Linear Algebra and its Applications 434 (2011) 1319–1324

Contents lists available at ScienceDirect

Linear Algebra and its Applications

journal homepage: www.elsevier.com/locate/laa

Natural density of rectangular unimodular integer matrices ☆

Gérard Maze, Joachim Rosenthal *, Urs Wagner

Mathematics Institute, University of Zürich, Winterthurerstr 190, CH-8057 Zürich, Switzerland

ARTICLE INFO

Article history:
Received 26 August 2010
Accepted 7 November 2010
Available online 4 December 2010

Submitted by H. Schneider

AMS classification:
15B36
11C20

Keywords:
Natural density
Unimodular matrices
Cesàro's Theorem
Quillen–Suslin's Theorem

ABSTRACT

In this paper, we compute the natural density of the set of $k \times n$ integer matrices that can be extended to an invertible $n \times n$ matrix over the integers. As a corollary, we find the density of rectangular matrices with Hermite normal form $[O_{k\times(n-k)}\ I_k]$. Connections with Cesàro's Theorem on the density of coprime integers and Quillen–Suslin's Theorem are also presented.

1. Introduction and main result

Given a commutative ring R with 1, the notion of invertible $n \times n$ matrix is well defined, and can be characterized by the condition that the determinant of such a matrix is a unit in R. Given a $k \times n$ matrix A, the question of whether it can be completed by an $(n-k) \times n$ matrix into an $n \times n$ invertible matrix over R has raised several interesting problems in the past. For instance, the celebrated Quillen–Suslin Theorem [14,15], previously known under the name Serre Conjecture, deals with the case when $R = \mathbb{F}[x_1, \ldots, x_l]$, with $\mathbb{F}$ a field and $k = 1$. It can be shown that this case also contains in essence the general case $1 \leqslant k \leqslant n$, see [18]. The theorem states that over this ring, the following three properties are equivalent [14,15,18]:

1. A can be completed into an $n \times n$ invertible matrix,
2. there exists an $n \times k$ matrix B such that $AB = I_k$, where I_k is the $k \times k$ identity matrix,
3. the $k \times k$ minors of A have no common zeros.

☆ Partially supported by SNF grant No. 121874 and Armasuisse.

* Corresponding author.
E-mail addresses: gmaze@math.uzh.ch (G. Maze), rosen@math.uzh.ch (J. Rosenthal), urs.wagner@math.uzh.ch (U. Wagner).

0024-3795/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2010.11.015

When the ring R is a PID it is a direct consequence of the Smith normal form (see e.g. [11]) that Conditions 1. and 2. are again equivalent and these conditions are equivalent to the fact that the gcd of the $k \times k$ minors is equal to 1.

In the sequel we will adopt the usual convention and call a $k \times n$ matrix A over some ring R *unimodular* as soon as A can be extended to $n \times n$ invertible matrix.

Over the ring of integers various related results on unimodular matrices are known in the literature. E.g. Zhan [19] showed that any partial $n \times n$ matrix with n given entries not lying on the same row or column can be completed into a unimodular matrix. Another result is due to Fang [6] who showed that, if the diagonal of a square matrix is let free, then it can be completed into a unimodular one.

Our focus in this paper will be on the "probability" that a random $k \times n$ integer matrix is unimodular. Related to our problem is a classical result due to Cesàro [1–3] (see also [16,17] and the historical remarks below) which states that the "probability" that two randomly chosen integers are coprime is $\frac{1}{\zeta(2)} = \frac{6}{\pi^2}$, where ζ denotes Riemann's zeta function. A re-statement of Cesàro's result is then: the probability that a random 1×2 integer matrix is unimodular is $\zeta(2)^{-1}$.

