Reading Group Notes

Joshua Agterberg

June 30, 2020

1 Notation

I use much of the same notation as in Mao et al. (2019), but I will not bold matrices. I will also have a few different eigen decompositions.

We have

- $P = \rho \Theta B \Theta^\top$, where $\Theta \in [0,1]^{n\times K}$ and B is full rank.
- P has eigendecomposition $P = U\Lambda U^\top$.
- V are the unique rows of the matrix such that $P = \Theta V$ (see their Lemma 2.3). This is $K \times K$ matrix whose rows correspond to pure nodes. If we had a stochastic blockmodel, each row of U would be a row of V.
- Everything else is the same.

2 Notes on Mao et al. (2019)

This paper is interesting as it is one of the first to tackle entrywise eigenvector deviation not for the sake of eigenvector deviation, but for a real statistical problem of estimating memberships. The results are very nice, but it seems to me that they depend more on the eigenstructure of the problem than the randomness inherent in the problem as in other entrywise works. Their main eigenvector deviation bound is presented in Theorem 3.1.

Theorem 1 (Theorem 3.1)**.** *If $n\rho_n = \Omega(\log(n))$, $\lambda_K(\Theta^\top\Theta) \geq \rho^{-1}$ and $|\lambda_{\min}(P)| \geq 4\sqrt{n\rho}\log(n)^\xi$ for some $\xi > 1$ and B is full rank there is at least one pure node from each community, then*

$$||\hat{U} - UU^\top\hat{U}||_{2,\infty} = O\left(\frac{\log^\xi(n)\psi(P)\sqrt{Kn}}{\sqrt{\rho}|\lambda_{\min}(B)|\lambda_K(\Theta^\top\Theta)^{1.5}}\right),$$

with probability at least $1 - O(Kn^{-2})$.

Their version of the theorem hides the log factor. Perhaps the weirdest thing is the appearance of this $\psi(P)$ term, which measures "how well the eigenvalues can be packed into bins." I think the idea behind $\psi(P)$ is that it allows for eigenvalues of vastly different order (instead of just the smallest eigenvalue). Therefore, this work explicitly studies the dependence on the eigenvalues of the P matrix in a way that a number of these sorts of bounds do not. This term allows to have a few eigenvalues of order $\log(n)$, a few eigenvalues of order $\sqrt{n}$ and a few eigenvalues of order n.

The rest of their terms are more or less standard, though they do have additional dependence on the eigenstructure of the problem. For example, we expect both a $\sqrt{\rho}$ and $\lambda_{\min}(B)$ in the denominator and a $\sqrt{K}$ in the numerator. The dependence on n is made a bit clearer by noting that $\lambda_K(\Theta^\top\Theta) = n_{\min}$ when we have a stochastic blockmodel, and when $n_{\min} \asymp n$, the ratio $\frac{n^{1/2}}{\lambda_K(\Theta^\top\Theta)^{1.5}} \asymp \frac{1}{n}$. However, their results allow finer control over the mixing terms $\Theta^\top\Theta$.

In Fan et al. (2019b), the authors assume that $\lambda_K(\Theta^\top\Theta) \geq c_0 n$ regardless of the sparsity factor, which is a stronger assumption than $\lambda_K(\Theta^\top\Theta) \geq \rho^{-1}$. For example, when $\rho \equiv 1$, and $\lambda_{\min}(B) = \Theta(1)$, then we can get consistency as long as $\lambda_K(\Theta^\top\Theta) \geq cn^{1/3+\varepsilon}$, which is much weaker. Note, however, that the results in Fan et al. (2019a) concern distributional asymptotics and hence may require more specific assumptions.

Note that I wrote up the version of the theorem that considers the difference from $\hat{U}$ to its projection onto U; i.e. $\hat{U} - UU^\top\hat{U}$. Recall from last week or the week before that $||U^\top\hat{U} - W_*|| = O\left(\frac{||A-P||}{\lambda_{\min}(P)}\right)^2 = O((n\rho_n)^{-1})$ which is negligible with respect to the obtained bound. In general projections are a bit easier to handle than orthogonal matrices, and $U^\top\hat{U}$ is tending to an orthogonal matrix anyways by the above argument.

The authors do not include a comparison to Cape et al. (2019a), but they do mention in the introduction that Cape et al. (2019a) requires the K to grow slower than poly-log in n. Furthermore, Cape et al. (2019a) has results for a more general model than the one in this paper, so they cannot take advantage of the eigenstructure. However, their results read that

$$||\hat{U} - UW_*||_{2,\infty} \leq O\left(\frac{K^{1/2}\log^{\xi}(n)}{\sqrt{\rho n}}||U||_{2,\infty}\right),$$

which is of the same order when $\psi(P) = \Theta(1)$, and $\lambda_{\min}(B)$ is held constant. But these results don't quite work in the setting that $\lambda_{min}(B) \to 0$ slower than the mixing coefficients, so it's not exactly the same.

