Explicit Formulas for the Results of Gaussian Elimination

D. Stott Parker
stott@cs.ucla.edu
Computer Science Department
University of California
Los Angeles, CA 90024-1596

July 21, 1995

Abstract

Gaussian Elimination is probably the best known and most widely used method for solving linear systems, computing determinants, and decomposing matrices. It is an important tool in the kernel of most computational linear algebra packages today. While the basic elimination procedure is simple to state and implement, it is complex enough that it is nontrivial to reason about.

In this paper we resurrect a characterization of the behavior of Gaussian Elimination, namely that every one of its intermediate results is a ratio of determinants of submatrices of the input matrix, and show its equivalence to another characterization in terms of Schur complements. These characterizations seem to be unfamiliar to numerical analysts, but deserve wider attention since they lead to a deep appreciation of Gaussian Elimination.

The characterizations can be used in obtaining bounds for roundoff error with Gaussian Elimination. The characterizations also help in better understanding the effects of roundoff error and pivoting.

1 Introduction

Today high-performance computers are often used for little more than solving large linear systems, and they run variants of Gaussian Elimination and LU decomposition constantly [8, 31]. It is a prominent topic in every text on computational linear algebra. A comprehensive overview of modern numerical software references up to the late-1980's can be found in Chapter 3 of [14].[1]

In its basic form, Gaussian Elimination is a sequence of transformations to an $n \times n$ square matrix $A = (a_{ij})$, reducing it to upper-diagonal form in n steps. It can be defined equationally, with the initial assignment $a_{ij}^{(1)} = a_{ij}$ and the recursive definition of $a_{ij}^{(k+1)}$ for $1 \leq k \leq n-1$:

$$a_{ij}^{(k+1)} = \begin{cases} 0 & i \geq k+1,\ j = k \\ a_{ij}^{(k)} - \left(a_{ik}^{(k)} \,/\, a_{kk}^{(k)} \right) a_{kj}^{(k)} & i \geq k+1,\ j \geq k+1 \\ a_{ij}^{(k)} & \text{otherwise.} \end{cases}$$

Below we will obtain explicit formulas that solve these equations.

[1] A good review of the historical development of the Gaussian Elimination algorithm is the brief summary of Schrijver [33, pp.38–41], which traces it back several thousand years. It seems improperly attributed to Gauss, both because it predates him and it is difficult to find in his works (Schrijver emphasizes Arts. 180 and 182 of Gauss' [12]). However Gauss was certainly accomplished at solving equations via elimination.

Often this definition is viewed as summarizing a program – an implementation of Gaussian Elimination. However: *here the definition is viewed as a set of equations, without any 'roundoff error'*; we will discuss roundoff error later. Gaussian Elimination is typically implemented with **for**–loops in a program requiring $n(n^2-1)/3$ assignments altogether:

for $k = 1$ **to** $n-1$ **do**
  for $i = k+1$ **to** n **do**
    begin
      $m_{ik} \ := \ a_{ik} \ / \ a_{kk};$ $\{ \quad a_{ik}^{(k)} \ / \ a_{kk}^{(k)} \quad \}$
      for $j = k+1$ **to** n **do**
        $a_{ij} \ := \ a_{ij} \ - \ m_{ik} \ \cdot \ a_{kj}$ $\{ \quad a_{ij}^{(k)} \ - \ a_{ik}^{(k)} / a_{kk}^{(k)} \ \cdot \ a_{kj}^{(k)} \quad \}$
    end

Suppose that this program is executed with infinite-precision computation. Starting with $A^{(1)} = A$, each iteration of the outer loop computes $A^{(k+1)} = (a_{ij}^{(k+1)})$ for $1 \leq k \leq n-1$, except that as written above the program does not zero the elements of A below the diagonal in the first k columns. (These elements are in precisely the same positions as the multipliers m_{ik}, and so many implementations store m_{ik} in a_{ik}.) With the program we can form the triangular matrices

$$L \;=\; \begin{pmatrix} 1 & & & & \\ m_{21} & 1 & & & \\ m_{31} & m_{32} & 1 & & \\ \vdots & \vdots & \vdots & \ddots & \\ m_{n1} & m_{n2} & m_{n3} & \cdots & 1 \end{pmatrix}, \qquad U \;=\; \begin{pmatrix} a_{11}^{(n)} & a_{12}^{(n)} & a_{13}^{(n)} & \cdots & a_{1n}^{(n)} \\ & a_{22}^{(n)} & a_{23}^{(n)} & \cdots & a_{2n}^{(n)} \\ & & a_{33}^{(n)} & \cdots & a_{3n}^{(n)} \\ & & & \ddots & \vdots \\ & & & & a_{nn}^{(n)} \end{pmatrix}$$

such that $A^{(n)} = U$ and $A = LU$. The computation reduces a general linear problem of the form $A\mathbf{x} = \mathbf{b}$ to a pair of back-substitution problems $L\mathbf{y} = \mathbf{b}$, $U\mathbf{x} = \mathbf{y}$. For a particularly good introduction see [10].

In practice, of course, implementations of Gaussian Elimination rely on finite-precision computations. Thus a program like the one above will find finite-precision approximations $\widehat{a}^{(k)}$ to $a^{(k)}$, and compute approximations $\widehat{L}$ and $\widehat{U}$ to L and U. We will discuss roundoff error later.

Despite its clear importance, there still remain a number of mysteries about the results of Gaussian Elimination [36]. Many of these revolve around its average- and worst-case roundoff error behavior, which has resisted analysis for fifty years [37]. The worst-case error growth for Gaussian Elimination and complete pivoting has only recently been determined for $n \leq 16$ [15], and until [7] in 1988 was determined only for $n \leq 4$. Because the roundoff errors are bounded by the sizes of intermediate results in $\widehat{A}^{(k)}$ [10], we can begin to solve these mysteries if we can say explicitly what Gaussian Elimination produces.

2 Results of Gaussian Elimination as Ratios of Determinants

Definition 1 *The* **submatrix of** A **with rows** $i_1, \ldots, i_p$ **and columns** $j_1, \ldots, j_q$ *(in that order) is denoted* $A[\, i_1, \ldots, i_p \,|\, j_1, \ldots, j_q \,]$*, and consists of the indicated elements of* A*:*

$$A[\, i_1, \ldots, i_p \,|\, j_1, \ldots, j_q \,] \quad = \quad \begin{pmatrix} a_{i_1,j_1} & a_{i_1,j_2} & \cdots & a_{i_1,j_q} \\ a_{i_2,j_1} & a_{i_2,j_2} & \cdots & a_{i_2,j_q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i_p,j_1} & a_{i_p,j_2} & \cdots & a_{i_p,j_q} \end{pmatrix}.$$

For example, the submatrix of A *with rows 3, 2 and columns 1, 3, 1, 4 is*

$$A[\, 3, 2 \,|\, 1, 3, 1, 4 \,] \quad = \quad \begin{pmatrix} a_{31} & a_{33} & a_{31} & a_{34} \\ a_{21} & a_{23} & a_{21} & a_{24} \end{pmatrix}.$$

Also, when $p = 0$ *or* $q = 0$ *the submatrix is 1 by convention, so* $A[\,|\,] \;=\; 1$.

Since it occurs often, we write

$$1..k \qquad \text{as synonymous for the sequence} \qquad 1,\, 2,\, \ldots,\, k.$$

Thus for example

$$A[\, 1..(k-1), i \,|\, 1..(k-1), j \,] \qquad \text{is synonymous for} \qquad A[\, 1, 2, \ldots, (k-1), i \,|\, 1, 2, \ldots, (k-1), j \,].$$

Gaussian Elimination works only when the $a_{kk}^{(k)}$ are nonzero, for $1 \le k \le n$. It is well-known that this is equivalent to the following (and we will show this later):

Definition 2 *An* $n \times n$ *matrix* A *is* **Gauss eliminable** *if det* $A[\, 1..k \,|\, 1..k \,] \neq 0$*,* $1 \le k \le n$*.*

For clarity, throughout this paper we will consider only Gauss eliminable matrices. Although the results here can be extended for pivoting and Moore-Penrose generalized matrix inverses, in this paper we will not consider such extensions.

Theorem 1 *If* A *is an* $n \times n$ *Gauss eliminable matrix, then:*
For $1 \le k < i \le n$*,*

$$\ell_{ik} \quad = \quad \frac{\mathit{det}\ A[\, 1..(k-1), i \,|\, 1..k \,]}{\mathit{det}\ A[\, 1..k \,|\, 1..k \,]}.$$

For $1 \le k < i \le n$*,* $1 \le k < j \le n$*,*

$$a_{ij}^{(k+1)} \quad = \quad \frac{\mathit{det}\ A[\, 1..k,\, i \,|\, 1..k,\, j \,]}{\mathit{det}\ A[\, 1..k \,|\, 1..k \,]}.$$

A complete proof of this theorem is presented in the appendix. After deriving it, we found Theorem 1 first in [22] (which gives similar formulas for LDU factorization) and [11], and have since found it in a number of references, such as [33] and [17] on the foundations of linear programming. The excellent treatment in Gantmacher's book [11] traces the identities to Grossman [16]. A much earlier reference seems possible, because much early work on determinants was aimed at understanding elimination; for example the identities resemble Cramer's rule of 1750.

We have not found Theorem 1 in any modern numerical analysis reference. One purpose of this paper is to call attention to it again, extend it, and clarify its role for Gaussian Elimination.

