AVERAGE-CASE STABILITY OF GAUSSIAN ELIMINATION*

LLOYD N. TREFETHEN† AND ROBERT S. SCHREIBER‡

Dedicated to the memory of Jim Wilkinson.

Abstract. Gaussian elimination with partial pivoting is unstable in the worst case: the "growth factor" can be as large as 2^{n-1}, where n is the matrix dimension, resulting in a loss of $n - 1$ bits of precision. It is proposed that an average-case analysis can help explain why it is nevertheless stable in practice. The results presented begin with the observation that for many distributions of matrices, the matrix elements after the first few steps of elimination are approximately normally distributed. From here, with the aid of estimates from extreme value statistics, reasonably accurate predictions of the average magnitudes of elements, pivots, multipliers, and growth factors are derived. For various distributions of matrices with dimensions $n \leq 1024$, the average growth factor (normalized by the standard deviation of the initial matrix elements) is within a few percent of $n^{2/3}$ for partial pivoting and approximately $n^{1/2}$ for complete pivoting. The average maximum element of the residual with both kinds of pivoting appears to be of magnitude $O(n)$, as compared with $O(n^{1/2})$ for QR factorization.

The experiments and analysis presented show that small multipliers alone are not enough to explain the average-case stability of Gaussian elimination; it is also important that the correction introduced in the remaining matrix at each elimination step is of rank 1. Because of this low-rank property, the signs of the elements and multipliers in Gaussian elimination are not independent, but are interrelated in such a way as to retard growth. By contrast, alternative pivoting strategies involving high-rank corrections are sometimes unstable even though the multipliers are small.

Key words. Gaussian elimination, stability, pivoting, growth factor, extreme values

AMS(MOS) subject classifications. 65F05, 65G05

Notation.

A	matrix in $\mathbb{R}^{n \times n}$,
σ_A	standard deviation of elements of A,
$A^{(k)}$	modified matrix before step k of elimination,
$\hat{A}^{(k)}$	modified matrix at step k after pivoting but before row operations,
$U = A^{(n)}$	final upper-triangular matrix,
$u_{kk} = \hat{a}_{kk}^{(k)}$	kth pivot,
m	$n + 1 - k$ (partial pivoting), $(n + 1 - k)^2$ (complete pivoting),
σ_k	standard deviation of elements $a_{ij}^{(k)}$ ($k \leq i, j \leq n$),
π_k	average absolute value of pivots u_{kk},
μ_k	standard deviation of multipliers $\hat{a}_{ik}^{(k)}/\hat{a}_{kk}^{(k)}$ ($k < i \leq n$),
$\rho, \tilde{\rho}$	growth factor, growth factor normalized by σ_A,
$W(m)$	extreme value or "winner" function for normal random variables,
N	sample size,
$\langle \cdot \rangle$	expected value.

0. Introduction. At the beginning of the computer era, it was feared that Gaussian elimination would be an ineffective method for solving systems of linear equations. A paper by Hotelling in 1943 [19] predicted that in the solution of $n \times n$ systems of the form $A^T A x = b$, errors might be amplified by as much as 4^{n-1}, so that a "78-rowed matrix would need to be carried to no less than 46 places to insure even an approximate

accuracy in the first decimal place." Another paper by Bargmann, Montgomery, and von Neumann in 1946 [1] stated that "very little is known about the stability of the methods so far described, [but] what information there is tends to indicate that these methods are unstable and that rounding errors accumulate so seriously that the methods are impractical for large values of n."

By the early 1950s, computational experience had revealed that these fears were groundless, and Gaussian elimination with partial pivoting rapidly became the universal algorithm for solving general dense systems of linear equations. Progress was also made on the theoretical side by Turing [30], von Neumann and Goldstine [31], and especially Wilkinson [32], [33], whose elegant arguments based on condition numbers and backward error analysis shed light on every aspect of the elimination process. The result of these developments is that a widespread view among numerical analysts nowadays, thirty years later, is roughly that "Wilkinson proved that Hotelling's prediction was too pessimistic."

This view is not entirely accurate, however, for a fundamental gap in our understanding remains. When Gaussian elimination with partial pivoting is performed on an $n \times n$ matrix A, the result is a factorization $PA = LU$, where P is a permutation matrix, L is unit lower triangular, and U is upper triangular. Let $\bar{x}$ denote the solution of a linear system $Ax = b$ computed in floating-point arithmetic. Wilkinson proved that under reasonable assumptions, the relative error in $\bar{x}$ satisfies

$$(0.1) \qquad \frac{\|\bar{x} - x\|_\infty}{\|x\|_\infty} \leq 4n^2 \kappa_\infty(A) \rho \varepsilon,$$

where ε is the machine precision, $\kappa_\infty(A)$ is the condition number of A in the supremum norm, and ρ is the **growth factor**,

$$(0.2) \qquad \rho = \frac{\max_{i,j,k} |a_{ij}^{(k)}|}{\max_{i,j} |a_{ij}|},$$

with $a_{ij}^{(k)}$ denoting the i, j element before the kth step of elimination [33]. (Results like (0.1) appear in various forms, with different definitions of ρ, norms, and polynomial factors; we have picked a representative one.) Unfortunately, ρ may be as large as 2^{n-1} (though no larger), as is proved by the simple example shown here for $n = 5$:

(0.3)

$$
\begin{bmatrix}
1 & & & & 1 \\
-1 & 1 & & & 1 \\
-1 & -1 & 1 & & 1 \\
-1 & -1 & -1 & 1 & 1 \\
-1 & -1 & -1 & -1 & 1
\end{bmatrix}
=
\begin{bmatrix}
1 & & & & \\
-1 & 1 & & & \\
-1 & -1 & 1 & & \\
-1 & -1 & -1 & 1 & \\
-1 & -1 & -1 & -1 & 1
\end{bmatrix}
\begin{bmatrix}
1 & & & & 1 \\
 & 1 & & & 2 \\
 & & 1 & & 4 \\
 & & & 1 & 8 \\
 & & & & 16
\end{bmatrix}.
$$

$$\qquad\qquad A \qquad\qquad\qquad\qquad L \qquad\qquad\qquad\qquad U$$

It follows that unless (0.1) is highly pessimistic, Gaussian elimination will be useless for certain matrices. And so it is.[1]

Thus Gaussian elimination is unstable in the worst case; the improvement from Hotelling to Wilkinson is merely from 4^{n-1} to 2^{n-1}. Why, then, is it successful in practice? Indeed, partial pivoting is so reliable that most of the software in use today—including

---

[1] Thanks to the integer entries and unit diagonal elements, experiments with this matrix A sometimes reveal no instability. To be sure of seeing it, choose a right-hand side with negative as well as positive entries, or perturb the elements of A slightly in such a way that the pivot sequence is preserved.

LINPACK [8]—does not even bother to monitor pivot growth, although that would be a fail-safe method of guarding against instability.

We propose that a partial answer would be obtained if we could show that Gaussian elimination is stable *on average*. Average-case analysis has not been popular in numerical linear algebra, partly because of the obvious fact that the matrices encountered in practical problems are by no means random. Indeed, some researchers have expressed the opinion that Gaussian elimination is stable in practice precisely because the matrices that occur in practice are better behaved than if they were random.[2] The purpose of this paper is to argue the opposite opinion. We believe that Gaussian elimination is stable because the matrices encountered in practice *are* random, to a sufficient degree, and that the essential reason examples such as (0.3) do not cause trouble is that they occupy a negligible proportion of the space of matrices.

We began this project with the optimistic conjecture that Gaussian elimination is stable on average for a combination of two reasons:

(1) The magnitudes of the multipliers are on average much less than 1;

(2) The signs of the multipliers and elements are effectively random and tend to cancel.

Both of these hypotheses are readily translated into quantitative predictions, but when carried out, it was quickly found that the two of them, taken together, are not enough to explain experimental observations. In actuality, as will be discussed in § 6, average growth factors in Gaussian elimination exhibit a mild $n^{2/3}$ dependence on n, at least for $n \leq 1024$, whereas (1) and (2) lead to a prediction on the order of $e^{n/4 \log n}$ (see eq. (5.4)). This paper can be viewed as an exploration of how (1) and (2) can be made precise, and modified where necessary, to explain this behavior. To summarize, §§ 2–4 show that hypothesis (1) is valid: simple estimates based on extreme value statistics give good predictions of observed multipliers, which are indeed on average small. The trouble lies in hypothesis (2), which must be corrected as follows:

(2′) The signs of the multipliers and elements are "better than random" from the point of view of cancellation.

For many distributions of matrices the multipliers and elements are *uncorrelated* in the sense that their covariances are zero—this follows from simple sign considerations—but they are not independent. On the contrary, there are relationships among them that conspire to retard growth. A tentative explanation of this phenomenon, together with a quantitative model of it, are proposed in § 5.

For a quick demonstration that the numbers produced by Gaussian elimination with pivoting are highly dependent, factor a random matrix A into $PA = LU$, i.e., $U = L^{-1}PA$, and you will find that $\|L^{-1}\|$ is reasonably small—33.2 in one experiment with $n = 256$. Now, randomize the signs of the elements of L and compute $\|L^{-1}\|$ again. It will be dramatically larger—in the same experiment, 2.7×10^8.

Our statistical arguments and numerical experiments indicate that for matrices that are random in various senses, both growth factors and computed residuals tend to be no

---

[2] For example, Gaussian elimination is particularly stable for ill-conditioned matrices, and some have suggested, with discretization of partial differential equations in mind, that its stability in practice comes about because most matrices arising in practice tend to be exceptionally ill-conditioned. However, this is not true; the average $n \times n$ matrix has condition number $O(n)$ or larger [11], [12], while the condition number for the standard discretization of Poisson's equation is $O(n)$ in two space dimensions and only $O(n^{2/3})$ in three dimensions. (See a similar remark on p. 460 of [1].) Even if it were true, this kind of argument could not explain the success of Gaussian elimination. If examples such as (0.3) were typical in the space of matrices, it would not be enough for *most* matrices arising in practice to be exceptionally well behaved; essentially *all* of them would have to be exceptional, which is highly implausible.

larger than $O(n)$ on average. This is true for matrices with elements drawn from a normal distribution and for various other classes of matrices too; in fact, the growth factors and residuals often depend only on the standard deviation of the initial matrix elements. Although the initial matrix elements may be far from normally distributed, a few steps of Gaussian elimination typically bring them toward that form. A more systematic summary of our results can be found in the final section.

An analogous problem of average- versus worst-case behavior—concerning speed rather than stability—appears in linear programming. The simplex method was invented in 1947, and it was soon recognized that the number of steps to convergence is usually small in practice, even though the worst-case behavior is exponential [20]. The problem of obtaining an average-case convergence result became well publicized beginning in 1963 [6, p. 160], and in recent years has been solved in various senses by Borgwardt, Smale, and others [3], [25], [26], [24].

The problem of stability of Gaussian elimination is an embarrassing theoretical gap at the heart of numerical analysis. We believe that it is also of practical importance. One reason is that pivoting conflicts with both sparsity preservation and parallelization, so that less stringent strategies such as threshold pivoting [10] and pairwise pivoting [27] are attracting increasing attention (see § 8). We can hardly assess these variants fully while our understanding of classical Gaussian elimination remains incomplete. A more basic reason is that as computers grow more powerful, n is getting bigger. Traditionally, polynomial factors like the n^2 term in (0.1) have been ignored as moderate in size and in any case generally pessimistic, but as n increases from 10^2 (Wilkinson?) to 10^3 (LIN-PACK?) to 10^4 (supercomputers?) to 10^6 (the year 2000?) and beyond, the need for a more quantitative understanding of stability will grow. Average-case modeling of error propagation is already a well-established tool, for example, in the study of fast Fourier transforms for digital signal processing [22].

