

ON THE GROWTH FACTOR OF RANDOM MATRICES

JOHN URSCHTEL

ABSTRACT. Gaussian elimination is the oldest and most popular method for solving a linear system. Its numerical stability for a given matrix is controlled by the growth factor, a measure of how large entries can become during elimination. In this work, we prove a number of new results regarding the growth factor of random matrices. First, we provide tight estimates for the growth factor of a matrix preconditioned by a Haar orthogonal matrix without pivoting and characterize the asymptotic distribution of the growth factor of Gaussian matrices without pivoting. Second, and most notably, we prove that the growth factor of an $n \times n$ Gaussian matrix under partial pivoting is rarely much larger than $\sqrt{n}$, resolving an old conjecture of Nick Trefethen. The same techniques used to prove this conjecture also provide an improved smoothed analysis of the growth factor under partial pivoting.

1. INTRODUCTION

Gaussian elimination is the premier technique for solving a linear system $A\mathbf{x} = \mathbf{b}$, by factoring A into the product of a lower and upper triangular matrix. In particular, given an $n \times n$ matrix A with non-zero leading principal minors, the procedure performs a series of rank-one transformations

$$A \rightarrow A/A_{[1],[1]} \rightarrow A/A_{[2],[2]} \rightarrow \dots \rightarrow A/A_{[n-1],[n-1]}.$$

The resulting LU factorization is encoded in the first row and column of the iterates $A/A_{[k],[k]}$,

$$Le_{k+1} = \begin{bmatrix} 0 \\ (A/A_{[k],[k]})_{11}^{-1} (A/A_{[k],[k]})e_1 \end{bmatrix} \quad \text{and} \quad e_{k+1}^T U = \begin{bmatrix} 0 & e_1^T (A/A_{[k],[k]}) \end{bmatrix}, \quad (1.1)$$

where M/X is the Schur complement of M with respect to the submatrix X , and $[n] = \{1, \dots, n\}$. While this algorithm is exact in exact arithmetic, issues may still arise in finite precision due to the accumulation of round-off error. The error in Gaussian elimination in finite precision is controlled by the growth factor

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

where $\|\cdot\|_{\max}$ is the maximum magnitude entry. If $\text{growth}(A)$ is not too large, then the LU factorization is stable¹ (e.g., roughly $\log(n \text{growth}(A))$ bits of precision suffice); see [8, Theorem 9.3 & Lemma 9.6]

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

Email address: urschel@mit.edu.

2020 *Mathematics Subject Classification.* 65F05, 15A23, 60B20.

Key words and phrases. Gaussian elimination, LU factorization, growth factor, partial pivoting, smoothed analysis.

¹One subtle point, which is often glossed over in the analysis of Gaussian elimination, is that the growth factor needed for error analysis is the growth factor under floating point arithmetic. In some cases, this quantity can be quite different from the growth factor in exact arithmetic, due to an accumulation of round-off error or a change in pivot selection. Still, for mathematical neatness, the majority of results regarding the growth factor are for exact arithmetic. There is little harm in this: The worst-case growth factor for partial pivoting is the same in exact and floating point arithmetic, and, given a reasonable number of bits of precision, the worst-case growth factor for complete pivoting is barely changed by floating point arithmetic [6, Theorem 1.5], and the growth factor of a Gaussian matrix under partial pivoting is increased by at most a small multiplicative constant with high probability [10, Section 7].

for details. For this reason, a great deal of research regards the behavior of the growth factor under various pivoting and preconditioning strategies.

1.1. Partial Pivoting. Gaussian elimination with partial pivoting is the standard linear system solver in nearly every programming language, including the `numpy.linalg.solve` in Python and the backslash command for a general matrix in Matlab and Julia. The partial pivoting procedure is a greedy rule for reordering rows during elimination to avoid creating large entries. In particular, at each step of elimination, the first row is (possibly) swapped with another, so that the largest magnitude entry in the first column lies in the first row. While algorithmically this is done iteratively, it is represented mathematically by left-multiplication by a single permutation matrix P that satisfies

$$|((PA)/(PA)_{[k],[k]})_{11}| = \|((PA)/(PA)_{[k],[k]})e_1\|_\infty \quad \text{for } k = 0, 1, \dots, n-1. \quad (1.2)$$

Let $\Pi(\cdot) : \text{GL}_n(\mathbb{R}) \rightarrow \text{GL}_n(\mathbb{R})$ denote the partial pivoting map, which left-multiplies a non-singular matrix A by the permutation matrix satisfying (1.2). If there are multiple, $\Pi(\cdot)$ breaks ties at each stage of elimination by choosing the eligible row with smallest index in the current row ordering. However, in this paper, such ambiguities occur with probability zero.

The growth factor under partial pivoting can be exponentially large. A fairly simple doubling argument shows that the growth factor under partial pivoting is at most 2^{n-1} . In 1961, Wilkinson produced a matrix construction that achieves this quantity in every dimension [27]. In spite of this, the relatively low computational cost and memory access requirements for implementing the partial pivoting rule (1.2), and the apparent rarity of a large growth factor, have made Gaussian elimination with partial pivoting the preferred method for solving a linear system. In an attempt to provide an explanation for this apparent rarity, Trefethen and Schreiber performed an empirical average-case analysis of Gaussian elimination under various pivoting strategies [23]. A matrix $G \in \mathbb{R}^{m \times n}$ is said to be Gaussian if its entries are independent standard normal random variables. Inspired by their analysis, Edelman performed further numerical experiments for Gaussian matrices under partial pivoting. He noted that the growth factor appeared to grow like $n^{\frac{1}{2}}$, but remarked that “proving such a result to be true is probably very difficult” [5]. Trefethen also observed this behavior and made the following conjecture, first at the 1995 International Congress on Industrial and Applied Mathematics, and then in his 1997 text *Numerical Linear Algebra*, coauthored with David Bau [22, pg. 169].

Conjecture 1.1 (Trefethen, [22, pg. 169]). *Let $G \in \mathbb{R}^{n \times n}$ be a Gaussian matrix and $\Pi(\cdot)$ be the partial pivoting map. For any $\alpha, M > 0$,*

$$\mathbb{P} \left[\text{growth}(\Pi(G)) > n^{\frac{1}{2} + \alpha} \right] < n^{-M}$$

for all n sufficiently large.

Trefethen later offered \$1000 for a resolution of Conjecture 1.1 [24]. The numerical evidence for this conjecture is overwhelming; see [22, Figures 22.1 & 22.2] and their improvement in [4]. The mathematical progress has been made on two fronts. First, Trefethen himself claimed an unpublished proof of the needed bound for the last pivot $|U_{nn}|$ [24], which was released 14 years later, in parallel with this manuscript [21]. Second, Huang and Tikhomirov proved that a Gaussian matrix under partial pivoting has growth at most n^c with high probability, for a large unspecified constant c [10].

Another approach aimed at the stability of partial pivoting is smoothed analysis, which considers the performance of an algorithm under small random perturbations to the input. Sankar proved that any matrix plus a small Gaussian perturbation, $A + \epsilon G$ where $\|A\|_2 \leq 1$, has growth factor under partial pivoting bounded by $(n/\epsilon)^{O(\log n)}$ with high probability [17]. For more information regarding Gaussian elimination with partial pivoting, we refer the reader to [22, 25].

1.2. Alternatives to Partial Pivoting. Complete pivoting has a growth factor which is only quasi-polynomial; see [1, 6, 19, 27] for details. However, due to the additional computational complexity required, partial pivoting remains preferred in practice. Demmel, Grigori, and Rusciano considered the benefits of avoiding pivoting altogether, by preconditioning a matrix via pre- and post-multiplication by a Haar orthogonal matrix. They proved that, for any $A \in \text{GL}_n(\mathbb{R})$, the expected log growth factor of $Q_1 A Q_2$, for independent $Q_1, Q_2 \sim \text{Haar}(\text{O}(n))$, is $O(\log n)$ [3, Corollary 5.19]. Butterfly matrices have also been considered as a potential preconditioner to limit growth [14, 15] and have the benefit of being easier to apply. More generally, the possibility of avoiding pivoting or preconditioning altogether has been considered for certain random matrix models. For instance, Yeung and Chan proved probabilistic bounds on $\|U\|_\infty/\|A\|_\infty$ and $\|L\|_\infty$ for a Gaussian matrix without pivoting [28]. This analysis was tightened and generalized to the smoothed analysis setting by Sankar, Spielman, and Teng [18]. In particular, for Gaussian elimination without pivoting $A + \epsilon G = LU$, with $\|A\|_2 \leq 1$ and $\epsilon \in (0, 1)$, they proved that

$$\mathbb{P}[\|L\|_\infty > x] \lesssim \frac{n^2}{x} \left(\frac{1}{\epsilon} + \sqrt{\log n} \right) \quad \text{and} \quad \mathbb{P} \left[\frac{\|U\|_\infty}{\|A + \epsilon G\|_\infty} > 1 + x \right] \lesssim \frac{1}{x} \min \left\{ \frac{n^2}{\epsilon}, n^{\frac{3}{2}} + \frac{n^{\frac{1}{2}}}{\epsilon^2} \right\}.$$

Somewhat remarkably, even matrices with entries sampled independently from a non-degenerate discrete distribution with finite support have a polynomial growth factor with constant probability [11].

1.3. Our Contributions. Here we prove a number of results regarding the growth factor of random matrices, both without pivoting and with partial pivoting. First, we prove an upper bound on the growth factor of a matrix left-multiplied by a Haar orthogonal matrix without pivoting.

Theorem 1.2. *Let $A \in \text{GL}_n(\mathbb{R})$ and $Q \sim \text{Haar}(\text{O}(n))$. Then*

$$\mathbb{P} \left[\text{growth}(QA) > \frac{14 n^{\frac{3}{2}} \sqrt{\log(en)}}{\delta} \right] < \delta \quad \text{for all } \delta \in (0, 1).$$

The $\frac{3}{2}$ exponent in the previous theorem is tight for any non-adaptive orthogonal preconditioner, illustrated by the asymptotic distribution of the growth factor of a Gaussian matrix without pivoting.

Theorem 1.3. *Let $G \in \mathbb{R}^{n \times n}$ be a Gaussian matrix. Then $n^{-\frac{3}{2}} \text{growth}(G) \xrightarrow{d} X$, where $X \sim \text{InvExp}(\frac{2}{3\sqrt{\pi}})$, i.e., $\Pr[X \leq x] = e^{-\frac{2}{3\sqrt{\pi}x}}$ for $x > 0$.*

Numerical experiments suggest that this convergence occurs quite rapidly, even for small values of n ; see Figure 1. While Theorems 1.2 and 1.3 are fairly straightforward to prove, the results themselves provide further context for the situation regarding partial pivoting, the main focus of this work. In particular, we note that orthogonal preconditioning, a reasonable alternative to partial pivoting, produces a larger growth factor than one expects to see with partial pivoting when applied to a Gaussian matrix. Furthermore, the tails of the bounds decay inverse polynomially, which is slower than one expects for partial pivoting. The following theorem, our main result, proves that the growth factor of a Gaussian matrix is at most $n^{\frac{1}{2} + o(1)}$ with overwhelming probability, that is, $1 - n^{-\alpha}$ for any α .

Theorem 1.4. *For every $\alpha > 0$, there exist constants C_α, n_α such that*

$$\mathbb{P} \left[\text{growth}(\Pi(G)) > \sqrt{n} e^{C_\alpha \sqrt{\log n}} \right] < n^{-\alpha} \quad \text{for all } n \geq n_\alpha,$$

where $G \in \mathbb{R}^{n \times n}$ is a Gaussian matrix and $\Pi(\cdot)$ is the partial pivoting map.

