ON THE SINGULARITY OF RANDOM MATRICES WITH INDEPENDENT ENTRIES

LAURENT BRUNEAU AND FRANÇOIS GERMINET

Abstract. We consider n by n real matrices whose entries are non-degenerate random variables that are independent but non necessarily identically distributed, and show that the probability that such a matrix is singular is $O(1/\sqrt{n})$. The purpose of this note is to provide a short and elementary proof of this fact using a Bernoulli decomposition of arbitrary non degenerate random variables.

1. Introduction

Let $M_n = (a_{ij})$ be a random $n \times n$ matrix, where the a_{ij} are independent (non necessarily identically distributed) real random variables. We assume that the r.v. a_{ij} satisfy the following uniform non-degeneracy property

(H) *There exists* $\rho \in]0, \frac{1}{2}[$ *such that for any* $i, j = 1, \cdots, n$, $\mathbb{P}(a_{ij} > x_{ij}^+) > \rho$ *and* $\mathbb{P}(a_{ij} < x_{ij}^-) > \rho$ *for some real numbers* $x_{ij}^- < x_{ij}^+$.

We provide an elementary proof of the following proposition.

Proposition 1. *Let* M_n *be an* $n \times n$ *matrix whose coefficients are independent random variables satisfying* (H). *Then* $\mathbb{P}(M_n \text{ is singular}) = O(1/\sqrt{n})$.

The study of the singularity of random matrices goes back, at least, to Komlós who showed in [Ko1] that $\mathbb{P}(M_n \text{ is singular}) = o(1)$ for independent and identically distributed (iid) Bernoulli entries, namely $a_{ij} = 0, 1$ with probability 1/2. Using Sperner's Lemma, Komlós noticed that the probability was $O(n^{-1/2})$ [B], a result which has been further extended in [Sl] to the case of iid entries equally distributed over a finite set. For iid Bernoulli entries, the conjecture is that $\mathbb{P}(M_n \text{ is singular}) = (c + o(1))^n$ with $c = \frac{1}{2}$. Such an exponential behaviour has been successively obtained and improved in [KKoS, TV1, TV2] up to $c = \frac{3}{4}$. The value $c = \frac{1}{2}$ still seems to be out of reach.

If one turns to general entries, Komlós proved in [Ko2] that $\mathbb{P}(M_n \text{ is singular}) = o(1)$ for independent *and identically distributed* non degenerate random variables. Furthermore, as pointed out by Tao and Vu in [TV1, Section 8], it follows from their analysis that $\mathbb{P}(M_n \text{ is singular}) = o(1)$ for independent non degenerate entries, provided Property (H) holds. Under the same hypothesis Proposition 1 asserts that $\mathbb{P}(M_n \text{ is singular}) = O(n^{-1/2})$.

But the main purpose of this note is to illustrate how the Bernoulli decomposition developed in [AGKW] may be used in order to extend results known for Bernoulli to the general case of independent non degenerate random variables. We perform that illustration by extending Komlós's argument as reproduced in [B], to independent random variables satisfying the Property (H). It is however not clear, at least to the authors, whether results in [TV1, TV2], and in particular Halàsz-type arguments, could be extended in a similar way.

2. Proof

Our approach relies on the following lemma which is essentially contained in [AGKW]. For the reader's convenience we sketch its proof in the appendix.

Lemma 2. *Let M_n be an $n \times n$ matrix whose coefficients are independent random variables satisfying* (H). *We can decompose the entries of the matrix M_n as follows: For all i, j, there exist two independent random variables w_{ij} and ϵ_{ij} and functions $f_{ij} :]0,1[\to \mathbb{R}$ and $\delta_{ij} :]0,1[\to]0,+\infty[$ such that*
1. ϵ_{ij} is a Bernoulli random variable with parameter $p_{ij} \in]0,1[$;
2. w_{ij} has the uniform distribution in $]0,1[$;
3. $a_{ij} = f_{ij}(w_{ij}) + \delta_{ij}(w_{ij})\epsilon_{ij}$.
Moreover, $p_{ij} \in]1-p_0, p_0[$ for all i, j, where $p_0 = 1-\rho$.

