r$  orthogonal matrix. Beyond asserting our result holds up to a suitable rotation, we give an explicit form of such orthogonal matrix.

Let  $H = U^T U^* \in \mathbb{R}^{r \times r}$ , and its singular value decomposition be  $H = \bar{U} \bar{\Sigma} \bar{V}^T$ , where  $\bar{U}, \bar{V} \in \mathbb{R}^{r \times r}$  are orthonormal matrices, and  $\bar{\Sigma} \in \mathbb{R}^{r \times r}$  is a diagonal matrix. Define an orthonormal matrix  $\text{sgn}(H) \in \mathbb{R}^{r \times r}$  as

$$\text{sgn}(H) := \bar{U} \bar{V}^T. \quad (2.5)$$

This orthogonal matrix is called the *matrix sign function* (Gross, 2011). Now we are able to extend the results in Section 1.2 to general eigenspaces.

**Theorem 2.1**—Under Assumptions A1–A4, with probability at least  $1 - \delta_0 - 2\delta_1$  we have

$$\begin{aligned} \|U\|_{2 \rightarrow \infty} &\lesssim (\kappa + \varphi(1)) \|U^*\|_{2 \rightarrow \infty} + \gamma \|A^*\|_{2 \rightarrow \infty} / \Delta^*, \\ \|U \text{sgn}(H) - AU^*(\Lambda^*)^{-1}\|_{2 \rightarrow \infty} &\lesssim \kappa(\kappa + \varphi(1))(\gamma + \varphi(\gamma)) \|U^*\|_{2 \rightarrow \infty} + \gamma \|A^*\|_{2 \rightarrow \infty} / \Delta^*, \\ \|U \text{sgn}(H) - U^*\|_{2 \rightarrow \infty} &\leq \|U \text{sgn}(H) - AU^*(\Lambda^*)^{-1}\|_{2 \rightarrow \infty} + \varphi(1) \|U^*\|_{2 \rightarrow \infty}. \end{aligned}$$

Here the notation  $\lesssim$  only hides absolute constants.

The third inequality is derived by simply writing  $U \text{sgn}(H) - U^*$  as a sum of the first-order error  $EU^*(\Lambda^*)^{-1}$  and higher-order error  $U \text{sgn}(H) - AU^*(\Lambda^*)^{-1}$ , and bounding  $EU^*(\Lambda^*)^{-1}$  by the row concentration Assumption A4. It will be useful for the noisy matrix completion problem. It is worth pointing out that Theorem 2.1 is applicable to any eigenvector of  $A$  that is not necessarily the leading one. This is particularly powerful in SBM (Section 3.2) where we need to analyze the second eigenvector. In addition, we do not need  $A^*$  to have low rank, although the examples to be presented have such structure. For low-rank  $A^*$ , estimation errors of all the eigenvectors can be well controlled by the following corollary of Theorem 2.1.

**Corollary 2.1**—Let Assumptions A1–A4 hold, and suppose that  $A^* = U^* \Lambda^* (U^*)^T$ . With probability at least  $1 - \delta_0 - 2\delta_1$ , we have

$$\begin{aligned} \|U\|_{2 \rightarrow \infty} &\lesssim (\kappa + \varphi(1)) \|U^*\|_{2 \rightarrow \infty}, \\ \|U \text{sgn}(H) - AU^*(\Lambda^*)^{-1}\|_{2 \rightarrow \infty} &\lesssim \kappa(\kappa + \varphi(1))(\gamma + \varphi(\gamma)) \|U^*\|_{2 \rightarrow \infty}, \\ \|U \text{sgn}(H) - U^*\|_{2 \rightarrow \infty} &\leq \|U \text{sgn}(H) - AU^*(\Lambda^*)^{-1}\|_{2 \rightarrow \infty} + \varphi(1) \|U^*\|_{2 \rightarrow \infty}. \end{aligned}$$

Here the notation  $\lesssim$  only hides absolute constants.

Corollary 2.1 directly follows from Theorem 2.1, inequality (2.4) and the fact that  $\kappa \leq 1$ . Below we use a simple example to illustrate the results above. Let  $A^* = \lambda^* u^* (u^*)^T$  be a rank-one matrix with  $\lambda^* > 0$  and  $\|u^*\|_2 = 1$ . Set  $r = 1$  and  $s = 0$ . This structure implies  $\kappa^* = \lambda^*$  and  $\kappa = 1$ . Suppose  $A$  has independent entries on and above the diagonal. Such  $A$  is usually called a *spiked Wigner matrix* in statistics and random matrix theory.

Let Assumptions A1-A4 hold. The first two inequalities in Corollary 2.1 are simplified as

$$\|u\|_\infty \lesssim (1 + \varphi(1)) \|u^*\|_\infty, \tag{2.6}$$

$$\|u - Au^*/\lambda^*\|_\infty \lesssim (\gamma + \varphi(\gamma))(1 + \varphi(1)) \|u^*\|_\infty. \tag{2.7}$$

In many applications,  $\varphi(1) \lesssim 1$  and  $\gamma = \alpha(1)$  as  $n$  goes to infinity. Then (2.6) controls the magnitude of the empirical eigenvector  $u$  by that of the true eigenvector  $u^*$  in the  $\ell^\infty$  sense. Furthermore, (2.7) has the same form as the main result in Theorem 1.1, stating that  $Au^*/\lambda^*$  is an  $\ell^\infty$  approximation of  $u$  with error much smaller than  $\|u^*\|_\infty$ . Therefore, it is possible to study  $u$  via its linearization  $Au^*/\lambda^*$ , which usually makes analysis much easier.

The regularity conditions in Theorem 1.1 imply our Assumptions A1-A4. In particular, the condition  $\|u^*\|_\infty \leq \gamma$  there is equivalent to Assumption A1. As a result, Theorem 1.1 with  $r = 1$  is a special case of Corollary 2.1 and hence of Theorem 2.1. It is not hard to generalize to  $r = \alpha(1)$ .

### 3. Applications

#### 3.1. $\mathbb{Z}_2$ -synchronization and spiked Wigner model

The problem of  $\mathbb{Z}_2$ -synchronization is to recover  $n$  unknown labels  $\pm 1$  from noisy pairwise measurements. This is a prototype of more general  $\text{SO}(d)$ -synchronization problems including phase synchronization and  $\text{SO}(3)$ -synchronization, in which one wishes to estimate the phases of signals or rotations of cameras/molecules, etc. Such problems arise in time synchronization of distributed networks (Giridhar and Kumar, 2006), calibration of cameras (Tron and Vidal, 2009), and cryo-EM (Shkolnisky and Singer, 2012).

Consider an unknown signal  $x \in \{\pm 1\}^n$ . Suppose we have independent measurements of the form  $Y_{ij} = x_i x_j + \sigma W_{ij}$ , where  $i < j$ ,  $W_{ij} \sim \mathcal{N}(0, 1)$  and  $\sigma > 0$ . We can define  $W_{ii} = 0$  and  $W_{ij} = W_{ji}$  for simplicity, and write our model into a matrix form as follows:

$$Y = xx^T + \sigma W, \quad x \in \{\pm 1\}^n. \tag{3.1}$$

This is sometimes called the Gaussian  $\mathbb{Z}_2$ -synchronization problem, in contrast to the one with  $\mathbb{Z}_2$ -noise, also known as the censored block model (Abbe et al., 2014a). This problem can be further generalized: each entry  $x_j$  is a unit-modulus complex number  $e^{i\theta_j}$ , if the goal is to estimate unknown angles from pairwise measurements; or, each entry  $x_j$  is an orthogonal matrix from  $\text{SO}(3)$ , if the goal is to estimate unknown orientations of molecules, cameras, etc. Here we focus on the simplest case  $x_j \in \{\pm 1\}$ .

