Law of random determinants

Van H. Vu

Department of Mathematics
Yale University

Joint work with H. Nguyen, UPenn

Let A_n be an n by n random matrix whose entries $a_{ij}, 1 \leq i, j \leq n$, are independent real random variables of zero mean and unit variance. We will refer to the entries a_{ij} as the *atom* variables.

The study of random determinants has a long and rich history. The earliest paper we find on the subject is a paper of Szekeres andTurán from 1937, in which they studied an extremal problem.

1950-60s: Turán, Forsythe and Tukey, Nyquist, Rice, and Riordan, Prékopa, Zurbenko: Compute fixed moments of $\det A_n$.

1950-60s: Turán, Forsythe and Tukey, Nyquist, Rice, and Riordan, Prékopa, Zurbenko: Compute fixed moments of $\det A_n$.

The explicit formula for higher moments get very complicated and in general not available, except in cases when the atom variables have some special distribution (Dembo 90s).

One can use the estimate for the moments and Markov inequality to obtain a upper bound on $|\det A_n|$. However, no lower bound was known for a long time. In particular, Erdös asked whether $\det A_n$ is non-zero with probability tending to one.

One can use the estimate for the moments and Markov inequality to obtain a upper bound on $|\det A_n|$. However, no lower bound was known for a long time. In particular, Erdös asked whether $\det A_n$ is non-zero with probability tending to one.

In 1967, Komlós addressed this question, proving that almost surely $|\det A_n| > 0$ for random Bernoulli matrices (where the atom variables are iid Bernoulli, taking values ± 1 with probability $1/2$). His method also works for much more general models.

One can use the estimate for the moments and Markov inequality to obtain a upper bound on $|\det A_n|$. However, no lower bound was known for a long time. In particular, Erdös asked whether $\det A_n$ is non-zero with probability tending to one.

In 1967, Komlós addressed this question, proving that almost surely $|\det A_n| > 0$ for random Bernoulli matrices (where the atom variables are iid Bernoulli, taking values ± 1 with probability $1/2$). His method also works for much more general models.

The upper bound on the probability $\det A_n = 0$ has been improved several times (Kahn-Komlós-Szemerédi 95, Tao-V. 05-07, Bourgain-V.-Wood 09). On the other hand, these arguments do not reveal further information about the values of $|\det A_n|$.

In 05, Tao-V. proved that for Bernoulli random matrices, with probability tending to one (as n tends to infinity)

$$\sqrt{n!}\exp(-c\sqrt{n\log n}) \leq |\det A_n| \leq \sqrt{n!}\omega(n) \qquad (1)$$

for any function $\omega(n)$ tending to infinity with n. This shows that almost surely, $\log|\det A_n|$ is $(\frac{1}{2}+o(1))n\log n$, but does not provide distributional information. The proof, however, generalizes for other distributions.

In 1968, Goodman studied random matrices with normal entries and observed that the determinant is product of independent χ-square random variables. Thus, $\log|\det A_n|$ is sum of independent random variables.

In 1968, Goodman studied random matrices with normal entries and observed that the determinant is product of independent χ-square random variables. Thus, $\log|\det A_n|$ is sum of independent random variables.

In 1980, Girko claimed the log-normal law

$$\frac{\log(|\det A_n|) - \frac{1}{2}\log(n-1)!}{\sqrt{\frac{1}{2}\log n}} \quad \to \quad \mathbf{N}(0,1). \tag{2}$$

under the additional assumption that the fourth moment of the atom variables is 3.

In 1968, Goodman studied random matrices with normal entries and observed that the determinant is product of independent χ-square random variables. Thus, $\log|\det A_n|$ is sum of independent random variables.

In 1980, Girko claimed the log-normal law

$$\frac{\log(|\det A_n|) - \frac{1}{2}\log(n-1)!}{\sqrt{\frac{1}{2}\log n}} \quad \rightarrow \quad \mathbf{N}(0,1). \qquad (2)$$

under the additional assumption that the fourth moment of the atom variables is 3.

In 1998, he claimed a stronger result which replaced the above assumption by the assumption that the atom variables have bounded $(4+\delta)$th moment.

In 1968, Goodman studied random matrices with normal entries and observed that the determinant is product of independent χ-square random variables. Thus, $\log|\det A_n|$ is sum of independent random variables.

