Analysis of High-Dimensional Data

Dolev Dublon
Moria Grohar
Shirel Zecharia
Roy Harel
Hadas Evers

March 6, 2024

Abstract

This Survey explores the fascinating world of high-dimensional data analysis, focusing on the properties of random matrices. These matrices play a crucial role in various fields such as computer science, statistics, and machine learning, especially concerning their singularity and the analysis of their smallest singular values.

1 Introduction

In the realm of high-dimensional data analysis, random matrices have captivated the attention of mathematicians and scientists due to their ubiquitous presence in various domains, including computer science, statistics, and machine learning. Understanding their behavior is paramount, particularly regarding their singularity (non-invertibility) and the analysis of their smallest singular values. This article delves into ten research papers that shed light on these aspects.

2 timeline

[2]	[10]	[3]	[7]	[5]	[4]	[1]	[6]	[9]	[8]
9.5.05	31.1.08	9.2.09	10.8.20	3.9.20	3.11.20	5.11.20	5.5.21	4.8.22	22.12.23

3 Mathematical Context Main

[2] The article "Random Symmetric Matrices Are Almost Surely Non-Singular" by K. Costello, T. Tao, and V. Vu presents a significant result in the field of random matrices. The main finding of the article is the proof that a random symmetric matrix Q_n with independent and identically distributed (with the same distribution) Bernoulli variables as its upper diagonal entries is almost surely non-singular, with a probability of $1 - O(n^{-1/8+\delta})$ for any $\delta > 0$. This result extends previous results for random matrices to

more general models of random matrices. The article presents the history of the non-singularity problem in random matrices, namely whether it is true that a random matrix A_n with independent Bernoulli variables is almost surely non-singular. This question was positively answered by Komlós in 1967, and later he generalized the result to more general models of random matrices. In a recent paper, Tao and Vu found a different proof for random matrices that provides a precise estimate for the absolute value of the determinant of the matrix A_n. Building upon these previous proofs, the authors develop a quadratic version of Littlewood-Offord type results concerning the concentration of random variables to prove the non-singularity of Q_n- a random symmetric matrix. This method allows researchers to overcome the challenge of the row and column transpose, which was a hurdle in previous proofs for random matrices due to the dependence between the row vectors of the matrix Q_n. The article raises open questions for future research in the field of random matrices:

Determinant Estimation: The article raises the question of estimating the determinant of random matrices. The estimation provided in the article is: $|det\ Q_n| = n^{(1/2-o(1))n}$.

Singularity Probability: Another open question raised in the article relates to estimating the probability that a random matrix is singular. The authors estimate that the probability of Q_n being a singular matrix is $(1/2 + o(1))^n$.

The quadratic variant of the Littlewood-Offord : Let Q be a quadratic random variable defined as: $Q = \sum_{1 \le i,j \le n} c_{ij} z_i z_j$ where z_i are random variables, $1, \dots, n = U_1 \cup U_2$ is a non-trivial partition, and S is a non-empty subset of U_1. For each $i \in S$, let d_i be the number of indices $j \in U_2$ such that $|c_{ij}| \ge 1$. If $d_i \ge 1$ for each $i \in S$, and I is an interval of length 1, then: $P(Q \in I) = O(|S|^{-1/2} + |S|^{-1} \sum_{i \in S} {d_i}^{-1/2})^{1/4}$

4 roy article1 summary

[10]

5 roy article2 summary

[3]

6 shirel Singularity of random symmetric matrices revisited summery

[1] Singularity of Random Symmetric Matrices: This paper studies the probability, denoted as $P(det(M_n) = 0)$, of a singular random $n \times n$ matrix M_n drawn uniformly from matrices with entries of -1 and 1. It's a long-standing problem with the conjecture that $P(det(M_n) = 0)$ goes to zero exponentially fast with n, written as $P(det(M_n) = 0) = (1 + o(1))n^2 2^{-n+1}$. Prior work established bounds on this probability but couldn't overcome a natural barrier of $exp(-c\sqrt{n}logn)$ for some constant c, where the randomness in the matrix isn't "reused."

This paper breaks the barrier by introducing a "rough" inverse Littlewood-Offord theorem, proving that $P(det(M_n) = 0) \le exp(-cpnlogn)$ for some constant c and sufficiently large n. The authors build upon previous work that divides vectors v into structured and unstructured and analyzes their contribution to $P(det(M_n) = 0)$. Their

key improvement lies in a simpler and stronger "rough" inverse Littlewood-Offord theorem, which defines concepts like
$N_{\mu}(w) := x \in \mathbb{Z}_p : P(X\mu(w) = x) > 2^{-1}P(X_{\mu}(w) = 0)$ to analyze the "neighborhood" of a vector w under a random walk $X_{\mu}(v) := \varepsilon_1 v_1 + ... + \varepsilon_n v_n$, where ε_i are independent and take values -1, 0, or 1 with equal probability $\mu/2$.