3 Results of Gaussian Elimination as Schur Complements

In this section we derive formulas for $a_{ij}^{(k+1)}$ using *Schur complements.* A quick literature survey shows that Schur complements are rapidly gaining appreciation. We show that they are intimately related to the determinantal identities above, and that in fact *every* value $a_{ij}^{(k+1)}$ produced by Gaussian Elimination is expressible as a Schur complement of a leading principal submatrix of A in a submatrix of A.

3.1 Basic Properties of Schur Complements

Definition 3 *If* $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ *is a block matrix in which A is nonsingular, the* **Schur complement of** A **in** M*, written* $(\ M\ /\ A\ )$*, is the matrix*

$$(\ M\ /\ A\ ) \quad = \quad D - CA^{-1}B.$$

More generally, for any nonsingular principal submatrix $M[\,\alpha\,|\,\alpha\,]$ where α is any (ordered) subsequence of $1, \ldots, n$, if α' is its (ordered) sequence complement in $1, \ldots, n$ define

$$(\ M\ /\ M[\,\alpha\,|\,\alpha\,]\ ) \quad = \quad M[\,\alpha'\,|\,\alpha'\,] \quad - \quad M[\,\alpha'\,|\,\alpha\,]\cdot(M[\,\alpha\,|\,\alpha\,])^{-1}\cdot M[\,\alpha\,|\,\alpha'\,]$$

Emilie Haynsworth pioneered interest in the Schur complement in a series of papers published between 1968 and 1974 [3, 6, 18, 19, 20, 27]. Cottle [5] surveys uses of Schur complements, including determinantal identities, pivoting, inverse matrices, inertia, etc. See also [21, pp.17–23].

Lemma 1 (Schur's identity for determinants)
If $M[\,\alpha\,|\,\alpha\,]$ is nonsingular and β is a subsequence of α' (so α and β are disjoint), then

$$\frac{det\ M[\,\alpha,\beta\,|\,\alpha,\beta\,]}{det\ M[\,\alpha\,|\,\alpha\,]} \quad = \quad det\ (\ M[\,\alpha,\beta\,|\,\alpha,\beta\,]\ /\ M[\,\alpha\,|\,\alpha\,]\ ).$$

This identity can be derived by forming a block LU decomposition of $M[\,\alpha,\beta\,|\,\alpha,\beta\,]$, which finds the diagonal entries of U to be $M[\,\alpha\,|\,\alpha\,]$ and $(M[\,\beta\,|\,\beta\,]\ -\ M[\,\beta\,|\,\alpha\,]\,(M[\,\alpha\,|\,\alpha\,])^{-1}\,M[\,\alpha\,|\,\beta\,])$.

Lemma 2 (Principal submatrices of M^{-1})
If α is any ordered sequence such that $M[\,\alpha\,|\,\alpha\,]$ is nonsingular,

$$M^{-1}\,[\,\alpha'\,|\,\alpha'\,] \quad = \quad (\ M\ /\ M[\,\alpha\,|\,\alpha\,]\ )^{-1}$$

where α' is the ordered complement of α'.

Note that in general $M^{-1}[\,\alpha'\,|\,\alpha'\,] \neq (M[\,\alpha'\,|\,\alpha'\,])^{-1}$. This is again derivable with block LU decomposition, the $[\,\alpha'\,|\,\alpha'\,]$ block of M^{-1} being $\left(M[\,\alpha'\,|\,\alpha'\,]\ -\ M[\,\alpha'\,|\,\alpha\,]\cdot(M[\,\alpha\,|\,\alpha\,])^{-1}\cdot M[\,\alpha\,|\,\alpha'\,]\right)^{-1}$.

Lemma 3 (Quotient Property of Schur Complements)
If we have the block decompositions

$$M \quad = \quad \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \qquad A \quad = \quad \begin{pmatrix} E & F \\ G & H \end{pmatrix}$$

where both A and E are nonsingular, then

$$(\ M\ /\ A\ ) \quad = \quad (\ (\ M\ /\ E\ )\ /\ (\ A\ /\ E\ )\ ).$$

This is proven in [6] with Lemma 1 and Sylvester's identity (Theorem 8 in the appendix). Other proofs using less sophisticated tools have been developed, e.g., by Ostrowski [27].

3.2 Results of Gaussian Elimination as Schur Complements

We show that every result of Gaussian Elimination is expressible explicitly as a Schur complement. Thus, any property of Schur complements gives a corresponding property of these results.

Theorem 2 *If A is an $n \times n$ Gauss eliminable matrix, then:*
For $1 \leq k < i \leq n$, $1 \leq k < j \leq n$,

$$\boxed{a_{ij}^{(k+1)} \;=\; a_{ij} \;-\; A[\,i\,|\,1..k\,]\;(\,A[\,1..k\,|\,1..k\,]\,)^{-1}\;A[\,1..k\,|\,j\,]}$$

$$i.e., \boxed{a_{ij}^{(k+1)} \;=\; (\;\; A[\,1..k, i\,|\,1..k, j\,] \;\;/\;\; A[\,1..k\,|\,1..k\,] \;\;).}$$

Proof
Follows directly from Theorem 1 and Lemma 1 (Schur's identity for determinants), where $M = A[\,1..k, i\,|\,1..k, j\,]$, α is the sequence $(1..k)$ and β is $(k+1)$. The right side of Schur's identity is then the determinant of a 1×1 matrix, with $M[\,\beta\,|\,\beta\,] = a_{ij}$.

This can also be derived from Theorem 1 directly, using the perturbation identity

$$\frac{det\ X}{det\ Y} \;=\; det\,(\,I \;+\; Y^{-1}\,(X - Y)\,)$$

where

$$X \;=\; A[\,1..k, i\,|\,1..k, j\,], \qquad Y \;=\; \begin{pmatrix} A[\,1..k\,|\,1..k\,] & \mathbf{0} \\ \mathbf{0}^\top & 1 \end{pmatrix}.$$

□

Definition 4 *Let $A_{(1)} = A$ and, for $1 \leq k < n$,*

$$A_{(k+1)} \;\stackrel{def}{=}\; A^{(k+1)}\,[\,(k+1), \ldots, n\,|\,(k+1), \ldots, n\,].$$

Thus $A_{(k+1)}$ is the lower right $(n-k) \times (n-k)$ submatrix of $A^{(k+1)}$ on which elimination steps $(k+1), \ldots, (n-1)$ still need to be performed. A direct corollary of Theorem 2 is the following:

Corollary 1 *If A is an $n \times n$ Gauss eliminable matrix, then for $1 \leq k < n$:*

$$\boxed{A_{(k+1)} \;=\; (\;A \,/\, A[\,1..k\,|\,1..k\,]\;).}$$

This is merely Theorem 2 repeated for all values i and j in $(k+1), \ldots, n$:

$$A_{(k+1)} \;=\; A[\,(k+1)..n\,|\,(k+1)..n\,] \;-\; A[\,(k+1)..n\,|\,1..k\,]\;(A[\,1..k\,|\,1..k\,])^{-1}\;A[\,1..k\,|\,(k+1)..n\,].$$

In [6] this identity is derived as a lemma.

Corollary 2 *If A is an $n \times n$ Gauss eliminable matrix, then for $1 \leq k < n$:*

$$\boxed{A_{(k+1)} \;=\; (\;A^{-1}\,[\,(k+1), \ldots, n\,|\,(k+1), \ldots, n\,]\,)^{-1}.}$$

Immediate from Corollary 1 and Lemma 2. This result is derived in [7] using a different method.

Corollary 3 *If A is an $n \times n$ Gauss eliminable matrix, then for $1 \leq k < n$:*

$$\boxed{A_{(k+1)} \;=\; (\;\; A_{(k)} \;\;/\;\; (\,a_{kk}^{(k)}\,) \;\;).}$$

This identity just combines Lemma 3 (the quotient property), Corollary 1, and Theorem 2. It gives an incremental algorithm for computing $A_{(k+1)}$ — which is, not surprisingly, Gaussian Elimination.

4 Examples

Several familiar matrices give an appreciation for the results above.

4.1 The Hilbert Matrix

The $n \times n$ **Hilbert matrix** $A = (a_{ij})$ is defined by

$$a_{ij} = \frac{1}{i+j-1}$$

with inverse $A^{-1} = (b_{ij})$ such that [35]

$$b_{ij} = \frac{(-1)^{i+j}\,(n+i-1)!\,(n+j-1)!}{(i+j-1)\,(\,(i-1)!\,(j-1)!\,)^2\,(n-i)!\,(n-j)!}.$$

Consider the 4×4 Hilbert matrix and inverse

$$A = \begin{pmatrix} 1 & 1/2 & 1/3 & 1/4 \\ 1/2 & 1/3 & 1/4 & 1/5 \\ 1/3 & 1/4 & 1/5 & 1/6 \\ 1/4 & 1/5 & 1/6 & 1/7 \end{pmatrix}, \qquad A^{-1} = \begin{pmatrix} 16 & -120 & 240 & -140 \\ -120 & 1200 & -2700 & 1680 \\ 240 & -2700 & 6480 & -4200 \\ -140 & 1680 & -4200 & 2800 \end{pmatrix}.$$

Gaussian Elimination obtains an LU factorization for A in 4 steps, as follows:

k	L	$U = A^{(k)}$
1	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 1/2 & 1/3 & 1/4 \\ 1/2 & 1/3 & 1/4 & 1/5 \\ 1/3 & 1/4 & 1/5 & 1/6 \\ 1/4 & 1/5 & 1/6 & 1/7 \end{pmatrix}$
2	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1/2 & 1 & 0 & 0 \\ 1/3 & 0 & 1 & 0 \\ 1/4 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 1/2 & 1/3 & 1/4 \\ 0 & 1/12 & 1/12 & 3/40 \\ 0 & 1/12 & 4/45 & 1/12 \\ 0 & 3/40 & 1/12 & 9/112 \end{pmatrix}$
3	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1/2 & 1 & 0 & 0 \\ 1/3 & 1 & 1 & 0 \\ 1/4 & 9/10 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 1/2 & 1/3 & 1/4 \\ 0 & 1/12 & 1/12 & 3/40 \\ 0 & 0 & 1/180 & 1/120 \\ 0 & 0 & 1/120 & 9/700 \end{pmatrix}$
4	$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1/2 & 1 & 0 & 0 \\ 1/3 & 1 & 1 & 0 \\ 1/4 & 9/10 & 3/2 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 1/2 & 1/3 & 1/4 \\ 0 & 1/12 & 1/12 & 3/40 \\ 0 & 0 & 1/180 & 1/120 \\ 0 & 0 & 0 & 1/2800 \end{pmatrix}$.

