How ordinary elimination became Gaussian elimination

Joseph F. Grcar

6059 Castlebrook Drive, Castro Valley, CA 94552-1645, USA

Abstract

Newton, in notes that he would rather not have seen published, described a process for solving simultaneous equations that later authors applied specifically to linear equations. This method — which Euler did not recommend, which Legendre called "ordinary," and which Gauss called "common" — is now named after Gauss: "Gaussian" elimination. Gauss's name became associated with elimination through the adoption, by professional computers, of a specialized notation that Gauss devised for his own least-squares calculations. The notation allowed elimination to be viewed as a sequence of arithmetic operations that were repeatedly optimized for hand computing and eventually were described by matrices.
© 2010 Elsevier Inc. All rights reserved.

Zusammenfassung

In Aufzeichnungen, die Newton lieber nicht der Veröffentlichung preisgegeben hätte, beschreibt er den Prozess für die Lösung von simultanen Gleichungen, den spätere Autoren speziell für lineare Gleichungen anwandten. Diese Methode — welche Euler nicht empfahl, welche Legendre "ordinaire" nannte, und welche Gauß "gewöhnlich" nannte — wird nun nach Gauß benannt: Gaußsches Eliminationsverfahren. Die Verbindung des Gaußschen Namens mit Elimination wurde dadurch hervorgebracht, dass professionelle Rechner eine Notation übernahmen, die Gauß speziell für seine eigenen Berechnungen der kleinsten Quadrate ersonnen hatte, welche zuließ, das Elimination als eine Sequenz von arithmetischen Rechenoperationen betrachtet wurde, die wiederholt für Handrechnungen optimisiert wurden und schließlich durch Matrizen beschrieben wurden.
© 2010 Elsevier Inc. All rights reserved.

1. Introduction

1.1. Overview

The familiar method for solving simultaneous linear equations, Gaussian elimination, originated independently in ancient China and early modern Europe. The details of the

Chinese development may be lost because of a tradition of anonymity in the primary text. In contrast, the more recent European origin is traceable through its appearance in comparatively few significant documents produced over a period of some four hundred years. This paper explains the European history of Gaussian elimination up to the invention of electronic computers.

The development has three phases: first came the "schoolbook" rule, second were methods that professional hand computers used primarily for least-squares work, and third was the adoption of matrix notation, which reconciled the schoolbook and professional methods. Section 1 summarizes the history and connects it to the present by documenting where in education and technology Gaussian elimination is found today.

Section 2 examines phase one, the period before Gauss during which Gaussian elimination was invented and became a staple of algebra textbooks. Beginning with a discussion of the Chinese origin of the technique, it continues to the European origin in the late Renaissance and an examination of the contributions of Newton and Rolle.

Section 3 introduces phase two, the period during which the need to solve simultaneous linear equations, and solve them repeatedly, arose for a clear social purpose: statistical inference. After Legendre and Gauss invented the method of least squares, Gauss started a tradition of seeking to improve the methods of calculation.

Section 4 continues phase two into the period after Gauss and describes the efforts of professional hand computers to ease their labor. This section focuses on the work of three men: Myrick Doolittle, a computer working at the United States Coast and Geodetic Survey; André Cholesky, a French military geodesist; and Prescott Crout, a professor of mathematics at the Mathematics Institute of Technology.

Section 5 treats phase three, the matrix explications of Gaussian elimination that arose for investigative purposes in the first half of the 20th century. It was understood from the matrix descriptions that the schoolbook methods of phase one, and the professional methods of phase two, were all essentially related. This phase begins with the Cracovian algebra of the astronomer, geodesist, and mathematician Tadeusz Banachiewicz. The latter was an important influence on the geodesist Henry Jensen, whose contribution is also discussed, along with those of the statistician Paul Sumner Dwyer and the mathematician Ewald Bodewig. The section concludes with a short discussion of the work of John von Neumann and Herman Goldstine.

1.2. Gaussian elimination today

Both elementary and advanced textbooks discuss Gaussian elimination. For example, the precalculus algebra textbook of Cohen et al. [2006, 743–747, Sect. 10.2] shows how "elementary operations" on equations produce an "equivalent system" in "upper-triangular form" that can be solved by "back-substitution" (Fig. 1, Eq. (1)). The distinguishing features are as follows.

1. The equations and variables may be rearranged so that the leading equation contains whichever variable is chosen to be the leading variable for immediate elimination.[1]
2. The leading equation is used (in various ways) to remove the leading variable from each of those following.

---

[1] A large coefficient is recommended for the leading variable in the leading equation.

Schoolbook method:

$$
\left.\begin{array}{r}
x + 2y + z = 3 \\
x + y + 2z = 9 \\
2x + y + z = 16
\end{array}\right\} \Rightarrow
\left.\begin{array}{r}
x + 2y + z = 3 \\
-y + z = 6 \\
-3y - z = 10
\end{array}\right\} \Rightarrow
\left.\begin{array}{r}
x + 2y + z = 3 \\
-y + z = 6 \\
-4z = -8
\end{array}\right\} \Rightarrow
\left.\begin{array}{r}
z = 2 \\
y = -4 \\
x = 9
\end{array}\right.
\tag{1}
$$

"Classic" professional method:

$$
\left.\begin{array}{c}
\begin{bmatrix} 1 & & \\ & 1 & \\ & 3 & 1 \end{bmatrix}^{-1}
\begin{bmatrix} 1 & & \\ 1 & 1 & \\ 2 & & 1 \end{bmatrix}^{-1}
\begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 2 \\ 2 & 1 & 1 \end{bmatrix} =
\begin{bmatrix} 1 & 2 & 1 \\ & -1 & 1 \\ & & -4 \end{bmatrix} \\[4pt]
\Downarrow \\[2pt]
\begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 2 \\ 2 & 1 & 1 \end{bmatrix} =
\begin{bmatrix} 1 & & \\ 1 & 1 & \\ 2 & 3 & 1 \end{bmatrix}
\begin{bmatrix} 1 & 2 & 1 \\ & -1 & 1 \\ & & -4 \end{bmatrix}
\end{array}\right\} \Rightarrow
\begin{array}{c}
\begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 2 \\ 2 & 1 & 1 \end{bmatrix}
\begin{bmatrix} x \\ y \\ z \end{bmatrix} =
\begin{bmatrix} 3 \\ 9 \\ 6 \end{bmatrix} \\[4pt]
\Downarrow \\[2pt]
\begin{bmatrix} 1 & & \\ 1 & 1 & \\ 2 & 3 & 1 \end{bmatrix}
\begin{bmatrix} 1 & 2 & 1 \\ & -1 & 1 \\ & & -4 \end{bmatrix}
\begin{bmatrix} x \\ y \\ z \end{bmatrix} =
\begin{bmatrix} 3 \\ 9 \\ 6 \end{bmatrix}
\end{array}
\tag{2}
$$

Fig. 1. Gaussian elimination in precalculus algebra and as currently understood by computing specialists.

3. These steps are then applied recursively to the following (modified) equations.
4. Once a single equation remains, the back-substitution is made.

The form of the algorithm employed in Eq. (1) is viewed as canonical: the leading equation remains unchanged while variables are removed by subtracting appropriate multiples of it from each equation below. This paper follows current usage by referring to any algorithm that is essentially equivalent to Eq. (1) as "Gaussian elimination," whatever its period or source; the phrase should not be interpreted as an attribution.

In contrast, the technical literature views Gaussian elimination as a method for factoring matrices. Elementary operations "reduce" the coefficient matrix of Eq. (1) to an "upper-triangular" matrix, thereby accomplishing a "triangular factorization," or "decomposition," from which the equations can be solved by "forward substitution" followed by backward substitution (Fig. 1, Eq. (2), but note that the arithmetical operations are identical to those in Eq. (1)). The advanced textbook of Petersen and Arbenz [2004, 107] explains that this algorithm is the standard test for the speed of computers in scientific work. Its widespread use in so large a field as scientific computing results in many algorithmic variants that are collectively called, simply, Gaussian elimination. The variations are designated by acronyms, adjectives, and eponyms. At this level of differentiation the version of Eq. (1) is named either "classic" elimination or "Doolittle's method." Nevertheless, advanced or specialized texts always begin by identifying exactly this algorithm as Gaussian elimination.[2]

Today, textbooks at all levels encourage the belief that Gauss introduced the method of Eq. (1) and that his work was somehow remarkable when compared to prior art. For example, Cohen et al. [2006, 743] claim that "Gauss used this technique" to analyze the orbit of

---

[2] For example, see Duff et al. [1986, 43–44], Golub and Van Loan [1996, 94], Higham [2002, 158–160], Petersen and Arbenz [2004, 23–25], and Stewart [1998, 148–154].

Pallas [Gauss, 1810], although "the essentials" had already appeared in ancient Chinese texts. Higham [2002, p. 187] acknowledges that "a variant" was used by the Chinese, but he finds "the first published appearance of Gaussian elimination" in another paper of Gauss [1809]. Some authors do allude to a possible European origin before Gauss. Farebrother [1988, 3] makes the remark that "Gauss's formalization" of the "traditional schoolbook method" appeared in his Pallas work. These inconsistent attributions raise the question: what exactly did Gauss contribute?

2. Before Gauss

2.1. Gaussian elimination in antiquity or not

Periods before Gauss are surveyed here to see where and when Gaussian elimination may have developed. It will be seen that consideration of any simultaneous linear equations is comparatively rare in the primary sources. Moreover, almost all early sources consist of terse "word problems" rather than discourses on solution methods.

The earliest written mathematics is believed to be "Babylonian" cuneiform tablets from the Tigris and Euphrates valley. Robson [2008, 343–344] counts around a thousand published tablets that record mathematics taught in scribal schools. Her inventory shows that most of these tablets contain arithmetical or metrological tables. Tablets VAT 8389 and 8391, from the Old Babylonian period, 2000–1600 BC, contain what are believed to be among the earliest problems that can be interpreted as systems of linear equations.[3]

The actual text of problem 1 on VAT 8389 and a literal restatement in modern English are given by Høyrup [2002, 77–82]. For brevity, Eq. (3) of Fig. 2 expresses the solution symbolically; see also Friberg [2007b, 334–335] and Katz [1998, 16]. The frequent admonition by the anonymous scribe to "keep it in your head" [see Høyrup, 2002, 79] is poignant evidence of the computing technology available to him. Høyrup [2002, 82] comments that the solution method is pedagogically superior to the method of double false position found later. Friberg [2007b, 334] reports that relatively few tablets pose linear problems. Indeed, the tablets summarized by Bashmakova and Smirnova [2000, 3] suggest that when "Babylonian" problems can be interpreted as simultaneous equations, usually one condition is nonlinear.

One of the main sources for Egyptian mathematics is the Rhind papyrus, which dates from around 1550 BC.[4] Problem 40 in the Rhind papyrus (Fig. 2, Eq. (4)) is sometimes mentioned as an early example of simultaneous linear equations [Dedron and Itard, 1959, 303]. Gillings [1972, 170–172], however, argues that the solution is based on knowledge of arithmetic progressions, the latter being a distinguishing feature of mathematics in ancient Egypt.

By far the most impressive treatment of simultaneous linear equations known from antiquity — and the only lengthy presentation prior to the late 17th century — is Chapter 8

---

[3] This dating is the widest possible based on Robson [1999, 6, 321]. Friberg [2007a, 440–441] explains the acquisition of VAT 8389 and 8391 by the Vorderasiatisches Museum in Berlin and their lost provenance. Neugebauer [1973, II, Pl. 24, 26] provides photographs of the tablets.

[4] The Rhind papyrus originated in Thebes and is now in the British Museum. For dating, provenance, and a view of the papyrus, see the museum Web site: `http://www.britishmuseum.org/explore/highlights/highlight_objects/aes/r/rhind_mathematical_papyrus.aspx`.

2000–1600 BC: Problem 1 on tablet VAT 8389: the total area of two fields is 1800 sar, the rent for one is 2 silà of grain per 3 sar, for the other is 1 silà per 2 sar, the total rent on the first exceeds that on the other by 500 silà.

$$
\begin{aligned}
x + y &= 1800 \\
\tfrac{2}{3} \cdot x - \tfrac{1}{2} \cdot y &= 500
\end{aligned}
\Rightarrow
\begin{aligned}
(900 + s) + (900 - s) &= 1800 \\
\tfrac{2}{3} \cdot (900 + s) - \tfrac{1}{2} \cdot (900 - s) &= 150 + \tfrac{7}{6} \cdot s \Rightarrow s = \tfrac{6}{7} \cdot 350, \text{ hence } x, y
\end{aligned}
\tag{3}
$$

1550 BC: Problem 40 of the Rhind papyrus: divide 100 into 5 parts so the common difference is the same and the three largest are 7 times the other two [Gillings, 1972, 170–172]. Solution: an arithmetic progression is understood from the phrase "common difference." The scribe sums $\left\{23, 17\tfrac{1}{2}, 12, 6\tfrac{1}{2}, 1\right\}$, which he somehow knows beforehand satisfies the $7 : 1$ condition, and upon obtaining 60, he then scales the terms by $\tfrac{5}{3}$ to enforce the 100 condition.

$$
\underbrace{23 + 17.5 + 12}_{52.5 = 7 \times 7.5} + \underbrace{6.5 + 1}_{7.5} = 60 \quad \Rightarrow \quad \left\{ 38\tfrac{1}{3}, \ 29\tfrac{1}{6}, \ 20, \ 10\tfrac{5}{6}, \ \tfrac{5}{3} \right\}
\tag{4}
$$

Before 200 AD: Problem 1 in chapter 8 of the *Nine Chapters* [Kangshen et al., 1999, 391 and 399–403]: from 3 top-grade rice paddies, 2 medium-grade, and 1 low-grade, the combined yield is 39 dou of grain, and so on for two more collections of paddies. What is the yield of a paddy of each grade? Solution:

$$
\begin{aligned}
3x + 2y + z &= 39 \\
2x + 3y + z &= 34 \\
x + 2y + 3z &= 26
\end{aligned}
\iff
\begin{matrix}
1 & 2 & 3 \\
2 & 3 & 2 \\
3 & 1 & 1 \\
26 & 34 & 39
\end{matrix}
\Rightarrow
\begin{matrix}
 & 3 & \\
4 & 5 & 2 \\
8 & 1 & 1 \\
39 & 24 & 39
\end{matrix}
\Rightarrow
\begin{matrix}
 & 3 & \\
 & 5 & 2 \\
36 & 1 & 1 \\
99 & 24 & 39
\end{matrix}
\Rightarrow
\begin{aligned}
z &= \tfrac{11}{4} \\
y &= \tfrac{17}{4} \\
x &= \tfrac{37}{4}
\end{aligned}
\tag{5}
$$

250 AD: Problem 19 in book 1 of the *Arithmetica* of Diophantus [Heath, 1910, 136]: find four numbers such that the sum of any three exceeds the fourth by a given amount. Solution (using symbols for clarity): let n_i be the numbers, $2h$ the sum of all four, and e_i the excesses. Then $2h - n_i = n_i + e_i$ or $n_i = h - e_i/2$. Summing the latter for all four numbers gives $2h = 4h - (e_1 + e_2 + e_3 + e_4)/2$. Thus h and then the n_i can be evaluated. As the problem would be expressed today:

$$
\begin{aligned}
-n_1 + n_2 + n_3 + n_4 &= e_1 \\
n_1 - n_2 + n_3 + n_4 &= e_2 \\
n_1 + n_2 - n_3 + n_4 &= e_3 \\
n_1 + n_2 + n_3 - n_4 &= e_4
\end{aligned}
\quad \Rightarrow \quad
n_i = \frac{e_1 + e_2 + \cdots + e_n}{4} - \frac{e_i}{2}
\tag{6}
$$

499 AD: Problem 29 in chapter 2 of the *Āryabhaṭīya* of Āryabhaṭa [Clark, 1930, 40; Plofker, 2009, 134]: to find several numbers when the results of subtracting each from their sum are known. Solution: sum the known differences and divide by the quantity of terms less one. The result is the sum of all the numbers, from which they can be determined.

$$
\begin{aligned}
0 + n_2 + n_3 + \cdots + n_n &= d_1 \\
n_1 + 0 + n_3 + \cdots + n_n &= d_2 \\
n_1 + n_2 + 0 + \cdots + n_n &= d_3 \\
\vdots \quad \vdots \quad \vdots \quad \ddots \quad \vdots \quad \vdots \\
n_1 + n_2 + n_3 + \cdots + 0 &= d_n
\end{aligned}
\quad \Rightarrow \quad
n_i = \frac{d_1 + d_2 + \cdots + d_n}{n - 1} - d_i
\tag{7}
$$

Fig. 2. Representations of all the known ancient problems, and their solutions, that might be interpreted in symbolic algebra as simultaneous linear equations.

of the *Jiuzhang Suanshu* (*Nine Chapters of the Mathematical Art*), a problem "book" anonymously and collectively written in China. An early signed commentary, including comments on Chapter 8, was written by Liu Hui, who is known to have lived in the 3rd century AD. The method for solving linear problems is therefore at least as old, and Hui asserts that it is much older, although no original text survives. According to Martzloff [1997, 128–131], Chapters 1–5 are known from a 13th-century copy, and Chapters 6–9 are reconstructed from 18th-century quotations of a lost 15th-century encyclopedia.

Mathematicians in ancient China represented numbers by counting rods. They organized elaborate calculations by placing the rods inside squares arranged in a rectangle.[5] In Chapter 8 of the *Nine Chapters*, each column of squares corresponds to a modern linear equation, so in order to obtain the coefficient tableau of modern mathematics, the ancient rectangle must be rotated counterclockwise by 90°. Problem 1 in Chapter 8 is frequently displayed as representative of the ancient calculation (Fig. 2, Eq. (5)). Chapter 8 includes the solution of 18 different systems of equations, all treated in the same systematic way. Unlike Eq. (1), when two columns of the table are combined to remove a number, each column is multiplied by the leading number in the other. Subtracting the right column from the left column then removes the leading number from the left while preserving integer coefficients.[6]

The main extant source for Greco-Roman algebra is the *Arithmetica* by Diophantus. An approximate date of composition is usually given as around 250 AD, but Schappacher [2005, 3, 9] believes Diophantus could have lived as late as 400 AD, and in any case, the sources of the extant texts cannot be traced any earlier than the 9th century.[7]

Heath [1910, 59] quotes Diophantus for a rule to solve problems with one unknown and one condition, which need not be linear: "If a problem leads to an equation in which any terms are equal to the same terms but have different coefficients, we must take like from like." The use of the word "equation" is of course an anachronism. Many problems in the *Arithmetica* involve more than one unknown. Heath writes that Diophantus "shows great adroitness" in expressing all the unknowns in terms of a single newly introduced value, "the ultimate result being that it is only necessary to solve a simple equation with one unknown quantity." The process is illustrated by the verbal solution given by Diophantus for Eq. (6) in Fig. 2. This problem is among the most elaborate of the few in the *Arithmetica* that are purely linear. As Heath suggests, the problem involves a special linear system whose solution is a repetitive formula and is thus amenable to special reasoning that would not suffice for the more general problems of the *Nine Chapters*.

According to Plofker [2009, 122, 317–318], the earliest surviving work of ancient Hindu mathematics is the *Āryabhaṭīya* of Āryabhaṭa who is known to have been born in 476 AD. Āryabhaṭa's linear problems are reminiscent of the *Arithmetica* but are stated more generally for any quantity of unknowns (Fig. 2, Eq. (7)). The translation by Clark [1930] has Āryabhaṭa explaining how to calculate the solution but not how to find it. The repetitive

---

[5] The Romans also arranged complicated calculations in tabular form, apparently using pebbles in place of rods. Neither Chinese nor Roman artifacts survive for these "table" calculations. Both were supplanted by abaci.

[6] For more information on Chapter 8, see Lay-Yong and Kangshen [1989].

[7] Of 13 supposed books in the *Arithmetica*, 1–6 are known from several copies in the original Greek Heath, 1910, 5], while 4–7 are known in Arabic translation [Sesiano, 1982], and the rest are lost. Three manuscripts are at the Biblioteca Apostolica Vaticana in Rome: Vat.gr.191, ff. 360–390v, Vat.gr.200, ff. 1–193, and Vat.gr.304, ff. 77–117v. Four manuscripts are at the Bibliothèque nationale de France in Paris: BnF grec 2378, 2379, 2380, and 2485.

conditions of the problem allow it to be reduced to a single unknown, as in the case of the problems of Diophantus.

In the medieval period, the equivalent of single polynomial equations, but notably not simultaneous equations, were solved by several Arabic-speaking mathematicians. Examples are found in the work of the encyclopedist Al-Khwarizmi of Baghdad and the writings of Leonardo of Pisa (Fibonacci), who travelled in the western Arabic world [Hogendijk, 1994, 78].

Symbolic algebra developed during the European Renaissance in the arithmetization of geometry and in the theory of equations. By the end of the 16th century, an audience had developed for textbooks that taught arithmetic, how to express "questions" (word problems in modern parlance) as symbolic equations, and methods for solving them. To obtain a comprehensive picture of algebra in the late Renaissance, Kloyda [1938] surveyed what may be all the algebra texts printed between 1550 and 1660. Only four of the 107 texts she examined discuss simultaneous linear equations, which suggests that such equations were of little interest during this period.

The four authors found by Kloyda who did solve linear systems display a rapid exploration of the possibilities of symbolic algebra. Peletier du Mans [1554, 107–112] illustrates the power of symbolic reasoning by solving a problem about three men and their money in two different ways. Peletier attributed the problem to Cardano [1545, 71–73], where it can be found solved by a mixture of verbal reasoning and equations in two variables. Peletier reviewed that argument, and then passed to a purely symbolic solution beginning from three variables and three equations that most directly express the conditions of the problem (Fig. 3, Eq. (8)). For no apparent reason, he took seven steps where three are needed. Buteo [1560] solved a similar problem by Gaussian elimination using three steps (Fig. 3, Eq. (9)). The method was the same double–multiply elimination used in the *Nine Chapters*. The solution is from the sixth, fifth, and third equations, whereas the *Nine Chapters* and the canonical elimination of Eqs. (1) would retain the sixth, fourth, and first equations. Buteo [1560, 193–196] also solved a problem with four equations and unknowns that is repeated by Gosselin [1577, 82 reverse to 84 obverse]. Eighty years later, Rahn [1659, 85–86] solved a system of three linear equations. The four authors found by Kloyda provided examples of Gaussian elimination, but they did not explicitly state the algorithm.