Theorem 1.4 implies Trefethen's Conjecture 1.1, as $e^{\sqrt{\log n}} = o(n^c)$ for all $c > 0$. The $\frac{1}{2}$ exponent in the theorem is correct, as $\log \|G\|_{\max} = o(\log n)$ and, by [13], $\log \|U\|_{\max} \geq \log(|\det(G)|^{\frac{1}{n}}) \geq \frac{1}{2} \log n - O(1)$ with probability $1 - o(1)$. This, combined with Theorem 1.4, implies the following.

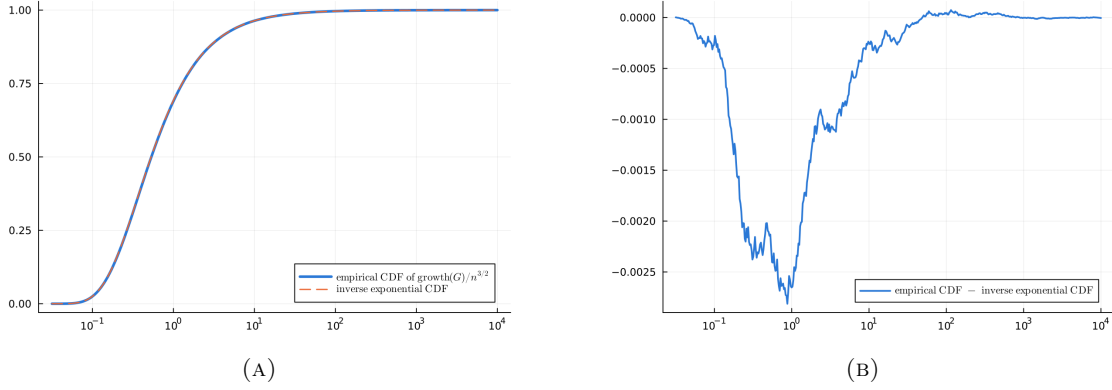


FIGURE 1. We compare the empirical distribution of one million samples of $n^{-\frac{3}{2}}\text{growth}(G)$, the scaled growth factor of a Gaussian matrix without pivoting, for $n = 64$, to the distribution of an inverse exponential with rate $2/(3\sqrt{\pi})$. In Figure (A), we compare their distributions directly, which are indistinguishable to the human eye. In Figure (B), we plot the difference. The KS distance is 2.8×10^{-3} , illustrating the rapid practical convergence that occurs even for small n . Unlike the scaled growth factor of a Gaussian matrix under partial pivoting, $n^{-\frac{1}{2}}\text{growth}(\Pi(G))$, which we prove is $n^{o(1)}$ with overwhelming probability, $n^{-\frac{3}{2}}\text{growth}(G)$ is $n^{\Theta(1)}$ with probability $n^{-O(1)}$. The empirical difference in tail distribution between no pivoting and partial pivoting is made readily apparent by comparing the above Figure (A) to Figure 2

Corollary 1.5. *Let $G \in \mathbb{R}^{n \times n}$ be a Gaussian matrix and $\Pi(\cdot)$ be the partial pivoting map. Then*

$$\frac{\log \text{growth}(\Pi(G))}{\log n} \xrightarrow{p} \frac{1}{2}.$$

Numerical experiments suggest some further refinements. While the evidence is not overwhelming, empirically, the fluctuations of $n^{-\frac{1}{2}}\text{growth}(G)$, properly shifted and normalized, seem to resemble Gumbel distributions. This would not be too surprising, as these distributions often appear when taking the maximum of some number of random variables. Somewhat surprisingly, as n increases, the empirical density of $n^{-\frac{1}{2}}\text{growth}(G)$ appears to narrow, leaving open the possibility that $n^{-\frac{1}{2}}\text{growth}(G)$ converges to a constant. See Figure 2 for details.

The proof of Theorem 1.4, suitably adjusted, also provides a smoothed analysis of partial pivoting.

Theorem 1.6. *For every $\alpha > 0$, there exist constants C_α, n_α such that, for all $n \geq n_\alpha$, every matrix $A \in \mathbb{R}^{n \times n}$ with $\|A\|_2 \leq 1$ and every $\epsilon \in (0, 1)$ satisfy*

$$\mathbb{P} \left[\text{growth}(\Pi(A + \epsilon G)) > n^{C_\alpha} \epsilon^{-\frac{2}{\log 2} \log \log n} \right] < n^{-\alpha},$$

where $G \in \mathbb{R}^{n \times n}$ is a Gaussian matrix and $\Pi(\cdot)$ is the partial pivoting map.

This shows that polynomially small perturbations produce $n^{O(\log \log n)}$ growth. While the correct behavior is likely $n^{O(1)}$, this is a significant improvement over the longstanding $n^{O(\log n)}$ bound.

1.4. Remainder of Paper. In Section 2, we prove that a Haar orthogonally preconditioned matrix rarely has growth much larger than $n^{\frac{3}{2}}$ (Theorem 1.2). In Section 3, we prove that the distribution of the growth of a Gaussian matrix without pivoting, divided by $n^{\frac{3}{2}}$, converges to an inverse exponential distribution (Theorem 1.3). In Section 4, we prove six technical lemmas regarding Gaussian matrices

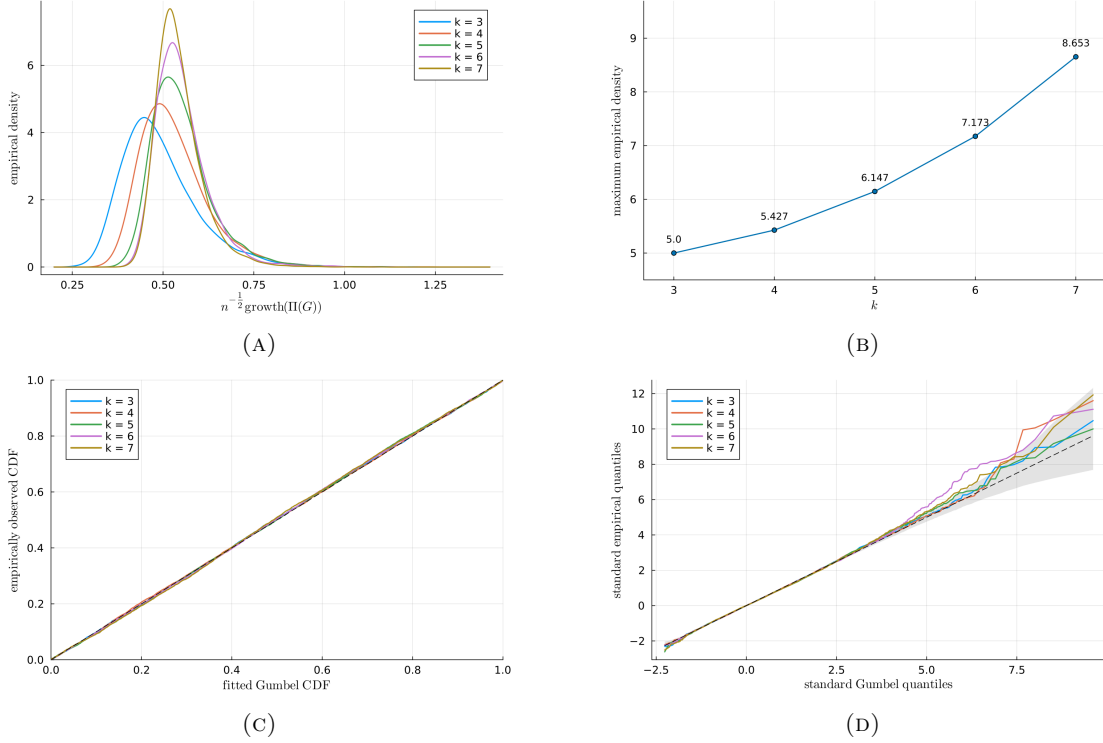


FIGURE 2. We compute 7500 samples of $n^{-\frac{1}{2}} \text{growth}(\Pi(G))$ for $n = 4^k$, $k = 3, \dots, 7$. In Figure (A), we plot their empirical densities. Note that the tails decay rapidly and the densities appear to narrow somewhat. In Figure (B), for each value of k , we plot the maximum empirical density in a region of length $1/100$, further illustrating this narrowing. The empirical densities look somewhat similar to Gumbel distributions. We fit each value of k individually, and, after proper shifting and scaling, compare them to their fitted counterparts. In Figures (C) and (D), we provide a P-P plot and Q-Q plot, respectively. The grey region in Figure (D) notes the area that the empirical quantiles of 7500 Gumbel samples would lie in, with 95% likelihood.

and singular values, which we use in Sections 5 and 6. In Section 5, we prove that the growth factor of a Gaussian matrix under partial pivoting is rarely much more than $n^{\frac{1}{2}}$ (Theorem 1.4). In Section 6, we prove that any matrix, of norm at most one, plus an ϵ sized Gaussian perturbation rarely has growth much larger than $n^{O(1)} \epsilon^{-O(\log \log n)}$ (Theorem 1.6). The proof of Theorem 1.6 is a consequence of a more general version of Theorem 1.4 (Theorem 6.1). However, for clarity, we present the slightly simpler proof of the Gaussian case first in Section 5. Then, in Section 6, we present a generalization of the same argument, with far greater brevity.

2. GROWTH OF ORTHOGONALLY PRECONDITIONED MATRICES WITHOUT PIVOTING

Here we prove that the growth factor of QA , where $A \in \text{GL}_n(\mathbb{R})$ is fixed and $Q \sim \text{Haar}(\text{O}(n))$, is usually not much bigger than $n^{\frac{3}{2}}$ (Theorem 1.2). Our main technical tool is the following lemma regarding the inverse of a submatrix of a Haar orthogonal matrix.

Lemma 2.1. *Let $Q \sim \text{Haar}(\text{O}(n))$, $S, T \subset [n]$ be fixed subsets of size m , and $\mathbf{x} \in \mathbb{S}^{m-1}(\mathbb{R})$ and $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k \in \mathbb{S}^{m-1}(\mathbb{R})$ be fixed for some $k \leq m$. Then*

$$\mathbb{P} \left[\max_{j \in [k]} |\mathbf{y}_j^T (Q_{S,T})^{-1} \mathbf{x}| > t \right] \leq \frac{6}{t} \sqrt{\frac{(n-m+1) \log(ek)}{m}} \quad \text{for all } t > 0.$$

Proof. Without loss of generality, we may take $B = Q_{[m],[m]}$ and $\mathbf{x} = \mathbf{e}_1$. First we identify the distribution of $(Q_{[m],[m]})^{-1} \mathbf{e}_1$. Note that $\mathbf{u} := Q_{:, [m]} (Q_{[m],[m]})^{-1} \mathbf{e}_1$ satisfies $\mathbf{u}_1 = 1$ and is in the intersection of the column space of $Q_{:, [m]}$ and $\text{span}\{\mathbf{e}_1, \mathbf{e}_{m+1}, \dots, \mathbf{e}_n\}$, which is almost surely one-dimensional. A unit vector in this intersection, after multiplication by a random ± 1 sign, is distributed uniformly on $\mathbb{S}^{n-m}(\mathbb{R})$, the unit sphere in this $(n-m+1)$ dimensional subspace, and so

$$\|(Q_{[m],[m]})^{-1} \mathbf{e}_1\|_2 = \|\mathbf{u}\|_2 \stackrel{d}{=} |\boldsymbol{\eta}_1|^{-1},$$

where $\boldsymbol{\eta} \sim \text{Unif}(\mathbb{S}^{n-m}(\mathbb{R}))$. By the orthogonal invariance of $Q_{[m],[m]}$, the direction of $(Q_{[m],[m]})^{-1} \mathbf{e}_1$ is uniform and independent of its length. Therefore, $(Q_{[m],[m]})^{-1} \mathbf{e}_1 \stackrel{d}{=} |\boldsymbol{\eta}_1|^{-1} \boldsymbol{\xi}$, where $\boldsymbol{\xi} \sim \text{Unif}(\mathbb{S}^{m-1}(\mathbb{R}))$ is independent of $\boldsymbol{\eta}$. By [2, Lemma 2.2],

$$\mathbb{P}[|\boldsymbol{\eta}_1| < a] \leq 2\sqrt{n-m+1} a \quad \text{and} \quad \mathbb{E} \left[\max_{j \in [k]} |\mathbf{y}_j^T \boldsymbol{\xi}| \right] \leq 3\sqrt{\frac{\log(ek)}{m}},$$

where the second estimate follows from a union bound and integration of their upper tail estimate. By independence,

$$\begin{aligned} \mathbb{P} \left[\max_{j \in [k]} |\mathbf{y}_j^T (Q_{[m],[m]})^{-1} \mathbf{x}| > t \right] &= \mathbb{E} \left[\mathbb{P} \left[|\boldsymbol{\eta}_1| < \frac{1}{t} \max_{j \in [k]} |\mathbf{y}_j^T \boldsymbol{\xi}| \mid \max_{j \in [k]} |\mathbf{y}_j^T \boldsymbol{\xi}| \right] \right] \\ &\leq \frac{2\sqrt{n-m+1}}{t} \mathbb{E} \left[\max_{j \in [k]} |\mathbf{y}_j^T \boldsymbol{\xi}| \right] \\ &\leq \frac{6}{t} \sqrt{\frac{(n-m+1) \log(ek)}{m}}. \end{aligned}$$

□

Repeated application of Lemma 2.1 is sufficient to prove our desired Theorem 1.2.