Note that in (3.1), both  $Y$  and  $W$  are symmetric matrices in  $\mathbb{R}^{n \times n}$ , and the data matrix  $Y$  has a noisy rank-one decomposition. This falls into the spiked Wigner model. The quality of an estimator  $\hat{x}$  is usually gauged either by its correlation with  $x$ , or by the proportion of labels  $x_i$  it correctly recovers. It has been shown that the information-theoretic threshold for a nontrivial correlation is  $\sigma = \sqrt{n}$  (Javanmard et al., 2016; Deshpande et al., 2015; Lelarge and Miolane, 2016; Perry et al., 2016), and the threshold for exact recovery (i.e.,  $\hat{x} = \pm x$  with probability tending to 1) is  $\sigma = \sqrt{\frac{n}{2 \log n}}$  (Bandeira et al., 2016).

When  $\sigma \leq \sqrt{\frac{n}{(2 + \epsilon) \log n}}$  ( $\epsilon > 0$  is any constant), it was proved by Bandeira et al. (2016) that semidefinite programming (SDP) finds the maximum likelihood estimator and achieves exact recovery. We are going to show that a very simple method, both conceptually and computationally, also achieves exact recovery. This method is outlined as follows:

1. Compute the leading eigenvector of  $Y$ , denoted by  $u$ ;
2. Take the estimate  $\hat{x} = \text{sgn}(u)$ .

Our next theorem asserts that the eigenvector-based method above succeeds in finding  $x$  consistently under  $\sigma \leq \sqrt{\frac{n}{(2 + \epsilon) \log n}}$ . Thus, under any regime where the MLE achieves exact recovery, our eigenvector estimator  $\hat{x}$  equals the MLE with high probability. This phenomenon also holds for the stochastic block model.

**Theorem 3.1**—*Suppose  $\sigma \leq \sqrt{\frac{n}{(2 + \epsilon) \log n}}$  for some  $\epsilon > 0$ . With probability  $1 - o(1)$ , the leading eigenvector of  $Y$  with unit  $\ell_2$  norm satisfies*

$$\sqrt{n} \min_{i \in [n]} \{s x_i u_i\} \geq 1 - \sqrt{\frac{2}{2 + \epsilon}} + \frac{C}{\sqrt{\log n}},$$

for a suitable  $s \in \{\pm 1\}$ , where  $C > 0$  is an absolute constant. As a consequence, our eigenvector-based method achieves exact recovery.

Note that our approach does not utilize the structural constraints  $|x_i| = 1, \forall i \in [n]$ ; whereas such constraints appear in the SDP formulation (Bandeira et al., 2016). A natural question is an analysis of both methods with an increased noise level  $\sigma$ . A seminal work by Javanmard et al. (2016) complements our story: the authors showed via non-rigorous statistical mechanics arguments that when  $\sigma$  is on the order of  $\sqrt{n}$ , the SDP-based approach outperforms the eigenvector approach. Nevertheless, with a slightly larger signal strength, there is no such advantage of the SDP approach.

When  $\sigma \asymp \sqrt{n}$ , general results for spiked Wigner models (Baik et al., 2005; Féral and Péché, 2007; Benaych-Georges and Nadakuditi, 2011) imply that  $\frac{1}{n} |u^T x|^2 \rightarrow 1 - \sigma^2/n$  for  $\sigma/\sqrt{n} < 1 - \epsilon$  with any small constant  $\epsilon > 0$ . Deshpande et al. (2015) proved that non-trivial correlation with  $x$  cannot be obtained by any estimator if  $\sigma/\sqrt{n} > 1 + \epsilon$ .

### 3.2. Stochastic Block Model

As is briefly discussed in Section 1, we focus on the symmetric SBM with two equally-sized groups. (Though the second eigenvector of  $A^*$  depends on relative sizes of the groups, our analysis only requires slight modification if groups have different sizes.) For simplicity, we allow for self-loops (i.e. edges from vertices to themselves) in the random graph, and it makes no much difference if they are excluded. In that case, the expectation of the adjacency matrix changes by a negligible quantity  $O(\log n/n)$  under the spectral norm and moreover, Assumptions A1–A4 still hold with the same parameters.

**Definition 3.1**—Let  $n$  be even,  $0 < q < p < 1$ , and  $J \subseteq [n]$  with  $|J| = n/2$ .  $SBM(n, p, q, J)$  is the ensemble of  $n \times n$  symmetric random matrices  $A = (A_{ij})_{i,j \in [n]}$  where  $\{A_{ij}\}_{1 \leq i, j \leq n}$  are independent Bernoulli random variables, and

$$\mathbb{P}(A_{ij} = 1) = \begin{cases} p, & \text{if } i \in J, j \in J \text{ or } i \in J^c, j \in J^c \\ q, & \text{otherwise} \end{cases}. \tag{3.2}$$

The community detection problem aims at finding the bi-partition  $(J, J^c)$  given only one realization of  $A$ . Let  $z_i = 1$  if  $i \in J$  and  $z_i = -1$  otherwise. We want to find an estimator  $\hat{z}$  for the unknown labels  $z \in \{\pm 1\}^n$ . Intuitively, the task is more difficult when  $p$  is close to  $q$ , and when the magnitudes of  $p, q$  are small. It is impossible, for instance, to produce any meaningful estimator when  $p = q$ . The task is also impossible when  $p$  and  $q$  are as small as  $o(n^{-2})$ , since  $A$  is a zero matrix with high probability.

As is already discussed in Section 1, under the regime  $p = \frac{a \log n}{n}, q = \frac{b \log n}{n}$  where  $a$  and  $b$  are constants independent of  $n$ , it is information theoretically impossible to achieve exact recovery (the estimate  $\hat{z}$  equals  $z$  or  $-z$  with probability tending to 1) when  $\sqrt{a} - \sqrt{b} < \sqrt{2}$ . In contrast, when  $\sqrt{a} - \sqrt{b} > \sqrt{2}$ , the goal is efficiently achievable. Further, it is known that SDP succeeds down to the threshold. Under the regime  $p = \frac{a}{n}, q = \frac{b}{n}$ , it is impossible to obtain nontrivial correlation (i.e. the correlation between  $\hat{z}$  and  $z$  is at least some positive constant  $\epsilon$ , as a random guess gets roughly half the signs correct and almost zero correlation with  $z$ ) between any estimator  $\hat{z}$  and  $z$  if  $(a - b)^2 < 2(a + b)$ , and when  $(a - b)^2 > 2(a + b)$ , nontrivial correlation can efficiently be obtained (Massoulié, 2014; Mossel et al., 2013).

Here we focus on the regime where  $p = a \frac{\log n}{n}, q = b \frac{\log n}{n}$  and  $a > b > 0$  are constants. Note that  $\mathbb{E}A$ , or equivalently  $A^*$ , is a rank-2 matrix. Its nonzero eigenvalues are  $\lambda_1^* = (p + q)n/2$  and  $\lambda_2^* = (p - q)n/2$ , whose associated eigenvectors are  $u_1^* = \frac{1}{\sqrt{n}} \mathbf{1}_n$  and  $u_2^* = \frac{1}{\sqrt{n}} \mathbf{1}_J - \frac{1}{\sqrt{n}} \mathbf{1}_{J^c}$ . As  $u_2^*$  is aligned with  $z$  and perfectly reveals the desired partition, the following vanilla spectral method is a natural candidate:

1. Compute  $u_2$ , the eigenvector of  $A$  corresponding to its second largest eigenvalue  $\lambda_2$ ;
2. Set  $\hat{z} = \text{sgn}(u_2)$ .



It has been empirically observed and conjectured that as soon as the signal strength  $\sqrt{a} - \sqrt{b}$  exceed the information threshold  $\sqrt{2}$ , the vanilla spectral method achieves exact recovery (Abbe et al., 2014b). Moreover, in regimes where exact recovery is impossible, Zhang and Zhou (2016) established the following minimax result. It has not been clear whether the vanilla spectral method achieves the minimax misclassification rate.