To verify Theorems 1 and 2, select for example $a_{34}^{(3)} = 1/120$.
According to Theorem 1,

$$a_{34}^{(2+1)} = \frac{det\ A[\,1,2,3\,|\,1,2,4\,]}{det\ A[\,1,2\,|\,1,2\,]} = \frac{1/1440}{1/12} = 1/120.$$

According to Theorem 2,

$$\begin{aligned} a_{34}^{(2+1)} &= a_{34} - A[\,3\,|\,1,2\,]\,(A[\,1,2\,|\,1,2\,])^{-1}\,A[\,1,2\,|\,4\,] \\ &= 1/6 - \begin{pmatrix} 1/3 & 1/4 \end{pmatrix} \begin{pmatrix} 4 & -6 \\ -6 & 12 \end{pmatrix} \begin{pmatrix} 1/4 \\ 1/5 \end{pmatrix} \\ &= 1/120. \end{aligned}$$

The values of $a_{ij}^{(k)}$ are all at most 1, and decrease rapidly on the diagonal. From the explicit formulas for $A = (a_{ij})$ and $A^{-1} = (b_{ij})$, Theorem 2 gives the semi-explicit

$$a_{ij}^{(k+1)} = \frac{1}{i+j-1} - \sum_{p,\,q=1}^{k} \frac{1}{i+p-1} \frac{(-1)^{p+q}\,(k+p-1)!\,(k+q-1)!}{(p+q-1)\,(\,(p-1)!\,(q-1)!\,)^2\,(k-p)!\,(k-q)!} \frac{1}{q+j-1}.$$

Finally, to verify Corollary 2, consider

$$A_{(3)} = A^{(3)}[\,3,4\,|\,3,4\,] = \begin{pmatrix} 1/180 & 1/120 \\ 1/120 & 9/700 \end{pmatrix}.$$

According to Corollary 2,

$$A_{(3)} = (\,A^{-1}[\,3,4\,|\,3,4\,]\,)^{-1} = \begin{pmatrix} 6480 & -4200 \\ -4200 & 2800 \end{pmatrix}^{-1} = \begin{pmatrix} 1/180 & 1/120 \\ 1/120 & 9/700 \end{pmatrix}.$$

4.2 Matrices of Turing and Wilkinson

In [38], Turing remarks that the matrices defined by

$$T_n = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ -1 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & -1 & -1 & \cdots & 1 & 0 \\ -1 & -1 & -1 & \cdots & -1 & 1 \end{pmatrix}, \qquad T_n^{-1} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 0 & \cdots & 0 & 0 \\ 2 & 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 2^{n-3} & 2^{n-4} & 2^{n-5} & \cdots & 1 & 0 \\ 2^{n-2} & 2^{n-3} & 2^{n-4} & \cdots & 1 & 1 \end{pmatrix}$$

have high condition number. It is plain from the definition here that

$$\kappa_1(\,T_n\,) = n\,2^{n-1}$$

where $\kappa_1(M) = \|\,M\,\|_1\ \|\,M^{-1}\,\|_1$ is the condition number of M using the column-sum norm $\|\cdot\|_1$. Later, his colleague Wilkinson [41, p.212] adapted this matrix to give a matrix W_n with worst-case error in LU decomposition with partial pivoting, where $L = T_n$:

$$W_n = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 1 \\ -1 & 1 & 0 & \cdots & 0 & 1 \\ -1 & -1 & 1 & \cdots & 0 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & -1 & -1 & \cdots & 1 & 1 \\ -1 & -1 & -1 & \cdots & -1 & 1 \end{pmatrix} = T_n \cdot \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 1 & 0 & \cdots & 0 & 2 \\ 0 & 0 & 1 & \cdots & 0 & 4 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 2^{n-1} \\ 0 & 0 & 0 & \cdots & 0 & 2^n \end{pmatrix}$$

Because $T_n[\,1..k\,|\,1..k\,] = T_k$, and also $T_n^{-1}[\,(n-m+1),\ldots,n\,|\,(n-m+1),\ldots,n\,] = T_m^{-1}$, Corollary 2 shows for example that the Schur complement of T_n with respect to its leading $k \times k$ submatrix is T_{n-k}:

$$(T_n)_{(k+1)} = \left(\,T_n^{-1}[\,(k+1),\ldots,n\,|\,(k+1),\ldots,n\,]\,\right)^{-1} = T_{n-k}.$$

Wilkinson's W_n is interesting because its leading principal submatrices have high condition numbers

$$\kappa_1(\,W_n[\,1..k\,|\,1..k\,]\,) = \kappa_1(\,T_k\,) = k\,2^{k-1} \qquad (1 \le k \le n-1)$$

but the matrix itself has low condition number $\kappa_1(W_n) = n$. Put a different way, W_n itself is not poorly conditioned, but almost all of its leading principal submatrices are.

4.3 A Recursive Hadamard Matrix

A **Hadamard matrix** is a square matrix whose entries are ± 1, such that the row vectors of A are mutually orthogonal. In the error analysis of Gaussian Elimination, Hadamard matrices are often studied because they yield large intermediate values of $a_{ij}^{(k)}$ [7]: **Hadamard's inequality** implies that an $n \times n$ matrix A with entries in the interval $[-1, 1]$ must satisfy $|det\ A| \leq n^{n/2}$, and equality is possible only if A is a Hadamard matrix.

Let us consider here only recursive Hadamard matrices defined by

$$H_1 \quad = \quad \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad H_{p+1} \quad = \quad H_1 \otimes H_p \quad = \quad \begin{pmatrix} H_p & H_p \\ H_p & -H_p \end{pmatrix} \qquad p \geq 1.$$

We can easily prove by induction for $p \geq 1$ that:

- ${H_p}^{-1} \;=\; 2^{-p}\, {H_p}^{\top} \;=\; 2^{-p}\, H_p$.
- The eigenvalues of H_p consist of 2^{p-1} instances of $2^{p/2}$ and 2^{p-1} instances of $-2^{p/2}$.
- For $p > 1$, $det\ H_p = (2^p)^{2^{p-1}}$. Thus $det\ H_p = n^{n/2}$ when $n = 2^p > 2$.
- If $H_p = (a_{ij})$, and $\beta_q(x)$ is the q-th bit in the binary representation of integer x, then

$$a_{ij} \quad = \quad \prod_{q=0}^{p-1} (-1)^{\;\beta_q(i-1)\;\beta_q(j-1)}.$$

The LDU decomposition of these matrices is obtainable recursively as

$$H_{p+1} \quad = \quad \begin{pmatrix} I & 0 \\ I & I \end{pmatrix} \begin{pmatrix} H_p & 0 \\ 0 & -2H_p \end{pmatrix} \begin{pmatrix} I & I \\ 0 & I \end{pmatrix}.$$

Correspondingly $(\; H_{p+1} \;/\; H_p \;) \;=\; -2H_p$. For $1 \leq k \leq n$, the resulting k-th diagonal entry, representing $a_{kk}^{(k)}$, is thus an integral power of (-2):

$$a_{kk}^{(k)} \quad = \quad \prod_{q=0}^{p-1} (-2)^{\;\beta_q(k-1)}.$$

Furthermore, the absolute value of every entry in the completed LDU decomposition will be either 0, or a power of 2. With this information, Theorem 1 implies that:

> If $A = H_p$, then for any i and j larger than k, $|det\ A[\,1..k, i \,|\, 1..k, j\,]|$ is either 0, or is a power of 2 not less than $|det\ A[\,1..k \,|\, 1..k\,]|$.

The value of $a_{ij}^{(k+1)}$ when $A = H_p$ and $k < n = 2^p$ can be obtained via Corollary 1's identity

$$A_{(k+1)} \;=\; A[\,(k+1)..n \,|\, (k+1)..n\,] \;-\; A[\,(k+1)..n \,|\, 1..k\,]\; (A[\,1..k \,|\, 1..k\,])^{-1}\; A[\,1..k \,|\, (k+1)..n\,]$$

with recursive evaluation of $(A[\,1..k \,|\, 1..k\,])^{-1}$ using Schur complements and the two facts that $H_p = 2^p\, {H_p}^{-1}$, and when $k < n/2$ the $k \times k$ initial or final submatrix of H_p is also that of H_{p-1}.