We wish to acknowledge several previous experimental studies of the behavior of Gaussian elimination for random matrices and related matters: by Goodman and Moler [16] (reported also in the LINPACK manual [8]), by Birkhoff and Gulati [2], and by MacLeod [21], [34] who presents detailed statistics from Gaussian elimination applied to random matrices of dimensions $n \leq 100$ with sample sizes 10,000. Higham and Higham have investigated general classes of matrices with large growth factors [18]. Many theoretical questions concerning eigenvalues and condition numbers of random matrices have recently been settled by Edelman [11], [12].

1. Preliminaries. Throughout this paper A denotes a real $n \times n$ matrix, and $A^{(k)}$, $1 \leq k \leq n$, is the modified matrix, with zeros below the diagonal in the first $k - 1$ columns, that remains before the kth step of Gaussian elimination. The end result is an upper-triangular matrix $U = A^{(n)}$. We denote by $\hat{A}^{(k)}$ the intermediate matrix obtained after pivoting but before elimination at step k; thus the kth elimination step has the form

$$\text{Step } k: \quad A^{(k)} \to \hat{A}^{(k)} \to A^{(k+1)} \quad (1 \leq k \leq n - 1).$$

The i, j entries of $A^{(k)}$ and $\hat{A}^{(k)}$ are denoted by $a_{ij}^{(k)}$ and $\hat{a}_{ij}^{(k)}$, respectively, and $\hat{a}_{kk}^{(k)} = a_{kk}^{(n)} = u_{kk}$ is the kth pivot element.

The growth factor ρ of (0.2) is intimately connected with the pivots u_{kk}: for complete pivoting (rows and columns) $\rho = \max_k |u_{kk}|/|u_{11}|$ exactly, and for partial pivoting (rows only) the details are more complicated but large growth is again usually associated with large pivots. On the other hand, a constraint on the size of the pivots is provided by Hadamard's inequality,

$$(1.1) \qquad \prod_{k=1}^{n} |u_{kk}| = |\det A| \leq \prod_{k=1}^{n} \|a_k\|,$$

where $\|a_k\|$ is the 2-norm of the kth column of A. If A is $\sqrt{n}$ times an orthogonal matrix (the factor $\sqrt{n}$ being introduced to make the standard deviation of the elements equal to 1), then (1.1) becomes

$$(1.2) \qquad \prod_{k=1}^{n} |u_{kk}| = (\sqrt{n})^n.$$

Similarly, if A is a random matrix with independent elements drawn from the standard normal distribution of mean 0 and standard deviation 1, a known result on expected determinants [15] gives

$$(1.3) \qquad \left\langle \prod_{k=1}^{n} |u_{kk}| \right\rangle = \sqrt{n!} \sim (2\pi n)^{1/4}(\sqrt{n/e})^n.$$

(Here and throughout the paper, $\langle \cdot \rangle$ denotes the expected value.) These observations imply that so long as the pivots are reasonably uniform in magnitude, they must be of a modest size, comparable to $\sqrt{n}$. Large pivots can occur only if the pivots are highly nonuniform, as in (0.3).

Sections 2–5 of this paper are devoted to investigating, by statistical arguments and numerical experiments, the dependence on n and k of the following quantities:

$$(1.4) \qquad \sigma_k = \langle (a_{ij}^{(k)})^2 \rangle^{1/2} \quad \text{standard deviation of elements } (k \le i, j \le n),$$

$$(1.5) \qquad \pi_k = \langle |\hat{a}_{kk}^{(k)}| \rangle = \langle |u_{kk}| \rangle \quad \text{average absolute value of pivots,}$$

$$(1.6) \qquad \mu_k = \langle (\hat{a}_{ik}^{(k)}/\hat{a}_{kk}^{(k)})^2 \rangle^{1/2} \quad \text{standard deviation of multipliers } (k < i \le n).$$

(In the definitions of σ_k and μ_k, i and j are any integers in the ranges indicated; for most distributions of matrices A of practical interest, symmetry implies that the statistics are independent of these indices.) We shall argue that for many distributions of matrices, σ_k and π_k grow slowly and steadily with k, never attaining very large values. Section 6 applies these results to investigate average growth factors and § 7 reports numerical experiments concerning average residuals.

Our experiments are based on eight classes of matrices:

normal   standard normal distribution of mean 0, variance 1,
$[-1, 1]$ uniform distribution on $[-1, 1]$,
$[0, 1]$ uniform distribution on $[0, 1]$,
$\{-1, 1\}$ discrete distribution with $p(-1) = p(1) = \frac{1}{2}$,
$\{0, 1\}$ discrete distribution with $p(0) = p(1) = \frac{1}{2}$,
symm.   symmetric matrices with elements from the standard normal dist.,
Toep.   Toeplitz matrices with elements from the standard normal dist.,
orth.   orthogonal matrices distributed by Haar measure.

In the first five cases, the elements of an individual matrix are independent samples from the distributions indicated, while the final three cases have dependent elements. The random orthogonal matrices are calculated by a sequence of Householder reflections as proposed by Stewart [28]; Haar measure is a name for the isotropic distribution of orthogonal matrices in which each column or row is uniformly distributed on the unit $(n-1)$-sphere.

In each of our experiments, matrices A of one or more dimensions n are selected at random from one of these classes, with the sample size N diminishing with n to keep the computing time within reasonable bounds. A typical set of dimensions and sample sizes are listed below, although for some of our experiments the samples were larger.

dimension n	2	4	8	16	32	64	128	256	512	1024
sample size N	4096	2048	1024	512	256	128	64	32	20	10

Our calculations have been carried out in single precision Fortran 77 on SUN workstations, IBM-compatible personal computers, a CRAY-2, and an Ardent Titan; no machine dependences were observed. Most of our experiments, but not all, have made use of the shuffled random number generators RAN1 and GASDEV in [23]. Plotting, data analysis, and hundreds of supporting tests of every kind were carried out with the superb matrix "workbench" program MATLAB, without whose powerful assistance a project of this kind would have been difficult indeed.

2. Elements. Our arguments begin with a fundamental observation: for many classes of matrices, the elements $a_{ij}^{(k)}$ at the kth step of Gaussian elimination tend to be normally distributed with mean 0. This statement is not exactly valid for $k > 1$, even if the elements of the initial matrix $A = A^{(1)}$ are themselves normally distributed, nor is it asymptotically valid in any limit such as $k, n \to \infty$, so far as we know, since the conditions of the central limit theorem are not satisfied by Gaussian elimination. Nevertheless, the hypothesis of normally distributed elements is often an excellent approximation, after the first few steps of elimination, even when the elements of A are not normally distributed.

Figure 2.1 provides evidence for this claim. For each half of the figure—partial and complete pivoting—1280 matrices of dimension 64 with normally distributed elements have been factored, and the elements $a_{ij}^{(k)}$ ($k \leq i, j \leq n$) in columns $k = 1, 9, 17, \cdots,$ 57 accumulated in bins. The data are plotted as asterisks after being rescaled to have
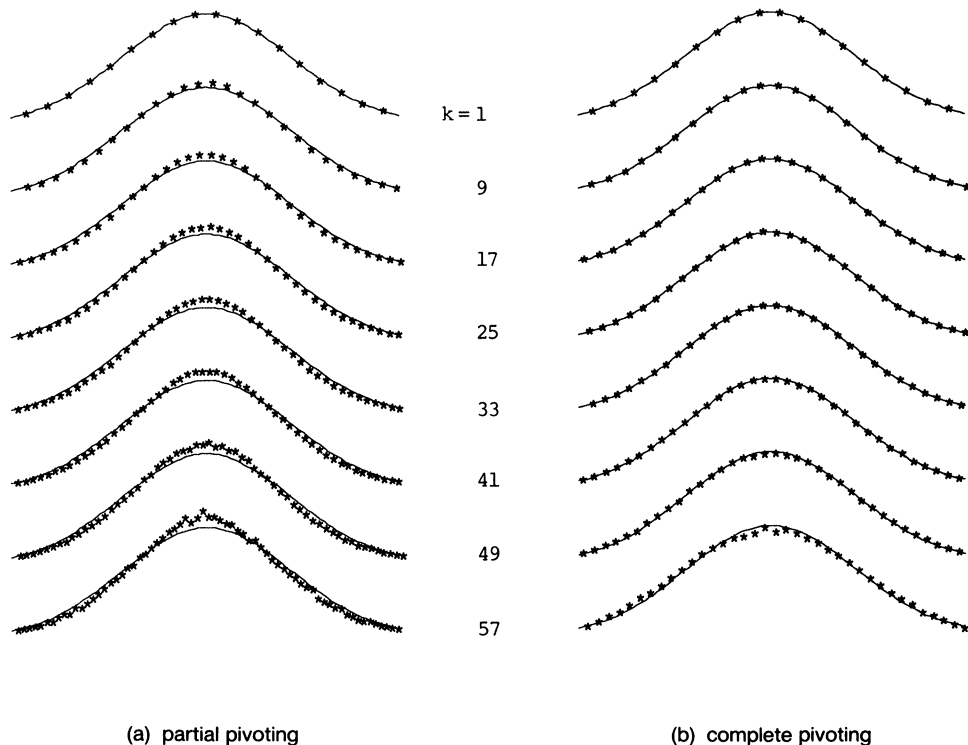

(a) partial pivoting   (b) complete pivoting

FIG. 2.1. *Distributions of elements* $a_{ij}^{(k)}$ *rescaled to have variance* 1, *for* $n = 64$: *observed* (∗) *and normal distribution* (———).

standard deviation 1, and the solid curves show the normal distribution for comparison. The agreement of the two is excellent. (The noise toward the end results from the smaller numbers of elements in the samples.) It is not perfect, however; evidently partial pivoting leads to a distribution that is slightly more peaked in the center than the normal distribution. Similar plots are obtained for other values of n and k.

Although the shape of the element distribution is roughly independent of k, its standard deviation σ_k grows considerably. This is visible in the increasing density of asterisks in Fig. 2.1 as $k \to n$, especially for partial pivoting. (The bins holding the raw data before rescaling were equally spaced.) This dependence of σ_k on k is essentially the growth that is the subject of this paper. We shall model it in § 5.

When matrices from nonnormal distributions are investigated, the results are often surprisingly similar to those of Fig. 2.1. As a modest example, Fig. 2.2 details the initial steps of Gaussian elimination for matrices with elements from the uniform $[-1, 1]$ distribution. At $k = 1$, the asterisks reveal the initial square wave, but by $k = 8$, the distributions have become very close to normal, and for higher k (not shown) they are barely distinguishable from those of Fig. 2.1. The same phenomenon occurs with most of the classes of matrices listed in the last section. The exception is orthogonal matrices, for which the element distribution is approximately normal for much of the elimination but changes to a pronounced bimodal form toward the end.