The few problems discussed here and shown in Figs. 2 and 3 are representative of the systems of linear equations solved from ancient times up until the latter years of the 17th century. During this period, simultaneous linear equations were seldom considered, and when they were, Gaussian elimination was absent except in the *Nine Chapters* and in the work of Peletier, Buteo, Gosselin, and Rahn. Since the latter are unlikely to have had been in receipt of communications from China, Gaussian elimination apparently developed independently in Europe as a natural exercise in symbolic algebra. A statement of the method and uses for it were yet to come.

2.2. *"This bee omitted by all that have writ introductions"*

Isaac Newton's work on algebra extended roughly from his appointment to the Lucasian professorship in 1669 until he began composing "the Principia" in 1684.[8] In 1669–1670 Newton wrote amendments for a Latin version of a Dutch algebra text, authored by

---

[8] Newton's mathematical papers have been published by Whiteside [1964–1967, 1968–1982]. Especially relevant to simultaneous equations in the latter set are Vol. 2 for the years 1667–1670 and Vol. 5 for 1683–1684.

Problem and solution of Peletier du Mans [1554, 111–112], with back substitution from the 10th, 5th, and 1st equations:

$$
\begin{array}{rll}
& 1. & 2R + A + B = 64 \\
& 2. & R + 3A + B = 84 \\
& 3. & R + A + 4B = 124 \\[4pt]
2. + 3. \Rightarrow & 4. & 2R + 4A + 5B = 208 \\
4. - 1. \Rightarrow & 5. & 3A + 4B = 144 \\
1. + 2. \Rightarrow & 6. & 3R + 4A + 2B = 148 \\
1. + 3. \Rightarrow & 7. & 3R + 2A + 5B = 188 \\
6. + 7. \Rightarrow & 8. & 6R + 6A + 7B = 336 \\
6 \times 3. \Rightarrow & 9. & 6R + 6A + 24B = 744 \\
9. - 8. \Rightarrow & 10. & 17B = 408
\end{array}
\tag{8}
$$

Problem and solution of Buteo [1560, 190–191], with back substitution from the 6th, 5th, and 3rd equations:

$$
\begin{array}{rll}
& 1. & 3A + B + C = 42 \\
& 2. & A + 4B + C = 32 \\
& 3. & A + B + 5C = 40 \\[4pt]
3 \times 2. - 1. \Rightarrow & 4. & 11B + 2C = 54 \\
3 \times 3. - 1. \Rightarrow & 5. & 2B + 14C = 78 \\
11 \times 5. - 2 \times 4. \Rightarrow & 6. & 150C = 750
\end{array}
\tag{9}
$$

Fig. 3. In the compilation by Kloyda [1938], these are two of the apparently only four Renaissance examples of solving linear systems by symbolic means, restated in modern notation. Peletier wrote an infix "p." (piu) for modern +, also "m." (meno) for −, and for equality he wrote the word. Buteo wrote "." or "," for +, and "[" for =. The problems appear to have been chosen to avoid negatives.

Kinckhuysen [1661], which John Collins planned to publish in England. Collins abandoned the project when other books appeared.[9] Newton himself lectured on algebra at Cambridge for 11 years, beginning with the 1673–1674 academic term.[10] During that time, he wrote and repeatedly revised an incomplete manuscript for his own algebra treatise, which was to be named *Arithmeticæ Universalis*. His last algebra manuscript was prepared in 1684 when, for unknown reasons, he suddenly honored the requirements of the Lucasian professorship by depositing with Cambridge University his lectures for the algebra course. The bulk of the notes were transcribed by his secretary from his previous algebra manuscripts. After Newton left academic life, his lectures were published in their original Latin (1707, 1722) and in translation (1720, 1728) under the intended title of his aborted treatise, *Universal Arithmetic*. Newton had no claim to material that the university had paid him to prepare; nevertheless, as explained by Whiteside [1968–1982, V, 8–11], he strongly objected to its publication, in case it should be misinterpreted as representing his latest research.[11] The second English edition, incorporating changes from Newton, appeared the year after his death.

---

[9] A cursory inspection of Google Books reveals that between 1650 and 1750 at least 40 algebra textbooks were published in England.
[10] Clark [2006, 82] describes the academic politics peculiar to Cambridge and Oxford that resulted in professorial lectures, including Newton's, being given "to the walls."
[11] The title pages of Newton [1707, 1720] pointedly identify no author.

Thanks to Whiteside's impressive scholarship in preparing Newton's papers, a portion of the textbook that is relevant to Gaussian elimination can be traced directly to Newton. In the much earlier commentary on Kinckhuysen, Newton scribbled a marginal note to Collins saying he intended to rectify a lacuna in contemporary textbooks: an explanation for solving collections of equations.

> Though this bee omitted by all that have writ introductions to this Art, yet I judge it very propper & necessary to make an introduction compleate.
>
> — Isaac Newton, marginal note, circa 1669–1670
> quoted by Whiteside [1968–1982, II, 400, n. 62]

Newton proposed to insert a new chapter that first explained the overall strategy of solving simultaneous equations and then listed the tactics by which it might be accomplished.[12]

> **Of the Transformation of two or more Æquations into one, in order to exterminate the unknown Quantities.**
>
> When, in the Solution of any Problem, there are more Æquations than one to comprehend the State of the Question, in each of which there are several unknown Quantities; those Æquations (two by two, if there are more than two) are to be so connected, that one of the unknown Quantities may be made to vanish at each of the Operations, and so produce a new Æquation.... And you are to know, that by each Æquation one unknown Quantity may be taken away, and consequently, when there are as many Æquations and unknown Quantities, all at length may be reduc'd into one, in which there shall be only one Quantity unknown.
>
> — Newton [1720, 60–61] and prior years

On the pages following this rule, Newton offered several methods for removing a variable from two equations, including "equating" and "substituting."

> **The Extermination of an unknown Quantity by an Equality of its Values.**
>
> When the quantity to be exterminated is only of one Dimension in both Æquations, both its Values are to be sought by the Rules already deliver'd, and the one made equal to the other.
>
> Thus, putting $a + x = b + y$ and $2x + y = 3b$, that y may be exterminated, the first Equation will give $a + x - b = y$, and the second will give $3b - 2x = y$. Therefore $a + x - b = 3b - 2x, \ldots$
>
> **The Extermination of an unknown Quantity, by substituting its Value for it.**
>
> When, at least, in one of the Æquations the Quantity to be exterminated is only of one Dimension, its Value is to be sought in that Æquation, and then to be substituted in its Room in the other Æquation. ...
>
> — Newton [1720, 61–62] and prior years

The context and accompanying examples make clear that Newton meant a general approach for solving simultaneous *nonlinear* equations. Newton considered only a few linear systems. In addition to the illustration quoted above, the textbook has a problem about the composition of alloys that is solved by extermination through equality of values [Newton, 1720, 75–77].[13]

---

[12] The new chapter for Kinckhuysen's textbook can be found in [Whiteside, 1968–1982, II, 400–411, parallel Latin and English texts]. The material appears also in the transcribed lecture notes [Whiteside, 1968–1982, V, 122–129, parallel texts]. That text was copied into the Latin and English editions of the unauthorized textbook.

[13] Newton's unpublished treatise has three linear problems about the cost of some goods (wheat, wine, and silk), about money divided among beggars, which also appears in the textbook, and the alloy problem [Whiteside, 1968–1982, V, 567–573, Probs. 3, 4, 7].

		Columne de direction		Columne de retour		
		Premiere Classe		*Premiere Classe*		
	1)	$x + y + z = 6$		$y = 2$	(1	$\Leftarrow y$ from left column 10
	2)	$x + y + v = 7$		$z = 1 + y$	(2	$\Leftarrow z$ from left column 8
	3)	$x + z + v = 8$		$v = 1 + z$	(3	$\Leftarrow v$ from left column 5
	4)	$y + z + v = 9$		$x = 6 - y - z$	(4	$\Leftarrow x$ from left column 1
		Seconde Classe		*Seconde Classe*		
x from 1 sub. into 2 $\Rightarrow$	5)	$v - z = 1$		$z = 3$	(5	$\Leftarrow y$ from 1 sub. into 2
x from 1 sub. into 3 $\Rightarrow$	6)	$v - y = 2$		$v = 1 + z$	(6	$\Leftarrow y$ from 1 sub. into 3
x from 1 sub. into 4 $\Rightarrow$	7)	$y + z + v = 9$		$x = 4 - z$	(7	$\Leftarrow y$ from 1 sub. into 4
		Troisiéme Classe		*Troisiéme Classe*		
v from 5 sub. into 6 $\Rightarrow$	8)	$z - y = 1$		$v = 4$	(8	$\Leftarrow z$ from 5 sub. into 6
v from 5 sub. into 7 $\Rightarrow$	9)	$y + 2z = 8$		$x = 1$	(9	$\Leftarrow z$ from 5 sub. into 7
		Quartriéme Classe		*Quartriéme Classe*		
z from 8 sub. into 9 $\Rightarrow$	10)	$3y = 6$		$x = 1$	(10	$\Leftarrow v$ from 8 sub. into 9

Fig. 4. Gaussian elimination as performed by Michel Rolle [1690, 47]. Rolle did not use "=" for equality.

Newton's comparatively accessible algebra textbook became, as characterized by Whiteside [1968–1982, V, 54–55, n. 1], "the most widely read and influential of his writings." Macomber [1923, 143–144] notes that "before the death of Newton there came to be a demand for suitable text books of algebra for the public schools; and during the 18th century, a number of texts appeared, all more closely resembling the algebra of Newton than those of earlier writers." Further, Macomber [1923, 132] finds that Newton's rule for solving simultaneous equations is "the earliest appearance of this method on record."[14] She refers to Newton's statement of Gaussian elimination, but not by that name, because it had not yet been named for Gauss, even as late as 1923.

During the 37 years when Newton's work lay unpublished, it was overtaken by another development. Michel Rolle [1690, 42–43] explained how to solve simultaneous, specifically linear, equations by "la Methode" of precise rules for Gaussian elimination. He arranged the calculation in two columns to highlight similarities in the two different phases of the work: algebra on the left, where he substituted formulas into variables, and arithmetic on the right, where he substituted numbers into variables (Fig. 4). The left column begins with the original equations, and records successively smaller systems by using the first equation in each system to remove the same variable from the other equations in that system. The right column begins with the first row of each reduced system, and chooses substitutions using the same sequence of rows as on the left. Rolle worked entirely in terms of substitution.

It is difficult to see Rolle's influence in subsequent algebra textbooks. His emphasis on substitution may survive in the "method of substitution," as Gaussian elimination is sometimes called, and the name for his second column, *retour*, may be present in *back-* or *backward* substitution, but these are speculations. Authors of influential French textbooks near the turn of the century, such as Bézout [1788], Lacroix [1800], and Bourdon [1828], appear to owe more to Newton than to Rolle.

---

[14] Macomber supported her claim by her own survey of algebra textbooks from the sixteenth to eighteenth century, which was corroborated by Kloyda [1938]. As noted, Newton himself in 1669–1670 believed that he had supplied the missing lesson, although his text remained unpublished until 1707. Macomber knew of Rolle's work of 1690, but she evidently believed that Newton's approach more resembled what was taught in early 20th-century schoolbooks.

2.3. Gaussian elimination in textbooks after Newton

The influence of Newton's textbook can be traced in three signature features. First is a series of lessons for removing one variable from a pair of equations. The practice of devoting separate lessons to each elimination process became a common pedantry. Newton and Rolle taught substitution, and Newton also had "extermination of an unknown quantity by an equality of its values," which became "comparison." Later authors added the rule used in the *Nine Chapters* that forms linear combinations of two equations, called "addition and/or subtraction." Second is a separate lesson for more than two simultaneous equations. For Newton it was a recursive rule for any number and all kinds of equations. Some later authors limited the rule to linear equations, repeated it for each type of elimination, or built on the two-equation case with separate rules for three and four equations. Third is Newton's use of the word *exterminate*. Whiteside [1968–1982, II, 401, n. 63] reports that Newton first wrote "elimino" and then replaced it in almost all instances by "extermino." The use of this terminology grew from the 18th to the 19th centuries and then abruptly vanished.

Among the earliest authors influenced by Newton was the banker Nathaniel Hammond [Macomber, 1923, 44]. He served as chief accountant for the Bank of England from 1760 to 1768 [Roberts and Kynaston, 1995], and his successful algebra textbook went to four editions between 1742 and 1772. Hammond got down to business by emphasizing his clear instructions for solving word problems.

> As the principal Difficulty in this Science, is acquiring the Knowledge of solving of Questions, I have given a great Variety of these in respect to Numbers and Geometry, and their solutions I chose to give in the most particular, distinct, and plain Manner; and for which the Reader will find full and explicit Directions.
>
> — Hammond [1742, vii]

Hammond [1742, 142, 219–220, and 296–297] may have been the first to replace Newton's recursive rule by a progression of rules for two, three, and four equations. The last of these is

> **The Method of resolving Questions, which contain four Equations, and four unknown Quantities.**
> 72. When the Question contains four Equations, and there are four unknown Quantities in each Equation; find the Value of one of the unknown Quantities in one of the given Equations, and for that unknown Quantity in the other three Equations write the Value of it, which then reduces the Question to three Equations, and three unknown Quantities.
>   Then find the Value of one of these three unknown Quantities in one of these three Equations, and for that unknown Quantity in the other two Equations write the Value of it, which reduces the Question to two Equations, and two unknown Quantities.
>   Then find the Value of one of the unknown Quantities in each of these two Equations, and make these Equations equal to one another, when we shall have an equation with only one unknown Quantity, which being reduced, will answer the Question. . . .
>   And in the same Manner may any other Question in the like Circumstances be answered.
>
> — Hammond [1742, 296–298]

The method for removing variables from two equations was equality of values, and from three or four equations it was substitution. Hammond [1742, 142], like Newton, used the word *exterminating*.

Thomas Simpson's algebra text is another early and very prominent example of Newton's influence. This popular textbook appeared in ten editions from 1745 to 1826. The definitive second edition states Newton's recursive rule separately for each type of elimination, and uses Newton's favored word:

> **Of the Extermination of unknown quantities, or the reduction of two, or more equations, to a single one.**
>
> …. This, in most cases, may be performed various ways, but the following are the most general.
>
> 1°. Observe which, of all your unknown quantities, is the least involved, and let the value of that quantity be found in each equation … looking upon all the rest as known; let the values thus found be put equal to each other …; whence new equations will arise, out of which the quantity will be totally excluded; with which new equations the operation may be repeated, and the unknown quantities exterminated, one by one, till, as last, you come to an equation containing only one unknown quantity.
>
> 2°. Or, let the value of the unknown quantity, which you would first exterminate, be found in that equation where in it is the least involved, considering all the other quantities as known, and … be substituted … in the other equations; and with the new equations thus arising repeat the operation, till you have only one unknown quantity in one equation.
>
> 3°. Or lastly, let the given equations be multiply'd or divided by such numbers or quantities, whether known or unknown, that the term which involves the highest power of the unknown quantity to be exterminated, may be the same in each equation; and then, by adding, or subtracting the equations … that term will vanish, and a new equation emerge, wherein the number of dimensions (if not the number of unknown quantities) will be diminished.
>
> — Simpson [1755, 63–64]

Simpson followed Newton in not limiting the discussion to linear equations. His hesitancy about the outcome of case 3° for polynomial equations suggests he may have been the originator of the addition and/or subtraction lesson. He includes 12 examples, of which 8 are linear and solved by adding or subtracting pairs of equations to remove a variable common to both.

The innovations of Newton and Rolle are underscored by the failure of some authors to formulate an explicit rule for solving simultaneous equations. For example, a contemporary of Hammond and Simpson, the sightless Lucasian professor Nicholas Saunderson [1761, 164–166], solved the three-equation problem of Peletier and Cardano using Gaussian elimination by the method of addition and/or subtraction. Saunderson explained that equations from which variables were removed should be grouped into *ranks*, which were equivalent to Rolle's *classes*, but Saunderson did not develop this terminology into a general prescription.

Leonhard Euler, in his algebra textbook, which was much admired for its concise style, provided a clear example of a missing rule. The text was originally written in German but appeared first in Russian in 1768 and eventually in many other languages.[15] Euler began with the compelling testimonial that the book was dictated for the instruction of his secretary, who mastered the subject from the text without additional instruction

---

[15] Heefer [2005, 3–4, 6] enumerates the editions of Euler's textbook. He traces most of Euler's exercises to an algebra textbook of 1525 by Christoff Rudolff which was reprinted by Michael Stifel in 1553.

[Euler, 1822, xxiii]. He included a chapter specifically for simultaneous linear equations. To find the values for two unknowns in two equations, he repeated the equality-of-values method.

> The most natural method of proceeding ... is, to determine, from both equations, the value of one of the unknown quantities, as for example x, and to consider the equality of these two values; for then we shall have an equation, in which the unknown quantity y will be found by itself.... Then, knowing y, we shall only have to substitute its value in one of the quantities that express x.
> — Euler [1771, Part 2, Sect. 1, Chap. 4, Art. 45], translated in Euler [1822, 206]

Euler continued with equality of values for three equations. However, he cautioned against adopting a rote approach, and therefore did not state a general algorithm.

> If there were more than three unknown quantities to determine, and as many equations to resolve, we should proceed in the same manner; but the calculation would often prove very tedious. It is proper, therefore, to remark, that, in each particular case, means may always be discovered of greatly facilitating the solution.
> — Euler [1771, Part 2, Sect. 1, Chap. 4, Art. 53], translated in Euler [1822, 211]

Euler's ambivalence toward Gaussian elimination became a dichotomy between research and instruction in the work of Étienne Bézout. The preface to Bézout's 1779 masterpiece on the theory of equations reviewed efforts to solve simultaneous polynomial equations. Bézout noted that Euler and Gabriel Cramer had studied how to "reach the final equation" in the reduction process that Newton and Rolle, both unmentioned, first made explicit. He then announced a way to "improve the elimination method for first-order equations" that takes advantage of zero coefficients [Bézout, 2006, xxiii and 138–144]. Although Bézout referred to this process as elimination, in modern terminology he evaluated determinant solution formulas by a novel method for sparse systems, those with many zero coefficients.

Bézout did not let the determinants in his research intrude on his successful textbooks, which he prepared for military academies. In these he taught Gaussian elimination. A posthumous and final version of his lessons began with "égalez ces deux valuers" to reduce two equations to one, and then continued to larger systems [Bézout, 1788, 53–67, Art. 65–73]. He solved three equations by equality of values, and unlike Euler, Bézout did not hesitate to recommend the method for a greater number of equations.

Equations of the first degree, with three or more unknowns

71. Once what we have said is well understood, it is easy to see how we should behave when the number of unknowns and equations is greater.

We always assume that we have as many equations as unknowns. If there are three, *we take in each the value of a single unknown, as if everything else were known. We then will match the first value to the second and first to third, or we will match the first to the second, and the second to third. We will, by this method, obtain two equations with two unknowns only, and which are treated by the preceding rule (66).*

If all the unknowns are not in every equation, the calculation would be easier, but it would always be done in a similar manner.

72. We see then that if there were a greater number of equations, the general rule would be: *Take in each equation, the value of one unknown; equate one of these values*

to the others, and you have one equation and one unknown less. Treat these new equations
as you have done for the first, and you will again have one equation and one unknown less.
Continue until you finally have one equation that has no more than one unknown.
                                — Bézout [1788, 57 and 59, original emphasis][16]

As the 18th century ended, Sylvestre Lacroix wrote a textbook that demonstrated that
the time when authors could neglect Gaussian elimination was passing. Domingues
[2008, 1–2] characterizes Lacroix as a minor mathematician but an astute textbook author
who sought to present the best approaches in an original, uniform style.[17] Lacroix's text-
book included lessons on simultaneous linear equations that underwent considerable revi-
sion between the second and fifth editions of 1800 and 1804. In the second edition Lacroix
discussed one and two unknowns in the same number of equations, and then passed to a
derivation of explicit, determinant-like formulas for the unknowns in systems of two, three,
and even four equations [Lacroix, 1800, 79–104, Art. 75–91]. In the fifth edition, between
these two rather different discussions, he inserted text similar to Hammond's for up to four
simultaneous linear equations. In the section title Lacroix made it plain that the method
could be applied to any number of equations.

Of the resolution of any given number of Equations of the First Degree, containing an equal
number of Unknown Quantities.
78. . . . if these unknown quantities are only of the first degree, [then] according to the
method adopted in the preceding articles, *we take in one of the equations the value of*
one of the unknown quantities, as if all the rest were known, and substitute this value in
all the other equations, which will then contain only the other unknown quantities.

This operation, by which we exterminate one of the unknown quantities, is called *elim-*
ination. In this way, if we have three equations with three unknown quantities, we deduce
from them two equations with two unknown quantities, which are to be treated as above;
and having obtained the values of the two last unknown quantities, we substitute them in
the expression for the value of the first unknown quantity.

If we have four equations with four unknown quantities, we deduce from them, in the
first place, three equations with three unknown quantities, which are to be treated in the

---

[16] **Des Équations du premier degré, à trois & á un plus grand nombre d'inconnues.**
71. Ce que nous venons de dire étant une fois bien conçu, il est facile de voir comment on doit se
conduire, lorsque le nombre des inconnues & des équations est plus considérable.

Nous supposerons toujours qu'on ait autant d'équations que d'inconnues. Si l'on en a trois, *on*
prenda dans chacune la valeur d'une même inconnue, comme si tout le rest étoit connu. On égalera
ensuite la première valeur à la seconde, & la première à la troisième; ou bien l'on égalera la première à
la seconde, & la seconde à la troisième. On aura, par ce procédé, deux équations à deux inconnues
seulement, & on les traitera par la règle précédene (66).