Remark 3. *It is of crucial importance for us (see (6) in the proof of Lemma 5) that $\delta_{ij} > 0$. We however do not need here a uniform bound from below on these δ_{ij}. In some situations, one actually does need such a uniform lower bound (see* [AGKW]*), in which case it is sufficient to modify* (H) *above and require the existence of $x_- < x_+$ independent of i, j.*

Thanks to Lemma 2 and since w_{ij} and ϵ_{ij} are independent r.v., we may adopt the following strategy to prove the proposition: 1. do the conditioning with respect to the variables w_{ij}, so that, given the w_{ij}'s, M_n becomes a sum of a constant matrix and of a random matrix with Bernoulli entries with probabilities $(1-p_{ij}, p_{ij})$ and amplitudes $\delta_{ij}(w_{ij})$; 2. estimate, with respect to the Bernoulli variables ϵ_{ij}, the probability that M_n is singular following the strategy of [B]; 3. take the expectation value with respect to the variables w_{ij}.

We shall denote by $\mathbb{P}^{(w)}$ the conditional probability with respect to the w_{ij} variables, i.e. $\mathbb{P}^{(w)}(\cdot) := \mathbb{P}(\cdot|\{w_{ij}\}_{i,j})$.

Following [B], we introduce the strong rank of a system of vectors $S = \{v_1, \cdots, v_n\}$, $\mathrm{sr}(S)$, to be the largest integer k such that any k of the v_j's are linearly independent. For an m by n matrix A, we denote by $\mathrm{sr}_c(A)$ and $\mathrm{sr}_r(A)$ to be, respectively, the strong rank of the system of columns and of rows of A.

The first ingredient of the proof is the following upper bound on the probability for an m by n matrix to have a "not too large" strong rank.

Lemma 4. *Let A be an m by n random matrix whose coefficients a_{ij} satisfy* (H), *and $w = (w_{ij})$ be given. Then*

$$\mathbb{P}^{(w)}(\mathrm{sr}_c(A) < k) \leq \begin{pmatrix} n \\ k \end{pmatrix} \frac{p_0^{m-k+1}}{1-p_0} \quad \text{and} \quad \mathbb{P}^{(w)}(\mathrm{sr}_r(A) < k) \leq \begin{pmatrix} m \\ k \end{pmatrix} \frac{p_0^{n-k+1}}{1-p_0}. \tag{1}$$

Proof. The second statement is clearly equivalent to the first one (applied to A^T). By definition of the strong rank, $\mathrm{sr}_c(A) < k$ if and only if there exists k columns of A which are linearly dependant. It thus suffices to show that for any $1 \leq i_1 < \cdots < i_k \leq n$,

$$\mathbb{P}^{(w)}(\mathrm{rank}\{v_{i_1}, \cdots, v_{i_k}\} < k) \leq \frac{p_0^{m-k+1}}{1-p_0}, \tag{2}$$

where the v_j denote the columns of A. Now we have

$$\begin{aligned}\mathbb{P}^{(w)}&(\text{rank}\{v_{i_1},\cdots,v_{i_k}\}<k) \qquad (3)\\ \leq\ & \mathbb{P}^{(w)}(v_{i_1}=0)+\sum_{j=1}^{k-1}\mathbb{P}^{(w)}(v_{i_{j+1}}\in \text{Span}\{v_{i_1},\cdots,v_{i_j}\}|\text{rank}\{v_{i_1},\cdots,v_{i_j}\}=j).\end{aligned}$$

Let B denote the m by j matrix whose columns are the vectors $v_{i_1},\cdots,v_{i_j}$. If B has rank j, without loss of generality we may decompose it as $B=\begin{pmatrix} C\\ D\end{pmatrix}$ where C is an invertible j by j matrix. In a similar way, let us decompose $v_{i_{j+1}}$ as $v_{i_{j+1}}=\begin{pmatrix} Y\\ Z\end{pmatrix}$, where Y has length j and Z length $m-j$. Note that for w given in the Bernoulli decomposition, the probability of each entry of Z taking a particular value is bounded by p_0.