If we define the misclassification rate as

$$r(\hat{z}, z) = \min_{s \in \{\pm 1\}} n^{-1} \sum_{i=1}^n \mathbf{1}_{\{\hat{z}_i \neq sz_i\}}, \tag{3.3}$$

then the results of Zhang and Zhou (2016) imply that

$$\inf_{\hat{z}} \sup \mathbb{E} r(\hat{z}, z) = \exp\left(- (1 + o(1)) \cdot (\sqrt{a} - \sqrt{b})^2 \frac{\log n}{2}\right), \tag{3.4}$$

where the supremum is taken over approximately equal-sized SBM with 2-blocks. Note that this parameter space is slightly different from our Definition 3.1, but as explained before, we can modify our proofs accordingly such that the same conclusions still hold. See the supplementary materials (Abbe et al., 2018) for further explanation of (3.4).

Here we prove that the vanilla spectral method indeed succeeds in exact recovery whenever it is information-theoretic possible, which resolves the conjecture of (Abbe et al., 2014b); and if it is not, vanilla spectral method achieves the optimal misclassification rate.

**Theorem 3.2**—(i) *If  $\sqrt{a} - \sqrt{b} > \sqrt{2}$  then there exists  $\eta = \eta(a, b) > 0$  and  $s \in \{\pm 1\}$  such that with probability  $1 - o(1)$ ,*

$$\sqrt{n} \min_{i \in [n]} sz_i(u_2)_i \geq \eta.$$

*As a consequence, our spectral method achieves exact recovery.*

(ii) *Let the misclassification rate  $r(\hat{z}, z)$  be defined in (3.3). If  $\sqrt{a} - \sqrt{b} \in (0, \sqrt{2}]$ , then*

$$\mathbb{E} r(\hat{z}, z) \leq n^{-(1 + o(1))(\sqrt{a} - \sqrt{b})^2/2}.$$

*This upper bound matches the minimax lower bound.*

The first part implies that, under the regime where the MLE achieves exact recovery, our eigenvector estimator is exactly the MLE with high probability. This proves Corollary 1.1 in the introduction. Moreover, the second part asserts that for more challenging regime where exact recovery is impossible, the eigenvector estimator has the optimal misclassification rate.

Before further explaining our results, we give a brief review of previous endeavors and an analysis of difficulties. Various papers have investigated this algorithm and its variants such as McSherry (2001), Coja-Oghlan (2006), Rohe et al. (2011), Sussman et al. (2012), Vu

(2014), Lelarge et al. (2015), Yun and Proutiere (2014), Yun and Proutiere (2016), Lei and Rinaldo (2015), Gao et al. (2015), among others. However, it is not known if the simple algorithm above achieves exact recovery down to the information-theoretic threshold, nor the optimal misclassification rate studied in Zhang and Zhou (2016) while below the threshold. An important reason for the unsettlement of this question is that the entrywise behavior of  $u_2$  is not fully understood. In particular, people have been focusing on the  $\ell_\infty$  error  $\|u_2 - u_2^*\|_\infty$ , which may well exceed  $\|u_2^*\|_\infty$  (see Theorem 3.3), suggesting that the algorithm may potentially fail by rounding on the incorrect sign. This is not necessarily the case—as errors could have larger magnitudes on the ‘good side’ of the signal range—but  $\|u_2 - u_2^*\|_\infty$  cannot capture this. To avoid suboptimal theoretical results, multi-round algorithms are popular choices in the literature (Coja-Oghlan, 2006; Vu, 2014), which typically have a preprocessing step of trimming and/or a postprocessing step refining the initial solution. Yun and Proutiere (2014) and Yun and Proutiere (2016) showed that such variants can achieve the exact recovery threshold. We are going to prove that the vanilla spectral algorithm alone achieves the threshold and the minimax lower-bound in one shot.

The key to proving Theorem 3.2 is the following first-order approximation result for  $u_2$  under the  $\ell_\infty$  norm, which is a consequence of Theorem 2.1.

**Corollary 3.1**—*If  $A \sim \text{SBM}\left(n, a\frac{\log n}{n}, b\frac{\log n}{n}, J\right)$ , then with probability  $1 - O(n^{-3})$  we have*

$$\min_{s \in \{\pm 1\}} \|u_2 - sAu_2^*/\lambda_2^*\|_\infty \leq \frac{C}{\sqrt{n \log \log n}}. \tag{3.5}$$

where  $C = C(a, b)$  is some constant depending only on  $a$  and  $b$ .

The above result holds for any constants  $a$  and  $b$ , and does not depend on the gap  $\sqrt{a} - \sqrt{b}$ . This fact will be useful for analyzing the misclassification rate. By Corollary 3.1, the  $\ell_\infty$  approximation error is negligible, and thus the analysis of vanilla spectral algorithm boils down to analyzing the entries in  $Au_2^*/\lambda_2^*$ , which are just weighted sums of Bernoulli random variables.

As a by-product, we can show that entrywise analysis through  $\|u_2 - u_2^*\|_\infty$  is not a good strategy. As is mentioned earlier, our sharp result for the eigenvector estimator stems from careful analysis of the linearized version  $Au_2^*/\lambda_2^*$  of  $u_2$ , and the approximation error  $\|u_2 - Au_2^*/\lambda_2^*\|_\infty$ . This is superior to direct analysis of the  $\ell_\infty$  perturbation  $\|u_2 - u_2^*\|_\infty$ , as the next theorem implies that  $\|u_2 - u_2^*\|_\infty > \|u_2^*\|_\infty$  is possible even if  $\text{sgn}(u_2) = \text{sgn}(u_2^*)$ .

**Theorem 3.3**—(Asymptotic lower bound for eigenvector perturbation). *Let  $J = [n/2]$  and  $A \sim \text{SBM}\left(n, a\frac{\log n}{n}, b\frac{\log n}{n}, J\right)$ , where  $a > b > 0$  are constants and  $n \rightarrow \infty$ . For any fixed  $\eta > 1$  with  $\eta \log \eta - \eta + 1 < 2/a$ , with probability  $1 - o(1)$  we have*

$$\sqrt{n}\|u_2 - u_2^*\|_\infty \geq \frac{a(\eta - 1)}{a - b}.$$

Now let us consider the case in Figure 1, where  $a = 4.5$  and  $b = 0.25$ . On the one hand, exact recovery is achievable since  $\sqrt{a} - \sqrt{b} > 1.62 > \sqrt{2}$ . On the other hand, by taking  $\eta = 2$  we get  $h(\eta) = 2\log 2 - 2 + 1 < 4/9 = 2/a$  and  $\frac{a(\eta-1)}{a-b} > 1.05$ . Theorem 3.3 implies

$$\lim_{n \rightarrow \infty} \mathbb{P}(\|u_2 - u_2^*\|_\infty > 1.05/\sqrt{n}) = 1.$$

In words, the size of fluctuation is consistently larger than the signal strength. As a result, by merely looking at  $\|u_2 - u_2^*\|_\infty$  we cannot expect sharp analysis of the spectral method in exact recovery.

Finally, we point out that it is not straightforward to develop a simple spectral method to achieve the information threshold for exact recovery in SBM with  $K > 2$  blocks. Spectral methods in this scenario (Rohe et al., 2011; Lei and Rinaldo, 2015) typically start with  $r > 1$  eigenvectors  $\{v_j\}_{j=1}^r \subseteq \mathbb{R}^n$  of some data matrix (e.g. the adjacency matrix or Laplacian matrix). Then, the  $n$  rows of  $V = (v_1, \dots, v_r) \in \mathbb{R}^{n \times r}$  are treated as embeddings of the  $n$  nodes into  $\mathbb{R}^r$ , from which one infers block memberships using clustering techniques. In our vanilla spectral method for 2 blocks, we only look at a single eigenvector and return the blocks based on signs of coordinates. This method always returns the same memberships (up to a global swap), even though the eigenvector is identifiable only up to a sign. When  $K > 2$  and  $r > 1$ , due to possible multiplicity of eigenvalues, the embeddings of  $n$  nodes may be identifiable only up to an orthonormal transform in  $\mathbb{R}^r$ . Such ambiguity causes trouble for effective clustering, although we can still study the embedding using Theorem 2.1. Due to space constraints, we put a brief discussion in the supplementary material Abbe et al. (2018).