Si toutes les inconnues n'entroient pas à la sois dans chaque équation, le calcul seroit plus simple,
mais se seroit toujours d'une manière analogue.

72. On voit par-là que s'il y avoit un plus grand nombre d'équations, la règle générale seroit. . . .
Prenez, dans chaque équation, la valeur d'une même inconnue; égalez l'une de ces valeurs à chacune des
autres, & vous aurez une équation & une inconnue de moins. Traitez ces nouvelles équations comme
vous venez de saire pour les premières, & vous aurez encore une équation & une inconnue de moins.
Continuez ainsi jusqu'à ce qu'enfin vous parveniez à n'avoir plus qu'une inconnue.

[17] The *Monthly Review* [1801, 471] recommended Lacroix's *Elemens d'algèbre* for "a great
variety of valuable methods; and the excellence of the selection forms, in our opinion, the chief
merit . . . ."

manner just described; having found the values of the three unknown quantities, we substitute them in the expression for the value of the first, and so on.

> — Lacroix [1804, 114, Art. 78, original emphasis],
> translated in Lacroix [1818, 84, Art. 78]

The American translator, John Farrar, inserted Newton's "exterminate" into the English text above, whereas the original is as follows.

> Cette opération, par laquelle on chasse une des inconnues, se nomme *élimination*.
> — Lacroix [1804, 114, Art. 78, original emphasis]

With this passage Lacroix initiated the characterization of Gaussian elimination as *elimination*. Dozens of 19th-century American textbooks repeated "[this] is called elimination" once Farrar's 1818 translation of Lacroix appeared.[18]

Following Lacroix, Louis Pierre Marie Bourdon was the next author of an influential algebra text that became grist for American translators. The fifth edition of Bourdon [1828, 79–80] identified three ways to eliminate a variable from two equations: Newton's equality of values or comparison, Newton and Rolle's substitution, and Simpson's addition and/or subtraction, the last being the most highly recommended. Bourdon taught that simultaneous equations should be solved by using elimination to replace m equations with $m - 1$ equations, then $m - 2$, and so on. Bourdon's text was translated by Edward Ross [1831] and abridged by Charles Davies [1835] to become *Davies' Bourdon*, which had "extensive circulation in all parts of the United States" [Cajori, 1890, 372].

3. Gaussian elimination at the time of Gauss

3.1. "*Without any desire to do things which are useful*"

As the preceding survey of European texts shows, Newton brought order to the treatment of simultaneous equations through his presentation of a method that evolved into schoolbook Gaussian elimination. That is not to say that simultaneous linear equations were more than pedagogic exercises through the period considered. Neither the scientist, Newton, nor the banker, Hammond, offered any compelling examples from their respective domains of expertise.

None of the secondary literature argues that solving algebraic equations was needed, and several authors intimate to the contrary. Neugebauer [1969, 71–72] points out that ancient economies required only arithmetic to function. The exercises in Chapter 8 of the *Nine Chapters* are plainly contrived. Although mathematics lessons can supply valuable information about daily life in the past [Libbrecht, 1973, 416], Katz [1997, 25] sees an "artificial" quality to algebra problems in Babylonian, Greek, Arabic, and European Renaissance texts. Hogendijk [1994] explains that Islamic civilization needed arithmetic for commerce and advanced mathematics for astronomy, but the mathematical studies that in hindsight were the most sophisticated, such as algebra, were undertaken for their own sake.

"By and large it is uniformly true in mathematics that there is a time lapse between a mathematical discovery and the moment when it is useful," John von Neumann [1954] opined in his measured prose, and in the meantime,"the whole system seems to function without any direction, without any reference to usefulness, and without any desire to do

---

[18] A search in Google Books for the phrase "is called elimination" finds many examples in the period 1820–1900.

things which are useful."[19] It is not often that a mathematical discovery is immediately useful, as was the case with the method of least squares. And it was the method of least squares that finally created a recurring need to solve simultaneous linear equations. When a use for schoolbook Gaussian elimination finally arose, Carl Friedrich Gauss invented something better for the problem at hand.

3.2. "*Méthode des moindres quarrés*"

The genesis of the method of least squares lay in a scientific question which in the 18th century had not been completely resolved: how to make accurate predictions from measurements. In the middle of the century, differences between astronomical observations and orbital formulas, derived from Newton's principles, at times had even cast doubt on the inverse square law of gravitation.[20] Both Euler and Pierre-Simon Laplace had speculated that the law of gravity might need modification for astronomical distances. Stigler [1986, 17, 28] explains that fitting the orbital formulas to more observations than were absolutely necessary had been inconceivable. Euler "distrusted the combination of equations, taking the mathematician's view that errors actually increase with aggregation rather than taking the statistician's view that random errors tend to cancel one another" [Stigler, 1986, 28]. A new paradigm was adopted once Tobias Mayer successfully applied *ad hoc* fitting methods to predict the lunar orbit. Laplace in particular then derived fitted orbits that both vindicated Newton and established the mathematics that Laplace employed, "analysis."[21] These successes raised the question: what was the best fitting method to use?[22]

Two inventors are recognized for the method of least squares. Gauss claimed that he knew in 1794 or 1795[23] "the principle that the sum of the squares of the differences between the observed and computed quantities must be a minimum." Thus when the dwarf planet Ceres was sighted and lost in 1801, he quickly found its orbit by procedures that included least squares methods. Gauss was reticent about the calculations in the hasty announcement of his results, although he did send an explanation in 1802 to his friend Wilhelm Olbers.[24] The explanation was inexplicably returned to Gauss only in 1805, yet in the interim Gauss had been publishing orbits for Ceres and other celestial bodies [Dunnington, 2004, 53 and 420–421].

Meanwhile, in an appendix to a long paper, Adrien-Marie Legendre posed the general problem of finding the most accurate parameterization furnished by a given set of observations.[25]

---

[19] Von Neumann was not an apologist for mathematics in the sense of G.H. Hardy. Dieudonné [1981, 89] called von Neumann "the last of the great mathematicians" who understood both mathematics and its uses.

[20] See Gillispie [1997, 29], Goldstine [1977, 142], and Stigler [1986, 30] for elaboration.

[21] Hawkins [1977b, Sect. 2] summarizes the status of analysis at the time of Lagrange and Laplace.

[22] Farebrother [1999] surveys the several fitting methods available at the end of the 18th century, including those used by Mayer and Laplace.

[23] Gauss [1809, Art. 186] wrote 1795, but Plackett [1972, 241] quotes Gauss recalling 1794. Gauss may not have remembered the exact date, because Dunnington [2004, 469] reports that Gauss began his famous diary of mathematical discoveries only in 1796.

[24] Heinrich Wilhelm Matthäus Olbers was a physician and an accomplished amateur astronomer.

[25] Stigler [1986, 13 and 15] recommends Legendre's appendix as among the most elegant introductions of a significant mathematical concept.

Legendre [1805, 72] observed that the problem often involves many systems of equations each of the form

$$E = a + bx + cy + fz + \text{etc.}$$

(his notation), where $a, b, c, f, \ldots$ are numbers that vary among the equations, and $x, y, z, \ldots$ are parameters common to all the equations. As in linear models today, the numbers in each equation represent the data of one observation; the unknowns are the model parameters. Legendre viewed the problem as finding values for the parameters that make E small. If the number of equations exceeds that of the unknowns (so that the equations cannot be solved exactly), then he suggested minimizing the sum of the squares of the E's. He called this overall process the *méthode des moindres quarrés* (modern *carrés*). The solution was found by differentiating the sum of squares to derive the "equations of the minimum" (modern normal equations). Since these simultaneous equations were linear and equinumerous with the variables, Legendre [1805, 73] said they could be solved by "ordinary methods" (par les méthodes ordinaires).

Over a decade passed between Gauss's own discovery and its publication. Although Gauss intended to publish immediately after Legendre, the manuscript was delayed by the Napoleonic wars. In *Theoria Motus* (1809), Gauss finally explained his process for orbital calculations, and he returned to the conceptual problem of fifty years earlier: how to justify values calculated from erroneous data. Lacking justification, he intimated that the only reason to minimize squares was convenience [Gauss, 1809, 220–221, Art. 186]. Rather than beginning by minimizing the discrepancy in the equations as Legendre had done, instead Gauss formulated "the expectation or probability that all these values will result together from observation" [Gauss, 1809, 210, Art. 175]. Assuming the errors in the unknowns follow a Gaussian or normal distribution (later terminology), Gauss showed that it was maximizing the expectation which implied that the sum of squares should be minimized. He then echoed Legendre's sentiment about the resulting calculation to find the parameters.

> We have, therefore, as many linear equations as there are unknown quantities to be determined, from which the values of the latter will be obtained by common elimination. (... per eliminationem vulgarem elicientur.)
>
> — Gauss [1809, 214, Art. 180], this translation Gauss [1857, 261]

There then followed, in the published work of Gauss and Laplace, a scientific dialogue that explored the nature of probability and estimation.[26] During this period, Gauss [1823] gave a second and unqualified justification for the method of least squares, now known as the Gauss–Markov theorem for the minimum variance linear unbiased estimator. Hald [2007, 98, 105–109] suggests that few contemporary readers, if any, understood all that Gauss and Laplace wrote. However, he considers that Gauss's first proof for the method of least squares, coupled with the sufficiency in many instances of the assumption of normally distributed errors, allowed the likes of Hagen [1867], Chauvenet [1868], and Merriman [1884] to maintain and extend a statistically respectable methodology of estimation from the time of Gauss through to the development of modern statistics.

---

[26] Three short monographs with historical emphasis but each on a slightly different aspect of this work — statistical fitting procedures, parametric statistical inference, and classical analysis of variance — have recently been written by Farebrother [1999], Hald [2007], and Clarke [2008], respectively. Histories with a wider scope include Gillispie [1997, Chap. 25], Goldstine [1977, Chaps. 4.10], and Stigler [1986, Chap. 4].

At the beginning of the 20th century a mathematics professor at the Massachusetts Institute of Technology could proudly announce that "scientific investigations of all kinds" relied on a mature computational technology called "The Adjustment of Observations" or "The Method of Least Squares" [Bartlett, 1900, 1 and 17].[27] The subject divided into two cases. Note that matrices were not used in these formulations until the mid-20th century.

Case 1. The "adjustment of indirect observations" had Legendre's original, overdetermined equations that today would be stated as $\min_x \|b - Ax\|_2$, where A is an $m \times n$ matrix, $m > n$, and b and x are compatibly sized column vectors. These problems were solved by reducing them to $A^t A x = c$ with $c = A^t b$, where ^{t} is transposition [Bartlett, 1900, 17, Arts. 23–25; Wright and Hayford, 1906, 93 Art. 74, 106 Art. 84].

Case 2. The "adjustment of conditioned observations" was formulated by Gauss [1826] to find minimum 2-norm solutions of underdetermined equations, $\min_{Ax=b} \|x\|_2$, where $m < n$. He reduced these problems to $AA^t u = b$ with $x = A^t u$. The latter equations for x were called the *correlate equations* [Bartlett, 1900, 111, Arts. 116–118; Wright and Hayford, 1906, 152 Art. 119, 243 Art. 179].

Each row of the overdetermined (Case 1) or underdetermined (Case 2) equations, $Ax = b$, was called a *condition*. The reduced forms of both problems were called *normal equations* and were solved by various forms of elimination.[28]

3.3. Gauss's formulation of the method of least squares

Gauss performed meticulous calculations almost as a leisure activity throughout his life [Dunnington, 2004, 227–228]. He thought calculations so important that he included them in his papers, striving to make them brief and intuitively clear. Accordingly, his publication that mentioned "common elimination" was followed with a detailed explanation. Rather than solve the simultaneous linear equations "of the minimum" as Legendre had done, instead Gauss [1810] formulated his calculations in terms of quadratic forms.

A canonical expression for quadratic forms had already been introduced by Lagrange [1759]. In modern matrix notation, a quadratic form is given by $x^t A x$ for a column vector of variables x and a symmetric matrix A. Lagrange substituted new variables for linear combinations of the original variables, which amounts to constructing a triangular matrix U of substitution coefficients so that $A = U^t D U$ for a diagonal matrix D. The quadratic form thus became a weighted sum of squares, $(Ux)^t D (Ux)$, which Lagrange used to ascertain local extrema. Such was the approach that Gauss took for least squares.[29] Since

---

[27] Bartlett [1900, v–vi] lists English, French, and German textbooks from the end of the 19th century. See Ghilani and Wolf [2006] for a treatment from the beginning of the 21st century.

[28] Stigler [1999, 415–420] reports that Gauss [1822, 84] appears to have used *Normalgleichungen* first but also just once and offhand. Gauss did not explain what he meant by the name. He could have simply meant "ordinary" equations.

[29] Gauss may have learned the quadratic construction from Lagrange. At the time when Gauss first considered least squares in 1795, he borrowed from the Göttingen library the journal volume that contains Lagrange's paper on extrema [Dunnington, 2004, 398]. Since Gauss did not need this volume for his major work from this period, which was on number theory (in which he did cite papers from the journal but none from this particular volume [Gauss, 1801, Art. 202, n. 9]), it seems likely that he borrowed the volume to consult Lagrange's paper to learn about minimizing quadratic forms.

neither Lagrange [1759] nor Gauss [1801] dealt with quadratic forms of more than three variables,[30] Gauss had yet to systematically extend the construction of the canonical expression to arbitrarily many unknowns, which he finally did in his papers about least squares.[31]

Gauss [1810, 22] wrote the overdetermined equations as

$$n + ap + bq + cr + ds + \ldots = w$$
$$n' + a'p + b'q + c'r + d's + \ldots = w'$$
$$n'' + a''p + b''q + c''r + d''s + \ldots = w''$$
$$\ldots$$

(original notation, but with "..." replacing "etc."). The symbols $n, a, b, c, \ldots$, with or without primes, are numbers. The purpose is to find values for the variables $p, q, r, \ldots$ to minimize

$$\Omega = ww + w'w' + w''w'' + \ldots.$$

Gauss introduced a bracket notation

$$[xy] = xy + x'y' + x''y'' + \ldots,$$

where the letter x either is y or alphabetically precedes y.[32] Gauss did not have a name for his bracket symbols; they would become known as "auxiliaries" because they represented intermediate values that were needed to solve the least-squares problem by Gauss's approach. This notation expressed the normal equations (name not yet introduced) as follows:

$$[an] + [aa]p + [ab]q + [ac]r + [ad]s + \ldots = 0$$
$$[bn] + [ab]p + [bb]q + [bc]r + [bd]s + \ldots = 0 \tag{10}$$
$$[cn] + [ac]p + [bc]q + [cc]r + [cd]s + \ldots = 0$$
$$\ldots$$

As Legendre [1805, 73] and he himself had done before [Gauss, 1809, 214], Gauss again remarked that these equations could be solved by elimination [Gauss, 1810, 22], but he did not explicitly perform that calculation. Instead, he noted that the brackets give the coefficients of the variables in the sum of squares, a quadratic form.

$$\Omega = [nn] + 2[an]p + 2[bn]q + 2[cn]r + 2[dn]s + \ldots$$
$$+ [aa]pp + 2[ab]pq + 2[ac]pr + 2[ad]ps + \ldots$$
$$+ [bb]qq + 2[bc]qr + 2[bd]qs + \ldots$$
$$+ [cc]rr + 2[cd]rs + \ldots$$
$$\ldots$$

---

[30] Hawkins [1977a, 83–85] summarizes the use of quadratic forms in Gauss [1801].

[31] A difficulty in extending the construction was the subscriptless notation of the time. Farebrother [1999, 161, n.] attributes double subscript notation to Cauchy in 1815 [Cauchy, 1905, 114]. Gauss first used numeral indices in 1826, but as superscripts [Gauss, 1826, 83].

[32] That is, "$[xy]$" is a composite symbol that Gauss meant to be used only when the letter "x" preceded or matched the letter "y" in alphabetic order.

Gauss extended his bracket notation to

$$[xy, 1] = [xy] - \frac{[ax][ay]}{[aa]}$$
$$[xy, 2] = [xy, 1] - \frac{[bx, 1][by, 1]}{[bb, 1]} \qquad (11)$$
$$[xy, 3] = [xy, 2] - \frac{[cx, 2][cy, 2]}{[cc, 2]}$$
$$\dots$$

and so on. The values of these formulas are the coefficients remaining after successive combinations of variables have been grouped into perfect squares. The first of these combinations of variables, A,

$$A = [an] + [aa]p + [ab]q + [ac]r + [ad]s + \dots$$
$$B = [bn, 1] + [bb, 1]q + [bc, 1]r + [bd, 1]s + \dots$$
$$C = [cn, 2] + [cc, 2]r + [cd, 2]s + \dots \qquad (12)$$
$$\dots,$$

simplifies the quadratic form by grouping all occurrences of the variable p into a perfect square, thus removing p from the balance of Ω:

$$\Omega - \frac{A^2}{[aa]} = [nn, 1] + 2[bn, 1]q + 2[cn, 1]r + 2[dn, 1]s + \dots$$
$$+ [bb, 1]qq + 2[bc, 1]qr + 2[bd, 1]qs + \dots$$
$$+ [cc, 1]rr + 2[cd, 1]rs + \dots$$
$$\dots.$$

If this process is repeated with B, C, $\dots$, then eventually

$$\Omega - \frac{A^2}{[aa]} - \frac{B^2}{[bb, 1]} - \frac{C^2}{[cc, 2]} - \dots = [nn, \mu],$$

where μ is the quantity of variables. Each combination A, B, C, $\dots$ has one less unknown than the preceding combination. Thus $A = 0$, $B = 0$, $C = 0$, $\dots$ can be solved in reverse order to obtain the values for $p, q, r, \dots$, also in reverse order, at which Ω attains its minimum, $[nn, \mu]$. In later theoretical discussions of least squares methods, e.g., Gauss [1826, Art. 13], Gauss always referred back to his 1810 paper for details of the calculations for transforming quadratic forms.

Gauss's contributions to the method of least squares became known almost immediately. By 1819 even a gymnasium prospectus, Paucker [1819], cited Legendre [1805] and Gauss [1809] for formulating the statistical inference problem (modern terminology), although Gauss was not cited for any particular method of calculating the inferences. Gauss's process which neatly tied together (what we now call) linear algebra, optimization theory, and his probabilistic justification for the inferences, seems to have been adopted as an algorithm slowly and by geodesists, not by mathematicians.

3.4. Geodesy for cartography

The original motivation for the computational developments in least squares was to cope with observational errors in two major scientific activities of the 19th century, and Gauss was involved in both. One was astronomy, which was pursued for navigation and timekeeping as well as for its intrinsic interest. Gauss [1809] described the calculations that he had invented to derive orbital formulas from a few observations. He reported the results of his own calculations before and after 1809 in many papers, which constituted the bulk of his early publications.[33]

The other scientific field was geodetic research for cartography, an activity continually and directly sponsored by governments. Indeed, the first scientific agency of the United States was the Coast Survey Office, founded in 1807.[34] Gauss became prominent in geodesy through the many papers he wrote during his protracted survey of Hanover. This small German state, roughly coincident with modern Lower Saxony, was Gauss's home for most of his life.

Friedrich Bessel warned Gauss that the toil of the survey would detract from his research [Dunnington, 2004, 120]. However, the time was well spent. Although Legendre invented the method of least squares to solve a geodetic problem, an English survey officer explained that it was Gauss from whom geodesists adopted the approach.

> If in effecting a [survey] triangulation one observed only just so many angles as were absolutely necessary to fix all the points, there would be no difficulty in calculating [the locations]; only one result would be arrived at. But it is the invariable custom to observe more angles than are absolutely needed, and it is these supernumerary angles which give rise to complex calculations. Until the time of Gauss and Bessel computers had simply used their judgement as they best could as to how to employ and utilize the supernumerary angles; the principal of least squares showed that a system of corrections ought to be applied, one to each observed bearing or angle, such that subject to the condition of harmonizing the whole work, the sum of their squares should be an absolute minimum. The first grand development of this principle is contained in this work of Bessel's.
>
> — Clarke [1880, 26–27]

Clarke refers to a Prussian triangulation to which Bessel applied the panoply of Gauss's techniques.[35] Bessel's endorsement of Gauss in a survey for a major government drew the attention of other geodesists.

As Clarke explained, a critical step in making maps was to reconcile the measurement errors in the raw data gathered by surveyors. At each vertex surveyors measured either the included angles of triangles or the azimuthal directions to nearby vertices; in the latter case the angles were differences of the directions.[36] Gauss [1826, Art. 22] stipulated that the angles

---

[33] Nievergelt [2001] explains the orbital calculations that were used in 19th-century astronomy, while Grier [2005] describes the institutional history of computing groups in national observatories where the calculations were made.

[34] Cajori [1929] traces the colorful early years of the Coast Survey. The enthusiasm for scientific research at the Coast Survey and the overlapping responsibilities of the younger Geological Survey were the subject of vehement debate in the government [United States Congress, 1886].

[35] Clarke [1880, 21] refers to Bessel and Baeyer [1838, 130], which is discussed by Torge [2001, 11].

[36] Each survey prescribed strict uniformity in taking measurements. The measuring instrument was and is called a *theodolite*. The book by Clarke [1880], the research paper by Nell [1881], the encyclopedia by Jordan [1895], and the textbook by Wright and Hayford [1906] have further explanations and more examples from the past.