In 1980, Girko claimed the log-normal law

$$\frac{\log(|\det A_n|) - \frac{1}{2}\log(n-1)!}{\sqrt{\frac{1}{2}\log n}} \quad \to \quad \mathbf{N}(0,1). \tag{2}$$

under the additional assumption that the fourth moment of the atom variables is 3.

In 1998, he claimed a stronger result which replaced the above assumption by the assumption that the atom variables have bounded $(4+\delta)$th moment.
No one understands the proofs.

We say that a random variable ξ satisfies condition **C0** (with positive constants C_1, C_2) if

$$P(|\xi| \geq t) \leq C_1 \exp(-t^{C_2}) \tag{3}$$

for all $t > 0$.

Theorem (Main theorem)

Assume that all atom variables a_{ij} satisfy condition **C0** *with some positive constants C_1, C_2. Then*

$$\frac{\log(|\det A_n|) - \frac{1}{2}\log(n-1)!}{\sqrt{\frac{1}{2}\log n}} \quad \xrightarrow{d} \quad \mathbf{N}(0,1). \tag{4}$$

We will actually prove the following equivalent form of (4),

$$\frac{\log(\det A_n^2) - \log(n-1)!}{\sqrt{2\log n}} \xrightarrow{d} \mathbf{N}(0,1). \tag{5}$$

We will actually prove the following equivalent form of (4),

$$\frac{\log(\det A_n^2) - \log(n-1)!}{\sqrt{2\log n}} \xrightarrow{d} \mathbf{N}(0,1). \tag{5}$$

The starting point is the "base times height" formula:

$$\det A_n^2 = \prod_{i=0}^{n-1} \Delta_{i+1}^2. \tag{6}$$

V_i be the subspace generated by the first i rows of A_n; Δ_{i+1} denotes the distance from $\mathbf{a}_{i+1}$ to V_i, where $\mathbf{a}_{i+1} = (a_{i+1,1}, \ldots, a_{i+1,n})$ is the $(i+1)$-th row vector of A_n.

$$\log\det A_n^2 = \sum_{i=0}^{n-1} \log \Delta_{i+1}^2. \tag{7}$$

If the a_{ij} are iid standard gaussian, Δ_{i+1}^2 are independent Chi-square random variables of degree $n-i$. Thus, the right hand side of (7) is a sum of independent random variables.

Notice that Δ_{i+1}^2 has mean $n-i$ and variance $O(n-i)$ and is very strongly concentrated. Thus, with high probability $\log \Delta_{i+1}^2$ is roughly $\log((n-i)+O(\sqrt{n-i}))$ and so it is easy to show that $\log \Delta_{i+1}^2$ has mean close to $\log(n-i)$ and variance $O(\frac{1}{n-i})$. So the variance of $\sum_{i=0}^{n-1} \log \Delta_{i+1}^2$ is $O(\log n)$. To get the precise value $\sqrt{2\log n}$ one needs to carry out some careful (but rather routine) calculation.

With some works, we can make two extra assumptions about M_n: The entries a_{ij} are bounded in absolute value by $\log^{\beta} n$ for some constant $\beta > 0$ and M_n has full rank with probability one.

Theorem

For any constant $\beta > 0$ following holds for any sufficiently large constant $\alpha > 0$. Let A_n be an n by n matrix whose entries $a_{ij}, i \leq n_0, 1 \leq j \leq n$, are independent real random variables of zero mean, unit variance, and bounded in absolute value by $\log^{\beta} n$ for some constant $\beta > 1$. Also, assume that the components of the last $\log^{\alpha} n$ rows of A are independent standard gaussian random variables. Then

$$\frac{\log(|\det A_n|) - \frac{1}{2}\log(n-1)!}{\sqrt{\frac{1}{2}\log n}} \quad \xrightarrow{d} \quad \mathbf{N}(0,1). \tag{8}$$

With some works, we can make two extra assumptions about M_n: The entries a_{ij} are bounded in absolute value by $\log^{\beta} n$ for some constant $\beta > 0$ and M_n has full rank with probability one.

Theorem

For any constant $\beta > 0$ following holds for any sufficiently large constant $\alpha > 0$. Let A_n be an n by n matrix whose entries $a_{ij}, i \leq n_0, 1 \leq j \leq n$, are independent real random variables of zero mean, unit variance, and bounded in absolute value by $\log^{\beta} n$ for some constant $\beta > 1$. Also, assume that the components of the last $\log^{\alpha} n$ rows of A are independent standard gaussian random variables. Then

$$\frac{\log(|\det A_n|) - \frac{1}{2}\log(n-1)!}{\sqrt{\frac{1}{2}\log n}} \xrightarrow{d} \mathbf{N}(0,1). \tag{8}$$

After this, we show that replacing the gaussian rows by the original ones does not change the law.

