On the Histories of Linear Algebra: The Case of Linear Systems

Christine Larson, Indiana University

There is a long-standing tradition in mathematics education to look to history to inform instructional design (see e.g. Avital, 1995; Dorier, 2000; Swetz, 1995). An historical analysis of the genesis of a mathematical idea offers insight into (1) the contexts that give rise to a need for a mathematical construct, (2) the ways in which available tools might shape the development of that mathematical idea, and (3) the types of informal and intuitive ways that students might conceptualize that idea. Such insights can be particularly important for instruction and instructional design in inquiry-oriented approaches where students are expected to reinvent significant mathematical ideas. Sensitivity to the original contexts and notations that afforded the development of particular mathematical ideas is invaluable to the instructor or instructional designer who aims to facilitate students' reinvention of such ideas. This paper will explore the ways in which instruction and instructional design in linear algebra can be informed by looking to the history of the subject.

Three of the most surprising things I have learned from my excursion into the literature on the history of linear algebra had to do with Gaussian elimination. The first surprise was learning that the idea behind Gaussian elimination preceded Gauss by over 2000 years – there is evidence that the Chinese were using an equivalent procedure to solve systems of linear equations as early as 200 BC (Katz, 1995; van der Waerden, 1983). The second surprise was that Gauss developed the method we now call Gaussian elimination without the use of matrices. In fact the term 'matrix' wasn't even coined until nearly 40 years after the idea we now call Gaussian elimination was introduced by Gauss in 1811. The third surprise was that Gauss developed the method we now call Gaussian elimination to find the best approximation to a solution to a system of equations that technically had no solution because there were twice as many equations as unknowns.

In order to help contextualize the central discussion of the two developments of Gaussian elimination, this paper is organized in five parts. First, I offer some brief theoretical background that informs this work. Second, I give a broad overview of several important historical developments in linear algebra. Third, I discuss the work of Gauss that gave rise to the method of solving systems of linear equations using what is now commonly referred to as Gaussian elimination. Fourth, I contrast Gauss's work with the development of a remarkably similar procedure developed in ancient China ~ 200 BC. Finally, I discuss implications for instruction that emerge from my analysis, with a particular emphasis on (1) the contexts that give rise to a need for ideas relating to linear systems of equations, (2) the ways in which available tools and representations shaped the development of those ideas, and (3) the identification of central, underlying ideas and questions that drove the development of a coherent theory of systems of linear equations.

THEORETICAL BACKGROUND

This work reflects the underlying view espoused by Freudenthal that mathematics is an inherently human activity that takes place within and relative to social and cultural contexts (1991). In this capacity, activity is said to be mathematical in nature when it aims to develop increasingly sophisticated and general ways of organizing, quantifying, characterizing, predicting, and modeling the world by either creating new mathematical tools for dealing pragmatically with challenging issues that exist within a social/cultural context, or by using the tools and language of the existing mathematical community to reason and problem solve (Lesh & Doerr, 2003).

Sfard (1991) claims that mathematical development, on both the historical level and on the individual cognitive level, tends to occur as a series of process-object shifts. These shifts involve first conceiving of a mathematical operation that yields a result (operational or process view) and eventually coming to conceive of the result of a mathematical operation as an object in and of itself (structural or object view). This distinction between process and object views informs my historical analysis.

IMPORTANT DEVELOPMENTS IN THE HISTORY OF LINEAR ALGEBRA

Historians of mathematics differ on what they view to be the most important contributions to the history of linear algebra (see, e.g. Dorier, 2000; Fearnley-Sandler, 1979; Kleiner, 2007). However, there seems to be consensus that the history of linear algebra lies in two related points of view. From one point of view, the development of a coherent, comprehensive characterization of systems of equations and their solutions is seen as a driving, underlying force behind the subject. I will refer to this as the "systems view." From the other point of view, the development with a formal, axiomatic way of algebraically defining relations among and operations on vectors is central to what we now consider to be linear algebra. I will refer to this as the "vector spaces and transformations view." I consider both approaches to be central to linear algebra, but I find the distinction to be useful for contextualizing my analysis and discussion of the history of linear algebra. This paper will focus primarily on the "systems view."