The structure of the leading submatrices $H_p[\,1..k \,|\, 1..k\,]$ is interesting. For $p \leq 5$, and probably in general, the absolute values of eigenvalues have maximum $2^{p/2}$ and minimum $2^{(\log_2^+ k)/2}$, where we define $\log_2^+ k$ to be the number of factors of 2 in k so that e.g. $\log_2^+ 2^q 3 = q$. Therefore we conjecture that for all p and k,

$$\kappa_2(H_p[\,1..k \,|\, 1..k\,]) \quad = \quad 2^{(p \;-\; \log_2^+ k)/2}$$

where $\kappa_2(M) = \|\, M \,\|_2\; \|\, M^{-1} \,\|_2$ is the condition number of M using the spectral norm $\|\cdot\|_2$.

4.4 The Rodman Matrix

If t is a real parameter, which we take to be in $[0,1]$, the $n \times n$ **Rodman matrix** [32] and its inverse are defined by

$$R_n \quad = \quad \begin{pmatrix} 1 & t & t & \cdots & t \\ t & 1 & t & \cdots & t \\ t & t & 1 & \cdots & t \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t & t & t & \cdots & 1 \end{pmatrix}, \qquad R_n^{-1} \quad = \quad \frac{-1}{\lambda_n\ \lambda_1} \begin{pmatrix} t-\lambda_n & t & t & \cdots & t \\ t & t-\lambda_n & t & \cdots & t \\ t & t & t-\lambda_n & \cdots & t \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t & t & t & \cdots & t-\lambda_n \end{pmatrix}.$$

R_n has maximum eigenvalue $\lambda_n = (1 + (n-1)\,t)$, and $n-1$ other eigenvalues $\lambda_1 = (1-t)$, so

$$det\ R_n \quad = \quad \lambda_n\ {\lambda_1}^{n-1} \quad = \quad (1+(n-1)\,t)\ \ (1-t)^{n-1}.$$

Either Theorem 1 or Theorem 2 then implies that Gaussian Elimination on R_n yields

$$a_{ij}^{(k+1)} \quad = \quad \begin{cases} (1+kt)\ (1-t)\ /\ (1+(k-1)\,t) & \text{if } i=j \\ (1-t)\ /\ (1+(k-1)\,t) & \text{if } i \neq j. \end{cases}$$

5 Applications

The results above improve one's understanding of the behavior of Gaussian Elimination. For example, Theorem 1 shows why the condition that matrix A be Gauss eliminable, i.e.,

$$det\ A[\,1..k\,|\,1..k\,] \quad \neq \quad 0 \qquad (1 \leq k \leq n),$$

is equivalent to requiring

$$a_{kk}^{(k)} \quad \neq \quad 0 \qquad (1 \leq k \leq n).$$

Also, the results give insight into *invariants* that govern each iteration of Gaussian Elimination, and into error bounds derivable for linear systems solvers based on Gaussian Elimination.

5.1 Understanding Properties that are Invariant under Gaussian Elimination

Theorem 1 shows that Gaussian Elimination repeatedly transforms $A_{(k)}$ to $A_{(k+1)}$ in a way that preserves the following invariant properties:

- $det\ A_{(k)} = a_{kk}^{(k)}\ det\ A_{(k+1)}$.
  This gives a direct way to show standard results about Gaussian Elimination and $det\ A$, such as that when $a_{kk}^{(k)}$ is nonzero for $1 \leq k \leq n$, then A is nonsingular and
  $$det\ A \quad = \quad a_{11}^{(1)}\ a_{22}^{(2)}\ \cdots\ a_{nn}^{(n)}.$$
- $A_{(k+1)}$ is Gauss eliminable if $A_{(k)}$ is Gauss eliminable.
- $A_{(k+1)}$ is symmetric if $A_{(k)}$ is symmetric.
- $A_{(k+1)}$ is positive definite if $A_{(k)}$ is positive definite [26].

These invariant properties can all be observed with the Hilbert and Rodman matrices just studied, because they are Gauss eliminable, symmetric, and positive definite.

Related to the final property, a nice application of Theorem 1 arises in determining whether a matrix is positive definite, using the following well-known characterization:

Theorem 3 *(Frobenius, 1876 [9, p.270])*
A necessary and sufficient condition for a real symmetric $n \times n$ matrix A to be positive definite is that det $A[\,1..k \,|\, 1..k\,] \; > \; 0$ for $1 \le k \le n$.

Thus for a real symmetric matrix A to be positive definite it is necessary and sufficient for Gaussian Elimination to find $a_{kk}^{(k)} > 0$ for $1 \le k \le n$.[2]

5.2 Bounds on Results of Gaussian Elimination

Theorem 2 leads directly to new bounds:

Theorem 4 *If A is an $n \times n$ Gauss eliminable matrix, then:*
For $1 \le k < i \le n$, $1 \le k < j \le n$,

$$|a_{ij}^{(k+1)}| \quad \le \quad |a_{ij}| \quad + \quad \kappa_2(\; A[\,1..k \,|\, 1..k\,]\;) \; \frac{\|\, A[\,i \,|\, 1..k]\,\|_2 \; \|\, A[\,1..k \,|\, j\,]\,\|_2}{\|\, A[\,1..k \,|\, 1..k\,]\,\|_2.}$$

Proof
Direct from Theorem 2:

$$\begin{aligned} |a_{ij}^{(k+1)}| \quad &= \quad |\; a_{ij} \quad - \quad A[\,i \,|\, 1..k]\; (A[\,1..k \,|\, 1..k])^{-1} \; A[\,1..k \,|\, j\,]\;| \\ &\le \quad |a_{ij}| \quad + \quad \|\, A[\,i \,|\, 1..k]\,\|_2 \; \|\, (A[\,1..k \,|\, 1..k])^{-1}\,\|_2 \; \|\, A[\,1..k \,|\, j\,]\,\|_2 \\ &= \quad |a_{ij}| \quad + \quad \kappa_2(\; A[\,1..k \,|\, 1..k\,]\;) \; \frac{\|\, A[\,i \,|\, 1..k]\,\|_2 \; \|\, A[\,1..k \,|\, j\,]\,\|_2}{\|\, A[\,1..k \,|\, 1..k\,]\,\|_2}. \end{aligned}$$

□

5.3 Backwards Error Analysis, Growth Factors, and Complete Pivoting

Our original motivation behind this paper was to get more exact bounds for the error in Gaussian Elimination. The roundoff error in the floating point implementation of Gaussian Elimination has a well-known backwards error bound:

Theorem 5 *(Wilkinson [39, 40])*
Suppose A is an $n \times n$ nonsingular matrix. Then the matrices $\widehat{L}$ and $\widehat{U}$ computed by Gaussian Elimination with pivoting in floating point with unit roundoff $\mathbf{u}$ satisfy $\widehat{L}\widehat{U} = A + E$, where E is a matrix of roundoff errors[3] *such that, if $\widehat{a}_{ij}^{(k)}$ is the computed floating point approximation to $a_{ij}^{(k)}$,*

$$\|\, E \,\|_\infty \quad \le \quad n^2 \; \max_{i,j,k} \; |\widehat{a}_{ij}^{(k)}| \quad \mathbf{u}.$$

Many researchers concentrate on bounding $\max_{i,j,k} \; |\widehat{a}_{ij}^{(k)}|$ as a result of theorems like this one. Because it has been difficult in the past to do this, such bounds were often swept under a carpet called a *growth factor* such as

$$\rho_n(A) \quad = \quad \frac{\max_{i,j,k} \; |\widehat{a}_{ij}^{(k)}|}{\max_{i,j} \; |a_{ij}|}.$$

---

[2]Ostrowski and Taussky showed that Frobenius' characterization is also sufficient for nonsymmetric A; see [2, p.59]. Thus $a_{kk}^{(k)} > 0$ is sufficient for A to be positive definite.

[3]Here the matrix infinity-norm is defined by $\|\, A \,\|_\infty = \max_{i=1}^n \; \sum_{j=1}^n \; |a_{ij}|$. The unit roundoff $\mathbf{u}$ on a machine with t-digit, base-β floating-point arithmetic is $\frac{1}{2}\beta^{1-t}$ when floating point operations are rounded, and β^{1-t} when they are truncated. The bound implicitly assumes $\beta^{1-t} \le 1/n$.

This growth factor is usually estimated informally, because how it can be bounded is still a famous open question; see [7, 23, 37] and the recent work [15]. Growth factors often appear along with condition numbers of matrices in error bounds for linear systems [14]. Recently there has been a good deal of research on growth factors, stressing that concern may be warranted for error growth in real problems when Gaussian Elimination with partial pivoting is used. For example, Wright [42] finds exponential error growth for a specific family of matrices arising in two-point boundary value problems. Both [14, p.122] and [23] present many other recent results on growth factors.

Von Neumann and Goldstine noted that when A is symmetric positive definite, $a_{ii}^{(k+1)} \leq a_{ii}^{(k)}$ and [26, p.1066] $\widehat{a}_{ii}^{(k+1)} \leq \widehat{a}_{ii}^{(k)}$ for $1 \leq k < i \leq n$, and putting this together with the fact that $\max_i a_{ii} = \max_{i,j} |a_{ij}|$ when A is symmetric positive definite, derived the bound

$$\max_{i,j,k} |\widehat{a}_{ij}^{(k)}| \quad \leq \quad \max_{i,j} |a_{ij}|.$$

Wilkinson proved that use of pivoting in Gaussian Elimination yields tighter bounds on roundoff error, notably his famous bound for complete pivoting [39]:[4]

$$|\widehat{a}_{ij}^{(k)}| \quad \leq \quad k^{1/2} \left(2^1\, 3^{1/2}\, 4^{1/3} \cdots k^{1/(k-1)}\right)^{1/2} \max_{i,j} |a_{ij}| \quad \sim \quad C\, k^{1/2}\, k^{(1/4)\ln k} \max_{i,j} |a_{ij}|.$$

Wilkinson complained that this bound appears quite loose and a tighter bound may be possible. Although it seems very difficult to improve on Wilkinson's bound, we can come up with the following bound involving condition numbers of submatrices. It suggests that growth factors can be approximated by condition numbers.