Proof of Theorem 1.2. Without loss of generality, by $Q \sim \text{Haar}(\text{O}(n))$ and the QR factorization, it suffices to consider upper triangular A . Then, by the block inversion formula,

$$(QA)/(QA)_{[k],[k]} = (Q/Q_{[k],[k]})A_{[k]^c,[k]^c} = (Q_{[k]^c,[k]^c})^{-T}A_{[k]^c,[k]^c},$$

where $[k]^c = [n] \setminus [k]$. Let $\alpha = \max_j \|A\mathbf{e}_j\|_2 \leq \sqrt{n}\|QA\|_{\max}$. Then every column of $\alpha^{-1}A$ has two-norm at most one. Applying Lemma 2.1 to each row of $\alpha^{-1}U$ and summing the failure probabilities, we obtain

$$\mathbb{P} \left[\frac{\|U\|_{\max}}{\|QA\|_{\max}} > t \right] \leq \frac{6\sqrt{n}}{t} \sum_{m=1}^n \sqrt{\frac{(n-m+1) \log(em)}{m}} \leq \frac{6n\sqrt{\log(en)}}{t} \sum_{m=1}^n \frac{1}{\sqrt{m}} \leq \frac{12n^{3/2}\sqrt{\log(en)}}{t}.$$

Next, we consider L . Note that, because A is upper triangular, QA and Q have the same lower triangular factor [19, proof of Lemma 5.1]. The subdiagonal entries of the i^{th} column of L have law $(g_{i+1}, \dots, g_n)/g_i$ for $g_i, \dots, g_n$ i.i.d. $N(0, 1)$, and therefore standard Gaussian estimates give, for $t \geq 1$,

$$\mathbb{P}[\|L\|_{\max} > t] \leq \sum_{i=1}^{n-1} \mathbb{P} \left[\max_{j>i} |L_{ji}| > t \right] \leq \frac{n-1}{t} \sqrt{\frac{2}{\pi}} \mathbb{E} \left[\max_{j \in [n]} |g_j| \right] \leq \frac{2(n-1)\sqrt{\log(en)}}{\sqrt{\pi}t}. \quad (2.1)$$

Combining the two probabilistic bounds for L and U completes the proof. □

3. GROWTH OF GAUSSIAN MATRICES WITHOUT PIVOTING

Here we prove that the asymptotic distribution of the growth factor of a Gaussian matrix without pivoting, divided by $n^{\frac{3}{2}}$, is inverse exponential. We require two lemmas. The first concerns the distribution of the largest entry in a row of U , which is related to beta random variables. The second concerns the maximum of the inverse square roots of independent beta random variables.

Lemma 3.1. *Let G be an $n \times n$ Gaussian matrix with LU factorization $G = LU$, and*

$$X_k = \|(-G_{k+1,[k]}G_{[k],[k]}^{-1}, 1)\|_2 \quad \text{and} \quad Y_k = \max_{j \geq k+1} \left| \frac{(-G_{k+1,[k]}G_{[k],[k]}^{-1}, 1)G_{[k+1],j}}{X_k} \right|$$

for $k \in [n-1]$. Then

$$\frac{1}{X_k^2} \sim \text{Beta}\left(\frac{1}{2}, \frac{k}{2}\right), \quad Y_k \stackrel{d}{=} \max_{j \in [n-k]} |g_j|, \quad \text{and} \quad \max_{j \geq k+1} |U_{k+1,j}| = X_k Y_k$$

for all $k \in [n-1]$, where $g_1, \dots, g_{n-1}$ are i.i.d. $N(0, 1)$. Furthermore, $X_1, \dots, X_{n-1}$ are mutually independent, and X_k is independent of Y_k for $k \in [n-1]$.

Proof. Let $\mathbf{x}_k^T = (-G_{k+1,[k]}G_{[k],[k]}^{-1}, 1)$ for $k \in [n-1]$. We have

$$U_{k+1,j} = (G/G_{[k],[k]})_{1,j-k} = \mathbf{x}_k^T G_{[k+1],j} \quad \text{for } k+1 \leq j,$$

from which $\max_{j \geq k+1} |U_{k+1,j}| = X_k Y_k$ for $k \in [n-1]$ follows immediately. Conditional on the first k columns, $\mathbf{x}_k^T/X_k$ is fixed, the remaining columns are Gaussian, and so their inner products with $\mathbf{x}_k^T/X_k$ are independent $N(0, 1)$. This gives the distribution of Y_k and its independence from X_k .

Next, consider X_k . The vector $\mathbf{x}_k/X_k$ spans $\text{null}(G_{[k+1],[k]}^T)$, and is therefore, after a random signing, uniformly distributed on the unit sphere $\mathbb{S}^k(\mathbb{R})$. X_k^{-2} is the squared magnitude of its last entry, and so X_k^{-2} is Beta distributed with parameters $\alpha = \frac{1}{2}$ and $\beta = \frac{k}{2}$. Finally, the independence of X_k to $X_{k+1}, \dots, X_{n-1}$ follows from considering the application of a Haar orthogonal matrix to the first $k+1$ rows of G . This operation preserves the Gaussian distribution of the matrix, fully randomizes X_k , but leaves the values of $X_{k+1}, \dots, X_{n-1}$ unchanged. $\square$

Lemma 3.2. *Let $X_1, \dots, X_{n-1}$ be independent positive random variables satisfying $X_k^{-2} \sim \text{Beta}(\frac{1}{2}, \frac{k}{2})$ for $k \in [n-1]$,*

$$c_k = \frac{2\Gamma(\frac{k+1}{2})}{\sqrt{\pi}\Gamma(\frac{k}{2})} \quad \text{for } k \in [n-1], \quad \text{and} \quad \alpha_n = \sum_{k=1}^{n-1} c_k.$$

Then $\alpha_n = \frac{2\sqrt{2}}{3\sqrt{\pi}} n^{\frac{3}{2}} + O(n^{\frac{1}{2}})$ and, for fixed $x > 0$,

$$\mathbb{P} \left[\max_{k \in [n-1]} X_k \leq \alpha_n x \right] \rightarrow e^{-\frac{1}{x}}.$$

Proof. Using the estimate $\Gamma(x+c)/\Gamma(x) = x^c + O_c(x^{c-1})$, we have

$$\alpha_n = \frac{2}{\sqrt{\pi}} \sum_{k=1}^{n-1} \frac{\Gamma(\frac{k+1}{2})}{\Gamma(\frac{k}{2})} = \sqrt{\frac{2}{\pi}} \sum_{k=1}^{n-1} k^{\frac{1}{2}} + O(n^{\frac{1}{2}}) = \frac{2\sqrt{2}}{3\sqrt{\pi}} n^{\frac{3}{2}} + O(n^{\frac{1}{2}}).$$

For n sufficiently large, $1/(\alpha_n x) < 1/2$ and $|(1-u^2)^{\frac{k}{2}-1} - 1| \leq Cku^2$ for all $k \in [n-1]$ and $0 \leq u \leq 1/(\alpha_n x)$, for some constant C . Therefore,

$$p_k := \mathbb{P}[X_k \geq \alpha_n x] = c_k \int_0^{\frac{1}{\alpha_n x}} (1-u^2)^{\frac{k}{2}-1} du = \frac{c_k}{\alpha_n x} (1 + O_x(n^{-2})).$$

Independence and the estimate $\log(1-p) = -p + O(p^2)$ imply our desired result

$$\log \mathbb{P} \left[\max_{k \in [n-1]} X_k \leq \alpha_n x \right] = - \sum_{k=1}^{n-1} p_k + O \left(\sum_{k=1}^{n-1} p_k^2 \right) = -\frac{1}{x} + O_x(n^{-1}).$$

□

Using Lemmas 3.1 and 3.2, we are now prepared to prove Theorem 1.3.

Proof of Theorem 1.3. Lemma 3.2 gives control of X_k . In order to estimate $\|U\|_{\max}$, we require concentration of Y_k . Let $g_1, g_2, \dots$ be i.i.d. $N(0, 1)$. We have

$$\mathbb{E} \left[\frac{\max_{j \in [m]} |g_j|}{\sqrt{2 \log m}} \right] \rightarrow 1 \quad \text{and} \quad \mathbb{E} \left[\left| \frac{\max_{j \in [m]} |g_j|}{\sqrt{2 \log m}} - \mathbb{E} \left[\frac{\max_{j \in [m]} |g_j|}{\sqrt{2 \log m}} \right] \right| \right] \leq \sqrt{2\pi},$$

by [26, Exercise 2.11c] and [7, Proposition 10.3], respectively. The quantity $\mathbb{E} \max_{j \in [m]} |g_j|$ is monotonic in m , and so, for every fixed $\epsilon \in (0, 1)$,

$$\sup_{\epsilon n \leq m \leq n} \mathbb{E} \left[\left| \frac{\max_{j \in [m]} |g_j|}{\sqrt{2 \log n}} - 1 \right| \right] \rightarrow 0 \quad \text{and} \quad \sup_{m \in [n]} \mathbb{E} \left[\left| \frac{\max_{j \in [m]} |g_j|}{\sqrt{2 \log n}} - 1 \right| \right] \leq C.$$

We have $X_k^{-2} \sim \text{Beta}(\frac{1}{2}, \frac{k}{2})$, and so $\mathbb{P}[X_k > t] \leq \frac{\pi c_k}{2t}$ for all $t > 0$. Then, by conditioning on each Y_k and using the independence of X_k and Y_k ,

$$\mathbb{P} \left[\max_{k \in [n-1]} X_k \left| \frac{Y_k}{\sqrt{2 \log n}} - 1 \right| > \delta \alpha_n \right] \leq \frac{\pi}{2\delta \alpha_n} \sum_{k=1}^{n-1} c_k \mathbb{E} \left[\left| \frac{Y_k}{\sqrt{2 \log n}} - 1 \right| \right] \rightarrow 0$$

for all $\delta > 0$. Note that the first row of U is the first row of G , and can be safely ignored when considering the scale $\alpha_n \sqrt{2 \log n}$. Therefore, by Lemma 3.2,

$$\mathbb{P} \left[\frac{\|U\|_{\max}}{\alpha_n \sqrt{2 \log n}} \leq x \right] \rightarrow e^{-\frac{1}{x}}.$$

Now, consider L . The subdiagonal entries of the i^{th} column of L have the same law as in the proof of Theorem 1.2, and so (2.1) holds here as well and $n^{-3/2} \|L\|_{\max} \rightarrow 0$ in probability.