From now on, then, we shall assume that at every step of elimination, the elements $a_{ij}^{(k)}$ are normally distributed with mean 0; only the standard deviation σ_k depends on k. For most of the argument, we shall further assume that the elements are independent, until we are forced to abandon that assumption at (5.5).
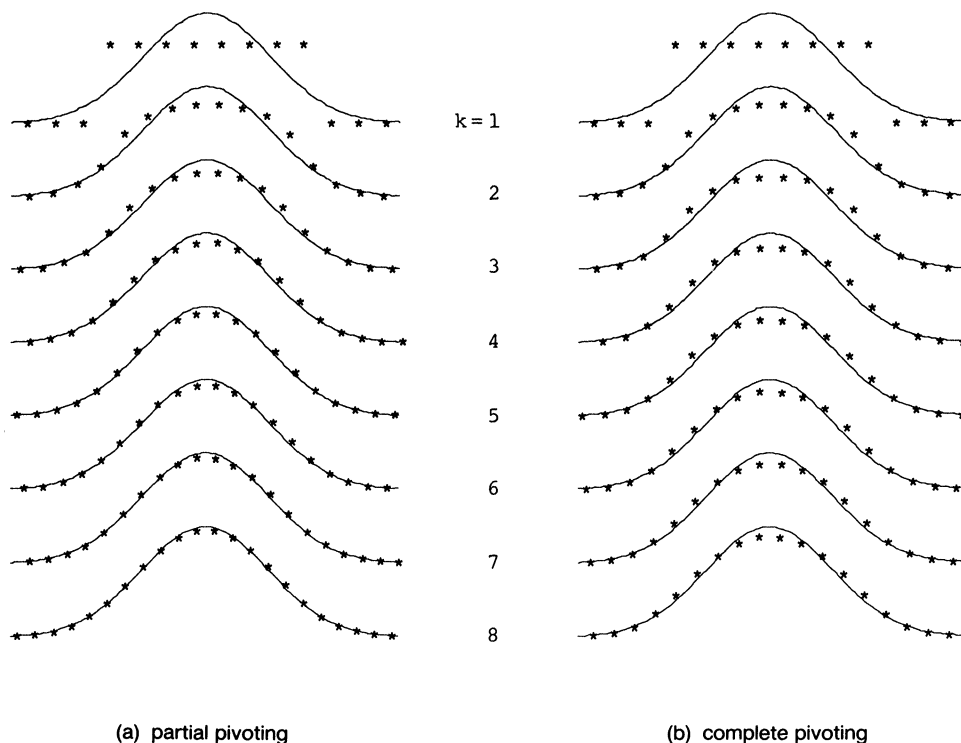

(a) partial pivoting   (b) complete pivoting

FIG. 2.2. *Similar to Fig. 2.1, but for matrices with elements uniformly distributed in* $[-1, 1]$. *Only the initial steps* $1 \leq k \leq 8$ *are shown.*
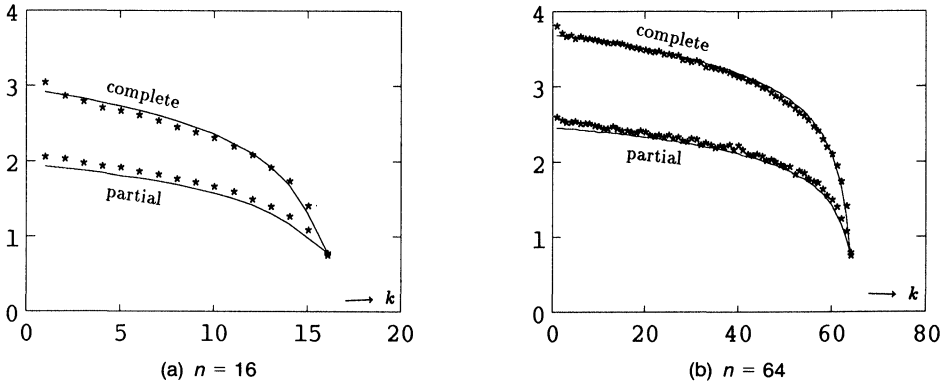
FIG. 3.1. *Average ratios* π_k / σ_k *of pivots to elements: observed* (∗) *and predicted* (———).

Except where otherwise indicated, the experiments reported in the remainder of this paper are based on matrices A with elements from the standard normal distribution.

3. Pivots. Even without knowing σ_k, the standard deviation of the elements at the kth step of elimination, we can predict π_k / σ_k, the size of the average pivot relative to σ_k. The pivot element $u_{kk} = \hat{a}_{kk}^{(k)}$ is the largest in absolute value among m contestants, where

$$(3.1) \qquad m = \begin{cases} n+1-k & \text{(partial pivoting)}, \\ (n+1-k)^2 & \text{(complete pivoting)}. \end{cases}$$

If the elements are normally distributed with standard deviation σ_k, the distribution of the pivots is a standard result from the field of the statistics of extreme values, going back to Tippett and Fisher in the 1920's [13], [17]. Let $W(m)$ (the "winner function") be defined as the mode of the distribution of the largest absolute value among m numbers taken from a normal distribution of mean 0, variance 1.[3] From equations 4.2.3(11, 15) of Gumbel [17], $W(m)$ is asymptotic to

$$(3.2) \qquad \alpha := \sqrt{2 \log (m\sqrt{2/\pi})}$$

as $m \to \infty$, and a more accurate estimate is[4]

$$(3.3) \qquad W(m) \approx \alpha \left(1 - \frac{2 \log \alpha}{1 + \alpha^2} \right)^{1/2} + O\left(\frac{1}{\log m} \right).$$

(We define $W(1) = \sqrt{2/\pi}$, the expected absolute value of a single normal variate.) Thus our model of Gaussian elimination makes the prediction

$$(3.4) \qquad \pi_k \approx \sigma_k W(m).$$

---

[3] We have chosen to work with the mode (the most frequent value) rather than the mean, although the two are asymptotic as $m \to \infty$. The reason is that the extreme value distribution is far from symmetric: for practical values of m the mode is several percent smaller than the median, which is several percent smaller than the mean. We shall be dividing by $W(m)$ to compute multipliers in the next section, and the mode is a convenient statistic that is relatively insensitive to this inversion.

[4] Gumbel has $m/2$ instead of m in (3.2), since he is concerned with the largest element in *signed* magnitude.

In the calculations and plots to follow we shall assume that each pivot element is exactly equal to $\pm\sigma_k W(m)$, although in actuality it is, of course, a random variable.

Figure 3.1 provides experimental confirmation of this prediction for matrices with normally distributed elements. For $n = 16$ and 64 and both partial and complete pivoting, the figure compares experimentally obtained ratios π_k/σ_k with the prediction $W(m)$, as a function of k. The agreement is not perfect, but it is quite good. Similar agreement is obtained with most of the other matrix distributions listed in § 1.

4. Multipliers. The previous two sections lead readily to a prediction of the distribution of multipliers at step k and of their standard deviation μ_k. First the pivot element $\hat{a}_{kk}^{(k)}$ is chosen and the rows and possibly columns permuted accordingly; we have assumed $\hat{a}_{kk}^{(k)}$ is equal to $\pm\sigma_k W(m)$. The multipliers are then the numbers $\hat{a}_{ik}^{(k)}/\hat{a}_{kk}^{(k)}$, and what we know about $\hat{a}_{ik}^{(k)}$ is that it comes from the normal distribution of mean 0, standard deviation σ_k, except with the tails beyond $\pm\sigma_k W(m)$ deleted and the total probability renormalized to compensate. That distribution has probability density function

$$\text{p.d.f.}\,(\hat{a}_{ik}^{(k)}) \approx \begin{cases} \dfrac{(1/\sqrt{2\pi})e^{-(x/\sigma_k)^2/2}}{\sigma_k \operatorname{erf}(W(m)/\sqrt{2})} & \text{for } |x| \leqq \sigma_k W(m), \\[2ex] 0 & \text{for } |x| > \sigma_k W(m), \end{cases}$$

where erf is the error function. (By (3.3) and standard estimates we have $1 - \operatorname{erf}(W(m)/\sqrt{2}) \sim W(m)/2m$ as $m \to \infty$.) The division by $\hat{a}_{kk}^{(k)} = \pm\sigma_k W(m)$ now gives the following approximate density distribution for the multipliers:

$$(4.1) \qquad \text{p.d.f.}\left(\frac{\hat{a}_{ik}^{(k)}}{\hat{a}_{kk}^{(k)}}\right) \approx \begin{cases} \dfrac{(1/\sqrt{2\pi})W(m)e^{-(xW(m))^2/2}}{\operatorname{erf}(W(m)/\sqrt{2})} & \text{for } |x| \leqq 1, \\[2ex] 0 & \text{for } |x| > 1. \end{cases}$$

This is a rather remarkable formula, for it asserts that the multiplier distribution is independent of everything except the length of the column on and below the diagonal, $n + 1 - k$ (which determines m by (3.1)). From (4.1), by integration by parts, we can further derive an approximation for the variance of the multipliers at step k:

$$(4.2) \qquad \mu_k^2 \approx \frac{1}{W(m)^2}\left(1 - \frac{\sqrt{2/\pi}\,W(m)e^{-W(m)^2/2}}{\operatorname{erf}(W(m)/\sqrt{2})}\right)$$

$$(4.3) \qquad \sim \frac{1}{2\log(m\sqrt{2/\pi})}.$$

For experimental confirmation of these predictions, Fig. 4.1 is patterned after Fig. 2.1, but shows both $n = 16$ and $n = 128$. This time the solid reference curves are not simply rescaled Gaussians, but the predicted multiplier distributions (4.1). The agreement with predictions is excellent for partial pivoting and reasonably good for complete pivoting. Note that the multipliers are smaller for large n and for complete pivoting.

5. Dependence on k. Sections 2–4 have proposed models of the behavior of elements, pivots, and multipliers at each step k, but did not consider how the scale of these quantities—σ_k and π_k—changes with k. We turn now to this question.

The first half of step k is the interchange of rows and possibly columns $A^{(k)} \to \hat{A}^{(k)}$, which moves some large element $a_{ij}^{(k)}$ to the pivot position $\hat{a}_{kk}^{(k)}$. In the case of complete pivoting, this repeated removal of the largest element from the submatrix $k \leqq i, j \leqq n$ has a pronounced retarding effect on element growth, especially toward the end of the

(a) partial pivoting, $n = 16$

(b) complete pivoting, $n = 16$

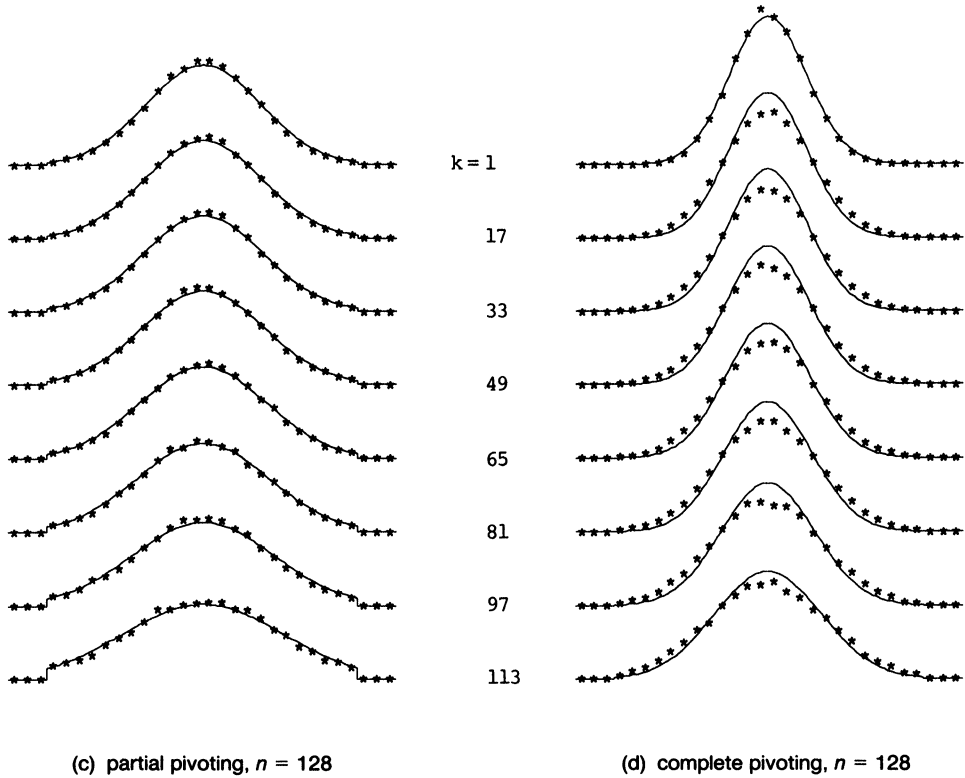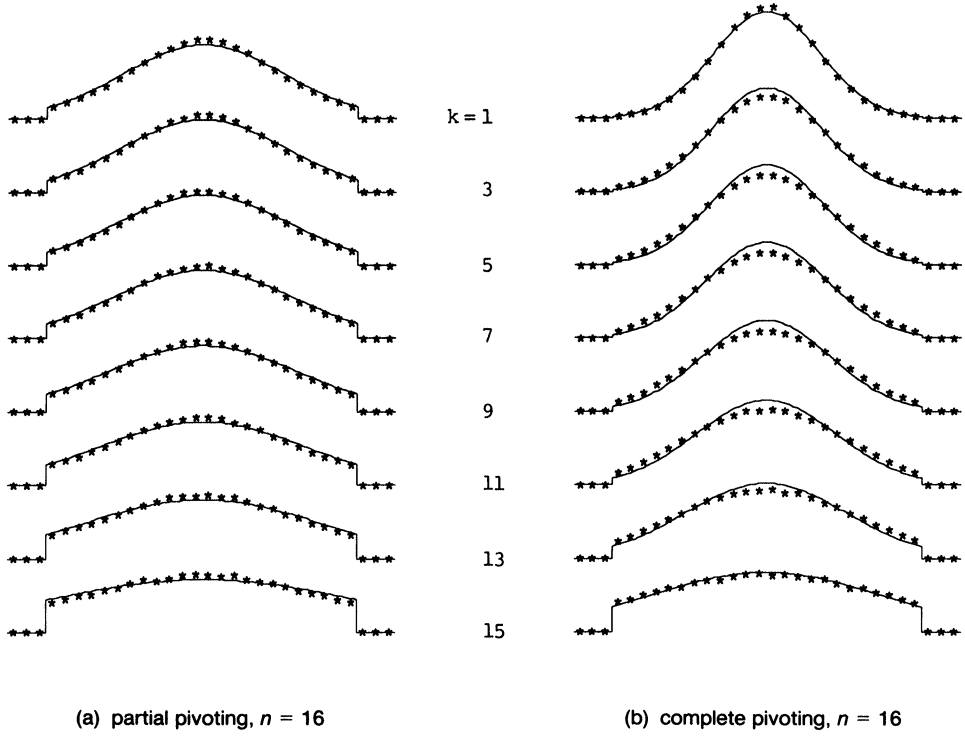(c) partial pivoting, $n = 128$