Determinants: System Solving Origins

With the exception of the solution methods developed around 200 BC in China, little progress in the development of a comprehensive theory of systems of linear equations and their solutions was made until the 1600s and 1700s when determinants emerged (separately) in both

Japan and Europe (Dorier, 2000; Mikami, 1914). Before the development of determinants, the ancient Chinese methods for solving systems of equations used counting boards to represent problem constraints in rectangular arrays and to specify the sequence of manipulations to be performed in order to solve a given system.

In 1683, Japanese mathematician Seki Kowa developed a version of determinants as part of a method for solving a nonlinear system of equations. This method was described in a way that relied on the positions of coefficients arranged in a rectangular array, an indicator of the strong influence of Chinese mathematics on Japanese mathematics. Mikami (1914) recreates Seki's illustration from the original manuscript for 2x2 and 3x3 cases as shown below in Figure 1 with the following explanation:

> The dotted and real lines, or the red and black lines in the original manuscript, are used to indicate the signs which the product of the elements connected by these lines, will take in the development, the dotted lines corresponding to the positive sign and the real lines to the negative sign, if all the elements be positive. (p. 12)

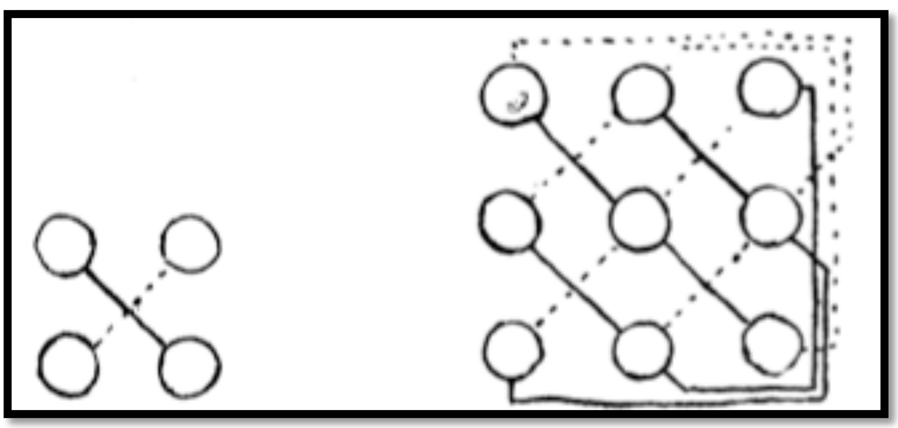


Figure 1: Mikami's recreation of Seki's diagrams

In 1750, Swiss mathematician Gabriel Cramer independently developed determinants as a way of specifying the solution to a system of linear equations as a closed form expression comprised purely of fixed by unspecified coefficients of the given system. Cramer's development of determinants leveraged the combinatorics of cleverly subscripted but unspecified coefficients. Katz (1995) shows how Cramer expresses a 3x3 linear system in 3 unknowns *x,y,z:*

$$A = Z_1 z + Y_1 y + X_1 x$$
$$A = Z_2 z + Y_2 y + X_2 x$$
$$A = Z_3 z + Y_3 y + X_3 x$$

Cramer then expresses the how the unknown values for *x,y,z* in terms of the coefficients. For instance, the value of *z* is given:

$$z = \frac{A_1 Y_2 X_3 - A_1 Y_3 X_2 - A_2 Y_1 X_3 + A_2 Y_3 X_1 + A_3 Y_1 X_2 - A_3 Y_2 X_1}{Z_1 Y_2 X_3 - Z_1 Y_3 X_2 - Z_2 Y_1 X_3 + Z_2 Y_3 X_1 + Z_3 Y_1 X_2 - Z_3 Y_2 X_1}$$

While Cramer does not explain how this result was obtained, he does give a general rule framed in terms of the combinatorics of the subscripts. Each unknown in an *nxn* system is solved by a fractional value, with numerators and denominators alike containing *n!* terms. Each of those terms:

> is composed of the coefficient letters, for example ZYX, always written in the same order, but the indexes are permuted in all possible ways. The sign is determined by the rule that if in any given permutation the number of times a larger number precedes a smaller number is even, then the sign is "+" otherwise it is "-". (Katz, 1995, p. 192)

Cramer also notes that a zero denominator indicates that the system does not have a unique solution, specifying that in the case then the numerators are zero, the system will have infinitely many solutions and that if any of the numerators are nonzero, the system will have no solution.