Finally, we have $\|G\|_{\max}/(2\sqrt{\log n}) \rightarrow 1$ in probability, and so, by Slutsky's theorem,

$$\frac{\|U\|_{\max}}{n^{3/2} \|G\|_{\max}} = \frac{\alpha_n}{\sqrt{2} n^{3/2}} \frac{\|U\|_{\max}}{\alpha_n \sqrt{2 \log n}} \frac{2\sqrt{\log n}}{\|G\|_{\max}} \xrightarrow{d} X.$$

□

4. SINGULAR VALUES AND GAUSSIAN MATRICES

Here, we recall and prove six technical results regarding singular values and Gaussian matrices required for our analysis of partial pivoting in Sections 5 and 6. For Section 5, the analysis of a Gaussian matrix under partial pivoting, we apply the below lemmas with $\Delta = 0$. We denote the singular values of a matrix $M \in \mathbb{R}^{m \times n}$ by $\sigma_1(M) \geq \dots \geq \sigma_{\min(m,n)}(M)$, and, for a square $M \in \mathbb{R}^{n \times n}$, let

$$\Sigma_k(M) = \sum_{i=1}^{n-k} \frac{1}{\sigma_i(M)^2}$$

denote the squared Frobenius norm of the inverse of a matrix after truncating the k largest singular values of the inverse. By convention, we set $\Sigma_k(M) = 0$ when $k \geq n$. Let $M^\dagger$ denote the Moore-Penrose pseudoinverse of a matrix $M \in \mathbb{R}^{m \times n}$.

Lemma 4.1. *Let $G \in \mathbb{R}^{m \times n}$ be Gaussian and $M \in \mathbb{R}^{n \times p}$, $\Delta \in \mathbb{R}^{m \times n}$ be fixed, with $\|\Delta\|_2 \leq K$. Then*

$$\mathbb{P} [\|(G + \Delta)M\|_F^2 > 2(\sqrt{m} + K)^2 \|M\|_F^2 + 4x \|M\|_2^2] \leq e^{-x} \quad \text{and} \quad (4.1)$$

$$\mathbb{P} [\sigma_k((G + \Delta)M) > \|M\|_F + (\sqrt{m} + K + \sqrt{2x/k}) \|M\|_2] \leq e^{-x} \quad (4.2)$$

for all $x > 0$ and $k \leq \min(m, p)$.

Proof. Let $f_k(A) := \sqrt{\sum_{i=1}^k \sigma_i(A)^2}$. Both $G \mapsto \|(G + \Delta)M\|_F$ and $G \mapsto f_k((G + \Delta)M)$ are $\|M\|_2$ -Lipschitz, and so, by [7, Proposition 10.1] and the triangle inequality,

$$\mathbb{E} \|(G + \Delta)M\|_F \leq (\sqrt{m} + K) \|M\|_F \quad \text{and} \quad \mathbb{E} f_k((G + \Delta)M) \leq \sqrt{k} (\|M\|_F + (\sqrt{m} + K) \|M\|_2).$$

The desired result follows from Gaussian concentration [7, Proposition 10.3] with deviation $\sqrt{2x} \|M\|_2$ and the bounds $(a + b)^2 \leq 2(a^2 + b^2)$ and $\sigma_k(A) \leq f_k(A)/\sqrt{k}$. $\square$

Lemma 4.2. *Let $G \in \mathbb{R}^{m \times n}$ be Gaussian and $\Delta \in \mathbb{R}^{m \times n}$ be fixed, with $m \leq r$ and $n \geq 3r$. Then*

$$\mathbb{P} \left[\|(G + \Delta)^\dagger\|_2^2 > \frac{2e^{\frac{2x}{r} + 2}}{r} \right] \leq e^{-x}$$

for all $x \geq 1$.

Proof. Let $Q \in \mathbb{R}^{n \times 2r}$ have orthonormal columns in $\text{null}(\Delta)$. Then $(G + \Delta)Q \in \mathbb{R}^{m \times 2r}$ is Gaussian and $(G + \Delta)(G + \Delta)^T \succeq (G + \Delta)QQ^T(G + \Delta)^T$, so $\|(G + \Delta)^\dagger\|_2 \leq \|((G + \Delta)Q)^\dagger\|_2 = \|(GQ)^\dagger\|_2$. Next, add $r - m$ Gaussian rows to GQ , which can only decrease the smallest singular value. When $r > 1$, apply [7, Proposition A.3] with $t = e\sqrt{2/r}e^{x/r}$ to the resulting $r \times 2r$ matrix. When $r = 1$, this is a standard result about χ_2^2 . $\square$

Lemma 4.3. *Let $G \in \mathbb{R}^{n \times n}$ be Gaussian and $\Delta \in \mathbb{R}^{n \times n}$ be fixed. Then*

$$\mathbb{P} \left[\sigma_{n-j}(G + \Delta) \leq \frac{cj\theta}{\sqrt{n}} \right] \leq n^{j+1} \theta^{\frac{1}{4}j^2}$$

for all $4 \leq j < n$ and $\theta \in (0, 1]$, for some absolute constant $c \in (0, 1]$.

Proof. This follows immediately from [12, (7) of Theorem 1.7] by taking $\gamma = 1/2$, $k = j + 1$, $\epsilon = cj\theta/(j + 1)$, using $\sup_a \mathbb{P}[|X - a| \leq \epsilon] \leq \sqrt{2/\pi} \epsilon$ for $X \sim N(0, 1)$, and choosing c sufficiently small. $\square$

Lemma 4.4. *Let $X \in \mathbb{R}^n$ be distributed as $N(\mathbf{v}, I_n)$ conditioned on lying in a convex set $\mathcal{K} \subset \mathbb{R}^n$ of positive Gaussian measure with $\mathcal{K} = -\mathcal{K}$. Then*

$$\|\mathbb{E}X\|_2 \leq \|\mathbf{v}\|_2 \quad \text{and} \quad \mathbb{E} \left[e^{\mathbf{u}^T(X - \mathbb{E}X)} \right] \leq e^{\frac{1}{2}\|\mathbf{u}\|_2^2} \quad \text{for all } \mathbf{u} \in \mathbb{R}^n. \quad (4.3)$$

For fixed $M \in \mathbb{R}^{n \times p}$ and $x > 0$,

$$\mathbb{P} \left[\|X^T M\|_2 > \sqrt{2\|M\|_F^2 + 4x\|M\|_2^2} + \|\mathbf{v}\|_2 \|M\|_2 \right] \leq e^{-x}. \quad (4.4)$$

Proof. Let $\varphi_n(\mathbf{x}) = (2\pi)^{-n/2} e^{-\|\mathbf{x}\|_2^2/2}$ be the density of a standard Gaussian vector and

$$f(\mathbf{y}) = \log \int_{\mathcal{K}} e^{\mathbf{y}^T \mathbf{x}} \varphi_n(\mathbf{x}) d\mathbf{x}.$$

Then $X_{\mathbf{y}}$, distributed as $N(\mathbf{y}, I_n)$ conditioned on lying in $\mathcal{K}$, has density $\mathbf{1}_{\mathcal{K}}(\mathbf{x}) e^{\mathbf{y}^T \mathbf{x} - f(\mathbf{y})} \varphi_n(\mathbf{x})$. In addition, $\nabla f(\mathbf{y}) = \mathbb{E}X_{\mathbf{y}}$ and $\nabla^2 f(\mathbf{y}) = \text{Cov}(X_{\mathbf{y}})$. Therefore $\nabla^2 f(\mathbf{y}) \succeq 0$. Next, we aim to upper bound this quantity. By [16, Theorem 6],

$$\log \int_{\mathbb{R}^n} \mathbf{1}_{\mathcal{K}}(\mathbf{x}) \varphi_n(\mathbf{x} - \mathbf{y}) d\mathbf{x} = f(\mathbf{y}) - \frac{1}{2} \|\mathbf{y}\|_2^2$$

is concave, and so $\nabla^2 f(\mathbf{y}) \preceq I_n$. By the symmetry of $\mathcal{K}$, $\nabla f(0) = 0$. We have

$$\mathbb{E}X_{\mathbf{y}} = \nabla f(\mathbf{y}) - \nabla f(0) = \int_0^1 \nabla^2 f(t\mathbf{y})\mathbf{y} dt,$$

and, by $0 \preceq \nabla^2 f(t\mathbf{y}) \preceq I_n$,

$$\|\mathbb{E}X_{\mathbf{y}}\|_2 \leq \int_0^1 \|\nabla^2 f(t\mathbf{y})\mathbf{y}\|_2 dt \leq \|\mathbf{y}\|_2.$$

Similarly,

$$\begin{aligned} \log \mathbb{E} \left[e^{\mathbf{u}^T (X_{\mathbf{y}} - \mathbb{E}X_{\mathbf{y}})} \right] &= f(\mathbf{y} + \mathbf{u}) - f(\mathbf{y}) - \mathbf{u}^T \mathbb{E}X_{\mathbf{y}} \\ &= \int_0^1 (1-t) \mathbf{u}^T \nabla^2 f(\mathbf{y} + t\mathbf{u})\mathbf{u} dt \\ &\leq \int_0^1 (1-t) \|\mathbf{u}\|_2^2 dt \\ &= \frac{1}{2} \|\mathbf{u}\|_2^2. \end{aligned}$$

By [9, Theorem 2.1] applied to $X_{\mathbf{y}} - \mathbb{E}X_{\mathbf{y}}$ with $A = (MM^T)^{1/2}$ and $\text{trace}((MM^T)^2) \leq \|M\|_2^2 \|M\|_F^2$,

$$\|(X_{\mathbf{y}} - \mathbb{E}X_{\mathbf{y}})^T M\|_2^2 \leq 2\|M\|_F^2 + 4x\|M\|_2^2$$

with probability at least $1 - e^{-x}$, which implies our final desired bound, by the triangle inequality. $\square$

Lemma 4.5. *Let $k < n$ and the fixed matrix $M \in \mathbb{R}^{n \times n}$ satisfy $\sigma_{n-k}(M) \geq \mu$ and $\Sigma_k(M) \leq 2k\mu^{-2}$ for some $0 < \mu \leq \sqrt{k}$. Let $G \in \mathbb{R}^{n \times 4k}$ be Gaussian, $\Delta \in \mathbb{R}^{n \times 4k}$ be fixed with $\|\Delta\|_2 \leq K$, and $B = [M \ (G + \Delta)]$. Then, for any $x \geq 1$,*

$$\|B^\dagger\|_2^2 \leq C\mu^{-2}(1 + K^2/k)e^{\frac{Cx}{k}} \quad \text{and} \quad (4.5)$$

$$\|B^\dagger\|_F^2 \leq Ck\mu^{-2}(1 + K^2/k)e^{\frac{Cx}{k}} \quad (4.6)$$

with probability at least $1 - 3e^{-x}$, for some absolute constant C .

Proof. Without loss of generality, let M be diagonal with $M_{ii} = \sigma_{n-i+1}(M)$. Let $\hat{M} = M_{[n] \setminus [k], [n] \setminus [k]}$, partition $G + \Delta = \begin{bmatrix} X \\ Y \end{bmatrix}$, where $X \in \mathbb{R}^{k \times 4k}$ and $Y \in \mathbb{R}^{(n-k) \times 4k}$, and let $Q = X^T (XX^T)^{-1/2} \in \mathbb{R}^{4k \times k}$.