In order to make the notion of probability precise we first remark that the uniform distribution over the set $\mathbb{Z}^m$ has little meaning. For this reason researchers often use the concept of *natural density* when stating probability results in $\mathbb{Z}^m$ or more general infinite modules R^m. In the following we briefly explain this concept. Let $S \subset \mathbb{Z}^m$ be a set. Define the upper (respectively lower) natural density as

$$\overline{\mathbb{D}}(S) = \limsup_{B\to\infty} \frac{|S \cap [-B, B[^m|}{(2B)^m}, \qquad \underline{\mathbb{D}}(S) = \liminf_{B\to\infty} \frac{|S \cap [-B, B[^m|}{(2B)^m}.$$

When both limits are equal one defines the natural density of the set S as:

$$\mathbb{D}(S) := \overline{\mathbb{D}}(S) = \underline{\mathbb{D}}(S). \tag{1}$$

The following properties of natural density are readily verified: If S^c denotes the complement of S then $\mathbb{D}(S^c) = 1 - \mathbb{D}(S)$, whenever S has a well-defined density. Similarly if $\{S_i\}_{i\in I}$ is a set of subsets of $\mathbb{Z}^m$ with well defined densities $\mathbb{D}(S_i)$ and if $S = \cup_{i\in I} S_i$, then:

$$\overline{\mathbb{D}}(S) \leqslant \sum_{i\in I} \mathbb{D}(S_i). \tag{2}$$

Readers interested in more background on the notion of natural densities of sets of integer matrices are referred to [8], even though this paper attributes the result of Cesàro to Mertens, something which was probably triggered by a remark in [7].

In fact the exact fatherhood of the result of Cesàro appears to be inexactly described in several occasions in the literature. The story behind this historical misunderstanding seems to be the following. The problem of evaluating the probability that two random integers are coprime appears in a 1881 question raised by Cesàro [1]. Two years later, Sylvester [16] and Cesàro [2] independently publish their solutions. Interestingly, two proofs are presented in [16]: Sylvester's own argument is based on Farey series, and a more "probabilistic" argument of Franklin is also presented with his permission. In the footnote of a 1888 paper [17], Sylvester publishes a similar proof, and mentions that Cesàro "claimed the prior publication" of the result. Sylvester's argument is based on Farey series, and the remark in [7] makes a connection between an earlier work of Mertens [12] of 1874 on the average value of Euler φ function and Farey series, which probably triggered the remark in [8] but at no point in [7] the result is attributed to Mertens. If we want to associate the aforementioned probability with the average value of φ, then it is legitimate to go back to an 1849 paper of Dirichlet [5] where the value $\frac{6}{\pi^2}$ appears for the first time. The case of k coprime integers, $k > 2$, is also presented for first time by Cesàro in 1884 [3]. The result is rediscovered in 1900 by Lehmer [9], apparently independently (see also [13]).

The major result of our paper is a matrix version of Cesàro's theorem:

Proposition 1. *Let $1 \leqslant k < n$ be integers. The natural density $\mathrm{d}_{k,n}$ of $k \times n$ unimodular matrices with integer entries is given by*

$$\mathrm{d}_{k,n} = \left(\prod_{j=n-k+1}^{n} \zeta(j) \right)^{-1}.$$

In other words, the "probability" that a "random" $k \times n$ integer matrix can be extended into a matrix in $\mathrm{GL}_n(\mathbb{Z})$ is given by $(\zeta(n) \cdot \cdots \cdot \zeta(n-k+1))^{-1}$. Since $\zeta(n)$ converges rapidly towards 1, if $k = n - d$, the above density converges rather fast to the limit d_d with

$$\mathrm{d}_1 = \left(\prod_{j=2}^{\infty} \zeta(j) \right)^{-1} = 0.43575707677\ldots \quad \text{and} \quad \mathrm{d}_2 = \zeta(2) \cdot \mathrm{d}_1, \ \mathrm{d}_3 = \zeta(2) \cdot \zeta(3) \cdot \mathrm{d}_1 \cdots$$

When n is not too small, say $n > 4$, d_1 is a good approximation of the proportion of integer matrices that can be completed by a row into an invertible matrix. We would like to point out that the proof of the above proposition is independent from the result of Cesàro and as such a new proof of his theorem is given when the above proposition is considered with $k = 1$. Proposition 1 above gives a probabilistic extension of the simplest case of the Quillen–Suslin Theorem. The concept of natural density does not extend naturally to the ring $\mathbb{F}[x_1, \ldots, x_n]$ and thus the existence of a direct extension of our result to the general case is unclear. In the next section we will prove Proposition 1.

