

LU FACTORIZATION OF DISCRETE RANDOM MATRICES

SAMUEL ORELLANA MATEO, JOHN URSCHTEL, AND NICHOLAS WEST

ABSTRACT. We consider the probability that a discrete random matrix $M_n(\xi)$ is *strongly non-singular*, meaning all its leading principal submatrices are non-singular. This property is equivalent to the existence of an LU factorization. We show that for any discrete random variable ξ with finite support and $|\xi|_\infty < 1$, there is a constant probability that $M_n(\xi)$ is strongly non-singular with a growth factor bounded by $n^{5/2+\delta}$. Furthermore, we provide a tight asymptotic lower bound for this probability as $|\xi|_\infty \rightarrow 0$. Finally, we provide exact counts for strongly non-singular binary matrices up to $n = 9$ and use these to derive improved upper bounds for the Bernoulli case.

1. INTRODUCTION

Gaussian elimination, the process in which a matrix is factored into the product of a lower and upper triangular matrix, is the oldest and most-used algorithm for solving a general linear system. The resulting factorization $A = LU$, where L is lower unitriangular and U is upper triangular, is unique. However, an LU factorization may not exist if a leading principal minor (excluding the determinant of the matrix itself) of A is zero. In this case, a permutation of the rows and/or columns of A or a suitable pre- and/or post-multiplication is needed, increasing computational costs. In addition, when performing Gaussian elimination in finite precision, round-off error can accumulate. The stability of the Gaussian elimination algorithm in finite precision is governed by the growth factor, defined by

$$\text{growth}(A) = \max \left\{ \|L\|_{\max}, \frac{\|U\|_{\max}}{\|A\|_{\max}} \right\},^1$$

where $\|\cdot\|_{\max}$ is the entrywise matrix infinity norm. A matrix is said to be *strongly non-singular* if all of its leading principal submatrices are non-singular. The existence of an LU factorization without pivoting for an $n \times n$ matrix is equivalent to the leading $(n-1) \times (n-1)$ principal submatrix being strongly non-singular.

In this paper, we consider an average case analysis of Gaussian elimination for discrete matrices. In particular, for a discrete random variable ξ , we consider the random matrix A with entries sampled independently from ξ and estimate the probability that A has an LU factorization (essentially, the probability that it is strongly non-singular) and the probability that A has a stable LU factorization in finite precision (i.e., a polynomially bounded $\text{growth}(A)$). This analysis is closely connected to two seemingly disjoint areas of research, the numerical stability of Gaussian elimination and the singularity problem for discrete matrices, which we detail below.

DEPARTMENT OF MATHEMATICS, DUKE UNIVERSITY, DURHAM, NORTH CAROLINA, 27708 USA.

DEPARTMENT OF MATHEMATICS, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, CAMBRIDGE, MA, 02139 USA.

E-mail addresses: samuel.orellanamateo@duke.edu, urschel@mit.edu, npwest00@mit.edu.

2020 *Mathematics Subject Classification.* 15A23, 15B52, 65F05.

Key words and phrases. Gaussian elimination, growth factor, random matrices.

¹The exact definition of the growth factor varies by source. The quantity $\max_k \|A^{(k)}\|_{\max}/\|A\|_{\max}$ and the pair $\|L\|_\infty$ and $\|U\|_\infty/\|A\|_\infty$ are also popular, with the former implicitly assuming a pivoting strategy that produces bounded entries in L . They are all equivalent in that they can all be used to bound the backward error of the LU factorization.

1.1. Numerical stability of Gaussian elimination. The numerical stability of Gaussian elimination has a long history, beginning, most notably, with the work of von Neumann and Goldstine [28] and later the work of Wilkinson [30, 31]. Wilkinson was the first to recognize how the numerical stability of Gaussian elimination is related to the growth factor, and proved upper bounds for $\text{growth}(A)$ under partial pivoting [31, pg. 212] and complete pivoting [30]. His upper bound of 2^{n-1} for partial pivoting is tight, and the large majority of research in this area is devoted to understanding the extent to which exponential growth is rare (see [7, 24, 26] and [25, Lecture 22] for more details regarding the average case and smoothed analysis of Gaussian elimination with partial pivoting). Wilkinson's quasi-polynomial upper bound for the growth factor under complete pivoting was quite pessimistic, but was only recently improved by an exponential constant [1]. A tight estimate for the growth factor under complete pivoting remains a major open problem. For more details on partial and complete pivoting, we refer the reader to [27] and [6, Section 1.1]. Significant attention has also been devoted to matrix transformations as an alternative to pivoting, see [4, 16, 17]. In contrast, Gaussian elimination without pivoting is significantly easier to analyze. It is mostly understood in average case and smoothed analysis settings [19, 32]. For instance, Sankar, Spielman, and Teng showed that, for a $n \times n$ Gaussian matrix A , $\mathbb{P}[\|U\|_\infty \geq Mn^{3/2}\|A\|_\infty] \lesssim 1/M$ [19, Theorem 4.3] and $\mathbb{P}[\|L\|_\infty \geq Mn^2] \lesssim 1/M$ [19, Theorem 4.4] (which implies that $\mathbb{P}[\text{growth}(A) \geq Mn^{5/2}] \lesssim 1/M$). In Theorem 1.1, we prove that, for discrete random matrices, the bound $\text{growth}(A) \lesssim n^{5/2+\delta}$ holds with constant probability for any $\delta > 0$.

1.2. Singularity of discrete random matrices. What is the probability that a binary matrix is singular? This folklore question has a long and rich history. Komlós was the first to show that this probability was $o_n(1)$ [10]. A sequence of works by Kahn, Komlós, and Szemerédi [9], Tao and Vu [20, 21], and Bourgain, Vu, and Wood [2] produced exponentially small bounds, with decreasingly small constants. Tikhomirov proved a tight bound of $(1/2 + o_n(1))^n$ [23]. However, the stronger conjecture that this probability is $(1 + o_n(1))n^2 2^{1-n}$ for a random $\{\pm 1\}$ -matrix (see [9, Equation 1]) remains open. The work of Bourgain, Vu, and Wood also addresses more general classes of discrete random matrices, proving a bound of $(\sqrt{p} + o_n(1))^n$ for matrices with iid random entries ξ with $\mathbb{P}[\xi = a] \leq p$ for all a . The work of Jain, Sah, and Sawhney [8] produced tight estimates for random matrices with non-uniform iid entries, and non-exact but improved estimates even in the uniform setting. Furthermore, the authors provided not only singularity estimates but bounds on the smallest singular value as well.

As noted above, the existence of an LU factorization is equivalent to all of the (strictly) leading principal submatrices being non-singular. As such, our analysis of the LU factorization of discrete matrices borrows many key results and techniques from this literature. In particular, the techniques of Subsection 2.2 are strongly inspired by [10], and we make use of [8, Theorem 1.4] and [2, Theorem 1.4] (see Corollaries 2.6 and 3.9) to handle leading principal minors of arbitrarily large dimension.

1.3. Our Contributions. In this work, we make three key contributions. For a discrete random variable ξ , let $|\xi|_\infty := \max_{a \in \mathbb{R}} \mathbb{P}[\xi = a]$ and $M_n(\xi)$ denote a $n \times n$ random matrix with entries independently distributed according to ξ . First, we prove that any discrete random matrix $M_n(\xi)$ has a constant probability of having a bounded growth factor:

Theorem 1.1. *Let $\delta > 0$ and ξ be a discrete random variable with finite support and $|\xi|_\infty < 1$. There exists a fixed constant $C_\xi < 1$ such that*

$$\mathbb{P}\left[\text{growth}(M_n(\xi)) > n^{\frac{5}{2}+\delta}\right] \leq C_\xi + o_n(1).$$

The exponent of $5/2 + \delta$ is most likely not tight, and experiments suggest that the correct exponent for $\text{growth}(A)$ (for both continuous and discrete random matrices) is $3/2$. Let $\Delta_k(A)$ denote the leading $k \times k$ principal minor of A and $[n] := \{1, \dots, n\}$. Recall, an $n \times n$ matrix A is said to be strongly

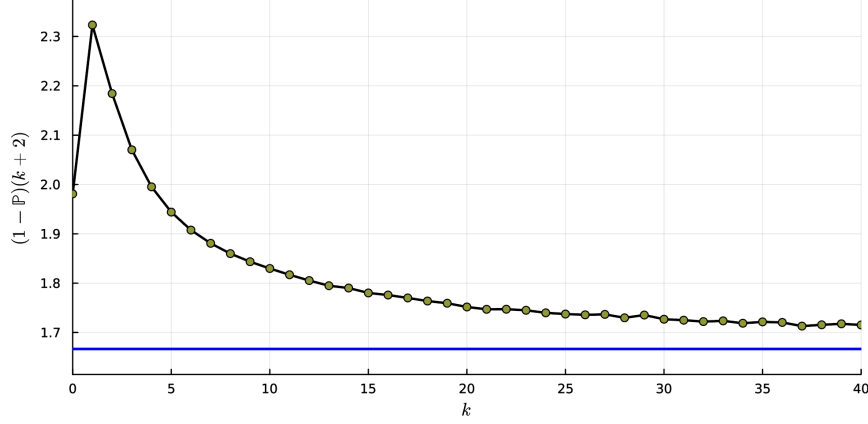


FIGURE 1. The black line is empirical data for $\frac{1 - \mathbb{P}[M_n(\xi) \text{ is strongly non-singular}]}{|\xi|_\infty}$, where $\xi \sim \text{Uniform}(\{2^j\}_{j=0}^k \cup \{0\})$. The blue line is the value $5/3$, which Theorem 1.3 proves the black line approaches in limit. We see that the theoretical estimate of Theorem 1.3 is quite close to experimental results even for small k .

non-singular if $\Delta_k(A) \neq 0$ for all $k \in [n]$. We note that, as a corollary, $M_n(\xi)$ has a constant probability of being strongly non-singular, or, equivalently, $M_{n+1}(\xi)$ has a constant probability of having an LU factorization:

Corollary 1.2. *Let ξ be a discrete random variable with finite support and $|\xi|_\infty < 1$. There exists a fixed constant C'_ξ such that*

$$\mathbb{P}[M_n(\xi) \text{ is strongly non-singular}] \geq C'_\xi > 0.$$

Of course, as $|\xi|_\infty$ tends to zero, the probability of a singular leading principal minor also tends to zero. Second, we give the exact rate at which the singularity probability does so. Let $M_n(p)$ denote an arbitrary $n \times n$ random matrix with independent entries ξ_{ij} , where each ξ_{ij} is a discrete random variable, possibly taking infinitely many values, with $|\xi_{ij}|_\infty \leq p$ for all $i, j \in [n]$. We prove the following result:

Theorem 1.3. *There exists a constant $\alpha > 0$ such that*

$$\mathbb{P}[M_n(p) \text{ is strongly non-singular}] \geq 1 - \frac{5}{3}p - O(p^{1+\alpha})$$

for all $p \in (0, 1)$, and this bound is tight. In particular, for $\xi \sim \text{Uniform}(\{2^j\}_{j=0}^{\lfloor 1/p \rfloor} \cup \{0\})$,

$$\mathbb{P}[M_n(\xi) \text{ is strongly non-singular}] = 1 - \frac{5}{3}p - O(p^{1+\alpha}).$$

The tight example of a random variable ξ uniformly sampled from zero and a geometric progression experimentally achieves the theoretical estimate of $1 - \frac{5}{3}p$ quite quickly (see Figure 1). We note that the generality of Theorem 1.3 comes at a cost: because the entries of $M_n(p)$ are only independent and not identical, statements regarding minimum singular values (or, equivalently, growth factors) are not possible without significant restrictions on the behavior of the entries.