(d) complete pivoting, $n = 128$

FIG. 4.1. *Distributions of multipliers* $\hat{a}_{ik}^{(k)} / \hat{a}_{kk}^{(k)}$: *observed* (∗) *and predicted* (——). *The cutoff points are* ± 1.
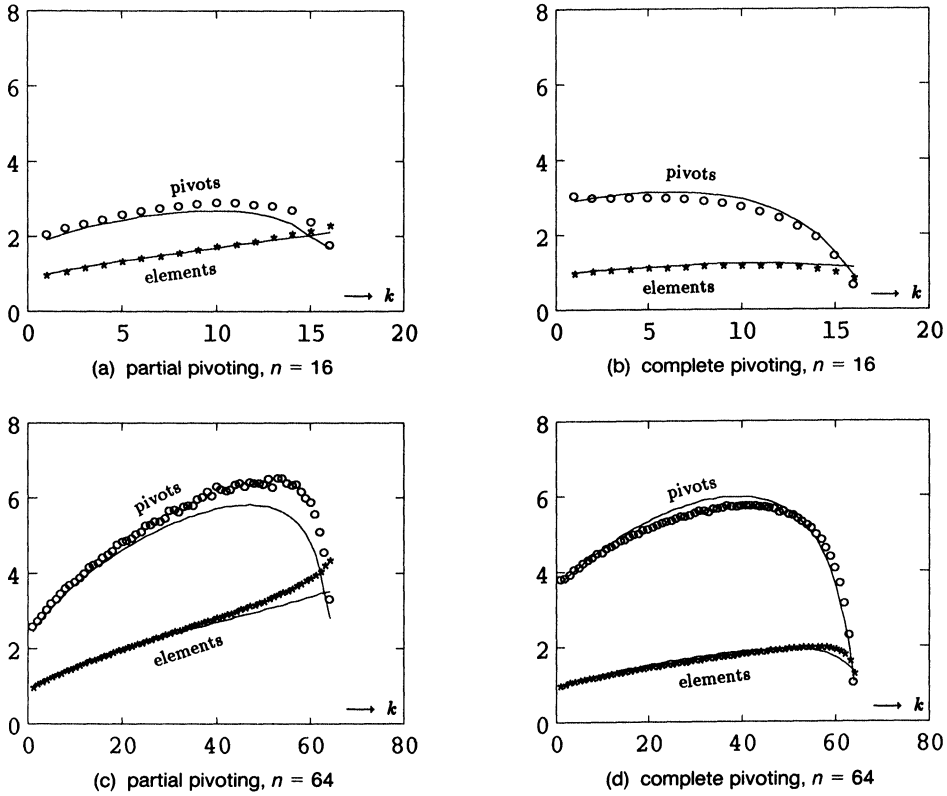
FIG. 5.1. *Average elements σ_k and pivots π_k: observed* (*, $\bigcirc$) *and predicted* (——).

elimination. At step k the elements $\alpha_{ij}^{(k)}$ with $k \leqq i, j \leqq n$ have variance σ_k^2; thus the expected sum of their squares is $m\sigma_k^2$ with $m = (n + 1 - k)^2$. When the pivot $\pm\sigma_k W(m)$ is removed from this collection, the remaining $m - 1$ elements have expected sum of squares $(m - W(m)^2)\sigma_k^2$. If $\hat{\sigma}_k$ denotes the standard deviation of the elements $\hat{a}_{ij}^{(k)}$ for $k < i, j \leqq n$, we conclude

$$(5.1) \qquad \hat{\sigma}_k^2 = \begin{cases} \sigma_k^2 & \text{(partial pivoting)}, \\ \sigma_k^2\left(\dfrac{m - W(m)^2}{m - 1}\right) & \text{(complete pivoting)}. \end{cases}$$

The downturn resulting from this mechanism is clearly apparent in Figs. 5.1(b), 5.1(d).[5]
The second half of step k is the elimination calculation $\hat{A}^{(k)} \to A^{(k+1)}$,

$$(5.2) \qquad a_{ij}^{(k+1)} := \hat{a}_{ij}^{(k)} - \frac{\hat{a}_{ik}^{(k)}}{\hat{a}_{kk}^{(k)}}\hat{a}_{kj}^{(k)} \qquad (k < i, j \leqq n).$$

By assumption $\hat{a}_{ij}^{(k)}$ and $\hat{a}_{kj}^{(k)}$ have variance $\hat{\sigma}_k^2$, and (4.2) gives a prediction of the vari-

---

[5] The growth-retarding mechanism just described is analogous to the cooling that occurs in evaporation of a liquid, in which the most energetic molecules escape from the surface, leaving those that remain behind a little less energetic on average. Further analogies can also be found between Gaussian elimination and statistical mechanics.

ance of $\hat{a}_{ik}^{(k)}/\hat{a}_{kk}^{(k)}$, denoted by μ_k^2. If all of these quantities were truly independent, (5.2) would imply that σ_{k+1}^2 was related to $\hat{\sigma}_k^2$ by

$$(5.3) \qquad\qquad \sigma_{k+1}^2 = \hat{\sigma}_k^2 + \mu_k^2 \hat{\sigma}_k^2,$$

thus completing our model of Gaussian elimination. But this formula is utterly inaccurate: it leads to a prediction for partial pivoting of nearly exponential growth,

$$(5.4) \qquad\qquad \frac{\sigma_n}{\sigma_1} \approx e^{n/(4 \log n)},$$

which fails to match experiments except for $n \approx 1$. Equation (5.4) is derived by iterating (5.1) and (5.3),

$$\frac{\sigma_n^2}{\sigma_1^2} = \prod_{k=1}^{n-1} (1 + \mu_k^2),$$

and then taking the logarithm and using (4.3) to obtain

$$\log \frac{\sigma_n}{\sigma_1} = \frac{1}{2} \sum_{k=1}^{n-1} \log (1 + \mu_k^2) \sim \frac{1}{2} \sum_{k=1}^{n-1} \mu_k^2 = \frac{1}{2} \sum_{m=2}^{n} \frac{1}{2 \log m} \sim \frac{n}{4 \log n}.$$

We have now reached the point where hypothesis (2) mentioned in the Introduction has failed us; it is time to replace it by some quantitative version of (2').

We have found that the following simple assumption is surprisingly accurate, at least until the last few steps of elimination: the variances σ_k^2 accumulate additively rather than multiplicatively according to the formula

$$(5.5) \qquad\qquad \sigma_{k+1}^2 = \hat{\sigma}_k^2 + \mu_k^2 \hat{\sigma}_1^2.$$

We do not have a rigorous justification of why (5.5) is an appropriate replacement for (5.3), but here is a heuristic one. Equation (5.5) amounts to the statement that the operations performed in Gaussian elimination do not compound, from the point of view of growth factors; it is as if the kth elimination step were applied to the original matrix $A = A^{(1)}$ rather than to $A^{(k)}$. Why should this be? Our best answer is to describe the following mechanism, which suggests that the growth introduced at one elimination step tends not to contribute to further growth at later steps. At step k, the correction subtracted from $\hat{A}^{(k)}$ by (5.2) is a rank-1 matrix. Taking the extreme, suppose this correction happened to be much larger than the elements it was being added to. Then the new matrix $A^{(k+1)}$ would be close to a matrix of rank one in its lower-right subsquare $k + 1 \leq i$, $j \leq n$. Consequently, the large numbers just introduced would vanish at step $k + 1$.

This argument is certainly not complete, nor precise enough to distinguish (5.5) from various other possible modifications of (5.3). But we believe the feedback mechanism it describes is essential to the stability of Gaussian elimination: large growth makes the remaining matrix close to a matrix of low rank, which in turn inhibits large growth. Note that in keeping with the distinction in the Introduction between (2) and (2'), the low-rank property would be destroyed if the signs of the correction matrix were randomized. Experiments with a "lobotomized Gaussian elimination" algorithm of this kind confirm that (5.3) and (5.4) then become accurate. See § 8 for the occurrence of this instability phenomenon in a computation of practical interest based on "parallel pivoting."

Equation (5.5) completes our model of average element and pivot growth as a function of k, which consists in its entirety of equations (3.1), (3.2), (3.3), (3.4), (4.2), (5.1), and (5.5).

Figure 5.1 compares predicted element and pivot sizes with experimental observations. For all our inexact assumptions, the agreement is remarkably good except at the very end of the elimination. We emphasize that all of the solid curves in Fig. 5.1 represent predictions from general principles, dependent on no adjustable parameters.

Figure 5.2 returns once again to nonnormal distributions of elements. The first two plots repeat Figs. 5.1(c), 5.1(d) for matrices with elements from the $\{0, 1\}$ distribution. Of the nonorthogonal distributions we have considered, this is as far from having normally distributed elements as any, but even so, the figure reveals that our element and pivot predictions are roughly valid after $k \approx 8$. Similar but generally better agreement is observed in most of the other cases. Figures 5.2(c), 5.2(d), however, repeat the same experiments for random orthogonal matrices, and the results are very different. In keeping with (1.2) and (1.3), we see that the geometric mean of the pivots has increased by a factor of approximately $\sqrt{e}$. Hypothesis (5.5) has failed in this case, although the growth is still far less rapid than (5.3) would predict.

6. Growth factors.
At last we are prepared to turn to the problem of average growth factors. We will begin with experiments, and then see how these can be related to the statistical model of the last four sections.

Figure 6.1 summarizes various theoretical and experimental results concerning the average growth factor $\langle \rho \rangle$ of (0.2), all plotted on a log-log scale. The highest curve shows the worst-case bound $\rho \leqq 2^{n-1}$ for partial pivoting, which we know by (0.3) is sharp.