Finally, their analysis requires a bit of complex analysis which can be, one might say, "complex." I think that these bounds can be obtained in other ways, but they might be equally as difficult to trace through, and the only real complex analysis you need is the residue theorem. The analysis in Abbe et al. (2017) does not require any contour integrals, but it is still involved with its leave-one-out analysis, which is a bit more of a probabilistic argument (as are similar types of analyses, e.g. Cai et al. (2019); Abbe et al. (2020)). In Lei (2019), he points out (very deep into the appendix) that the analysis Mao et al. (2019) is slightly suboptimal since they do a bound analogous to

$$\left|\int_a^b f(x)dx\right| \leq (b-a)\sup_{x\in(a,b)}|f(x)|.$$

I suspect that to get these sorts of bounds, one must either do some complex analysis or use it "under the hood," such as in the series expansions in Cape et al. (2019a); Eldridge et al. (2018); Tang (2018); Tang et al. (2017). If you go the proof for the series expansion in Bhatia, you just end up having to do complex analysis anyways. I think the reason Lei (2019) obtains very tight bounds is that he does both complex analysis and leave one out analysis.

3 Discussion Points

- How do these results compare to those in Cape et al. (2019b)?
- Why might studying $UU^\top\hat{U}$ be easier than studying UW_* where $W_* = sgn(U^\top\hat{U})$?
- What do we think about the complex analytic techniques? Do we like them? Are they too much work for not much gain?
- These results demonstrate (to me) how eigenstructure plays a role in the analysis – how might other matrix analytic structure help or hurt us? What about if we have less or even no matrix analytic structure?
- Why do we assume B is full-rank? Should we?

References

E. Abbe, J. Fan, K. Wang, and Y. Zhong. Entrywise Eigenvector Analysis of Random Matrices with Low Expected Rank. 2017. arXiv: 1709.09565.

E. Abbe, J. Fan, and K. Wang. An \$\ell_p\$ theory of PCA and spectral clustering. *arXiv:2006.14062 [cs, math, stat]*, June 2020. URL http://arxiv.org/abs/2006.14062. arXiv: 2006.14062.

C. Cai, G. Li, Y. Chi, H. V. Poor, and Y. Chen. Subspace Estimation from Unbalanced and Incomplete Data Matrices: \$\ell_{2,\infty}\$ Statistical Guarantees. *arXiv:1910.04267 [cs, math, stat]*, Oct. 2019. URL http://arxiv.org/abs/1910.04267. arXiv: 1910.04267.

J. Cape, M. Tang, and C. E. Priebe. Signal-plus-noise matrix models: eigenvector deviations and fluctuations. *Biometrika*, 106(1):243–250, Mar. 2019a. ISSN 0006-3444. doi: 10.1093/biomet/asy070. URL https://academic.oup.com/biomet/article/106/1/243/5280315.

J. Cape, M. Tang, and C. E. Priebe. The two-to-infinity norm and singular subspace geometry with applications to high-dimensional statistics. *The Annals of Statistics*, 47(5):2405–2439, Oct. 2019b. ISSN 0090-5364, 2168-8966. doi: 10.1214/18-AOS1752. URL https://projecteuclid.org/euclid.aos/1564797852.

J. Eldridge, M. Belkin, and Y. Wang. Unperturbed: spectral analysis beyond Davis-Kahan. In *Algorithmic Learning Theory*, pages 321–358, Apr. 2018. URL http://proceedings.mlr.press/v83/eldridge18a.html.

J. Fan, Y. Fan, X. Han, and J. Lv. Asymptotic Theory of Eigenvectors for Large Random Matrices. *arXiv:1902.06846 [math, stat]*, Feb. 2019a. URL http://arxiv.org/abs/1902.06846. arXiv: 1902.06846.

J. Fan, Y. Fan, X. Han, and J. Lv. SIMPLE: Statistical Inference on Membership Profiles in Large Networks. *arXiv:1910.01734 [math, stat]*, Oct. 2019b. URL http://arxiv.org/abs/1910.01734. arXiv: 1910.01734.

L. Lei. Unified \$\ell_{2\rightarrow\infty}\$ Eigenspace Perturbation Theory for Symmetric Random Matrices. *arXiv:1909.04798 [math, stat]*, Sept. 2019. URL http://arxiv.org/abs/1909.04798. arXiv: 1909.04798.

X. Mao, P. Sarkar, and D. Chakrabarti. Estimating Mixed Memberships with Sharp Eigenvector Deviations. *arXiv:1709.00407 [physics, stat]*, Nov. 2019. URL http://arxiv.org/abs/1709.00407. arXiv: 1709.00407.

M. Tang. The eigenvalues of stochastic blockmodel graphs. *arXiv:1803.11551 [cs, stat]*, Mar. 2018. URL http://arxiv.org/abs/1803.11551. arXiv: 1803.11551.

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:1710.10936 [stat]*, Oct. 2017. URL http://arxiv.org/abs/1710.10936. arXiv: 1710.10936.