Non-Asymptotic Theory of Random Matrices
Lecture 16: Invertibility of Gaussian Matrices and Compressible/Incompressible Vectors

Date: February 27, 2007

Lecturer: Roman Vershynin Scribe: Matthew Herman

1 Background and Motivation

We begin this lecture by asking why should an arbitrary $n \times n$ Gaussian matrix A be invertible? That is, does there exist a lower bound on the smallest singular value

$$s_n(A) \;=\; \inf_{x \in S^{n-1}} \|Ax\|_2 \;\geq\; \frac{c}{\sqrt{n}}$$

where $c > 0$ is an absolute constant. There are two reasons (or cases) which we will pursue in this lecture.

1. In Lecture 15 we saw that the invertibility of rectangular (i.e., non-square) Gaussian matrices yields invertibility of A for all *sparse* vectors. Specifically, we derived **Sparse Lemma 5** which stated that

   There exists an absolute constant $\delta \in (0,1)$ *such that with probability* $1 - e^{-cn}$
   $$\inf_{\substack{x \in S^{n-1} \\ (\delta n)\text{-sparse}}} \|Ax\|_2 \;\geq\; c\sqrt{n}. \tag{1}$$

2. Suppose that the rows $X_1, \ldots, X_n$ of A are "very" linearly independent. This has a geometric interpretation as we saw the **Distance Lemma** of Lecture 14. Let the hyperplane $H_k = \operatorname{span}(X_j)_{j \neq k}$. Then
   $$\mathbb{P}(\operatorname{dist}(X_k, H_k) < \varepsilon) \sim \varepsilon$$
   and it follows that
   $$\|Ax\|_2 \geq \max_k |x_k| \cdot \operatorname{dist}(X_k, H_k) \tag{2}$$
   for all $x \in S^{n-1}$. This yields invertibility of A for all *spread* vectors, e.g., if $|x_k| \sim \frac{1}{\sqrt{n}}$, then $\|Ax\|_2 \geq \frac{1}{\sqrt{n}} \cdot \text{const}$.

In some sense these two cases represent opposite extremes of vectors on the unit sphere. Our strategy will be to decompose the unit sphere into sets of vectors that are either sparse, spread, or neither. These fall into two mutually exclusive classes: compressible vectors and incompressible vectors.

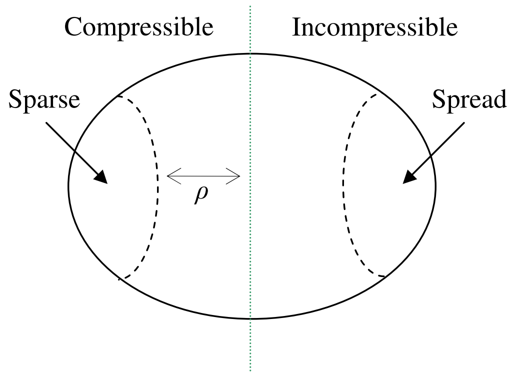

Figure 1: Decomposition of the unit sphere

Definition 1 (Compressible vector)**.** *Let $\delta, \rho \in (0,1)$. A vector $x \in S^{n-1}$ is called compressible if its distance to the set of (δn)-sparse vectors is less than or equal to ρ.*

Notation (Compressible and incompressible sets). The set of compressible vectors with respect to δ, ρ is denoted as $Comp = Comp(\delta, \rho)$. The set of incompressible vectors, denoted as $Incomp$, is the complement of $Comp$ on the unit sphere, i.e., $Incomp = S^{n-1} \backslash Comp$.

Equivalently, $x \in Comp$ if and only if there exists a (δn)-sparse vector y such that $\|x-y\|_2 \leq \rho$. This says that most of the energy (or "information") of vector x is contained in just δn of its n coordinates.

In signal processing, a signal $x \in \mathbb{R}^n$ is compressible if and only if its coefficients decay "fast." That is, by arranging (x_k) in non-increasing order $x_1^* \geq x_2^* \geq \ldots \geq x_n^*$, then x is compressible if and only if $|x_k^*| \leq Ck^{-1/p}$ for all k where C is an absolute constant and $p \in (0,1)$. For instance, (x_k) can be the Fourier or wavelet coefficients of the signal. Thus we see that most of the signal's information is carried in just δn coefficients. So $x \approx (\delta n)$-sparse vector which implies that x is compressible.