Fig. 5. Gauss [1826, 86, Art. 23] used a small portion of a triangulation of Holland to illustrate survey adjustments. He took the data from [Krayenhoff, 1827], a book in his personal library. The map seen here is from the earlier edition [Krayenhoff, 1813]. Courtesy of the Bancroft Library, University of California, Berkeley.

satisfy three kinds of conditions: (1) the sum of angles around an interior vertex equals 2π radians, (2) the sum of angles in a triangle equals π plus the spherical excess, and (3) *side conditions* that chain together linearized sine laws for circuits of triangles with common edges. The unknowns are perturbations to the measurements intended to make the angles satisfy the conditions. Gauss formed a side condition from the triangles around each interior vertex, in which case (it is easy to see) even an ideal net consisting of f nonoverlapping triangles and v vertices has $3f$ angles, but only $3f - 2v + 4$ conditions. Thus, in general, the adjustment problems were under- not overdetermined. In his last major theoretical work on least squares, Gauss [1826] introduced the solution method given in Section 3.2 as Case 2. He illustrated the method by readjusting a small part of the Dutch triangulation (Figs. 5 and 6). In comparison, the British Isles triangulation was quite irregular [Ordnance Survey, 1858, II, Pl. xviii]. Most surveys could not afford to follow a highly regular triangulation and also had missing measurements from inaccessible vertices (e.g., mountaintops) or blocked sight lines.

As the 19th century progressed, the growing use of least squares methods created a recurring need to solve dauntingly large problems. Of necessity these large problems were broken into smaller subproblems.

> In the principal triangulation of Great Britain and Ireland there are 218 stations, at 16 of which there are no observations, the number of observed bearings is 1554, and the number of equations of condition 920.[37] The reduction of so large a number of observations

---

[37] In Stigler [1986, 158] the numbers 1554 and 920 are mistakenly switched.
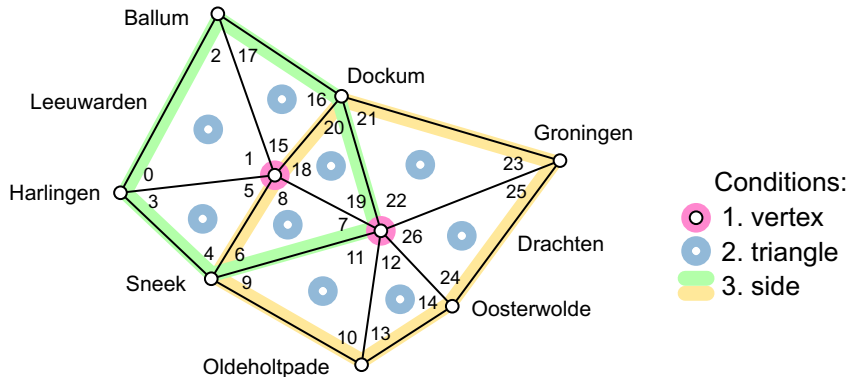
Fig. 6. The portion of Fig. 5 that Gauss readjusted. He had to find adjustments for 27 angles, but he had only 13 conditions for them to satisfy, consisting of 2 vertex conditions, 9 triangle conditions, and 2 side conditions. Jordan [1895, 489] reproduces a similar figure.

> in the manner we have been describing [i.e., least-squares Case 2] would have been quite impossible, and it was necessary to have recourse to methods of approximation....
>
> ... the network covering the kingdom was divided into a number of blocks, each presenting a not unmanageable number of equations of condition. One of these being corrected or computed independently of the others, the corrections so obtained were substituted (as far as they entered) in the equations of condition of the next block, and the sum of the squares of the remaining equations in that figure made a minimum. The corrections thus obtained for the second block were substituted in the third and so on. ... The number of blocks is 21, in 9 of them the number of equations of condition is not less than 50, and in one case the number is 77. These calculations — all in duplicate — were completed in two years and a half — an average of eight computers being employed....
>
> In connection with so great a work successfully accomplished, it is but right to remark how much it was facilitated by the energy and talents of the chief computer, Mr. James O'Farrell.
>
> <div align="right">— Clarke [1880, 237, 243]</div>

The systems of linear equations for the subproblems were still very large for hand computing. In these situations, professional computers found it useful to emulate Gauss's calculations of 50 years earlier.

4. Perfecting Gaussian elimination for professional computers

4.1. Overview of perfecting calculations

"Numerical mathematics stayed as Gauss left it until World War II," concluded Goldstine [1972, 287] in his history of the subject. In the 120 years after Gauss [1826], there were at best a dozen noteworthy publications on solving simultaneous linear equations. The two main thrusts were simplifications for professional computers, which began with Gauss himself, and matrix interpretations. These developments overlap only slightly chronologically, and they did not influence each other until the middle of the 20th century. The former will be addressed here and the latter in the following section. Four authors, Gauss, Doolittle, Cholesky, and Crout, demonstrably influenced professional hand-computing practice for Gaussian elimination.

4.2. Gauss's convenient notation

The triangulation of the great Ordnance Survey [1858] and the formulation of its adjustment [Clarke, 1880] are preserved in detail, but exactly how the calculations were performed is missing. Since the details of major calculations were not archived, whether and why Gauss's solution method was used — and what it was considered to be — must be inferred from sources such as William Chauvenet [1868].[38] There appear to have been three advantages to Gauss's method of solving the normal equations (described in Section 3.3).

First, the bracket notation conceptually separated the algebra from the arithmetic, so the workflow could be addressed. Gauss relieved computers of the tedium of having to rewrite equations, and in so doing, he enabled them to consider how best to organize their work.

> By whatever method of elimination is performed, we shall necessarily arrive at the same final values of the unknown quantities; but when the number of equations is considerable, the method of substitution, with Gauss's convenient notation, is universally followed.
>
> — Chauvenet [1868, 514]

It may come as some surprise to learn that when Gauss performed Gaussian elimination, he simply listed all the numbers in the order in which he computed them, using his bracket notation to identify the values: $[cd] = 1.13382$, $[cd, 1] = 1.09773$, $[cd, 2] = 1.11063$, etc. (Fig. 7). Dunnington [2004, 138] echoed Bessel in regretting how much time Gauss spent calculating for his interminable survey projects: Gauss estimated that he used one million numbers!

Second, in contrast to the method depicted in Fig. 1 and prescribed in algebra textbooks, Gauss realized economies by avoiding duplicate calculations for symmetric equations.[39]

> By means of a peculiar notation proposed by Gauss, the elimination by substitution is carried on so as to preserve throughout the symmetry which exists in the normal equations.
>
> — Chauvenet [1868, 530]

Chauvenet even foreshadowed modern practice for measuring the work of a calculation by counting how many bracket or auxiliary values had to be computed: 156 for eight normal equations, etc.

Third, Gauss included refinements such as estimates of precision, variances, and weights for the unknowns, all expressed in his bracket notation. The difficulty of making changes to the formulas, whose complexity was compounded by their relationship to Gauss's statistical theories, and the demonstrable benefits of Gauss's ideas, engendered a reluctance to alter his computational prescriptions. His efficient method for overdetermined problems by itself was conceptually difficult because the solution of the normal equations, $AA^t u = b$, was not the solution of the problem, $x = A^t u$. Gauss's bracket notation was still being taught a hundred years after his 1810 paper.[40] At the beginning of the 20th century, Wright and Hayford

---

[38] Chauvenet is remembered as a founder of the United States Naval Academy and by the Mathematical Association of America's Chauvenet Prize.

[39] The normal equations are symmetric in the sense that the same coefficient is attached to the jth variable in the kth equation, and vice versa.

[40] Bartlett [1900], Johnson [1905], and Wright and Hayford [1906] offer instruction on the use of bracket notation. For continued emphasis on the advantages of preserving symmetry, see Palmer [1912, 84–85] and Golub and Van Loan [1996, 133, Chap. 4].

DISQVISITIO DE ELEMENTIS ELLIPTICIS PALLADIS etc. **25**

aequationibus incognitae p, q, r, s etc. determinari debebunt, quod ordine inuerfo facillime effici poterit, quum manifefto vltima aequatio vnicam incognitam implicet, penultima duas et fic porro. Simul haec methodus eo nomine fe commendat, quod valor minimus aggregati Ω fponte inde innotefcit, quippe qui manifefto eft $= (nn, \mu)$.

14.

Applicemus iam haecce praecepta ad exemplum noftrum, vbi p, q, r, s etc. funt d L, d٦, dπ, dφ, dΩ, di. Calculo accurate abfoluto hosce valores numericos inueni:

[nn] = 148848	[cn,1] = − 119,31	[de,2] = − 0,46088
[an] = − 371,09	[dn,1] = − 125,18	[df,2] = − 0,17205
[bn] = − 580104	[en,1] = + 72,52	[ee,2] = + 2,24325
[cn] = − 113,45	[fn,1] = − 43 22	[ef,2] = − 0,37841
[dn] = + 268,53	[bb,1] = + 2458225	[ff,2] = + 5,61686
[en] = + 94,26	[bc,1] = + 62,13	Hinc porro
[fn] = − 31,81	[bd,1] = − 510,58	[nn,3] = + 99034
[aa] = + 5,91569	[be,1] = + 213,84	[dn,3] = + 25,66
[ab] = + 7203,91	[bf,1] = + 73,45	[en,3] = + 74,23
[ac] = − 0,09344	[cc,1] = + 0,71769	[fn,3] = + 2,75
[ad] = − 2,28516	[cd,1] = + 1,09773	[dd,3] = + 9,29213
[ae] = − 0,34664	[ce,1] = − 0,05852	[de,3] = − 0,36175
[af] = − 0,18194	[cf,1] = + 0,26054	[df,3] = − 0,57384
[bb] = + 10834225	[dd,1] = + 11,12064	[ee,3] = + 2,23754
[bc] = − 49,06	[de,1] = − 0,50528	[ef,3] = − 0,35532
[bd] = − 3229,77	[df,1] = − 0,18790	[ff,3] = + 5,52342
[be] = − 198,64	[ee,1] = + 2,26185	Hinc eodem modo
[bf] = − 143,05	[ef,1] = − 0,37202	[nn,4] = + 98963
[cc] = + 0,71917	[ff,1] = + 5,61905	[en,4] = + 75,23
[cd] = + 1,13382		[fn,4] = + 4,33
[ce] = + 0,06400	Atque hinc fimili modo	[ee,4] = + 2,22346
[cf] = + 0,26341	[nn,2] = + 117763	[ef,4] = − 0,37766
[dd] = + 12,00340	[cn,2] = − 115,81	[ff,4] = + 5,48798
[de] = − 0,37137	[dn,2] = − 153,95	Hinc
[df] = − 0,11762	[en,2] = + 84,57	[nn,5] = + 96418
[ee] = + 2,28215	[fn,2] = − 39,08	[fn,5] = + 17,11
[ef] = − 0,36136	[cc,2] = + 0,71612	[ff,5] = + 5,42383
[ff] = + 5,62456	[cd,2] = + 1,11063	Atque hinc tandem
Hinc porro deduxi	[ce,2] = − 0,06592	[nn,6] = + 96364
[nn,1] = + 125569	[cf,2] = + 0,25868	
[bn,1] = − 138534	[dd,2] = + 11,01463	

C. F. Gaufs de element. ellipt. Pallad. T. I.   D   Habe-

Fig. 7. How Gauss performed Gaussian elimination. He wrote down the numbers in the order he computed them, using his bracket notation to identify the values [Gauss, 1810, 25].

[1906] saw professional computers using only two methods to solve normal equations: either the brackets of Gauss or the tables of Doolittle.

4.3. Doolittle, legendary computer

Myrick Hascall Doolittle (Fig. 11) was a computer who solved Gauss's normal equations to adjust triangulations at the United States Coast and Geodetic Survey for a period of 38 years from 1873 to 1911.[41] He is still remembered for his professional acumen as a human computer. Demonstrating the efficacy of his methods, Doolittle [1878, 117] notes that he solved 41 normal equations in a week using paper and pencil. For perspective, Fox [1987, 676] reports that four mathematicians, Alan Turing among them, needed about two weeks to solve 18 equations with "desk computing equipment" in 1946.[42] This does

---

[41] For further information on Doolittle, see the biographical sketch by Grier [2005, 78–79] and the primary biographical sources quoted at length by Farebrother [1987].

[42] Using Gaussian elimination to find n unknowns in n equations requires about $2n^3/3$ arithmetic operations. The approximately 3,888 operations for $n = 18$ compares well with the 4,800 minutes in 10 working days. The question is thus not why Fox et al. were slower, but rather how Doolittle was so quick.

not, however, imply that Doolittle was a numerical savant. A colleague, Mr. J. G. Porter, duplicated the calculation in the same time to check for errors.

Doolittle's speed stemmed from streamlining the work for hand computing. His one paper on the subject [Doolittle, 1878] contained a small numerical example that is restated here in Gauss's bracket notation to clarify the calculation. Corresponding to Gauss's Eq. (10), Doolittle's normal equations were as follows:

$$
\begin{aligned}
0 &= [aa]w + [ab]x + [ac]y + [ad]z + [an] \\
0 &= [ab]w + [bb]x + [bc]y + [bd]z + [bn] \\
0 &= [ac]w + [bc]x + [cc]y + [cd]z + [cn] \\
0 &= [ad]w + [bd]x + [cd]y + [dd]z + [dn].
\end{aligned}
\tag{13}
$$

Doolittle expressly planned the calculation to derive Newton's and Rolle's substitution formulas for each variable, which correspond to rearrangements of Gauss's equations $A = 0$, $B = 0$, ... Doolittle called these formulas the "explicit functions" for the variables. He was able to colocate many of the numbers by completing the formula for a given variable before undertaking any calculations for the next. Among the practices that he introduced, is placing the numbers of the calculation in tables to identify them. Doolittle kept the coefficients of the substitution formulas in Table A (Fig. 8), while he used a second Table, B, to record in columns the sums that give the values in Table A. Exactly how Doolittle conducted the work may be lost. "For the sake of perspicuity," he noted, "I have here made some slight departures from actual practice" [Doolittle, 1878, 117].

The salient feature of Doolittle's tables is a reduction in the labor of division and multiplication. All divisions reduce to multiplications through reciprocals formed (one per variable) in the first column of Table A. All multiplications have a single multiplier repeatedly applied to several multiplicands in a row of Table A and recorded in another row of Table B. For example, Row 8 in Table B results from a single multiplier in Table A (Row 2 Column y) applied to several multiplicands in Row 1, beginning at the same column (y) and moving rightward. The reduction of work occurs because Doolittle [1878, 117] performed multiplication using the three-digit tables of Crelle [1864]. Since he reused the multiplier for an entire row of calculations, Doolittle could open Crelle's tables to the one page for that multiplier and all the multiplicands. Schott [1879, 93] emphasized that using multiplication tables was innovative, noting that "logarithms are altogether dispensed with."[43]

The back-substitutions was performed with similar economy. Doolittle [1878, 116] distinguished between numbers used once and those used many times, so he prepared for the substitutions by copying the reciprocals and "explicit function" coefficients from Table A to Table C (Fig. 9). The value of z was available in the final row of Table A. The remaining variables were evaluated in Table D, where each row consists of one multiplier applied to one column, this time, of Table C. The sums of the columns in Table D give the other variables.

Doolittle's method included several contributions for which he is not now credited. He owed his speed in part to using just three-digit arithmetic in the multiplication tables, and hence everywhere in the calculation. This could, as Doolittle appreciated, introduce

---

[43] This comment suggests that computers generally did use logarithms to evaluate Gauss's brackets. Although the logarithm of a sum is not usually expressed in terms of logarithms of the summands, the subtraction in Eq. (11) could be done via special tables of "Gauss's logarithms."

Table A

step	recip-rocal	w	x	y	z	absolute term
step 1.			$[aa]$	$[ab]$	$[ac]$	$[ad]$	$[an]$
2.	$\dfrac{-1}{[aa]}$	$w=$	$-\dfrac{[ab]}{[aa]}$	$-\dfrac{[ac]}{[aa]}$	$-\dfrac{[ad]}{[aa]}$	$-\dfrac{[an]}{[aa]}$
5.				$[bb,1]$	$[bc,1]$	$[bd,1]$	$[bn,1]$
6.	$\dfrac{-1}{[bb,1]}$		$x=$	$-\dfrac{[bc,1]}{[bb,1]}$	$-\dfrac{[bd,1]}{[bb,1]}$	$-\dfrac{[bn,1]}{[bb,1]}$
10.					$[cc,2]$	$[cd,2]$	$[cn,2]$
11.	$\dfrac{-1}{[cc,2]}$			$y=$	$-\dfrac{[cd,2]}{[cc,2]}$	$-\dfrac{[cn,2]}{[cc,2]}$
16.						$[dd,3]$	$[dn,3]$
17.	$\dfrac{-1}{[dd,3]}$				$z=$	$-\dfrac{[dn,3]}{[dd,3]}$

Table B

step	x	y	z	absolute term
step 3.	$[bb]$	$[bc]$	$[bd]$	$[bn]$
4.	$-\dfrac{[ab]}{[aa]}\times[ab]$	$-\dfrac{[ab]}{[aa]}\times[ac]$	$-\dfrac{[ab]}{[aa]}\times[ad]$	$-\dfrac{[ab]}{[aa]}\times[an]$
7.		$[cc]$	$[cd]$	$[cn]$
8.		$-\dfrac{[ac]}{[aa]}\times[ac]$	$-\dfrac{[ac]}{[aa]}\times[ad]$	$-\dfrac{[ac]}{[aa]}\times[an]$
9.		$-\dfrac{[bc,1]}{[bb,1]}\times[bc,1]$	$-\dfrac{[bc,1]}{[bb,1]}\times[bd,1]$	$-\dfrac{[bc,1]}{[bb,1]}\times[bn,1]$
12.			$[dd]$	$[dn]$
13.			$-\dfrac{[ad]}{[aa]}\times[ad]$	$-\dfrac{[ad]}{[aa]}\times[an]$
14.			$-\dfrac{[bd,1]}{[bb,1]}\times[bd,1]$	$-\dfrac{[bd,1]}{[bb,1]}\times[bn,1]$
15.			$-\dfrac{[cd,2]}{[cc,2]}\times[cd,2]$	$-\dfrac{[cd,2]}{[cc,2]}\times[cn,2]$

Fig. 8. How Doolittle [1878, 115] performed Gaussian elimination, transcribed from his numerical example using Gauss's bracket notation to identify the quantities. The step numbers indicate the order of forming the rows. Each row in Table B is a multiple, by a single number, of a partial row in Table A. Rows 5, 10, and 16 in Table A are sums of rows 3–4, 7–9. and 12–15 in Table B, respectively.

severe rounding errors.[44] Thus, an important aspect of Doolittle's method was the ability to correct the three-digit approximate solution with comparatively little extra work. Since the angle adjustment problem itself corrected numbers that were approximately known — the

---

[44] Rounding errors are relatively much smaller for modern computing equipment that adheres to international standards, which require using roughly 7 or 14 decimal digits in calculations [Institute of Electrical and Electronics Engineers, 1985].

Table C

recip-rocal	x	y	z
$\dfrac{-1}{[aa]}$	$-\dfrac{[ab]}{[aa]}$	$-\dfrac{[ac]}{[aa]}$	$-\dfrac{[ad]}{[aa]}$
$\dfrac{-1}{[bb,1]}$		$-\dfrac{[bc,1]}{[bb,1]}$	$-\dfrac{[bd,1]}{[bb,1]}$
$\dfrac{-1}{[cc,2]}$			$-\dfrac{[cd,2]}{[cc,2]}$
$\dfrac{-1}{[dd,3]}$			

Table D

w	x	y	z
$-\dfrac{[an]}{[aa]}$	$-\dfrac{[bn,1]}{[bb,1]}$	$-\dfrac{[cn,2]}{[cc,2]}$	$\dfrac{[dn,3]}{[dd,3]}$
$z\times-\dfrac{[ad]}{[aa]}$	$z\times-\dfrac{[bd,1]}{[bb,1]}$	$z\times-\dfrac{[cd,2]}{[cc,2]}$	column sum = z
$y\times-\dfrac{[ac]}{[aa]}$	$y\times-\dfrac{[bc,1]}{[bb,1]}$	column sum = y	
$x\times-\dfrac{[ad]}{[aa]}$	column sum = x		
column sum = w			

Fig. 9. How Doolittle performed back-substitution. Table C he copied from Table A. The sum of each column in Table D evaluates the "explicit function" for a different variable.

measured angles — it may have seemed natural to further correct the adjustments. Both Doolittle [1878] and Schott [1879, 93] described the correction process without giving it a name; today it is called *iterative improvement* or *iterative refinement*.[45]

In describing the refinement process, Doolittle wrote w_1, x_1, y_1, z_1 for the values obtained from Tables A–D. When these approximations are substituted into the normal equations they give residual values.

$$r_1 = [aa]w_1 + [ab]x_1 + [ac]y_1 + [ad]z_1 + [an]$$
$$r_2 = [ab]w_1 + [bb]x_1 + [bc]y_1 + [bd]z_1 + [bn]$$
$$r_3 = [ac]w_1 + [bc]x_1 + [cc]y_1 + [cd]z_1 + [cn]$$
$$r_4 = [ad]w_1 + [bd]x_1 + [cd]y_1 + [dd]z_1 + [dn].$$

Doolittle emphasized that this particular calculation had to be carried out to a high degree of accuracy, after which the residuals r_1, r_2, r_3, r_4 (my notation) could be rounded to three digits for the remaining steps. In Table E (Fig. 10) he repeated the calculations on the residuals that he had performed on the constant terms of the normal equations in the final column of Table B. In Table F (Fig. 10) he duplicated the back-substitution of Table D, but in this case for the corrections, which Doolittle named w_2, x_2, y_2, z_2. The corrected solutions, $w = w_1 + w_2$, $x = x_1 + x_2$, etc., were accurate to two or three digits in Doolittle's example. Note that these solutions still must be entered into the correlate equations to obtain the angle adjustments.

Another innovation, "one of the principal advantages," was a provision for including new equations and variables. Wright and Hayford [1906, 117–118] provide a clearer text than Doolittle [1878, 17–18], whose description is somewhat brief. Doolittle suggested that large problems could be solved by successively including conditions in the minimization

---

[45] This process has both practical and theoretical importance in numerical analysis [Higham, 2002, 231, Chap. 12]. In matrix notation that is consistent with the equations (13), suppose $\bar{x}$ is an approximate solution to the simultaneous linear equations $Ax + b = 0$, where A and b are a given matrix and vector, and where x is the true vector solution. The correction to $\bar{x}$ can be obtained by solving $Ae + r = 0$ where $r = A\bar{x} + b$ is called the *residual* vector. It is straightforward to see that $x = \bar{x} + e$, but the calculation of e will incur some errors, so in principle the correction process may need to be repeated — hence *iterative* improvement.