Then, $v_{i_{j+1}}\in \text{Span}\{v_{i_1},\cdots,v_{i_j}\}$ iff there exists a vector $u=(u_1,\cdots,u_j)^T$ such that $Bu=v_{i_{j+1}}$ and hence iff $Cu=Y$ and $Du=Z$. But since C is invertible we finally get $v_{i_{j+1}}\in \text{Span}\{v_{i_1},\cdots,v_{i_j}\}$ iff $Z=DC^{-1}Y$. Therefore we have

$$\begin{aligned}&\mathbb{P}^{(w)}(v_{i_{j+1}}\in \text{Span}\{v_{i_1},\cdots,v_{i_j}\}|\text{rank}\{v_{i_1},\cdots,v_{i_j}\}=j) \qquad (4)\\ &\quad=\mathbb{E}_Y^{(w)}\left(\mathbb{P}^{(w)}(Z=DC^{-1}Y|Y)\right)\leq \mathbb{E}_Y^{(w)}(p_0^{m-j})=p_0^{m-j}, \qquad (5)\end{aligned}$$

where $\mathbb{E}_Y^{(w)}$ denotes the conditional expectation with respect to the variables w and over the random vector Y. Inserting (4) into (3) and noting that $\mathbb{P}^{(w)}(v_{i_1}=0)\leq p_0^m$, this proves (2). □

The second ingredient of the proof is the following improvement of (2).

Lemma 5. *Let $v_1,\cdots,v_k\in\mathbb{R}^n$ ($k<n$) be linearly independent and $X=(a_1,\cdots,a_n)$ a random vector whose coefficients satisfy* (H). *Suppose that* $\text{sr}_r(A)=s$ *where A is the matrix whose columns are the v_j's. Then $\mathbb{P}^{(w)}(X\in \text{Span}\{v_1,\cdots,v_k\})\leq Cp_0^{n-k-1}/\sqrt{s}$.*

The above lemma relies on the following generalization of the Littlewood and Offord problem to the case of non necessarily identically distributed r.v. and which is an immediate consequence of an extended version of Sperner's lemma (see [AGKW], Lemma 3.2).

Lemma 6. *If $\alpha_1,\cdots,\alpha_s$ are non zero real numbers, $b\in\mathbb{R}$ and $\epsilon_1,\cdots,\epsilon_s$ independent Bernoulli random variables with parameter $p_i\in]1-p_0,p_0[$, then*

$$\mathbb{P}(\alpha_1\epsilon_1+\cdots+\alpha_s\epsilon_s=b)=O(1/\sqrt{s}).$$

Proof of Lemma 5. Let B denote the (n by $k+1$) matrix A augmented with the column vector X, and let $r_1,\cdots,r_n$ denote the rows of B. If $X\in \text{Span}\{v_1,\cdots,v_k\}$ then B has rank k, so that without loss of generality we may assume that $r_1,\cdots,r_k$ are linearly independent, and that the others r_j depend on these. In particular

$$\sum_{i=1}^{k+1}\gamma_i a_i=0,$$

where $\gamma_{k+1} = 1$ and, because $\mathrm{sr}_r(B) \geq \mathrm{sr}_s(A) = s$, at least s of the others γ_i are non-zero. Thus, using Lemma 6, we have, recalling $\delta(w_i) > 0$,

$$\mathbb{P}^{(w)} \left(\sum_{i=1}^{k+1} \gamma_i a_i = 0 \right) = \mathbb{P}^{(w)} \left(\sum_{i=1}^{k+1} \gamma_i \delta(w_i) \epsilon_i = - \sum_{i=1}^{k+1} \gamma_i f(w_i) \right) \leq C/\sqrt{s+1}. \quad (6)$$

Finally, in the same way as in the proof of (4), the a_i for $k+2 \leq i \leq n$ are uniquely determined by $a_1, \cdots, a_k$, and thus each of them has a probability at most p_0 to take a particular value. □

Proof of Proposition 1. By Lemma 2 we have

$$\mathbb{P}(\mathrm{rank}(M_n) < n) = \mathbb{E}_{\{w_{ij}\}_{i,j}} \left(\mathbb{P}^{(w)}(\mathrm{rank}(M_n) < n \right).$$