7 moria Singularity of discrete random matrices revisited summery

[6] The article discusses the singularity of discrete n $\times$ n random matrices $M_n(\xi)$, focusing on matrices with non-constant real-valued random variables ξ. It explores the probability of singularity $\mathbb{P}[M_n \text{ is singular}] = \mathbb{P}[\text{zero row or column}] + (1 + o_n(1))\mathbb{P}[\text{two equal (up to sign) rows or columns}]$ in these random matrices and confirm a conjecture related to this topic. They provide precise results for various scenarios, including cases involving Bernoulli distributions:
For Bernoulli Distribution with $\rho \in (0, 1/2)$: $\mathbb{P}[M_n \text{ is singular}] = 2n(1-\rho)^n + (1 + o_n(1))n(n-1)(\rho^2 + (1-\rho)^2)^2$.
For Bernoulli Distribution with $\rho \in (1/2, 1)$: $\mathbb{P}[M_n \text{ is singular}] = (1 + o_n(1))n(n-1)(\rho^2 + (1-\rho)^2)^2$. One key aspect of the study is the analysis of the contribution of the 'compressible' part of the unit sphere into 'structured' and 'unstructured' components to the lower tail of the smallest singular value of the matrices. By examining this contribution, the researchers aim to gain a deeper understanding of the processes occurring in these random matrices and identify the impact of specific subsets on the singular behavior of the matrices. In this article the novelties are: Unlike "structured" vectors with similar components, "unstructured" vectors in random matrices have diverse values. The authors exploit this non-uniformity to analyze them. They introduce a novel "multi-slice" theorem to handle these vectors, overcoming challenges of dependence and non-integer values. This method offers a powerful tool for understanding how unstructured vectors influence the invertibility of random matrices, building upon previous work on simpler cases.

TODO::::NOT FINISHED

8 moria on the smallest singular value of symmetric random matrices summary

[4] In this article the authors explores the behavior of the smallest singular value in n $\times$ n random symmetric matrices $A_n(ij) = A_n(ji)$. They investigate M_n random matrix $n \times n$ each of the entries is an independent copy of a sub-Gaussian random variables ξ with mean 0 and variance 1. The smallest singular value of M_n is denoted as $s_n(M_n)$, defined as: $s_n(M_n) = inf_{v \in \mathbb{S}^{n-1}} \|Mv\|_2$ where $\mathbb{S}^{n-1}$ is the unit sphere in $\mathbb{R}^n$ **(from The Littlewood-Offord problem and invertibility of random matrices)**. The main result shows that for a random symmetric matrix A_n with sub-Gaussian entries, the probability of $s_n(A_n)$ being less than $\epsilon/\sqrt{n}$ is bounded by: $P[s_n(A_n) \leq \epsilon/\sqrt{n}] \leq C\epsilon^{1/8} + 2e^{-cn^{1/2}}$ for all $\epsilon \geq 0$, where C and c are constants depending on the sub-Gaussian norm of ξ. When ξ is a Rademacher random variable, the probability bound becomes: $P[s_n(A_n) \leq \epsilon/\sqrt{n}] \leq O(\epsilon^{1/8} + \exp(-\Omega((\log n)^{1/4} n^{1/2})))$ It also mentions the Median Regularized Least Common

Denominator (MRLCD) and the Median Threshold. These notions improve upon the Regularized Least Common Denominator (RLCD) by efficiently utilizing the arithmetic structure of vectors, particularly when many projections of a vector are arithmetically unstructured. They demonstrate that the MRLCD and median threshold have level sets that can be covered by sufficiently small nets at the appropriate scale, which is crucial for their applications. These new concepts can replace RLCD in various applications and are expected to provide better quantitative estimates.

9 shirel On the smoothed analysis of of the smallest singular value with discrete noise

[5]

10 dolev article1 summery of graph

[9]

11 shirel on graph article

[7]

12 Haddas last article

[8]

13 Results

// in the mean time we don't fill this

14 References

References

[1] Marcelo Campos et al. "Singularity of random symmetric matrices revisited". In: *arXiv preprint arXiv:2011.03013* (2020).

[2] Kevin Costello, Terence Tao, and Van Vu. "Random symmetric matrices are almost surely non-singular". In: *arXiv preprint math/0505156* (2005).

[3] Kevin P Costello. "Bilinear and Quadratic Variants on the Littlewood-Offord Problem". In: *arXiv preprint arXiv:0902.1538* (2009).

[4] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney. "On the smallest singular value of symmetric random matrices". In: *arXiv preprint arXiv:2011.02344* (2020).

[5] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney. “On the smoothed analysis of the smallest singular value with discrete noise”. In: *arXiv preprint arXiv:2009.01699* (2020).

[6] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney. “Singularity of discrete random matrices”. In: *Geometric and Functional Analysis* 31 (2021), pp. 1160–1218.

[7] Matthew Kwan and Lisa Sauermann. “An algebraic inverse theorem for the quadratic Littlewood-Offord problem, and an application to Ramsey graphs”. In: *arXiv preprint arXiv:1909.02089* (2019).

[8] Matthew Kwan and Lisa Sauermann. “Resolution of the quadratic Littlewood–Offord problem”. In: *arXiv preprint arXiv:2312.13826* (2023).

[9] Matthew Kwan et al. “Anticoncentration in Ramsey graphs and a proof of the Erdős-McKay conjecture”. In: *arXiv e-prints* (2022), arXiv–2208.

[10] Mark Rudelson and Roman Vershynin. “The Littlewood–Offord problem and invertibility of random matrices”. In: *Advances in Mathematics* 218.2 (2008), pp. 600–633.