

# Asymptotic Gaussian Fluctuations of Spectral Clustering Eigenvectors

Arun Kadavankandy\*, Romain Couillet\*<sup>†</sup>

\*CentraleSupélec, Université Paris-Saclay, Gif-sur-Yvette, France.

email@centralesupelec.com

<sup>†</sup>GIPSA Lab, Université Grenoble-Alpes, St-Martin-d’Hères, France.

email@gipsa-lab.com

**Abstract**—In this article, we analyze the asymptotic distribution of the eigenvectors used in spectral clustering of random graphs and in kernel spectral clustering of high dimensional Gaussian random vectors. For dense random graphs drawn from the Stochastic Block Model (SBM), we prove that the isolated dominant eigenvectors of the modularity matrix behave asymptotically like Gaussian random vectors with independent components. As opposed to previous works on SBM eigenvectors, we deal with a more challenging and practically meaningful growth rate of the edge probabilities. Similarly for kernel clustering of a two-class Gaussian mixture we prove the asymptotic Gaussianity of the finite-dimensional marginals of the single isolated eigenvector. We present two practical applications of our results: predicting the classification accuracy of clustering algorithms, and speeding up the convergence of the final Expectation Maximization (EM) clustering using an improved initialization.

**Index Terms**—spectral clustering, kernel spectral clustering.

## I. INTRODUCTION

The distributions of eigenvalues and eigenvectors of random matrix ensembles form an important research topic in random matrix theory [2], [3], [15], [17]. In this article, we focus on the dominant eigenvector distributions of the random matrix ensembles that arise in community detection and clustering. Spectral clustering for community partitioning or data grouping has two main steps outlined below [18]:

1. The spectral phase: from the spectral analysis of a suitable kernel or graph matrix, identify and compute the isolated eigenvalues and their corresponding eigenvectors;
2. The clustering phase: Apply a clustering algorithm such as Expectation Maximization (EM) or k-means to the eigenvectors from step 1. to determine the hidden classes [12].

In the analysis of spectral clustering the matrices in the spectral phase take on the form of “spiked random matrices” that have been extensively studied in the random matrix literature [5], [8], [9]. A spiked random matrix is of the form  $\mathbf{H} = \mathbf{W} + \mathbf{P}$ , where  $\mathbf{W}$  can be a Hermitian random matrix with zero mean and i.i.d. upper triangular entries (i.e., a Wigner matrix) or a random covariance matrix of the form  $\mathbf{Z}\mathbf{Z}^T$ , where  $\mathbf{Z}$  is a rectangular matrix with independent zero mean columns [3], [9]. The matrix  $\mathbf{P}$  is a deterministic matrix of finite rank. While the limiting spectral measure of the eigenvalues of  $\mathbf{H}$  is the same as that of  $\mathbf{W}$ , it may have extra isolated eigenvalues depending on the eigenvalues of  $\mathbf{P}$  relative to that of  $\mathbf{W}$  [6].

In [8], [9] the authors develop central limit theorems (CLT) for the dominant eigenvalues of spiked Wigner matrices for general families of component distributions. Deep properties of spiked covariance models are developed and applied to kernel spectral clustering of high dimensional Gaussian vectors in [10], [11] and to principle component analysis in [14]. Limiting distributions of the dominant eigenvalues of spiked covariance models have been studied in [4], [19]. The results in the present article complement these existing results by characterizing the asymptotic distributions of the eigenvector components for the spiked random matrices appearing in graph spectral clustering and kernel spectral clustering.

In [2], [16] the authors proved the asymptotic Gaussianity of the dominant eigenvectors of SBM adjacency matrices. However the growth rate of SBM edge probabilities was chosen so that the eigenvalues of the corresponding  $\mathbf{P}$  matrix grow with the graph size. This specific case is “easier” to analyze and leads to trivial asymptotic clustering. In contrast, in the present paper we analyze a “non-trivial” regime of SBM edge probabilities [1], i.e., the community edge probabilities decay with graph size and  $\mathbf{P}$  ends up having finite-valued eigenvalues. This scenario is more realistic since it leads to a finite asymptotic error rate of classification.

The main contribution of our paper is threefold. Using random matrix techniques we characterize exactly the finite-dimensional dominant eigenvector distributions of the modularity matrix of an SBM graph and the inner-product kernel of a two-class Gaussian mixture model. We then use these distributions to characterize the asymptotic classification errors of community partitioning and clustering. Finally, we show that we can smartly initialize the EM parameters used to cluster the eigenvectors, in order to accelerate the convergence of spectral clustering. These facts are corroborated by suitable simulations.