Table E

1.	2.	3.	4.
r_1	r_2	r_3	r_4
column sum = s_1	$s_1 \times -\dfrac{[ab]}{[aa]}$	$s_1 \times -\dfrac{[ac]}{[aa]}$	$s_1 \times -\dfrac{[ad]}{[aa]}$
	column sum = s_2	$s_2 \times -\dfrac{[bc,1]}{[bb,1]}$	$s_2 \times -\dfrac{[bd,1]}{[bb,1]}$
		column sum = s_3	$s_3 \times -\dfrac{[cd,2]}{[cc,2]}$
			column sum = s_4

Table F

w	x	y	z
$s_1 \times \dfrac{-1}{[aa]}$	$s_2 \times \dfrac{-1}{[bb,1]}$	$s_3 \times \dfrac{-1}{[cc,2]}$	$s_4 \times \dfrac{-1}{[dd,3]}$
$z_2 \times -\dfrac{[ad]}{[aa]}$	$z_2 \times -\dfrac{[bd,1]}{[bb,1]}$	$z_2 \times -\dfrac{[cd,2]}{[cc,2]}$	column sum = z_2
$y_2 \times -\dfrac{[ac]}{[aa]}$	$y_2 \times -\dfrac{[bc,1]}{[bb,1]}$	column sum = y_2	
$x_2 \times -\dfrac{[ad]}{[aa]}$	column sum = x_2		
column sum = w_2			

Fig. 10. How Doolittle performed iterative refinement. Table E corresponds to the final column of Table B. The calculations in Table B apply to the constant terms, while in Table E they apply to the residuals. Table F is similar in construction to Table D.

problem, that is, by appending equations and variables to the normal equations. He recommended ordering the conditions to preserve zeroes in the elimination, and he suggested an ordering based on the geometric interpretation of the conditions. Doolittle thus anticipated the work on sparse matrix factorizations that would be done a hundred years later, for example by George et al. [1981] and Duff et al. [1986].

Dwyer [1941a, 112] remarked that "from Doolittle down to the present" no formal proof had ever been offered that Doolittle's tables do solve the normal equations. Some justification is needed because Doolittle did not explicitly reduce to zero the coefficients of eliminated variables. For example, the combination of the first and second equations with the third involves a choice of multipliers to eliminate w and x. In Rows 8 and 9 of Table B, Doolittle ignored the variables to be eliminated, yet he applied the proper multipliers to update coefficients for the retained variables y and z. This saving is possible because, thanks to the underlying symmetry, Doolittle knew that the multipliers are also the coefficients of his "explicit functions." Dwyer [1941b] gave one proof, and other explanations could be given, such as expanding Gauss's bracket formulas for the coefficients in his equations $A = 0$, $B = 0$, ... The relationship between Gauss's brackets and Gaussian elimination seems to have been known, as evidenced by Chauvenet [1868, 530]. Although Doolittle did not express his method symbolically, he undoubtedly had the training and ability to do so.[46]

Doolittle's reticence in publishing his method may be because discussing computing methods per se was thought to be neither appropriate (as judged by the mathematical community) nor desirable (from the standpoint of the computer). Grier [2005, 156] concluded from his study of hand computers that until the 20th century computing was a craft skill passed from masters to apprentices. A reluctance to disclose methods is consistent with Grier's picture of journeymen computers. Indeed, Schott [1879, 93] emphasized that Doolittle's paper of 1878 was written at the express request of the Coast Survey Director. The Washington Star [1913] reported that Doolittle "contributed numerous papers on his favourite subject" to the Philosophical Society of Washington, yet of Doolittle's three

---

[46] After receiving a Bachelor's degree at Antioch College, Doolittle taught mathematics there, and then he studied under Benjamin Peirce at Harvard College.

publications listed in the bibliography compiled by Gore [1889, 365], calculating was discussed only in [Doolittle, 1878].

Doolittle's 1878 paper made a strong impression on computers. The claims made for the "Coast Survey method" by Doolittle and by Schott [1879] caused Werner [1883] to examine the paper immediately and skeptically. He calculated Doolittle's tables three different ways: with Crelle's multiplication tables, with logarithms, and with the Thomas Arithmometer, an early calculating machine.[47] The Doolittle tables evidently passed muster, because they were reprinted for many years. Jordan [1895, iv, 65] remarked on the "old, classic" notation of Gauss and then exhibited without attribution a calculation using logarithms and Doolittle's Table B.[48] Wright and Hayford [1906] offered the Gauss brackets and the Doolittle tables as competing approaches. They did not understand the subtleties, because they touted the use of Crelle's three-digit tables (p. 120), yet they omitted the iterative refinement.

Doolittle continued to be associated with computing in the 20th century, even after computers adopted mechanical calculators and least squares became known as regression analysis. Howard Tolley, a computer at the Coast and Geodetic Office who became an official at the United States Department of Agriculture, politely scolded economists and statisticians when they neglected to credit geodesists [Grier, 2005, 159–164]. Tolley and Ezekiel [1927, 497] cited several textbooks teaching Doolittle's method, and they illustrated what were still essentially Doolittle's tables by a calculation that, by then, was done with a Monroe four-function electric calculator.[49] Paul Dwyer [1941a] (who is discussed in Section 5) found enough similarities between Doolittle's calculations and a class of matrix computing methods to name the methods after Doolittle, in which form Doolittle is commemorated to this day.

4.4. Mechanical calculators

Just as multiplication tables made possible Doolittle's method in the last quarter of the 19th century, calculating machines enabled other methods at the beginning of the next century.[50] The most pressing need was help with multiplication, a tedious operation and the one most frequently encountered. The multiplying machines most common in scientific calculations were based first on Leibniz stepped drums and then on pinwheel gears [Williams, 1982, xiv–xv].[51]

The first commercial multiplier was the Thomas Arithmometer [van Eyk, 1854], based on Leibniz drums. Widespread acceptance came only in the late 1870s, and the mechanism was never sufficiently rugged to be entirely reliable in the 19th century [Johnson, 1997, 17 and 19]. Indeed, the machines that the Coast Survey first acquired in 1890 performed only addi-

---

[47] The greater accuracy of the latter two methods obviated the need for iterative refinement, which was neglected by later authors.

[48] Willhelm Jordan's encyclopedic reference work on geodesy was constantly updated. He credited the material to scarcely anyone by name except for Gauss. Althoen and McLaughlin [1987] explain that it is this Jordan, from Hanover, who was responsible for the Gauss–Jordan algorithm. Katz [1988] has a biographical sketch of Jordan.

[49] Tolley and Ezekiel [1927, 500] reported that Miss Helen Lee, a computer, could use a form of Doolittle's method to solve five normal equations with six-digit numbers (two whole and four fractional) in 50 minutes (or 40 minutes if rounded to two fractional digits).

[50] The same of course can be said of electronic computers later in the 20th century. Innovation in mathematical practice can thus depend on manufacturing.

[51] Baxendall [1929] and Murray [1948, 1961] describe the mechanisms, and Marguin [1994] shows pictures of the computing instruments.

tion [Grier, 2005, 93]. Thus, 19th-century technology offered little improvement over logarithm and multiplication tables.

The first mass-produced multiplying machines were based on the version of the pinwheel mechanism that Willgodt Odhner patented in 1890 [Apokin, 2001]. The explosive growth in the market for these machines soon supported many European manufacturers (and resellers) such as Brunsviga, Dactyle, and Original Odhner. Whipple [1914, 84] reports that by 1912 Brunsviga alone had manufactured 20,000 pinwheel calculators.

The ability to accumulate sums of products was concomitant with machine multiplication. Multiplying calculators performed "long multiplication" in which the multiplicand shifts, to align with different digits of the multiplier, for repeated addition to an accumulator.[52] The accumulator need not contain zero at the start, so a product could increment whatever value was already there; a sum of products could be formed in this way.

Calculations that were arranged to accumulate products were both faster and more accurate. Doolittle's Table B could be discarded because it recorded only products that were to be summed. Further, the accumulator had more fractional digits than the multiplicand register, which was also the addend for addition. Any number removed from the accumulator had to be rounded before it could be reentered later. Therefore, as Dwyer [1941a, 112] explained, summing products without reentry was "theoretically more accurate" because it gave "the approximation resulting from a number of operations rather than the combination of approximations resulting from the operations."

Electrified versions of calculators began to appear after 1920 [Swartzlander, 1995]. Availability was limited by cost during the Great Depression and then by rationing during World War II. The Mathematical Tables Project gave most computers only paper and pencils because the cost of an electric machine rivaled the annual salary of a computer [Grier, 2005, 220].[53] In the United States, just three manufacturers of multiplying machines survived after the war: Friden from San Leandro, California, Monroe of Orange, New Jersey, and the ultra-quick and -quiet Marchant "silent speed" calculators from Oakland, California [Chase, 1980].

Another multiplying device was the multiplying punch that was introduced in the 1930s. It was a variation of commercial punch card equipment, and access was again limited by high cost. These machines could solve 10 simultaneous linear equations in about four hours [Verzuh, 1949, 462]. A literature developed for using them in scientific and statistical calculations, but nothing suggests that it influenced Gaussian elimination.[54]

4.5. Cholesky: machine algorithm

The first algorithm intended for a machine appears to be that developed by the military geodesist and World War I casualty André-Louis Cholesky (Fig. 12).[55] Like Doolittle,

---

[52] In fact, contrary to paper and pencil multiplication, the accumulator and multiplier moved rather than the multiplicand.

[53] The celebrated but short-lived American Mathematical Tables Project project commenced in 1938 to aid the unemployed and closed in 1948, having aided the war effort.

[54] For examples of punch card computing methods besides Verzuh see Brandt [1935] and Eckert [1940]. The latter was the famous "Orange Book" of calculating methods for punch card equipment.

[55] For a biography and a discussion of his work see Brezinski and Gross-Cholesky [2005]; see also Brezinski [2006].

Fig. 11. Myrick Hascall Doolittle, 1830–1911, circa 1862. Courtesy of the Antiochiana collection at the Olive Kettering Library of Antioch University.

Fig. 12. André-Louis Cholesky, 1875–1918, at the start of his student days at the École Polytechnique, 1895–1897. Courtesy of the Archives of the Bibliothèque Centrale of the École Polytechnique.

Cholesky calculated Gauss's angle adjustments, which were formulated as an underdetermined least squares problem (Case 2 of the least squares method described in Section 3.2). Commandant Benoit [1924] published his colleague's method posthumously.[56] A similar but not identical manuscript was recently found among Cholesky's military papers and has now been published [Cholesky, 2005].

Although Cholesky's method is most simply expressed using matrices, which were known when he developed it, he did not use them, so his invention is better understood in algebraic notation. Cholesky wrote the condition equations as [Benoit, 1924, Eq. 1]

$$a_1 x_1 + a_2 x_2 + a_3 x_3 + \ldots + a_n x_n + K_1 = 0$$

$$b_1 x_1 + b_2 x_2 + b_3 x_3 + \ldots + b_n x_n + K_2 = 0$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots \tag{14}$$

$$\ell_1 x_1 + \ell_2 x_2 + \ell_3 x_3 + \ldots + \ell_n x_n + K_p = 0,$$

---

[56] Benoit served with Cholesky as a military geodesist and also published a eulogy for him.

where $n > p$. He wrote the normal equations as [Benoit, 1924, Eq. 5],

$$a_1^1 \lambda_1 + a_1^2 \lambda_2 + \ldots + a_1^p \lambda_p + K_1 = 0$$
$$a_2^1 \lambda_1 + a_2^2 \lambda_2 + \ldots + a_2^p \lambda_p + K_2 = 0$$
$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \tag{15}$$
$$a_p^1 \lambda_1 + a_p^2 \lambda_2 + \ldots + a_p^p \lambda_p + K_p = 0,$$

where $a_j^k = a_k^j$ is the sum of products of coefficients in the jth and kth conditions; for example, $a_2^p = b_1 \ell_1 + \ldots + b_n \ell_n$.

Cholesky's remarkable insight was that, because many underdetermined systems share the same normal equations, for any normal equations there may be some condition equations that can be directly solved more easily [Benoit, 1924, 70]. He found his alternate equations in the convenient triangular form (introducing new unknowns, y in place of x, and new coefficients, β in place of a)[57]

$$\beta_1^1 y_1 \qquad\qquad\qquad\qquad + K_1 = 0$$
$$\beta_2^1 y_1 + \beta_2^2 y_2 \qquad\qquad\qquad + K_2 = 0$$
$$\beta_3^1 y_1 + \beta_3^2 y_2 + \beta_3^3 y_3 \qquad\qquad + K_3 = 0 \tag{16}$$
$$\vdots \qquad \vdots \qquad \vdots \qquad\quad \ddots \qquad \vdots$$
$$\beta_p^1 y_1 + \beta_p^2 y_2 + \beta_p^3 y_3 + \ldots + \beta_p^p y_p + K_p = 0.$$

Cholesky discovered that the coefficients in Eq. (16) are given by straightforward formulas [Benoit, 1924, 72]:

$$\beta_i^i = \sqrt{a_i^i - (\beta_i^1)^2 - (\beta_i^2)^2 - \ldots - (\beta_i^{i-1})^2}$$
$$\beta_{i+r}^i = \frac{a_{i+r}^i - \beta_i^1 \beta_{i+r}^1 - \beta_i^2 \beta_{i+r}^2 - \ldots - \beta_{i+r}^{i-1} \beta_{i+r}^{i-1}}{\beta_i^i}. \tag{17}$$

The solution λ of the normal equations expresses the solution of the condition equations as a linear combination of the (transposed) coefficients in the condition equations. For Cholesky, these combinations become a system of equations to be solved for λ [Benoit, 1924, Eq. 7]:

$$y_1 = \beta_1^1 \lambda_1 + \beta_2^1 \lambda_2 + \ldots + \beta_p^1 \lambda_p$$
$$y_2 = \qquad\quad \beta_2^2 \lambda_2 + \ldots + \beta_p^2 \lambda_p$$
$$\vdots \qquad\qquad\qquad \ddots \qquad \vdots \tag{18}$$
$$y_p = \qquad\qquad\qquad\qquad \beta_p^p \lambda_p$$

---

[57] Benoit [1924] unfortunately used α for the new coefficients, which here has been replaced by β to distinguish them more easily from the original coefficients a.

		λ_1	λ_2	...	λ_{i-1}	λ_i	...	λ_{p-1}	λ_p	K	λ
1	β_1^1	a_1^1	a_1^2		a_1^{i-1}	a_1^i		a_1^{p-1}	a_1^p	K_1	λ_1
2	β_2^1	β_2^2	a_2^2		a_2^{i-1}	a_2^i		a_2^{p-1}	a_2^p	K_2	λ_2
i	β_i^1	β_i^2	β_i^3		β_i^i	a_i^i		a_i^{p-1}	a_i^p	K_i	λ_i
								a_{p-1}^{p-1}	a_{p-1}^p		
p	β_p^1	β_p^2	β_p^3		β_p^i			β_p^p	a_p^p	K_p	λ_p
y	y_1	y_2	y_3		y_i			y_p			

Fig. 13. How Cholesky may have organized the calculation to solve the normal equations by what is now known as the Cholesky factorization, from the foldout table of Benoit [1924]. The left column and top row are labels; the other positions would be occupied by numbers. The coefficients, *a*, and the constants terms, *K*, of the normal equations are placed for reference above the diagonal. The coefficients, *β*, of Cholesky's manufactured condition equations are written below the diagonal. The solution, *y*, of his condition equations stands in the bottom row. The solution, *λ*, of the normal equations is recorded in the right column. Some additional rows and columns for arithmetic checks and for accuracy estimates have been omitted.

Because the original condition Eqs. (14) and Cholesky's Eqs. (16) have the same normal equations (15), the quantities λ are the same for both problems. Once obtained, λ can be used to evaluate x as usual (from the transpose of the original coefficients), thereby solving the original condition equations. Thus Cholesky's method was first to form his new coefficients using Eq. (17), then to solve (16) by forward substitution for y, next to solve (18) by backward substitution for λ, and finally to evaluate x:

$$
\begin{aligned}
x_1 &= a_1\lambda_1 + b_1\lambda_2 + \ldots + \ell_1\lambda_p \\
x_2 &= a_2\lambda_1 + b_2\lambda_2 + \ldots + \ell_2\lambda_p \\
x_3 &= a_3\lambda_1 + b_3\lambda_2 + \ldots + \ell_3\lambda_p \\
&\;\;\vdots \qquad\quad \vdots \qquad\quad \vdots \qquad\qquad \vdots \\
x_n &= a_n\lambda_1 + b_n\lambda_2 + \ldots + \ell_n\lambda_p.
\end{aligned}
\tag{19}
$$

Benoit published Cholesky's method in his own format with revised notation and with a table showing how the calculation was conducted (Fig. 13). He indicated that the ith column of new coefficients (from β_i^i to β_p^i) should be calculated before going on to the next column. The table is more compact than Doolittle's tables because much of the intermediate work is not recorded: each β_{i+r}^i is a sum of products that can be accumulated with a machine. Both Benoit and Cholesky mentioned using calculators to form these sums, and Cholesky specifically referred to the Dactyle brand of the Odhner design. Cholesky [2005] reported solving 10 equations with five-digit numbers in four to five hours: he was as quick as a multiplying punch, although he worked with fewer digits.

For 20 years after the publication by Benoit [1924], Cholesky's method was seldom discussed. It was, however, used in Sweden during this time by the geodesist Tryggve Rubin [1926], who augmented it to produce other statistically relevant quantities.[58]

> The normal equations have been solved by the Cholesky-Rubin method, which offers the advantage that the solution is easily effected on a calculating machine (Cholesky) and that the most probable values and their mean errors are derived simultaneously (Rubin).
> — Ahlmann and Rosenbaum [1933, 30]

Jensen [1944, 22], writing in a Danish geodetic publication, remarked that Cholesky's method of solving the normal equations "ought to be more generally used." Soon thereafter it was independently discovered in matrix form by Dwyer and was put to use in America (see Section 5).

4.6. Crout: "each element is determined by one continuous machine operation"

Prescott Crout (Fig. 14) was a professor of mathematics at the Massachusetts Institute of Technology with an interest in mathematics for electrical engineering. In a paper that is a model of brevity, Crout [1941a] listed the gamut of uses for linear equations that had developed after least squares, and he described a new method of solving general systems of equations rather than only normal equations from the method of least squares. Like Cholesky, Crout formulated his method to emphasize sums of products.

Crout [1941a, 1239] explained "the method was originally obtained by combining the various processes which comprise Gauss's method, and adapting them for use with a computing machine." He wrote the coefficients in a rectangular matrix. Crout meant "matrix" in the sense of "table" with the constant terms in the final column; he did not employ matrix algebra. His method consisted of a few terse rules for transforming the numbers. So much had changed in the application of Newton's and Rolle's elimination rules that all semblance of symbolic algebra had vanished. Crout's pithy instructions are here restated more expansively.

1. The first column is left unchanged. The first row to the right of the diagonal is divided by the diagonal entry.
2. An entry on or below the diagonal is reduced by the sum of products between entries to the left in its row and the corresponding entries above in its column. This calculation is permitted only after all those entries themselves have been transformed.
3. As in 2, similarly for an entry above the diagonal except it is lastly divided by the diagonal entry in its row.
4. On completing the above teps (condensing further instructions), conduct a back-substitution using the coefficients above the diagonal and the final column of transformed constants.

These steps were obviously intended for use by computers who had calculators that could accumulate sums of products. Like Doolittle before him, Crout included instructions to improve the accuracy of the solution by an unnamed iterative refinement process.[59] He

---

[58] Brezinski [2006, 286] claims incorrectly that after Benoit [1924] "a period of 20 years followed without any mention of the work."

[59] Wilkinson [1960, 49] commented that the sum of product methods as originally stated could give poor results because they omitted provisions to reorder the equations and variables, such as stated in Section 1.2. This contingency is unnecessary for normal equations, but it can be important for others.

Fig. 14. Prescott Durand Crout, 1907–1984, circa 1936. Courtesy MIT Museum.

Fig. 15. Paul Sumner Dwyer, 1901–1982, circa 1960s. Courtesy of Prof. Ingram Olkin of Stanford University.

concluded the paper with a rigorous proof that the method necessarily solves the intended equations.

In the same year Paul Dwyer [1941a, 108] (Fig. 15) published a survey of computing methods which included a method similar to Crout's, the "abbreviated method of single division." Dwyer did not find the method in the literature but rather synthesized it to complete his classification scheme: for normal equations ("symmetric") or not, with "single division" or not, and "abbreviated" or not. The latter was Dwyer's name for methods that accumulated sums. Waugh and Dwyer [1945, 259] reported "great recent interest" in such "compact" methods, but they cited no recent literature other than their own papers and those of Prescott Crout.

Crout's method was rapidly adopted by those with access to calculators, and in the following decade his name appeared dozens of times in the literature. Crout himself reprised his paper [Crout, 1941b] for a series of manuals issued by the Marchant Company, a leading calculator manufacturer, and the Company prepared very detailed instructions for Crout's method in a subsequent manual [Marchant Calculating Machine Company, 1941]. Wilson [1952, 352] particularly recommended the Marchant manuals to academic researchers. Black [1949] cited both Crout [1941a] and Doolittle as interpreted by Dwyer [1941b] in describing how to use calculators for engineering work.

5. Describing Gaussian elimination with matrices

5.1. Overview of matrix descriptions

The milieux of Gauss, Doolittle, Cholesky, and Crout, who worked on Gaussian elimination through symbolic algebra, was eventually brought to an end by those who worked with matrix algebra. The interpretation of Gaussian elimination through matrices led to a consolidation by showing how all the computing variations were trivially related through what are now called matrix decompositions. Henceforth one formalism would suffice for all methods.