2. Proof of the main result

Let us fix some notations for the rest of the article. Let $1 \leqslant t \leqslant k < n$ be integers. Given a $k \times n$ matrix A with integer entries, the determinants of the $\binom{n}{t}\binom{k}{t}$ $t \times t$ submatrices of A formed by the intersection of any subset of t column and t row vectors of A are called the t-minors of A. When $t = k$, they are called the full rank minors of A. The set of primes is denoted by $\mathbb{P}$ and when no confusion is possible, the set $\mathbb{Z}^{k\times n}$ will be identified with $\mathbb{Z}^{kn}$.

The strategy of the proof of the above proposition is to localize the computation of the density at every prime, and then lift the information up in order to extract the exact density over $\mathbb{Z}^{kn}$. The proof of the above proposition relies on the next lemmas. As noted before, the PID version of Quillen–Suslin's Theorem gives directly the following lemma:

Lemma 2. *A $k \times n$ matrix A with integer entries is unimodular if and only if the full-rank minors of A are coprime.*

Over a field $\mathbb{F}$, the rank of a $k \times n$ matrix, i.e., the maximal number of rows that are linearly independent over $\mathbb{F}$, is equal to the largest integer t such that there exists a non-zero $t \times t$ minor. A classical computation shows that over the prime finite field $\mathbb{F} = \mathbb{Z}/p\mathbb{Z}$, if F_p is the set of full rank $k \times n$ matrices, its cardinality satisfies $|F_p| = \prod_{j=0}^{k-1}(p^n - p^j)$. See [10] for the details. This can be extended as follows:

Lemma 3. *Let S be a finite set of prime numbers. The density of $k \times n$ matrices with integer entries for which the* gcd *of the full rank minors are coprime to all primes in S is given by*

$$\prod_{j=n-k+1}^{n} \prod_{p\in S} \left(1 - \frac{1}{p^j}\right).$$

Proof. Let us call $E_S \subset \mathbb{Z}^{kn}$ the set of $k\times n$ matrices for which the gcd of the full rank minors are coprime to all primes in S. Let $N = \prod_{p\in S} p$ and recall the Chinese remainder theorem $(\mathbb{Z}/N\mathbb{Z}) \cong \prod_{p\in S} (\mathbb{Z}/p\mathbb{Z})$. Let B be an integer that will go to infinity in the sequel. Write $B = qN + r$, with $q, r \in \mathbb{N}$, $0 \leqslant r < N$

and consider the map ϕ obtained as the composition of maps

$$[-qN, qN[^{kn} \longrightarrow (\mathbb{Z}/N\mathbb{Z})^{kn} \longrightarrow \prod_{p\in S} (\mathbb{Z}/p\mathbb{Z})^{kn},$$

where the first map is the quotient modulo N and the second is the induced homomorphism given by the Chinese remainder theorem. Because of the above remark we have

$$\phi(E_S \cap [-qN, qN[^{kn}) = \prod_{p\in S} F_p.$$

The first map is a $(2q)^{kn}$-to-1 map, i.e., each fiber contains exactly $(2q)^{kn}$ elements, and the second map is an isomorphism. Thus

$$|E_S \cap [-qN, qN[^{kn}| = (2q)^{kn} \cdot \prod_{p\in S} |F_p|.$$

We have the disjoint union $[-B, B[^{kn} = [-qN, qN[^{kn} \sqcup \left([-B, B[^{kn} \backslash [-qN, qN[^{kn}\right)$. The difference of hypercubes has a volume bounded by the volume of $2kn$ hyper-rectangles of side area B^{kn-1} and height r which gives $0 \leqslant |[-B, B[^{kn} \backslash [-qN, qN[^{kn}| < 2knrB^{kn-1}$ and therefore we have

$$|E_S \cap [-B, B[^{kn}| = |E_S \cap [-qN, qN[^{kn}| + \rho$$

with $0 \leqslant \rho < 2knrB^{kn-1}$. Thus

$$\begin{aligned}\frac{|E_S \cap [-B, B[^{kn}|}{(2B)^{kn}} &= \frac{|E_S \cap [-qN, qN[^{kn}|}{(2qN)^{kn}} \frac{(2qN)^{kn}}{(2B)^{kn}} + \frac{\rho}{(2B)^{kn}} \\ &= \prod_{p\in S} \frac{|F_p|}{p^{nk}} \cdot \left(1 - \frac{r}{B}\right)^{kn} + O(1/B).\end{aligned}$$