## II. EIGENVECTOR ANALYSIS FOR SBM GRAPHS

### A. Graph Model

In this section we describe the SBM model, which is an undirected random graph model with inherent community structure. Consider a graph drawn from a  $K$ -community SBM. Let  $\mathbf{A}$  be its adjacency matrix, i.e.,  $A_{ij} \in \{0, 1\}$  equals 1 if there exists an edge between nodes  $i$  and  $j$ . Then the

$A_{ij}, 1 \leq i < j \leq n$ , form a collection of independent Bernoulli random variables such that  $p_{ij} = \mathbb{P}(A_{ij} = 1)$  is given by

$$p_{ij} = q_0 \left( 1 + \frac{M_{C_i C_j}}{\sqrt{n}} \right),$$

where  $q_0 < 1$  is a constant,  $M_{C_i C_j} = \mathcal{O}(1)$ , and  $C_i \in \{1, 2, \dots, K\}$  is the community to which node  $i$  belongs. Thus  $\mathbf{M} \in \mathbb{R}^{K \times K}$  is the matrix of intercommunity edge probabilities. For  $K = 2$  we have:

$$\mathbf{M} = \begin{bmatrix} p_1 & q \\ q & p_2 \end{bmatrix},$$

for some  $p_1, p_2, q > 0$ . Let  $n_a$  be the number of nodes in community  $a, 1 \leq a \leq K$ , and  $c_a = n_a/n = \mathcal{O}(1)$ . Let  $\mathbf{1}_n \in \mathbb{R}^n$  denote the vector of all ones. Without loss of generality, we assume that the nodes of the graph are grouped such that the first  $n_1$  nodes belong to community 1, the next  $n_2$  nodes belong to community 2 and so on. Let  $\mathbf{j}_a \in \mathbb{R}^n$  be the membership vector of community  $a$ , i.e.,  $[j_a]_i = 1$  if  $i \in a$ , and zero otherwise.

In [1] community partitioning using the leading eigenvectors of the modularity matrix

$$\mathbf{B} = \mathbf{A} - \frac{\mathbf{d}\mathbf{d}^T}{\mathbf{d}^T \mathbf{1}},$$

was considered, where  $\mathbf{d} = \mathbf{A}\mathbf{1}$  is the vector of node degrees. Introduced in [13], the modularity matrix defines a widely used community partitioning heuristic [7].

### B. Asymptotic SBM Eigenvector Distribution

It is convenient for the analysis to introduce the notations  $\bar{\mathbf{1}} = \mathbf{1}_n/\sqrt{n}$  and  $\bar{\mathbf{X}} := \mathbf{X}/\sqrt{n}$ , where  $\mathbf{X} = \mathbf{A} - \mathbb{E}\mathbf{A}$ . Under these notations, when  $K \geq 2$  we can express  $\mathbf{B}$  of a  $K$ -community SBM as follows [1]

$$\mathbf{B} = \bar{\mathbf{X}} + \frac{q_0}{n} \mathbf{J} \widetilde{\mathbf{M}} \mathbf{J}^T - \bar{\mathbf{1}} \bar{\mathbf{1}}^T \bar{\mathbf{X}} - \bar{\mathbf{X}} \bar{\mathbf{1}} \bar{\mathbf{1}}^T + \mathbf{\Delta}, \quad (1)$$

where  $\mathbf{\Delta}$  represents a random matrix such that  $\|\mathbf{\Delta}\|_2 = \mathcal{O}(n^{-1/2})$  almost surely.<sup>1</sup> Here  $\mathbf{J} = [\mathbf{j}_1 \ \mathbf{j}_2 \ \dots \ \mathbf{j}_K] \in \mathbb{R}^{n \times K}$  and  $\widetilde{\mathbf{M}} = (\mathbf{I} - \mathbf{1}\mathbf{c}^T)\mathbf{M}(\mathbf{I} - \mathbf{c}\mathbf{1}^T)$ , where  $\mathbf{c} = [c_1 \ c_2 \ \dots \ c_K]^T \in \mathbb{R}^K$ . It can be verified that  $\widetilde{\mathbf{M}}$  has rank  $K - 1$ . Let  $\theta_1 > \theta_2 > \theta_3 \dots > \theta_{K-1}$  be its  $K - 1$  non-zero eigenvalues with eigenvectors  $\bar{\mathbf{r}}_i, 1 \leq i \leq K - 1$ . We define  $\bar{\mathbf{u}}_i = \mathbf{J}\bar{\mathbf{r}}_i$ . The matrix  $\mathbf{B}$  thus has the form of a spiked Wigner matrix as studied for example in [8].

