RANDOM MATRICES: l_1 CONCENTRATION AND DICTIONARY LEARNING WITH FEW SAMPLES

KYLE LUH (JOINT WORK WITH VAN VU)

Department of Mathematics, Yale University

Abstract. Let X be a sparse random matrix of size $n \times p$ ($p \gg n$). We prove that if $p \geq Cn \log^4 n$, then with probability $1 - o(1)$, $\|X^T v\|_1$ is close to its expectation for all vectors $v \in \mathbb{R}^n$ (simultaneously). The bound on p is sharp up to the polylogarithmic factor.

The study of this problem is directly motivated by an algorithmic application. Let A be an $n \times n$ matrix, X be an $n \times p$ matrix and $Y = AX$. A challenging and important problem in data analysis, motived by dictionary learning and other practical problems, is to recover both A and X, given Y. Under normal circumstances, it is clear that this problem is underdetermined. However, in the case when X is sparse and random, Spielman, Wang and Wright showed that one can recover both A and X efficiently from Y with high probability, given that p (the number of samples) is sufficiently large. Their method works for $p \geq Cn^2 \log^2 n$ and they conjectured that $p \geq Cn \log n$ suffices. The bound $n \log n$ is sharp for an obvious information theoretical reason. The matrix concentration result verifies the Spielman et. al. conjecture up to a $\log^3 n$ factor.

Our proof of the concentration result is based on two ideas. The first is an economical way to apply the union bound. The second is a refined version of Bernstein's concentration inequality for sum of independent variables. Both have nothing to do with random matrices and are applicable in general settings.

E-mail address: kyle.luh@yale.edu.

Key words and phrases. Dictionary learning, matrix concentration.