### 3.3. Matrix completion from noisy entries

Matrix completion based on partial observations has wide applications including collaborative filtering, system identification, global positioning, remote sensing, etc., see Candès and Plan (2010). A popular version is the “Netflix problem”, where one is given an incomplete table of customer ratings and wants to predict the missing entries. This could be useful for targeted recommendation in the future. Since it has been intensively studied in the past decade, our brief review below is by no means exhaustive. Candès and Recht (2009), Candès and Tao (2010), and Gross (2011) focused on exact recovery of low-rank matrices based on noiseless observations. More realistic models with noisy observations were studied in Candès and Plan (2010), Keshavan et al. (2010b), Koltchinskii et al. (2011), Jain et al. (2013) and Chatterjee (2015).

As an application of Theorem 2.1, we are going to study a model similar to the one in Chatterjee (2015) where both sampling scheme and noise are random. It can be viewed as a statistical problem with missing values. Suppose we have an unknown signal matrix  $M^* \in \mathbb{R}^{n_1 \times n_2}$ . For each entry of  $M^*$ , we have a noisy observation  $M_{ij}^* + \varepsilon_{ij}$  with probability  $p$ , and have no observation otherwise. Let  $M^{obs} \in \mathbb{R}^{n_1 \times n_2}$  record our observations, with missing entries treated as zeros. We consider the rescaled partial observation matrix  $M =$

$M^{obs}/p$  for simplicity. It is easy to see that  $M$  is an unbiased estimator for  $M^*$ , and hence a popular starting point for further analysis. The definition of our model is formalized below.

**Definition 3.2**—Let  $M^* \in \mathbb{R}^{n_1 \times n_2}$ ,  $p \in (0, 1)$  and  $\sigma > 0$ . We define  $NMC(M^*, p, \sigma)$  to be the ensemble of  $n_1 \times n_2$  random matrices  $M = (M_{ij})_{i \in [n_1], j \in [n_2]}$  with

$$M_{ij} = (M_{ij}^* + \varepsilon_{ij})I_{ij}/p, \text{ where } \{I_{ij}, \varepsilon_{ij}\}_{i \in [n_1], j \in [n_2]} \text{ are jointly independent,}$$

$$\mathbb{P}(I_{ij} = 1) = p = 1 - \mathbb{P}(I_{ij} = 0) \text{ and } \varepsilon_{ij} \sim N(0, \sigma^2).$$

Let  $r = \text{rank}(M^*)$  and  $M^* = U^* \Sigma^* V^{*T}$  be its singular value decomposition (SVD), where  $U^* \in \mathcal{O}_{n_1 \times r}$ ,  $V^* \in \mathcal{O}_{n_2 \times r}$ ,  $\Sigma^* = \text{diag}(\sigma_1^*, \dots, \sigma_r^*)$  is diagonal, and  $\sigma_1^* \geq \dots \geq \sigma_r^*$ . We are interested in estimating  $U^*$ ,  $V^*$ , and  $M^*$ . The rank  $r$  is assumed to be known, which is usually easily estimated otherwise, see Keshavan and Oh (2009) for example. We work on a very simple spectral algorithm that often serves as an initial estimate of  $M^*$  in iterative methods.

1. Compute the  $r$  largest singular values  $\sigma_1 \dots \sigma_r$  of  $M$ , and their associated left and right singular vectors  $\{u_j\}_{j=1}^r$  and  $\{v_j\}_{j=1}^r$ . Define  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_r)$ ,  $U = (u_1, \dots, u_r) \in \mathcal{O}_{n_1 \times r}$  and  $V = (v_1, \dots, v_r) \in \mathcal{O}_{n_2 \times r}$ .
2. Return  $U$ ,  $V$  and  $U\Sigma V^T$  as estimators for  $U^*$ ,  $V^*$ , and  $M^*$ , respectively.

Note that the matrices in Definition 3.2 are asymmetric in general, due to the rectangular shape and independent sampling. Hence, Theorem 2.1 is not directly applicable. Nevertheless, it could be tailored to fit into our framework by a ‘‘symmetric dilation’’ trick. See the supplementary materials (Abbe et al., 2018) for details. Below we present our results.

**Theorem 3.4**—Let  $M \sim NMC(M^*, p, \sigma)$ ,  $n = n_1 + n_2$ ,  $\kappa = \sigma_1^*/\sigma_r^*$ ,  $H = \frac{1}{2}(U^T U^* + V^T V^*)$ , and  $\eta = (\|U^*\|_2 \rightarrow \infty \vee \|V^*\|_2 \rightarrow \infty)$ . There exist constants  $C$  and  $C'$  such that the followings hold. Suppose  $p \geq 6 \frac{\log n}{n}$  and  $\kappa \frac{n(\|M^*\|_{\max} + \sigma)}{\sigma_r^*} \sqrt{\frac{\log n}{np}} \leq 1/C$ . With probability at least  $1 - C/n$ , we have

$$\begin{aligned} (\|U\|_2 \rightarrow \infty \vee \|V\|_2 \rightarrow \infty) &\leq C' \kappa \eta, \\ (\|U \text{sgn}(H) - U^*\|_2 \rightarrow \infty \vee \|V \text{sgn}(H) - V^*\|_2 \rightarrow \infty) &\leq C' \eta \kappa^2 \frac{n(\|M^*\|_{\max} + \sigma)}{\sigma_r^*} \sqrt{\frac{\log n}{np}}, \\ \|U \Sigma V^T - M^*\|_{\max} &\leq C' \eta^2 \kappa^4 (\|M^*\|_{\max} + \sigma) \sqrt{\frac{n \log n}{p}}. \end{aligned}$$

To our best knowledge, the results for singular vectors are the first of this type for the spectral algorithm. Our bound on  $\|U \Sigma V^T - M^*\|_{\max}$  is a by-product of that, and a similar result was derived by Jain and Netrapalli (2015) using a different approach.

There are two reasons why entrywise type bounds are important. First, in applications such as recommender systems, it is often desirable to have uniform guarantees for all individuals. If we directly use existing  $\ell_2$ -type inequalities to control entrywise errors, the resulting

bounds can be highly sub-optimal in high dimensions. Thus new results are needed. Second, in algorithms based on non-convex optimization (Keshavan et al., 2010b; Sun and Luo, 2016; Jain and Netrapalli, 2015), entrywise bounds are critical for the analysis of initializations and iterations. After the first draft of this paper came out, the entrywise bounds on singular subspaces were applied by Ma et al. (2017) as a guarantee for spectral initialization. The relevance of entrywise bounds goes well beyond matrix completion; see Section 1.5.

For the rest of this subsection, we will illustrate the results in Theorem 3.4 by comparing them with existing ones based on Frobenius norm.

Suppose  $p > c \frac{\log n}{n}$  for some large constant  $c > 0$ . By Theorems 1.1 and 1.3 of Keshavan et al. (2010b), an upper bound for the root-mean squared error (RMSE) gives:

$$\frac{1}{n} \|U\Sigma V^T - M^*\|_F \lesssim (\|M^*\|_{\max} + \sigma) \sqrt{\frac{r}{np}}. \quad (3.6)$$

This implies that the spectral algorithm is rate-optimal when  $\sigma \gtrsim \|M^*\|_{\max}$  as Candès and Plan (2010) established a lower bound  $\frac{1}{n} \|\widehat{M} - M^*\|_F \gtrsim \sigma \sqrt{\frac{r}{np}}$  for any estimator  $\widehat{M}$ . On the other hand, our Theorem 3.4 asserts that

$$\|U\Sigma V^T - M^*\|_{\max} \lesssim_{\kappa, r, \eta} (\|M^*\|_{\max} + \sigma) \sqrt{\frac{\log n}{np}}.$$

where  $\lesssim_{\kappa, r, \eta}$  hides a factor  $O(\kappa, r, \eta \sqrt{n/r})$  that is not large if certain matrix incoherence structure is assumed; see Candès and Recht (2009) for example. Note that our result recovers (3.6) up to a factor of  $\sqrt{\log n}$ , since  $\|X\|_F \leq \sqrt{n_1 n_2} \|X\|_{\max}$  always holds for any  $X$  of size  $n_1 \times n_2$ .