*Theorem 1:* Let  $J \leq K - 1$  be such that  $\theta_J < 2\sigma$  and  $\theta_{J-1} > 2\sigma$ . Further assume that each  $\theta_i$  is different. Let  $\mathbf{v}_i^B$  be the eigenvector corresponding to  $\lambda_i^B$ , the  $i^{\text{th}}$  largest eigenvalue of  $\mathbf{B}$ . Then:

$$\begin{aligned} \lim_{n \rightarrow \infty} \Pr \left( \prod_{1 \leq i \leq J, l \in S} \theta_i \left( \sqrt{n} v_{il}^B - \frac{\sqrt{\theta_i^2 - \sigma^2}}{\theta_i} \sqrt{n} \bar{u}_{il} \right) \geq x_{il} \right) \\ = \prod_{1 \leq i \leq J, l \in S} Q \left( \frac{x_{il}}{\sigma} \right), \end{aligned}$$

<sup>1</sup>For a Hermitian matrix  $\mathbf{A}$ ,  $\|\mathbf{A}\|_2 = \max_i |\lambda_i(\mathbf{A})|$ .

for  $x_{il} \in \mathbb{R}$  and any finite set  $S$ , where  $\sigma^2 = q_0(1 - q_0)$ .

This result tells us that the dominant eigenvector components, corresponding to those deterministic eigenvalues  $\theta_i$  that satisfy a threshold condition, have asymptotically Gaussian distributions around the components of  $\bar{\mathbf{u}}_i$ , which encode the community memberships. Furthermore, their variances are inversely proportional to  $\theta_i$ . Therefore, the larger the various  $\theta_i$ , the more separated the dominant eigenvalues are from the bulk spectral edge (given by  $2\sigma$ ), and the lower the asymptotic variance of the eigenvector components. Also the eigenvectors are independent from each other asymptotically. We shall corroborate this fact in simulations.

Next we present a corollary of the above theorem for  $K = 2$ .

*Corollary 1:* Let  $\mathbf{v}^B$  be the dominant eigenvector corresponding to the largest eigenvalue  $\lambda^B$  of  $\mathbf{B}$ . We assume  $\theta > 2\sigma$ . Then  $\lambda^B$  converges to  $\theta + \frac{\sigma^2}{\theta}$  as  $n \rightarrow \infty$  almost surely. Besides, in the limit, any finite number of components are jointly independent Gaussians with the following entry-wise distributions:

$$\begin{aligned} \theta \left( \sqrt{n} v_i^B - \beta \sqrt{\frac{1 - c_1}{c_1}} \right) &\xrightarrow{n \rightarrow \infty} \mathcal{N}(0, \sigma^2) \quad \text{for } 1 \leq i \leq n_1 \\ \theta \left( \sqrt{n} v_i^B + \beta \sqrt{\frac{c_1}{1 - c_1}} \right) &\xrightarrow{n \rightarrow \infty} \mathcal{N}(0, \sigma^2) \quad \text{for } n_1 + 1 \leq i \leq n, \end{aligned}$$

in distribution, where  $\theta := q_0(p_1 + p_2 - 2q)c_1(1 - c_1)$  and  $\beta = \sqrt{1 - \frac{\sigma^2}{\theta^2}}$ .

The parameter  $\theta$  determines the separability of the communities. It is larger when the edge-probabilities  $p_1, p_2$  within each community are larger than the inter-community edge probability  $q$ . As the asymptotic variance is inversely proportional to  $\theta^2$ , we achieve a smaller classification error when  $\theta$  is larger, i.e., the communities are more separable.

### III. KERNEL SPECTRAL CLUSTERING

We similarly apply the asymptotic analysis of the dominant eigenvector to a linear inner product kernel. For simplicity, we analyze a Gaussian mixture with two classes but the general principle can be straightforwardly extended to multiple classes.