Theorem 6 *When Gaussian Elimination with complete pivoting is used on an $n \times n$ nonsingular matrix A, and the index ordering $1, \ldots, k$ reflects the final ordering after pivoting,*

$$|a_{ij}^{(k)}| \quad \leq \quad k \quad \kappa_2(\, A[\,1..(k-1)\,|\,1..(k-1)]\,) \quad \max_{i,j} |a_{ij}|.$$

Proof

With the index ordering obtained by complete pivoting, $\max_{i,j} |a_{ij}| = |a_{11}|$, so Theorem 4 implies:

$$\begin{aligned}
|a_{ij}^{(k)}| \quad &\leq \quad |a_{ij}| \quad + \quad \kappa_2(\, A[\,1..(k-1)\,|\,1..(k-1)]\,) \ \frac{\|\, A[\,i\,|\,1..(k-1)]\,\|_2 \ \|\, A[\,1..(k-1)\,|\,j\,]\,\|_2}{\|\, A[\,1..(k-1)\,|\,1..(k-1)]\,\|_2} \\
&\leq \quad |a_{ij}| \quad + \quad \kappa_2(\, A[\,1..(k-1)\,|\,1..(k-1)]\,) \ \frac{(\sqrt{(k-1)}\ |a_{11}|\,)\,(\sqrt{(k-1)}\ |a_{11}|\,)}{|a_{11}|} \\
&= \quad |a_{ij}| \quad + \quad (k-1)\ \ \kappa_2(\, A[\,1..(k-1)\,|\,1..(k-1)]\,) \ \ |a_{11}| \\
&\leq \quad k \quad \kappa_2(\, A[\,1..(k-1)\,|\,1..(k-1)]\,) \ \ \max_{i,j} |a_{ij}|.
\end{aligned}$$

The bound $\|\, A[\,1..(k-1)\,|\,1..(k-1)]\,\|_2 \geq |a_{11}|$ holds because $\|\, A\,\|_2 = \max_{\|\mathbf{x}\|_2=\|\mathbf{y}\|_2=1} |\mathbf{y}^* A \mathbf{x}|$, and we can select $\mathbf{x} = \mathbf{y} = (1, 0, \ldots, 0)$. The final inequality follows since always $\kappa_2(\cdot) \geq 1$. □

Theorems 5 and 6 inspire the question: *how closely does $\widehat{a}_{ij}^{(k)}$ approximate $a_{ij}^{(k)}$?* [5]

---

[4]The same bound is derivable for $\widehat{a}_{ij}^{(k)}$ with Theorem 1, since $\prod_{i=k}^{n} \widehat{a}_{ii}^{(i)} = det\,\widehat{A}_{(k)}$ and Hadamard's bound for determinants says $det\,\widehat{A}_{(k)} \leq (\sqrt{n-k+1}\ \widehat{a}_{kk}^{(k)})^{n-k+1}$, because $\max_{i,j\,\geq k} |\widehat{a}_{ij}^{(k)}| = |\widehat{a}_{kk}^{(k)}|$ with complete pivoting. Minor improvements on this bound are possible; for a good discussion, see [7, pp.496–498].

[5]That $\widehat{a}_{ij}^{(k)}$ and $a_{ij}^{(k)}$ might diverge is important. This divergence is not considered in [7, 15], which tacitly assume $\widehat{a}_{ij}^{(k)} = a_{ij}^{(k)}$ in specifying the worst-case growth factor. Both of these papers also ignore the roundoff error that arises in their growth factor nonlinear optimization computations. To be complete, it appears any definition of growth factor must take into account the floating-point system being used.

In fact, it can happen that $\widehat{a}_{ij}^{(k)}$ and $a_{ij}^{(k)}$ diverge, i.e., that $\widehat{U}$ and U differ by quite a bit. Divergence occurs when the ratios $|a_{ij}^{(k)}/a_{ij}^{(k+1)}|$ are large, i.e., considerable cancellation occurs in computing $a_{ij}^{(k+1)}$.

To see this, define the error $\eta_{ij}^{(k)}$ for $1 \le k \le i, j \le n$ by $a_{ij}^{(k)} + \eta_{ij}^{(k)} = \widehat{a}_{ij}^{(k)}$. Then:

$$\begin{aligned}
\widehat{a}_{ij}^{(k+1)} &= fl[\ \widehat{a}_{ij}^{(k)} - \widehat{a}_{ik}^{(k)} \div \widehat{a}_{kk}^{(k)} \times \widehat{a}_{kj}^{(k)}\ ] \\
&= \widehat{a}_{ij}^{(k)}\,(1+\delta_3) - \frac{\widehat{a}_{ik}^{(k)}}{\widehat{a}_{kk}^{(k)}}\,(1+\delta_1)\ \widehat{a}_{kj}^{(k)}\,(1+\delta_2)(1+\delta_3) \\
&= (a_{ij}^{(k)} + \eta_{ij}^{(k)})\,(1+\delta_3) - \frac{(a_{ik}^{(k)} + \eta_{ik}^{(k)})\,(a_{kj}^{(k)} + \eta_{kj}^{(k)})}{(a_{kk}^{(k)} + \eta_{kk}^{(k)})}\ (1+\delta_1)\,(1+\delta_2)\,(1+\delta_3)
\end{aligned}$$

where the δ's are individual roundoff errors, bounded by the unit roundoff error $\mathbf{u}$. This gives

$$\begin{aligned}
\eta_{ij}^{(k+1)} &= \widehat{a}_{ij}^{(k+1)} - a_{ij}^{(k+1)} \\
&= \frac{a_{kk}^{(k)}\eta_{ij}^{(k)} + a_{ij}^{(k)}\eta_{kk}^{(k)} - a_{ik}^{(k)}\eta_{kj}^{(k)} - a_{kj}^{(k)}\eta_{ik}^{(k)} - a_{ij}^{(k+1)}\eta_{kk}^{(k)} + \eta_{ij}^{(k)}\eta_{kk}^{(k)} - \eta_{ik}^{(k)}\eta_{kj}^{(k)}}{a_{kk}^{(k)} + \eta_{kk}^{(k)}} + O(\mathbf{u}).
\end{aligned}$$

This recurrence shows how errors might grow exponentially at first (until it is no longer true that $|\eta^{(k)}| \ll |a^{(k)}|$). since it reads $\eta^{(k+1)} \simeq 5\eta^{(k)}$ when the $a^{(k)}$ values are of comparable magnitudes, the $\eta^{(k)}$ values are of comparable magnitudes, and we ignore subscripts and signs.

Potentially, then, $\widehat{A}^{(k)}$ can diverge from $A^{(k)}$, and $\widehat{U}$ can diverge from U. In other words, we may have explosive 'forwards error' growth. This analysis is reminiscent of the gloomy exponential bounds of Hotelling [24, p.7] which helped inspire the groundbreaking work of Turing [38] and von Neumann and Goldstine [26] in establishing 'backwards error' bounds like Theorem 5, which showed that $\widehat{L}\widehat{U}$ cannot diverge significantly from A. In other words, Gaussian elimination gives excellent backwards error growth, but its forwards error growth may be exponential.

n	k												
	2	3	4	5	6	7	8	9	10	11	12	13	14
4	15.3	14.2	12.8										
5	15.3	14.2	12.7	11.9									
6	15.3	14.2	12.7	11.5	10.8								
7	15.3	14.2	12.6	11.4	10.8	9.07							
8	15.3	14.2	12.6	11.3	10.8	8.68	7.86						
9	15.3	14.2	12.6	11.3	10.8	8.44	7.37	6.69					
10	15.3	14.2	12.6	11.2	10.8	8.29	7.05	6.11	5.76				
11	15.3	14.2	12.6	11.2	10.4	8.18	6.83	5.78	4.85	4.10			
12	15.3	14.2	12.6	11.2	10.4	8.10	6.67	5.50	4.48	4.10	2.78		
13	15.3	14.2	12.6	11.2	10.4	8.04	6.55	5.31	4.14	3.86	2.06	.646	
14	15.3	14.2	12.6	11.2	10.3	8.00	6.44	5.16	3.89	3.86	1.50	.646	−.201

Table 1: $-\log_{10} \max_{ij} |\eta_{ij}^{(k)}|$ using 15-digit decimal arithmetic on the Hilbert matrix

Exponential growth actually does arise with the Hilbert matrix, and gives a dramatic appreciation of how $\widehat{a}_{ij}^{(k)}$ can diverge from $a_{ij}^{(k)}$. Table 1 shows that for the 15-digit decimal approximation of the Hilbert matrix of size $n \geq 14$, some computed entries in $\widehat{U}$ have zero significant digits. In the Hilbert matrix, the values $a_{ij}^{(k)}$, $a_{ik}^{(k)}$, $a_{kj}^{(k)}$, $a_{kk}^{(k)}$ are typically of comparable magnitude, and all are significantly larger than $a_{ij}^{(k+1)}$. In particular the ratios $|a_{ij}^{(k)}/a_{ij}^{(k+1)}|$ can get as large as $4k^2$, which is the value they take on the diagonal $i = j = k+1$. In other words, enormous cancellation occurs in computing

$$a_{ij}^{(k+1)} \quad = \quad \frac{a_{ij}^{(k)} \cdot a_{kk}^{(k)} \; - \; a_{ik}^{(k)} \cdot a_{kj}^{(k)}}{a_{kk}^{(k)}}$$

with exact arithmetic. However there is no corresponding cancellation in $\eta_{ij}^{(k+1)}$, which amounts to a randomly-weighted sum of $\eta_{ij}^{(k)}$, $\eta_{ik}^{(k)}$, $\eta_{kj}^{(k)}$, $\eta_{kk}^{(k)}$ (as long as $|\eta^{(k)}| \ll |a^{(k)}|$). Thus as k increases the values $|\eta^{(k)}|$ increase, while cancellation causes the values $|a^{(k)}|$ to decrease rapidly. Within a few iterations these values reach parity.