We also compare the estimation errors of singular vectors under the Frobenius norm and the max-norm. On the one hand, the perturbation inequality in Wedin (1972) and spectral norm concentration yield the following.

$$\begin{aligned} \max\{\|U \operatorname{sgn}(H) - U^*\|_F, \|V \operatorname{sgn}(H) - V^*\|_F\} &\lesssim \sqrt{r} \|M - M^*\|_2 / \sigma_r^* \\ &\lesssim \frac{\sqrt{rn/p} (\|M^*\|_{\max} + \sigma)}{\sigma_r^*} \lesssim \frac{n \|M^*\|_{\max}}{\sigma_r^*} \left(1 + \frac{\sigma}{\|M^*\|_{\max}}\right) \sqrt{\frac{r}{np}}. \end{aligned} \quad (3.7)$$

On the other hand, by our entry-wise bound in Theorem 3.4 we have

$$\begin{aligned} &\sqrt{n} \max\{\|U \operatorname{sgn}(H) - U^*\|_{2 \rightarrow \infty}, \|V \operatorname{sgn}(H) - V^*\|_{2 \rightarrow \infty}\} \\ &\lesssim \kappa, r, \eta \frac{n \|M^*\|_{\max}}{\sigma_r^*} \left(1 + \frac{\sigma}{\|M^*\|_{\max}}\right) \sqrt{\frac{r \log n}{np}}. \end{aligned} \quad (3.8)$$

where, as before,  $\lesssim_{\kappa, r, \eta}$  hides a factor  $O(\kappa, r, \eta \sqrt{n/r})$  that is usually not large. Therefore, we also recover (3.7) up to a factor of  $\sqrt{\log n}$ , since  $\|X\|_F \leq \sqrt{nr} \|X\|_{\max}$  holds for any  $X$  of size  $n \times r$ . Note that our goal is to derive good max-norm bounds rather than improving Frobenius-

norm bounds. The comparisons above demonstrate that our bounds have the ‘correct’ order. To a certain extent, our results better portrait the behavior of spectral algorithm and provide more information than their Frobenius counterparts.

## 4. Numerical experiments

### 4.1. $\mathbb{Z}_2$ -synchronization

We present our numerical results for the phase transition phenomenon of  $\mathbb{Z}_2$ -synchronization —see Figure 5. Fix  $q_1 = 500^{1/50}$  and  $q_2 = 2^{1/10}$ . For each  $n$  in the geometric sequence  $\{2, 2q_1, 2q_1^2, \dots, 2q_1^{50}\}$  (rounded to the nearest integers), and each  $\sigma$  in the geometric sequence  $\{q_2^{-32}, q_2^{-31}, \dots, q_2^{50}\}$ , we compare our eigenvector-based estimator  $\hat{x}$  with the unknown signal  $x$ , and report the proportion of success (namely  $\hat{x} = \pm x$ ) out of 100 independent runs in the heat map.

A theoretical curve  $\sigma = \sqrt{\frac{n}{2 \log n}}$  is added onto the heat map. It is clear that below the curve, the eigenvector approach almost always recovers the signal perfectly; and above the curve, it fails to recover the signal.

### 4.2. Stochastic Block Model

Now we present our simulation results for exact recovery and misclassification rates of SBM. The phase transition phenomenon of SBM is exhibited on the left of Figure 6. In this simulation,  $n$  is fixed as 300, and parameters  $a$  ( $y$ -axis) and  $b$  ( $x$ -axis) vary from 0 to 30 and 0 to 10, with increments 0.3 and 0.1 respectively. We compare the labels returned by our eigenvector-based method with the true cluster labels, and report the the proportion of success (namely  $\hat{z} = \pm z$ ) out of 100 independent runs. As before, lighter pixels represent higher chances of success. Two theoretical curves  $\sqrt{a} - \sqrt{b} = \pm \sqrt{2}$  are also added onto the heat map. Clearly, theoretical predictions match numerical results.

The right plot of Figure 6 shows misclassification rates of our eigenvector approach with a fixed parameter  $b$  and a varying parameter  $a$ , where  $a$  is not large enough to reach the exact recovery threshold. We fix  $b = 2$ , and increase  $a$  from 2 to 8 by 0.2 for three different choices of  $n$  from  $\{100, 500, 5000\}$ . Then we calculate the mean misclassification rates  $\mathbb{E}r(\hat{z}, z)$  averaged over 100 independent runs, and plot  $\log \mathbb{E}r(\hat{z}, z) / \log n$  against varying  $b$ . We also add a theoretical curve (with no markers), whose  $y$ -coordinates are  $-(\sqrt{a} - \sqrt{b})^2 / 2$ ; see Theorem 3.2 (ii). It is clear that with  $n$  tending to infinity, the curves of mean misclassification rates move closer to the theoretical one.

### 4.3. Matrix completion from noisy entries

Finally we come to experiments of matrix completion from noisy entries. The performance of the spectral algorithm in terms of root-mean squared error (RMSE) has already been demonstrated in Keshavan et al. (2010b), among others. In this part, we focus on the comparison between the maximum entrywise errors and RMSEs, for both the singular vectors and the matrix itself. The settings are mainly adopted from Candès and Plan (2010)

and Keshavan et al. (2010b). Each time we first create a rank- $r$  matrix  $M^* \in \mathbb{R}^{n \times n}$  using the product  $M_L M_R^T$ , where  $M_L, M_R \in \mathbb{R}^{n \times r}$  have i.i.d.  $N(0, 20/\sqrt{n})$  entries. Then, each entry of  $M^*$  is picked with probability  $p$  and contaminated by random noise drawn from  $N(0, \sigma^2)$ , independently of others. While increasing  $n$  from 500 to 5000 by 500, we choose  $p = \frac{10 \log n}{n}$ , fix  $r = 5$  and  $\sigma = 1$ . All the data presented in the plot are averaged over 100 independent experiments.

In support of our discussions in Section 3.3, Figure 7 shows that the following two ratios

$$R_{\text{mat}} = \frac{\|U\Sigma V^T - M^*\|_{\max}}{\eta^2 \sqrt{\log n} \cdot \|U\Sigma V^T - M^*\|_F},$$

$$R_{\text{vec}} = \frac{\max\{\|U \text{sgn}(H) - U^*\|_2 \rightarrow \infty, \|V \text{sgn}(H) - V^*\|_2 \rightarrow \infty\}}{\eta \sqrt{\log n} \cdot \max\{\|U \text{sgn}(H) - U^*\|_F, \|V \text{sgn}(H) - V^*\|_F\}},$$

approximately remain constant as  $n$  grows. Here the RMSEs  $n^{-1} \|U\Sigma V^T - M^*\|_F$  and  $n^{-1/2} \max\{\|U \text{sgn}(H) - U^*\|_F, \|V \text{sgn}(H) - V^*\|_F\}$  are scaled by  $(\sqrt{n}\eta)^2 \sqrt{\log n}$  and  $(\sqrt{n}\eta) \sqrt{\log n}$ , respectively. Hence our analysis is sharp, and the perturbations are obviously delocalized among the entries.

## 5. Discussions

We have developed first-order approximations for eigenvectors and eigenspaces with small  $\ell_\infty$  errors under random perturbations. These results lead to sharp guarantees for three statistical problems.

Several future directions deserve exploration. First, the main perturbation theorems are currently stated only for symmetric matrices. We think it may be possible to extend the current analysis to SVD of general rectangular matrices, which has broader applications such as principal component analysis. Second, there are many other graph-related matrices beyond adjacency matrices, including graph Laplacians and non-backtracking matrices, which are important both in theory and in practice. Third, we believe our assumption of row- and column-wise independence can be relaxed to block-wise independence, which is relevant to cryo-EM and other problems.

