Analysis of High-Dimensional Data

Dolev Dublon

Moria Grohar

Shirel Zecharia

Roy Harel

Hadas Evers

February 28, 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 Mathematical Context - Main

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, $1, \ldots, n = U_1 \cup U_2$ 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: $P(Q \in I) = O(|S|^{-1/2} + |S|^{-1} \sum_{i \in S} d_i^{-1/2})^{1/4}$

3 roy article1 summary

4 roy article2 summary

5 shirel Singularity of random symmetric matrices revisited summery

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 $X_\mu(v) := \varepsilon_1 v_1 + \ldots + \varepsilon_n v_n$, where ε_i are independent and take values -1, 0, or 1 with equal probability $\mu/2$.

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