Von Neumann [1954, 488] cited matrices as an example of the delay between a mathematical discovery and its use outside mathematics. Hawkins [1975, 1977a, 1977b] found that matrix algebra was independently invented by Eisenstein (1852), Cayley (1858), Laguerre (1867), Frobenius (1878), and Sylvester (1881) to clarify subjects such as determinants, quadratic forms, and the elementary divisors that Weierstrass used to study ordinary differential equations. The first significant application of matrices outside pure mathematics was Heisenberg's matrix (quantum) mechanics in 1925. Although accounts of Gaussian elimination without matrix notation continued well into mid-century,[60] from the end of the 1930s several authors offered matrix descriptions within a short span of time. The citation analysis in Fig. 16 identifies six groups or individuals who made unprecedented contributions: Banachiewicz, Frazer et al., Jensen, Dwyer, Satterthwaite, and von Neumann and Goldstine.

5.2. Toeplitz: first matrix decomposition

David Hilbert's study of integral equations inspired his students Erhard Schmidt and Otto Toeplitz to promote infinite matrices as a representation for linear operators on function spaces.[61] As part of that research, Toeplitz used determinants to examine the invertibility of infinite matrices, "If one uses the symbolism of matrix calculus (see for example Frobenius)," then any finite symmetric matrix S with all leading principal determinants not zero has a matrix U with

$$U'U = S^{-1} \quad \text{equivalently} \quad U^{-1}U'^{-1} = S \tag{20}$$

(original notation), where $'$ is transposition [Toeplitz, 1907, 102].[62] In this way Toeplitz first exhibited what now is called the Cholesky factor, U^{-1}, though he left the entries in a computationally useless form given by determinants. It was clear from the determinantal formulas that U was a lower triangular matrix. Taussky and Todd [2006, 198] suggested that Eq. (20) was the first expression of any such formula in matrix notation. Moreover, the equation may be unique to Toeplitz because his U and S are related by inversion. Taussky and Todd thought a proof might be based on formulas of Gantmacher [1959, 39].

---

[60] For example Aitken [1937, 1951], Crout [1941a, 1941b], Dwyer [1941b], Hotelling [1943], Bargmann et al. [1946, 422–423], and MacDuffee [1949].

[61] Another Hilbert protégé, John von Neumann, showed the infinite matrix approach to be fundamentally inadequate [Bernkopf, 1968, 330].

[62] Toeplitz remarked that Lagrange and Gauss knew such a decomposition for quadratic forms, and he cited Jacobi [1857] for a similar decomposition of bilinear forms.
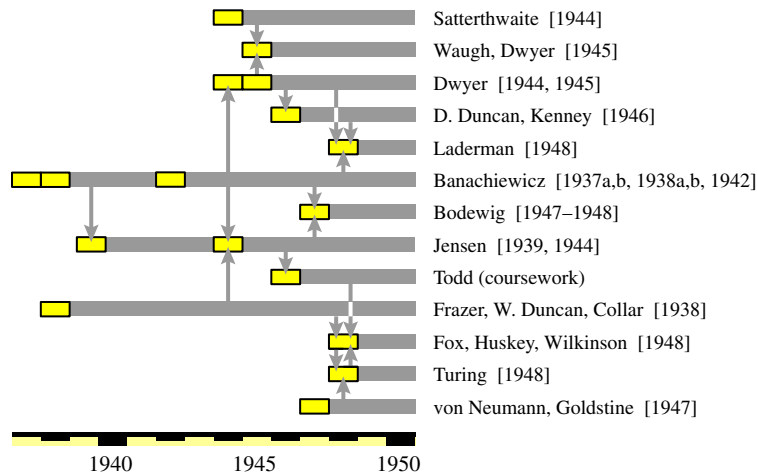
Fig. 16. Citation patterns among the earliest authors who mention Gaussian elimination in the form of matrix decomposition. Arrows show the flow of information to the citing papers at the head from the cited oeuvre at the tail. For example, Dwyer in 1944 cited one or more works of Banachiewicz. Boxes indicate published papers, except in the case of Todd, who taught a course in 1946 that is described by Taussky and Todd [2006].

5.3. Banachiewicz: Cracovian algebra

One path to matrix algebra that has not been recognized is a motivation deriving from computing that prompted the Cracovian algebra of the polymath Tadeusz Banachiewicz (Fig. 17).[63] Cracovians began as a notational scheme for astronomical calculations and became an algebra different from Cayley's for matrices of arbitrary dimension. The distinguishing feature of the Cracovian algebra is a column-by-column product that is more natural when calculating with columns of figures by hand.

> It must, however, be conceded that in practice it is easier to multiply column by column than to multiply row by column .... It may, in fact, be said, that the computations are made by cracovians and the theory by matrices.
>
> — Jensen [1944, 5]

Banachiewicz posed least-squares problems in terms of Cracovians for the purpose of improving computations.[64] Jensen [1939, 3, 19] reports hearing Banachiewicz advocate this approach at meetings of the Baltic Geodetic Commission as early as 1933. Banachiewicz [1938b] independently discovered Cholesky's method, and although later than Cholesky, he initially had greater impact. Banachiewicz inspired the work of Jensen, and he was widely cited: by Jensen [1944, 45], Dwyer [1944, 89], Cassinis [1946, 78], Bodewig [1947–1948, Part V, 90], Laderman [1948], again by Dwyer [1951, 103], and by Forsythe [1953, 301]. Nevertheless, he was neglected by the influential computational mathematician Householder [1956, 1964].

---

[63] The first of several biographical sketches of Banachiewicz, a concentration camp survivor, was given by Witkowski [1955].

[64] For descriptions of the algorithms and additional references to the original papers, see Kocinski [2004].

Fig. 17. Tadeusz Banachiewicz, 1882–1954, at the 1946 meeting of the International Astronomical Union in Copenhagen. Courtesy of Professor Adam Strzałkowski of the Jagiellonian University.
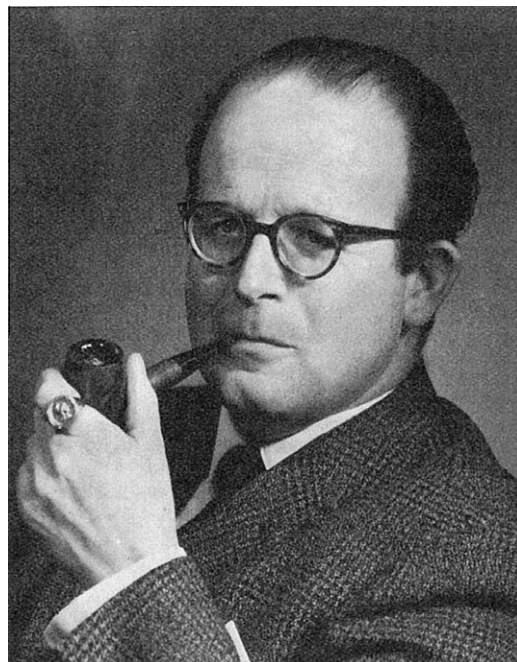
Fig. 18. Henry Jensen, 1915–1974. Courtesy University of Copenhagen.

5.4. Frazer, Duncan, and Collar: elementary matrices

The work of R.A. Frazer, W.J. Duncan, and A.R. Collar exemplifies the practical aspects of the source for matrix algebra that Hawkins [1977b] finds in Weierstrass's study of linear differential equations.[65] In this case the stimulus for Frazer's interest in matrices may have been Henry Baker, with whom Frazer studied as a Cambridge undergraduate.[66] Collar and Duncan used matrix notation in their approximate formulations of structural problems in the precursor to the finite element method [Felippa et al., 2001]. The three collaborators thus found it natural to translate into questions about matrices the problems they studied for Frazer's aerodynamics section at the National Physical Laboratory, notably problems of airframe vibration or flutter. Their textbook about matrices in engineering analysis became an international standard [Pugsley, 1961, 80].

Frazer et al. [1938, 96–99] viewed elimination as "building up the reciprocal matrix in stages by elementary operations," which could produce a triangular matrix "such that its reciprocal can be found easily." They demonstrated column elimination of a 4×4 matrix a (their notation). They had $aM_1M_2M_3 = \tau$, where the M_i are "post multipliers" that effect the elementary operations and τ is "the final triangular matrix," and they remarked that

---

[65] See Pugsley [1961], Relf [1961], and Bishop [1987] for biographies of Frazer, Duncan, and Collar, respectively.

[66] Pugsley [1961] remarks on the influence of Cayley's student Baker, whose work included matrix formulations of differential equations, such as Baker [1903].

$M_1M_2M_3$ was itself a triangular matrix but "opposite-handed" from τ. However, they did not continue their presentation to the modern conclusion, neither commenting on matrix factoring nor writing out a factorization such as $a = \tau(M_1M_2M_3)^{-1}$.

The modern literature retains the emphasis on elementary operations that are accomplished through multiplication by *elementary matrices*. Jensen [1944, 13–15] borrowed the approach to establish the connection between Gaussian elimination and triangular factoring. He restated it in the more conventional form of row operators acting on the left but using the same notation M_i. Textbooks now use this row exposition in terms of "premultiplication" to establish the relationship between Gaussian elimination and matrix factoring, although they attribute it to neither Frazer et al. nor Jensen. For example, Golub and Van Loan [1996, 94–95] use, remarkably, the same notation M_i but ahistorically call it a "Gauss transformation."

5.5. Jensen: synthesis and conduit

Two groups of papers, by Jensen[67] (Fig. 18) and by Bodewig, helped to establish the matrix interpretation of Gaussian elimination by describing many algorithms independent of origin in a common notation. Jensen [1939, 3, 19] characterized his contribution as extending to "matrix symbolism" the results of Banachiewicz for least squares problems and for the normal equations. Jensen [1944, 11] found that three algorithms for solving the normal equations were similar in that they could be interpreted as "reducing the matrix in question to a triangular matrix:" the "Gauss'ian algorithm," the Cracovian method, and Cholesky's method.[68] To emphasize the similarities among the triangular factoring algorithms, Jensen used pictograms for triangular matrices with zeroes either *under* $\triangledown$ or *over* $\triangle$ the main diagonal (original terminology and pictures). His primary interest was the normal equations with coefficient matrix $A^*A = N$ for a rectangular matrix A, where $*$ is transposition. Jensen [1944, 15, Eq. 15; 22, Eq. 3] explained that the "Gauss'ian algorithm" amounted to $N = \triangle\triangledown$, where $\triangle$ and $\triangledown$ are not related by transposition (he refers to Gauss's calculation for the symmetric normal equations), while in Cholesky's method the factors are so related.

What Jensen [1944, 13–16] called the "Gauss'ian algorithm" was his original synthesis of three different approaches. He began with his own row-oriented version of Frazer et al.'s transformation of a matrix to triangular form. He conducted the transformation symbolically by using Gauss's brackets as the matrix entries, thereby connecting the transformations to the work of Gauss. Although Jensen did not mention Doolittle, he recommended that the calculated numbers should "be conveniently tabulated" not in matrices but rather in a form that resembled Doolittle's Table B.

Jensen's survey was evidently read. The use of the row formulation of Frazer et al.'s elementary transformations in modern textbooks has already been noted. Similarly, Jensen's recommendation that Cholesky's method "ought to be more generally used than is the case" [Jensen, 1944, 22] was heard by Fox et al. [1948] and by Turing [1948] through John Todd, who taught the method from Jensen's paper in 1946 [Taussky and Todd, 2006, 197].

---

[67] Henry Jensen did his work on elimination at the Danish Geodætisk Institut, and he later joined the faculty of the University of Copenhagen.

[68] The other, nontriangular methods that Jensen discussed were solution by determinants, the method of equal coefficients, Boltz's method, and Kruger's method. He attributed the method of equal coefficients to B.I. Clasen in 1888; it is now more commonly attributed to Jordan as the Gauss–Jordan method, cf. footnote 48.

5.6. Dwyer: abbreviated Doolittle method and square root method

Paul Sumner Dwyer (Fig. 15) was a professor at the University of Michigan and a president of the Institute of Mathematical Statistics.[69] Dwyer collaborated with an official at the United States Department of Agriculture named Frederick Waugh.[70] The Department participated in creating a field of economics, known as econometrics [Fox, 1989], whose research methodology then consisted of data analysis by, essentially, the method of least squares. The partnership with Waugh began in the 1940s and exposed Dwyer to computers financed by government resources. This meant he could expect calculations to be mechanized, which placed him at the forefront of computing practice.

Dwyer [1941a] began from a comparison of solution methods that emphasized sources in American and English statistics. He considered 22 papers from 1927 to 1939 and suggested that their bibliographies should be consulted for an even more thorough picture of the subject. Dwyer's review painstakingly uncovered the similarities between proliferating methods distinguished by minor changes in the placement of numbers in tables. For example, he noted that a method of Deming [1928] was equivalent, except for superficial changes, to a "method of pivotal condensation" of Aitken [1937], both of which he included under the rubric "method of single division" (Fig. 19). The differences among these methods seem pedestrian until one has to choose the most effective way to calculate and record all the numbers in the tables by hand. Nevertheless, in another paper, Waugh and Dwyer [1945, 260] summarized the field more succinctly, observing that the methods are more or less the same, except that "Crout divides the elements of each row by the leading element while we divide the elements of columns." Dwyer [1941a, 111–112] credited to others, including Waugh, the observation that accumulating calculators made it unnecessary to record the series terms in Doolittle's Table B. Dwyer called the streamlined procedure the "abbreviated method of single division — symmetric" or the "abbreviated Doolittle method."

Dwyer [1944] independently interpreted Gaussian elimination as matrix factoring. His primary interest was the Case 1 least-squares problem described in Section 3.2. Beginning from the coefficient matrix A of the normal equations, he showed that the abbreviated Doolittle method was an "efficient way of building up" some "so called triangular" matrices S and T with $A - S^t T = 0$ [Dwyer, 1944, 86]. Rather than as the product of the elementary matrices of Frazer et al. [1938] and Jensen [1944], Dwyer obtained the decomposition from a sum of vector outer-products [Dwyer, 1944, 86, Eqs. 21, 23]. He remarked that this formula would be the "key" to "a more general theory" [ibid., 8] of decompositions, which Waugh and Dwyer [1945] subsequently developed to invert nonsymmetric matrices. For the normal equations it was possible to choose $S = T$ for a "square root" method [Dwyer, 1944, 88] (modern Cholesky method), which Dwyer [1945] developed in a later paper. He added that he found no other matrix algebra interpretation of solving equations except from Banachiewicz [1942] who, he noted, also had a square root method [Dwyer, 1944, 89].

Dwyer particularly influenced computers in the United States. Laderman [1948] reported that Duncan [1946] popularized Dwyer's square root method, and that it was even used at the Mathematical Tables Project. European mathematicians such as Fox [1954], however, preferred to apply Cholesky's name. Since Dwyer's many papers and his book on linear

---

[69] For biographical information on Dwyer, see the article by his longtime colleague Cecil Craig [1972].

[70] Fox [1989, 68–69] summarizes Waugh's career at the Department.

	x_1	x_2	x_3	x_4	r. h. s.	
1	1.0000	.4000	.5000	.6000	.2000	original equations
2	.4000	1.0000	.3000	.4000	.4000	
3	.5000	.3000	1.0000	.2000	.6000	
4	.6000	.4000	.2000	1.000	.8000	
$1 \Rightarrow$ 5	1.0000	.4000	.5000	.6000	.2000	elimination, part 1
$5,2 \Rightarrow$ 6		.8400	.1000	.1600	.3200	
$5,3 \Rightarrow$ 7		.1000	.7500	$-.1000$	.5000	
$5,4 \Rightarrow$ 8		.1600	$-.1000$	.6400	.6800	
$6 \Rightarrow$ 9		1.0000	.1190	.1905	.3810	elimination, part 2
$9,7 \Rightarrow$ 10			.7381	$-.1190$	.4619	
$9,8 \Rightarrow$ 11			$-.1190$	.6095	.6190	
$10 \Rightarrow$ 12			1.0000	$-.1612$	.6258	elimination, part 3
$12,11 \Rightarrow$ 13				.5903	.6935	
$13 \Rightarrow$ 14				1.0000	1.1748	back-substitution
$11,14 \Rightarrow$ 15			1.0000		.8152	
$9,14,15 \Rightarrow$ 16		1.0000			.0602	
$5,14,15,16 \Rightarrow$ 17	1.0000				$-.9366$	

Fig. 19. A form of Gaussian elimination that Paul Dwyer [1941a] called "the method of single division" and which he found equivalent, except for cosmetic changes, to the "method of pivotal condensation" of Aitken [1937], and to an earlier method of Deming [1928]. The nonunitary numbers in rows 5, 9, and 12 are the upper diagonal entries in Crout's table.

equations [Dwyer, 1951] always invoked the memory of Doolittle, Dwyer's own name was never attached to either of the computing methods that he championed.

5.7. Satterthwaite: three-factor decomposition

Dwyer [1941a] had concluded his survey of methods with a process for solving equations with multiple right sides that he attributed to R.A. Fisher. The presentation inspired the statistician Franklin Satterthwaite [1944] to calculate inverse matrices.[71] He linked the tables of the calculation to inverses through the triple "factorization" $A = (R_1 + I) S_1 (I + T_1)$, where S_1 is diagonal, and R_1 and T_1 are respectively "pre-" and "postdiagonal" (strictly triangular) [Satterthwaite, 1944, 374, Eq. 3.1; 376, Sect. 5]. In consideration of having to round the numbers that were written into the tables, Satterthwaite showed that the process was accurate provided that A was close to being an identity matrix. He therefore suggested using the process to improve inverses: if $F \approx A^{-1}$, then calculate $(FA)^{-1}F$. His paper appears to have had little influence. It received only 10 citations, of which 4 were by Dwyer and Waugh.

5.8. Bodewig: literature surveys

The mathematician Ewald Bodewig (Fig. 20) followed Jensen's approach on a grand scale in a five-part paper that summarized in matrix notation all methods for solving linear equations that were known up until 1947 [Bodewig, 1947–1948].[72] His interesting bibliog-

---

[71] Satterthwaite [1944] published between Dwyer [1944] and Waugh and Dwyer [1945] and in the same journal, but he cited only Dwyer's papers from 1941.

[72] During the same period Bodewig [1944, 1948, 1949] wrote evaluations for *Mathematical Reviews* of all three of the papers by Jensen [1944], von Neumann and Goldstine [1947], and Turing [1948]. Shortly thereafter Bodewig became distraught over the paltry wages paid to mathematicians and announced his semiretirement from mathematics in a letter that received much comment [Bodewig, 1950; Betts et al., 1951].

Fig. 20. Ewald Konrad Bodewig. Courtesy Archives of the Mathematisches Forschungsinstitut Oberwolfach.

Fig. 21. John von Neumann, 1903–1957, circa March, 1947. Courtesy of the Los Alamos National Laboratory Archives.

raphy lay at the end of the paper but, like Jensen's, it was not comprehensive because he neglected authors such as Crout, Doolittle, and Dwyer. He took Jensen too literally, because he labeled triangular matrices "left" or "right" in order to indicate the location of the *zeroes* instead of the location of the *non*zeroes: Bodewig's "right" was $\searrow = \mathfrak{D}_r$ and "left" was $\triangledown = \mathfrak{D}_l$ where $\mathfrak{D}$ stood for Dreiecksmatrix [Bodewig, 1947–1948, Part I, 444]. He repeated Jensen's row version of Frazer et al.'s presentation of Gaussian elimination, and he emphasized their summary formula using Jensen's pictograms, $\searrow\square = \triangledown$ [Bodewig, 1947–1948, Part I, 449]. He also followed Jensen [1944] in describing Cholesky's method as $\mathfrak{S} = \mathfrak{D}_l'\mathfrak{D}_l$ for a symmetric $\mathfrak{S}$, where $'$ is transpose [Bodewig, 1947–1948, Part I, 450].

5.9. Von Neumann and Goldstine: the combination of two tricks

Among the first authors to describe Gaussian elimination in terms of matrix algebra, John von Neumann (Fig. 21) and his collaborator Herman Goldstine were alone in making a nontrivial use of matrix decomposition. They devised a decomposition method related to Gaussian elimination that was appropriate for use with the first electronic computers, which were then being built. They established bounds on the rounding errors of the computation in terms of the ratio of the largest to the smallest singular values of the coefficient matrix. Their result, which is beyond the scope of the present discussion, marks the beginning of modern research in computational mathematics. The ratio that they discovered is now called the matrix condition number.

Von Neumann and Goldstine were also the only authors to show exactly how the "traditional schoolbook method" [Farebrother, 1988, 3] of Eq. (1) calculates triangular factors of the coefficient matrix. They described the elimination algorithm as creating a sequence of

ever-smaller reduced matrices and vectors from $Ax = y$ [von Neumann and Goldstine, 1947, 1051],

$$A = A^{(1)},\ A^{(2)},\ A^{(3)},\ \ldots,\ A^{(n)}, \qquad y = y^{(1)},\ y^{(2)},\ y^{(3)},\ \ldots,\ y^{(n)},$$

where the rows and columns of $A^{(i)}$ and $y^{(i)}$ are numbered from i to n.[73] For $i = 1, \ldots, n-1$ the computation is

$$A_{j,k}^{(i+1)} = A_{j,k}^{(i)} - A_{j,i}^{(i)} A_{i,k}^{(i)} / A_{i,i}^{(i)} \qquad \text{for } j, k > i \qquad (21)$$

$$y_j^{(i+1)} = y_j^{(i)} - (A_{j,i}^{(i)} / A_{i,i}^{(i)}) y_i^{(i)} \qquad \text{for } j > i. \qquad (22)$$

Next, the algorithm solves by substitution the equations $B'x = z$, where the entries of B' and z are chosen from the reduced matrices and vectors (the first row of each). For a matrix C of the multipliers $A_{j,i}^{(i)} / A_{i,i}^{(i)}$ with $j \geqslant i$ (note the unit diagonal), von Neumann and Goldstine summed Eq. (22) over i and rearranged to give $Cz = y$. From this equation and $B'x = z$ they concluded that $CB' = A$.

