Non-Asymptotic Theory of Random Matrices
Lecture 1: Background, Techniques, Methods

Lecturer: Roman Vershynin Scribe: Deanna Needell

Thursday, January 4, 2006

1 Two Types of Statements

This course will focus on the non-asymptotic theory of random matrices. To emphasize the notion of non-asymptotic theory, we first demonstrate two types of probabilistic statements: asymptotic and non-asymptotic. This first example exhibits an asymptotic statement about sums of independent identically distributed (i.i.d.) random variables.

Example 1 (Central Limit Theorem)**.** *Let $X_1, X_2, \dots,$ be symmetric ± 1-valued random variables (called Bernoulli random variables). Then by the Central Limit Theorem as $n \to \infty$,*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \overset{dist.}{\to} g, \tag{1}$$

where g is a N(0,1) standard normal random variable. Recall that convergence in distribution is equivalent to saying that for any $t \in \mathbb{R}$, as $n \to \infty$,

$$\mathbb{P}\Big(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i > t\Big) \to \mathbb{P}(g > t) = \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-t^2/2} dt. \tag{2}$$

The above example exhibits an asymptotic statement since it says something about the limit as $n \to \infty$. Non-asymptotic statements are statements that hold for all n, or at least for all $n > n_0$ where n_0 is some constant. The next example demonstrates such a non-asymptotic statement known as Bernstein's Inequality.

Example 2 (Bernstein's Inequality)**.** *Let $X_1, X_2, \dots$ be as in Example 1. For any n and $t > 0$,*

$$\mathbb{P}\Big(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i > t\Big) \le 2\mathsf{e}^{-t^2/2}. \tag{3}$$

2 Non-Asymptotic Techniques of Random Matrices

We will look at random matrices A with random i.i.d. entries a_{ij}. For example, A could be the random matrix whose entries are i.i.d. standard normal random variables (random Gaussian matrix), or i.i.d. symmetric Bernoulli random variables (random Bernoulli matrix). Henceforth assume $N > n$. A will be a tall $N \times n$ or a flat $n \times N$ matrix. As we will see later, non-asymptotic techniques have applications in Geometric and Functional Analysis, Computer Science, Approximation Theory, and Engineering (specifically Signal Processing and Coding). From the Functional Analysis viewpoint, we will look at these matrices as linear operators. In the case where A is $n \times N$, A is a linear operator $A : \mathbb{R}^N \to \mathbb{R}^n$. The study of these kinds of matrices is in particular important to Dimension Reduction, in which a finite set of points in $\mathbb{R}^N$ is mapped into the lower dimensional space $\mathbb{R}^n$ with the hopes that the distances are approximately preserved. In the case where A is $N \times n$, A is the embedding $A : \mathbb{R}^n \to \mathbb{R}^N$.

An important property of random matrices and one that we will be studying is the spectrum of such matrices, and in particular, their singular values. Recall that the singular values of A are the eigenvalues of $\sqrt{A^*A}$. When A is $N \times n$ we order the singular values as $s_1(A) \geq s_2(A) \geq \ldots \geq s_n(A)$. Denote the Euclidean operator norm by $\|\cdot\|$ and recall then that $s_1(A) = \|A\|$ and $s_n(A) = \frac{1}{\|A^{-1}\|}$ where by A^{-1} we mean the inverse of A restricted to the image of A. So then we have that for every x,

$$s_n(A)\|x\| \leq \|Ax\| \leq s_1(A)\|x\|. \tag{4}$$

Thus to say A is an almost-isometry we need only place a lower bound on $s_n(A)$ and an upper bound on $s_1(A)$. This type of bounding has an immediate application in Computer Science for example, when $A : \ell_2^n \to \ell_1^N$. If such an operator is an almost-isometry then the ℓ_2-norm of a vector can be approximately calculated by simply finding its ℓ_1-norm. This eliminates the need for the square root operation, something that is more expensive. It has been shown using the Volume Ration Method that random matrices with $N = 2n$ satisfy this property. Below are a few results on bounds of the smallest and biggest singular values. The first result is due to Gordon [2, 3, 4]. See also [1].

Theorem 3 (Gordon). *If A is an $n \times n$ random Gaussian matrix then with high probability*

$$s_1(A) \lesssim 2\sqrt{n}. \tag{5}$$

Theorem 4 (Litvak, Pajor, Rudelson, Tomczak-Jaegermann [?]). *For random Bernoulli rectangular matrices, the smallest singular value obeys*

$$s_n(A) \geq c_{n/N}\sqrt{n} \tag{6}$$

where $c_{n/N}$ depends only on the ratio n/N.

The next theorem is a very recent result of Rudelson and Vershynin.

Theorem 5 (Rudelson, Vershynin). *For random Gaussian and Bernoulli square matrices, with high probability the smallest singular value obeys*

$$s_n(A) \gtrsim \frac{1}{\sqrt{n}}. \tag{7}$$

We would also like to know the probability that such a matrix is non-singular. That is, we would like an upper bound on $\mathbb{P}(s_n(A) = 0)$. For Gaussian matrices this probability is precisely 0, and for Bernoulli matrices the tight upper bound is conjectured to be $(\frac{1}{2} + o(1))^n$. The following result was proved in 1995 for Bernoulli matrices.

Theorem 6 (Kahn-Komlós-Szemerédi [5]). *Let A be a random Bernoulli matrix. Then there is a constant $c \in (0,1)$ so that*

$$\mathbb{P}(s_n(A) = 0) \leq c^n. \tag{8}$$

We will prove that for any $\epsilon > 0$, $\mathbb{P}(s_n(A) < \frac{\epsilon}{\sqrt{n}}) \leq C\epsilon + c^n$ which will imply the above two theorems.

References

[1] K. Davidson and S. Szarek. Local operator theory, random matrices and banach spaces. *Handbook of the Geometry of Banach Spaces*, 1:317, 2001.

[2] Y. Gordon. On dvoretzky's theorem and extensions of slepian's lemma. *Israel seminar on geometrical aspects of functional analysis*, 2:25, 1983.

[3] Y. Gordon. Some inequalities for guasian processes and applications. *Israel J. Math*, 50:265, 1985.

[4] Y. Gordon. Majorization of gaussian processes and geometric applications. *Probab. Theory Related Fields*, 91:251, 1992.

[5] J. Kahn, J. Komlós, and E. Szemerédi. On the probability that a random ± 1-matrix is singular. *Journal of the American Math. Society*, 8:223, 1995.