Lemma 2 (Incompressible vectors are spread)**.** *Let $x \in Incomp$. Then there exists $\sigma \subset (1, \ldots, n)$ such that $|\sigma| \geq c_0 n$ and*

$$\frac{c_1}{\sqrt{n}} \;\leq\; |x_k| \;\leq\; \frac{c_2}{\sqrt{n}}$$

for all $k \in \sigma$.

Proof. Exercise. □

Note that here, c_0, c_1, c_2 depend only on δ and ρ.

2 Invertibility for compressible and incompressible vectors

With the definitions and lemmas of the previous section we can now address invertibility for compressible and incompressible vectors.

2.1 Compressible vectors

Invertibility for compressible vectors follows directly from the Sparse Lemma in (1) and approximation. Specifically, note that every $x \in Comp$ can be expressed as $x = y + z$ where y is a (δn)-sparse vector and $\|z\|_2 \leq \rho$. Then from the reverse triangle inequality we have

$$\|y\|_2 \;\geq\; \|x\|_2 - \|z\|_2 \;\geq\; 1 - \rho$$

and

$$\begin{aligned}\inf_{x \in Comp} \|Ax\|_2 \;&\geq\; \inf_{\substack{y:(\delta n)\text{-sparse} \\ \|y\|_2 \geq 1-\rho}} \|Ay\|_2 \;-\; \sup_{\|z\|_2 \leq \rho} \|Az\|_2 \\ &=\; a \;-\; b.\end{aligned}$$

By the Sparse Lemma we have

$$a \;\geq\; (1-\rho)c\sqrt{n}$$

with probability $1 - e^{-cn}$, and

$$b \;=\; \rho\,\|A\| \;\leq\; C\rho\sqrt{n}$$

also with probability $1 - e^{-cn}$. Hence, with probability $1 - 2e^{-cn}$ we have

$$\begin{aligned} a - b \;&\geq\; \Big((1-\rho)c - C\rho\Big)\sqrt{n} \\ &\geq\; c'\sqrt{n}\end{aligned}$$

if $\rho > 0$ is chosen as a sufficiently small absolute constant.

Now we can revise (1) for general compressible vectors:

Lemma 3 (Compressible vectors)**.** *There exists absolute constants $\delta, \rho \in (0,1)$ such that with probability* $1 - 2e^{-cn}$

$$\inf_{x \in Comp} \|Ax\|_2 \;\geq\; c\sqrt{n}.$$

2.2 Incompressible vectors

There are many more incompressible than compressible vectors, so this is intuitively harder to analyze. We make the following assumptions:

a) Incompressible vector x has $c_0 n$ coordinates $\sim \frac{1}{\sqrt{n}}$ (from Lemma 2)

b) $\mathrm{dist}(X_k, H_k) > \varepsilon$ with probability $1 - c\varepsilon$ (from last lecture).

Then we expect that $(1 - c\varepsilon)n$ of these have a distance greater than ε. Take the intersection of these sets of coordinates. It is non-empty because $c_0 n + (1 - c\varepsilon)n > n$ (with ε small enough). Thus, if $k \in$ intersection, then $\|Ax\|_2 \geq \frac{1}{\sqrt{n}}\varepsilon$.

Lemma 4 (Invertibility via distance to a hyperplane)**.**

$$\mathbb{P}\Big(\inf_{x \in Incomp} \|Ax\|_2 < \frac{\rho\varepsilon}{\sqrt{n}}\Big) \;\leq\; \frac{1}{\delta}\,\mathbb{P}\Big(dist(X_1, H_1) < \varepsilon\Big)$$

since the X_k are i.i.d..

