Név: **Kovács Sebestyén**
NEPTUN-kód: **FP959I**
Témavezető: **Backhausz Ágnes Mariann**

Math project 2
Random matrices, perturbations and their applications in statistics
(Véletlen mátrixok, perturbációk és statisztikai alkalmazásaik)

1 Introduction

In Multivariate statistics and in PCA (Principal Component Analysis) it is really important to understand the behaviour of data matrices. In applications we often try to observe a low rank matrix that is the sum of a sign matrix, a sparse matrix and a random noise matrix. Here it is really important to estimate the signal matrix punctually. By this the singular vector decomposition of matrices can help us a lot, so the role of the singular vectors and values of matrices is dominant, because it can help us filter the noise matrix. The applications of these terms are meaningful today, for example when we want to see the driving factors in the stock market or when we want to separate the moving object from the background by video surveillance, and also if we are interested in determining the location of sensor nodes with an unknown position. In these cases the singular value decomposition can solve these tasks, as we can read it in [4].

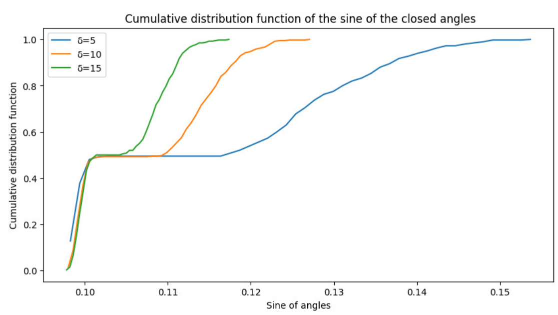

Figure 1: Sine of closed angles of the first singular vector of A and $A + E$ with a simple two-rank matrix A and Bernoulli matrix E.

2 Perturbed random matrices

2.1 Angles of singular vectors

This semester I made various simulations about theorems that I read in articles, to be able to understand the meaning and strength of these theorems. Now I would like to define the singular vectors and the values of matrices.

Definition. The Singular Value Decomposition (SVD) says if $A \in \mathbb{R}^{d_1 \times d_2}$, $r(A) = n$ then there exists only one $U \in \mathbb{R}^{d_1 \times n}$, $\Sigma \in \mathbb{R}^{n \times n}$ and $V \in R^{d_2 \times n}$ such that

$$A = U\Sigma V^T = \sum_{i=1}^{n} \sigma_i u_i v_i^T,$$

where U and V matrices are orthogonal with u_1, u_2,... u_n and v_1, v_2, ... v_n column vectors and Σ is a diagonal matrix with

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n > 0$$

in the main diagonal. We say that v_i and σ_i are the i-th singular vector and value of matrix A (for $i = 1 \ldots n$).

I will now observe how differently the first singular vector of the signal matrix behaves from the new low rank matrix, if the distribution of the noise matrix is Bernoulli, i.e. $P(E_{ij} = 1) = P(E_{ij} = -1) = 1/2$ with independent entries. While making Figure 1 I simulated 400 independent Bernoulli matrices and I added three easy deterministic sparse matrices to them, only with four nonzero components. The sizes of these matrices were 400×400. The rank of the deterministic matrices was two. I calculated the first singular vectors of the deterministic matrices and of the new matrices, i.e. the deterministic matrices plus Bernoulli matrices with the program. After this I calculated the sine of the closed angles of these vectors and plotted its cumulative distribution functions. Picture 1 is similar to the figure in the Wang's, Vu's and O'Rourke's article in [5] on the 31st page. The reason is that the ninth theorem of this article says that if the difference of the first singular value and the second singular value of the deterministic matrix is large enough then the sine of the closed angles of these vectors will be small with high probability. But one can see a difference between my figure and the diagram of the article, namely the cumulative distribution function converges to 1 faster in my figure. In my opinion it is so because we did not work with the same deterministic matrices, my matrices were really easy sparse matrices so the convergence could be faster in this situation.

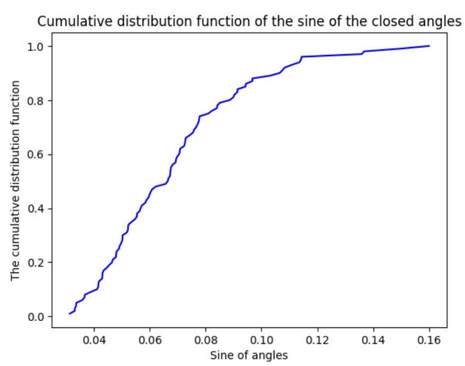

Figure 2: Sine of closed angles with Wishart matrix A and Bernoulli matrix E.