Finally, we have

$$\mathbb{D}(E_S) = \lim_{B\to\infty} \prod_{p\in S} \frac{|F_p|}{p^{nk}} \cdot \left(1 - \frac{r}{B}\right)^{kn} + O(1/B) = \prod_{p\in S} \frac{|F_p|}{p^{nk}} = \prod_{j=n-k+1}^{n} \prod_{p\in S} \left(1 - \frac{1}{p^j}\right). \quad \square$$

Extending the definition of E_S, let us call $E_t \subset \mathbb{Z}^{kn}$ the set of $k \times n$ matrices for which the gcd of the full rank minors are coprime with the first t primes $2, 3, \ldots, p_t$. Note that the sequence E_t is a decreasing sequence of sets, i.e., $E_i \subset E_j$ if $i \geqslant j$ and if $E = \cap_{t\in\mathbb{N}} E_t$, then Lemma 2 implies that E is the set of $k \times n$ unimodular matrices with integer entries. In order to prove Proposition 1, we have to prove that $\mathbb{D}(E)$ exists and compute its value $\mathrm{d}_{k,n} = \mathbb{D}(E)$. Since we know $\mathbb{D}(E_t)$ for all t, it is tempting to prove the proposition by simply letting t going to infinity in the expression given by Lemma 3, but this is an invalid argument in general. Indeed the example of $E_t = [t, \infty[$ with $E = \emptyset$ shows that it is possible to have a sequence of sets $E_i \subset E_j$ if $i \geqslant j$ with $E = \cap_{t\in\mathbb{N}} E_t$ and $\mathbb{D}(E) = 0 \neq 1 = \lim_{t\to\infty} \mathbb{D}(E_t)$ since $\mathbb{D}(E_t) = 1, \forall t$. The next lemma describes how to avoid this pathological case. Let us recall that for any real sequences a_n, b_n, $\liminf a_n + \liminf b_n \leqslant \liminf(a_n + b_n)$ and $\limsup(a_n + b_n) \leqslant \limsup a_n + \limsup b_n$ and $\limsup -a_n = -\liminf a_n$.

Lemma 4. *Let E_t be a sequence of decreasing sets in $\mathbb{Z}^m$ such that $\mathbb{D}(E_t)$ exists for all t and converges to* d. *Let $E = \cap_{t\in\mathbb{N}} E_t$. If $\lim_{t\to\infty} \overline{\mathbb{D}}(E_t \setminus E) = 0$, then $\mathbb{D}(E)$ exists and is equal to* d.

Proof. We will use the disjoint union $E_t = E \sqcup (E_t \setminus E)$. Since

$$|E_t \cap [-B, B[^m| = |E \cap [-B, B[^m| + |(E_t \setminus E) \cap [-B, B[^m|,$$

we have

$$\frac{|E \cap [-B, B[^m|}{(2B)^m} = \frac{|E_t \cap [-B, B[^m|}{(2B)^m} + \frac{-|(E_t \setminus E) \cap [-B, B[^m|}{(2B)^m}.$$

Taking the lim inf and the lim sup, and since $\lim_{B\to\infty} \frac{|E_t \cap [-B,B[^m|}{(2B)^m} = \mathbb{D}(E_t)$, we have

$$\mathbb{D}(E_t) - \overline{\mathbb{D}}(E_t \setminus E) = \underline{\mathbb{D}}(E) \leqslant \overline{\mathbb{D}}(E) = \mathbb{D}(E_t) - \underline{\mathbb{D}}(E_t \setminus E).$$

The result follows when $t \to \infty$ since $0 \leqslant \underline{\mathbb{D}}(E_t \setminus E) \leqslant \overline{\mathbb{D}}(E_t \setminus E) \longrightarrow 0$. □