Let $0 < \beta < \alpha < 1$ to be specified. Let $C_1, \cdots, C_n$ denote the column vectors of M_n and write E_k for the event that $C_1, \cdots, C_k$ are linearly independent and C_{k+1} depends on them. We then have

$$\mathbb{P}^{(w)}(\mathrm{rank}(M_n) < n) \leq \mathbb{P}^{(w)}(\mathrm{sr}_c(M_n) < \alpha n) + \sum_{k=\alpha n}^{n-1} \mathbb{P}^{(w)}(E_k).$$

Indeed, if $\alpha n \leq \mathrm{sr}_c(M_n) < n$ there exists $k \geq \alpha n$ such that $C_1, \cdots, C_k$ are independent but C_{k+1} does depend on them.

Fix now $\alpha n \leq k < n$, and denote by A_k the n by k matrix whose columns are $C_1, \cdots, C_k$. We have then

$$\mathbb{P}^{(w)}(E_k) \leq \mathbb{P}^{(w)}(\mathrm{sr}_r(A_k) < \beta n) + \mathbb{P}^{(w)}(E_k | \mathrm{sr}_r(A_k) \geq \beta n).$$

Using Lemmas 4 and 5, we thus get

$$\begin{aligned}
&\mathbb{P}^{(w)}(\mathrm{rank}(M_n) < n) \\
&\leq \begin{pmatrix} n \\ \alpha n \end{pmatrix} \frac{p_0^{n(1-\alpha)+1}}{1-p_0} + \sum_{k=\alpha n}^{n-1} \left(\begin{pmatrix} n \\ \beta n \end{pmatrix} \frac{p_0^{k-\beta n+1}}{1-p_0} + C p_0^{n-k-1} / \sqrt{\beta n} \right) \\
&\leq \frac{p_0}{1-p_0} \begin{pmatrix} n \\ \alpha n \end{pmatrix} p_0^{(1-\alpha)n} + \frac{p_0}{(1-p_0)^2} \begin{pmatrix} n \\ \beta n \end{pmatrix} p_0^{(\alpha-\beta)n} + \frac{C}{(1-p_0)\sqrt{\beta n}} \\
&\leq C' \left(\mathrm{e}^{n(h(\alpha)+(1-\alpha)\ln p_0)} + \mathrm{e}^{n(h(\beta)+(\alpha-\beta)\ln p_0)} + \frac{1}{\sqrt{n}} \right),
\end{aligned}$$

where $h(x) = -x\ln(x) - (1-x)\ln(1-x)$ is the entropy function and we made used of the Stirling formula to get the last line. It finally suffices to take $0 < \beta < \alpha < 1$ small enough so that $h(\alpha) + (1-\alpha)\ln p_0$ and $h(\beta) + (\alpha - \beta)\ln p_0$ are both strictly negative. □

3. Appendix

For the reader's convenience, we recall the basic material from [AGKW] and show how to extract from (H) the desired uniform estimates on the Bernoulli decomposition. Namely, we prove Lemma 2.

Let X be a real random variable. We denote by μ its law and by F its distribution function: $F(u) = \mu(]-\infty, u])$. We set, for any $t \in]0,1[$,

$$G(t) := \inf\{u, F(u) \geq t\}. \quad (7)$$

Note that

$$G(t) \leq u \iff F(u) \geq t, \quad (8)$$

so that, if t is a random variable with uniform distribution in]0,1[, X and $G(t)$ have the same law ($G(t)$ can be seen as a parametrization of X). For $p \in]0,1[$ given, following [AGKW, Proof of Theorem 2.1], we set for $t \in]0,1[$:

$$\begin{aligned} Y_1(t) &:= G((1-p)t) \\ Y_2(t) &:= G(1-p+pt). \end{aligned} \tag{9}$$