Finally, we pay special attention to the classical case of random $\{0, 1\}$ -matrices, i.e., $M_n(\xi)$ where $\xi \sim \text{Ber}(p)$ equals 1 with probability p , and 0 with probability $1 - p$. Corollary 1.2 provides an implicit constant lower bound for this setting. Here we provide exact counts for dimension at most nine,

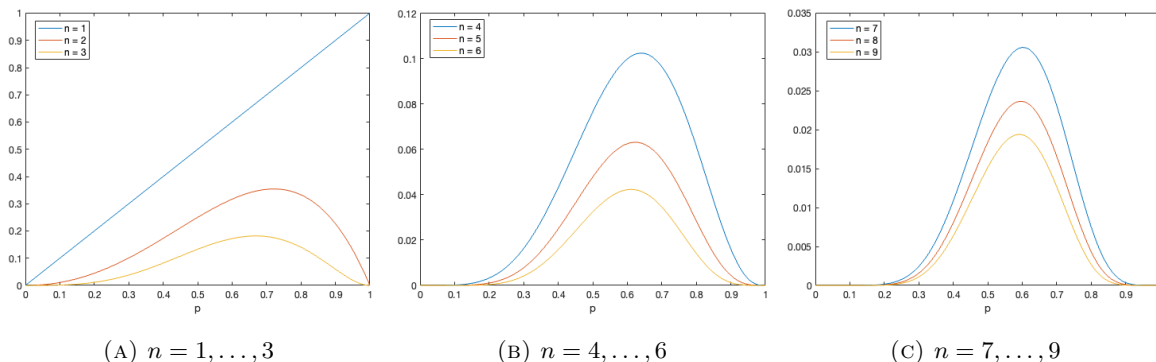


FIGURE 2. Plots of $N_n(p) := \mathbb{P}[M_n(\text{Ber}(p)) \text{ is strongly non-singular}]$ for $n = 1, \dots, 9$.

detailed experimental results for larger dimensions (in exact arithmetic), and theoretical upper bounds for sufficiently large dimensions.

The exact number of strongly non-singular binary matrices, known as sequence OEIS A125587 by the Online Encyclopedia of Integer Sequences, was previously known up to dimension 6, given by the sequence

$$1, 4, 68, 5008, 1603232, 2224232640,$$

with $n = 1, \dots, 4$ computed by Sloane and Vaishapayan and $n = 5, 6$ computed by McKay [13]. We are not aware if the methods used by Sloane, Vaishapayan, and McKay are publicly available or if they differ from an exhaustive search over the 2^{n^2} binary matrices.

In Section 4, we derive an algorithm that uses group theoretic and graph theoretic techniques to drastically improve upon exhaustive search; it relies on the construction of canonical representatives of strongly non-singular binary matrices which form equivalence classes under row and column permutations and an efficient search over the “children” of these matrices formed by appending a row and column. In addition to extending the above sequence to $n = 9$, it provides fine-grained sparsity data by reporting the number of strongly non-singular binary matrices of dimension n with k non-zero entries. These computations were performed on the Duke University Department of Computer Science cluster compute nodes, which are each equipped with 128 CPU threads and 1.15 TB of RAM. Utilizing a dedicated allocation of 100 CPU threads and 500 GB of RAM, evaluating the canonical equivalence classes up to dimension 8×8 required 30.5 days of computation, and computing the 9×9 extensions required an additional 7.2 days. We summarize our results in the following theorem:

Theorem 1.4. *Let N_n denote the number of strongly non-singular matrices in $\{0, 1\}^{n \times n}$ and $N_{n,k}$ denote the number of strongly non-singular matrices in $\{0, 1\}^{n \times n}$ with exactly k non-zero entries. We have*

$$\begin{aligned} N_7 &= 13\,288\,620\,860\,672, \\ N_8 &= 339\,148\,010\,221\,571\,072, \\ N_9 &= 36\,646\,054\,311\,185\,413\,881\,216, \end{aligned}$$

and the exact values of $N_{n,k}$ for all $n \leq 9$ are reported in Tables 1 and 2.

Using the values $N_{n,k}$ for $n \leq 9$, we may exactly compute the functions

$$N_n(p) := \mathbb{P}[M_n(\text{Ber}(p)) \text{ is strongly non-singular}]$$

for $n \leq 9$. We do so and plot the results in Figure 2.

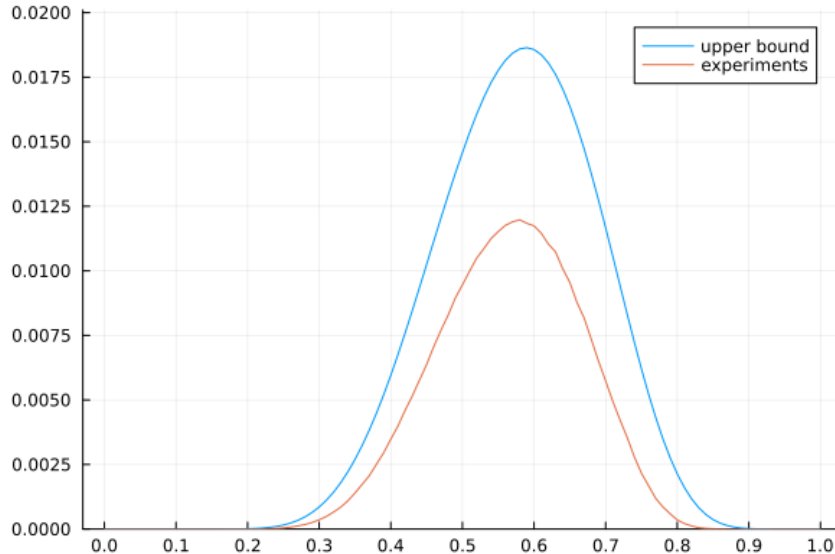


FIGURE 3. Comparison of the analytical upper bound and experimental estimates for $N_{30}(p)$. The theoretical upper bound is evaluated from the exact 9×9 counts. Empirical values were generated via Monte Carlo simulation, testing 1001600 random 30×30 extensions drawn from the set of strongly non-singular 4×4 base matrices.

Exact values of strongly non-singular matrices for small dimensions can be extrapolated to upper bounds for larger dimensions. We perform this extrapolation using the data $N_{n,k}$ in dimension $n = 9$. Our upper bounds are always within an additive factor of at most 0.68% to what sampling suggests. For example, for a uniformly random $\{0, 1\}$ -matrix of dimension $n = 30$, we show that the probability of being strongly non-singular is bounded above by $1.45 \dots \%$, whereas experiments strongly suggest that this value is approximately 0.94%. Proving reasonably tight lower bounds (compared to the implicit constant in Corollary 1.2) appears to be much more difficult. Such an improvement would likely require the creation of new techniques, as all known results that correctly capture the exponential rareness of singularity (i.e., the correct exponential rate) hide large, complicated constants. We leave the following question of providing effective lower bounds for this specific setting as an open problem:

Open Problem 1.5. *Prove that $\lim_{n \rightarrow \infty} \mathbb{P}[M_n(\text{Ber}(\frac{1}{2})) \text{ is strongly non-singular}] > 0.5\%$.*

1.4. Remainder of paper. In Section 2, we prove Theorem 1.1. In Section 3, we prove Theorem 1.3. In Section 4, we detail an algorithm to exactly count the number of $n \times n$ strongly non-singular binary matrices. In Section 5, we detail our methodology for extrapolating our exact computations for dimension nine to upper bounds for $n \times n$ strongly non-singular binary matrices.

2. LU FACTORIZATION OF DISCRETE RANDOM MATRICES

Here we provide a proof of Theorem 1.1. Our technique consists of handling small, medium, and large values of n separately, and then relating singular value estimates to growth factor estimates.

2.1. Bounding n small. A very simple argument provides a loose, but sufficiently good first estimate for small values of n .

Lemma 2.1.

$$\mathbb{P}[\Delta_k(M_n(\xi)) \neq 0 \text{ for all } k \in [n]] \geq (1 - |\xi|_\infty)^n$$

Proof. We proceed by induction. By definition, $\mathbb{P}[\xi \neq 0] \geq 1 - |\xi|_\infty$. Now, suppose $\Delta_k(M_n(\xi)) \neq 0$ for all $k \in [\ell]$ with probability at least $(1 - |\xi|_\infty)^\ell$. Assuming that $\Delta_k(M_n(\xi)) \neq 0$ for all $k \in [\ell]$, the quantity $\Delta_{\ell+1}(M_n(\xi))$ is linear in the entry $[M_n(\xi)]_{\ell+1, \ell+1}$ with a non-zero coefficient, and therefore $\Delta_{\ell+1}(M_n(\xi))$ can equal zero with probability at most $|\xi|_\infty$. $\square$

2.2. Bounding n moderate. Here we apply a version of a technique introduced by Kómlós [10], to produce a bound on invertibility that decays slowly. For the remainder of the subsection, let us write $M_{n+1}(\xi)$ in block notation as

$$M_{n+1}(\xi) = \begin{pmatrix} A & \mathbf{x} \\ \mathbf{y}^T & \alpha \end{pmatrix}, \quad \text{where } A \in \mathbb{R}^{n \times n}, \mathbf{x}, \mathbf{y} \in \mathbb{R}^n, \alpha \in \mathbb{R}.$$

Conditional on the invertibility of A , $M_{n+1}(\xi)$ is singular if and only if $\alpha = \mathbf{y}^T A^{-1} \mathbf{x}$. We aim to bound $\mathbb{P}_{\mathbf{x}, \mathbf{y}, \alpha}[\alpha = \mathbf{y}^T A^{-1} \mathbf{x}]$ for a fixed invertible A , by breaking our analysis into two cases: either the number of non-zeros of $A^{-1} \mathbf{x}$, denoted $\text{nnz}(A^{-1} \mathbf{x})$, is small, which is rare, or $\text{nnz}(A^{-1} \mathbf{x})$ is large, which means it's unlikely that $\mathbf{y}^T A^{-1} \mathbf{x} = \alpha$.

First, we produce a bound on the probability of the random vector $\mathbf{x} \in \mathbb{R}^n$ lying in some fixed subspace of a given dimension.

Lemma 2.2. *Let W be a fixed d -dimensional subspace of $\mathbb{R}^n$ and $\mathbf{x} \in \mathbb{R}^n$ be a random vector with iid entries sampled from a discrete random variable ξ . Then $\mathbb{P}[\mathbf{x} \perp W] \leq |\xi|_\infty^d$.*

Proof. Let $\{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_d\}$ be an orthonormal basis for W and $Q = (\mathbf{w}_1 \mathbf{w}_2 \dots \mathbf{w}_d) \in \mathbb{R}^{n \times d}$. The condition that $\mathbf{x}$ is orthogonal to W is equivalent to $Q^T \mathbf{x} = 0$. The matrix Q has rank d , and, without loss of generality, we may suppose that the first d columns of Q^T are linearly independent. Partitioning Q^T and $\mathbf{x}$ as

$$Q^T = (Q_L \ Q_R) \quad \text{and} \quad \mathbf{x} = \begin{pmatrix} \mathbf{x}_L \\ \mathbf{x}_R \end{pmatrix}, \quad \text{where } Q_L \in \mathbb{R}^{d \times d}, Q_R \in \mathbb{R}^{d \times (n-d)}, \mathbf{x}_L \in \mathbb{R}^d, \mathbf{x}_R \in \mathbb{R}^{n-d},$$

the condition $Q^T \mathbf{x} = 0$ becomes $\mathbf{x}_L = -Q_L^{-1} Q_R \mathbf{x}_R$, i.e., the first d entries of $\mathbf{x}$ are uniquely determined by the last $n - d$. The probability that $\mathbf{x}_L$ equals $-Q_L^{-1} Q_R \mathbf{x}_R$ is at most $|\xi|_\infty^d$, as desired. $\square$

Using the above lemma, we can now effectively bound the probability that $\text{nnz}(A^{-1} \mathbf{x})$ is small.