Finally, in our examples, the spectral algorithm is strongly consistent if and only if the MLE is, though the latter can be NP-hard to compute in general. It would be interesting to see how general this phenomenon is, in view of better understanding the statistical and computational tradeoffs.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

The authors thank Harrison Zhou, Amit Singer, Nicolas Boumal, Yuxin Chen and Cong Ma for helpful discussions.



The research was supported by NSF CAREER Award CCF-1552131, ARO grant W911NF-16-1-0051, NSF CSOI CCF-0939370.

The research was supported by NSF grants DMS-1662139 and DMS-1712591, NIH grant R01-GM072611-11 and ONR grant N00014-19-1-2120.

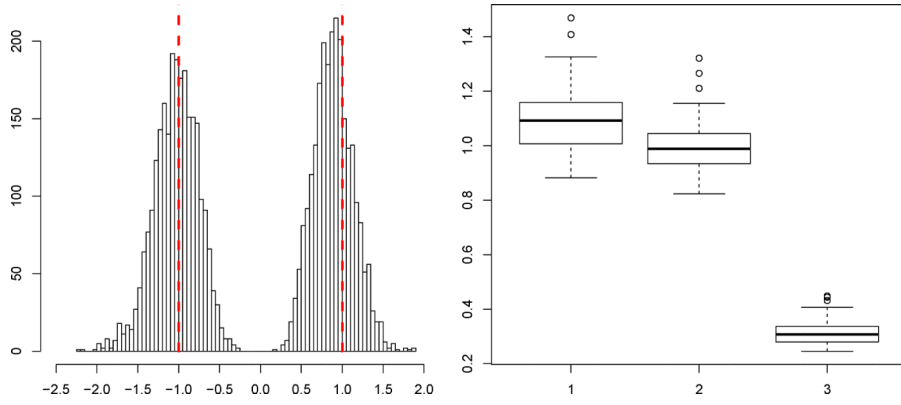
## References

- Abbe E (2017). Community detection and stochastic block models: recent developments. arXiv preprint arXiv:1703.10146.
- Abbe E, Bandeira AS, Bracher A and Singer A (2014a). Decoding binary node labels from censored edge measurements: Phase transition and efficient recovery. *IEEE Transactions on Network Science and Engineering* 1 10–22.
- Abbe E, Bandeira AS and Hall G (2014b). Exact recovery in the stochastic block model. arXiv preprint arXiv:1405.3267.
- Abbe E, Bandeira AS and Hall G (2016). Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory* 62 471–487.
- Abbe E, Fan J, Wang K and Zhong Y (2018). Supplement to “entrywise eigenvector analysis of random matrices with low expected rank”. Submitted to the *Annals of Statistics*.
- Abbe E and Sandon C (2015). Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*. IEEE.
- Abbe E and Sandon C (2017). Proof of the achievability conjectures in the general stochastic block model. *Communications on Pure and Applied Mathematics*.
- Agarwal N, Bandeira AS, Koiliaris K and Kolla A (2015). Multisection in the stochastic block model using semidefinite programming. arXiv preprint arXiv:1507.02323.
- Amini AA and Levina E (2014). On semidefinite relaxations for the block model. arXiv preprint arXiv:1406.5647.
- Baik J, Ben Arous G and Péohé S (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability* 1643–1697.
- Bandeira A, Boumal N and Singer A (2016). Tightness of the maximum likelihood semidefinite relaxation for angular synchronization. *Mathematical Programming* 1–23.
- Bandeira AS (2015). Random Laplacian matrices and convex relaxations. *Foundations of Computational Mathematics* 1–35.
- Banks J, Moore C, Neeman J and Netrapalli P (2016). Information-theoretic thresholds for community detection in sparse networks. In *Conference on Learning Theory*.
- Benaych-Georges F and Nadakuditi RR (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics* 227 494–521.
- Bickel PJ (1975). One-step Huber estimates in the linear model. *Journal of the American Statistical Association* 70 428–434.
- Bordenave C, Lelarge M and Massoulié L (2015). Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*. IEEE.
- Candés EJ, Li X, Ma Y and Wright J (2011). Robust principal component analysis? *Journal of the ACM (JACM)* 58 11.
- Candés EJ, Li X and Soltanolkotabi M (2015). Phase retrieval via Wirtinger flow: Theory and algorithms. *IEEE Transactions on Information Theory* 61 1985–2007.
- Candés EJ and Plan Y (2010). Matrix completion with noise. *Proceedings of the IEEE* 98 925–936.
- Candés EJ and Recht B (2009). Exact matrix completion via convex optimization. *Foundations of Computational Mathematics* 9 717.
- Candés EJ and Tao T (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory* 56 2053–2080.
- Cape J, Tang M and Priebe CE (2017). The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. arXiv preprint arXiv:1705.10735.

- Chatterjee S (2015). Matrix estimation by universal singular value thresholding. *The Annals of Statistics* 43 177–214.
- Chin Y, Fan J, Ma C and Wang K (2017). Spectral method and regularized MLE are both optimal for top-K ranking. arXiv preprint arXiv:1707.09971.
- Chin P, Rao A and Vu V (2015). Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery. In *Conference on Learning Theory*.
- Coja-Oghlan A (2006). A spectral heuristic for bisecting random graphs. *Random, Structures & Algorithms* 29 351–398.
- Cucuringu M, Lipman Y and Singer A (2012). Sensor network localization by eigenvector synchronization over the Euclidean group. *ACM Transactions on Sensor Networks (TOSN)* 8 19.
- Davis C and Kahan WM (1970). The rotation of eigenvectors by a perturbation. III. *SIAM Journal on Numerical Analysis* 7 1–46.
- Decelle A, Krzakala F, Moore C and Zdeborová L (2011). Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications. *Phys. Rev. E* 84 066106.
- Deshpande Y, Abbe E and Montanari A (2015). Asymptotic mutual information for the two-groups stochastic block model. arXiv preprint arXiv:1507.08685.
- Eldridge J, Belkin M and Wang Y (2017). Unperturbed: spectral analysis beyond Davis-Kahan. arXiv preprint arXiv:1706.06516.
- Fan J, Wang W and Zhong Y (2016). An  $\ell_0$  eigenvector perturbation bound and its application to robust covariance estimation. arXiv preprint arXiv:1603.03516.
- Féral D and Péché S (2007). The largest eigenvalue of rank one deformation of large wigner matrices. *Communications in mathematical physics* 272 185–228.
- Gao C, Ma Z, Zhang AY and Zhou HH (2015). Achieving optimal misclassification proportion in stochastic block model. arXiv preprint arXiv:1505.03772.
- Ge R, Lee JD and Ma T (2016). Matrix completion has no spurious local minimum. In *Advances in Neural Information Processing Systems*.
- Giridhar A and Kumar PR (2006). Distributed clock synchronization over wireless networks: Algorithms and analysis. In *Decision and Control, 2006 45th IEEE Conference on*. IEEE.
- Gross D (2011). Recovering low-rank matrices from few coefficients in any basis. *IEEE Transactions on Information Theory* 57 1548–1566.
- Guédon O and Vershynin R (2016). Community detection in sparse networks via Grothendieck’s inequality. *Probability Theory and Related Fields* 165 1025–1049.
- Hajek B, Wu Y and Xu J (2016). Achieving exact cluster recovery threshold via semidefinite programming: Extensions. *IEEE Transactions on Information Theory* 62 5918–5937.
- Holland PW, Laskey KB and Leinhardt S (1983). Stochastic blockmodels: First steps. *Social Networks* 5 109–137.
- Jain P and Netrapalli P (2015). Fast exact matrix completion with finite samples. In *Conference on Learning Theory*.
- Jain P, Netrapalli P and Sanghavi S (2013). Low-rank matrix completion using alternating minimization. In *Proceedings of the Forty-fifth Annual ACM Symposium on Theory of Computing*. ACM.
- Javanmard A, Montanari A and Ricci-Tersenghi F (2016). Phase transitions in semidefinite relaxations. *Proceedings of the National Academy of Sciences* 113 E2218–E2223.
- Keshavan RH, Montanari A and Oh S (2010a). Matrix completion from a few entries. *IEEE Transactions on Information Theory* 56 2980–2998.
- Keshavan RH, Montanari A and Oh S (2010b). Matrix completion from noisy entries. *Journal of Machine Learning Research* 11 2057–2078.
- Keshavan RH and Oh S (2009). A gradient descent algorithm on the Grassman manifold for matrix completion. arXiv preprint arXiv:0910.5260.
- Koltchinskii V, Lounici K and Tsybakov AB (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *The Annals of Statistics* 39 2302–2329.