The terrible forward roundoff errors for the Hilbert matrix stand in stark contrast to the fact that its growth factor is 1! This growth factor follows from von Neumann and Goldstine's result above, since this matrix is symmetric positive definite. So, by Theorem 5 it has outstandingly good backwards error bounds. With the Hilbert matrix, in other words, $\widehat{A}^{(k)}$ diverges from $A^{(k)}$, yet the product $\widehat{L}\widehat{U}$ will be very close to A.

In general, $\widehat{A}^{(k)}$ will diverge from $A^{(k)}$ if there is a great deal of cancellation in computing $A^{(k+1)}$ from $A^{(k)}$. We can estimate the degree of cancellation by estimating $|a_{ij}^{(k)}/a_{ij}^{(k+1)}|$. The ratio here can also be reexpressed by Theorem 1 as

$$\left(\frac{a_{ij}^{(k)}}{a_{ij}^{(k+1)}}\right) \quad = \quad \frac{det\ A[\,1..(k-1), i \,|\, 1..(k-1), j\,]}{det\ A[\,1..(k-1) \,|\, 1..(k-1)\,]} \; \frac{det\ A[\,1..k \,|\, 1..k\,]}{det\ A[\,1..k, i \,|\, 1..k, j\,]} \quad = \quad \left(\frac{a_{kk}^{(k)}}{a_{kk\leftrightarrow ij}^{(k)}}\right)$$

where we define

$$a_{kk\leftrightarrow ij}^{(k)} \quad = \quad \frac{det\ A[\,1..(k-1), i, k \,|\, 1..(k-1), j, k\,]}{det\ A[\,1..(k-1), i \,|\, 1..(k-1), j\,]}$$

— the diagonal value that would be obtained by using $[i, j]$ as the k-th pivot instead of $[k, k]$. Thus Gaussian Elimination gives better agreement between $\widehat{U}$ and U when pivots are chosen to be as large as possible. In other words, if complete pivoting is used, $|a_{ij}^{(k)}/a_{ij}^{(k+1)}|$ is reduced, and assuming the values $a_{ij}^{(k)}$ are all of comparable magnitudes, the errors between entries in $\widehat{U}$ and U are reduced. This gives further justification for using complete pivoting.

5.4 Statistical Error Analysis

Theorem 1 suggests possible extensions to the results in the statistical analysis of Trefethen and Schreiber [37]. First, Theorem 1 gives a way to *formally explain* (or, viewing things negatively, *formally disprove*) their empirical observation that the values $\widehat{a}_{ij}^{(k)}$ display normal distributions. Even though determinants of random matrices have been proven to display normal distributions [13], the values $\widehat{a}_{ij}^{(k)}$ are ratios of these, and ratios of normally distributed random variables are not normally distributed. (It also explains, however, how one would naturally take them for normal distributions.)

Second, the theorem helps explain why random orthogonal matrices constructed so as to be 'uniform' in [37] (either with Euler angles having a Haar measure distribution, so that rows and columns are uniformly distributed on the unit $(n-1)$-sphere, or just with uniformly distributed Euler angles) will tend to have high growth factors: because their entries are small (have expected value near zero), $det\,A[\,1..k \,|\, 1..k\,]$ will be near 0 for most values of k. However, for $k = n$ this determinant must suddenly climb to 1! So, as k approaches n, the pivot values must suddenly change from ratios of small values to ratios having large numerators, and large growth factors result. Put another way, $\kappa(A)$ must be 1, but $\kappa(\,A[\,1..k\,|\,1..k\,]\,)$ will be large for some intermediate values of k. (The Wilkinson matrix W_n studied earlier had a similar property.) Theorem 1 also suggests a way to avoid this phenomenon: use a distribution of Euler angles concentrated around $\frac{\pi}{4}$, say, avoiding small sines or cosines and therefore avoiding very small random determinants.

The Hadamard matrix H_p studied earlier illustrates this point. Note that choosing $\theta = \frac{\pi}{4}$ in a reflection gives a constant multiple of the Hadamard matrix

$$\begin{pmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{pmatrix} \quad = \quad \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \quad = \quad \frac{1}{\sqrt{2}}\, H_1.$$

More generally, for $p > 1$ the Hadamard matrix H_p exhibits the limited condition number growth just described. If as conjectured earlier

$$\kappa_2(H_p[\,1..k\,|\,1..k\,]) \quad = \quad 2^{(p \;-\; \log_2^+ k)/2}$$

where $\log_2^+(k)$ is the number of factors of 2 in k, the condition number of leading submatrices is at most $2^{p/2} = \sqrt{n}$.

5.5 New Pivoting Strategies

The formulas above give tools for analyzing existing pivoting strategies and developing new ones. They make it possible to derive strategies that are provably good with respect to some measure.

Theorem 1 also suggests possible extensions to the views of pivoting presented by Poole and Neal [28, 29, 30], who among other things argue that good pivots have the property that their row is nearly orthogonal to the hyperplane generated by the rows of the previous pivots. The results here suggest we should choose pivot elements $(i_1, j_1), (i_2, j_2), \ldots, (i_n, j_n)$ so that the progression of partial volumes $det\; A[\,i_1, \ldots, i_k\,|\,j_1, \ldots, j_k\,]$ $(1 \le k \le n)$ does not vary widely — a similar view, but not exactly the same.

6 Conclusions

This paper presents a number of identities about the results $a_{ij}^{(k)}$ of Gaussian Elimination in a common framework. Theorems 4 and 6 appear to be new. Also Theorem 2 and Corollary 1 are new in their emphasis here (although the equivalence of Schur complements and determinantal ratios was derived as a Lemma in [6]). Beyond this, the existence of explicit formulas gives new perspectives on Gaussian Elimination, and suggests new avenues for implementing or analyzing it.

To the author's frustration, most of the formulas in this paper turned up in the literature after being (re)derived by the author. This experience suggests that it is time for an up-to-date literature survey on Gaussian Elimination.

We have found no modern numerical linear algebra text that mentions results like these in the study of Gaussian Elimination. The results do seem very useful in understanding computational properties of Gaussian Elimination, particularly in combination with recent analyses of the algorithm. For example, Theorem 6 can be useful in obtaining tighter error bounds, and in approximating growth factors with condition numbers. The identities here should help in the ongoing clarification of the circumstances under which concern over roundoff error is warranted, and in attacking other remaining mysteries of Gaussian Elimination.

Also, the Schur complement has gradually gained in recognition as a useful tool in various aspects of matrix analysis. The results collected here provide evidence that this trend should continue.

Acknowledgement

I am indebted to Brad Pierce for his very careful readings of this paper, and many insightful suggestions for improving it. Among other things, it is only because of Brad's patience and persistence that I realized the importance of distinguishing between $a_{ij}^{(k)}$ and $\widehat{a}_{ij}^{(k)}$.

Appendix: Determinantal Formulas for Gaussian Elimination

To prove Theorem 1, we need several 19th-century identities concerning determinants. We take the time to prove them carefully because they are instructive, and they clarify the relationship between Gaussian Elimination and determinants. Also, the proofs dispel the magic that otherwise surrounds these identities.

Theorem 7 *(Chiò, 1853 [4])*
If A is an $n \times n$ Gauss eliminable matrix, then

$$det\ A \quad = \quad \frac{1}{a_{11}^{n-2}}\ det \begin{pmatrix} det\ A[1,2\,|\,1,2] & det\ A[1,2\,|\,1,3] & \cdots & det\ A[1,2\,|\,1,n] \\ det\ A[1,3\,|\,1,2] & det\ A[1,3\,|\,1,3] & \cdots & det\ A[1,3\,|\,1,n] \\ \vdots & \vdots & \ddots & \vdots \\ det\ A[1,n\,|\,1,2] & det\ A[1,n\,|\,1,3] & \cdots & det\ A[1,n\,|\,1,n] \end{pmatrix}.$$

Proof
When A is an $n \times n$ Gauss eliminable matrix, $a_{11} \neq 0$. The theorem can then be proven with invariance of the determinant under elementary row operations. Multiplying all but the first row of A by a_{11} augments the determinant by a_{11}^{n-1}. Then Gaussian Elimination on the first column leaves this determinant value undisturbed.