Lemma 2.3. *Let $A \in \mathbb{R}^{n \times n}$ be a fixed invertible matrix, and $\mathbf{x} \in \mathbb{R}^n$ be a random vector with iid entries sampled from a discrete random variable ξ . Then*

$$\mathbb{P}[\text{nnz}(A^{-1} \mathbf{x}) \leq k] \leq \exp\{k \log(n) + (n - k) \log |\xi|_\infty\} \quad \text{for all } k \geq 3.$$

Proof. For a subset $I \subset [n]$, let E_I be the event

$$E_I := \{(A^{-1} \mathbf{x})_i = 0 \text{ for all } i \in I\} = \{\mathbf{x} \perp \text{span}\{\text{row}_i(A^{-1}) : i \in I\}\}.$$

The event $\text{nnz}(A^{-1} \mathbf{x}) \leq k$ implies that there exists an index set I of size $|I| = n - k$ for which E_I occurs, or, equivalently, there is some $|I| = n - k$ for which $\mathbf{x}$ is orthogonal to the $(n - k)$ -dimensional subspace $\text{span}\{\text{row}_i(A^{-1}) : i \in I\}$. Using Lemma 2.2 and a union bound,

$$\mathbb{P}[\text{nnz}(A^{-1} \mathbf{x}) \leq k] \leq \sum_{|I|=n-k} \mathbb{P}[E_I] \leq \binom{n}{k} |\xi|_\infty^{n-k} \leq \left(\frac{en}{k}\right)^k |\xi|_\infty^{n-k}.$$

Noting that $k \geq 3$, our desired result follows immediately

$$\begin{aligned} \mathbb{P}[\text{nnz}(A^{-1}\mathbf{x}) \leq k] &\leq \exp\{k(\log n + 1 - \log k) + (n - k) \log |\xi|_\infty\} \\ &\leq \exp\{k \log n + (n - k) \log |\xi|_\infty\}. \end{aligned}$$

□

When $\text{nnz}(A^{-1}\mathbf{x})$ is not too small, we can make use of the following Littlewood-Offord type inequality of Rogozin to conclude that $\alpha = \mathbf{y}^T A^{-1}\mathbf{x}$ is rare.

Corollary 2.4 (Corollary of [18, Theorem 2]). *Let $\mathbf{z} \in \mathbb{R}^m$ be a fixed vector with k non-zero entries, and $\mathbf{y} \in \mathbb{R}^m$ be a random vector with iid entries sampled from a discrete random variable ξ . Then*

$$\sup_{a \in \mathbb{R}} \mathbb{P}[\mathbf{y}^T \mathbf{z} = a] \leq \frac{C}{\sqrt{k(1 - |\xi|_\infty)}}$$

for an absolute constant C .

Combining Lemma 2.2 and Corollary 2.4, we can obtain a gentler decay of the probability of all leading principal minors being non-zero.

Lemma 2.5. *Let $A \in \mathbb{R}^{n \times n}$ be a fixed non-singular matrix, and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$ have iid entries sampled from a discrete random variable ξ . Then there exists some $n_0 \in \mathbb{N}$ such that*

$$\mathbb{P}\left[\det \begin{pmatrix} A & \mathbf{x} \\ \mathbf{y}^T & \alpha \end{pmatrix} = 0\right] \leq \frac{1}{n^{1/3}} \quad \text{for all } n \geq n_0.$$

Proof. We break our analysis into two cases depending on the size of $\text{nnz}(A^{-1}\mathbf{x})$. By Lemma 2.2, $\text{nnz}(A^{-1}\mathbf{x}) \leq k$ with probability at most $\exp\{k \log n + (n - k) \log |\xi|_\infty\}$ (for $k \geq 3$). By Corollary 2.4, if $\text{nnz}(A^{-1}\mathbf{x}) > k$, then $(\mathbf{y}^T \quad \alpha) \begin{pmatrix} A^{-1}\mathbf{x} \\ -1 \end{pmatrix} = 0$ with probability at most $C((k + 2)(1 - |\xi|_\infty))^{-1/2}$ for some fixed constant C . Altogether, we have that, for $k = \lfloor (n \log |\xi|_\infty^{-1}) / (2 \log(n |\xi|_\infty^{-1})) \rfloor$,

$$\begin{aligned} \mathbb{P}\left[\det \begin{pmatrix} A & \mathbf{x} \\ \mathbf{y}^T & \alpha \end{pmatrix} = 0\right] &\leq \exp\{k \log n + (n - k) \log |\xi|_\infty\} + \frac{C}{\sqrt{(k + 2)(1 - |\xi|_\infty)}} \\ &\leq \exp\{-\tfrac{1}{2}n \log |\xi|_\infty^{-1}\} + \frac{C\sqrt{2 \log(n |\xi|_\infty^{-1})}}{\sqrt{(1 - |\xi|_\infty) n \log |\xi|_\infty^{-1}}}. \end{aligned}$$

For n sufficiently large, this quantity is at most $1/n^{1/3}$. □

2.3. Bounding n large. To treat n sufficiently large and produce bounds on singular values, we recall the following corollary of a result of Jain, Sah, and Sawhney.

Corollary 2.6 (Corollary of [8, Theorem 1.4]). *Let ξ be a discrete random variable taking finitely many values. There exists a constant $C_\xi > 0$ such that for any fixed $\epsilon > 0$ and for all sufficiently large n and all $t \geq 0$,*

$$\mathbb{P}[\sigma_{\min}(M_n(\xi)) \leq t/\sqrt{n}] \leq C_\xi t + (1 + \epsilon)^n |\xi|_\infty^n.$$

2.4. From singular values to growth factor. The standard technique for estimating the growth factor of a random matrix is through estimates for the smallest singular values of leading sub-matrices. The below lemma does not exactly appear anywhere in the literature, and so we provide a short proof for completeness.

Lemma 2.7. *Let $A \in \mathbb{R}^{n \times n}$. Then*

$$\text{growth}(A) \leq 1 + \|A\|_{\max} \max_{k \in [n-1]} k \sigma_{\min}^{-1}(A_{[k], [k]}).$$

Proof. Let $A_{S,T}$ denote the submatrix of A indexed by the rows $S \subset [n]$ and $T \subset [n]$, and $A = LU$ be the LU factorization of A , given by

$$L_{ij} = \begin{cases} \frac{\det(A_{[j-1] \cup \{i\}, [j]})}{\det(A_{[j], [j]})} & i \geq j \\ 0 & i < j \end{cases} \quad \text{and} \quad U_{ij} = \begin{cases} A_{1j} & i = 1 \\ A_{ij} - A_{i, [i-1]} A_{[i-1], [i-1]}^{-1} A_{[i-1], j} & 1 < i \leq j \\ 0 & i > j \end{cases}.$$

We may easily bound the largest magnitude entry of U using the smallest singular value

$$\|U\|_{\max} \leq \max_{k \in [n-1]} \|A\|_{\max} + k \|A\|_{\max}^2 \sigma_{\min}^{-1}(A_{[k], [k]}).$$

For an entry L_{ij} , $i > j$, we have

$$L_{ij} = \frac{\det(A_{[j-1] \cup \{i\}, [j]})}{\det(A_{[j], [j]})} = \sum_{k=1}^j A_{ik} (-1)^{j+k} \frac{\det(A_{[j-1], [j] \setminus k})}{\det(A_{[j], [j]})} = \sum_{k=1}^j A_{ik} [A_{[j], [j]}]_{kj}^{-1} = A_{i, [j]} A_{[j], [j]}^{-1} \mathbf{e}_j,$$

implying that

$$\|L\|_{\max} \leq \max \left\{ \|A\|_{\max} \max_{k \in [n-1]} \sqrt{k} \sigma_{\min}^{-1}(A_{[k], [k]}), 1 \right\}.$$

From here, our desired result follows immediately

$$\text{growth}(A) = \max \left\{ \|L\|_{\max}, \frac{\|U\|_{\max}}{\|A\|_{\max}} \right\} \leq 1 + \|A\|_{\max} \max_{k \in [n-1]} k \sigma_{\min}^{-1}(A_{[k], [k]}).$$

□

2.5. Proof of Theorem 1.1. Using Lemmas 2.1 and 2.5 and Corollary 2.6, we are now prepared to prove Theorem 1.1.

Proof of Theorem 1.1. By Lemmas 2.1 and 2.5, a lower bound for the probability of having all non-vanishing leading principal minors for $n > n_0$ is given by

$$(1 - |\xi|_{\infty})^{n_0} \prod_{k=n_0+1}^n \left(1 - \frac{1}{k^{1/3}} \right) \geq (1 - |\xi|_{\infty})^{n_0} \exp\{-2n^{2/3}\}.$$

Choose $0 < \delta' < \delta$. Since $|\xi|_{\infty} < 1$, there exists an $\epsilon > 0$ such that $\gamma := (1 + \epsilon)|\xi|_{\infty} < 1$. For some sufficiently large $n_1 > n_0$ and sufficiently small constant $\hat{c}$, Corollary 2.6 is valid for this ϵ and $t = \hat{c} k^{-1-\delta'}$, and we have

$$\sum_{k=n_1}^{\infty} \gamma^k + \sum_{k=n_1}^{\infty} \frac{\hat{c} C_{\xi}}{k^{1+\delta'}} < (1 - |\xi|_{\infty})^{n_0} \exp\{-2n_1^{2/3}\}$$

giving a lower bound of

$$(1 - |\xi|_{\infty})^{n_0} \exp\{-2n_1^{2/3}\} - \sum_{k=n_1}^{\infty} \gamma^k - \sum_{k=n_1}^{\infty} \frac{\hat{c} C_{\xi}}{k^{1+\delta'}} > 0$$

for the probability that all leading principal minors are non-zero and the smallest singular value of all the $k \times k$, $k \geq n_1$, leading principal minors are all at least $\hat{c}/k^{\frac{3}{2}+\delta'}$. Since ξ is finitely supported, $\|A\|_{\max} \leq M$ for some constant M . Applying Lemma 2.7 gives $\text{growth}(A) \leq 1 + \frac{M}{\hat{c}} n^{\frac{5}{2}+\delta'}$. Taking n large enough so that $1 + \frac{M}{\hat{c}} n^{\frac{5}{2}+\delta'} \leq n^{\frac{5}{2}+\delta}$ and $n^{\frac{3}{2}+\delta'}$ is larger than the inverse of any of the smallest singular values of the leading principal minors of length at most n_1 completes the proof. □

3. LU FACTORIZATION OF DISCRETE RANDOM MATRICES WITH ANTI-CONCENTRATED ENTRIES

Here we provide a proof of Theorem 1.3. Again, our technique consists of considering small ($n = 1, 2$), medium $n = O(\log 1/p)$, and large n separately.

3.1. Bounding $n = 1, 2$. Analysis of dimension $n = 1$ is obvious, as $\mathbb{P}[\Delta_1(M_1(p)) = 0] = \mathbb{P}[\xi_{11} = 0] \leq p$. However, already when $n = 2$, some work is required. First, we prove the following helpful lemmas regarding the collision probability $\sum_{\alpha} \mathbb{P}[\xi = \alpha]^2$ for two i.i.d. copies of a discrete random variable ξ .

Lemma 3.1. *Let X, X', Y, Y' be independent discrete random variables taking finitely many values, with X, X' and Y, Y' each identically distributed. Then*

$$\sum_{\alpha} \mathbb{P}[X + Y = \alpha]^2 \leq \sqrt{\left(\sum_{\beta} \mathbb{P}[X + X' = \beta]^2\right) \left(\sum_{\gamma} \mathbb{P}[Y + Y' = \gamma]^2\right)}$$

and $\sum_{\beta} \mathbb{P}[X + X' = \beta]^2$ is a convex function of the weights $\mathbb{P}[X = \alpha]$, $\alpha \in \text{supp}(X)$.