The distances Δ_i are no longer independent. But we will transform (essentially) the RHS of (7) into a sum of martingale differences.

Theorem (Central limit theorem for martingales; Brown 71)

Assume that $X_1, \ldots, X_m$ are martingale differences with respect to the nested σ-algebra $\mathcal{E}_0, \mathcal{E}_1, \ldots, \mathcal{E}_{m-1}$. Let $v_m^2 := \sum_{i=0}^{m-1} \mathbf{E}(X_{i+1}^2 | \mathcal{E}_i)$, and $s_m^2 := \sum_{i=1}^{m} \mathbf{E}(X_i^2)$. Assume that

(i) $v_m / s_m \to 1$ *in probability;*

(ii) *(Lindeberg condition) for every* $\epsilon > 0$

$$s_m^{-2} \sum_{i=0}^{m-1} \mathbf{E}(X_{i+1}^2 \mathbf{1}_{|X_{i+1}| \geq \epsilon s_m}) \to 0 \text{ as } m \to \infty.$$

Then

$$\frac{\sum_{i=0}^{m-1} X_{i+1}}{s_m} \xrightarrow{d} \mathbf{N}(0,1).$$

Condition on the first i rows $\mathbf{a}_1, \ldots, \mathbf{a}_i$, we can view Δ_{i+1} as the distance from a random vector to $V_i := \mathrm{Span}\,(\mathbf{a}_1, \ldots, \mathbf{a}_i)$. Since M_n has full rank with probability 1, $\dim V_i = i$.

Lemma (Distance Lemma; Tao-V.05)

For any constant $\beta > 0$ there is a constant $C_3 > 0$ depending on β such that the following holds. Assume that $V \subset \mathbf{R}^n$ is a subspace of dimension $\dim(V) \leq n - 4$. Let $\mathbf{a}$ be a random vector whose components are independent variables of zero mean and unit variance and absolute values at most $\log^\beta n$. Denote by Δ the distance from $\mathbf{a}$ to V. Then we have

$$\mathbf{E}(\Delta^2) = n - \dim(V) = n - i$$

and for any $t > 0$

$$\mathbf{P}(|\Delta - \sqrt{n - \dim(V)}| \geq t) = O\big(\exp(-\frac{t^2}{\log^{C_3} n}\big).$$

Set $n_0 := n - \log^\alpha n$ where α is a sufficiently large constant (which may depend on β). We will use short hand k_i to denote $n - i$, the co-dimension of V_i (and the expectation of Δ_i^2). We consider $\sum_i \log \Delta_{i+1}^2 with 0 \leq i < n_0$.

$$\begin{aligned}\log \frac{\Delta_{i+1}^2}{k_i} &= \log(1 + \frac{\Delta_{i+1}^2 - k_i}{k_i}) \\ &= \frac{\Delta_{i+1}^2 - k_i}{k_i} - \frac{1}{2}(\frac{\Delta_{i+1}^2 - k_i}{k_i})^2 + R_{i+1} \\ &:= X_{i+1} - \frac{X_{i+1}^2}{2} + R_{i+1},\end{aligned}$$

where

$$X_{i+1} := \frac{\Delta_{i+1}^2 - k_i}{k_i}, \text{ and } R_{i+1} := \log(1 + X_{i+1}) - (X_{i+1} - \frac{X_{i+1}^2}{2}).$$

Lemma (Small error terms)

The contribution of R_{i+1} is negligible.

Lemma (Main Lemma)

$$\frac{\sum_{0 \le i < n_0} (X_{i+1} - \frac{X_{i+1}^2}{2}) + \log n}{\sqrt{2 \log n}} \xrightarrow{d} \mathbf{N}(0,1).$$

Lemma (Contribution of gaussian rows)

$$\frac{\sum_{i \ge n_0} \log \frac{\Delta_{i+1}^2}{n-i}}{\sqrt{2 \log n}} \xrightarrow{d} 0.$$

A rough upper bound on X_{i+1}:

By Distance Lemma and choosing α sufficiently large, we have with probability at least $1 - O(\exp(-\log^2 n))$ (the probability here is with respect to the random $i+1$th row, fixing the first i rows arbitrarily)

$$|X_{i+1}| = O(k_i^{-3/8}) = O((n-i)^{-3/8}) = o(1). \tag{9}$$

Proof of Main Lemma: Opening

Denote by $P_i = (p_{st}(i))_{s,t}$ the projection matrix onto the orthogonal complement $V_i^{\perp}$.

$$Trace(P_i) = \sum_s p_{ss}(i) = k_i. \tag{10}$$

Also, as P_i is a projection, $P_i^2 = P_i$. Comparing the traces

$$\sum_{s,t} p_{st}(i)^2 = \sum_s p_{ss}(i) = k_i \tag{11}$$

$$X_{i+1} = \frac{\|P_i \mathbf{a}_{i+1}\|^2 - k_i}{k_i} = \frac{\sum_{s,t} p_{st}(i) a_s a_t - k_i}{k_i} := \sum_{s,t} q_{st}(i) a_s a_t - 1,$$

where $a_1, \ldots, a_n$ are the coordinates of the vector $\mathbf{a}_{i+1}$ and

$$q_{st}(i) := \frac{p_{st}(i)}{k_i}.$$

We have $\sum_s q_{ss}(i) = 1$ and $\sum_{s,t} q_{st}^2 = \frac{1}{k_i}$.

Using $\mathbf{E}a_s = 0$ and $\mathbf{E}a_s^2 = 1$, and the fact that a_s are mutually independent:

$$\mathbf{E}(X_{i+1}^2|\mathcal{E}_i) = \frac{2}{k_i} - \sum_s q_{ss}(i)^2(3 - \mathbf{E}{a_s}^4). \qquad (12)$$

$$Y_{i+1} := X_{i+1} - \frac{X_{i+1}^2}{2} + \frac{1}{k_i} - \frac{1}{2}\sum_s q_{ss}(i)^2(3 - \mathbf{E}{a_s}^4)$$

$$Z_{i+1} := \frac{1}{2}\sum_s q_{ss}(i)^2(3 - \mathbf{E}{a_s}^4).$$

The point is that $\mathbf{E}(Y_{i+1}|\mathcal{E}_i) = 0$ and so we can use the CLT for martingale differences to show

$$\frac{\sum_{i<n_0} Y_{i+1}}{\sqrt{2\log n}} \xrightarrow{d} \mathbf{N}(0,1). \tag{13}$$

To complete the proof, we also show that the sum of the Z_i is negligible

$$\frac{\sum_{i<n_0} Z_{i+1}}{\sqrt{2\log n}} \xrightarrow{d} 0. \tag{14}$$

Proof of Main Lemma: Mid game

Main technical tool.

Lemma (Error term)

With probability $1 - O(n^{-100})$ *we have*

$$\sum_{i<n_0} \sum_s q_{ss}(i)^2 = O(\log\log n) = o(\sqrt{2\log n}).$$

Lemma (Delocalization lemma)

For any constant $\beta > 0$ *the following holds for all sufficiently large constant* $\alpha > 0$. *Assume that the components of* $\mathbf{a}_1, \dots, \mathbf{a}_{k_0}$, *where* $k_0 := n - n\log^{-4\alpha} n$, *are independent random variables of mean zero, variance and bounded in absolute value by* $\log^{\beta} n$. *Then with probability* $1 - O(n^{-100})$, *the following holds for all unit vectors* $\mathbf{v}$ *of the space* $V_{k_0}^{\perp}$

$$\|\mathbf{v}\|_\infty = O(\log^{-2\alpha} n).$$

$$S = \sum_{i \le n - n \log^{-4\alpha} n} \sum_s q_{ss}(i)^2 + \sum_{n - n \log^{-\alpha} n \le i < n_0} \sum_s q_{ss}(i)^2$$
$$:= S_1 + S_2.$$

Note that

$$\sum_s q_{ss}(i)^2 \le \sum_{s,t} q_{st}^2(i) = \sum_s \frac{p_{st}^2(i)}{k_i^2} = \frac{1}{k_i} = \frac{1}{(n-i)}.$$

And thus

$$\sum_s q_{ss}(i)^2 \le \frac{1}{(n-i)}.$$

Hence,

$$S_1 \le \sum_{i \le n - n \log^{-4\alpha} n} \sum_s q_{ss}(i)^2 \le \sum_{i \le n - n \log^{-4\alpha} n} \frac{1}{(n-i)} = O(\log \log n).$$