$$det\ A \quad = \quad \frac{1}{a_{11}^{n-1}}\ det \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{11}a_{21} & a_{11}a_{22} & a_{11}a_{23} & \cdots & a_{11}a_{2n} \\ a_{11}a_{31} & a_{11}a_{32} & a_{11}a_{33} & \cdots & a_{11}a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{11}a_{n1} & a_{11}a_{n2} & a_{11}a_{n3} & \cdots & a_{11}a_{nn} \end{pmatrix}$$

$$= \quad \frac{1}{a_{11}^{n-1}}\ det \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & (a_{11}a_{22}-a_{21}a_{12}) & (a_{11}a_{23}-a_{21}a_{13}) & \cdots & (a_{11}a_{2n}-a_{21}a_{1n}) \\ 0 & (a_{11}a_{32}-a_{31}a_{12}) & (a_{11}a_{33}-a_{31}a_{13}) & \cdots & (a_{11}a_{3n}-a_{31}a_{1n}) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & (a_{11}a_{n2}-a_{n1}a_{12}) & (a_{11}a_{n3}-a_{n1}a_{13}) & \cdots & (a_{11}a_{nn}-a_{n1}a_{1n}) \end{pmatrix}$$

$$= \quad \frac{1}{a_{11}^{n-2}}\ det \begin{pmatrix} det\ A[1,2\,|\,1,2] & det\ A[1,2\,|\,1,3] & \cdots & det\ A[1,2\,|\,1,n] \\ det\ A[1,3\,|\,1,2] & det\ A[1,3\,|\,1,3] & \cdots & det\ A[1,3\,|\,1,n] \\ \vdots & \vdots & \ddots & \vdots \\ det\ A[1,n\,|\,1,2] & det\ A[1,n\,|\,1,3] & \cdots & det\ A[1,n\,|\,1,n] \end{pmatrix}.$$

□

Chiò's pivotal condensation process uses this identity iteratively, repeatedly reducing the evaluation of an $n \times n$ determinant to an $(n-1) \times (n-1)$ determinant. Householder [22, p.141] remarks "the essential identity of Gaussian Elimination and the method of Chió for evaluating a determinant will be recognized by anyone acquainted with both".

When taken to the limit, Chiò's process yields Sylvester's identity:

Theorem 8 *(Sylvester, 1851 [34])*
Let A be an $n \times n$ Gauss eliminable matrix. Then for $1 \le k < i \le n$, $1 \le k < j \le n$:

$$\begin{aligned} & \mathit{det}\ A[\,1..(k-1) \,|\, 1..(k-1)\,] \ \cdot\ \mathit{det}\ A[\,1..(k-1), k, i \,|\, 1..(k-1), k, j\,] \\ & \quad = \ \mathit{det} \begin{pmatrix} \mathit{det}\ A[\,1..(k-1), k \,|\, 1..(k-1), k\,] & \mathit{det}\ A[\,1..(k-1), k \,|\, 1..(k-1), j\,] \\ \mathit{det}\ A[\,1..(k-1), i \,|\, 1..(k-1), k\,] & \mathit{det}\ A[\,1..(k-1), i \,|\, 1..(k-1), j\,] \end{pmatrix}. \end{aligned}$$

Proof
We need only prove the memorable identity obtained with $k = (n-1)$:

$$\begin{aligned} & \mathit{det}\ A[\,1..(n-2) \,|\, 1..(n-2)\,] \ \cdot\ \mathit{det}\ A[\,1..n \,|\, 1..n\,] \\ & \quad = \ \mathit{det} \begin{pmatrix} \mathit{det}\ A[\,1..(n-1) \,|\, 1..(n-1)\,] & \mathit{det}\ A[\,1..(n-1) \,|\, 1..(n-2), n\,] \\ \mathit{det}\ A[\,1..(n-2), n \,|\, 1..(n-1)\,] & \mathit{det}\ A[\,1..(n-2), n \,|\, 1..(n-2), n\,] \end{pmatrix}. \end{aligned}$$

This identity is equivalent to the statement of the theorem, since when $k < (n-1)$ we can replace A with the $(k+1) \times (k+1)$ submatrix $A[\,1..(k-1), k, i \,|\, 1..(k-1), k, j\,]$, and view n as $(k+1)$. (With this in mind, Sylvester stated the theorem in a more general form, using expressions such as $\mathit{det}\ A[\,a_1, \ldots, a_n \,|\, \alpha_1, \ldots, \alpha_n\,]$ where $a_1, \ldots, a_n, \alpha_1, \ldots, \alpha_n$ represent arbitrary indices.)

The theorem can be proved by a careful induction on n. It uses the statement of the theorem as an induction hypothesis, but establishes only the memorable identity above, relying on their equivalence to complete the proof.

For the basis $n = 2$ (forcing $k = 1$), the theorem asserts that

$$\mathit{det}\ A[\,|\,] \cdot \mathit{det}\ A[\,1,2 \,|\, 1,2\,] \quad = \quad \mathit{det}\ A[\,1 \,|\, 1\,] \cdot \mathit{det}\ A[\,2 \,|\, 2\,] \ - \ \mathit{det}\ A[\,2 \,|\, 1\,] \cdot \mathit{det}\ A[\,1 \,|\, 2\,]$$

i.e., simply that $1 \cdot (a_{11}a_{22} - a_{21}a_{12}) = a_{11} \cdot a_{22} - a_{21} \cdot a_{12}$.

For the induction step, the induction hypothesis (the statement of the theorem for $k < (n-1)$) is sufficient to prove the memorable identity. In fact, the final case is the final result of Chiò's process, and this is reflected in its proof. We claim that for $2 \le k \le n-1$:

$$\begin{aligned} \mathit{det}\ A \ &= \ 1/(\mathit{det}\ A[\,1..(k-1) \,|\, 1..(k-1)\,])^{n-k} \\ & \cdot\ \mathit{det} \begin{pmatrix} \mathit{det}\ A[\,1..(k-1), (k) \,|\, 1..(k-1), k\,] & \cdots & \mathit{det}\ A[\,1..(k-1), (k) \,|\, 1..(k-1), n\,] \\ \mathit{det}\ A[\,1..(k-1), (k+1) \,|\, 1..(k-1), k\,] & \cdots & \mathit{det}\ A[\,1..(k-1), (k+1) \,|\, 1..(k-1), n\,] \\ \vdots & \ddots & \vdots \\ \mathit{det}\ A[\,1..(k-1), (n) \,|\, 1..(k-1), k\,] & \cdots & \mathit{det}\ A[\,1..(k-1), (n) \,|\, 1..(k-1), n\,] \end{pmatrix}. \end{aligned}$$

The case $k = 2$ is Theorem 7, and the case $k = (n-1)$ is the memorable identity above, so proving the claim will establish the theorem.

The claim can be proven with a nested induction on k, using Chiò's basis $k = 2$. The induction hypothesis allows us to implement Chiò's process, increasing k from 2 to $n-1$. Each step reduces the right side of the claim, $1/(\mathit{det}\,A[\,1..(k-1) \,|\, 1..(k-1)\,])^{n-k}$ multiplied by the determinant of an $(n-k+1) \times (n-k+1)$ submatrix, to $1/(\mathit{det}\,A[\,1..k \,|\, 1..k\,])^{n-(k+1)}$ multiplied by the determinant of an $(n-k) \times (n-k)$ submatrix. The [1,1] entry of the former submatrix is $\mathit{det}\,A[\,1..k \,|\, 1..k\,]$. Following the proof of Theorem 7, the determinant of this submatrix is $1/(\mathit{det}\,A[\,1..k \,|\, 1..k\,])^{n-(k+1)}$ multiplied by the determinant of an $(n-k) \times (n-k)$ matrix, in which each entry is a 2×2 determinant. The induction hypothesis can then be used to replace the $(i-k+1), (j-k+1)$-th 2×2 determinant (because its entries are determinants of $k \times k$ submatrices with $k < (n-1)$) by the product of two terms, $\mathit{det}\,A[\,1..(k-1) \,|\, 1..(k-1)\,]$ and $\mathit{det}\,A[\,1..(k-1), k, i \,|\, 1..(k-1), k, j\,]$. Factoring out the terms $\mathit{det}\,A[\,1..(k-1) \,|\, 1..(k-1)\,]$ cancels the previous leading multiplicative factor $1/(\mathit{det}\,A[\,1..(k-1) \,|\, 1..(k-1)\,])^{n-k}$ and yields the claim for $k+1$. □

This brings us to the theorem we set out to prove: *all of the results of Gaussian Elimination are ratios of two determinants of submatrices of A.*

Theorem 1 *If A is an $n \times n$ Gauss eliminable matrix, then:*
For $1 \leq k < i \leq n$,

$$\ell_{ik} \quad = \quad \frac{det\ A[\,1..(k-1), i \,|\, 1..k\,]}{det\ A[\,1..k \,|\, 1..k\,]}.$$

For $1 \leq k < i \leq n$, $1 \leq k < j \leq n$,

$$a_{ij}^{(k+1)} \quad = \quad \frac{det\ A[\,1..k,\ i \,|\, 1..k,\ j\,]}{det\ A[\,1..k \,|\, 1..k\,]}.$$

Proof
By simultaneous induction on n. For $n = 2$, we require only verification that

$$\ell_{21} \quad = \quad \frac{a_{21}^{(1)}}{a_{11}^{(1)}} \quad = \quad \frac{a_{21}}{a_{11}} \quad = \quad \frac{det\ A[\,2 \,|\, 1\,]}{det\ A[\,1 \,|\, 1\,]}$$

as stated, and that

$$a_{22}^{(2)} \quad = \quad a_{22}^{(1)} - \frac{a_{21}^{(1)}}{a_{11}^{(1)}} \cdot a_{12}^{(1)} \quad = \quad a_{22} - \frac{a_{21}}{a_{11}} \cdot a_{12} \quad = \quad \frac{a_{11}a_{22} \ - \ a_{21}a_{12}}{a_{11}} \quad = \quad \frac{det\ A[\,1,2 \,|\, 1,2\,]}{det\ A[\,1 \,|\, 1\,]}.$$