Proof. Here, we are concerned only with probabilities involving pairwise sums of X, X', Y, Y' , and so, by the Frieman isomorphism lemma [22, Lemma 5.25], we may assume that X, X', Y, Y' are all supported on $\mathbb{Z}$. By Parseval's theorem and the discrete convolution theorem [14, Section 2.9],

$$\begin{aligned} \sum_{\alpha} \mathbb{P}[X + Y = \alpha]^2 &= \frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{\alpha} \mathbb{P}[X + Y = \alpha] e^{i\alpha\theta} \right|^2 d\theta \\ &= \frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{\beta} \mathbb{P}[X = \beta] e^{i\beta\theta} \right|^2 \left| \sum_{\gamma} \mathbb{P}[Y = \gamma] e^{i\gamma\theta} \right|^2 d\theta \\ &\leq \sqrt{\left(\frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{\beta} \mathbb{P}[X = \beta] e^{i\beta\theta} \right|^4 d\theta\right) \left(\frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{\gamma} \mathbb{P}[Y = \gamma] e^{i\gamma\theta} \right|^4 d\theta\right)} \\ &= \sqrt{\left(\frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{\delta} \mathbb{P}[X + X' = \delta] e^{i\delta\theta} \right|^2 d\theta\right) \left(\frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{\epsilon} \mathbb{P}[Y + Y' = \epsilon] e^{i\epsilon\theta} \right|^2 d\theta\right)} \\ &= \sqrt{\left(\sum_{\delta} \mathbb{P}[X + X' = \delta]^2\right) \left(\sum_{\epsilon} \mathbb{P}[Y + Y' = \epsilon]^2\right)}. \end{aligned}$$

Using the same argument,

$$\sum_{\delta} \mathbb{P}[X + X' = \delta]^2 = \frac{1}{2\pi} \int_0^{2\pi} \left| \sum_{\beta} \mathbb{P}[X = \beta] e^{i\beta\theta} \right|^4 d\theta.$$

From here, convexity follows quickly, as $\sum_{\beta} \mathbb{P}[X = \beta] e^{i\beta\theta}$ is a linear function of $\mathbb{P}[X = \beta]$, $|\cdot|^4$ is convex, the convex function of a linear function is convex, and integrals of convex functions are convex. $\square$

Lemma 3.2. *Let $s_1 < \dots < s_n$ and $C_{ij} = |\{(k, \ell) \mid s_i + s_j = s_k + s_{\ell}\}|$. Then*

$$\phi(m) := |\{(i, j) \mid C_{ij} \geq m\}| \leq n^2 - m(m-1).$$

In addition, if X and X' are i.i.d. discrete random variables supported on $\leq n$ points with $\max_{\alpha} \mathbb{P}[X = \alpha] \leq p$, then

$$\sum_{\alpha} \mathbb{P}[X + X' = \alpha]^2 \leq \frac{(2n^2 + 1)np^4}{3}.$$

Proof. If $s_i \leq s_j$, $s_k \leq s_\ell$, $i \neq k$, and $s_i + s_j = s_k + s_\ell$, then either $i > k$ and $j < \ell$, or $i < k$ and $j > \ell$. Therefore, in order for $C_{ij} \geq m$, for $i \leq j$, we must have

$$|(n - j) - (i - 1)| \leq n - m.$$

Exactly $m(m - 1)$ pairs fail to satisfy this condition for $m \leq n$, and every pair fails to satisfy this condition for $m > n$.

Now, consider (X, X') , supported on $\{s_1, \dots, s_n\}^2$. We have $\mathbb{P}[(X, X') = (s_i, s_j)] \leq p^2$ and

$$\sum_{\alpha} \mathbb{P}[X + X' = \alpha]^2 \leq \sum_{m=1}^n \frac{\phi(m) - \phi(m+1)}{m} (mp^2)^2 = p^4 \sum_{m=1}^n \phi(m).$$

By our above bound on $\phi(m)$,

$$\sum_{\alpha} \mathbb{P}[X + X' = \alpha]^2 \leq p^4 \sum_{m=1}^n n^2 - m(m-1) = \frac{(2n^2 + 1)np^4}{3}.$$

□

We are now prepared to provide a tight estimate for the case $n = 2$.

Lemma 3.3.

$$\mathbb{P}[\Delta_2(M_n(p)) = 0 \mid \Delta_1(M_n(p)) \neq 0] \leq \frac{2}{3}p + O(p^2)$$

Proof. Conditioning on $\Delta_1(M_n(p)) \neq 0$ is equivalent to replacing ξ_{11} by a new random variable with ξ'_{11} satisfying $|\xi'_{11}|_{\infty} \leq p/(1-p) \leq p + O(p^2)$. Therefore, we may safely ignore the conditioning in the lemma statement. In addition, we may also assume that the entries ξ_{ij} are finitely supported, at negligible cost. The probability that any entry is zero and causes a vanishing minor is bounded by $O(p^2)$, allowing us to condition on all entries being non-zero. Let $\eta_{ij} = \log |\xi_{ij}|$. Because ignoring signs strictly upper bounds the probability of a multiplicative collision, we can apply Cauchy-Schwarz to obtain,

$$\begin{aligned} \mathbb{P}[\Delta_2(M_n(p)) = 0] &= \sum_{\alpha} \mathbb{P}[\eta_{11} + \eta_{22} = \alpha] \mathbb{P}[\eta_{12} + \eta_{21} = \alpha] \\ &\leq \left(\sum_{\alpha} \mathbb{P}[\eta_{11} + \eta_{22} = \alpha]^2 \right)^{1/2} \left(\sum_{\alpha} \mathbb{P}[\eta_{12} + \eta_{21} = \alpha]^2 \right)^{1/2}, \end{aligned}$$

and so it suffices to bound $\sum_{\alpha} \mathbb{P}[\eta_{11} + \eta_{22} = \alpha]^2$. By Lemma 3.1, we may instead bound the quantity $\sum_{\alpha} \mathbb{P}[\eta_{11} + \hat{\eta}_{11} = \alpha]^2$, where η_{11} and $\hat{\eta}_{11}$ are independent and identically distributed. In addition, also by Lemma 3.1, $\sum_{\alpha} \mathbb{P}[\eta_{11} + \hat{\eta}_{11} = \alpha]^2$ is a convex function of the weights $\mathbb{P}[\eta_{11} = \beta]$. Therefore, it is maximized at a boundary point, i.e., the weights of all support points equal either p or 0 , save for a single support point, which has weight $1 - \lfloor 1/p \rfloor p$. Applying Lemma 3.2 with $\lfloor 1/p \rfloor$ support points completes the proof

$$\sum_{\alpha} \mathbb{P}[\eta_{11} + \hat{\eta}_{11} = \alpha]^2 \leq \frac{(2\lfloor 1/p \rfloor^2 + 1)\lfloor 1/p \rfloor p^4}{3} = \frac{2}{3}p + O(p^2).$$

□

3.2. Bounding n moderate. When $n = O(\log 1/p)$ and greater than 2, we may make use of the following generalization of the Elkes-Szabó theorem for the intersection of semi-algebraic varieties and grids to higher dimensions.

Theorem 3.4 ([3, Theorem A]). *Let $s \geq 4$, $Q \subset \mathbb{R}^s$ be semi-algebraic, of description complexity D , and such that the projection of Q to any $s - 1$ coordinates is finite-to-one. Then exactly one of the following holds:*

- (1) *There exists a constant c depending only on s, D such that: for any $n \in \mathbb{N}$ and finite $S_i \subset \mathbb{R}$ with $|S_i| = n$ for $i \in [s]$ we have $|Q \cap (S_1 \times \dots \times S_s)| \leq cn^{s-4/3}$.*
- (2) *There exists open sets $U_i \subset \mathbb{R}$, $i \in [s]$, an open set $V \subset \mathbb{R}$ containing 0, and analytic bijections with analytic inverses $\pi_i : U_i \rightarrow V$ such that $\pi_1(x_1) + \dots + \pi_s(x_s) = 0 \Leftrightarrow Q(x_1, \dots, x_s)$ for all $x_i \in U_i$, $i \in [s]$.*

In particular, we aim to apply the above theorem to the intersection of the set of singular 3×3 matrices and a grid $\prod_{i,j=1}^3 S_{ij} \subset \mathbb{R}^{3 \times 3}$. First, we prove the following result regarding a C^2 function whose zero level set locally matches that of a linear function.

Lemma 3.5. *Let f be C^2 in an open neighborhood of $0 \in \mathbb{R}^n$, g be a non-zero linear function on $\mathbb{R}^n$, $f(0) = g(0) = 0$, $\nabla f(0) \neq 0$, and, in an open neighborhood of 0, $f(x) = 0$ if and only if $g(x) = 0$. Then there exists a C^1 function in an open neighborhood of 0, u , with $f(x) = g(x)u(x)$ and $u(0) \neq 0$.*

Proof. Because g is linear, there is a change of basis Φ of x such that, for $\alpha \in \mathbb{R}$ and $y \in \mathbb{R}^{n-1}$, $\Phi(\alpha, y) = x$ and $\alpha = g(x)$. Let $F(\alpha, y) = f(\Phi(\alpha, y))$. Near the origin, we have $F(0, y) = 0$, $F(0, 0) = 0$, $\nabla F(0, 0) \neq 0$, $F^{-1}(0) = \{\alpha = 0\}$, and

$$F(\alpha, y) = F(\alpha, y) - F(0, y) = \int_0^\alpha \frac{\partial F}{\partial \alpha}(\beta, y) d\beta = \alpha \int_0^1 \frac{\partial F}{\partial \alpha}(\tau \alpha, y) d\tau.$$

Let $w(\alpha, y) = \int_0^1 \frac{\partial F}{\partial \alpha}(\tau \alpha, y) d\tau$ and note that, because F is C^2 near the origin, w is also C^1 near the origin. Letting $u(x) = w(g(x), \Phi^{-1}(x)|_y)$, we have

$$f(x) = F(g(x), \Phi^{-1}(x)|_y) = g(x)u(x).$$

Finally, we note that

$$u(0) = \int_0^1 \frac{\partial F}{\partial \alpha}(0, 0) d\tau = \frac{\partial F}{\partial \alpha}(0, 0)$$

and the tangent space of the zero level set at the origin is $\{(\alpha, y) \mid \alpha = 0\}$, therefore, $\frac{\partial F}{\partial \alpha}(0, 0) \neq 0$. $\square$

Lemma 3.6. *Let $\mathcal{S}^* = \{A \in \mathbb{R}^{3 \times 3} : \det(A) = 0 \text{ and no } 2 \times 2 \text{ minor is zero}\}$. There exists a constant c such that, for any $S_{ij} \subset \mathbb{R}$ with $|S_{ij}| = n$ for $i, j \in [3]$, we have $|\mathcal{S}^* \cap (\prod_{i,j=1}^3 S_{ij})| \leq cn^{23/3}$.*

Proof. We apply Theorem 3.4 to the semi-algebraic set $\mathcal{S}^*$. The determinant is linear in any entry, and because no 2×2 minor is zero for matrices in $\mathcal{S}^*$, fixing any 8 entries of $A \in \mathcal{S}^*$ leaves a non-trivial linear equation for the 9th entry, which has one solution. Thus, the projection of $\mathcal{S}^*$ to any 8 coordinates is at finite-to-one, satisfying the hypothesis of Theorem 3.4.

It suffices to show that $\mathcal{S}^*$ does not satisfy property (2) of Theorem 3.4. Let us suppose property (2) holds. Note that there must exist a matrix $\hat{A} \in U_1 \times \dots \times U_9$ with $\text{rank}(\hat{A}) = 2$ and no vanishing 2×2 minors, as the subset of matrices violating these conditions has dimension at most 7, while $\mathcal{S}^*$ has dimension 8. Let $B = \pi(\hat{A})$, with, without loss of generality, $\pi(\hat{A}) = 0$. For $B \in V$, $f(B) := \sum_{i,j=1}^3 B_{ij} = 0$ if and only if $\det(\pi^{-1}(B)) = 0$. Because $\text{rank}(\hat{A}) = 2$, $\nabla \det(\pi^{-1}(0)) \neq 0$, and, by

Lemma 3.5, $\det(\pi^{-1}(B)) = f(B)u(B)$ for some C^1 function u with $u(0) \neq 0$. Computing the mixed second derivatives of both sides, we find that

$$\frac{\partial^2 \det(\pi^{-1}(B))}{\partial B_{ij} \partial B_{k\ell}}(0) = \frac{\partial u}{\partial B_{ij}}(0) + \frac{\partial u}{\partial B_{k\ell}}(0) \quad \text{for } (i, j) \neq (k, \ell).$$

However, by the linearity of the determinant in each row and column, we have $\frac{\partial^2 \det(\pi^{-1}(B))}{\partial B_{ij} \partial B_{\ell k}}(0) = 0$ when (i, j) and (k, ℓ) are in the same row or column. The resulting system of equations in derivatives of u implies that $\frac{\partial u}{\partial B_{ij}}(0) = 0$ for all (i, j) , and therefore all mixed derivatives of $\det(\pi^{-1})$ vanish at 0.