Now consider the matrix

$$\hat{B} = \begin{bmatrix} XQ & 0 \\ YQ & \hat{M} \end{bmatrix} \quad \text{with} \quad \hat{B}^{-1} = \begin{bmatrix} (XQ)^{-1} & 0 \\ -\hat{M}^{-1}YQ(XQ)^{-1} & \hat{M}^{-1} \end{bmatrix}.$$

We have

$$BB^T - \hat{B}\hat{B}^T = \begin{bmatrix} M_{[k], [k]} M_{[k], [k]}^T & 0 \\ 0 & 0 \end{bmatrix} + (G + \Delta)(I - QQ^T)(G + \Delta)^T \succeq 0,$$

and so we may instead work with $\hat{B}$ and bound the blocks of $\hat{B}^{-1}$. By Lemma 4.1 applied to $(YQ)^T \hat{M}^{-1}$, conditional on X , and Lemma 4.2 applied to X ,

$$\|\hat{M}^{-1}YQ\|_F^2 \leq 2(\sqrt{k} + K)^2 \|\hat{M}^{-1}\|_F^2 + 4x\|\hat{M}^{-1}\|_2^2 \leq C\mu^{-2}(k^2 + kK^2 + x),$$

$$\|\hat{M}^{-1}YQ\|_2^2 \leq \left(\|\hat{M}^{-1}\|_F + (\sqrt{k} + K + \sqrt{2x})\|\hat{M}^{-1}\|_2 \right)^2 \leq C\mu^{-2}(k + K^2 + x), \quad \text{and}$$

$$\|(XQ)^{-1}\|_2^2 = \|X^\dagger\|_2^2 \leq \frac{2e^2}{k} e^{2x/k}$$

with probability at least $1 - 3e^{-x}$. Combining this with the bounds

$$\begin{aligned}\|\hat{B}^{-1}\|_F^2 &\leq (k + \|\hat{M}^{-1}YQ\|_F^2)\|(XQ)^{-1}\|_2^2 + 2k\mu^{-2} \quad \text{and} \\ \|\hat{B}^{-1}\|_2^2 &\leq 3[(1 + \|\hat{M}^{-1}YQ\|_2^2)\|(XQ)^{-1}\|_2^2 + \mu^{-2}],\end{aligned}$$

and the inequality $1 + x/k + K^2/k \leq (1 + K^2/k)e^{x/k}$ completes the proof. $\square$

Lemma 4.6. *Let $M \in \mathbb{R}^{n \times (n+k)}$ have rank n , the columns of $Q \in \mathbb{R}^{(n+k) \times k}$ be an orthonormal basis for $\text{null}(M)$, $B \in \mathbb{R}^{k \times (n+k)}$, and $j < k/2$. Then*

$$\sigma_{n+k-2j} \left(\begin{bmatrix} M \\ B \end{bmatrix} \right) \geq \left(\|M^\dagger\|_2 + \frac{1 + \sigma_{j+1}(BM^\dagger)}{\sigma_{k-j}(BQ)} \right)^{-1}. \quad (4.7)$$

Proof. Let $W \in \mathbb{R}^{k \times (k-2j)}$ have orthonormal columns orthogonal to the bottom j left singular directions of BQ and top j left singular directions of $BM^\dagger$, so that

$$\sigma_{k-2j}(W^T BQ) \geq \sigma_{k-j}(BQ) \quad \text{and} \quad \|W^T BM^\dagger\|_2 \leq \sigma_{j+1}(BM^\dagger).$$

We have

$$\sigma_{n+k-2j} \left(\begin{bmatrix} M \\ B \end{bmatrix} \right) \geq \sigma_{n+k-2j} \left(\begin{bmatrix} M \\ W^T B \end{bmatrix} \right) = \sigma_{\min} \left(\begin{bmatrix} M \\ W^T B \end{bmatrix} \right),$$

and so it suffices to consider the system $M\mathbf{x} = \mathbf{a}$, $W^T B\mathbf{x} = \mathbf{b}$ and control the size of the solution $\mathbf{x} = M^\dagger \mathbf{a} + Q(W^T BQ)^\dagger (\mathbf{b} - (W^T BM^\dagger)\mathbf{a})$. We have

$$\begin{aligned}\|\mathbf{x}\|_2 &\leq \|M^\dagger\|_2 \|\mathbf{a}\|_2 + \frac{\|\mathbf{b}\|_2 + \|W^T BM^\dagger\|_2 \|\mathbf{a}\|_2}{\sigma_{k-2j}(W^T BQ)} \\ &\leq \left(\|M^\dagger\|_2 + \frac{1 + \|W^T BM^\dagger\|_2}{\sigma_{k-2j}(W^T BQ)} \right) \sqrt{\|\mathbf{a}\|_2^2 + \|\mathbf{b}\|_2^2} \\ &\leq \left(\|M^\dagger\|_2 + \frac{1 + \sigma_{j+1}(BM^\dagger)}{\sigma_{k-j}(BQ)} \right) \sqrt{\|\mathbf{a}\|_2^2 + \|\mathbf{b}\|_2^2}.\end{aligned}$$

$\square$

5. GROWTH OF GAUSSIAN MATRICES UNDER PARTIAL PIVOTING

Here, we prove that the growth factor of an $n \times n$ Gaussian matrix G under partial pivoting is rarely much more than $\sqrt{n}$. The proof technique is based on control of most of the singular values of the leading submatrices $\Pi(G)_{[k],[k]}$. Our technique, at a high level, is heavily inspired by that of Huang and Tikhomirov [10], with a number of key refinements and differences in our spectral recursion and handling of extreme singular values. To obtain the correct $\sqrt{n}$ behavior, we exploit the only guaranteed benefit of partial pivoting, that is, that $\|L\|_{\max} = 1$. Other than exploiting this key fact, the majority of our analysis is devoted to showing that the greedy nature of partial pivoting does not create a situation which is too much worse than no pivoting at all. While simulations suggest that partial pivoting is in fact better than no pivoting for many of our desired estimates, such as $\Sigma_r(\Pi(G)_{[k],[k]})$, proving this seems significantly more challenging than the growth conjecture itself.

Our argument involves revealing pivot choices or columns of G iteratively and conditioning. Let π_t be the ordered list of the first t pivot indices, $\pi_t^c = [n] \setminus \{\pi_t(1), \dots, \pi_t(t)\}$ ordered in increasing order, $T_t = G_{\pi_t, [t]}$, $\mathcal{H}_t$ record the matrix T_t and the pivot indices π_t , and $\mathcal{F}_t$ record the matrix $G_{[n], [t]}$. We say a quantity is measurable based on some information if it can be determined from it. In the coming analysis, we use $\mathcal{H}_t$ to argue about non-pivot rows and $\mathcal{F}_t$ to reveal new columns. Note that many of the results presented are stated in slightly greater generality than needed here, to aid in the smoothed analysis in the following section. Throughout this section, the notation C_x , for some variable x , denotes a positive constant depending only on x , whose value may increase from one occurrence to the next.

The exception is a constant used to define a threshold hold, which is fixed once chosen, i.e., D_γ below is fixed. For brevity, let $\lambda = \log n$.

5.1. A High-Level Outline of the Proof. Before turning to the proof itself, we present a brief high-level sketch of the argument. The partial pivoting rule (1.2) implies that $\|L\|_{\max} = 1$, and $\|G\|_{\max}$ is at least one with exponentially high probability. Therefore, it suffices to show $\|U\|_{\max}$ is rarely much bigger than $n^{\frac{1}{2}}$. We do so by proving an even stronger statement, that is, that $\max_t \|G/T_t\|_{\max}$, the largest entry of any of the Schur complements computed during elimination, is rarely much bigger than $n^{\frac{1}{2}}$. By (1.1), the entries of U are a subset of the entries of the Schur complements. Let P_t be the permutation matrix that reorders the original row indices as (π_t, π_t^c) , with the pivot indices in pivot order and the remaining indices in increasing order. Let

$$E_t = [-G_{\pi_t^c, [t]} T_t^{-1} \ I] P_t \in \mathbb{R}^{(n-t) \times n}$$

be the elimination matrix after t steps of Gaussian elimination, and consider an arbitrary entry of G/T_t , corresponding to row $\mathbf{e}^T$ of E_t and column index $j > t$ of G . Then, conditional on $\mathcal{F}_t$, this entry is normally distributed

$$\mathbf{e}^T G_{[n], j} \mid \mathcal{F}_t \sim N(0, \|\mathbf{e}\|_2^2),$$

as the j^{th} column of G is independent of $\mathcal{F}_t$. Therefore, controlling the norms of the rows of E_t controls the entries of G/T_t through a union bound over Gaussian variables. Simply put, to prove Conjecture 1.1, it suffices to show that, with overwhelming probability, $\|\mathbf{e}\|_2^2$ is not much bigger than n .

Ideally, one could control the row norms of E_t by controlling the singular values of T_t . Indeed, control of $\|T_t^{-1}\|_F^2$ would be sufficient, as, conditional on $\mathcal{H}_t$, each row of $G_{\pi_t^c, [t]}$ is distributed as a Gaussian conditioned to lie in the convex set

$$\mathcal{K}_t = \left\{ \mathbf{x} \mid \|\mathbf{x} \hat{U}^{-1}\|_\infty \leq 1, \text{ where } T_t = \hat{L} \hat{U} \right\}, \quad (5.1)$$

which, by Lemma 4.4, retains enough ‘‘Gaussianity’’ for our purposes. One might hope that $\|T_t^{-1}\|_F^2 \lesssim n$, which would follow from, say, $\sigma_{t-j}(T_t) \gtrsim (j+1)/\sqrt{n}$ for all j . Unfortunately, $\frac{1}{n} \|G^{-1}\|_F^2$ has a heavy tail and an infinite mean. It is true, however, that, by Lemma 4.3, $\Sigma_r(G_{[t], [t]}) \lesssim n^{1+o(1)}$ for $r = \lceil \sqrt{\lambda} \rceil$ with all but superpolynomially small probability. Unsurprisingly, numerical results suggest that partial pivoting only helps matters, that is, $\Sigma_r(T_t)$ behaves better than $\Sigma_r(G_{[t], [t]})$; see Figure 3. We prove, in Proposition 5.5, that it is not too much worse. The proof of Proposition 5.5 uses only two properties of partial pivoting: that the k^{th} pivot is chosen after seeing only the first k columns of the matrix, and removing some number of rows not selected in the first k steps and re-running Gaussian elimination without them does not change the first k pivots chosen. In particular, the result is equally valid for any adversarial pivoting scheme with these two properties.

Armed with the information that $\Sigma_r(T_t)$ is not too much more than n , we can obtain a bound on the row norms of E_{t+4r} via a rank argument for E_t combined with an argument that $4r$ additional steps of Gaussian elimination smooth the r bad directions of E_t without worsening the situation too much in some other way. In Lemma 5.2, we show that E_t can be written as the sum of two parts, one whose row norms are controlled by $\Sigma_r(T_t)$ and one of rank at most r . In Lemma 5.3, we prove that the randomness introduced by revealing $4r$ additional Gaussian columns smooths out the bad low rank part of E_t and that, by the greedy property of partial pivoting, the elimination matrix for these additional $4r$ steps has row one-norms of size at most 2^{4r} , which, for $r = \lceil \sqrt{\lambda} \rceil$, are $n^{o(1)}$. This bounds the row norms of E_{t+4r} . Using this observation, in Proposition 5.1, we prove that control on $\Sigma_r(T_t)$ for $t \in [n]$ implies control on the growth factor under partial pivoting.

Combining Proposition 5.1 (control on $\Sigma_r(T_t)$, $t \in [n]$, implies control on growth) with Proposition 5.5 (control on $\Sigma_r(T_t)$, $t \in [n]$) completes the proof.