Cramer's approach would have been impossible without a shift in the use of algebraic notation introduced by French mathematician François Viète. Literal symbols (e.g. using a variable such as 'x' to represent an unknown fixed quantity, a fixed but unspecified quantity, or a varying quantity) were in use as early ~300 BC Greece by Euclid, but their use was not modernized until 1591 when French mathematician Viete introduced the convention of using vowels to represent unknown quantities and consonants to represent quantities which are known but unspecified. This advance can be thought of as the specification of the idea of a parameter. The particulars of his convention are no longer in use, but the distinction that arose was pivotal in shaping contemporary algebraic notation (Boyer, 1985).

For both Seki and Cramer, the way in which their notational system structured the coefficients shaped the way the determinant was specified. Seki's articulation of the determinant leveraged the physical arrangement of the coefficients in order to specify the operations to be performed on those coefficients. In this way, it seems that Seki's notational system afforded a process view of determinants because they were specified in terms of operations to be performed. Cramer's articulation of the determinant relied on the clever use of subscripting in the coefficients to create a closed form expression. I posit that Cramer's notational system affords an object view of determinants more strongly that does Seki's. Cramer's observation about what the value of the determinant reveals offers further evidence that he is conceptualizing the determinant as a mathematical object.

Instructional Implication: Reinventing Determinants

Historically, determinants were developed to help express solutions to systems of linear equations in terms of their coefficients. While it is not the case that the idea of determinants emerged from an explicit goal of computing a single value that would reveal whether a system had a unique solution, such a framing is potentially useful from a pedagogical point of view. Instructionally, one might present an opportunity for students to reinvent the notion of a determinant using the framing suggested below in Figure 2. A similar approach to reinventing determinants has been used in the context of inquiry-oriented differential equations (Rasmussen, 2002).

Determine under what conditions the following system of equations has a unique solution.

$$\begin{cases} ax + by = k_1 \\ cx + dy = k_2 \end{cases}$$

Figure 2: Creating a need for determinants (2x2 case)

One way to see if this system has a unique solution is to see if the lines are not parallel, which is easily done by putting both forms in slope-intercept form (the ever-popular "y=mx+b" from high school algebra). This yields the following (assuming both b and d are nonzero):

$$\begin{cases} y = -\frac{a}{b}x + \frac{k_1}{b} \\ y = -\frac{c}{d}x + \frac{k_2}{d} \end{cases}$$

In order for these lines to be non-parallel, we require that $\frac{a}{b} \neq \frac{c}{d}$. (Note that this requirement can be thought of as a comparison of the multiplicative relationship between the coefficients of x and y in the first equation with the multiplicative relationship between the coefficients of x and y in the second equation, i.e. comparing the relationship of the entries of one ROW to those of another ROW.) Equivalently, we can require that $ad \neq bc$, or equivalently, that $ad-bc \neq 0$. A benefit of the last two characterizations is that they allow us to relax the requirement on b and d.

Another way to see when this is by comparing the multiplicative relationship between the coefficients of the x's in the two equations with the multiplicative relationship between the coefficients of the y's in the two equations. In other words, we can require that $\frac{a}{c} \neq \frac{b}{d}$, which yields the same requirement as before – namely, that $ad \neq bc$, or equivalently, that $ad-bc \neq 0$

In this way, we can see the **determinant as a tool that measures whether a system of equations has a unique solution**. We can also observe that the uniqueness of the solution to the system of equations does NOT depend on the constant values (which we could choose to think of as the y-intercepts).

It is worth noting that in three or more dimensions, the mathematics becomes more complicated because one must deal with linear combinations and not just scalar multiples (much like when dealing with span and linear independence). Katz (1995) suggests Maclaurin's 1729 approach if students are to derive determinants for a 3x3 system: "given three equations in three unknowns *x, y, z,* he solved the first and second equations for *y* (treating *x* as a constant), then the first and third equations, then equated the two values and found *z*."