However, this is a contradiction, as $\hat{A}$ has rank two, and so $\frac{\partial^2 \det(\hat{A})}{\partial A_{ij} \partial A_{k,\ell}}$ is non-zero for some $(i, j) \neq (k, \ell)$. Because $\pi'_{ij}(0) \neq 0$ for all (i, j) , the chain rule implies that the mixed derivatives of $\det(\pi^{-1}(B))$ cannot all vanish. $\square$

Remark 3.7. The proof of Lemma 3.6 crucially requires that $n \geq 3$. Note that, by Lemma 3.3, an analogous statement is not true for $n = 2$. In particular, while the mixed derivatives in each row and column do vanish, because there are only two entries in each row and column, this is not sufficient to conclude that all first derivatives vanish. For example, for an entry-wise positive matrix in $\mathbb{R}^{2 \times 2}$, logarithmic bijections recognize singularity:

$$\det(A) = 0 \text{ for } A \in (0, \infty)^{2 \times 2} \iff \log(A_{11}) - \log(A_{12}) - \log(A_{21}) + \log(A_{22}) = 0.$$

Now, taking

$$B = \begin{pmatrix} \log(A_{11}) & -\log(A_{12}) \\ -\log(A_{21}) & \log(A_{22}) \end{pmatrix}$$

and $\hat{A} = \mathbf{1}\mathbf{1}^T$, the mixed derivatives of $\det(\pi^{-1}(B)) = e^{B_{11}+B_{22}} - e^{-(B_{12}+B_{21})}$ along a row or column do vanish, but the first derivatives of

$$u(B) = \frac{e^{B_{11}+B_{22}} - e^{-(B_{12}+B_{21})}}{B_{11} + B_{12} + B_{21} + B_{22}},$$

appropriately defined at indefinite points, do not vanish at $B = 0$: $\partial u / \partial B_{11}(0) = \partial u / \partial B_{22}(0) = 1/2$ and $\partial u / \partial B_{12}(0) = \partial u / \partial B_{21}(0) = -1/2$.

Using the above lemma, we can now bound $\mathbb{P}[\Delta_n(M_n(p)) = 0 \mid \Delta_{n-3}(M_n(p)) \neq 0]$.

Lemma 3.8. Let $n \geq 3$. Then $\mathbb{P}[\Delta_n(M_n(p)) = 0 \mid \Delta_{n-3}(M_n(p)) \neq 0] = O(p^{27/26})$.

Proof. Let $A = M_n(p)$ be written in block notation as

$$\begin{aligned} A &= \begin{pmatrix} B & \mathbf{x} & \mathbf{y} & \mathbf{z} \\ \mathbf{u}^T & \xi_{11} & \xi_{12} & \xi_{13} \\ \mathbf{v}^T & \xi_{21} & \xi_{22} & \xi_{23} \\ \mathbf{w}^T & \xi_{31} & \xi_{32} & \xi_{33} \end{pmatrix} \\ &= \begin{pmatrix} I & & & \\ \mathbf{u}^T B^{-1} & 1 & & \\ \mathbf{v}^T B^{-1} & & 1 & \\ \mathbf{w}^T B^{-1} & & & 1 \end{pmatrix} \begin{pmatrix} B & \mathbf{x} & \mathbf{y} & \mathbf{z} \\ \xi_{11} - \mathbf{u}^T B^{-1} \mathbf{x} & \xi_{12} - \mathbf{u}^T B^{-1} \mathbf{y} & \xi_{13} - \mathbf{u}^T B^{-1} \mathbf{z} \\ \xi_{21} - \mathbf{v}^T B^{-1} \mathbf{x} & \xi_{22} - \mathbf{v}^T B^{-1} \mathbf{y} & \xi_{23} - \mathbf{v}^T B^{-1} \mathbf{z} \\ \xi_{31} - \mathbf{w}^T B^{-1} \mathbf{x} & \xi_{32} - \mathbf{w}^T B^{-1} \mathbf{y} & \xi_{33} - \mathbf{w}^T B^{-1} \mathbf{z} \end{pmatrix}, \end{aligned}$$

where $B \in \mathbb{R}^{(n-3) \times (n-3)}$ and $\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^{n-3}$. Let us condition on a (non-singular) realization of B and of the vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}, \mathbf{x}, \mathbf{y}, \mathbf{z}$, and consider the matrix

$$C = \begin{pmatrix} \xi_{11} - \mathbf{u}^T B^{-1} \mathbf{x} & \xi_{12} - \mathbf{u}^T B^{-1} \mathbf{y} & \xi_{13} - \mathbf{u}^T B^{-1} \mathbf{z} \\ \xi_{21} - \mathbf{v}^T B^{-1} \mathbf{x} & \xi_{22} - \mathbf{v}^T B^{-1} \mathbf{y} & \xi_{23} - \mathbf{v}^T B^{-1} \mathbf{z} \\ \xi_{31} - \mathbf{w}^T B^{-1} \mathbf{x} & \xi_{32} - \mathbf{w}^T B^{-1} \mathbf{y} & \xi_{33} - \mathbf{w}^T B^{-1} \mathbf{z} \end{pmatrix}$$

Then, after conditioning, $C = M_3(p)$ and $\det(A)$ is non-zero if and only if $\det(C)$ is non-zero. Therefore, we may restrict ourselves to the unconditional case of $n = 3$ and the matrix C . We decompose our analysis into a variety of cases that, together, bound $\mathbb{P}[\det(C) = 0]$. We have

$$\mathbb{P}[C_{ij} = C_{\ell k} = 0 \text{ for } (i, j) \neq (\ell, k)] = O(p^2)$$

and

$$\mathbb{P}[C_{11}C_{22} = C_{12}C_{21} \neq 0 \text{ and } C_{12}C_{23} = C_{22}C_{13} \neq 0] = O(p^2),$$

and so the probability that C has a vanishing minor of order two and $\det(C) = 0$ is $O(p^2)$. Now, we may restrict ourselves to realizations of C with no vanishing minors of order two, and further divide our analysis based on whether any entry of C equals a “rare” value. Using the same argument from Lemma 2.1,

$$\mathbb{P}\left[\det(C) = 0, C_{11}C_{22} \neq C_{12}C_{21}, C_{33} = \alpha \text{ for } \alpha \text{ s.t. } \mathbb{P}[C_{33} = \alpha] \leq p^{27/26}\right] = O(p^{27/26}).$$

If no entry of C equals a “rare” value, then we may simply use Lemma 3.6 to obtain

$$\begin{aligned} \mathbb{P}\left[\det(C) = 0, C_{ij} = \alpha_{ij} \text{ for } \alpha_{ij} \text{ s.t. } \mathbb{P}[C_{ij} = \alpha_{ij}] > p^{27/26} \text{ for } i, j = 1, 2, 3\right] &\leq \left[\frac{1}{p^{27/26}}\right]^{23/3} p^9 \\ &= O(p^{27/26}). \end{aligned}$$

□

3.3. Bounding n large. When n is sufficiently large relative to $\log 1/p$, we may simply make use of the following corollary of a result of Bourgain, Vu, and Wood.

Corollary 3.9 (Corollary of [2, Theorem 1.4]). *There exists fixed constants $\delta \in (0, 1)$ and $n_0 \in \mathbb{N}$ such that $\mathbb{P}[\Delta_n(M_n(p)) = 0] \leq \delta^n$ for all $p \leq 1/2$ and $n \geq n_0$.*

3.4. Proof of Theorem 1.3. Using Lemmas 3.3 and 3.8, and Corollary 3.9, we are now prepared to prove Theorem 1.3.

Proof of Theorem 1.3. Let $\ell = \max\left\{n_0, \left\lceil \frac{2 \log p}{\log \delta} \right\rceil\right\}$, where n_0 and δ are the constants of Corollary 3.9. We have

$$\begin{aligned} \mathbb{P}[\Delta_k(M_n(p)) = 0 \text{ for some } k \in [n]] &\leq \mathbb{P}[\Delta_1(M_n(p)) = 0] + \mathbb{P}[\Delta_2(M_n(p)) = 0 \mid \Delta_1(M_n(p)) \neq 0] \\ &\quad + \sum_{k=3}^{\ell} \mathbb{P}[\Delta_k(M_n(p)) = 0 \mid \Delta_{k-3}(M_n(p)) \neq 0] \\ &\quad + \sum_{k=\ell+1}^n \mathbb{P}[\Delta_k(M_n(p)) = 0]. \end{aligned}$$

Clearly, the first term is at most p . By Lemma 3.3, the second term is at most $\frac{2}{3}p + O(p^2)$. By Lemma 3.8, the third term is at most $O(\ell p^{27/26})$. Finally, by Corollary 3.9, and noting that our choice of ℓ gives $\delta^\ell \leq p^2$,

$$\sum_{k=\ell+1}^n \mathbb{P}[\Delta_k(M_n(p)) = 0] \leq \sum_{k=\ell+1}^n \delta^k \leq \frac{\delta^{\ell+1}}{1-\delta} \leq \frac{\delta^\ell}{1-\delta} \leq \frac{p^2}{1-\delta} = O(p^2).$$

To verify tightness, note $\mathbb{P}[\Delta_1(M_n(\xi)) = 0] = p + O(p^2)$. Given $\Delta_1(M_n(\xi)) \neq 0$, $\Delta_2(M_n(\xi)) = 0$ is dominated by the collision $\xi_{11}\xi_{22} = \xi_{12}\xi_{21}$ among non-zero entries. Taking base-2 logarithms reduces this directly to the uniform additive collision problem of Lemma 3.3, giving a probability of $\frac{2}{3}p + O(p^2)$. Since higher-order minors contribute at most $O(p^{27/26})$, the probability of a singular leading principal minor is exactly $\frac{5}{3}p + O(p^{27/26})$. □

4. EXACT COUNTING OF STRONGLY NON-SINGULAR $\{0, 1\}$ -MATRICES

Here we detail an algorithm for exactly counting the number of $n \times n$ strongly non-singular $\{0, 1\}$ -matrices for $n = 1, 2, \dots, 9$ via enumeration. The associated code and computational results are available at the GitHub repository [15]. The naive enumeration of $n \times n$ strongly non-singular $\{0, 1\}$ -matrices is computationally intractable for $n \geq 9$, as the search space size 2^{n^2} grows super-exponentially. To address this, we implement a counting algorithm based on identifying equivalence classes of matrices under row and column permutations. This approach reduces the space sufficiently to compute exact counts up to $n = 9$.

4.1. Equivalence Classes and Recursive Construction. We define an equivalence relation on the set of $n \times n$ binary matrices $\{0, 1\}^{n \times n}$. Two matrices $A, B \in \{0, 1\}^{n \times n}$ are equivalent, denoted $A \sim B$, if there exist permutation matrices P and Q such that $A = PBQ$ or $A = PB^TQ$. Let

$$\Omega_n := \{A \in \{0, 1\}^{n \times n} \mid \det(\Delta_k(A)) \neq 0 \forall k \in [n]\}$$

be the set of $n \times n$ strongly non-singular $\{0, 1\}$ -matrices. Given the set of equivalence classes $\Omega_n / \sim$ and their sizes, we detail an algorithm to construct $\Omega_{n+1} / \sim$ and their sizes.

We say a matrix C is an extension of $A \in \Omega_n$ if

$$C = \begin{pmatrix} A & \mathbf{x} \\ \mathbf{y}^T & \alpha \end{pmatrix}, \quad \text{where } \mathbf{x}, \mathbf{y} \in \{0, 1\}^n, \alpha \in \{0, 1\}.$$

The matrix C is strongly non-singular if and only if $A \in \Omega_n$ and $\det(C) = \alpha \cdot \det(A) - \mathbf{y}^T \text{adj}(A) \mathbf{x} \neq 0$. When this is the case, we say C is a child of A , and denote the set of strongly non-singular extensions of a matrix A by $\text{Children}(A)$.