Proof of Proposition 1. We will use the previous lemma. The set $E_t \setminus E$ is the set of $k \times n$ matrices A for which there exists a prime p with $p > p_t$ so that p divides the gcd of the full rank minors of A. For each prime p, let us define H_p to be the set of $k \times n$ matrices whose gcd of the full rank minors is divisible by p. Then $E_t \setminus E = \cup_{p>p_t} H_p$. Note that Lemma 3 applied to $S = \{p\}$ implies that H_p has a density equal to

$$\mathbb{D}(H_p) = 1 - \mathbb{D}(H_p^c) = 1 - \prod_{j=n-k+1}^{n} \left(1 - \frac{1}{p^j}\right).$$

By induction on the number of factors, one readily verifies that for real numbers $0 < x_j < 1$, we have the inequality $\prod_{j=0}^{n}(1 - x_j) > 1 - \sum_{j=0}^{n} x_j$, which applied to the above product gives

$$\mathbb{D}(H_p) = 1 - \prod_{j=n-k+1}^{n} \left(1 - \frac{1}{p^j}\right) < \sum_{j=n-k+1}^{n} \frac{1}{p^j} < \frac{1}{p^{n-k}(p-1)} < \frac{2}{p^2}.$$

Finally, we have

$$\overline{\mathbb{D}}(E_t \setminus E) = \overline{\mathbb{D}}(\cup_{p>p_t} H_p) \leqslant \sum_{p>p_t} \mathbb{D}(H_p) < \sum_{p>p_t} \frac{2}{p^2}$$

which shows that $\lim_{t\to\infty} \overline{\mathbb{D}}(E_t \setminus E) = 0$ since the last series is the tail of the convergent series $\sum_{p\in\mathbb{P}} \frac{2}{p^2}$. We can therefore apply the previous lemma and conclude that $\mathrm{d}_{k,n} = \mathbb{D}(E)$ exists and is equal to

$$\mathrm{d}_{k,n} = \mathbb{D}(E) = \prod_{j=n-k+1}^{n} \prod_{p\in\mathbb{P}} \left(1 - \frac{1}{p^j}\right) = \prod_{j=n-k+1}^{n} \zeta(j)^{-1}. \quad \square$$

3. Concluding remarks and further results

Proposition 1 does not cover the square case $k = n$. Indeed, the zeta function is not defined when $n - k + 1 = 1$ since ζ has a pole of order 1 at $x = 1$. The following result covers the square case.

Lemma 5. *The natural density of $n \times n$ unimodular matrices is* $\mathrm{d}_{n,n} = 0$.

Before we give a proof we remark that the product formula in Proposition 1 naturally has an extension to zero as $\lim_{x\to 1}(\zeta(x))^{-1} = 0$.

Proof. Each $n \times n$ matrix with $n^2 - 1$ entries in the range $[-B, B[$ can be completed by at most two values in order for this matrix to be unimodular, due to the Lagrange expansion of the determinant, which must be ± 1. As such there are at most $2(2B)^{n^2-1}$ unimodular matrices with entries in $[-B, B[$. The conclusion follows since $\mathrm{d}_{n,n} \leqslant \lim_{B\to\infty} 2(2B)^{n(n-1)}/(2B)^{n^2} = 0$. □

The result of Proposition 1 can also be used in the determination of the natural density of $k \times n$ matrices whose Hermite normal form (HNF) is very simple. Recall that the HNF of a $k \times n$ matrix A is the unique $k \times n$ matrix H of the following form

$$\begin{pmatrix} 0 & 0 & \dots & 0 & h_1 & h_{1,2} & \dots & h_{1,n-k} \\ 0 & 0 & \dots & 0 & 0 & h_2 & \dots & h_{2,n-k} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & h_k \end{pmatrix},$$

where $0 \leqslant h_{i,j} < h_i$, such that there exists $U \in \mathrm{Gl}_n(\mathbb{Z})$ with $A = HU$, see e.g. [4]. We have then the following result.