(a) partial pivoting, $\{0, 1\}$ matrices

(b) complete pivoting, $\{0, 1\}$ matrices

(c) partial pivoting, orthogonal matrices
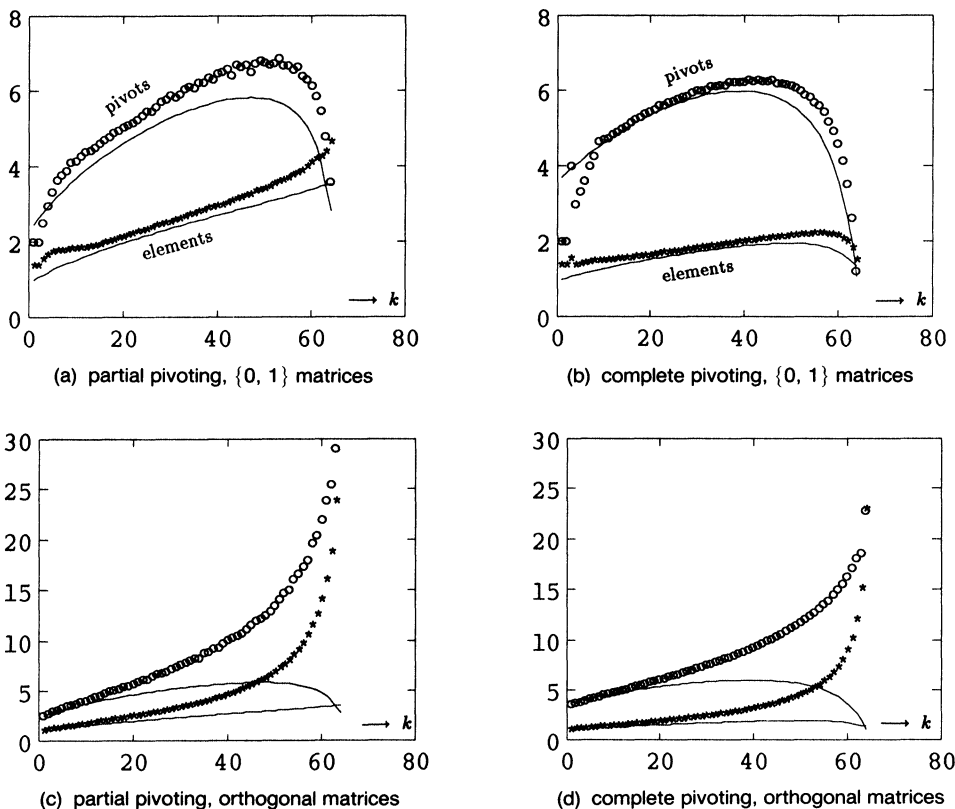
(d) complete pivoting, orthogonal matrices

FIG. 5.2. *Repetition of Figs.* 5.1(c), 5.1(d) *for random* $\{0, 1\}$ *and orthogonal matrices,* $n = 64$.
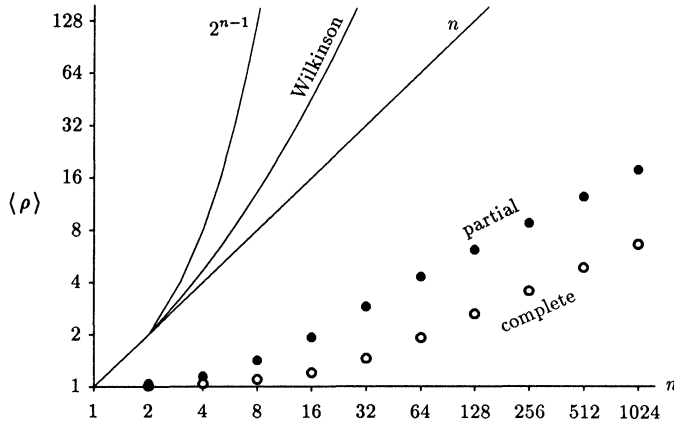
FIG. 6.1. *Average growth factors* $\langle \rho \rangle$. *The solid curves represent various theoretical worst-case bounds.*

The next curve shows the best available worst-case bound for complete pivoting, $\rho \leqq$ $\sqrt{n}\,(2^{1}3^{1/2}\cdots n^{1/(n-1)})^{1/2} \sim Cn^{1/2+1/4\log n}$, due to Wilkinson [32], which is known to be not sharp.[6] The straight line shows the bound $\rho \leqq n$ that was conjectured by Wilkinson for real matrices with complete pivoting [33, p. 213], which has never been proved except for $n \leqq 5$ [4], [7], [18]. Below these curves, we have plotted two sets of experimental values of $\langle \rho \rangle$ based on matrices with random elements from the standard normal distribution.[7] It is evident that the average growth factors for both partial and complete pivoting grow sublinearly with n and lie well below all of the worst-case bounds.[8]

The pattern in these data can be made more apparent if we modify the definition of ρ. Rather than dividing by the maximum element of A, let us divide by the standard deviation σ_A of the initial element distribution,

$$(6.1) \qquad \tilde{\rho} = \frac{\max_{i,j,k}|a_{ij}^{(k)}|}{\sigma_A}.$$

(σ_A is not the same as σ_1, unless the elements of A have mean 0: the former is a true standard deviation, while the latter is defined in (1.4) relative to 0.) For matrices with elements from a uniform distribution, this modification will increase $\langle \rho \rangle$ by a constant factor, whereas for matrices with normally distributed elements, the factor is approximately $W(n^2) = O(\sqrt{\log n})$. Figure 6.2 repeats the experimental data of Fig. 6.1, but showing $\langle \tilde{\rho} \rangle$ instead of $\langle \rho \rangle$. The data points lie strangely close to two straight lines:

$$(6.2) \qquad \text{partial pivoting:} \langle \tilde{\rho} \rangle \approx n^{2/3}, \qquad \text{complete pivoting:} \langle \tilde{\rho} \rangle \approx n^{1/2}.$$

---

[6] The proof of Wilkinson's bound is a reasonably straightforward recursive application of (1.1).

[7] In Fig. 6.1, the last two data points in each sequence ($n = 512$ and $n = 1024$) are fabricated by extrapolation; in our computer experiments we neglected to measure these numbers beyond $n = 256$. All the data in Fig. 6.2 are genuine, however, and since the two figures are nearly equivalent, the extrapolations are unlikely to be far wrong, so we have included the extra points in Fig. 6.1 to make the comparison clearer.

[8] Goodman and Moler [8], [16] report that in the LU factorization of 10,000 random matrices of dimensions $10 \leqq n \leqq 50$ drawn from four different distributions, the largest growth factor encountered was $\rho \approx 23$. MacLeod [21], who increased n to 100, observed a maximum growth factor $\rho \approx 35$. We regret to say that in our own experiments, we were so focused on average-case behavior that we neglected to measure the largest growth factor.
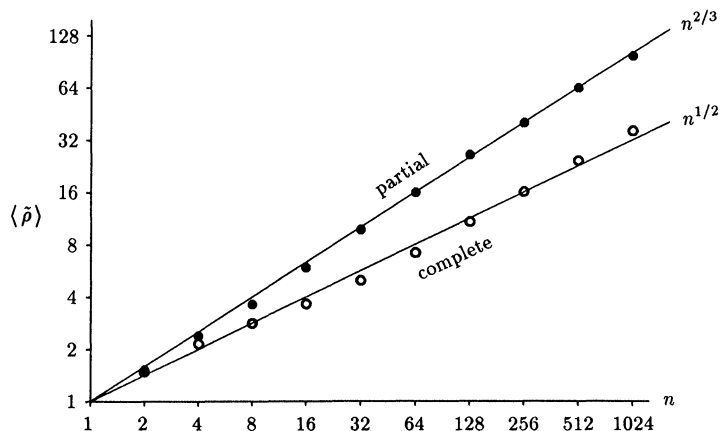
FIG. 6.2. *Average normalized growth factors* $\langle \tilde{\rho} \rangle$. *The solid lines are purely empirical*.

In this observation there is not even a constant factor to worry about—the fractional power of n is multiplied by 1! Despite these surprisingly close agreements, however—especially for partial pivoting and $n^{2/3}$—we do not claim that the approximations (6.2) are asymptotically valid as $n \to \infty$.

The data from Figs. 6.1 and 6.2 are recorded in Table 6.1. The sampling errors in this and subsequent tables probably range from about 1 percent for small n to more like 5 percent for large n.

Average growth factors change remarkably little when we turn to other distributions of matrices. Tables 6.2 and 6.3 list observed growth factors $\langle \tilde{\rho} \rangle$ for Gaussian elimination with partial and complete pivoting for the eight distributions of matrices listed in § 1. For larger n, except in the case of random orthogonal matrices, the numbers are nearly independent of the matrix distribution—so much so that a plot would be uninformative. Thus (6.2) appears to continue to hold with the constant factor 1, independently of the matrix distribution—a remarkable degree of a regularity that would have been obscured had we not normalized by σ_A in (6.1).

TABLE 6.1
Average growth factors $\langle \rho \rangle$ *and* $\langle \tilde{\rho} \rangle$.

	$\langle \rho \rangle$		$\langle \tilde{\rho} \rangle$	
n	Partial pivoting	Complete pivoting	Partial pivoting	Complete pivoting
2	1.04	1.01	1.52	1.48
4	1.15	1.04	2.39	2.15
8	1.42	1.10	3.63	2.82
16	1.93	1.20	5.92	3.64
32	2.89	1.45	9.77	4.97
64	4.31	1.91	15.9	7.17
128	6.14	2.62	26.3	10.8
256	8.74	3.56	40.0	16.1
512	–	–	63.7	24.3
1024	–	–	97.3	36.1

TABLE 6.2
Average growth factors $\langle \tilde{\rho} \rangle$ for matrices from various distributions, partial pivoting.

n	Normal	$[-1, 1]$	$[0, 1]$	$\{-1, 1\}$	$\{0, 1\}$	Symm.	Toep.	Orth.
2	1.52	1.48	2.77	1.50	1.87	1.40	1.41	1.59
4	2.39	2.23	3.33	2.02	2.13	2.26	2.12	2.78
8	3.63	3.50	4.06	3.61	4.18	3.60	3.37	5.20
16	5.92	5.85	6.17	6.63	6.86	6.06	5.83	10.3
32	9.77	9.67	9.85	9.86	9.91	10.0	10.3	20.7
64	15.9	15.5	16.3	15.7	16.6	16.5	18.3	42.5
128	26.3	24.6	25.2	24.9	25.9	25.6	30.1	81.4

The observation that orthogonal matrices fare worse in Gaussian elimination is not new, but goes back at least to Wilkinson (cf. Fig. 5.2). For example, the extreme case of element growth under complete pivoting in any example yet devised is achieved by Hadamard matrices—multiples of orthogonal matrices with elements ± 1—for which $\rho \geq |u_{nn}| = n$ (proof by Cramer's rule [4]). See [7] and [18] for more on this subject.

It remains to relate these observations to our statistical model of the past four sections.

To begin the discussion, let us for the first time take a look at the effect of Gaussian elimination on individual matrices rather than just averages. Figures 6.3(a), 6.3(b) show pivots $|u_{kk}|$ from the factorization of a single matrix with $n = 64$ compared with the prediction π_k of Fig. 5.1. Figs. 6.3(c), 6.3(d) superimpose the pivots $|u_{kk}|$ from 25 such matrices. These four plots show vividly that our average-case predictions have definite relevance even to an individual matrix, for although $|u_{kk}|$ oscillates considerably, its overall trend follows the predicted average. They also show that the extent of the oscillation is much greater for partial than complete pivoting.

The growth factor $\tilde{\rho}$ will, in general, be larger than $\max_k \pi_k$, since $\tilde{\rho}$ is a maximum while $\max_k \pi_k$ is a maximum of an average. Figure 6.3 suggests that for complete pivoting the excess is typically modest, whereas for partial pivoting it may be quite substantial. These considerations explain how it is possible that the average size of π_k can be insensitive to the type of pivoting (in keeping with (1.3)) while the growth factor still varies significantly.

