Non-Asymptotic Theory of Random Matrices

Lecture 15: Invertibility of Square Gaussian Matrices, Sparse Vectors

Lecturer: Roman Vershynin Scribe: Brendan Farrell

Thursday, February 22, 2006

1 Invertibility of Square Gaussian Matrices

Let A be an $n \times n$ square matrix with i.i.d. standard Gaussian entries. Recall that

$$s_1(A) = \max_{x:\|x\|_2=1} \|Ax\|_2 = \|A\|_2 = O(\sqrt{n})$$

w.h.p.,

$$s_n(A) = \min_{x:\|x\|_2=1} \|Ax\|_2 = \frac{1}{\|A^{-1}\|_2},$$

and

$$\mathbb{E}[s_n(A)] = \sqrt{n} - \sqrt{n} = 0.$$

If A is $n \times (n-1)$, then

$$\mathbb{E}[s_n(A)] \geq \sqrt{n} - \sqrt{n-1} \sim \frac{1}{\sqrt{n}}.$$

In the 1940's von Neumann predicted that $s_n(A) \sim 1/\sqrt{n}$. His motivation was solving a system of linear equations, $Ax = b$, with n equations and n unknowns. The solution $x = A^{-1}b$, however, is inexact because b is subject to roundoff and other errors. Rather than the true b, one must work with a noisy vector $\tilde{b}$, and so one actually computes $\tilde{x} = A^{-1}\tilde{b}$. Therefore, the error is $\|x - \tilde{x}\|_2 = \|A^{-1}(b - \tilde{b})\|_2 \leq \|A^{-1}\|_2 \cdot \|b - \tilde{b}\|_2$. An upperbound on $\|A^{-1}\|_2$ is given by a lower bound on $s_n(A)$.

In 1985 Smale conjectured that

$$\mathbb{P}(s_n(A) \leq \frac{\epsilon}{\sqrt{n}}) \sim \epsilon.$$

This implies, first, that $\mathbb{E}[s_n(A)]$ and $\mathbb{M}[s_n(A)] \sim 1/\sqrt{n}$. Second, this implies that $s_n(A)$ is not concentrated.

Alan Edelman proved this conjecture in 1988 [1] by using an explicit formula for the joint density of singular values $s_1(A), s_2(A), ..., s_n(A)$ of an

$m \times n$ ($m \geq n$), Gaussian matrix. Set $\lambda_k = s_k^2(A)$. Then λ_k are the eigenvalues of A^*A, ordered $\lambda_1 \geq, ..., \lambda_n$. the density is given by

$$dens(\lambda_1, ..., \lambda_n) = K_{m,n} \exp(-\frac{1}{2}\sum_{k=1}^{n}) \prod_{k=1}^{n} \lambda_k^{\frac{m-n-1}{2}} \prod_{j<k}(\lambda_j - \lambda_k).$$

Edelman "integrated out" $\lambda_1, ..., \lambda_{n-1}$ to obtain the explicit density for λ_n. This proved Smale's conjecture for matrices over $\mathbb{R}$. For $\mathbb{C}$, Edelman proved that λ_n is distributed identically with $\chi_2^2/\sqrt{n}$, where $\chi_2^2 = g_1^2 + g_2^2$ and $g_1, g_2 \sim \mathcal{N}(0,1)$ are independent. Spielman-Teng conjectured (ICM 2002, [3]), that for Bernoulli matrices $s_n(A) \sim 1/\sqrt{n}$ w.h.p..

Theorem 1 (Rudelson-Vershynin 2006 [2]). *$s_n(A) \sim 1/\sqrt{n}$ w.h.p. for all subgaussian matrices.*

First we consider the question: Why is the Gaussian square matrix invertible? That is, why is it nonsingular with probability 1?

Full rank $\Leftrightarrow$ all the rows are linearly independent
$\Leftrightarrow$ each row does not lie in the span of the other rows

Let X_k be the k^{th} row of A ($X_k = Ae_n$) and H_k the span of the remaining rows. The reason for good invertibility of A is $\text{dist}(x_k, H_k) \geq$ ... We have $s_n(A) = \min_{x:\|x\|_2=1} \|Ax\|_2$, $x = (x_1, ..., x_n)$, and $Ax = \sum_{i=1}^{n} x_k X_k$.

$$\begin{aligned} \|Ax\|_2 &\geq \text{dist}(\sum_{l=1}^{n} x_l X_l, H_k) \\ &= \text{dist}(x_k X_k, H_k) \\ &= |x_k| \cdot \text{dist}(X_k, H_k). \end{aligned}$$

Fact: $\|Ax\|_2 \geq \max_k |x_k| \cdot \text{dist}(X_k, H_k)$.
We then need to prove lower bounds for both $|x_k|$ and $\text{dist}(X_k, H_k)$.