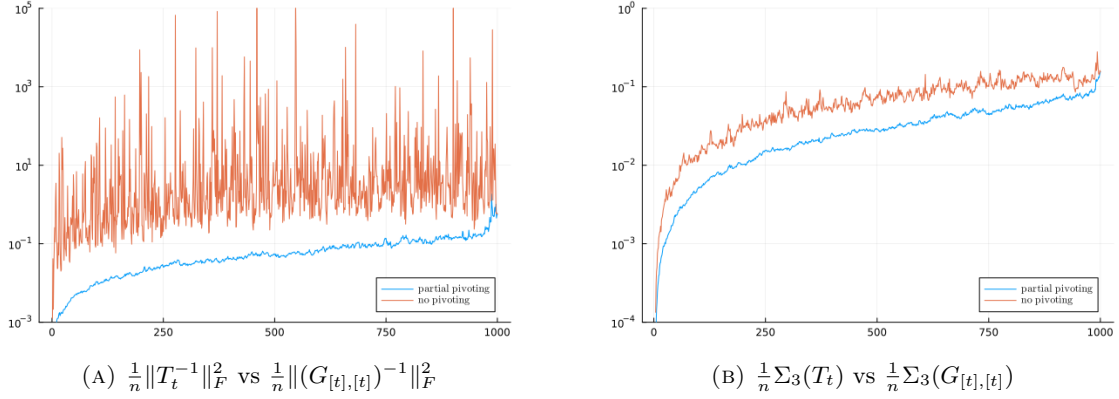


FIGURE 3. We compute the singular values of the inverses of the leading submatrices of a single 1000×1000 Gaussian matrix, both before and after partial pivoting. In Figure (A), the squared Frobenius norms of the inverses without pivoting are poorly controlled. With partial pivoting, the situation is much better, but still this quantity can get quite large near $t = n$. In Figure (B), we consider the Frobenius norms of the truncated inverses, with their $\lceil \sqrt{\lambda} \rceil = 3$ larger singular values removed. These quantities are much better controlled, both with and without pivoting, though partial pivoting still performs better than no pivoting at all for t not too close to n .

5.2. Controlling the Truncated Inverse Bounds Growth. To bound the growth factor, it suffices to control the squared Frobenius norms of the inverses of the leading pivot blocks after removing their r largest singular values.

Proposition 5.1. *Let $p > 0$ and $r < n/8$, and suppose that, for some deterministic $\tau, \eta > 0$,*

$$\mathbb{P}[\Sigma_r(T_t) \leq \tau \text{ for all } t < n] \geq 1 - \eta.$$

Then, for all sufficiently large n , depending only on p ,

$$\mathbb{P}\left[\text{growth}(\Pi(G)) > \sqrt{1 + \tau} e^{C_p(r + \frac{\lambda}{r})}\right] \leq \eta + n^{-p}.$$

In order to prove our desired proposition, we require the following two lemmas. The first decomposes E_t into a controlled part and a part of rank at most r . The second shows that adding $4r$ columns successfully treats the uncontrolled portion, and, for r small enough, does not have enough time to do additional harm during those $4r$ steps.

Lemma 5.2. *For every $t < n$ and $r \geq 1$, there is a $\mathcal{F}_t$ -measurable decomposition*

$$E_t = X_t + Y_t Q_t^T,$$

where $Q_t \in \mathbb{R}^{n \times \min(t, r)}$ satisfies $Q_t^T Q_t = I$. If $t \leq r$, every row of X_t has norm one. If $t > r$ and the $\mathcal{H}_t$ -measurable condition $\Sigma_r(T_t) \leq \tau$ holds, then

$$\mathbb{P}\left[\max_i \|(X_t)_{i, [n]}\|_2^2 > 1 + (2 + 4x)\tau \mid \mathcal{H}_t\right] \leq n e^{-x} \quad \text{for all } x > 0.$$

Proof. When $t \leq r$, we may simply take $X_t = [0 \ I] P_t$ and put the rest in the rank r part. When $t > r$, consider a singular value decomposition

$$T_t^{-1} = Z_t + H V^T,$$

where $V^T V = I_r$ and Z_t contains the $t - r$ smallest inverse singular values. Now, let

$$X_t = [-G_{\pi_t^c, [t]} Z_t I] P_t, \quad Y_t = -G_{\pi_t^c, [t]} H, \quad \text{and} \quad Q_t = P_t^T \begin{bmatrix} V \\ 0 \end{bmatrix}.$$

We have $\|Z_t\|_2^2 \leq \|Z_t\|_F^2 = \Sigma_r(T_t) \leq \tau$, and each row of X_t has squared two-norm $1 + \|G_{i, [t]} Z_t\|_2^2$. Note that, conditional on $\mathcal{H}_t$, the unpivoted rows at step t are distributed as a standard Gaussian conditioned to lie in $\mathcal{K}_t$. Applying Lemma 4.4 with zero mean and taking a union bound provides our desired result. $\square$

Lemma 5.3. *Let $n - t > 4r$, condition on $\mathcal{F}_t$, and let $E_t = X_t + Y_t Q_t^T$ be as in Lemma 5.2. If every row of X_t has two-norm at most ζ , then, for every $q \geq 1$,*

$$\max_i \|(E_{t+4r})_{i, [n]}\|_2 \leq \zeta e^{C_q(r + \frac{\lambda}{r})},$$

except with conditional probability at most Cn^{-q} , where C and C_q do not depend on $\|Y_t\|_2$.

Proof. Let $G_\star = G_{[n], [t+4r] \setminus [t]}$ contain the next $4r$ columns of the matrix, ordered according to E_t , and consider the result of $4r$ further steps of elimination applied to

$$E_t G_\star = X_t G_\star + Y_t (Q_t^T G_\star) \in \mathbb{R}^{(n-t) \times 4r}.$$

Note that, conditional on $\mathcal{F}_t$, $Q_t^T G_\star \in \mathbb{R}^{\min(t, r) \times 4r}$ is Gaussian. Let P permute the rows indexed by π_t^c so that the next $4r$ pivot rows come first, in pivot order, and the remaining rows follow, in the order of π_{t+4r}^c . Let the matrix $J \in \mathbb{R}^{(n-t-4r) \times (n-t)}$ encode the row operations performed during these additional $4r$ elimination steps, with its remaining rows returned to the order of π_{t+4r}^c . In particular,

$$J = [-(PE_t G_\star)_{[n-t] \setminus [4r], [4r]} ((PE_t G_\star)_{[4r], [4r]})^{-1} I_{n-t-4r}] P.$$

Notably, $JE_t = E_{t+4r}$ and $JE_t G_\star = 0$, the result of elimination. Therefore,

$$0 = JE_t G_\star = JX_t G_\star + JY_t (Q_t^T G_\star) = (JX_t G_\star (Q_t^T G_\star)^\dagger + JY_t) (Q_t^T G_\star),$$

and so

$$E_{t+4r} = JX_t + JY_t Q_t^T = J(X_t - (X_t G_\star) (Q_t^T G_\star)^\dagger Q_t^T).$$

Note that we have removed the dependence on Y_t and its size, replacing it by the controlled portion $(X_t G_\star)$ and a Gaussian pseudoinverse $(Q_t^T G_\star)^\dagger$. Because J encodes $4r$ steps of partial pivoting, it is the product of $4r$ matrices, each of whose rows are of the form $\mathbf{e}_i - \ell_i \mathbf{e}_j$ for $|\ell_i| \leq 1$, and so $\|J_{i, [n-t]}\|_1 \leq 2^{4r}$ for all rows. Therefore, by the triangle inequality,

$$\begin{aligned} \|(E_{t+4r})_{i, [n]}\|_2 &\leq \sum_{j=1}^{n-t} |J_{ij}| \|(X_t - X_t G_\star (Q_t^T G_\star)^\dagger Q_t^T)_{j, [n]}\|_2 \\ &\leq 2^{4r} \left(\zeta + \max_i \|(X_t G_\star)_{i, [4r]}\|_2 \|(Q_t^T G_\star)^\dagger\|_2 \right) \end{aligned}$$

for all i . Each row of $X_t G_\star$ has independent Gaussian entries with variance at most ζ^2 . By Lemmas 4.1 and 4.2, with $x = (q+1)\lambda$, we have

$$\max_i \|(X_t G_\star)_{i, [4r]}\|_2^2 \leq C_q \zeta^2 (4r + \lambda) \quad \text{and} \quad \|(Q_t^T G_\star)^\dagger\|_2^2 \leq \frac{C_q}{r} e^{C_q \lambda / r}$$

with conditional probability at least $1 - Cn^{-q}$. Combining these two bounds proves our desired result. $\square$

Using Lemmas 5.2 and 5.3, the proof of our desired proposition follows fairly quickly.

Proof of Proposition 5.1. Let $q = p + 4$ and apply Lemma 5.2 at every starting stage $t < n - 4r$ with $x = q\lambda$. Then

$$\max_i \|(X_t)_{i,[n]}\|_2^2 \leq C_q(1 + \lambda)(1 + \tau)$$

for all such t , with probability at least $1 - \eta - n^{2-q}$, where η accounts for the truncated inverse bound failing, and $n^{1-q} \times n$ accounts for each of the desired bounds failing. Now, apply Lemma 5.3 at each of these starting stages and take a union bound. For $u \leq 4r$, the deterministic bound $\|(E_u)_{i,[n]}\|_2 \leq 2^u$ suffices. We have

$$\max_{u \leq n} \max_i \|(E_u)_{i,[n]}\|_2 \leq \sqrt{1 + \tau} e^{C_q(r + \frac{\lambda}{r})}$$

with probability at least $1 - \eta - Cn^{2-q}$, after absorbing the $\sqrt{1 + \lambda}$ into the exponential. At a fixed stage t , the row bound is $\mathcal{F}_t$ -measurable and each Schur complement entry divided by its elimination row norm is a standard Gaussian. Applying the standard bound $\mathbb{P}[|X| > h] \leq 2e^{-\frac{1}{2}h^2}$ for $X \sim N(0, 1)$ for $h = \sqrt{2q\lambda}$ and union bounding over all the Schur complement entries, which contain all the entries of U , we get that

$$\|U\|_{\max} \leq \sqrt{1 + \tau} e^{C_p(r + \frac{\lambda}{r})}$$

with probability at least $1 - \eta - Cn^{3-q}$. Note that throughout this argument, we take union bounds over stage-wise conditional estimates, and do not use estimates conditioned on success at all stages.

Finally, by partial pivoting, $\|L\|_{\max} = 1$, and $\|G\|_{\max} \geq 1$ with all but an exponentially small probability, completing the proof. $\square$

5.3. Extending a Selected Block. In order to bound $\Sigma_r(T_t)$, we require a lemma that bounds the event that a leading pivot block has a good truncated inverse but its extension has too many small singular values, which we apply inductively in the following subsection.

Lemma 5.4. *Let $m + 4k \leq n$, $k < m$, $d \geq 16$, and $k \geq 100d$. For every $\beta \geq 1$, there is a constant C_β such that, for every $0 < \mu < \sqrt{k}$,*

$$\mathbb{P} \left[\sigma_{m-k}(T_m) \geq \mu, \quad \Sigma_k(T_m) \leq 2k\mu^{-2}, \quad \text{and} \quad \sigma_{m+4k-d}(T_{m+4k}) \leq \frac{\mu d}{k} e^{-C_\beta(1 + \frac{k\lambda}{d^2})} \right] \leq n^{-\beta}.$$

Proof. Let $s = 4k$ and suppose the $\mathcal{F}_m$ -measurable event

$$\sigma_{m-k}(T_m) \geq \mu \quad \text{and} \quad \Sigma_k(T_m) \leq 2k\mu^{-2}$$

holds. Conditional on $\mathcal{F}_m$, the pivots π_m are fixed and $G_{\pi_m, [m+s]}$ is the fixed block T_m and s additional Gaussian columns. By Lemma 4.5 with $x = (\beta + 1)\lambda$,

$$\|(G_{\pi_m, [m+s]})^\dagger\|_2^2 \leq \mu^{-2} e^{C_\beta(1 + \frac{\lambda}{k})} \quad \text{and} \quad \|(G_{\pi_m, [m+s]})^\dagger\|_F^2 \leq k\mu^{-2} e^{C_\beta(1 + \frac{\lambda}{k})} \quad (5.2)$$

with probability at least $1 - 3n^{-(\beta+1)}$. Note that this concerns the extension of the rows specifically chosen by π_m , and so this cost is only paid once.