We can estimate $\langle \tilde{\rho} \rangle$ as follows. Figures 5.1 and 6.3 suggest that very roughly, the last $n/2$ steps of Gaussian elimination are equally likely to contribute the largest element $a_{ij}^{(k)}$. (The crudeness of this estimate is not so important, since $W(m)$ depends very weakly on m.) These final $n/2$ steps generate a total of $\sim n^3/24$ new elements $a_{ij}^{(k)}$. Therefore we estimate

$$(6.3) \qquad \langle \tilde{\rho} \rangle \approx W(n^3/24) \max_k \sigma_k,$$

TABLE 6.3
Average growth factors $\langle \tilde{\rho} \rangle$ for matrices from various distributions, complete pivoting.

n	Normal	$[-1, 1]$	$[0, 1]$	$\{-1, 1\}$	$\{0, 1\}$	Symm.	Toep.	Orth.
2	1.48	1.42	2.77	1.50	1.87	1.38	1.39	1.59
4	2.15	1.98	3.27	2.02	2.14	2.06	1.97	2.63
8	2.82	2.75	3.50	3.40	3.76	2.83	2.74	4.30
16	3.64	3.70	4.13	4.10	4.15	3.73	3.84	7.26
32	4.97	5.11	5.34	5.48	5.55	5.10	5.72	13.0
64	7.17	7.40	7.49	7.73	7.88	7.34	8.94	23.3
128	10.8	11.0	11.2	11.3	11.4	11.0	13.9	43.3

(a) partial pivoting, one matrix

(b) complete pivoting, one matrix

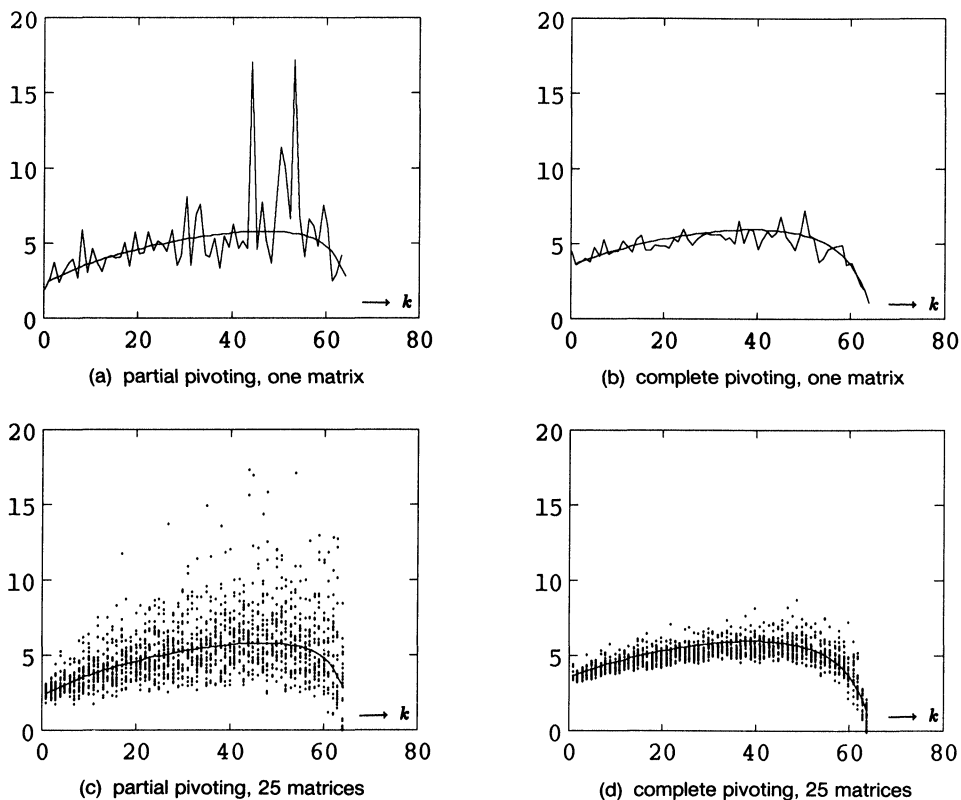(c) partial pivoting, 25 matrices

(d) complete pivoting, 25 matrices

FIG. 6.3. *Pivots* $|u_{kk}|$ *for matrices from the standard normal distribution*, $n = 64$. *The solid lines represent the same predictions as in Figs.* 5.1(c), 5.1(d).

where σ_k is the predicted value derived in (5.5). Figure 6.4 compares this prediction with the lines $n^{1/2}$ and $n^{2/3}$ of Fig. 6.2. The agreement is not bad! The predictions for partial pivoting are somewhat too low, however, which reflects the fact that our predicted values of σ_k were too low toward the end of elimination (Fig. 5.1).
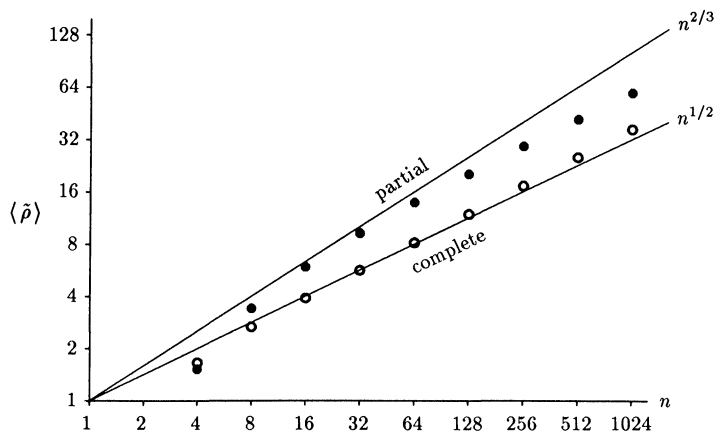

FIG. 6.4. *Predicted average growth factors* $\langle \tilde{\rho} \rangle$. *The solid lines are for comparison with Fig.* 6.2.

What about asymptotics as $n \to \infty$? We must be cautious here, for as is well known, extreme value statistics for <u>normal</u> distributions are approached painfully slowly. <u>But as</u> $n \to \infty$, $W(n^3/24) = O(\sqrt{\log n})$, by (3.2) and (3.3), and $\max_k \sigma_k = O(\sqrt{n}/\sqrt{\log n})$, by (4.2) and (5.5). Thus the natural conjecture appears to be

$$(6.4) \qquad\qquad \langle \tilde{\rho} \rangle = O(\sqrt{n}) \quad \text{as } n \to \infty?$$

for both partial pivoting and complete pivoting, despite (6.2). This guess is tidy but hardly astonishing, in the light of (1.3).

7. Residuals. Our next set of experiments concerns the actual errors introduced by Gaussian elimination, and also by QR factorization, as measured by residuals computed in double precision. Let an $n \times n$ matrix A be factored in one of the following ways:

$A = PLU$ (Gaussian elimination with partial pivoting),
$A = P_1LUP_2$ (Gaussian elimination with complete pivoting),
$A = QR$ (QR factorization),
$A = QRP$ (QR factorization with column pivoting),

where L is unit lower triangular, U and R are upper triangular, Q is orthogonal, and P, P_1, and P_2 are permutation matrices. The QR factorizations are carried out by Householder reflections, and as is customary, the vector associated with these reflections is stored rather than an explicit matrix Q. Let $\bar{L}$, $\bar{U}$, $\bar{R}$, and so on denote the matrices obtained in floating-point arithmetic, and define the *residual* for Gaussian elimination with partial pivoting by $E = A - \bar{P}\bar{L}\bar{U}$, and similarly for the other factorizations.

After a factorization has been carried out, we measure the size of E by its maximum element normalized by σ_A and also by machine epsilon:

$$(7.1) \qquad\qquad E_{\max} = \frac{\max_{i,j} |e_{ij}|}{\sigma_A \varepsilon}.$$

At the end of a series of N factorizations, we compute the average $\langle E_{\max} \rangle$ as usual.

Figure 7.1 begins with the most important case of the standard normal distribution, showing computed quantities $\langle E_{\max} \rangle$ as a function of n for Gaussian elimination with
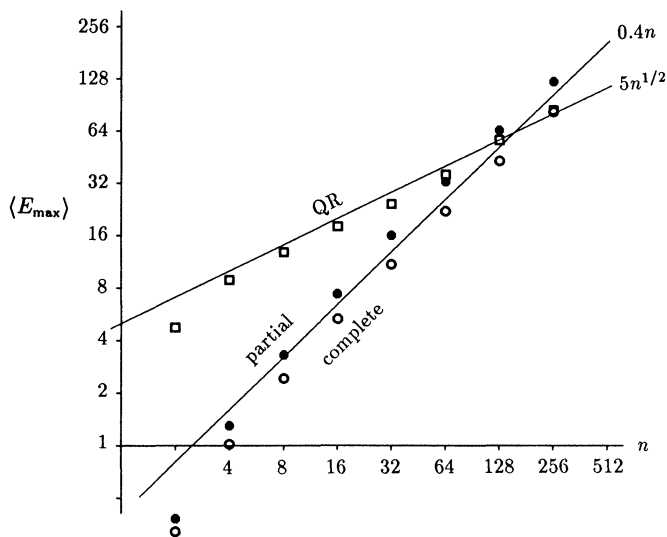

FIG. 7.1. *Average maximum residual element* $\langle E_{\max} \rangle$. *The solid lines are empirical*.

partial and complete pivoting and for QR factorization without pivoting. The data from the figure, together with corresponding numbers for QR factorization with column pivoting, are listed in Table 7.1. They suggest, somewhat surprisingly in the light of the last section, that the maximum residual elements in partial and complete pivoting differ only by a constant factor: both satisfy $\langle E_{max} \rangle \approx Cn$. They also suggest that QR factorization is asymptotically more stable than either form of Gaussian elimination, with $\langle E_{max} \rangle \approx Cn^{1/2}$. Thus it would appear that in practice, there may be little difference between partial and complete pivoting in Gaussian elimination, but not because both are entirely stable; apparently both suffer mildly from the unstable effects of pivot growth. The same conclusion is obtained if one measures the residual matrix E in ways other than by its maximum element.

TABLE 7.1

Average maximum residual element $\langle E_{max} \rangle$.

	Gaussian elim.		QR factorization	
n	Partial pivoting	Complete pivoting	No pivoting	Column pivoting
2	0.43	0.33	4.75	5.10
4	1.47	1.13	8.87	8.42
8	3.64	2.68	12.8	12.5
16	8.13	5.71	18.0	16.7
32	17.2	11.4	24.2	24.9
64	36.0	22.6	35.8	35.6
128	73.0	44.6	56.4	54.6
256	134.	86.2	84.3	–

TABLE 7.2

$\langle E_{max} \rangle$—*various matrix distributions, partial pivoting.*

n	Normal	$[-1, 1]$	$[0, 1]$	$\{-1, 1\}$	$\{0, 1\}$	Symm.	Toep.	Orth.
2	0.43	0.47	0.75	0.00	0.00	0.46	0.46	0.58
4	1.47	1.52	1.92	0.00	0.00	1.45	1.45	2.01
8	3.64	3.67	4.24	0.12	0.27	3.71	3.63	5.08
16	8.13	8.03	8.80	5.62	6.32	8.33	8.35	12.2
32	17.2	17.3	17.5	16.9	17.6	17.2	17.9	28.2
64	36.0	33.6	35.1	34.5	35.5	36.3	37.6	68.7
128	73.0	71.6	72.1	69.1	71.5	72.2	74.1	149.

TABLE 7.3

$\langle E_{max} \rangle$—*various matrix distributions, complete pivoting.*

n	Normal	$[-1, 1]$	$[0, 1]$	$\{-1, 1\}$	$\{0, 1\}$	Symm.	Toep.	Orth.
2	0.33	0.42	0.68	0.00	0.00	0.40	0.41	0.58
4	1.13	1.32	1.75	0.00	0.00	1.19	1.27	1.78
8	2.68	2.98	3.61	0.28	0.64	2.79	3.00	4.39
16	5.71	6.03	6.82	4.86	5.34	5.85	6.25	9.25
32	11.4	11.5	12.4	11.5	11.9	11.4	12.7	19.9
64	22.6	22.1	24.2	23.9	24.4	22.9	26.2	42.0
128	44.6	46.5	43.3	44.8	46.8	43.5	50.9	88.2

As with the growth factors of the last section, our observations concerning residuals are closely duplicated for random matrices from many other distributions. Tables 7.2 and 7.3 reveal that once again, only orthogonal matrices among the classes we have examined behave much differently.