Euler's Inclusive Dependence

A second important development in the theory of systems of linear equations also took place in France in 1750: Leonhard Euler questioned whether a system of *n* linear equations with *n* unknowns has a unique solution, using the following system of equations as a counterexample: $3x - 2y = 5$ and $4y = 6x - 10$ (Dorier, 2000). This observation was made as part of a discussion of Cramer's paradox, which deals with the number of points of intersection of algebraic curves and the number of points needed to determine an algebraic curve. Euler gives additional examples with larger systems, and points out that it is possible for an equation to be "comprised of" or "contained in" others (Dorier, 2000, p. 7). (These are quotes of a translation of his description.) Dorier tags this notion of Euler's with the term "*inclusive dependence,"* pointing out that our modern notion of linear dependence is more carefully defined and more sophisticated. However, Euler's observation certainly raised an issue that contributed to the development of our current conception of linear dependence.

Euler's observation is significant in that it marks a qualitative shift in perspective from his predecessors' process view of solutions – his reasoning adopts the notion of a solution to a system of equations as a mathematical object with its own properties (in this case, uniqueness). This view differs from earlier perspectives where mathematical reasoning focused primarily on the development and implementation of processes for solving linear systems.

Other Important Developments

In 1811, Gauss developed a method of least squares for finding the best approximate solution to an indeterminate system of linear equations that had 12 equations and 6 unknowns. It was in this context that he outlined the method of Gaussian elimination, and he developed this method without the use of matrices. His discussion in this and earlier works reflects a complete understanding of the conditions under which a system has no solution, a unique solution, and infinitely many solutions. For instance, Gauss gives a detailed explanation of the relationship between elimination and the nature of the solution set to a system of linear equations in his 1809 *Theoria Motus* (*Theory on the Motion of Heavenly Bodies moving in Conic Sections*),

> 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. Let us see now, whether this elimination is always possible, or whether the solution can become indeterminate, or even impossible. It is known, from the theory of elimination, that the second or third case will occur when one of the equations… being omitted, an equation can be formed from the rest, either identical with the omitted one or inconsistent with it, or, which amounts to the same thing, when it is possible to assign a linear function $\alpha P+\beta Q+\gamma R+\delta S+$ etc., which is identically either equal to zero, or, at least, free from all the unknown quantities. (Gauss, 1809/1857, p. 269).

Thus we see that Gauss did have a full understanding that any system of linear equations can have no solution, a unique solution, or infinitely many solutions. Furthermore, he explains how one can identify the nature of the solution set based on the elimination process. He does not give a detailed explanation of Gaussian elimination in this 1809 work, but one appears in an 1811 piece. Gaussian elimination as it appears in the 1811 work will be discussed in greater detail later in this paper. Recall that Gauss does not use matrices for notating systems of equations or performing elimination – a fact that is surprising to many due to the common use of matrix notation in teaching the idea in today's linear algebra courses.

Unlike in Eastern traditions, matrices did not come into use in Western mathematics until the late 1800s. The term matrix was coined in 1850 by English mathematician James Joseph Sylvester, who was doing work with determinants. In 1857, Sylvester's friend and colleague Arthur Cayley published his Treatise on the Theory of Matrices. In this treatise, Cayley introduced matrices from a systems of equations point of view (as shown below in Figure 3), and then proceeds to develop his theory of matrices as mathematical objects that can be added, multiplied, inverted, and so on.

$\begin{aligned} X &= ax + by + cz, \\ Y &= a'x + b'y + c'z, \\ Z &= a''x + b''y + c''z, \end{aligned}$	$(X, Y, Z) = \left(\begin{array}{ccc} a, & b, & c \\ a', & b', & c' \\ a'', & b'', & c'' \end{array} \right)(x, y, z)$

Figure 3: Cayley introduces matrices from a systems point of view

In 1875, Frobenius offered a definition for linear dependence and independence that worked for both equations and n-tuples. According to Dorier (2000), this treatment of equations and n-tuples as (in some sense) equivalent objects in terms of linearity served as a significant step toward the contemporary treatment of vectors in linear algebra.