1. $\|x\|_2 = 1$. So for some k, $|x_k| \geq 1/\sqrt{n}$.

2. For $\text{dist}(X_k, H_k)$, we use the Distance Lemma of Lecture 14.

X_k and H_k are independent, so we condition on H_k. That is, the probability $\mathbb{P}_{X_k}$ is w.r.t. X_k for H_k fixed. The Distance Lemma states

$$\mathbb{P}_{X_k}(\text{dist}(X_k, H_k) < \epsilon) \leq c\epsilon.$$

Thus $\mathbb{P}(\text{dist}(X_k, H_k) < \epsilon) \leq C\epsilon$. We take the union bound over $k = 1, ..., n$:

$$\mathbb{P}(\exists\, k : \text{dist}(X_k, H_k) < \epsilon) \leq C\epsilon n.$$

Define the event

$$\mathcal{E} = \{\text{dist}(X_k, H_k) > \epsilon \; \forall \; k = 1, ..., n\};$$

then $\mathbb{P}(\mathcal{E}^c) = C\epsilon n$. If $\mathcal{E}$ holds, then by the fact above,

$$\|Ax\|_2 \geq \max_k |x_k| \cdot \epsilon \geq \epsilon/n.$$

This holds for all x, $\|x\|_2 = 1$. Thus, by taking the min over all such x, $\mathcal{E} \Rightarrow s_n(A) \geq \epsilon/\sqrt{n}$. We have shown

$$\mathbb{P}(s_n(A) < \frac{\epsilon}{\sqrt{n}}) \leq \mathbb{P}(\mathcal{E}^c) \leq c\epsilon n.$$

Equivalently,

$$\mathbb{P}(s_n(A) < \frac{\epsilon}{n^{3/2}}) \leq C\epsilon.$$

Theorem 2. $s_n(A) \geq n^{-3/2}$ *w.h.p.*

This bound is polynomial, but not sharp.

2 Invertibility of Sparse Vectors

Consider a vector $x \in \mathbb{R}^n$, $\|x\|_2 = 1$, $\text{supp}(x) \subset \{1, 2, ..., n/2\} = I$. Then Ax is equivalent to $A_I x$, where A_I is the restriction of A to the columns given by I. Now A_I has dimension $n \times \frac{n}{2}$, which is rectangular, and thus well invertible.

Definition 3 (Sparse vectors)**.** *A vector* $x \in \mathbb{R}^n$ *is called* k-sparse *if* $|\text{supp}(x)| \leq k$.

If I is fixed, $|I| = \delta n$, for some $\delta \in (0, 1)$. Then the smallest singular value of A_I is distributed $\sim \sqrt{n} - \sqrt{\delta n} > c\sqrt{n}$ w.h.p..

Corollary 4 (to a Theorem of Lecture 11)**.** *Let B be an $n \times \delta n$ Gaussian matrix and $\delta < 1/2$. Then*

$$\mathbb{P}(\min_{x:\|x\|_2=1} \|B\|_2 \geq c\sqrt{n}) > 1 - \binom{n}{\delta n} e^{-cn}.$$

By Stirling's formula, $\binom{n}{k}^k \leq \binom{en}{k}^k$. Thus $\binom{n}{\delta n} \leq \frac{\epsilon}{\delta}^{\delta n} = e^{\log(\frac{\epsilon}{\delta})\delta n} < e^{\frac{cn}{2}}$ for some $\delta =$const.

Lemma 5 (Invertibility of Sparse Vectors)**.**

$$\mathbb{P}\left(\min_{x:\|x\|_2=1,\delta n-\text{sparse}} \|Ax\|_2 \geq c\sqrt{n}\right) \geq 1 - e^{-cn}.$$

The Lemma follows from Corollary 4 and the comments following it. In the next lecture we will consider compressible vectors; that is, vectors which are not sparse, but are well approximated by sparse vectors.

References

[1] A. Edelman. Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.*, 9(4):543–560, 1988.

[2] M. Rudelson and R. Vershynin. Preprint. 2006.

[3] D. A. Spielman and S.-H. Teng. Smoothed analysis of algorithms. In *Proceedings of the International Congress of Mathematicians, Vol. I (Beijing, 2002)*, pages 597–606, Beijing, 2002. Higher Ed. Press.