8. Alternative pivoting strategies. The previous sections have examined "classical" Gaussian elimination with partial or complete pivoting, and concluded that these algorithms are highly stable on average. In this final section we shall look more superficially at three variants of Gaussian elimination based on alternative pivoting strategies: "threshold," "pairwise," and "parallel" pivoting. All of these variants are less stable than partial or complete pivoting, and the last turns out to be markedly unstable for large n even though the multipliers are all less than 1 in magnitude. Table 8.1 summarizes our conclusions, which are based on experiments with $n \leq 1024$. The most interesting observation is that as discussed in earlier sections, the stability of Gaussian elimination depends not only on the size of the multipliers, but also on whether the corrections introduced at each step are of low rank.

TABLE 8.1
Summary of experimental results for various pivoting strategies.

Pivoting strategy	Size of multipliers	Rank of corrections	Average-case stability
partial or complete	≤ 1	1	highly stable
threshold	$\leq \tau^{-1}$	1	reasonably stable for larger τ
pairwise	≤ 1	low	reasonably stable
parallel	≤ 1	$n/2$	unstable

We begin with *threshold pivoting*, a well-known idea that is discussed by various authors (for a discussion and references see [10]). The idea is to require only that

$$(8.1) \qquad |\hat{a}_{kk}^{(k)}| \geq \tau |\hat{a}_{ik}^{(k)}|, \qquad i > k,$$

where $\tau \in [0, 1]$ is a parameter. For $\tau = 1$ this is partial pivoting, and for $\tau = 0$ it is no pivoting at all; of course in practice τ is taken to be positive. The motivation behind threshold pivoting is that it allows for more than one row to be a candidate for the pivot row, and some other criterion, such as sparsity, can be used to make the choice. With this strategy, the multipliers are at most τ^{-1} and the growth factor satisfies $\rho \leq (1 + \tau^{-1})^{n-1}$. As with partial or complete pivoting, each step involves an elimination operation of rank 1.

Several authors have espoused ways to choose τ, with recommended choices being as low as 0.01 [29] or as high as 0.25 [5]. Duff [9], [10] reports an experiment with four sparse matrices and arrives at the interesting conclusion that $\tau = 0.1$ affords both good reduction of fill-in and loss of only one to two digits of accuracy in the solution, whereas smaller τ (0.01 or less) can be disastrous to accuracy and may actually increase the fill-in. To explain this counterintuitive observation, he notes that when τ is small the variance of elements becomes large, so that the number of elements that satisfy (8.1) becomes small.

We performed a brief series of experiments using dense matrices of dimensions $n \leq 128$, with independent elements from the standard normal distribution and with $\tau \in \{0.5, 0.25, 0.1, 10^{-2}, 10^{-4}, 10^{-8}\}$. The sample sizes were approximately as listed in § 1, and at each step the pivot row was simply taken to be the first candidate satisfying
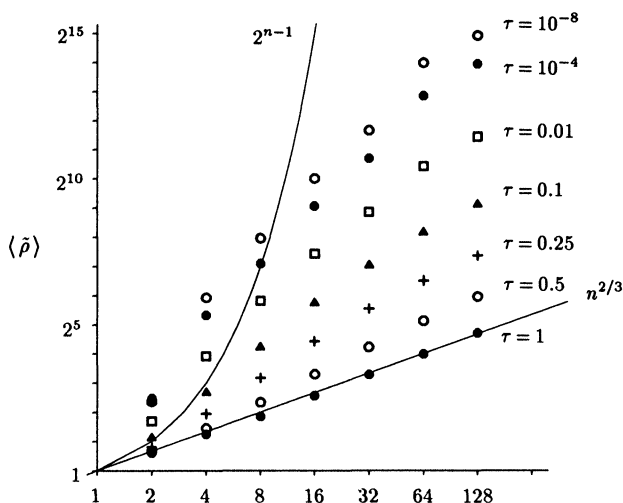
FIG. 8.1. *Average growth factors* $\langle \tilde{\rho} \rangle$ *for threshold pivoting with various thresholds* τ. *The solid lines are for comparison with earlier figures.*

(8.1). In Fig. 8.1, the observed average growth factor $\langle \tilde{\rho} \rangle$ is plotted against n for each of the values of τ, with the curves $n^{2/3}$ and 2^{n-1} shown for comparison. The numbers are listed in Table 8.2. These experiments support the conclusion that for larger values of τ, threshold pivoting is reasonably safe; the growth factors are nowhere near the worst-case bound 2^{n-1}. Of course, in applications involving sparse matrices the behavior may be different.

What about the limit $\tau = 0$—no pivoting? The data for $\tau = 10^{-8}$ in Fig. 8.1 are not much different from what would have been observed in the same experiment with $\tau = 0$, but there is an important mathematical difference nonetheless: although any single experiment will yield a finite result with probability 1, the expected growth factor is infinite in the absence of pivoting. (This is obvious; we need only consider the very first division $a_{21}^{(1)}/a_{11}^{(1)}$.) For a meaningful theory of the statistical behavior of Gaussian elimination without pivoting, we would have to employ a different measure of average-case growth such as $\exp(\langle \log \tilde{\rho} \rangle)$, as in the study of expected condition numbers [11], [26].

Another well-known variant of Gaussian elimination is *pairwise* or *neighbor pivoting*, in which only adjacent rows are interchanged or eliminated. Here is the algorithm. The scope of each control structure is indicated by indentation, and *row* (i) denotes the

TABLE 8.2
Average growth factors $\langle \tilde{\rho} \rangle$ *for threshold pivoting.*

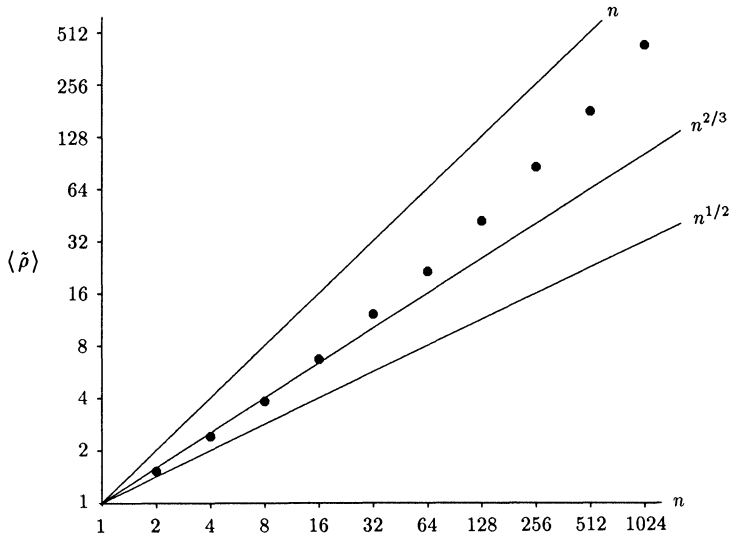| n	$\tau = 0.5$	$\tau = 0.25$	$\tau = 10^{-1}$	$\tau = 10^{-2}$	$\tau = 10^{-4}$	$\tau = 10^{-8}$
2	1.58	1.75	2.06	3.24	5.54	5.16
4	2.74	3.86	6.07	15.1	39.9	60.7
8	5.07	9.05	17.7	56.1	136.	249.
16	9.86	21.6	50.9	172.	535.	1030.
32	18.7	46.6	124.	464.	1660.	3210.
64	34.8	90.6	270.	1370.	7340.	16100.
128	62.2	164.	523.	2670.	15600.	30900.

FIG. 8.2. *Average growth factors* $\langle \tilde{\rho} \rangle$ *for pairwise pivoting. The solid lines are for comparison with earlier figures.*

elements $\{ a_{i,k}, a_{i,k+1}, \cdots, a_{i,n} \}$:

> **for** $k := 1$ **to** $n - 1$
>   **for** $i := n$ **to** $k + 1$ **step** -1
>   **if** $|a_{i,k}| > |a_{i-1,k}|$ **then**
>   exchange *row* (i) and *row* $(i - 1)$
>   *row* $(i) :=$ *row* $(i) - (a_{i,k}/a_{i-1,k})*$*row* $(i - 1)$

This algorithm is of interest for parallel computing because it avoids the search for a maximum required by partial and complete pivoting—a substantial bottleneck for parallel computations—yet keeps all multipliers less than 1 in magnitude. Sorensen has obtained a worst-case bound 4^{n-1} on the growth factor [27].

Our experiments involved matrices of dimensions $n = 2, 4, \cdots, 1024$ with independent elements from the standard normal distribution. The sample sizes were 10,000 for $n \leq 8$, 1000 for $16 \leq n \leq 128$, and 250 for $n \geq 512$. Figure 8.2 plots the observed

TABLE 8.3

Average growth factors $\langle \tilde{\rho} \rangle$ *for complete, partial, pairwise, and parallel pivoting.*

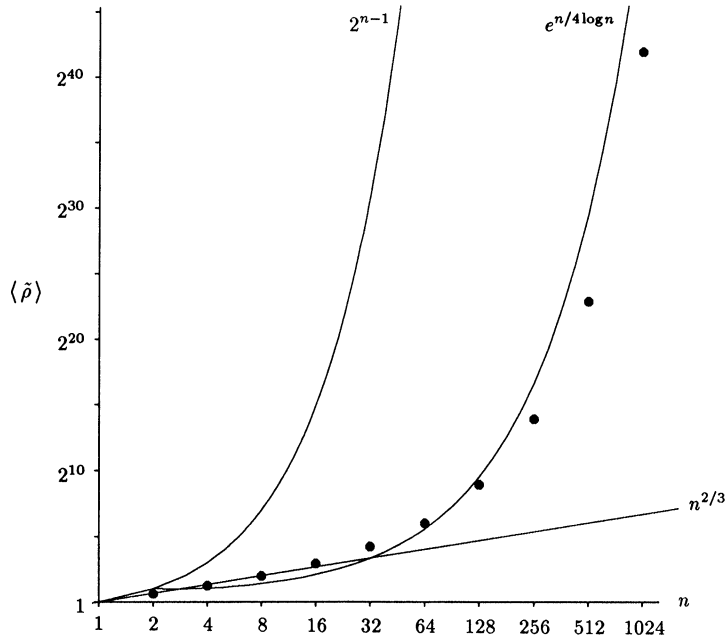| n	Complete pivoting	Partial pivoting	Pairwise pivoting	Parallel pivoting
2	1.48	1.52	1.52	1.55
4	2.15	2.39	2.41	2.41
8	2.82	3.63	3.83	3.94
16	3.64	5.92	6.68	7.67
32	4.97	9.77	12.1	18.7
64	7.17	15.9	21.3	64.1
128	10.8	26.3	41.8	483.
256	16.1	40.0	85.2	1.53×10^4
512	24.3	63.7	179.	7.72×10^6
1024	36.1	97.3	432.	4.2×10^{12}

FIG. 8.3. *Average growth factors* $\langle \tilde{\rho} \rangle$ *for parallel pivoting*.

average growth factors $\langle \tilde{\rho} \rangle$ as a function of n, and the numbers are listed in Table 8.3. Evidently pairwise pivoting is quite stable on average, though not as stable as partial or complete pivoting. We explain this by observing that first, the magnitudes of the multipliers are somewhat bigger but still much less than 1 on average; second, the corrections introduced at each step are still on average of low rank, although not of rank 1.