- Koltchinskii V, Lounici K et al. (2016). Asymptotics and concentration bounds for bilinear forms of spectral projectors of sample covariance. In *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, vol. 52. Institut Henri Poincaré.
- Koltchinskii V and Xia D (2016). Perturbation of linear forms of singular vectors under gaussian noise. In *High Dimensional Probability VII*. Springer, 397–423.
- Krzakala F, Moore C, Mossel E, Neeman J, Sly A, Zdeborová L and Zhang P (2013). Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences* 110 20935–20940.
- Lei J and Rinaldo A (2015). Consistency of spectral clustering in stochastic block models. *The Annals of Statistics* 43 215–237.
- Lelarge M, Massoulié L and Xu J (2015). Reconstruction in the labelled stochastic block model. *IEEE Transactions on Network Science and Engineering* 2 152–163.
- Lelarge M and Miolane L (2016). Fundamental limits of symmetric low-rank matrix estimation. arXiv preprint arXiv:1611.03888.
- Ma C, Wang k., Chi Y and Chen Y (2017). Implicit regularization in nonconvex statistical estimation: Gradient descent converges linearly for phase retrieval, matrix completion and blind deconvolution. arXiv preprint arXiv:1711.10467.
- Massoulié L (2014). Community detection thresholds and the weak ramanujan property. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*. ACM.
- McSherry F (2001). Spectral partitioning of random graphs. In *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*. IEEE
- Montanari A and Sen S (2016). Semidefinite programs on sparse random graphs and their application to community detection. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*. ACM.
- Mossel E, Neeman J and Sly A (2013). A proof of the block model threshold conjecture. arXiv preprint arXiv:1311.4115.
- Mossel E, Neeman J and Sly A (2014). Consistency thresholds for binary symmetric block models. arXiv preprint arXiv:1407.1591.
- Ng AY, Jordan MI and Weiss Y (2002). On spectral clustering: Analysis and an algorithm. In *Advances in Neural Information Processing Systems*.
- O'Rourke S, Vu V and Wang K (2018). Random perturbation of low rank matrices: Improving classical bounds. *Linear Algebra and its Applications* 540 26–59.
- Perry A, Wein AS, Bandeira AS and Moitra A (2016). Optimality and suboptimality of PCA for spiked random matrices and synchronization. arXiv preprint arXiv:1609.05573.
- Rayleigh JWSB (1896). *The theory of sound*, vol. 2. Macmillan.
- Rohe K, Chatterjee S and Yu B (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics* 39 1878–1915.
- Rosen DM, Carlone L, Bandeira AS and Leonard JJ (2016). A certifiably correct algorithm for synchronization over the special Euclidean group. arXiv preprint arXiv:1611.00128.
- Schrödinger E (1926). Quantisierung als eigenwertproblem. *Annalen der Physik* 385 437–490.
- Shi J and Malik J (2000). Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 888–905.
- Shkolnisky Y and Singer A (2012). Viewing direction estimation in cryo-EM using synchronization. *SIAM journal on imaging sciences* 5 1088–1110.
- Singer A (2011). Angular synchronization by eigenvectors and semidefinite programming. *Applied and Computational Harmonic Analysis* 30 20–36. [PubMed: 21179593]
- Stewart G and Sun J (1990). *Matrix perturbation theory*.
- Sun R and Luo Z-Q (2016). Guaranteed matrix completion via non-convex factorization. *IEEE Transactions on Information Theory* 62 6535–6579.
- Sussman DL, Tang M, Fishkind DE and Priebe CE (2012). A consistent adjacency spectral embedding for stochastic blockmodel graphs. *Journal of the American Statistical Association* 107 1119–1128.

- Tron R and Vidal R (2009). Distributed image-based 3-D localization of camera sensor networks. In Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on. IEEE.
- Vu V (2014). A simple SVD algorithm for finding hidden partitions. arXiv preprint arXiv:1404.3918.
- Wedin P-Å (1972). Perturbation bounds in connection with singular value decomposition. BIT Numerical Mathematics 12 99–111.
- Xia D and Zhou F (2017). The  $\ell_0$  perturbation of hosvd and low rank tensor denoising. arXiv preprint arXiv:1707.01207.
- Yun S-Y and Proutiere A (2014). Accurate community detection in the stochastic block model via spectral algorithms. arXiv preprint arXiv:1412.7335.
- Yun S-Y and Proutiere A (2016). Optimal cluster recovery in the labeled stochastic block model. In Advances in Neural Information Processing Systems.
- Zhang AY and Zhou HH (2016). Minimax rates of community detection in stochastic block models. The Annals of Statistics 44 2252–2280.
- Zhong Y (2017). Eigenvector under random perturbation: A nonasymptotic Rayleigh-Schrödinger theory. arXiv preprint arXiv:1702.00139.
- Zhong Y and Boumal N (2018). Near-optimal bounds for phase synchronization. SIAM Journal on Optimization 28 989–1016.



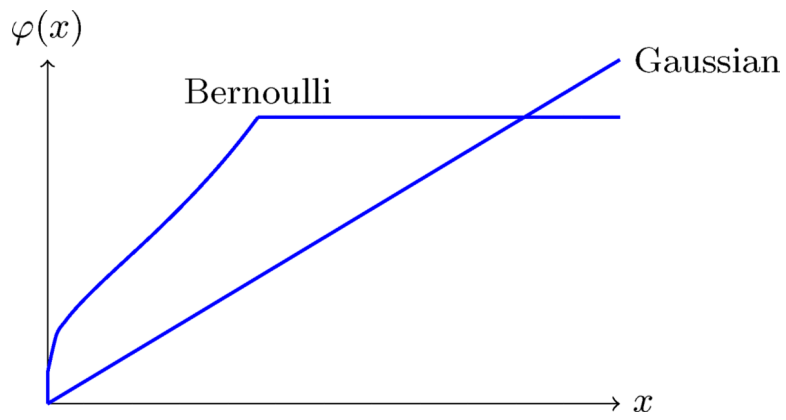
**Fig 1:** The second eigenvector and its first-order approximation in SBM. **Left:** The histogram of coordinates of  $\sqrt{n}u_2$  computed from a single realization of adjacency matrix  $A$ , where  $n$  is 5000,  $a$  is 4.5 and  $b$  is 0.25. Exact recovery is expected as coordinates form two well-separated clusters. **Right:** boxplots showing three different distance/errrors (up to sign) over 100 realizations: (i)  $\sqrt{n}\|u_2 - u_2^*\|_\infty$ , (ii)  $\sqrt{n}\|Au_2^*/\lambda_2^* - u_2^*\|_\infty$ , (iii)  $\sqrt{n}\|u_2 - Au_2^*/\lambda_2^*\|_\infty$ .  $Au_2^*/\lambda_2^*$  is a good approximation of  $u_2$  under  $\ell_\infty$  norm even though  $\|u_2 - u_2^*\|_\infty$  may be large.