#### A. The Model

Consider  $n$  jointly Gaussian random vectors denoted by  $\mathbf{y}_i \in \mathbb{R}^p, 1 \leq i \leq n$  with covariance matrix given by  $c^2 \mathbf{I}_p$  and  $\mathbf{Y} = [\mathbf{y}_1 \ \mathbf{y}_2 \ \dots \ \mathbf{y}_n] \in \mathbb{R}^{p \times n}$ . We further assume a fraction  $c_1$  of these vectors have mean  $\boldsymbol{\mu}$  and the rest have mean  $-\boldsymbol{\mu}$ . Without loss of generality, the vectors are arranged so that the first  $c_1 n$  of them are in community 1. Let  $\mu = \|\boldsymbol{\mu}\|_2$  by a slight abuse of notation.<sup>2</sup> We consider the high dimensional regime where both  $n$  and  $p$  are large and grow at the same rate, i.e.,  $c = p/n = \mathcal{O}(1)$  as  $n \rightarrow \infty$ . In order to group these points in an unsupervised manner, we apply kernel spectral clustering with the inner product kernel  $\mathbf{K} = \frac{1}{p} \mathbf{Y}^T \mathbf{Y}$ . Let  $\mathbf{w}_i = \mathbf{y}_i - \mathbb{E}\mathbf{y}_i$ . Then we can write

$$\mathbf{K} = \overline{\mathbf{W}}^T \overline{\mathbf{W}} + \overline{\mathbf{W}}^T \boldsymbol{\mu} \boldsymbol{\mu}^T + \mathbf{s} \boldsymbol{\mu}^T \overline{\mathbf{W}} + \|\boldsymbol{\mu}\|_2^2 \mathbf{s} \mathbf{s}^T, \quad (2)$$

<sup>2</sup>For a vector  $\mathbf{x}$ ,  $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1} x_i^2}$ , the 2-norm.

where

$$\overline{\mathbf{W}} = \frac{1}{\sqrt{p}} [\mathbf{w}_1 \quad \mathbf{w}_2 \quad \dots \quad \mathbf{w}_n] \quad \text{and} \quad \mathbf{s} = \frac{1}{\sqrt{p}} (\mathbf{j}_1 - \mathbf{j}_2).$$

We can thus retrieve the classes by merely estimating  $\mathbf{s}$ .

### B. Asymptotic Distribution of Kernel Eigenvector

Let  $\mathbf{v}^K$  be the dominant eigenvector of  $\mathbf{K}$ . In the following theorem, we present the limiting distribution of  $\mathbf{v}^K$  when the eigenvalue of the deterministic component in (2) satisfies a threshold condition.

*Theorem 2:* If  $\mu > \frac{c\zeta}{\sqrt{c}}(1 + \sqrt{c})$ , as  $n \rightarrow \infty$ , any finite number of components of  $\sqrt{n}\mathbf{v}^K$  are asymptotically independent and satisfy:

$$\sqrt{nv_i^K} \xrightarrow{n \rightarrow \infty} \begin{cases} \mathcal{N}\left(+\gamma, \frac{\zeta^2}{\mu^2} \left(c\zeta^2 + \frac{\mu^2}{\mu^2 + \zeta^2}\right)\right) & \text{if } i \leq n_1 \\ \mathcal{N}\left(-\gamma, \frac{\zeta^2}{\mu^2} \left(c\zeta^2 + \frac{\mu^2}{\mu^2 + \zeta^2}\right)\right) & \text{otherwise,} \end{cases} \quad (3)$$

in distribution where  $\gamma = \sqrt{\frac{\mu^2}{\mu^2 + \zeta^2} \left(1 - \frac{c\zeta^4}{\mu^4}\right)}$ .

According to this theorem the asymptotic distributions have the same absolute mean in the two communities, but with different signs. The variance is also the same for all components.

Finally, in order to cluster the points we need to apply EM or k-means as described in Section I, but a single-dimensional version, as opposed to the  $p$ -dimensional EM on the original data. From the asymptotic distribution in Theorem 2, we can see that the price we pay for the reduced complexity is the higher error rate. This can be explained as follows. Roughly the error rate in a Gaussian classification problem is an increasing function of the ratio of the square of mean to the variance. This ratio is  $\frac{\|\mu\|^2}{\zeta^2} = \frac{\mu^2}{\zeta^2}$  in the original  $p$  dimensional space, whereas in the limiting one-dimensional Gaussian problem of

Theorem 2 it equals  $\frac{\mu^2}{\zeta^2} \left(\frac{\frac{\mu^2}{\mu^2 + \zeta^2} \left(1 - \frac{c\zeta^4}{\mu^4}\right)}{\frac{\mu^2}{\mu^2 + \zeta^2} + c\zeta^2}\right)$ , which is less than  $\frac{\mu^2}{\zeta^2}$  for any  $c > 0$  and decreases as  $c$  increases. This is the cost incurred by the dimensionality reduction.