Theorem 6. *The density of $k \times n$ matrices whose Hermite normal form is the block matrix $[O_{k\times(n-k)}\ I_k]$ is $\mathrm{d}_{k,n}$.*

Proof. The result follows from the fact that the set E of $k \times n$ unimodular matrices over $\mathbb{Z}$ is the equal to the set L of $k \times n$ matrices whose HNF is the block matrix $[O_{k\times(n-k)}\ I_k]$. Let us show that $E \subset L$. If $A \in E$ then there exists a unimodular square matrix B such that A consists in the last k rows of B. The HNF of B is the identity matrix I_n. Indeed, it is a upper diagonal square matrix with integer entries whose determinant is 1, which forces the diagonal entries to be 1. The size condition of the coefficient of the HNF forces the entries above the diagonal to be 0. Thus we have $BU^{-1} = I_n$ for some $U \in \mathrm{Gl}_n(\mathbb{Z})$, which gives $AU^{-1} = [O_{k\times(n-k)}\ I_k]$. The uniqueness of the HNF shows that $A \in L$. Let us show now that $L \subset E$ by showing that the gcd of the $k \times k$ minors of a matrix in L are coprime. The equation $A = [O_{k\times(n-k)}\ I_k]U$ shows that the $k \times k$ minors of A are equal to the $k \times k$ minors of $[O_{k\times(n-k)}\ I_k]$ which are 0 or 1, and thus coprime. This finishes the proof of the corollary. □

Taking into account that the Smith normal form of a matrix [4] is obtained from the Hermite normal form via row operations, an immediate consequence of this theorem is:

Corollary 7. *The density of $k \times n$ matrices whose Smith normal form is the block matrix $[O_{k\times(n-k)}\ I_k]$ is $\mathrm{d}_{k,n}$.*

References

[1] E. Cesàro, Question proposée 75, Mathesis 1 (1881) 184.
[2] E. Cesàro, Question 75 (Solution), Mathesis 3 (1883) 224–225.
[3] E. Cesàro, Probabilité de certains faits arithméthiques, Mathesis 4 (1884) 50–151.
[4] H. Cohen, A course in computational algebraic number theory, Springer Graduate Texts in Mathematics, 1995.
[5] P.G.L. Dirichlet, Über die bestimmung der mittleren Werthe in der Zahlentheorie, Abhand. Ak. Wiss. Berlin, 1849, pp. 63–83. Reprinted in Werke, vol. 2, pp. 49–66.
[6] M. Fang, On the completion of a partial integral matrix to a unimodular matrix, Linear Algebra Appl. 422 (1) (2007) 291–294.
[7] G.H. Hardy, E.M. Wright, An Introduction to the Theory of Numbers, fifth ed., Oxford University Press, 1979.
[8] A.J. Hetzel, J.S. Liew, K.E. Morrison, The probability that a matrix of integers is diagonalizable, Amer. Math. Monthly 114 (6) (2007) 491–499.
[9] D.N. Lehmer, Asymptotic evaluation of certain totient sums, Amer. J. Math. 22 (4) (1900) 293–335.
[10] R. Lidl, H. Niederreiter, Finite fields, second ed., University Press, Cambridge, 1997.
[11] C.C. MacDuffie, The Theory of Matrices, Chelsea Publ. Co., New York, 1946.
[12] F. Mertens, Über einige asymptotische Gesetze der Zahlentheorie, J. Reine Angew. Math. 77 (1874) 46–62.
[13] J.E. Nymann, On the probability that k positive integers are relatively prime, J. Number Theory 7 (1975) 406–412.
[14] D. Quillen, Projective modules over polynomial rings, Invent. Math. (36) (1976) 167–171.
[15] A.A. Suslin, Projective modules over polynomial rings are free, Soviet Math. 4 (17) (1976) 1160–1164.
[16] J.J. Sylvester, Sur le nombre de fractions ordinaires inégales qu'on peut exprimer en se servant de chiffres qui n'excède pas un nombre donné, C. R. Acad. Sci. Paris XCVI (1883) 409–413, Reprinted in H.F. Baker (Ed.), The Collected Mathematical Papers of James Joseph Sylvester, vol. 4, Cambridge University Press, p. 86..
[17] J.J. Sylvester, On certain inequalities relating to prime numbers, Nature 38 (1888) 259–262.
[18] D.C. Youla, P.F. Pickel, The Quillen–Suslin theorem and the structure of n-dimensional elementary polynomial matrices, IEEE Trans. Circuits Systems 31 (6) (1984) 513–518.
[19] X. Zhan, Completion of a partial integral matrix to a unimodular matrix, Linear Algebra Appl. 414 (1) (2006) 373–377.