The theorem then follows from the induction hypothesis. First,

$$\begin{aligned}
\ell_{ik} &= a_{ik}^{(k)} \ / \ a_{kk}^{(k)} \\
&= \left(\frac{det\ A[\,1..(k-1), i \,|\, 1..(k-1), k\,]}{det\ A[\,1..(k-1) \,|\, 1..(k-1)]} \right) \ / \ \left(\frac{det\ A[\,1..(k-1), k \,|\, 1..(k-1), k\,]}{det\ A[\,1..(k-1) \,|\, 1..(k-1)]} \right) \\
&= \frac{det\ A[\,1..(k-1), i \,|\, 1..(k-1), k\,]}{det\ A[\,1..(k-1), k \,|\, 1..(k-1), k\,]} \\
&= \frac{det\ A[\,1..(k-1), i \,|\, 1..k\,]}{det\ A[\,1..k \,|\, 1..k\,]}
\end{aligned}$$

as claimed. Also,

$$\begin{aligned}
a_{ij}^{(k+1)} &= a_{ij}^{(k)} \ - \ \ell_{ik} \cdot a_{kj}^{(k)} \\
&= \frac{det\ A[\,1..(k-1), i \,|\, 1..(k-1), j\,]}{det\ A[\,1..(k-1) \,|\, 1..(k-1)]} - \frac{det\ A[\,1..(k-1), i \,|\, 1..(k-1), k\,]}{det\ A[1..k \,|\, 1..k]} \cdot \frac{det\ A[\,1..(k-1), k \,|\, 1..(k-1), j\,]}{det\ A[\,1..(k-1) \,|\, 1..(k-1)]} \\
&= \frac{det\ A[\,1..(k-1), i \,|\, 1..(k-1), j\,] \ \cdot \ det\ A[\,1..k \,|\, 1..k] \ - \ det\ A[\,1..(k-1), i \,|\, 1..k] \ \cdot \ det\ A[\,1..k \,|\, 1..(k-1), j\,]}{det\ A[\,1..(k-1) \,|\, 1..(k-1)] \ \cdot \ det\ A[\,1..k \,|\, 1..k]} \\
&= \frac{det\ A[\,1..(k-1) \,|\, 1..(k-1)] \ \cdot \ det\ A[\,1..k, i \,|\, 1..k, j\,]}{det\ A[\,1..(k-1) \,|\, 1..(k-1)] \ \cdot \ det\ A[\,1..k \,|\, 1..k]} \qquad \text{(by Theorem 8)} \\
&= \frac{det\ A[\,1..k, i \,|\, 1..k, j\,]}{det\ A[\,1..k \,|\, 1..k]}.
\end{aligned}$$

□

References

[1] A.C. Aitken, *Determinants and Matrices* (ninth ed.), Edinburgh: Oliver and Boyd, Ltd., 1956.

[2] E.F. Beckenbach, R. Bellman, *Inequalities (2nd revised printing)*, NY: Springer-Verlag, 1965.

[3] D. Carlson, E. Haynsworth, T. Markham, "A generalization of the Schur complement by means of the Moore-Penrose inverse", *SIAM J. Appl. Math.* **8**:1, 169–175, 1974.

[4] F. Chiò, "Mémoire sur les fonctions connues sous le nom de résultantes ou de déterminants". Turin, 32pp, 1853. (citation in [22]) (I could not find this reference; the presentation here is adapted from [9, pp.129–132], [1, pp.45–59], and from [25, pp.79–80].)

[5] R.W. Cottle, "Manifestations of the Schur Complement", *Linear Algebra and its Applications* **8**, 189–211, 1974.

[6] D. Crabtree, E. Haynsworth, "An Identity for the Schur Complement of a Matrix", *Proc. Amer. Math. Soc.* **22**, 364–366, 1969.

[7] J. Day, B. Peterson, "Growth in Gaussian Elimination", *Am. Math. Monthly* **95**, 489–513, 1988.

[8] J. Dongarra, I.S. Duff, D.C. Sorensen, H.A. van der Vorst, *Solving Linear Systems on Vector and Shared Memory Computers*, Philadelphia, PA: SIAM Press, 1991.

[9] H. Eves, *Elementary Matrix Theory*, NY: Dover Books, 1966.

[10] G.E. Forsythe, C.B. Moler, *Computer Solution of Linear Algebraic Systems*, Prentice-Hall, 1967.

[11] F.R. Gantmacher, *The Theory of Matrices, Volume One*. NY: Chelsea Publishing Co, 1959.

[12] C.F. Gauss, *Theoria Motus Corporum Coelestium*, 1809. Reprinted in *Werke* vol. 7, 1870. Gesellschaft der Wissenschaften zu Göttingen.

[13] V.L. Girko, *Theory of Random Determinants*, Boston: Kluwer Academic Publishers, 1990.

[14] G.H. Golub, C.F. Van Loan, *Matrix Computations: Second Edition*, Baltimore: Johns Hopkins University Press, 1989.

[15] N. Gould, "On Growth in Gaussian Elimination with Complete Pivoting", *SIAM J. Matrix Anal. Appl.* **12**:2, 354–361, April 1991.

[16] D.P. Grossman, "On the problem of a numerical solution of systems of simultaneous linear algebraic equations" (in Russian), *Uspekhi Matematicheskikh Nauk* **5**:3(37), 87–103, 1950.

[17] M. Grötschel, L. Lovász, A. Schrijver, *Geometric Algorithms and Combinatorial Optimization*, 2nd corrected edition, Springer-Verlag, 1993.

[18] E. Haynsworth, "Determination of the inertia of a partitioned Hermitian matrix", *Linear Algebra and Appl.* **1**, 73–81, 1968.

[19] E. Haynsworth, "Applications of an inequality for the Schur complement", *Proc. Amer. Math. Soc.* **24**, 512–516, 1970.

[20] E. Haynsworth, "Reduction of a matrix using properties of the Schur complement", *Linear Algebra and Appl.* **3**, 23–29, 1970.

[21] R.A. Horn, C.R. Johnson, *Matrix Analysis*, NY: Cambridge University Press, 1992.

[22] A.S. Householder, *The Theory of Matrices in Numerical Analysis*, NY: Dover Books, 1964.

[23] N.J. Higham, D.J. Higham, "Large Growth Factors in Gaussian Elimination with Pivoting", *SIAM J. Matrix Anal. Appl.* **10**:3, 155–164, April 1989.

[24] H. Hotelling, "Some new methods in matrix inversion", *Ann. Math. Stat.* **14**, 1–34, 1943.

[25] T. Muir, *The Theory of Determinants, in Historical Order of Development; Volume Two: the period 1841 to 1860*, NY: Dover Press, 1960.

[26] J. von Neumann, H.H. Goldstine, "Numerical inverting of matrices of high order", *Bull. Amer. Math. Soc.* **53**, 1021–1099, 1947.

[27] A. Ostrowski, "A new proof of Haynsworth's quotient formula for Schur complements", *Linear Algebra and Appl.* **4**, 389-392, 1971.

[28] G. Poole, L. Neal, "A Geometric Analysis of Gaussian Elimination. I", *Linear Algebra Appl.* **149**, 249–272, 1991.

[29] G. Poole, L. Neal, "Gaussian Elimination: When is Scaling Beneficial?", *Linear Algebra Appl.* **162**, 309–324, 1992.

[30] G. Poole, L. Neal, "A Geometric Analysis of Gaussian Elimination. II", *Linear Algebra Appl.* **173**, 239–264, 1992.

[31] Y. Robert, *The Impact of Vector and Parallel Architectures on the Gaussian Elimination Algorithm*, NY: Halsted Press, 1990.

[32] R.D. Rodman, "A Note on a Set of Test Matrices for Inversion", *Comm. ACM* **6**, 515 (1963).

[33] A. Schrijver, *Theory of Linear and Integer Programming*, NY: J. Wiley & Sons, 1986.

[34] J.J. Sylvester, "On the Relation between the Minor Determinants of Linearly Equivalent Quadratic Functions", *Philosophical Magazine* **1**, 295–305, 1851. Reprinted in *The Collected Mathematical Papers of James Joseph Sylvester*, Volume I (1837–1853), H.F. Baker, ed., 241–250, NY: Chelsea Publishing Company, 1973. See also [1, pp.103–107], and [25, pp.58–61].

[35] J. Todd, "The Condition of Finite Segments of the Hilbert Matrix", in O. Taussky, ed., *Contributions to the Solution of Systems of Linear Equations and the Determination of Eigenvalues*, Washington: National Bureau of Standards, 1954.

[36] L.N. Trefethen, "Three mysteries of Gaussian elimination", *ACM SIGNUM Newsletter* **20**:4, October 1985.

[37] L.N. Trefethen, R.S. Schreiber, "Average-Case Stability of Gaussian Elimination", *SIAM J. Matrix Anal. Appl.* **11**:3, 335–360, July 1990.

[38] A.M. Turing, "Rounding-Off Errors in Matrix Processes", *Quart. J. Mech.* **1**, 287–308, 1948.

[39] J.H. Wilkinson, "Error analysis of direct methods of matrix inversion", *J. ACM* **8**, 281–330, 1961.

[40] J.H. Wilkinson, *Rounding Errors in Algebraic Processes*, NJ: Prentice-Hall, Inc., 1963.

[41] J.H. Wilkinson, *The Algebraic Eigenvalue Problem*, London: Oxford University Press, 1965.

[42] S.J. Wright, "A Collection of Problems for which Gaussian Elimination with Partial Pivoting is Unstable", *SIAM J. Sci. Comput.* **14**:1, 231–238, January 1993.