## IV. PRACTICAL APPLICATIONS

In the following sections we describe two practical applications of the eigenvector CLTs previously derived.

### A. Asymptotic Classification Error

In both community partitioning and kernel clustering applications, it is important to determine the asymptotic rate of correct classification. In this section we demonstrate how the eigenvector distributions can be used to obtain the asymptotic classification errors. We restrict our presentation to the two community SBM, but the general principle applies irrespective of the underlying model, as long as the eigenvector distributions converge. As for the two-community SBM, we can achieve community partitioning by thresholding the entries of the scaled dominant eigenvector  $\mathbf{v}^B$ , i.e., we can estimate the community membership vector  $\mathbf{j}_1$  for example, as:

$$\hat{\mathbf{j}}_1 = \mathbb{1}_{\{\sqrt{n}\mathbf{v}^B > \tau\}},$$

where  $\mathbb{1}_A$  is the indicator function or the characteristic function for the set  $A$ , and  $\tau$  is an appropriately chosen threshold.

In order to evaluate the asymptotic error rate of classification we need to evaluate the limits of the empirical distribution functions, for  $x \in \mathbb{R}$ ,

$$F_1(x) = \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbb{1}_{\sqrt{nv_i^B} > x}, \quad F_2(x) = \frac{1}{n_2} \sum_{i=n_1+1}^n \mathbb{1}_{\sqrt{nv_i^B} > x}. \quad (4)$$

For a given threshold  $x$  the average classification error can be computed as

$$p_e(x) = c_1(1 - F_1(x)) + (1 - c_1)F_2(x).$$

We then have the following theorem.

*Theorem 3:* Under the assumptions of Corollary 1, we can show

$$\lim_{n \rightarrow \infty} F_1(x) = Q\left(\frac{1}{\sigma} \left(x - \beta \sqrt{\frac{1 - c_1}{c_1}}\right)\right) \quad (5)$$

$$\lim_{n \rightarrow \infty} F_2(x) = Q\left(\frac{1}{\sigma} \left(x + \beta \sqrt{\frac{c_1}{1 - c_1}}\right)\right). \quad (6)$$

Thus the asymptotic eigenvector distributions allow us to characterize exactly the asymptotic error rate of community partitioning as a function of the graph parameters. For a certain model, these parameters can be roughly estimated, as shown in the next section, and hence one can estimate the asymptotic error rate even without knowing the ground-truth communities.

### B. Smart Initialization of EM

EM is a recursive algorithm used to retrieve the underlying classes of a Gaussian mixture model [12]. It recursively updates the estimates for the per-class means, covariance matrices and class assignments until convergence. The convergence properties of EM depends heavily on the initial values of the estimates [20]. In this section, we show a way to initialize mean and variance estimates of EM using our knowledge of the asymptotic dominant eigenvector distributions. We focus on a two-community SBM, given that the general principle holds true for more complicated cases. To apply Corollary 1, we need estimates for  $q_0$  and  $\theta$ . The former can be estimated from the average graph degree and the latter can be estimated from the dominant eigenvalue (Corollary 1). We obtain

$$\hat{q}_0 = \frac{2}{n(n-1)} \sum_{i < j} A_{ij}; \quad \hat{\theta} = \frac{\hat{\lambda} + \sqrt{\hat{\lambda}^2 - 4\hat{q}_0(1 - \hat{q}_0)}}{2}$$

where  $\hat{\lambda}$  is the observed dominant eigenvalue of  $\mathbf{B}$ . We further take  $c_1 = 1/2$  since the community sizes are unknown. In the simulations, we compare the performance of this algorithm to EM with random initial values.