Lemma 4.1 (Isomorphism Preservation). *Let $A, B \in \Omega_n$ and $A \sim B$. There exists a bijection $\phi : \text{Children}(A) \rightarrow \text{Children}(B)$ such that, for every child C of A , $C \sim \phi(C)$.*

Proof. By definition, there exist permutation matrices P and Q such that either $A = PBQ$ or $A = PB^TQ$. Without loss of generality, suppose $A = PBQ$ (the case $A = PB^TQ$ is very similar) and let C be an extension of A :

$$C = \begin{pmatrix} A & \mathbf{x} \\ \mathbf{y}^T & \alpha \end{pmatrix}.$$

We define the corresponding extension of B as $D = \begin{pmatrix} B & \mathbf{x}' \\ \mathbf{y}'^T & \alpha \end{pmatrix}$, where $\mathbf{x}' = P^T \mathbf{x}$ and $\mathbf{y}' = Q \mathbf{y}$. It suffices to show that C is a child if and only if D is a child. Note that $\det(A) = \det(P) \det(B) \det(Q) = \pm \det(B)$, and so

$$\begin{aligned} \det(D) &= \det(B)(\alpha - \mathbf{y}'^T B^{-1} \mathbf{x}') \\ &= \det(B)(\alpha - (Q \mathbf{y})^T (P^T A Q^T)^{-1} (P^T \mathbf{x})) \\ &= \det(B)(\alpha - \mathbf{y}^T Q^T (Q A^{-1} P) P^T \mathbf{x}) \\ &= \det(B)(\alpha - \mathbf{y}^T A^{-1} \mathbf{x}) = \pm \det(C). \end{aligned}$$

Thus, $D \in \Omega_{n+1}$ if and only if $C \in \Omega_{n+1}$. In addition, for extended permutations $\mathcal{P} = \begin{pmatrix} P & 0 \\ 0 & 1 \end{pmatrix}$ and $\mathcal{Q} = \begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix}$,

$$\mathcal{P} D \mathcal{Q} = \begin{pmatrix} P & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} B & P^T \mathbf{x} \\ \mathbf{y}^T Q^T & \alpha \end{pmatrix} \begin{pmatrix} Q & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} PBQ & PP^T \mathbf{x} \\ \mathbf{y}^T Q^T Q & \alpha \end{pmatrix} = \begin{pmatrix} A & \mathbf{x} \\ \mathbf{y}^T & \alpha \end{pmatrix} = C.$$

Algorithm 1 Generation of Strongly Non-Singular Representatives of Dimension $n + 1$

Require: $\Omega_n / \sim$: Set of canonical representatives of size $n \times n$ and their orbit counts N_A .
Ensure: $\Omega_{n+1} / \sim$: Set of canonical representatives of size $(n + 1) \times (n + 1)$ and counts.

```

1: Initialize an empty hash map  $\mathcal{H}$  to store  $\Omega_{n+1} / \sim$ 
2: for each  $(A, N_A) \in \Omega_n / \sim$  do
3:   for each extension  $\mathbf{x}, \mathbf{y} \in \{0, 1\}^n$  and  $\alpha \in \{0, 1\}$  do
4:      $C \leftarrow \begin{pmatrix} A & \mathbf{x} \\ \mathbf{y}^T & \alpha \end{pmatrix}$ 
5:     if  $\det(C) \neq 0$  then ▷ Computed via Bareiss/Faddeev-Leverrier
6:        $K \leftarrow \text{GETCANONICAL}(C)$ 
7:        $\mathcal{H}[K] \leftarrow \mathcal{H}[K] + N_A$  ▷ Aggregate counts for the equivalence class
8:     end if
9:   end for
10: end for
11: return  $\mathcal{H}$ 

```

Thus $C \sim D$, and the map $(\mathbf{x}, \mathbf{y}, \alpha) \mapsto (P^T \mathbf{x}, Q\mathbf{y}, \alpha)$ is a bijection between the extension spaces. The case $A = PB^T Q$ is similar, with corresponding bijection $(\mathbf{x}, \mathbf{y}, \alpha) \mapsto (Q\mathbf{y}, P^T \mathbf{x}, \alpha)$. $\square$

This lemma allows us to process only one representative per equivalence class. Let $[A]$ be the image of A in the map $\Omega_n \rightarrow \Omega_n / \sim$. If a representative A has count N_A (the size of its orbit times the number of previous parents), and produces a child C with multiplicity k (number of valid $(\mathbf{x}, \mathbf{y}, \alpha)$ extensions reducing to C), we add $k \cdot N_A$ to the count of the class of C . Thus we can construct $\Omega_{n+1} / \sim$ from $\Omega_n / \sim$. We detail this procedure in Algorithm 1.

To efficiently store and compare the equivalence classes, we require a canonical form function $\mathcal{K} : \{0, 1\}^{n \times n} \rightarrow \{0, 1\}^{n \times n}$ such that $A \sim B \iff \mathcal{K}(A) = \mathcal{K}(B)$. We compute this by mapping the matrix equivalence problem to the graph isomorphism problem, utilizing the individualization-refinement paradigm [11, 12].

4.2. Bipartite Graph Mapping. Let $A \in \{0, 1\}^{n \times n}$ be a binary matrix. We associate A with a bipartite graph $\Gamma(A) = (V, E)$, where the vertex set $V = R \cup C$ consists of row indices $R = \{r_1, \dots, r_n\}$ and column indices $C = \{c_1, \dots, c_n\}$. The edge set is defined by $E = \{(r_i, c_j) \mid A_{ij} = 1\}$.

The equivalence relation $A = PBQ$ (under row and column permutations) corresponds to the isomorphism of $\Gamma(A)$ and $\Gamma(B)$ under permutations that preserve the ordered partition (R, C) . To account for the transpose case $A = PB^T Q$, we must also consider isomorphisms that map $R \rightarrow C$ and $C \rightarrow R$.

We define the canonical form $\mathcal{K}(A)$ as the lexicographically minimal adjacency matrix obtained by reading the bits of the bipartite graph sequentially. To handle the transpose, we compute the canonical labeling twice: once starting with the ordered vertex partition (R, C) , and once with (C, R) . The global canonical form is the lexicographical minimum of these two results.

4.3. Partition Refinement (1-WL). To reduce the search space, we apply the 1-dimensional Weisfeiler-Leman (1-WL) algorithm [29], also known as color refinement, to transform an ordered partition π of the vertex set into an equitable partition $\mathcal{R}(\Gamma, \pi)$.

Definition 4.2 (Equitable Partition [11]). *A partition π is equitable with respect to a graph Γ if for all cells $V_i, V_j \in \pi$ and all vertices $u, v \in V_i$, the number of neighbors of u in V_j equals the number of neighbors of v in V_j .*

Algorithm 2 Canonical Form via Individualization-Refinement

```

1: procedure GETCANONICAL( $C$ )
2:    $\Gamma \leftarrow \text{BipartiteGraph}(C)$  ▷ Vertices  $V = R \cup C$  (rows and columns)
3:    $K_{RC} \leftarrow \text{SEARCHTREE}(\Gamma, (R, C), \emptyset, 0)$ 
4:    $K_{CR} \leftarrow \text{SEARCHTREE}(\Gamma, (C, R), \emptyset, 0)$ 
5:   return  $\min(K_{RC}, K_{CR})$ 
6: end procedure

7: procedure SEARCHTREE( $\Gamma, \pi, G, \text{depth}$ )
8:    $\pi_{\text{eq}} \leftarrow \text{1-WL}(\Gamma, \pi)$ 
9:   if  $\pi_{\text{eq}}$  is completely discrete then
10:     $M \leftarrow \text{AdjacencyMatrix}(\Gamma, \pi_{\text{eq}})$ 
11:    if  $M < M_{\text{best}}$  then
12:       $M_{\text{best}} \leftarrow M$ 
13:       $\pi_{\text{best}} \leftarrow \pi_{\text{eq}}$ 
14:    else if  $M == M_{\text{best}}$  then
15:       $\gamma \leftarrow \text{ExtractAutomorphism}(\pi_{\text{eq}}, \pi_{\text{best}})$ 
16:       $G \leftarrow \langle G, \gamma \rangle$  ▷ Add new generator to the automorphism group
17:    end if
18:    return
19:  end if
20:   $W \leftarrow \text{First non-trivial cell in } \pi_{\text{eq}}$ 
21:   $\mathcal{O} \leftarrow \text{Orbits}(W, G^{(\text{depth})})$  ▷ Group vertices by current stabilizer subgroup
22:  for each representative  $v \in \mathcal{O}$  do
23:     $\pi' \leftarrow \text{Individualize}(\pi_{\text{eq}}, v)$  ▷ Split  $v$  into its own singleton cell
24:     $\text{SEARCHTREE}(\Gamma, \pi', G, \text{depth} + 1)$ 
25:  end for
26:  return  $M_{\text{best}}$ 
27: end procedure

```

The refinement procedure $\mathcal{R}$ iteratively splits cells based on vertex degrees relative to other cells until the partition is equitable. It is a well-established result that this refinement procedure is a canonical invariant; that is, any isomorphism mapping $\Gamma(A)$ to $\Gamma(B)$ maps $\mathcal{R}(\Gamma(A), \pi)$ to $\mathcal{R}(\Gamma(B), \pi^\gamma)$ [11].

4.4. Search Tree and Automorphism Pruning. Because equitable partitions are not necessarily discrete, refinement alone cannot distinguish all vertices. We disambiguate symmetries using a backtracking search tree $T(\Gamma, \pi)$. Each node ν in the tree represents a partition. Child nodes are generated by fixing a vertex v from the first non-trivial cell and re-applying the refinement procedure $\mathcal{R}$.

To make this search computationally tractable, we employ the automorphism pruning strategy utilized by the **nauty** and **Traces** software packages [12]. During the traversal of the search tree, whenever a leaf node yields an adjacency matrix identical to the current lexicographical best, an automorphism of the graph is discovered. We maintain a generating set for the automorphism group $G = \text{Aut}(\Gamma)$ discovered during the search.

Lemma 4.3 (Pruning Lemma [12]). *Let ν_1 and ν_2 be nodes in the search tree T , where ν_1 is visited before ν_2 in the lexicographical search order. If there exists a discovered automorphism $\gamma \in G$ such that $\nu_2 = \nu_1^\gamma$, then the subtree rooted at ν_2 contains no canonical candidates lexicographically smaller than those in the subtree of ν_1 , and may be safely discarded.*

4.5. Group-Theoretic Computations. The efficiency of Lemma 4.3 relies on the ability to quickly compute orbits of the stabilizer subgroup without enumerating all group elements. Following standard computational group theory practices [5], we represent the automorphism group G using a stabilizer chain.

When branching on a cell $W \in \pi$ at depth k in the search tree, we compute the orbits of W under the current stabilizer subgroup $G^{(k)}$. This is done efficiently using a Union-Find data structure on the Schreier generator graph [5, Section 3.6]. By branching on only one representative vertex from each orbit, we drastically prune symmetric branches of the search tree. New automorphisms discovered at the leaves are integrated into the stabilizer chain in polynomial time.

4.6. Computing Canonical Matrices. Combining the ingredients of Subsections 4.2, 4.3, 4.4, and 4.5, we describe our procedure for computing canonical matrices in Algorithm 2.

Theorem 4.4. *The algorithm correctly computes the canonical form $\mathcal{K}(A)$, returning the exact lexicographical minimum of the equivalence class of A under row permutations, column permutations, and transposition.*

Proof. The backtracking search tree exhaustively covers the permutation space $S_n \times S_n$. The 1-WL refinement restricts this space without eliminating valid isomorphisms [11]. Furthermore, the orbit pruning strategy is mathematically guaranteed to preserve at least one path to the lexicographically minimal leaf [12, Theorem 2.15]. By executing this search for both initial partitions (R, C) and (C, R) and taking the minimum, the algorithm correctly accounts for the transpose equivalence, ensuring the true global minimum of the equivalence class is found. $\square$

4.7. Exact Counts for $n \leq 9$. Tables 1 and 2 present the exact number of strongly non-singular $n \times n$ binary matrices, $N_{n,k}$, parameterized by the number of non-zero entries k , for dimensions $n \in \{1, \dots, 9\}$. The full dataset in JSON format is hosted at [15]. We note that, by computing a representative for each equivalence class and the count of each class, one may compute any statistics about the set, not just non-zero counts. The interested user can do so simply by running the publicly available code detailed in this section and available at [15].