> We may therefore interpret the elimination method as ... the combination of two tricks: First, it decomposes A into a product of two semi-diagonal matrices ... [and second] it forms their inverses by a simple, explicit, inductive process.
> — von Neumann and Goldstine [1947, 1053][74]

With this passage von Neumann and Goldstine finally expressed in a new idiom what Newton and Rolle had first explained over two hundred years earlier.

> And you are to know, that by each Æquation one unknown Quantity may be taken away, and consequently, when there are as many Æquations and unknown Quantities, all at length may be reduc'd into one, in which there shall be only one Quantity unknown.
> — Newton [1720, 60–61] and prior years

Von Neumann and Goldstine found a lack of symmetry in what is now called the "classic" algorithm because the first factor always had 1's on its main diagonal. They divided the second factor by its diagonal to obtain $B' = DB$; hence $A = CDB$, which they said was "a new variant" of Gaussian elimination [von Neumann and Goldstine, 1947, 1031], which is now written $A = LDU$,

$$L_{j,i} = A_{j,i}^{(i)} / A_{i,i}^{(i)} \qquad D_{i,i} = A_{i,i}^{(i)} \qquad U_{i,k} = A_{i,k}^{(i)} / A_{i,i}^{(i)} \qquad j, k \geqslant i, \qquad (23)$$

where $A_{j,k}^{(i)}$ are the entries of the reduced matrices given by Eq. (21).

5.10. Decompositions and diagonals

In summary, four groups of authors took conceptually different routes to interpret Gaussian elimination as constructing triangular decompositions in Cayleyan matrix algebra. Inspired by Banachiewicz, Jensen [1944] obtained a triangular decomposition from products of the elementary matrices of Frazer et al. [1938]. Dwyer [1944] expressed a matrix as a sum of vector outer-products from which a triangular decomposition followed. Satterthwaite [1944] found entries of the factors in tables used to solve equations with multiple right hand sides. Von Neumann and Goldstine [1947] chose entries of the triangular factors

---

[73] It is customary to neglect the possibility of reordering the equations and variables in presentations such as this to avoid complicating the notation.
[74] Von Neumann and Goldstine wrote "semi-diagonal" for "triangular."

directly from numbers that occurred in the "schoolbook" or symbolic algebra form of Gaussian elimination.

What accounts in part for the existence of multiple inventors of elimination algorithms is the fact that the triangular matrix decompositions are not unique. The triple term decomposition, $A = LDU$, is unique only from the requirement that all the diagonal entries of L and U equal 1. The decompositions $A = L(DU)$ and $A = (LD)U$ that differently apportion the diagonal are known today by the names Doolittle and Crout, respectively. If A is a symmetric and positive definite matrix, then $L = U$ and the decomposition $A = (LD^{1/2})(LD^{1/2})^t$ is known by the name Cholesky.[75] The Doolittle naming convention is misleading because the calculations of Doolittle — and Gauss also, of course — applied only to symmetric, positive definite matrices. All these names are anachronisms because the inventors did not use matrix notation.

6. Conclusions

6.1. Summary

An algorithm equivalent to Gaussian elimination appeared in mathematical texts from ancient China. A few instances of the method emerged during the late Renaissance in European textbooks, but since they were geographically scattered, it seems unlikely they were influenced by contact with the Orient. Newton [1720], writing circa 1670, and Rolle [1690] separately gave rules for successively eliminating variables to solve simultaneous equations, and they chose words other than *elimination* to describe their work. The features of Newton's rule can be traced through successive algebra texts from many authors, resulting in a fairly standardized presentation by 1800. Lacroix [1804] emphatically called this schoolbook process "elimination."

At the turn of the century, both Legendre and Gauss remarked that the textbook rule could be used to solve least-squares problems that were formulated as normal equations. Not content with that, Gauss [1810] devised a special solution process. His approach was adopted by professional hand computers working in geodesy, and was then replaced (Doolittle, Cholesky) to make better use of computing aids, including manual calculators. Astronomers, geodesists, and statisticians (Banachiewicz, Jensen, Dwyer) endeavored to improve least-squares calculations through studies based on matrix algebra. Nothing entirely new resulted, although Dwyer's independently discovered "square root" (Cholesky) method was widely adopted in the United States.

These methods, including that of Gauss, cannot substitute for schoolbook elimination if other than normal equations are of interest. A need to solve more general equations arose in engineering in the 20th century. Frazer et al. [1938] proposed a formalism for calculations in terms of "elementary" matrix operations, while Crout [1941a], without matrices, devised a version of the schoolbook method that was publicized by a calculator manufacturer. Dwyer [1944] and Satterthwaite [1944] observed that any matrix could be decomposed into triangular factors, which the latter and Waugh and Dwyer [1945] exploited to calculate inverse matrices. Von Neumann and Goldstine [1947] finally interpreted schoolbook elimination as a matrix decomposition for the purpose of studying the accuracy of calculations

---

[75] Similarly, what Gauss computed might be characterized as $A = (LD)D^{-1}(LD)^t$ because he only retained his bracket values, which give the entries of (LD) and those of D on the diagonal.

made by electronic computers. Gaussian elimination subsequently became the subject of a large literature patterned after their study.

6.2. Remembering Gauss imperfectly

At the beginning of the 19th century, the triumph of calculating the location of Ceres had earned the youthful Gauss fame enough to realize his wish for a life free of teaching mathematics. Nevertheless, by the end of that century, histories of mathematics would neglect Gauss's applied work.[76] Grattan-Guinness [2004, 177] explains the circumstances: the establishment of the research university as a social institution, which occurred roughly during Gauss's lifetime [Clark, 2006], increased employment for credentialed mathematicians, which both sped the professionalization of the subject and coincided with a preference for pure over applied mathematics. The same taste was manifest in historical scholarship, he adds, in that pure subjects became more thoroughly chronicled than applications.

Grattan-Guinness argues for the existence of two recollections of the mathematical past: history recounts the development of ideas in the context of contemporary associations, while heritage remembers reinterpreted work that embodies the state of mathematical knowledge. The example of Gaussian elimination cautions that much of applicable mathematics may lack both: no mathematical history because applications may be deemed the purview of nonmathematical faculties, no heritage because applications might not enter or remain in the corpus. As a result, significant developments may come from what appears to be the periphery of the mathematical community. The interpretation of Gaussian elimination through matrix algebra, which is the most important development in understanding the subject for present purposes, arose as much outside as inside mathematics.

Gauss's methods for determining orbits and solving inference problems indelibly linked calculations to his reputation among astronomers and geodesists. Whereas in the first half of the 19th century algebra textbooks referred only to elimination, from the second half of the 19th century reference books for astronomy and geodesy always cited Gauss in order to recommend their least-squares calculations. His bracket notation for what came to be called auxiliaries was correctly regarded as pertinent to the solution of just normal equations. For example, Chauvenet [1868, 530] had the "elimination of unknown quantities from the normal equations ... according to Gauss," and Liagre [1879, 557] "l'*élimination* des inconnues entre les équations du minimum (équations normales)" by "les coefficients auxiliaires de Gauss."

The invention of electronic computers led to new university faculties for instruction in the new way of computing, which drew people of diverse training to the field. They extrapolated opinions of its history from whatever heritage they knew. Geodesy at one time accounted for the bulk of the simultaneous linear equations that were solved professionally, so its terminology may have appeared to be authoritative. The citations that precisely described the contribution of Gauss had already been shortened to an unspecific "Gauss's procedure." This usage was misinterpreted as attributing ordinary, common, schoolbook elimination to Gauss, but only after World War II (Table 1). Von Neumann [1947] was apparently the last prominent mathematician to simply write "elimination" as Lacroix [1804] and Gauss [1809] had done. There is no harm in naming elimination after Gauss because he did initiate the development of professional methods which are of considerable importance today, but it should be under-

---

[76] Bühler [1981, 46] and Dunnington [2004, 405–410] explain Gauss's career choices; Matthiessen [1878] and Cajori [1894] are examples of early histories of mathematics.

Table 1
Demonstrating the gradual evolution to "Gaussian elimination," in these quotations from sources that span many years, the nomenclature for solving simultaneous linear equations is aligned at the end of the second-to-last column.

Year	Author	Page	Uses Gauss's brackets (✓)	Nomenclature	Remainder of quotation
1690	Rolle	34		Des Substitutions	
		42		De la Methode	
1720	Newton	60		Transformation	of two or more Equations into one
and prior years				in order to exterminate	the unknown Quantities
1742	Hammond	142		is called the *exterminating*	[of] *an unknown Quantity*
1755	Simpson	63		Of the Extermination	of unknown quantities, or
				the reduction	of two, or more equations, to a single one
1771	Euler	II. 27		Der natürlichste Weg	bestehet nun darin
1804	Lacroix	114		se nomme *élimination*	
1805	Legendre	73		par les méthodes	ordinaires
1809	Gauss	214		per eliminationem	vulgarem
1818	Lacroix	84		is called *elimination*	
1822	Euler	216		the most natural method	of proceeding
1828	Bourdon	79		La méthode d'elimination	
1835	Davies	93		quantities may be eliminated	by the following rule
1846	Clark	102		Of elimination	when there are three or more equations
		103		Reduce	the equations
1868	Chauvenet	530	✓	Elimination	of the unknown quantities by
				the method of substitution,	according to Gauss
1879	Liagre	557	✓	Pour procéder a l'*élimination*	... coefficients auxiliaires de Gauss
1884	Merriman	51		the method of substitution	due to Gauss by which is preserved the symmetry
1888	Doolittle (not	43	✓	In the elimination	it will be convenient to employ
	the Doolittle)			the method of substitution	using ... notation proposed by Gauss.
1895	Jordan	80	✓	**Gauss schen Elimination**	
1900	Bartlett	97	✓	"Method of Substitution"	proposed by Gauss
1905	Johnson	120	✓	method of substitution	as developed by Gauss, which has the advantage
					of preserving, ... in the elimination, the symmetry
1906	Wright and Hayford	106	✓	method of substitution	introduced by Gauss and the Doolittle method
1907	Helmert	120	✓	Algorithmus	von C. F. Gauss
1912	Palmer	92	✓	Gauss's method	of solution with logarithms
1924	Benoit	24		les méthodes ordinaires	y compris celle de Gauss
1927	Tolley and Ezekiel			Gauss himself developed a	
		107		direct process of elimination	which was much shorter than
					solution by determinants
1941	Crout	1235		Gauss's method	
1941	Dwyer	450	(mentions)	notation	... suggested by Gauss
1943	Hotelling	3		Doolittle method	
1944	Jensen	13	✓	Gauss'ian algorithm	
1947	Bodewig	447		Gaussschen Methode	
1947	von Neumann and	1049		the conventional elimination	method
	Goldstine				
1948	Fox, Huskey, and	149		Gaussian algorithm	
	Wilkinson				
1948	Turing	287		"Gauss elimination	process"
1953	Forsythe	301		systematic elimination	of unknowns in the fashion of high-school
					algebra, as described by Gauss
		315		**Gaussian elimination**	(pivotal condensation)
1953	Householder	68		method of elimination	by addition and subtraction
1960	Wilkinson	45		**Gaussian elimination**	
1964	Householder	253		**Gaussian elimination**	
1977	Goldstine	287		he [Gauss] discovered the	
				method of **Gaussian elimination**	for solving systems of linear equations

Note. Some of these excerpts contain opinions about the history of the subject that are not necessarily accurate.

stood that "Gaussian elimination" is an honorific and not an attribution for a simple concept to which surprisingly many people have contributed.

Acknowledgments

References

Ahlmann, H.W., Rosenbaum, L., 1933. Scientific results of the Swedish-Norwegian Arctic expedition in the Summer of 1931, Part I–III. Geografiska Annaler 15, 1–68.

Aitken, A.C., 1937. Studies in practical mathematics I. The evaluation, with application, of a certain triple product matrix. Proc. Roy Soc. Edinburgh 57, 172–181.

Aitken, A.C., 1951. Determinants and Matrices, seventh ed. Oliver and Boyd, Edinburgh.

Althoen, S.C., McLaughlin, R., 1987. Gauss–Jordan reduction: a brief history. American Mathematical Monthly 94, 130–142.

Apokin, I.A., 2001. Willgodt Theophil Odhner and his "Arifmometr:" the beginning of the mass-production of calculating machines. In: Trogemann, G., Nitussov, A.Y., Ernst, W. (Eds.), Computing in Russia. Vieweg, Braunschweig/Wiesbaden, pp. 39–46.

Baker, H.F., 1903. On the integration of linear differential equations. Proc. London Math. Soc. 35, 333–378.

Banachiewicz, T., 1937a. Sur la résolution numérique d'un système d'équations linéaires. Bull. intern. de l'Acad. Polonaise, Série A. Sc. Math., 350.

Banachiewicz, T., 1937b. Zur Berechnung der Determinanten wie auch der Inversen, und zu darauf basierten Auflösung linearer Gleichungen. Acta Astronomica 3, 41–67.

Banachiewicz, T., 1938a. Méthode de résolution numérique des équations linéaires, du calcul des déterminants et des inverses, et de réduction des formes quadratiques. Bull. intern. de l'Acad. Polonaise, Série A. Sc. Math., 393–404.

Banachiewicz, T., 1938b. Principes d'une nouvelle technique de la méthode des moindres carrés. Bull. intern. de l'Acad. Polonaise, Série A. Sc. Math., 134–135.

Banachiewicz, T., 1942. An outline of the Cracovian algorithm of the method of least squares. Astr. Jour. 50, 38–41.

Bargmann, V., Montgomery, D., von Neumann, J., 1946. Solution of linear systems of high order. In: Taub [1963], vol. 5. pp. 421–477.

Bartlett, D.P., 1900. General Principles of the Method of Least Squares with Applications, second ed. Private printing, Boston.

Bashmakova, I.G., Smirnova, G.S., 2000. The Beginnings and Evolution of Algebra (Tr. A. Shenitzer). Vol. 23 of Dolciani Mathematical Expositions. Mathematical Association of America, Washington.

Baxendall, D., 1929. Calculating machines. In: Encyclopedia Britannica, 14th ed., vol. 4. pp. 548–553.

Benoit, 1924. Note sur une méthode de résolution des équations normales provenant de l'application de la méthode des moindres carrés a un système d'équations linéaires en nombre inférieur a celui des inconnues. — Application de la méthode a la résolution d'un système defini d'équations linéaires (Procédé du Commandant Cholesky). Bulletin géodésique (2), 67–77. The author is identified only as Commandant Benoit.

Bernkopf, M., 1968. A history of infinite matrices. Arch. Hist. Exact Sci. 4 (4), 308–358.

Bessel, F.W., Baeyer, J.J., 1838. Gradmessung in Ostpreussen und ihre Verbindung mit Preussischen und Russischen Dreiecksketten. Königlichen Akademie der Wissenschaften, Berlin.

Betts, G.L., Bauer, D., Tiedjens, V.A., Landsberg, H.E., Holter, N.J., 1951. Scholars and the root of all evil. Science 113, 330–333.

Bézout, E., 1788. Cours de mathématiques, à l'usage du corps royal de l'artillerie. Vol. 2 of 4. Ph.D. Pierres, Paris.

Bézout, E., 2006. General theory of algebraic equations (Tr. E. Feron). Princeton University Press, Princeton. First French edition 1779.

Bishop, R.E.D., 1987. Arthur Roderick Collar. 22 February 1908–12 February 1986. Biogr. Mems Fell. R. Soc. 33, 164–185.

Black, A.N., 1949. Further notes on the solution of algebraic linear simultaneous equations. The Quarterly Journal of Mechanics and Applied Mathematics 2 (3), 321–324.

Bodewig, E., 1944. Mathematical Reviews 7, 488. Review of Jensen [1944].

Bodewig, E., 1947–1948. Bericht über die vershiedenen Methoden zur Lösung eines Systems linear Gleichungen mit reellen Koffizienten. I through V. Indagationes Mathematicae. Part I, 9(4):441–452 (1947). Part II, 9(5): 518–530 (1947). Part III, 9(5):611–621 (1947). Part IV, 10(1):24–35 (1948). Part V, 10(1):82–90 (1948). Also published in Koninklijke Nederlandsche Akademie van Wetenschappen, Proceedings. Part I, 50(7–8):930–941 (1947). Part II, 50(9–10):1104–1116 (1947). Part III, 50(9–10):1285–1295 (1947). Part IV, 51(1–2):53–64 (1948). Part V, 51(1–2):211–219 (1948).

Bodewig, E., 1948. MR0024235. Mathematical Reviews 9, 471b. Review of von Neumann and Goldstine [1947].

Bodewig, E., 1949. MR0028100. Mathematical Reviews 10, 405c. Review of Turing [1948].

Bodewig, E., 1950. The payment of the learned man. Science 112, 538–539.

Bourdon, M., 1828. Élémens D'Algèbre, fifth ed. Bachelier, Paris.

Brandt, A.E., 1935. Uses of the progressive digit method. Practical Application of the Punch Card Method in Colleges and Universities. Columbia University Press, New York, pp. 423–436.

Brezinski, C., 2006. The life and work of André Cholesky. Numerical Algorithms 43 (3), 279–288.

Brezinski, C., Gross-Cholesky, M., 2005. La vie et les travaux d'André Louis Cholesky. Bulletin de la Société des amis de la bibliothèque de l'Ecole polytechnique (SABIX) 39, 7–32. Available from: <http://www.sabix.org/bulletin/sabixb39.htm>.

Brezinski, C., Wuytack, L. (Eds.), 2001. Numerical Analysis: Historical Developments in the 20th Century. North-Holland, Amsterdam.

Britton, J.L. (Ed.), 1992. Collected Works of A.M. Turing: Pure Mathematics. North-Holland, Amsterdam.

Bühler, W.K., 1981. Gauss: A Biographical Study. Springer-Verlag, New York.

Buteo, J., 1560. Logistica. Paris. The author is also known as Jean Borrel.

Cajori, F., 1890. The Teaching and History of Mathematics in the United States. Government Printing Office, Washington. Bureau of Education Circular of Information No. 3.

Cajori, F., 1894. A History of Mathematics. MacMillan and Co., New York.

Cajori, F., 1929. The Chequered Career of Ferdinand Rudolph Hassler, First Superintendent of the United States Coast Survey; A Chapter in the History of Science in America. The Christopher Publishing House, Boston.

Cardano, G., 1545. Artis Magnae Sive de Regulis Algebraicis. Translated by T.R. Witmer as The Great Art or The Rules of Algebra, MIT Press, Cambridge, 1968.

Cassinis, G., 1946. Risoluzione dei sìstemi di equazioni algebriche lineari. Rendiconti del Seminario Matematico e Fisico di Milano 17, 62–78.

Cauchy, A.L., 1905. Œuvres complètes d'Augustin Cauchy, vol. 1. Gauthier-Villars, Paris.

Chase, G.C., 1980. History of mechanical computing machinery. Annals of the History of Computing 2 (3), 198–226 (reprinted from the Proceedings of the Association of Computing Machinery, 1952, with corrections and foreward by I. Bernard Cohen).

Chauvenet, W., 1868. Treatise on the Method of Least Squares. J.B. Lippincott & Co., Philadelphia.

Cholesky, A.-L., 2005. La résolution numérique des systèmes d'équations linéaires. Bulletin de la Société des amis de la bibliothèque de l'Ecole polytechnique (SABIX) 39. Available from: <http://www.sabix.org/bulletin/sabixb39.htm>. From an original written 1910.

Clark, D.W., 1846. Elements of Algebra. Harper & Brothers, New York.

Clark, W., 2006. Academic Charisma and the Origins of the Research University. University of Chicago Press, Chicago.

Clark, W.E., 1930. The Āryabhaṭīya of Āryabhaṭa: An Ancient Indian Work on Mathematics and Astronomy. University of Chicago Press, Chicago (reprinted by Kessinger Publishing, 2006).

Clarke, A.R., 1880. Geodesy. Clarendon Press, Oxford.

Clarke, B.R., 2008. Linear Models: The Theory and Application of Analysis of Variance. Wiley Series in Probability and Statistics. Wiley, Hoboken.

Coast and Geodetic Survey, 1881. Report of the Superintendent of the Coast and Geodetic Survey Showing the Progress of the Work During the Fiscal Year Ending with June, 1878. Government Printing Office, Washington.

Cohen, D., Lee, T., Sklar, D., 2006. Precalculus: With Unit-circle Trigonometry, fourth ed. Thomson Brooks/Cole, Belmont.

Craig, C.C., 1972. Remarks concerning Paul S. Dwyer. In: Tracy, D.S. (Ed.), Symmetric Functions in Statistics: Proceedings of a Symposium in Honor of Professor Paul S. Dwyer. University of Windsor, Windsor, pp. 3–10.

Crelle, A.L., 1864. Rechentafeln welche alles Multipliciren und Dividiren mit Zahlen unter Tausend ganz ersparen: Bei grösseren Zahlen aber die Rechnung erleichtern und sicherer machen, second ed. G. Reimer, Berlin.

Crout, P.D., 1941a. A short method for evaluating determinants and solving systems of linear equations with real or complex coefficients. Transactions of the American Institute of Electrical Engineers 60, 1235–1241.

Crout, P.D., 1941b. A short method for evaluating determinants and solving systems of linear equations with real or complex coefficients. Marchant Methods MM-182, Marchant Calculating Machine Company, Oakland.

Davies, C., 1835. Elements of Algebra: Translated from the French of M. Bourdon. Wiley & Long, New York.

Dedron, J., Itard, J., 1959. Mathématiques et Mathématiciens. Magnard, Paris.

Deming, H.G., 1928. A systematic method for the solution of simultaneous linear equations. Amer. Math. Monthly 35, 360–363.

Dieudonné, J., 1981. Von Neumann, Johann. In: Gillispie, C.C. (Ed.), . In: Dictionary of Scientific Biography, vol. 14. Charles Scribner's Sons, New York, pp. 89–92.