Now, to treat the next s pivot rows, we fix a set S , $|S| = s$, and consider the result of m steps of elimination with partial pivoting without the rows indexed by S . Let π_m^S be the first m pivots in that situation. Conditional on all the rows outside of S , $G_{\pi_m^S, [m+s]}$ is fixed and $G_{S, [m+s]}$ is Gaussian. In particular, whether (5.2) holds for $G_{\pi_m^S, [m+s]}$ is already determined. We do not condition on S being the next pivot set. Now, on the event that (5.2) does hold for $G_{\pi_m^S, [m+s]}$, let the columns of Q_S be an orthonormal basis for $\text{null}(G_{\pi_m^S, [m+s]})$. Using Lemma 4.1 with $x = (s + \beta + 1)\lambda$ and (5.2),

$$\sigma_{\lfloor d/2 \rfloor + 1}(G_{S, [m+s]}(G_{\pi_m^S, [m+s]})^\dagger) \leq \sqrt{k}\mu^{-1} e^{C_\beta(1 + \frac{\lambda}{d})}$$

with probability at least $1 - n^{-s-\beta-1}$. By Lemma 4.3 with $\theta = e^{-C_\beta(1+\frac{k\lambda}{d^2})}$, $j = \lfloor d/2 \rfloor$ and C_β a large enough constant so that $s^{j+1}\theta^{\frac{1}{4}j^2} \leq n^{-s-\beta-1}$,

$$\sigma_{s-\lfloor d/2 \rfloor}(G_{S,[m+s]}Q_S) \geq \frac{d}{\sqrt{k}} e^{-C_\beta(1+\frac{k\lambda}{d^2})}$$

with probability at least $1 - n^{-s-\beta-1}$.

Because $2\lfloor d/2 \rfloor \leq d$, $\mu \leq \sqrt{k}$, and $k \geq 100d$, by Lemma 4.6,

$$\sigma_{m+s-d} \left(\begin{bmatrix} G_{\pi_m^S, [m+s]} \\ G_{S, [m+s]} \end{bmatrix} \right) \geq \frac{\mu d}{k} e^{-C_\beta(1+\frac{k\lambda}{d^2})} \quad (5.3)$$

for a sufficiently large constant C_β , where we have absorbed the factor $e^{C_\beta \frac{\lambda}{d}}$ in the numerator because $\lambda/d \leq k\lambda/d^2$. There are at most n^s choices for S . Therefore, the probability that some candidate satisfies (5.2) but fails (5.3) is at most $2n^{-\beta-1}$. Applying this to the actual next s pivot rows S , for which $\pi_m^S = \pi_m$, completes the proof. $\square$

5.4. A Bound for the Truncated Inverse. Ideally, $\sigma_{t-k}(T_t) \gtrsim k/\sqrt{n}$ for $k \geq r$, which would imply $\Sigma_r(T_t) \lesssim n \sum_{k=r}^\infty 1/k^2 \lesssim n/r$. The following proposition proves such a bound, with an additional exponential term, due to the additional exponential term in Lemma 5.4.

Proposition 5.5. *For every $\gamma > 0$, there exist constants C_γ, n_γ such that, for $n \geq n_\gamma$,*

$$\max_{k < n} \Sigma_r(T_k) \leq \frac{C_\gamma n}{r} e^{C_\gamma(\frac{\lambda}{r} + \log \log n)}.$$

with probability at least $1 - n^{-\gamma}$, simultaneously for all $r \geq \lceil \sqrt{\lambda} \rceil$.

Proof. For an integer $d \geq \lceil \sqrt{\lambda} \rceil$, consider the recursion $k(d) = \max(100d, \lceil d^2/\lambda \rceil)$, with $d_0 = d$ and $d_{i+1} = k(d_i)$, and let

$$h(d) = \sum_{i|d_i < n} \left(1 + \frac{\lambda}{d_i}\right) \quad \text{and} \quad g(d) = \frac{d}{\sqrt{n}} e^{-D_\gamma h(d)}.$$

We use a factor of 100 so that Lemma 5.4 applies, and use the quadratic term so that the sequence grows quickly once $d > \lambda$. The quantity $h(d)$ tracks the total exponential loss accumulated over each stage from Lemma 5.4, and $g(d)$ provides the iterated thresholds for singular values coming from Lemma 5.4. Note that $h(d)$ is non-increasing and $g(d)/d$ is non-decreasing in d . We aim to prove that

$$\sigma_{t-d}(T_t) \geq g(d) \quad \text{for all } \lceil \sqrt{\lambda} \rceil \leq d < t \leq n, \quad (5.4)$$

which implies our desired result. In particular, if (5.4) holds, then, for every $t > r \geq \lceil \sqrt{\lambda} \rceil$,

$$\Sigma_r(T_t) \leq \sum_{d=r}^{t-1} \frac{1}{g(d)^2} \leq \frac{r^2}{g(r)^2} \sum_{d=r}^\infty \frac{1}{d^2} \leq \frac{2r}{g(r)^2} = \frac{2n}{r} e^{2D_\gamma h(r)}.$$

In addition,

$$h(r) \leq C \left(\frac{\lambda}{r} + \log \log n \right),$$

as $\sum_i \lambda/d_i \leq \lambda/r \sum_{i \geq 0} 100^{-i} \leq 2\lambda/r$, and $h(d)$ involves at most $O(\log \log n)$ steps, $O(\log \log n)$ for d_i to exceed λ and another $O(\log \log n)$ steps for d_i to exceed n .

What remains is to prove (5.4) holds with probability $1 - n^{-\gamma}$. We do so by induction on t . Let $\beta = \gamma + 2$, D_γ be sufficiently large, and n_γ be large enough that $\sqrt{\lambda} \geq 16$.

First, suppose that $k(d) \geq t/8$. There are $O(n^t)$ choices of t rows, and so we may bound $\sigma_{t-d}(T_t)$ by union bounding over $G_{S,[t]}$ for all $|S| = t$. In particular, by Lemma 4.3 applied $O(n^t)$ times with $\theta = c^{-1}e^{-C_\gamma(1+\frac{\lambda}{d})}$, the probability that

$$\sigma_{t-d}(T_t) < g(d) = \frac{d}{\sqrt{n}} e^{-D_\gamma h(d)} \leq \frac{cd}{\sqrt{t}} \left(c^{-1} e^{-C_\gamma(1+\frac{\lambda}{d})} \right)$$

is at most

$$\exp \left((t+d+1)\lambda - \frac{d^2}{4} \left[C_\gamma \left(1 + \frac{\lambda}{d} \right) - \log \frac{1}{c} \right] \right).$$

By assumption, $t\lambda \leq 8k(d)\lambda \leq C(d\lambda + d^2)$, so this probability is at most $n^{-\beta}$ for C_γ sufficiently large.

Now, consider the case where $k(d) < t/8$, and let $u = t - 4k(d)$, so that $u > t/2 > k(d)$. If our desired result (5.4) holds at stage u , then

$$\sigma_{u-k}(T_u) \geq g(k) \quad \text{and} \quad \Sigma_k(T_u) \leq \sum_{\ell=k}^{u-1} g(\ell)^{-2} \leq 2kg(k)^{-2}.$$

Note that $g(k) \leq k/\sqrt{n} \leq \sqrt{k}$, so we may apply Lemma 5.4 with $\mu = g(k)$, providing our desired inductive step with all but probability $n^{-\beta}$. Taking a union bound over the $< n^2$ pairs (t, d) , the total failure probability is at most $n^2 \times n^{-\beta} = n^{-\gamma}$, completing the proof. $\square$

5.5. A Proof of Theorem 1.4. Using Propositions 5.1 and 5.5, Theorem 1.4 follows quickly.

Proof of Theorem 1.4. Let $r = \lceil \sqrt{\log n} \rceil$. By Proposition 5.5, with exponent $\gamma = \alpha + 1$,

$$\max_{k \leq n} \Sigma_r(\Pi(G)_{[k],[k]}) \leq ne^{C_\alpha \sqrt{\log n}}$$

with probability at least $1 - n^{-(\alpha+1)}$. Combining this with Proposition 5.1, with exponent $p = \alpha + 1$, gives our desired result. $\square$

6. A SMOOTHED ANALYSIS OF PARTIAL PIVOTING

Here, we extend the analysis of previous section to the smoothed setting $A + \epsilon G$, where $A \in \mathbb{R}^{n \times n}$ is fixed and $\epsilon > 0$. We prove that the growth factor of such a matrix is at most $n^{O(1)} \epsilon^{-O(\log \log n)}$ with overwhelming probability.

6.1. Adapting the Argument for General Means. The argument in Section 5 carries over to the general mean setting, up to some additional loss. By the scale invariance of the growth factor, we may replace $A + \epsilon G$ by $A/\epsilon + G$, so that our matrix remains unit variance but with a possibly very large mean. Let $d_0 = d$, $d_{i+1} = k(d_i)$, and define

$$f_K(d) = \prod_{i>0, d_i < n} \left(1 + \frac{K^2}{d_i} \right) \quad \text{and} \quad g_K(d) = \frac{d}{\sqrt{n}} \frac{e^{-D_\gamma h(d)}}{f_K(d)}.$$

In the proof, we set $K = \epsilon^{-1} \|A\|_2$. Tracking the argument, we lose an additional $f_K(r)^2$ when bounding the truncated inverses (the adaptation of Proposition 5.5) and an additional $(1+K)^3$ when converting truncated inverse control to control on the growth (the adaptation of Proposition 5.1). Note that when $\|A\|_2 = 0$, both of these terms are one, recovering the original argument in Section 5.

Theorem 6.1. *For every $\alpha > 0$, there exist constants C_α, n_α such that, for all $n \geq n_\alpha$, every $\Delta \in \mathbb{R}^{n \times n}$, $\epsilon > 0$, and every $\lceil \sqrt{\lambda} \rceil \leq r < n/8$,*

$$\mathbb{P} \left[\text{growth}(\Pi(\Delta + \epsilon G)) > C_\alpha \left(1 + \frac{\|\Delta\|_2}{\epsilon} \right)^3 \sqrt{\frac{n}{r}} e^{C_\alpha(r+\frac{\lambda}{r})} f_{\frac{\|\Delta\|_2}{\epsilon}}(r) \right] < n^{-\alpha}.$$

Proof. Without loss of generality, let $\epsilon = 1$. We define $B = \Delta + G$, π_t to be the first t pivot indices of B , $T_t = B_{\pi_t, [t]}$, and $K = \|\Delta\|_2$. $\mathcal{H}_t$ records π_t and T_t , and $\mathcal{F}_t$ records $B_{[n], [t]}$. The elimination matrices are given by

$$E_t = [-B_{\pi_t^c, [t]} T_t^{-1} I] P_t$$

and we use the decomposition $E_t = X_t + Y_t Q_t^T$ from Lemma 5.2, where $T_t^{-1} = Z_t + H V^T$, with

$$X_t = [-B_{\pi_t^c, [t]} Z_t I] P_t, \quad Y_t = -B_{\pi_t^c, [t]} H, \quad \text{and} \quad Q_t = P_t^T \begin{bmatrix} V \\ 0 \end{bmatrix}.$$