We then let

$$f(t) := Y_1(t), \tag{10}$$

$$\delta(t) := Y_2(t) - Y_1(t), \tag{11}$$

so that, if ϵ is a Bernoulli variable with probabilities $(1-p,p)$ and t a random variable with uniform distribution in $]0,1[$, we have (in law)

$$X = f(t) + \delta(t)\epsilon. \tag{12}$$

That $\inf_{]0,1[} \delta(t) > 0$ is immediate if $Y_2(0) - Y_1(1) = G(1-p+0) - G(1-p) > 0$, which turns out to be the case if X is a Bernoulli itself (choosing p to be its Bernoulli parameter). If X takes at least 3 values, then it is enough to note in full generality that there exists (at least one) $p \in]0,1[$ so that $T_1 > T_2$ where

$$\begin{aligned} T_1 &= \inf\{t \in]0,1[: Y_1(t) = G(1-p)\} && \text{(arrival time of } Y_1\text{)}, \\ T_2 &= \sup\{t \in]0,1[: Y_2(t) = G(1-p+0)\} && \text{(departure time of } Y_2\text{)}. \end{aligned}$$

The latter implies that $\delta(t) > 0$ for all t.

Assume now that X satisfies the estimates of Property (H), with points $x_- < x_+$. We set $p_- = \mu(]-\infty, x_-[)$, $p_+ = \mu(]x_+, +\infty[)$. We show that $p = 1-p_-$ is a possible choice. Thanks to (H), $p \geq p_+ > \rho$, and $1-p = p_- > \rho$, so that $p \in]\rho, 1-\rho[$.

We always have $G(1-p) \leq x_- \leq G(1-p+0)$. If $G(1-p) < x_-$ then $\delta(t) \geq x_- - G(1-p) > 0$. Suppose $G(p_-) = x_-$. We claim that $T_1 = 1 > T_2$. It is easy to see that $T_2 \leq (p-p_+)/p < 1$. It remains to show that $T_1 = 1$. Suppose $T_1 < 1$. For any $t \in]T_1, 1[$ and for any $u < x_-$, one has $x_- = G(p_- t) > u$. Then (8) implies that $F(u) < p_- t$, and thus we get the following contradiction

$$p_- = \mu(]-\infty, x_-[) = \sup_{u < x_-} F(u) \leq p_- t < p_-. \tag{13}$$

Acknowledgement

It is the pleasure of the authors to thank Abel Klein for his comments on a preliminary version of this note. F.G. also thanks Michael Aizenman and Simone Warzel for their warm hospitality in Princeton.

References

[AGKW] M. Aizenman, F. Germinet, A. Klein, S. Warzel, *On Bernoulli decompositions for random variables, concentration bounds and spectral localization*, preprint arXiv:0707.0095v1 [math.PR], to appear in Prob. Th. and Rel. Fields.

[B] B. Bollobas, Random graphs, Academic Press 1985.

[KKoS] J. Kahn, J. Komlós, E. Szemerédi, *On the probability that a random* ± 1 *matrix is singular*, J. Amer. Math. Soc. **8**, 223-240 (1995).

[Ko1] J. Komlós, *On the determinant of* $(0,1)$ *matrices*, Studia. Sci. Math. Hungar. **2**, 7-22 (1967).

[Ko2] J. Komlós, *On the determinant of random matrices*, Studia. Sci. Math. Hungar. **3**, 387-399 (1968).

[Sl] A. Slinko, *A generalization of Komlós's theorem on random matrices*, New Zealand J. Math. **30** no. 1, 81-86 (2001).

[TV1] T. Tao, V. Vu, *On Random* ± 1 *matrices: Singularity and Determinant*, Random Structures and Algorithms **28**, 1-23 (2006).

[TV2] T. Tao, V. Vu, *On the singularity of random Bernoulli matrices*, J. Amer. Math. Soc. **20**, 603-628 (2007).

L.B.: Univ Cergy-Pontoise, CNRS UMR 8088, Département de Mathématiques, F-95000 Cergy-Pontoise.

F.G.: Univ Cergy-Pontoise, CNRS UMR 8088, IUF, Département de Mathématiques, F-95000 Cergy-Pontoise.

E-mail address: `laurent.bruneau@u-cergy.fr`
E-mail address: `francois.germinet@u-cergy.fr`