So we have

$$S_1 = o(\sqrt{2 \log n}).$$

To bound S_2, note that

$$p_{ss}(i) = \mathbf{e}_s^T P_i \mathbf{e}_s = \|P_i \mathbf{e}_s\|^2 = |\langle \mathbf{e}_s, \mathbf{v} \rangle|^2$$

for some unit vector $\mathbf{v} \in V_i^{\perp}$.
Thus if $i \geq n - n \log^{-4\alpha} n$, then $V_i^{\perp} \subset V_{k_0}^{\perp}$. By Delocalization Lemma:

$$p_{ss}(i) \leq \|\mathbf{v}\|_\infty^2 = O(\log^{-4\alpha} n). \tag{15}$$

It follows that

$$\begin{aligned} S_2 &\leq \sum_{n - n \log^{-4\alpha} n \leq i < n_0} \max_s p_{ss}(i) \sum_s \frac{p_{ss}(i)}{(n-i)^2} \\ &= O(\log^{-4\alpha} n) \sum \frac{1}{(n-i)} = O(\log^{-4\alpha+1} n). \end{aligned}$$

Proof of Delocalization Lemma.

By the union bound, it suffices to show that $|v_1| = O(\log^{-2\alpha} n)$ with probability at least $1 - O(n^{-101})$, where v_1 is the first coordinate of $\mathbf{v}$.
Let B be the matrix formed by the first k_0 rows $\mathbf{a}_1, \ldots, \mathbf{a}_{k_0}$ of A.
Assume that $\mathbf{v} \in V_{k_0}^{\perp}$, then

$$B\mathbf{v} = 0.$$

Let $\mathbf{w}$ be the first column of B, and B' be the matrix obtained by deleting $\mathbf{w}$ from B. Clearly,

$$v_1 \mathbf{w} = -B'\mathbf{v}', \tag{16}$$

where $\mathbf{v}'$ is the vector obtained from $\mathbf{v}$ by deleting v_1.

Lemma (Singular Values Lemma)

For any constant $\beta > 0$ the following holds for all sufficiently large constant $\alpha > 0$. Let A_n be a random matrix of size n by n, where the entries a_{ij} are independent random variables of mean zero, variance one and bounded in absolute value by $\log^{\beta} n$. Then for any $n/\log^{\alpha} n \leq k \leq n/2$, there exist $2k$ singular values of M in the interval $[0, ck/\sqrt{n}]$, for some absolute constant c, with probability at least $1 - O(n^{-101})$.

Proof. Tao-V. 06; using Guionnet-Zeitouni approach based on Talagrand's inequality.

By the interlacing law and Singular Value Lemma, we conclude that B' has $n - k_0$ singular values in the interval $[0, c(n-k_0)/\sqrt{n}]$ with probability $1 - O(n^{-101})$.

Let H be the space spanned by the left singular vectors of these singular values, and let π be the orthogonal projection on to H. By definition, the spectral norm of $\pi B'$ is bounded,

$$\|\pi B'\| \leq c(n-k_0)/\sqrt{n}.$$

Thus (16) implies that

$$|v_1| \|\pi \mathbf{w}\| \leq c(n-k_0)/\sqrt{n}. \tag{17}$$

On the other hand, since the dimension of H is $n - k_0$, Distance Lemma implies that $\|\pi \mathbf{w}\| \geq \sqrt{n-k_0}/2$ with probability $1 - 4\exp(-(n-k_0)/16) = 1 - O(n^{-\omega(1)})$.

It thus follows from (17) that with the desired prob.

$$|v_1| = O(\log^{-2\alpha} n).$$

Proof of the Main Lemma: End Game

Fact

Let X be a random variable depending on the first $i+1$ rows of A_n such that $|X| = \log^{O(1)} n$ with probability one and conditioned on the first i rows $|X| \le f$ with probability at least $1 - n^{-50}$. Then $|\mathbf{E}(X|\mathcal{E}_i)| \le f + n^{-50} \log^{O(1)} n$ with probability one. Consequently $|\mathbf{E}X| \le f + n^{-50} \log^{O(1)} n$.

$$Y_{i+1} := X_{i+1} - \frac{X_{i+1}^2}{2} + \frac{1}{k_i} - \frac{1}{2}\sum_s q_{ss}(i)^2(3 - \mathbf{E}{a_s}^4)$$