First, we convert the truncated inverse bound into a bound on growth (i.e., modify Proposition 5.1). Suppose $\Sigma_r(T_t) \leq \tau$ for every $t < n$ with probability at least $1 - \eta$. We aim to bound growth under this hypothesis. Conditional on $\mathcal{H}_t$, each non-pivoted row in the first t coordinates is $N(\Delta_{i, [t]}, I)$ restricted to same origin-symmetric convex set $\mathcal{K}_t$ as before. For $t > r$, on the $\mathcal{H}_t$ measurable event that $\Sigma_r(T_t) \leq \tau$, Lemma 4.4 and a union bound implies that

$$\mathbb{P} \left[\max_{i \in \pi_t^c} \|B_{i, [t]} Z_t\|_2 > \left(\sqrt{2 + 4q\lambda} + K \right) \sqrt{\tau} \mid \mathcal{H}_t \right] < n^{1-q},$$

and so, for C_q large enough and $R = C_q(1 + K^2 + \lambda)(1 + \tau)$,

$$\mathbb{P} \left[\max_i \|(X_t)_{i, [n]}\|_2^2 > R \mid \mathcal{H}_t \right] < n^{1-q}$$

on that same event. Again, for $t \leq r$, the rows of X_t have two-norm one, so the bound always holds. Now, let

$$B_\star = B_{[n], [t+4r] \setminus [t]}, \quad G_\star = G_{[n], [t+4r] \setminus [t]}, \quad \text{and} \quad \Delta_\star = \Delta_{[n], [t+4r] \setminus [t]},$$

and consider $4r$ additional steps of Gaussian elimination. Each row of $X_t B_\star$ has independent Gaussian entries, with variance at most R and mean row two-norm at most $K\sqrt{R}$. Also, conditional on $\mathcal{F}_t$,

$$Q_t^T B_\star = Q_t^T \Delta_\star + Q_t^T G_\star,$$

where $Q_t^T G_\star$ is Gaussian. By Lemmas 4.1 and 4.2, conditioned on the event $\max_i \|(X_t)_{i, [n]}\|_2^2 \leq R$,

$$\max_i \|(X_t B_\star)_{i, [4r]}\|_2^2 \leq C_q R(K^2 + r + \lambda) \quad \text{and} \quad \|(Q_t^T B_\star)^\dagger\|_2^2 \leq C r^{-1} e^{C_q \frac{\lambda}{r}},$$

except with conditional probability at most Cn^{-q} . Again, let J record the next $4r$ steps of elimination, with the non-pivot rows returned to increasing order. As in Lemma 5.3,

$$E_{t+4r} = J(X_t - X_t B_\star (Q_t^T B_\star)^\dagger Q_t^T) \quad \text{and} \quad \max_i \|J_{i, [n-t]}\|_1 \leq 2^{4r}.$$

Taking a union bound, together with the deterministic bound for $u \leq 4r$, we have

$$\max_{u < n} \max_i \|(E_u)_{i, [n]}\|_2 \leq (1 + K)^2 \sqrt{1 + \tau} e^{C_q(r + \frac{\lambda}{r})}$$

with probability at least $1 - \eta - Cn^{2-q}$. For a row e^T of E_u and a column $j > u$, conditional on $\mathcal{F}_u$,

$$e^T B_{[n], j} \sim N(e^T \Delta_{[n], j}, \|e\|_2^2).$$

Using the bound $|e^T \Delta_{[n], j}| \leq K \|e\|_2$ and applying a Gaussian tail bound over the $O(n^3)$ Schur complement entries loses an additional factor $K + \sqrt{2q\lambda}$. The same probabilistic argument lower bounding the max norm of the matrix also holds for $\|\Delta + G\|_{\max}$, and still $\|L\|_{\max} = 1$. All together, we obtain

$$\text{growth}(\Pi(\Delta + G)) \leq (1 + K)^3 \sqrt{1 + \tau} e^{C_q(r + \frac{\lambda}{r})}$$

with probability $1 - \eta - Cn^{3-q}$ at least. This completes the first half of the proof.

What remains is to produce a modified truncated inverse bound (i.e., modify Proposition 5.5). First, we adapt Lemma 5.4. Restricting to the $\mathcal{F}_m$ -measurable event $\sigma_{m-k}(T_m) \geq \mu$ and $\Sigma_k(T_m) \leq 2k\mu^{-2}$. Lemma 4.5 multiplies both bounds in (5.2) by v_k . As in Lemma 5.4, we fix a set S of size $s = 4k$ and

condition on all the other rows. Again, π_m^S is the first m pivots after deleting S , $j = \lfloor d/2 \rfloor$, and here the columns of Q_S form an orthonormal basis for $\text{null}(B_{\pi_m^S, [m+s]})$. On the event that $B_{\pi_m^S, [m+s]}$ does satisfy the above two bounds, Lemma 4.1 with $x = (s + \beta + 1)\lambda$ implies that

$$\sigma_{j+1}(B_{S, [m+s]}(B_{\pi_m^S, [m+s]})^\dagger) \leq \sqrt{k}\mu^{-1}v_k e^{C_\beta(1+\frac{\lambda}{d})},$$

except with conditional probability at most $n^{-s-\beta-1}$. The small singular value estimate for $B_{S, [m+s]}Q_S$ is unchanged, as Lemma 4.3 is unchanged by a non-zero mean. Together, the new output threshold for Lemma 5.4 is

$$\frac{\mu d}{kv_k} e^{-C_\beta(1+\frac{k\lambda}{d^2})}.$$

The quantity $f_K(r)$ tracks the accumulation of the additional term v_k in the denominator. Take $\beta = \gamma + 2$ and choose D_γ sufficiently large. Because $f_K(d)$ is non-increasing in d , $g_K(d)/d$ is non-decreasing still. The same argument holds for the product recurrence $k(d) < n$, now with

$$\frac{g_K(d)}{g_K(k)} = \frac{d}{kv_k} e^{-D_\gamma(1+\frac{\lambda}{d})},$$

which leads to the worse bound

$$\max_{t < n} \Sigma_r(T_t) \leq \frac{2n}{r} e^{2D_\gamma h(r)} f_K(r)^2 \leq \frac{C_\gamma n}{r} e^{C_\gamma(\frac{\lambda}{r} + \log \log n)} f_K(r)^2$$

with probability at least $1 - n^{-\gamma}$.

The result follows by combining both arguments, with $\gamma = \alpha + 1$, $\eta = n^{-\gamma}$, and $q = \alpha + 4$. $\square$

6.2. Proof of Theorem 1.6. Our desired Theorem 1.6 follows immediately from Theorem 6.1.

Proof of Theorem 1.6. Consider Theorem 6.1 with $\Delta = A$ and $r = \lceil e^4 \lambda \rceil$, for which $r + \lambda/r = O(\lambda)$. Let n_α be large enough so that $r < n/8$. Let m be the last index for which $d_m < n$ in the chain starting with $d_0 = r$. Because $d_{i+1} \geq d_i^2/\lambda$,

$$\log(d_i/\lambda) \geq 2^i \log(r/\lambda) \geq 2^{i+2}.$$

Therefore, $2^{m+2} < \lambda$ and $m+2 < \log(\lambda)/\log 2$. Let $K = \epsilon^{-1}\|A\|_2$. There are m factors in $f_K(r)$, each at most $2\epsilon^{-2}$, and so

$$(1+K)^3 f_K(r) \leq 2^{m+3} \epsilon^{-(2m+3)} \leq 2\lambda \epsilon^{-(2/\log 2) \log \lambda}.$$

$\square$

ACKNOWLEDGEMENTS

The author thanks Louisa Thomas for improving the style of presentation, Alan Edelman for many conversations on all aspects of Gaussian elimination, Nick Trefethen and Konstantin Tikhomirov for conversations regarding the average-case analysis of partial pivoting, and Rikhav Shah for conversations regarding orthogonal preconditioning. The author also thanks Nick Trefethen for comments on an earlier unpublished draft. This material is based upon work supported by the National Science Foundation under Grant No. DMS-2513687.

AI USAGE STATEMENT

The research presented here is the result of human–AI interaction involving ChatGPT Sol 5.6 & Astra 6. For example, the proof of Conjecture 1.1 is the result of a human-only proof sketch for the bottom right $o(n)$ entries of U_{ij} , which was improved to a proof for all entries through AI interaction. The other results presented here follow a similar trajectory: a combination of some human-only ideas and expertise which was improved through AI interaction. The results of Sections 2 and 3, though quite interesting to the author, are not technically difficult, and likely did not need human input. There is reason to suggest, however, that the proof of Conjecture 1.1 was resilient to standard use of an AI system. Conjecture 1.1 was listed on the well-known GitHub repository OpenProblemsInNLA, for which roughly half of the > 200 open problems have been solved, largely with autonomous AI systems [20]. Conjecture 1.1 was given a designation of open, challenging, and interesting to the community and remained unsolved prior to this manuscript. The author welcomes detailed questions regarding the role of AI in this work. The entire manuscript was written by the author.

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*, 57(5):1369–1387, 2025.
- [2] Sanjoy Dasgupta and Anupam Gupta. An elementary proof of a theorem of Johnson and Lindenstrauss. *Random Structures & Algorithms*, 22(1):60–65, 2003.
- [3] 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.
- [4] Tobin A Driscoll and Kara L Maki. Searching for rare growth factors using multicanonical Monte Carlo methods. *SIAM Review*, 49(4):673–692, 2007.
- [5] Alan Edelman and Walters Mascarenhas. On the complete pivoting conjecture for a Hadamard matrix of order 12. *Linear and Multilinear Algebra*, 38(3):181–187, 1995.
- [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] Nathan Halko, Per-Gunnar Martinsson, and Joel A Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011.
- [8] Nicholas J Higham. *Accuracy and stability of numerical algorithms*. SIAM, 2 edition, 2002.
- [9] Daniel Hsu, Sham Kakade, and Tong Zhang. A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, 17(52):1–6, 2012.
- [10] Han Huang and Konstantin Tikhomirov. Average-case analysis of the Gaussian elimination with partial pivoting. *Probability Theory and Related Fields*, 189(1):501–567, 2024.
- [11] Samuel Orellana Mateo, John Urschel, and Nicholas West. LU factorization of discrete random matrices. *arXiv preprint arXiv:2608.08998*, 2026.
- [12] Hoi H Nguyen. Random matrices: Overcrowding estimates for the spectrum. *Journal of Functional Analysis*, 275(8):2197–2224, 2018.
- [13] Hoi H Nguyen and Van Vu. Random matrices: Law of the determinant. *The Annals of Probability*, 42(1):146–167, 2014.
- [14] Douglass Stott Parker. Random butterfly transformations with applications in computational linear algebra. Technical Report CSD-950023, Computer Science Department, UCLA, 1995.
- [15] 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.
- [16] András Prékopa. On logarithmic concave measures and functions. *Acta Scientiarum Mathematicarum*, 34:335–343, 1973.
- [17] Arvind Sankar. *Smoothed analysis of Gaussian elimination*. PhD thesis, Massachusetts Institute of Technology, 2004.
- [18] 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.
- [19] Rikhav Shah and John Urschel. Entry growth in Gaussian elimination. *arXiv preprint arXiv:2608.19189*, 2026.
- [20] Alex Townsend. Open Problems in Numerical Linear Algebra, 2026. GitHub repository.

- [21] Lloyd N Trefethen. Instability of Gaussian elimination is exponentially rare (proof of partial result). *arXiv preprint*, 2026.
- [22] Lloyd N Trefethen and David Bau, III. *Numerical linear algebra*. SIAM, 1997.
- [23] 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.
- [24] Nick Trefethen. The smart money’s on numerical analysts. *SIAM News*, 45(9), 2012.
- [25] John Urschel. Numerical stability in Gaussian elimination. *Notices of the American Mathematical Society*, 72(6):605–613, 2025.
- [26] Martin J Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press, 2019.
- [27] James Hardy Wilkinson. Error analysis of direct methods of matrix inversion. *Journal of the ACM (JACM)*, 8(3):281–330, 1961.
- [28] 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.