Observation 5. *Note that there is no union bound as before which is why the RHS has no dependence on n.*

Proof. Let $\mathbf{1}_E$ be the indicator function on the set E. Recall that its expected value is just $\mathbb{E}\,\mathbf{1}_E = \mathbb{P}(E)$. We expect the number of coordinates that satisfy $\mathrm{dist}(X_k, H_k) < \varepsilon$ to be

$$\begin{aligned}
\mathbb{E}\,|\{k : \mathrm{dist}(X_k, H_k) < \varepsilon\}| \;&=\; \mathbb{E}\sum_{k=1}^{n} \mathbf{1}_{\{\mathrm{dist}(X_k, H_k) < \varepsilon\}} \\
&=\; \sum_{k=1}^{n} \mathbb{P}\Big(\mathrm{dist}(X_k, H_k) < \varepsilon\Big) \\
&=\; pn
\end{aligned}$$

where $p = \mathbb{P}\Big(\text{dist}(X_k, H_k) < \varepsilon\Big)$.

Consider the set $\sigma_{bad} = \{$ at least δn distances where $\text{dist}(X_k, H_k) < \varepsilon\}$ to be an undesirable or "bad" set. Then using Markov's inequality we have

$$\mathbb{P}(\sigma_{bad}) \;=\; \mathbb{P}\Big(|\{k : \text{dist}(X_k, H_k) < \varepsilon\}| \geq \delta n\Big) \;\leq\; \frac{pn}{\delta n} \;=\; \frac{p}{\delta}.$$

Now consider the "good" set $\sigma_1 = \{k : \text{dist}(X_k, H_k) \geq \varepsilon\}$. The "good" event $U = \{|\sigma_1| \geq (1-\delta)n\}$. Then the probability of its complement is $\mathbb{P}(U^{\complement}) \leq \frac{p}{\delta}$, so "many distances are big."

On the other hand, many coordinates of x are big too, so

$$x \in Incomp \quad \Rightarrow \quad \sigma_2 = \{k : |x_k| > \frac{\rho}{\sqrt{n}}\} \text{ with } |\sigma_2| \geq \delta n. \quad \text{(Ex.)}$$

If the good event U occurs, then

$$\sigma_1 \cap \sigma_2 \neq \emptyset$$

because

$$|\sigma_1| + |\sigma_2| > n.$$

Therefore, there exists k such that $\text{dist}(X_k, H_k) \geq \varepsilon$ and $|x_k| > \frac{\rho}{\sqrt{n}}$. Hence,

$$\|Ax\|_2 \geq \frac{\rho}{\sqrt{n}}\,\varepsilon.$$

Observation 6. *By the Distance Lemma*

$$\mathbb{P}\Big(\mathit{dist}(X_1, H_1) < \varepsilon\Big) \leq C\varepsilon.$$

Thus the RHS $\leq \frac{C}{\delta}\,\varepsilon \leq c_1\varepsilon$.

□

We arrive at the following theorem.

Theorem 7 (Invertibility of Gaussian matrices)**.** *Let A be an $m\times n$ Gaussian matrix. Then for all $\varepsilon > 0$*

$$\mathbb{P}\Big(s_n(A) < \frac{\varepsilon}{\sqrt{n}}\Big) \;<\; C\varepsilon.$$

In particular, the median of $s_n(A) \sim \frac{1}{\sqrt{n}}$ is sharp!

Proof.

$$\mathbb{P}\Big(\inf_{x\in S^{n-1}}\|Ax\|_2 < \frac{\varepsilon}{\sqrt{n}}\Big) \le \mathbb{P}\Big(\inf_{x\in Comp}\|Ax\|_2 < \frac{\varepsilon}{\sqrt{n}}\Big) + \mathbb{P}\Big(\inf_{x\in Incomp}\|Ax\|_2 < \frac{\varepsilon}{\sqrt{n}}\Big)$$

$$\le (1-2e^{-cn}) + C_1\varepsilon.$$

□

Observation 8. *The $2e^{-cn}$ term can be removed by considering*

$$\mathbb{P}\Big(\inf_{x\in Comp}\|Ax\|_2 < c\sqrt{n}\Big)$$

(i.e., $\frac{\varepsilon}{\sqrt{n}} \le c\sqrt{n}$) with care.

Observation 9. *There is a simpler proof for Gaussian matrices [1]. But it does* not *generalize to subgaussian matrices like the results of today's lecture.*

References

[1] Daniel A Sankar. A, Spielman and Shang-Hua Teng. Smoothed analysis of the condition numbers and growth factors of matrices. *SIAM J. Matrix Anal. Appl.*, 28(2):446–476, 2006.