## V. SIMULATIONS

In a first experiment, we compare the empirical histograms of the components of two dominant eigenvectors of the modularity matrix for an SBM graph with three communities to the theoretical Gaussian pdfs. The graph has 5000 nodes and the

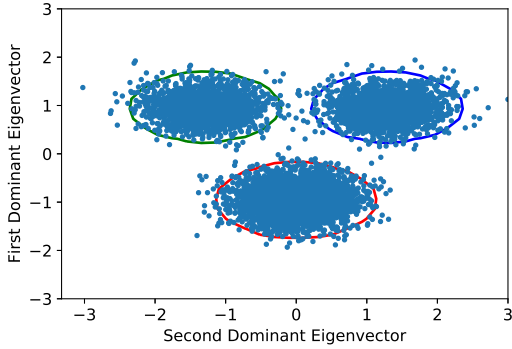


Fig. 1. Comparison of the empirical distributions of dominant eigenvector for 2 community SBM

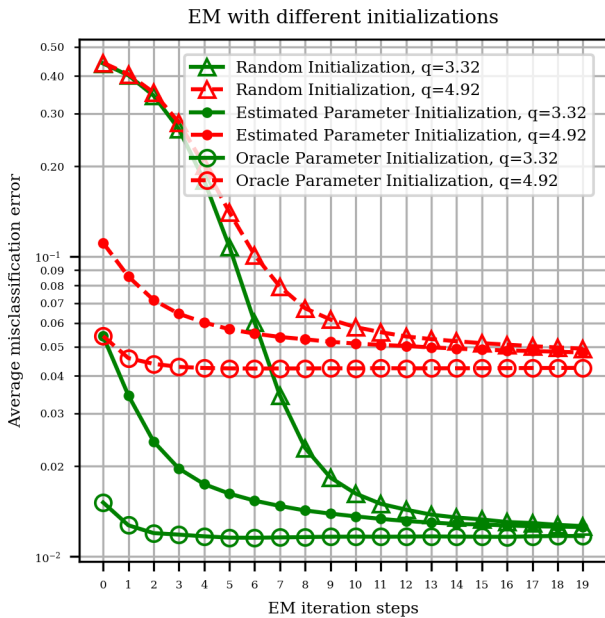


Fig. 2. Comparison of convergence speeds of EM

following parameters:  $p_1 = p_2 = p_3 = 20, q = 0.3, q_0 = 0.2$  and the scaled community sizes are  $c_1 = 1/2, c_2 = c_3 = 1/4$ . In Figure 1 a scatter plot shows the first against the second dominant eigenvectors of  $\mathbf{B}$  along with the level curve at 3 standard deviations of a two dimensional Gaussian distribution with independent components and means and variances as predicted by Theorem 1. The eigenvector components are evidently independent and can be verified to have the predicted means and variances.

Next, we examine the convergence speed of EM with the initialization scheme proposed in Section IV-B. We take a two-community SBM with  $p_1 = p_2 = 12, q_0 = 0.2$  and  $c_1 = 2/3$ . Figure 2 shows EM with the proposed initialization achieves an error an order of magnitude smaller than EM with random initialization quickly, although they both have the same limit-

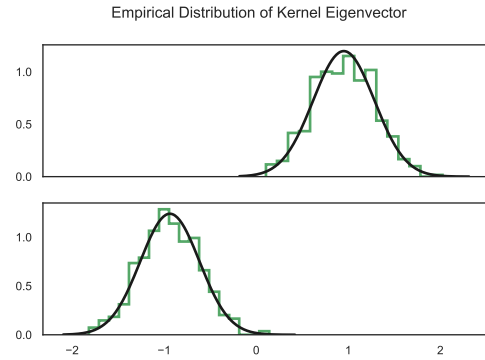


Fig. 3. Comparison of the empirical distributions of the dominant eigenvector of  $\mathbf{K}$ , a component in Community 1 (top); component in community 2 (bottom)

ing error performance. We achieved this improvement without any knowledge of the underlying community structure. The graphs are obtained by averaging over 100 independent runs of the algorithm. Thus we obtain non-trivial acceleration by the proposed initialization of the mean and variance estimates of EM. Finally, we confirm the theoretical distributions of the kernel eigenvector. We consider a simple case of clustering with two clusters of equal size and variance. For  $\sigma^2 = .16$  and a mean vector of norm  $\mu = 1.2$ , we obtain the histograms for the eigenvector in Figure 3.