Finally, what we call *parallel pivoting* is a (nonstandard) variant of Gaussian elimination in which as many as $n/2$ elements are eliminated in parallel. For example, if $n = 2m$, we first eliminate $a_{j+m,1}$ for $j = 1, 2, \cdots, m$ by subtracting a multiple of row j from row $j + m$. These two rows are exchanged first if necessary in order to keep the multiplier no greater in magnitude than 1. Here is the algorithm:

```
for k := 1 to n − 1
    nelts := n − k                  /* number below diagonal */
    while nelts > 0
        n2 := (nelts + 1)/2         /* number to eliminate */
        nelts := nelts − n2         /* number remaining */
        for i := 1 to n2
            row1 := k + i − 1        /* pivot row */
            row2 := k + i + nelts
            if |a_row2,k| > |a_row1,k| then exchange row (row1) and row (row2)
            row (row2) := row (row2) − (a_row2,k/a_row1,k)*row (row1)
```

We tested this algorithm on matrices of orders $n = 2, 4, 8, \cdots, 1024$ from the standard normal distribution. Except for $n = 1024$, where only two matrices were factored, the sample sizes were at least 100 matrices. Figure 8.3 plots the observed growth factors $\langle \tilde{\rho} \rangle$ as a function of n, together with the curves $n^{2/3}$, $e^{n/(4 \log n)}$ (equation (5.4)), and 2^{n-1} for comparison. The data were listed already in Table 8.3 above. Clearly, this parallel pivoting strategy is unstable. We explain this by observing that first, the multipliers

are bigger than in standard Gaussian elimination (although still no greater than 1); second and more important, the corrections introduced at each step are of high rank, so that there are no favorable dependences among signs to retard growth. The rough agreement of the data with the curve $e^{n/(4 \log n)}$ suggests that perhaps this particular pivoting strategy, unlike partial or complete pivoting, approximately satisfies hypotheses (1) and (2) of the Introduction.

9. Conclusions. Is Gaussian elimination with partial pivoting stable on average? Everything we know on the subject indicates that the answer is emphatically yes, and that one needs no hypotheses beyond statistical properties to account for the success of this algorithm during nearly half a century of digital computation.

This paper has presented a model of the average-case behavior of Gaussian elimination supported by extensive experiments. Although no theorems have been proved, we believe that there is reasonably good evidence for the following conclusions. These statements are approximate, not exact, and they apply to the average case for many, but not all, distributions of matrices. Except where otherwise indicated, they apply to Gaussian elimination with either partial or complete pivoting.

(1) For QR factorization with or without column pivoting, the average maximum element of the residual matrix is $O(n^{1/2})$, whereas for Gaussian elimination it is $O(n)$. This comparison reveals that Gaussian elimination is mildly unstable, but the instability would only be detectable for very large matrix problems solved in low precision. For most practical purposes Gaussian elimination is highly stable on average. (§§ 6, 7)

(2) The statistical behavior of Gaussian elimination depends on the standard deviation of the initial matrix elements, but is otherwise insensitive to the matrix distribution. In particular, the statements below apply equally to random matrices with elements from normal, uniform, or discrete distributions, as well as to random symmetric and Toeplitz matrices (but not to random orthogonal matrices). (§§ 2–6)

(3) For $n \leq 1024$, the average growth factor (normalized by the standard deviation of the initial elements) is within a few percent of $n^{2/3}$ for partial pivoting and is approximately $n^{1/2}$ for complete pivoting. (§ 6)

(4) After the first few steps of Gaussian elimination, the remaining matrix elements are approximately normally distributed, regardless of whether they started out that way. (§ 2)

(5) The average magnitudes of pivots relative to elements at each step of elimination can be predicted by extreme value statistics. The distribution of multipliers at each step can then be predicted based on the pivot magnitudes. (§§ 3, 4)

(6) The signs of the elements and multipliers are not independent, and their dependence is essential to the stability of Gaussian elimination. It results from the fact that each step of elimination introduces a rank-1 correction to the remaining matrix, which provides a feedback mechanism that inhibits potential element growth and instability. (§§ 5, 8)

(7) This dependence of elements and multipliers can be modeled by hypothesizing that the corrections added at each step of elimination accumulate additively rather than multiplicatively. The resulting predictions of growth factors agree reasonably well with observations. (§§ 5, 6)

(8) By contrast, nonclassical variants of Gaussian elimination involving higher-rank elimination steps are sometimes markedly unstable, even though the multipliers are small. (§ 8)

Spencer, Gil Strang, and Dan Zwillinger. Special thanks go to Kapil Mathur and Anne Trefethen of Thinking Machines Corporation for catching an error in our computation of residuals in an earlier draft; these two are in the process of obtaining results for larger n with the aid of a Connection Machine. Trefethen would also like to acknowledge the generous support of Iain Duff and John Reid during a visit to the Harwell Atomic Energy Research Establishment (January 1987) and of Bill Morton during a visit to the Oxford Computing Laboratory (Summer 1987).

Note added in proof. This paper has focused mainly on averages, not distributions. Ultimately, however, it is the tail of the growth factor distribution that is of greatest concern. Experiments leave little doubt that the tail decays exponentially, and to illustrate, the following figure is a histogram of computed growth factors ρ in an experiment involving partial pivoting applied to $N = 20,000$ matrices of dimension $n = 32$ with normally distributed elements. Note the logarithmic scale. We leave it to others to determine how close such figures are to standard distributions such as the extreme value distribution. A. J. MacLeod has previously carried out experiments in this line [21], [34], and further statistical analysis of pivoting data is being carried out by D. Hoaglin in the Dept. of Statistics, Harvard University.
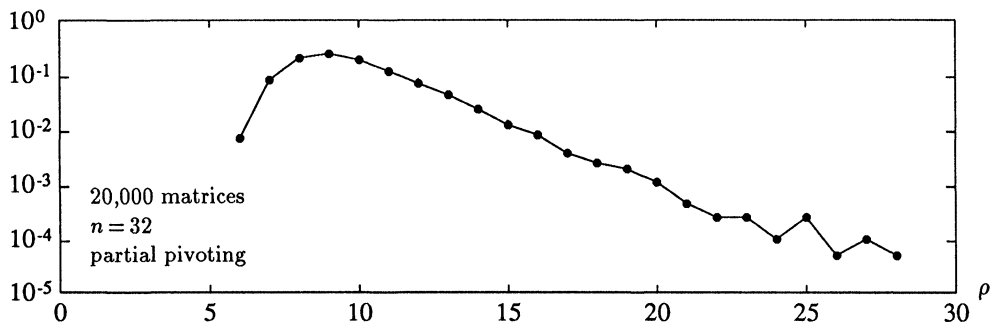

FIG. *Partial pivoting growth factor distribution based on* 20,000 *matrices of dimension* $n = 32$.

REFERENCES

[1] V. BARGMANN, D. MONTGOMERY, AND J. VON NEUMANN, *Solution of linear systems of high order*, Princeton, 1946. Reprinted in von Neumann's *Collected Works*, Vol. 5, A. H. Taub, ed., Pergamon, Elmsford, NY 1963.

[2] G. BIRKHOFF AND S. GULATI, *Isotropic distributions of test matrices*, J. Appl. Math. Phys. (ZAMP) 30 (1979), pp. 148–158.

[3] K. H. BORGWARDT, *The Simplex Method: A Probabilistic Analysis*, Springer-Verlag, Berlin, New York, 1987.

[4] C. W. CRYER, *Pivot size in Gaussian elimination*, Numer. Math., 12 (1968), pp. 335–345.

[5] A. R. CURTIS AND J. K. REID, *The solution of large sparse unsymmetric systems of linear equations*, J. Inst. Maths. Appl., 8 (1971), pp. 344–353.

[6] G. DANTZIG, *Linear Programming and Extensions*, Princeton University Press, Princeton, NJ, 1963.

[7] J. DAY AND B. PETERSON, *Growth in Gaussian elimination*, Amer. Math. Monthly, 95 (1988), pp. 489–513.

[8] J. J. DONGARRA, J. R. BUNCH, C. B. MOLER, AND G. W. STEWART, LINPACK *Users' Guide*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1979.

[9] I. S. DUFF, *Practical comparisons of codes for the solution of sparse linear systems*, in Sparse Matrix Proceedings 1978, I. S. Duff and G. W. Stewart, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1979.

[10] I. S. DUFF, A. M. ERISMAN, AND J. K. REID, *Direct Methods for Sparse Matrices*, Clarendon Press, Oxford, 1986.

[11] A. EDELMAN, *Eigenvalues and condition numbers of random matrices*, SIAM J. Matrix Anal. Appl., 9 (1988), pp. 543–560.

[12] ———, *Eigenvalues and condition numbers of random matrices*, Ph.D. thesis, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, 1989.

[13] R. A. FISHER AND L. H. C. TIPPETT, *Limiting forms of the frequency distribution of the largest or smallest member of a sample*, Proc. Cambridge Philos. Soc., 24 (1928), pp. 180–190.

[14] G. E. FORSYTHE AND C. B. MOLER, *Computer Solution of Linear Algebraic Systems*, Prentice-Hall, Englewood Cliffs, NJ, 1967.

[15a] V. L. GIRKO, *The central limit theorem for random determinants*, Theory Probab. Appl., 24 (1979), pp. 729–740.

[15b] ———, Theory Probab. Appl., 26 (1981), pp. 521–531.

[16] J. T. GOODMAN AND C. B. MOLER, *Three numerical experiments with Gaussian elimination*, Linpack Working Note #8, Applied Mathematics Division, Argonne National Laboratory, Argonne, IL, 1977.

[17] E. J. GUMBEL, *Statistics of Extremes*, Columbia University Press, New York, 1958.

[18] N. J. HIGHAM AND D. J. HIGHAM, *Large growth factors in Gaussian elimination with pivoting*, SIAM J. Matrix Anal. Appl., 10 (1989), pp. 155–164.

[19] H. HOTELLING, *Some new methods in matrix calculation*, Ann. Math. Statist., 14 (1943), pp. 1–34.

[20] V. KLEE AND G. MINTY, *How good is the simplex algorithm?*, in Inequalities III, O. Shisha, ed., Academic Press, New York, 1972.

[21] A. J. MACLEOD, *The distribution of the growth factor in Gaussian elimination with partial pivoting*, unpublished Tech. Report, Department of Mathematics and Statistics, Paisley College, Renfrewshire, Scotland.

[22] A. V. OPPENHEIM AND R. W. SCHAFER, *Digital Signal Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1975.

[23] W. H. PRESS, B. P. FLANNERY, S. A. TEUKOLSKY, AND W. T. VETTERLING, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, London, 1986.

[24] R. SHAMIR, *The efficiency of the simplex method: a survey*, Management Sci., 33 (1987), pp. 301–334.

[25] S. SMALE, *On the average number of steps in the simplex method of linear programming*, Math. Programming, 27 (1983), pp. 241–262.

[26] ———, *On the efficiency of algorithms of analysis*, Bull. Amer. Math. Soc., 13 (1985), pp. 87–121.

[27] D. C. SORENSEN, *Analysis of pairwise pivoting in Gaussian elimination*, IEEE Trans. Comput., 34 (1984), pp. 274–78.

[28] G. W. STEWART, *The efficient generation of random orthogonal matrices with an application to condition estimators*, SIAM J. Numer. Anal., 17 (1980), pp. 403–409.

[29] J. T. TOMLIN, *Pivoting for size and sparsity in linear programming inversion routines*, J. Inst. Maths. Appl., 10 (1972), pp. 289–295.

[30] A. M. TURING, *Rounding-off errors in matrix processes*, Quart. J. Mech. Appl. Math., 1 (1948), pp. 287–308.

[31a] J. VON NEUMANN AND H. H. GOLDSTINE, *Numerical inverting of matrices of high order*, Bull. Amer. Math. Soc., 53 (1947), pp.1021–1099; von Neumann's *Collected Works*, Vol. 5, A. H. Taub, ed., Pergamon, Elmsford, NY, 1963.

[31b] ———, *Numerical inverting of matrices of high order, Part II*, Proc. Amer. Math. Soc., 2 (1951), pp. 188–202; von Neumann's Collected Works, Vol. 5, A. H. Taub, ed., Pergamon Press, Elmsford, NY, 1963.

[32] J. H. WILKINSON, *Error analysis of direct methods of matrix inversion*, J. Assoc. Comput. Mach., 8 (1961), pp. 281–330.

[33] ———, *The Algebraic Eigenvalue Problem*, Clarendon Press, Oxford, 1965.

[34] A. J. MACLEOD, *Some statistics on Gaussian elimination with partial pivoting*, ACM SIGNUM Newsletter, 24, nos. 2–3 (1989), pp. 10–14.