In figure 2 one can see another cumulative distribution function. Here I simulated a Wishart matrix as a data matrix with rank 2. I added 100 independent random Bernoulli matrices to them and got 100 new matrices, and finally calculated the sine of the closed angles of the first singular vector of my Wishart matrix and the first singular vectors of the new matrices. The cumulative distribution function is about these sine values. Here one can see that now the convergence to 1 is a little bit slower than earlier. Maybe it's so because a Wishart matrix is a more complex one. It can be checked that the cumulative distribution function achieves 1 at approximately 0.16.

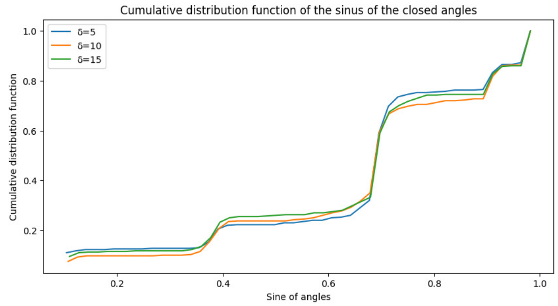

Figure 3: Sine of closed angles with more complex two-rank matrix A and Bernoulli matrix E.

In figure 3 I simulated again 3 deterministic matrices and added 400 independent random Bernoulli matrices to them similarly to the first situation. Here the difference is that now I worked with two rank matrices which are less sparse. The first two columns were linearly independent and the other columns were the linear combinations of them. It can be observed that at this time these matrices are more similar to one another so the difference of the first and the second singular value does not play such an important role as it does in the first case.

2.2 l^∞ eigenvector bounds

To study the strength of a theorem on perturbed low rank matrices [4], I prepared simulations about the following problem. Here I assumed that we have $A' = A + E$ where A is a signal matrix and E is the sum of a sparse and a noise matrix. E was always a Bernoulli matrix, and the signal matrix was chosen from normal and Wishart distribution. I made the singular value decomposition for A' and A too, so I got V' and V matrices where

$$A' := U'\Sigma'V'^T = \sum_{i=1}^{n} \sigma_i' u_i' v_i'^T \qquad \text{and} \qquad A := U\Sigma V^T = \sum_{i=1}^{n} \sigma_i u_i v_i^T$$

where

$$U, U' \in \mathbb{R}^{d_1 \times n} \qquad \text{and} \qquad V, V' \in \mathbb{R}^{d_2 \times n}.$$

Let $\sigma_1, \dots \sigma_n$ be the singular values of A. We will define $\mu(U)$ and $\mu(V)$ as follows:

$$\mu(U) := \frac{d_1}{n} \cdot \max_i \sum_{j=1}^{n} U_{ij}^2 \qquad \text{and} \qquad \mu := \mu(V) := \frac{d_2}{n} \cdot \max_i \sum_{j=1}^{n} V_{ij}^2.$$

Let us define A_r, where this is the best rank-r approximation of A under the Forbenius norm:

$$A_r := \sum_{i=1}^{r} \sigma_i u_i v_i^T.$$

For the sake of completeness we define two norms for a matrix $M = [M_{ij}] \in \mathbb{R}^{d_1 \times d_2}$ as follows:

$$\|M\|_{\max} := \max_{ij} |M_{ij}|, \qquad \|M\|_\infty := \max_i \sum_{j=1}^{d_2} |M_{ij}|.$$

Now we can formulate our

Theorem 1. (cf. [4]) We suppose that $\delta > ||E||_2$ and $\sigma_r - \varepsilon = \Omega(r^3\mu^2||E||_\infty)$, where $\varepsilon = ||A - A_r||_\infty$. If A is symmetric and for any $i = 1, \ldots r$ the interval $[\sigma_i - \delta, \sigma_i + \delta]$ does not contain any singular values of A other than σ_i, then

$$||V' - V||_{\max} = \mathcal{O}\left(\frac{r^4\mu^2||E||_\infty}{(\sigma_r - \varepsilon)\sqrt{n}} + \frac{r^{\frac{3}{2}}\mu^{\frac{1}{2}}||E||_\infty}{\delta\sqrt{n}}\right).$$

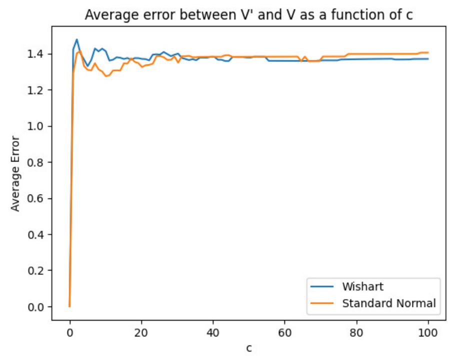

Figure 4: Simulation results for Theorem 1 with $0 \leq c \leq 100$.