5. AN UPPER BOUND FOR STRONGLY NON-SINGULAR $\{0, 1\}$ -MATRICES

Recall $N_n(p) = \mathbb{P}[M_n(\text{Ber}(p)) \text{ is strongly non-singular}]$ and, using Tables 1 and 2, is exactly computable for $n \leq 9$ (see Figure 2). In particular,

$$N_n(p) = \sum_{k=0}^{n^2} N_{n,k} p^k (1-p)^{n^2-k},$$

where $N_{n,k}$ is the number of $n \times n$ strongly non-singular $\{0, 1\}$ -matrices with exactly k ones. For example, the total number of 9×9 strongly non-singular $\{0, 1\}$ -matrices is

$$\sum_{k=9}^{73} N_{9,k} = 36,646,054,311,185,413,881,216,$$

and therefore

$$N_9(1/2) = \frac{36,646,054,311,185,413,881,216}{2^{81}} \approx 0.015156.$$

Here, using the data $N_{n,k}$ for $n = 9$ from Table 2, we produce an upper bound for $N_n(p)$ for $n > 9$.

Theorem 5.1. *Let $n > 9$ and $p \in (0, 1)$. Define $q = 1 - p$, $\alpha = \min\{p, q\}$, and $\beta = \max\{p, q\}$. Then*

$$N_n(p) \leq \sum_{w=9}^{73} N_{9,w} p^w q^{81-w} \prod_{k \in T_p} [1 - (2q^k - q^{2k-1} + 2((p^2 + q^2)\alpha^{k-10} - q^k \beta^{k-10})g(w) - \beta^{2k-19}g(w)^2)],$$

n	k	Count	n	k	Count	n	k	Count
1	1	1	6	21	270526947	8	10	3514
2	2	1	6	22	188703604	8	11	97244
2	3	3	6	23	115926816	8	12	1986607
3	3	1	6	24	58314572	8	13	31504564
3	4	9	6	25	25124264	8	14	395478728
3	5	29	6	26	8586990	8	15	3943148852
3	6	19	6	27	2411996	8	16	31369251177
3	7	10	6	28	484830	8	17	203223435056
4	4	1	6	29	72648	8	18	1084492529346
4	5	18	6	30	6484	8	19	4890576176528
4	6	143	6	31	396	8	20	18805194537373
4	7	592	7	7	1	8	21	63047260972722
4	8	1061	7	8	63	8	22	185050195830538
4	9	1380	7	9	1960	8	23	484312155800218
4	10	961	7	10	39830	8	24	1130001175634995
4	11	644	7	11	585347	8	25	2388141629501388
4	12	174	7	12	6458571	8	26	4557722772711510
4	13	34	7	13	53744776	8	27	7957465424434324
5	5	1	7	14	337996072	8	28	12664141178519307
5	6	30	7	15	1675103574	8	29	18570778015470134
5	7	425	7	16	6614988082	8	30	24957086783562282
5	8	3640	7	17	21743417216	8	31	31061367927548198
5	9	19512	7	18	59637219644	8	32	35567140229358595
5	10	62018	7	19	141545976180	8	33	37814938253601916
5	11	144352	7	20	288911603858	8	34	37074021913720550
5	12	232426	7	21	521858329294	8	35	33779665724027360
5	13	308685	7	22	827024639200	8	36	28393427281756235
5	14	306022	7	23	1176966346333	8	37	22180209800850350
5	15	246385	7	24	1482186878151	8	38	15968137418716642
5	16	160032	7	25	1693609745344	8	39	10660173316494306
5	17	81122	7	26	1724441624018	8	40	6535322648553720
5	18	29238	7	27	1594098556619	8	41	3700038252393912
5	19	8074	7	28	1317019173583	8	42	1911489482762620
5	20	1154	7	29	988559392212	8	43	905530224323564
5	21	116	7	30	661909038532	8	44	388067409532546
6	6	1	7	31	400500045128	8	45	150995283933710
6	7	45	7	32	214045604384	8	46	52554074390450
6	8	985	7	33	101997188760	8	47	16368776680422
6	9	13745	7	34	42257609448	8	48	4468830973592
6	10	132987	7	35	15337468014	8	49	1067165274292
6	11	900789	7	36	4735733876	8	50	217193007336
6	12	4226503	7	37	1234197450	8	51	37312365232
6	13	15051553	7	38	261591224	8	52	5233480948
6	14	40635867	7	39	44391892	8	53	591886560
6	15	90221539	7	40	5601004	8	54	51707652
6	16	161124379	7	41	536708	8	55	3482368
6	17	247776435	7	42	32972	8	56	156888
6	18	313666787	7	43	1352	8	57	4616
6	19	351245123	8	8	1			
6	20	329076355	8	9	84			

TABLE 1. The exact values of $N_{n,k}$ for $n \in \{1, \dots, 8\}$

k	Count	k	Count	k	Count
9	1	31	96719476982407744092	53	83333465510680978826
10	108	32	186503724910597960072	54	40330856830106979690
11	5838	33	335422734825487477468	55	18193563381572837212
12	210252	34	562215751430485988940	56	7606822625464081816
13	5652213	35	882789394906258039156	57	2945819219466818342
14	120044988	36	1296704229926641416276	58	1049968464418967194
15	2072228096	37	1790188101242490383464	59	343639220977425084
16	29407465584	38	2318475316973368470040	60	102439451887121968
17	344097864320	39	2828243308508161667208	61	27695978794734600
18	3333236753396	40	3243012884297962327568	62	6722470912110816
19	27016651978056	41	3507731778243932969769	63	1455146825624512
20	184889958771428	42	3570249044441509661900	64	277089814762980
21	1085361104090754	43	3430496716635991967290	65	45966172670268
22	5517513084920674	44	3103130069080774429632	66	6526518924368
23	24656256324744968	45	2649888111684432089913	67	783702562020
24	97542620062118552	46	2130133085694111554620	68	77872212440
25	345944420050724384	47	1615543159590023404340	69	6345678472
26	1104797971319879050	48	1152109135371983825440	70	409749896
27	3209822652195236944	49	774135602738705469776	71	20609608
28	8504988786860401576	50	488154989553138214988	72	711944
29	20726783710640352922	51	289298664178677485460	73	15760
30	46495824680544165784	52	160449071628568146100		

TABLE 2. The exact values of $N_{n,k}$ for $n = 9$

where $g(w) = 9p^{\frac{w}{9}}q^{9-\frac{w}{9}}$ and

$$T_p = \{k \in \mathbb{N} : 10 \leq k \leq n, (p^2 + q^2)\alpha^{k-10} \geq q^k \beta^{k-10} + (p + 8\alpha)\beta^{2k-11}\}.$$

Proof. Let $M := M_n(\text{Ber}(p))$ and let M_k denote the leading $k \times k$ submatrix of M . To compute an upper bound for $N_n(p)$ for $n > 9$, we condition on the realization of M_9 , and utilize data regarding its non-zero entries. We produce an upper bound for the weaker condition that M_9 is strongly non-singular, M has no zero columns or zero rows, and that none of the first nine rows and columns of M are repeated. Let S be the set of all 9×9 strongly non-singular matrices, $\mathbf{r}_k$ and $\mathbf{c}_k$ be the k^{th} row and k^{th} column, respectively, of M_k , and

$$\mathcal{E}_k = \{\mathbf{r}_k = \mathbf{0}\} \cup \{\mathbf{c}_k = \mathbf{0}\} \cup \bigcup_{i=1}^9 \{\mathbf{r}_k = \mathbf{r}_i\} \cup \bigcup_{i=1}^9 \{\mathbf{c}_k = \mathbf{c}_i\}.$$

For a fixed $A \in S$, strong non-singularity implies $\mathcal{E}_k^c$ for every $k \geq 10$. Therefore,

$$\mathbb{P}[M \text{ is strongly non-singular} \mid M_9 = A] \leq \prod_{k=10}^n (1 - \mathbb{P}[\mathcal{E}_k \mid M_9 = A, \mathcal{E}_{10}^c, \dots, \mathcal{E}_{k-1}^c])$$

Consequently,

$$N_n(p) \leq \sum_{A \in S} \mathbb{P}[M_9 = A] \prod_{k=10}^n (1 - \mathbb{P}[\mathcal{E}_k \mid M_9 = A, \mathcal{E}_{10}^c, \dots, \mathcal{E}_{k-1}^c])$$

and so it suffices to lower bound

$$\mathbb{P}[\mathcal{E}_k \mid M_9 = A, \mathcal{E}_{10}^c, \dots, \mathcal{E}_{k-1}^c]$$

We do so by inclusion-exclusion, and note that no three of the events $\{\mathbf{r}_k = \mathbf{0}\}$, $\{\mathbf{c}_k = \mathbf{0}\}$, $\{\mathbf{r}_k = \mathbf{r}_i\}$, $i = 1, \dots, 9$, and $\{\mathbf{c}_k = \mathbf{c}_i\}$, $i = 1, \dots, 9$ can occur simultaneously, as A has distinct non-zero rows and

columns, and $\mathbf{c}_k$ and $\mathbf{r}_k$ can each take only one value. In addition, events $\{\mathbf{r}_k = \mathbf{r}_i\}$ and $\{\mathbf{r}_k = \mathbf{r}_j\}$ are disjoint for $i, j \in \{1, \dots, 9\}$, $i \neq j$, as A has independent rows (and similarly for column events). Let $\hat{\mathcal{E}}_k$ be the event $M_9 = A$ and $\mathcal{E}_{10}^c, \dots, \mathcal{E}_k^c$. Let ρ_i and σ_i denote the number of non-zero entries in row and column i of A , respectively. Then

$$\begin{aligned} \mathbb{P}[\mathbf{r}_k = \mathbf{0} \mid \hat{\mathcal{E}}_{k-1}] &= \mathbb{P}[\mathbf{c}_k = \mathbf{0} \mid \hat{\mathcal{E}}_{k-1}] = q^k, \\ \mathbb{P}[\mathbf{r}_k = \mathbf{r}_i \mid \hat{\mathcal{E}}_{k-1}] &\geq (p^2 + q^2)\alpha^{k-10}p^{\rho_i}q^{9-\rho_i}, \\ \mathbb{P}[\mathbf{c}_k = \mathbf{c}_i \mid \hat{\mathcal{E}}_{k-1}] &\geq (p^2 + q^2)\alpha^{k-10}p^{\sigma_i}q^{9-\sigma_i}, \\ \mathbb{P}[\mathbf{r}_k = \mathbf{0}, \mathbf{c}_k = \mathbf{0} \mid \hat{\mathcal{E}}_{k-1}] &= q^{2k-1}, \\ \mathbb{P}[\mathbf{r}_k = \mathbf{0}, \mathbf{c}_k = \mathbf{c}_i \mid \hat{\mathcal{E}}_{k-1}] &\leq q^k\beta^{k-10}p^{\sigma_i}q^{9-\sigma_i}, \\ \mathbb{P}[\mathbf{r}_k = \mathbf{r}_i, \mathbf{c}_k = \mathbf{0} \mid \hat{\mathcal{E}}_{k-1}] &\leq q^k\beta^{k-10}p^{\rho_i}q^{9-\rho_i}, \\ \mathbb{P}[\mathbf{r}_k = \mathbf{r}_i, \mathbf{c}_k = \mathbf{c}_j \mid \hat{\mathcal{E}}_{k-1}] &\leq \beta^{2k-19}p^{\rho_i+\sigma_j}q^{18-\rho_i-\sigma_j}, \end{aligned}$$

where our inequalities are the result of the following analysis:

- $\{\mathbf{r}_k = \mathbf{r}_i\}$ ($\{\mathbf{c}_k = \mathbf{c}_i\}$ is similar): At step k , the target row $\mathbf{r}_i$ consists of the fixed i^{th} row of A , a continuation in the 10^{th} to $(k-1)^{\text{th}}$ entries, and the new entry $M_{i,k}$. The new row $\mathbf{r}_k$ matches $\mathbf{r}_i$ if the first nine entries of $\mathbf{r}_k$ match the i^{th} row of A , a probability $p^{\rho_i}q^{9-\rho_i}$ event, the middle $k-10$ entries match, at least an α^{k-10} probability event, and the last entry matches, a probability $p^2 + q^2$ event.
- $\{\mathbf{r}_k = \mathbf{0}, \mathbf{c}_k = \mathbf{c}_i\}$ ($\{\mathbf{r}_k = \mathbf{r}_i, \mathbf{c}_k = \mathbf{0}\}$ is similar): The event $\{\mathbf{r}_k = \mathbf{0}\}$ has probability q^k , and, conditioning on $M_{k,i} = M_{k,k} = 0$, the probability of $\{\mathbf{c}_k = \mathbf{c}_i\}$ is at most $p^{\sigma_i}q^{9-\sigma_i}\beta^{k-10}$.
- $\{\mathbf{r}_k = \mathbf{r}_i, \mathbf{c}_k = \mathbf{c}_j\}$: Both events holding requires $M_{k,k} = A_{i,j}$, a probability at most β event. The first nine entries of $\mathbf{r}_k$ and $\mathbf{c}_k$ match $\mathbf{r}_i$ and $\mathbf{c}_j$ with probability $p^{\rho_i+\sigma_j}q^{18-\rho_i-\sigma_j}$. The remaining $k-10$ entries of $\mathbf{r}_k$ and $\mathbf{c}_k$ match $\mathbf{r}_i$ and $\mathbf{c}_j$ with probability at most $\beta^{2(k-10)}$.