## VI. CONCLUSIONS

In this article, we have proved the asymptotic Gaussianity of the eigenvectors used in spectral clustering of SBM and kernel spectral clustering of high dimensional Gaussian random vectors. We have also seen that these distributions have practical application aside from their theoretical interest. As a next step it would be interesting to extend this treatment to more complex and realistic graph models and data models such as the Degree-corrected Stochastic Block Model and Gaussian mixtures with a more general covariance structure than dealt with in this article. It would be interesting to investigate how to speed up the clustering for these more realistic models, since for large datasets clustering can be resource intensive. In deriving the results for this paper, we were able to characterize the asymptotic distribution of the eigenvector object, which implicitly depends upon the corresponding random matrix. One could then investigate if other dimensionality reduction algorithms could also be analyzed with similar techniques. We leave these questions open for future works.

## REFERENCES

- [1] H. T. Ali and R. Couillet, "Improved spectral community detection in large heterogeneous networks," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 8344–8392, 2017.
- [2] A. Athreya, V. Lyzinski, D. J. Marchette, C. E. Priebe, D. L. Sussman, and M. Tang, "A central limit theorem for scaled eigenvectors of random dot product graphs," *arXiv preprint arXiv:1305.7388*, 2013.
- [3] Z. Bai, B. Miao, G. Pan *et al.*, "On asymptotics of eigenvectors of large sample covariance matrix," *The Annals of Probability*, vol. 35, no. 4, pp. 1532–1572, 2007.

- [4] Z. Bai and J.-f. Yao, "Central limit theorems for eigenvalues in a spiked population model," in *Annales de l'IHP Probabilités et statistiques*, vol. 44, no. 3, 2008, pp. 447–474.
- [5] F. Benaych-Georges and R. Couillet, "Spectral analysis of the gram matrix of mixture models," *ESAIM: Probability and Statistics*, vol. 20, pp. 217–237, 2016.
- [6] F. Benaych-Georges and R. R. Nadakuditi, "The singular values and vectors of low rank perturbations of large rectangular random matrices," *Journal of Multivariate Analysis*, vol. 111, pp. 120–135, 2012.
- [7] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.
- [8] M. Capitaine, C. Donati-Martin, D. Féral *et al.*, "Central limit theorems for eigenvalues of deformations of wigner matrices," in *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, vol. 48, no. 1. Institut Henri Poincaré, 2012, pp. 107–133.
- [9] M. Capitaine, C. Donati-Martin, D. Féral, M. Février *et al.*, "Free convolution with a semicircular distribution and eigenvalues of spiked deformations of wigner matrices," *Electron. J. Probab.*, vol. 16, no. 64, pp. 1750–1792, 2011.
- [10] R. Couillet, F. Benaych-Georges *et al.*, "Kernel spectral clustering of large dimensional data," *Electronic Journal of Statistics*, vol. 10, no. 1, pp. 1393–1454, 2016.
- [11] A. Kammoun and R. Couillet, "Subspace kernel spectral clustering of large dimensional data," *submitted to Annals of Applied Probability*, 2017.
- [12] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal processing magazine*, vol. 13, no. 6, pp. 47–60, 1996.
- [13] M. E. Newman, "Modularity and community structure in networks," *Proceedings of the national academy of sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.
- [14] D. Passemier, Z. Li, and J. Yao, "On estimation of the noise variance in high dimensional probabilistic principal component analysis," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 79, no. 1, pp. 51–67, 2017.
- [15] J. W. Silverstein, "On the eigenvectors of large dimensional sample covariance matrices," *Journal of multivariate analysis*, vol. 30, no. 1, pp. 1–16, 1989.
- [16] M. Tang, J. Cape, and C. E. Priebe, "Asymptotically efficient estimators for stochastic blockmodels: the naive mle, the rank-constrained mle, and the spectral," *arXiv preprint arXiv:1710.10936*, 2017.
- [17] T. Tao and V. Vu, "Random matrices: Universal properties of eigenvectors," *Random Matrices: Theory and Applications*, vol. 1, no. 01, p. 1150001, 2012.
- [18] U. Von Luxburg, "A tutorial on spectral clustering," *Statistics and computing*, vol. 17, no. 4, pp. 395–416, 2007.
- [19] Q. Wang, Z. Su, and J. Yao, "Joint clt for several random sesquilinear forms with applications to large-dimensional spiked population models," *arXiv preprint arXiv:1402.6064*, 2014.
- [20] L. Xu and M. I. Jordan, "On convergence properties of the EM algorithm for gaussian mixtures," *Neural computation*, vol. 8, no. 1, pp. 129–151, 1996.