Domingues, J.C., 2008. Lacroix and the Calculus. In: Vol. 35 of Science Networks Historical Studies. Birkhäuser, Basel.

Doolittle, C.L., 1888. A Treatise on Practical Astronomy as Applied to Geodesy and Navigation, second ed. John Wiley & Sons, New York.

Doolittle, M.H., 1878. Method employed in the solution of normal equations and in the adjustment of a triangularization. In: Coastand Geodetic Survey [1881]. pp. 115–120.

Duff, I.S., Erisman, A.M., Reid, J.K., 1986. Direct Methods for Sparse Matrices. Oxford University Press, Oxford.

Duncan, D.B., Kenney, J.F., 1946. On the Solution of Normal Equations and Related Topics. Edwards Brothers, Ann Arbor, Pamphlet.

Dunnington, G.W., 2004. Carl Friedrich Gauss: Titan of Science. The Mathematical Association of America, Washington. (First ed., 1955).

Dwyer, P.S., 1941a. The solution of simultaneous equations. Psychometrika 6 (2), 101–129.

Dwyer, P.S., 1941b. The Doolittle technique. Ann. Math. Statist. 12 (4), 449–458.

Dwyer, P.S., 1944. A matrix presentation of least squares and correlation theory with matrix justification of improved methods of solution. Ann. Math. Statist. 15 (1), 82–89.

Dwyer, P.S., 1945. The square root method and its use in correlation and regression. J. Amer. Statist. Assoc. 40, 493–503.

Dwyer, P.S., 1951. Linear Computations. John Wiley & Sons, New York.

Eckert, W.J., 1940. Punched Card Methods in Scientific Computation. Thomas J. Watson Astronomical Computing Bureau, New York.

Euler, L., 1771. Anleitung zur Algebra. Lund.

Euler, L., 1822. Elements of Algebra (Tr. F. Horner), third ed. Longman and Co., London.

Farebrother, R.W., 1987. A memoir of the life of M.H. Doolittle. Bulletin of the Institute of Mathematics and its Application 23, 102.

Farebrother, R.W., 1988. Linear Least Squares Computations. Marcel Dekker, New York.

Farebrother, R.W., 1999. Fitting Linear Relationships: A History of the Calculus of Observations 1750–1900. Springer Series in Statistics. Springer-Verlag, New York.

Felippa, C.A., 2001. A historical outline of matrix structural analysis: a play in three acts. Computers & Structures 79 (14), 1313–1324.

Forsythe, G.E., 1953. Solving linear algebraic equations can be interesting. Bull. Amer. Math. Soc. 59, 299–329.

Fox, K.A., 1989. Agricultural economists in the econometric revolution: institutional background, literature and leading figures. Oxford Economic Papers, New Series 41 (1), 53–70.

Fox, L., 1954. Practical solution of linear equations and inversion of matrices. Applied Mathematics Series 39, U.S. Bureau of Standards.

Fox, L., 1987. James Hardy Wilkinson. 27 September 1919–5 October 1986. Biogr. Mems Fell. R. Soc. 33, 670–708.

Fox, L., Huskey, H.D., Wilkinson, J.H., 1948. Notes on the solution of algebraic linear systems of equations. Quart. J. Mech. Appl. Math. 1 (2), 150–173.

Frazer, R.A., Duncan, W.J., Collar, A.R., 1938. Elementary Matrices and Some Applications to Dynamics and Differential Equations. Cambridge University Press, Cambridge.

Friberg, J., 2007a. Amazing Traces of a Babylonian Origin in Greek Mathematics. World Scientific, Hackensack.

Friberg, J., 2007b. A Remarkable Collection of Babylonian Mathematical Texts. Springer-Verlag, New York.

Gantmacher, F.R., 1959. The Theory of Matrices, vol. 1. Chelsea Publishing, New York.

Gauss, C.F., 1801. Disquisitiones Arithmeticae. Leipzig.

Gauss, C.F., 1809. Theoria Motus Corporum Coelestium in Sectionibus Conicis Solum Ambientium. Perthes and Besser, Hamburg.

Gauss, C.F., 1810. Disquisitio de elementis ellipticis Palladis. Commentationes Societatis Regiae Scientiarum Gottingensis recentiores: Commentationes classis mathematicae 1 (1808–1811), 1–26.

Gauss, C.F., 1822. Anwendung der Wahrscheinlichkeitsrechnung auf eine Aufgabe der practischen Geometrie. Astronomische Nachrichten 1 (6), 81–86.

Gauss, C.F., 1823. Theoria Combinatorionis Observationum Erroribus Minimis Obnoxiae. Heinrich Dieterich, Göttingen.

Gauss, C.F., 1826. Supplementum theoriae combinationis observationum erroribus minimis obnoxiae. Commentationes Societatis Regiae Scientiarum Gottingensis recentiores: Commentationes classis mathematicae 6 (1823–1827), 57–98.

Gauss, C.F., 1857. Theory of the Motion of the Heavenly Bodies Moving about the Sun in Conic Sections (Tr. C.H. Davis). Little, Brown, and Company, Boston.

George, A., Liu, J.W.H., 1981. Computer Solution of Large Sparse Positive Definite Systems. Prentice-Hall, Englewood Cliffs.

Ghilani, C.D., Wolf, P.R., 2006. Adjustment Computations: Spatial Data Analysis, fourth ed. John Wiley & Sons, Hoboken.

Gillings, R.J., 1972. Mathematics in the Time of the Pharaohs. MIT Press, Cambridge.

Gillispie, C.C., 1997. Pierre-Simon Laplace, 1749–1827: A Life in Exact Science. Princeton University Press, Princeton.

Goldstine, H.H., 1972. The Computer from Pascal to von Neumann. Princeton University Press, Princeton.

Goldstine, H.H., 1977. A History of Numerical Analysis from the 16th through the 19th Century. In: Vol. 2 of Studies in the History of Mathematics and Physical Sciences. Springer-Verlag, New York.

Golub, G.H., Van Loan, C.F., 1996. Matrix Computations, third ed. The Johns Hopkins University Press, Baltimore.

Gore, J.H., 1889. A Bibliography of Geodesy. Government Printing Office, Washington, United States Coast and Geodetic Survey. Appendix No. 16 — Report for 1887.

Gosselin, G., 1577. De Arte Magna. Paris.

Grattan-Guinness, I. (Ed.), 1994. Companion Encyclopedia of the History and Philosophy of the Mathematical Sciences. Routledge, London.

Grattan-Guinness, I., 2004. The mathematics of the past: distinguishing its history from our heritage. Historia Math. 31, 163–185.

Grier, D.A., 2005. When Computers were Human. Princeton University Press, Princeton.

Hagen, G., 1867. Grundzüge der Wahrscheinlichkeits-Rechnung. Ernst & Korn, Berlin.

Hald, A., 2007. A History of Parametric Statistical Inference from Bernoulli to Fisher, 1713–1935. Sources and Studies in the History of Mathematics and Physical Sciences. Springer, New York.

Hammond, N., 1742. The elements of algebra. London.

Hawkins, T., 1975. Cauchy and the spectral theory of matrices. Historia Math. 2, 1–29.

Hawkins, T., 1977a. Another look at Cayley and the theory of matrices. Archives Internationales d'Histoire des Sciences 27 (100), 82–112.

Hawkins, T., 1977b. Weierstrass and the theory of matrices. Arch. Hist. Exact Sci. 17 (2), 119–164.

Heath, T.L., 1910. Diophantus of Alexandria, A Study in the History of Greek Algebra, second ed. Cambridge University Press, Cambridge.

Heefer, A., 2005. The rhetoric of problems in algebra texbooks from Pacioli to Euler. Thesis, Ghent University Centre for Logic and Philosophy of Science, Ghent. Available from: <http://logica.ugent.be/albrecht/thesis/AlgebraRhetoric.pdf>.

Helmert, F.R., 1907. Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate, second ed. B.G. Teubner, Leipzig.

Higham, N.J., 2002. Accuracy and Stability of Numerical Algorithms, second ed. SIAM, Philadelphia.

Hogendijk, J.P., 1994. Pure mathematics in Islamic civilization. In: Grattan-Guinness [1994]. pp. 70–79.

Horsburgh, E.M. (Ed.), 1914. Handbook of the Napier tercentenary celebration, or, Modern instruments and methods of calculation. G. Bell, London (reprinted with a new introduction by M.R. Williams, Charles Babbage Institute reprint series for the history of computing, vol, 3, Tomash Publishers, Los Angeles, 1982).

Hotelling, H., 1943. Some new methods in matrix calculation. Ann. Math. Statist. 14 (1), 1–34.

Householder, A.S., 1953. Principles of Numerical Analysis. McGraw-Hill, New York.

Householder, A.S., 1956. Bibliography on numerical analysis. J. ACM 3 (2), 85–100.

Householder, A.S., 1964. The Theory of Matrices in Numerical Analysis. Blaisdell, New York (reprinted by Dover, New York, 1975).

Høyrup, J., 2002. Lengths, Widths, Surfaces: A Portrait of Old Babylonian Algebra and its Kin. Springer-Verlag, New York.

Institute of Electrical and Electronics Engineers, 1985. IEEE Standard for Binary Floating-Point Arithmetic. ANSI/IEEE Std 754-1985, New York.

Jacobi, C., 1857. Über eine elementare Transformation eines in Bezug auf jedes von zwei Variablen-Systemen linearen und homogenen Ausdrucks. J. Reine Angew. Math. 53, 265–270.

Jensen, H., 1939. Herleitung einiger Ergebnisse der Ausgleichsrechnung mit Hilfe von Matrizen. Meddelelse 13, Geodætisk Institut, Copenhagen.

Jensen, H., 1944. An attempt at a systematic classification of some methods for the solution of normal equations. Meddelelse 18, Geodætisk Institut, Copenhagen.

Johnson, S., 1997. Making the arithmometer count. Bulletin of the Scientific Instrument Society 52, 12–21.

Johnson, W.W., 1905. The Theory of Errors and Method of Least Squares, first ed. John Wiley & Sons, New York.

Jordan, W., 1895. Handbuch der Vermessungskunde, fourth ed., vol. 1. J.B. Metzler Verlag, Stuttgart.

Kangshen, S., Crossley, J.N., Lun, A.W.-C., 1999. The Nine Chapters of the Mathematical Art Companion and Commentary. Oxford University Press, New York.

Katz, V.J., 1988. Who is the Jordan of Gauss–Jordan. Mathematics Magazine 61 (2), 99–100.

Katz, V.J., 1997. Algebra and its teaching: an historical survey. The Journal of Mathematical Behavior 16 (1), 25–38.

Katz, V.J., 1998. A History of Mathematics: An Introduction, second ed. Addison-Wesley, Reading.

Kinckhuysen, G., 1661. Algebra ofte Stel-konst Beschreven Tot dienst van de leerlinghen. Passchier Van Wesbusch, Haarlem.

Kloyda, M.T.a.K., 1938. Linear and Quadratic Equations 1550–1660. Edwards Brothers, Ann Arbor, University of Michigan lithoprint dissertation.

Kocinski, J., 2004. Cracovian Algebra. Nova Science Publishers, Hauppauge.

Krayenhoff, C.R.T., 1813. Verzameling van Hydrographische en Topographische Waarnemigen in Holland. Doorman en Comp., Amsterdam.

Krayenhoff, C.R.T., 1827. Précis historique des opérations géodésique et Astronomiques, faites en Hollande. Imprimerie de l'état., La Haye.

Lacroix, S.F., 1800. Elemens d'algèbre, à l'usage de l'Ecole centrale des Quatre-Nations, second ed. Impr. de Crapelet, chez Duprat, Paris.

Lacroix, S.F., 1804. Elemens d'algèbre, à l'usage de l'Ecole centrale des Quatre-Nations, fifth ed. Chez Courcier, Paris.

Lacroix, S.F., 1818. Elements of Algebra (Tr. J. Farrar). University Press, Cambridge, Massachusetts.

Laderman, J., 1948. The square root method for solving simultaneous linear equations. Mathematical Tables and other Aids to Computation 3 (21), 13–16.

Lagrange, J.L., 1759. Researches sur la méthode de maximis et minimis. Miscellanea Taurinensia 1. Journal title varies (reprinted in Serret, J.-A., 1867, Œuvres de Lagrange, vol. 1, Gauthier–Villars, Paris, 1–16).

Lay-Yong, L., Kangshen, S., 1989. Methods of solving linear equations in traditional China. Historia Mathematica 16, 107–122.

Legendre, A.M., 1805. Nouvelle méthode pour la détermination des orbites des comètes. Chez Didot, Paris.

Liagre, J.B.J., 1879. Calcul des probabilités et théorie des erreurs avec des applications aux sciences d'observation en général et a la géodésie en particulier, second ed. Muquardt, Bruxelles.

Libbrecht, U., 1973. Chinese Mathematics in the Thirteenth Century. MIT Press, Cambridge.

MacDuffee, C.C., 1949. Vectors and Matrices, third ed. Vol. 7 of Carus Mathematical Monographs. The Mathematical Association of America, Menasha (first ed., 1943).

Macomber, G.L., 1923. The influence of the English and French writers of the sixteenth, seventeenth, and eighteenth centuries on the teaching of algebra, M.A. thesis, University of California, Berkeley.

Marchant Calculating Machine Company, 1941. Notes on the Use of the Marchant Calculator for Solution of Simultaneous Equations by the Method of Prescott D. Crout as Described in Marchant Method MM-182. Marchant Methods MM-183, Oakland.

Marguin, J., 1994. Histoire des instruments et machines à calculer: trois siècles de mécanique pensante, 1642–1942. Hermann, Paris.

Martzloff, J.-C., 1997. A History of Chinese Mathematics (Tr. S.S. Wilson). Springer, Berlin.

Matthiessen, L., 1878. Grundzüge der antiken und modernen Algebra der litteralen Gleichungen. B.G. Teubner, Leipzig.

Merriman, M., 1884. A Text Book on the Method of Least Squares. John Wiley & Sons, New York.

Murray, F.J., 1948. The Theory of Mathematical Machines, second ed. King's Crown Press, New York.

Murray, F.J., 1961. Mathematical Machines, 2 vols. Columbia University Press, New York.

Nell, P., 1881. Schleiermacher's Methode der Winkelausgleichung in einen Dreiecksnetze. Zeitschrift für Vermussungswesen 10 (1), 1–11, 109–121.

Neugebauer, O., 1969. The Exact Sciences in Antiquity, second ed. Dover, New York.

Neugebauer, O., 1973. Mathematische Keilschrift-Texte, second ed., vol. 2 and 3 in one book. Springer-Verlag, Berlin.

Newton, I., 1707. Arithmetica Universalis. London.

Newton, I., 1720. Universal Arithmetick: or, a Treatise of Arithmetical Composition and Resolution. To which is added, Dr. Halley's Method of finding the Roots of Equations Arithmetically. Senex, Taylor, et al., London. Translated from the Latin by the late Mr. Raphson, and Revised and Corrected by Mr. Cunn. The 1728 edition is reproduced in Whiteside [1964–1967, v. 2].

Nievergelt, Y., 2001. A tutorial history of least squares with applications to astronomy and geodesy. In: Brezinski and Wuytack [2001]. pp. 77–112.

Ordnance Survey, 1858. Ordnance Trigonometrical Survey of Great Britain and Ireland — Account of the observations and calculations of the principal triangulation, 2 vols. Eyre and Spottiswoode, London.

Palmer, A.d.F., 1912. The Theory of Measurements. McGraw-Hill Book Company, New York.

Paucker, M.G., 1819. Ueber die Anwendung der Methode der kleinsten Quadratsumme auf physikalische Beobachtungen. Johann Friedrich Steffenhagen und Sohn, Mitau, pamphlet. Program zur Erfffnung des Lehrkursus auf dem Gymnasium illustre zu Mitau.

Peletier du Mans, J., 1554. L'Algebre. Lyon.

Petersen, W., Arbenz, P., 2004. Introduction to Parallel Computing: A Practical Guide with Examples in C. Oxford University Press, Oxford.

Plackett, R.L., 1972. Studies in the history of probability and statistics, XXIX: the discovery of the method of least squares. Biometrika 59 (2), 239–251.

Plofker, K., 2009. Mathematics in India. Princeton University Press, Princeton.

Pugsley, A.G., 1961. Robert Alexander Frazer. 1891–1959. Biogr. Mems Fell. R. Soc. 7, 75–84.

Rahn, J.H., 1659. Teutsche Algebra. Zurich.

Relf, E.F., 1961. William Jolly Duncan. 1894–1960. Biogr. Mems Fell. R. Soc. 7, 37–51.

Roberts, R., Kynaston, D. (Eds.), 1995. The Bank of England: Money, Power and Influence 1694–1994. Oxford University Press, Oxford.

Robson, E., 1999. Mesopotamian Mathematics, 2100–1600 BC: Technical Constants in Bureaucracy and Education. Vol. 14 of Oxford Editions of Cuneiform Texts. Oxford University Press, Oxford.

Robson, E., 2008. Mathematics in Ancient Iraq: A Social History. Princeton University Press, Princeton.

Rolle, M., 1690. Traité d'algèbre; ou principes generaux pour resoudre les questions de mathematique. E. Michallet, Paris.

Ross, E.C., 1831. Elements of Algebra Translated from the French of M. Bourdon for the Use of the Cadets of the U.S. Military Academy. E.B. Clayton, New York.

Rubin, T., 1926. Et nytt sätt att lösa normalekvationer (A new method of solving normal equations). Svensk Lantmäteritidskrift (1), 3–9.

Satterthwaite, F.E., 1944. Error control in matrix calculation. Ann. Math. Statist. 15 (4), 373–387.

Saunderson, N., 1761. Selected Parts of Professor Saunderson's Elements of Algebra: For the Use of Students at the Universities, second ed. London.

Schappacher, N., 2005. Diophantus of Alexandria: a text and its history. Available from: <http://www-irma.u-strasbg.fr/~schappa/NSch/Publications_files/Dioph.pdf>.

Schott, C.A., 1881. Appendix No. 8. In: Coastand Geodetic Survey [1881]. pp. 92–94.

Sesiano, J., 1982. Books IV to VII of Diophantus' Arithmetica in the Arabic translation attributed to Qustā ibn Lūqā. Springer-Verlag, Heidelberg.

Simpson, T., 1755. A Treatise of Algebra, second ed. John Nourse, London.

Stewart, G.W., 1998. Matrix Algorithms 1: Basic Decompositions. SIAM, Philadelphia, Pennsylvania.

Stigler, S.M., 1986. The History of Statistics: The Measurement of Uncertainty before 1900. Harvard University Press, Cambridge.

Stigler, S.M., 1999. Statistics on the Table: the History of Statistical Concepts and Methods. Harvard University Press, Cambridge.

Swartzlander, E., 1995. Calculators. IEEE Ann. Hist. Comput. 17 (3), 75–77.

Taub, A.H. (Ed.), 1963. John von Neumann Collected Works. Macmillan, New York.

Taussky, O., Todd, J., 2006. Cholesky, Toeplitz and the triangular factorization of symmetric matrices. Numerical Algorithms 41, 197–202.

The Monthly Review, 1801. Review of Elémens d'Algèbra 35, 470–476.

Toeplitz, O., 1907. Die Jacobische Transformation der quadratischen Formen von unendlichvielen Veränderlichen. Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse, 101–109.

Tolley, H.R., Ezekiel, M., 1927. The Doolittle method for solving multiple correlation equations versus the Kelley–Salisbury "iteration" method. J. Amer. Statist. Assoc. 22 (160), 497–500.

Torge, W., 2001. Geodesy, third ed. Walter de Gruyter.

Turing, A.M., 1948. Rounding-off errors in matrix processes. The Quarterly Journal of Mechanics and Applied Mathematics 1 (3), 287–308 (reprinted in Britton [1992]).

United States Congress, 1886. Testimony Before the Joint Commission . . . . Government Printing Office, Washington.

van Eyk, J.A., 1854. Reken-werktuigen. De Volksvlijt, pp. 387–396.

Verzuh, F., 1949. The solution of simultaneous linear equations with the aid of the 602 calculating punch. Mathematical Tables and other Aids to Computation 3 (27), 453–462.

von Neumann, J., 1954. The role of mathematics in the sciences and in society. In: Taub [1963, v. 6, pp. 477–490].

von Neumann, J., Goldstine, H.H., 1947. Numerical inverting of matrices of high order. Bulletin of the American Mathematical Society 53 (11), 1021–1099 (reprinted in Taub [1963, v. 5, pp. 479–557]).

Washington Star, June 28, 1913. Myrick Hascall Doolittle obituary, Part 1 page 7.

Waugh, F.V., Dwyer, P.S., 1945. Compact computation of the inverse of a matrix. Ann. Math. Statist. 16 (3), 259–271.

Werner, W., 1883. Ueber die Methode der "Coast and Geodetic Survey" zur Auflösung von Normalgleichungen. Der Civilingenieur 29, 116–126.

Whipple, F.J.W., 1914. Calculating machines. In: Horsburgh [1914]. pp. 69–123.

Whiteside, D.T. (Ed.), 1964–1967. The Mathematical Works of Isaac Newton, 2 vols. Johnson Reprint Corporation, New York and London.

Whiteside, D.T. (Ed.), 1968–1982. The Mathematical Papers of Isaac Newton, 8 vols. Cambridge University Press, Cambridge.

Wilkinson, J.H., 1960. Rounding errors in algebraic processes. In: Proceedings of the International Conference on Information Processing, UNESCO, Paris, 15–20 June 1959. Information Processing. R. Oldenbourg, Munich and Butterworths, London, pp. 44–53.

Williams, M.R., 1982. Introduction. In: Horsburgh [1914]. pp. ix–xxi.

Wilson, E.B., 1952. An Introduction to Scientific Research. McGraw-Hill, New York.

Witkowski, J., 1955. The Life and Work of Professor Dr. Thaddeus Banachiewicz. Acta Astronomica Series C 5, 85–94.

Wright, T.W., Hayford, J.F., 1906. The Adjustment of Observations by the Method of Least Squares with Applications to Geodetic Work. D. Van Nostrand, New York.