Author Manuscript

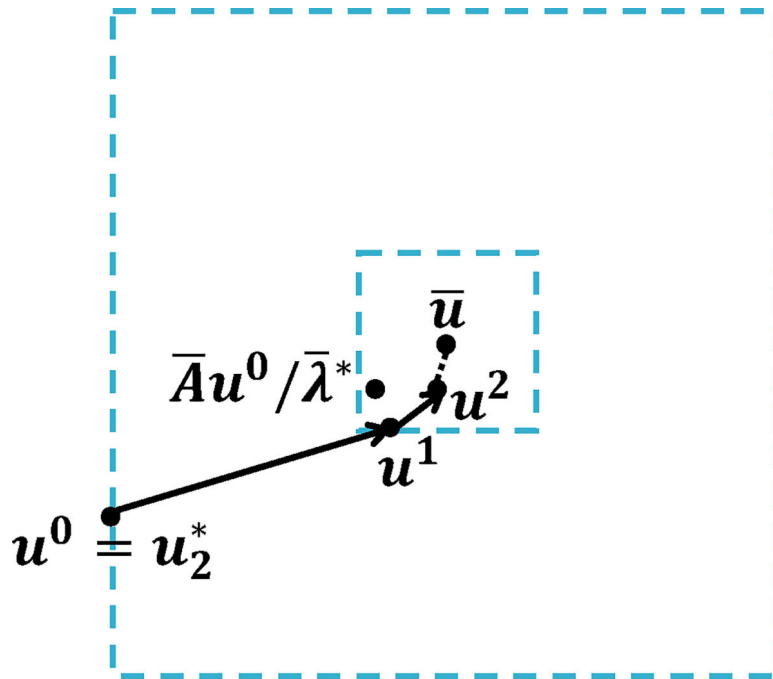
Author Manuscript

Author Manuscript

Author Manuscript

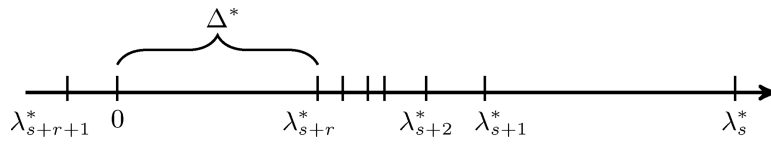


**Fig 2:**  
Typical choices of  $\varphi$  for Gaussian noise and Bernoulli noise.

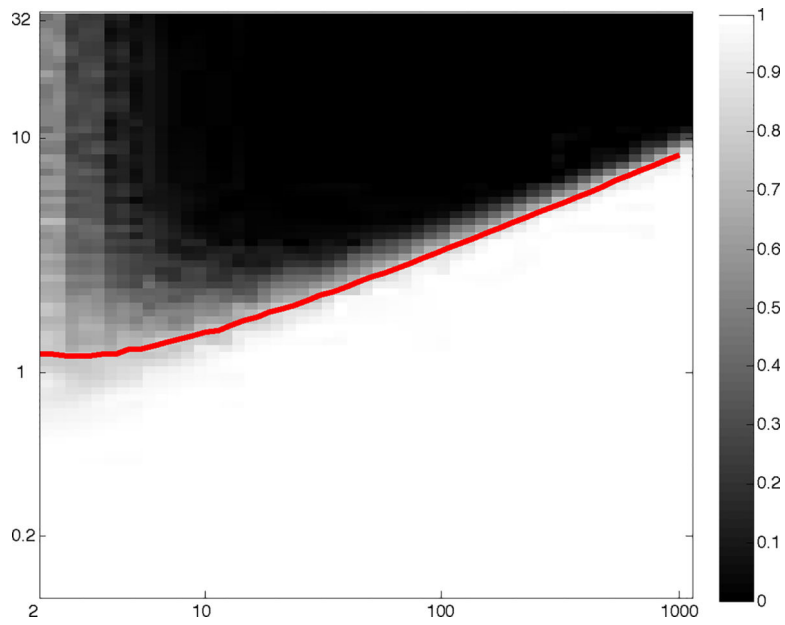


**Fig 3:** Error decay in power iterations. The larger and smaller squares represent  $\ell_\infty$  balls centered at  $\bar{u}$ , with radii  $\|u^0 - \bar{u}\|_\infty$  and  $\|u^1 - \bar{u}\|_\infty$ , respectively.

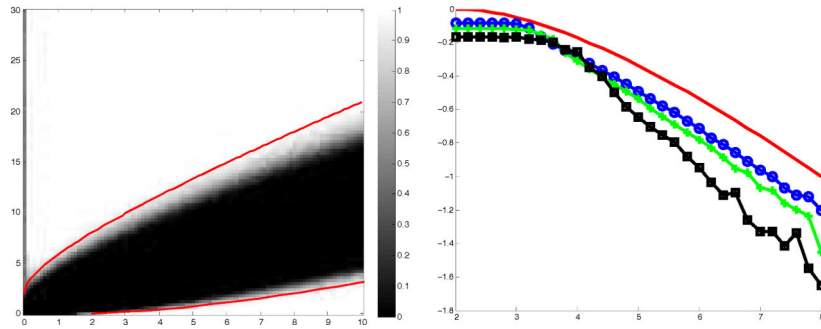




**Fig 4:**  
Eigen-gap \*



**Fig 5:**  
Phase transition of  $\mathbb{Z}_2$ -synchronization: the  $x$ -axis is the dimension  $n$ , and the  $y$ -axis is  $\sigma$ .  
Lighter pixels refer to higher proportions of runs that  $\hat{x}$  recovers  $x$ . The red curve shows the theoretical boundary  $\sigma = \sqrt{\frac{n}{2 \log n}}$ .



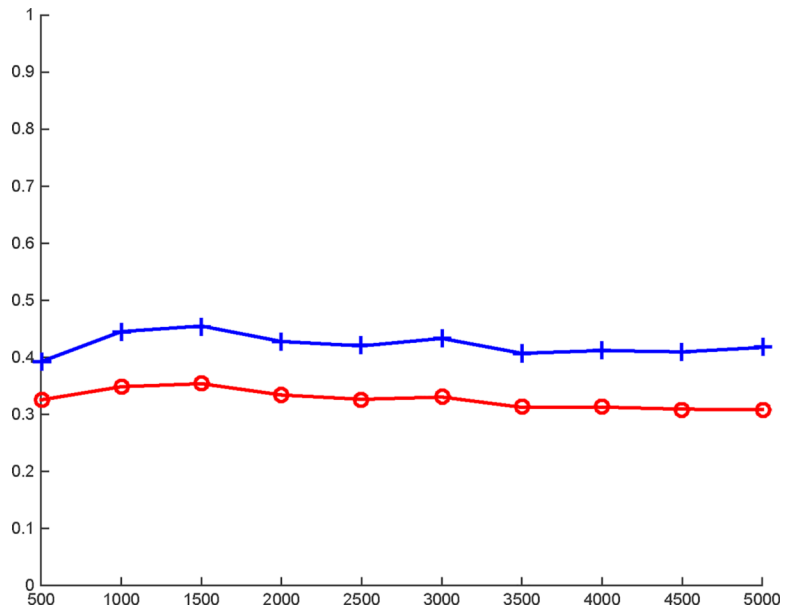
**Fig 6:** Vanilla spectral method for SBM. **Left:** phase transition of exact recovery. The  $x$ -axis is  $b$ , the  $y$ -axis is  $a$ , and lighter pixels represent higher chances of success. Two red curves  $\sqrt{a} - \sqrt{b} = \pm \sqrt{2}$  represent theoretical boundaries for phase transition, matched by numerical results. **Right:** mean misclassification rates on the logarithmic scale with  $b = 2$ . The  $x$ -axis is  $a$ , varying from 2 to 8, and the  $y$ -axis is  $\log \mathbb{E}r(\hat{z}, z) / \log n$ . No marker: theoretical curve; circles:  $n = 5000$ ; crosses:  $n = 500$ ; squares:  $n = 100$ .

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Fig 7:**  $R_{\text{mat}}$  and  $R_{\text{vec}}$  in matrix completion from noisy entries. The  $x$ -axis is  $n$ , varying from 500 to 5000 by 500, and the  $y$ -axis is the ratio. Crosses and circles stand for  $R_{\text{mat}}$  and  $R_{\text{vec}}$ , respectively.