I illustrated this theorem with two diagrams in order to understand how good the upper estimation of the theorem is. I drawn E from Bernoulli distribution and A from Wishart and standard normal distribution twenty times independently from one another. I determined the singular value decomposition of $(A' = A + c \cdot E, A)$ for every $c \in [0, 100]$, I got (V', V) matrices and plotted the mean of the twenty cases of $||V' - V||_{\max}$ depending on the function of c. In figure 4 one can see that the distribution does not play an important role here, because the norms of the differences behave similarly in the two distributions.

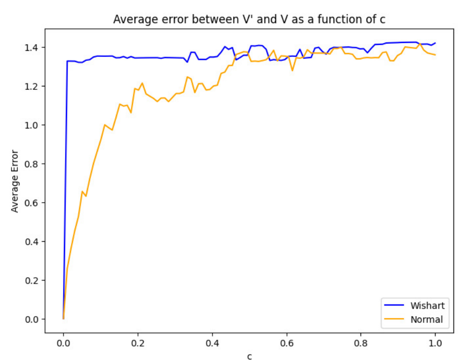

Figure 5: Simulation results for Theorem 1 with $0 \leq c \leq 1$.

But if we enlarge our simulation upon the interval of $[0, 1]$, then it is easy to see in figure 5 that the difference of the distributions counts, because the standard normal distribution achieves the average error 1.3 between V' and V by $c = 0.5$ much more slower than the Wishart distribution does it. Trivially both curves start from zero, for $c = 0$, $V' = V$ because of the uniqueness of singular value decomposition. It can be noticed that the condition $\sigma_r - \varepsilon = \Omega(r^3\mu^2||E||_\infty)$ of the theorem does not fulfil in Wishart distribution, because when c is near to zero, then the average error between V' and V does not change linearly. We know that $||cE||_\infty = c \cdot ||E||_\infty$ so if we calculate the average error with $c \cdot E$ instead of E then the upper estimation should depend on the quantities in the Ordo marking only linearly.

3 Further models

I read two other articles this semester. In [3] it was interesting that in case of some assumptions the authors could give a strong estimation about the norm of the noise matrix and in the symmetric case they could bound from below the norm of the noise matrix with the distance of the estimated leading eigenvalue (estimated first singular value) and the right leading eigenvalue (right first singular value). In [1] the authors examined the convergence of the corresponding singular values of the low rank matrices and could give the limit in some special cases. In all these cases and by PCA the singular value decomposition can help us to maximize functions of matrices with some conditions.

4 Future plans

The general goal is a deeper understanding of the applications of perturbed random matrices in statistics. We have seen by the simulations that the structure of the matrix can affect the behaviour of its singular vectors. I would like to understand this phenomenon more profoundly with new simulations. I plan to survey the newest articles of the literature of random perturbed matrices, and last but not least, I would like to apply the above to real data.

References

[1] Benaych-Georges, F.; Nadakuditi, R. R.: *The singular values and vectors of low rank perturbations of large rectangular random matrices*, J. Multivariate Anal. **111** (2012), 120–135.

[2] Blum, A.; Hopcroft, J.; Kannan, R.: *Foundations of Data Science*, Cambridge University Press, 2020.
(`https://doi.org/10.1017/9781108755528`, cf. `https://www.cs.cmu.edu/~venkatg/teaching/CStheory-infoage/book-chapter-4.pdf`)

[3] Chen, Y.; Cheng, Ch..; Fan, Y.: *Asymmetry helps: Eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices*, Annals of statistics, 2021, 49.1: 435.
(`https://doi.org/10.1214/20-aos1963`, cf. `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8300484/pdf/nihms-1639565.pdf`)

[4] Fan, Jianqing; Wang, Weichen; Zhong, Yiqiao: *An l^∞ eigenvector perturbation bound and its application to robust covariance*, J. Mach. Learn. Res. 18 (2017), Paper No. 207, 42 pp.
(`https://www.jmlr.org/papers/volume18/16-140/16-140.pdf`)

[5] O'Rourke, S.; Vu, V.; Wang, K.: *Random perturbation of low rank matrices: Improving classical bounds*, Linear Algebra and its Applications **540** (2016), 26–59.