Now, let $R(\rho) = \sum_{i=1}^9 p^{\rho_i}q^{9-\rho_i}$ and $C(\sigma) = \sum_{i=1}^9 p^{\sigma_i}q^{9-\sigma_i}$. We obtain the lower bound

$$\mathbb{P}[\mathcal{E}_k \mid \hat{\mathcal{E}}_{k-1}] \geq 2q^k + (p^2 + q^2)\alpha^{k-10}(R(\rho) + C(\sigma)) - q^{2k-1} - q^k\beta^{k-10}(R(\rho) + C(\sigma)) - \beta^{2k-19}R(\rho)C(\sigma)$$

Note that $R(\rho)$ and $C(\sigma)$ are both upper bounded by $W := (p+8\alpha)\beta^8$. To see this, let $f(t) = p^tq^{9-t}$. If $p \leq q$, then f is decreasing, and since every row of A is nonzero, $\rho_i \geq 1$ for all i , so

$$R(\rho) = \sum_{i=1}^9 f(\rho_i) \leq 9f(1) = 9pq^8 = (p+8\alpha)\beta^8.$$

If $p \geq q$, then f is increasing, and since the rows of A are pairwise distinct, at most one row can have $\rho_i = 9$ while the remaining eight rows have $\rho_i \leq 8$, so

$$R(\rho) \leq f(9) + 8f(8) = p^9 + 8p^8q = (p+8\alpha)\beta^8.$$

The same argument applies to $C(\sigma)$. Let

$$F_k(x, y) := 2q^k - q^{2k-1} + ((p^2 + q^2)\alpha^{k-10} - q^k\beta^{k-10})(x+y) - \beta^{2k-19}xy.$$

The lower bound above says

$$\mathbb{P}[\mathcal{E}_k \mid \hat{\mathcal{E}}_{k-1}] \geq F_k(R(\rho), C(\sigma)).$$

If $k \in T_p$, then the defining inequality for T_p gives

$$(p^2 + q^2)\alpha^{k-10} - q^k\beta^{k-10} \geq (p+8\alpha)\beta^{2k-11} = \beta^{2k-19}W.$$

Thus, for $0 \leq x, y \leq W$,

$$\begin{aligned}\frac{\partial F_k}{\partial x}(x, y) &= ((p^2 + q^2)\alpha^{k-10} - q^k \beta^{k-10}) - \beta^{2k-19}y \geq 0, \\ \frac{\partial F_k}{\partial y}(x, y) &= ((p^2 + q^2)\alpha^{k-10} - q^k \beta^{k-10}) - \beta^{2k-19}x \geq 0.\end{aligned}$$

Hence F_k is non-decreasing in each variable on $[0, W]^2$.

Now let $w := \text{nnz}(A)$. The function $f(r) = p^r q^{9-r}$ is convex, and $\rho_1 + \dots + \rho_9 = w$; therefore Jensen's inequality gives $R(\rho) \geq g(w)$. Similarly, since $\sigma_1 + \dots + \sigma_9 = w$, we have $C(\sigma) \geq g(w)$. Because $R(\rho), C(\sigma) \leq W$ and F_k is nondecreasing in each variable for $k \in T_p$, we obtain

$$\mathbb{P}[\mathcal{E}_k \mid \hat{\mathcal{E}}_{k-1}] \geq F_k(g(w), g(w)),$$

i.e.

$$\mathbb{P}[\mathcal{E}_k \mid \hat{\mathcal{E}}_{k-1}] \geq 2q^k - q^{2k-1} + 2((p^2 + q^2)\alpha^{k-10} - q^k \beta^{k-10})g(w) - \beta^{2k-19}g(w)^2.$$

Since each factor $1 - \mathbb{P}[\mathcal{E}_k \mid \hat{\mathcal{E}}_{k-1}]$ lies in $[0, 1]$, omitting the factors with $k \notin T_p$ can only increase the product. Grouping the matrices $A \in S$ by their number w of non-zero entries gives the stated bound. $\square$

We plot the upper bound of Theorem 5.1 and experimental results for $n = 30$ in Figure 3. The above bound simplifies significantly for $p = 1/2$, the case of greatest interest.

Corollary 5.2. *Let $n > 9$. Then*

$$N_n(1/2) \leq \frac{N_9}{2^{81}} \prod_{k=10}^n \left(1 - \frac{20}{2^k} + \frac{200}{2^{2k}}\right).$$

Proof. When $p = 1/2$, $g(w) = 9 \times 2^{-9}$ for every w , and $T_p = \{10, \dots, n\}$. Because $p = q = \alpha = \beta = 1/2$, each factor in the product simplifies as

$$1 - (2q^k - q^{2k-1} + 2((p^2 + q^2)\alpha^{k-10} - q^k \beta^{k-10})g(w) - \beta^{2k-19}g(w)^2) = 1 - \frac{20}{2^k} + \frac{200}{2^{2k}}.$$

$\square$

For $n = 30$ (the dimension used in Figure 3), Corollary 5.2 gives the reported upper bound of approximately 1.45...%.

ACKNOWLEDGEMENTS

This project was conducted as part of the MIT Summer Research Program (MSRP) during the summer of 2025. The authors thank Louisa Thomas for improving the style of presentation, and Anna Brandenburger and Byron Chin for their insightful discussions. We would like to thank Mehtaab Sawhney for his suggestions regarding proof techniques for Corollary 1.2 and Dmitrii Zakharov for pointing us to Elkes-Szabó type theorems, which proved invaluable for proving Lemma 3.6. Dmitrii also suggested that an alternate technique, involving the Szemerédi-Trotter theorem, may be possible. This material is based upon work supported by the National Science Foundation under grant no. DMS-2513687.

REFERENCES

- [1] Ankit Bisain, Alan Edelman, and John Urschel. A new upper bound for the growth factor in Gaussian elimination with complete pivoting. *Bulletin of the London Mathematical Society*, 2025.
- [2] Jean Bourgain, Van H Vu, and Philip Matchett Wood. On the singularity probability of discrete random matrices. *Journal of Functional Analysis*, 258(2):559–603, 2010.
- [3] Artem Chernikov, Ya’acov Peterzil, and Sergei Starchenko. Model-theoretic Elekes–Szabó for stable and o-minimal hypergraphs. *Duke Mathematical Journal*, 173(3):419–512, 2024.
- [4] James Demmel, Laura Grigori, and Alexander Rusciano. An improved analysis and unified perspective on deterministic and randomized low-rank matrix approximation. *SIAM Journal on Matrix Analysis and Applications*, 44(2):559–591, 2023.
- [5] John D. Dixon and Brian Mortimer. *Permutation Groups*, volume 163 of *Graduate Texts in Mathematics*. Springer New York, NY, 1996.
- [6] Alan Edelman and John Urschel. Some new results on the maximum growth factor in Gaussian elimination. *SIAM Journal on Matrix Analysis and Applications*, 45(2):967–991, 2024.
- [7] Han Huang and Konstantin Tikhomirov. Average-case analysis of the Gaussian elimination with partial pivoting. *Probability Theory and Related Fields*, pages 1–67, 2024.
- [8] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney. Singularity of discrete random matrices. *Geometric and Functional Analysis*, 31(5):1160–1218, 2021.
- [9] Jeff Kahn, János Komlós, and Endre Szemerédi. On the probability that a random ± 1 -matrix is singular. *Journal of the American Mathematical Society*, 8(1):223–240, 1995.
- [10] János Komlós. On the determinant of (0-1) matrices. *Studia Scientiarum Mathematicarum Hungarica*, 2:7–21, 1967.
- [11] Brendan D McKay. Practical graph isomorphism. *Congressus Numerantium*, 30:45–87, 1981.
- [12] Brendan D McKay and Adolfo Piperno. Practical graph isomorphism, II. *Journal of Symbolic Computation*, 60:94–112, 2014.
- [13] OEIS Foundation Inc. The on-line encyclopedia of integer sequences, sequence A125587. <https://oeis.org/A125587>, 2026.
- [14] Alan V Oppenheim. *Discrete-time signal processing*. Pearson Education India, 1999.
- [15] Samuel Orellana Mateo, John Urschel, and Nicholas West. LU-factorization-of-discrete-random-matrices. <https://github.com/Samuel-O-M/lu-factorization-of-discrete-random-matrices>, 2026.
- [16] Douglass Stott Parker. *Random butterfly transformations with applications in computational linear algebra*. UCLA Computer Science Department, 1995.
- [17] John Peca-Medlin and Thomas Trogdon. Growth factors of random butterfly matrices and the stability of avoiding pivoting. *SIAM Journal on Matrix Analysis and Applications*, 44(3):945–970, 2023.
- [18] Boris Alexeyevich Rogozin. An estimate for concentration functions. *Theory of Probability & Its Applications*, 6(1):94–97, 1961.
- [19] Arvind Sankar, Daniel A Spielman, and Shang-Hua Teng. Smoothed analysis of the condition numbers and growth factors of matrices. *SIAM Journal on Matrix Analysis and Applications*, 28(2):446–476, 2006.
- [20] Terence Tao and Van Vu. On random ± 1 matrices: singularity and determinant. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 431–440, 2005.
- [21] Terence Tao and Van Vu. On the singularity probability of random Bernoulli matrices. *Journal of the American Mathematical Society*, 20(3):603–628, 2007.
- [22] Terence Tao and Van H Vu. *Additive combinatorics*, volume 105. Cambridge University Press, 2006.
- [23] Konstantin Tikhomirov. Singularity of random Bernoulli matrices. *Annals of Mathematics*, 191(2):593–634, 2020.
- [24] Lloyd N Trefethen. The smart money’s on numerical analysts. *SIAM News*, 45(9):1–5, 2012.
- [25] Lloyd N Trefethen and David Bau. *Numerical linear algebra*. SIAM, 2022.
- [26] Lloyd N Trefethen and Robert S Schreiber. Average-case stability of Gaussian elimination. *SIAM Journal on Matrix Analysis and Applications*, 11(3):335–360, 1990.
- [27] John Urschel. Numerical stability in Gaussian elimination. *NOTICES OF THE AMERICAN MATHEMATICAL SOCIETY*, 72(6).
- [28] John Von Neumann and Herman Heine Goldstine. Numerical inverting of matrices of high order. 1947.
- [29] Boris Weisfeiler and Andrei A. Lehman. A reduction of a graph to a canonical form and an algebra arising during this reduction. *Nauchno-Tekhnicheskaya Informatsiya, Ser. 2*, 9:12–16, 1968.
- [30] James Hardy Wilkinson. Error analysis of direct methods of matrix inversion. *Journal of the ACM (JACM)*, 8(3):281–330, 1961.
- [31] J.H. Wilkinson. *The Algebraic Eigenvalue Problem*. Clarendon Press, Oxford, 1965.
- [32] Man-Chung Yeung and Tony F Chan. Probabilistic analysis of Gaussian elimination without pivoting. *SIAM Journal on Matrix Analysis and Applications*, 18(2):499–517, 1997.