Consider $\mathbf{E}(Y_{i+1}^2|\mathcal{E}_i)$. Expand Y_{i+1}^2 into sum of $4^2 = 16$ terms. The dominating term will be $\mathbf{E}(X_{i+1}^2|\mathcal{E}_i)$. By Lemma 8, we have with probability $1 - n^{-100}$ that

$$\sum_{i<n_0} \mathbf{E}(X_{i+1}^2|\mathcal{E}_i) = \sum_{i<n_0} \frac{2}{k_i} + O(\sum_{i<n_0}\sum_s q_{ss}(i)^2) = 2\log n + O(\log\log n). \tag{18}$$

Using Fact 11, we have

$$\sum_{i<n_0} \mathbf{E}(X_{i+1}^2) = 2\log n + O(\log\log n). \tag{19}$$

We now show that the contribution of all other terms is negligible by combining (9), Error term Lemma and Fact 11.
Consider X_{i+1}^3. By (9) we have that for any first i rows, with probability $1 - \exp(-\Omega(\log^2 n))$

$$|X_{i+1}|^3 = O(k_i^{-9/8}).$$

By Fact 11, we conclude that with probability one $\mathbf{E}(X_{i+1}^3|\mathcal{E}_i) = O(k_i^{-9/8})$ and $\mathbf{E}X_{i+1}^3 = O(k_i^{-9/8})$. Since $j^{-9/8}$ is summable, the contribution of X_{i+1}^3 in both v_m and s_m is only $O(1)$ and negligible.

Consider $X_{i+1}^2 \sum_s q_{ss}(i)^2(3 - \mathbf{E}(a_s^4(i))$. Again by (9), with probability $1 - \exp(-\Omega(\log^2 n))$ (conditioned on the first i rows)

$$|X_{i+1}^2 \sum_s q_{ss}(i)^2(3 - \mathbf{E}(a_s^4(i))| = O(k_i^{-6/8} \sum_s q_{ss}(i)^2).$$

By Fact 11 with probability one

$$\mathbf{E}(X_{i+1}^2 \sum_s q_{ss}(i)(3 - \mathbf{E}(a_s^4(i))|\mathcal{E}_i) = O(k_i^{-6/8} \sum_s q_{ss}(i)^2).$$

Since k_i is at least $\log^\alpha n$, we have

$$\sum_{i<n_0} |\mathbf{E}(X_{i+1}^2 \sum_s q_{ss}(i)^2(3-\mathbf{E}(a_s^4(i))|\mathcal{E}_i)| = O(\log^{-\frac{6}{8}\alpha} n \sum_i \sum_s q_{ss}^2(i)).$$

By Error term Lemma, this sum is, with probability $1 - n^{-100}$,

$$O(\log^{-\frac{6}{8}\alpha} n \log\log n) = o(1).$$

So $\sum_i |\mathbf{E}(X_{i+1}^2 \sum_s q_{ss}(i)(3 - \mathbf{E}(a_s^4(i))| = o(1)$

Lindeberg condition. Recall that $s_m = (1 + o(1))\sqrt{2 \log n}$; the definition of Y_{i+1} implies that if $|Y_{i+1}| \geq \epsilon s_m$ holds then one of the following three events must hold

$$|X_{i+1}| \geq \frac{\epsilon}{4}\sqrt{\log n}; |X_{i+1}|^2 \geq \frac{\epsilon}{4}\sqrt{\log n}; \sum_s q_{ss}(i)^2 \geq \frac{\epsilon}{4}\sqrt{\log n}.$$

The third event holds with probability 0 as $\sum_s q_{ss}^2 \leq \sum_{s,t} q_{st}^2 = \frac{1}{k_i} < 1$. Furthermore, the first event implies the second one. On the other hand, by (9), this event holds with probability at most $1 - \exp(-\log^2 n)$. Thus, we can conclude that

$$\mathbf{P}(|Y_{i+1}| \geq \epsilon s_m) \leq \exp(-\log^2 n).$$

Since $|X_{i+1}| = O(\|\mathbf{a}_{i+1}\|^2) = O(n \log^{O(1)} n)$ with probability one, it follows that $|Y_{i+1}| = O(n^2 \log^{O(1)} n)$ with probability one. Therefore

$$\sum \mathbf{E}(Y_{i+1}^2 \mathbf{I}_{|Y_{i+1}| \geq \epsilon s(m)}) \leq n^5 \exp(-\log^2 n) \log^{O(1)} n = o(1).$$

Open questions.

Rate of convergence ?

Hermitian case ? (some work in progress with Tao.)

Permanent ? (not clear even in the gaussian case)