In 1888, Italian mathematician Giuseppe Peano formalized the first modern definition of vector spaces and linear transformations. Peano's formal definition of vector spaces did not gain much attention or popularity until 1918 when Weyl "essentially repeated Peano's axioms" in his book *Space-Time-Matter*, and articulated an important relationship between a “systems view” and a “vector spaces and transformations view” of linear algebra. "Weyl then brings the subject of linear algebra full circle, pointing out that by considering the coefficients of the unknowns in a system of linear equations in n unknowns as vectors, 'our axioms characterize the basis of our operations in the theory of linear equations'." (Katz, 1995, p. 204).

The remainder of the paper will focus on the development of Gaussian elimination by Gauss in Europe in the 1800s, and the development of a remarkably similar procedure in ancient China. Both of these accounts endeavor to describe the context(s) that created a need to solve systems of linear equations and the representations used to notate and manipulate those systems. The final section of the paper will discuss the ways in which these historical insights might serve to inform instruction and instructional design.

EUROPEAN DEVELOPMENT OF GAUSSIAN ELIMINATION

Gauss developed his method of Gaussian elimination in the context of astronomy. He was working to determine information about the elliptical orbit of an asteroid named Pallas, which was discovered in 1802 by Heinrich Olbers. At the time, Pallas was considered to be a planet. Gauss had a set of observational measurements, collected over a number of years, which could be used to determine the eccentricity and inclination of the orbit of Pallas. In order to do so Gauss used his data set, together with current theories of astronomy, to create a system of linear equations with 6 unknowns and 11 equations. (He actually began with 12 equations, but one of them seemed wildly inaccurate, so he discarded it.) The system carried conflicting constraints that arose due to measurement error. The first two equations from his system are shown below in Figure 4.

$$
\begin{array}{rcrrrrrrr}
0 & = & - & 183.93'' & + & 0.79363dL & + & 143.66dZ & + & 0.39493d\pi \\
 & & & & + & 0.95920d\phi & - & 0.18856d\Omega & + & 0.17387di; \\
0 & = & - & 6.81'' & - & 0.02658dL & + & 46.71dZ & + & 0.02658d\pi \\
 & & & & - & 0.20858d\phi & + & 0.15946d\Omega & + & 1.25782di;
\end{array}
$$

Figure 4: The first 2 out of 12 linear equations in 6 unknowns that Gauss wanted to solve

In order to find the "best" approximation to a solution to this system of equations, Gauss developed a method of least squares. Gaussian elimination was developed as part of this method. Gauss explains the importance of considering the closest "solution" to systems that do not have a solution:

> If the astronomical observations and other quantities, on which the computation of orbits is based, were absolutely correct, the elements also, whether deduced from three or four observations, would be strictly accurate (so far indeed as the motion is supposed to take place exactly according to the laws of Kepler), and, therefore, if other observations were used, they might be confirmed, but not corrected. But since all our measurements and observations are nothing more than approximations to the truth, the same must be true of all calculations resting upon them, and the highest aim of all computations made concerning concrete phenomena must be approximate, as nearly as practicable, to the truth. But this can be accomplished in no other way than by a suitable combination of more observations than the number absolutely requisite for the determination of the unknown quantities. (Gauss, 1809/1857, p. 249)

Gauss's 1809 *Theoria Motus* offers an overview of his least squares method, and a much more elaborate explanation of the Gaussian elimination portion of this method is given in his 1811 piece *Disquisitio de Elementis Ellipticis Palladis*. He explains the setup of his method of least squares:

> Let *V, V', V''*, etc. be functions of the unknown quantities *p, q, r, s,* etc., μ the number of those functions, and ν the number of the unknown quantities; and let us suppose that the values of the functions found by direct observation are $V=M$, $V'=M'$, $V''=M''$, etc. Generally speaking, the determination of the unknown quantities will constitute a problem, indeterminate, determinate, or more than determinate, according as $\mu<\nu$, $\mu=\nu$, or $\mu>\nu$. We shall confine ourselves here to the last case, in which, evidently, an exact representation of all the observations

would only be possible when they were all absolutely free from error… this cannot, in the nature of things, happen. (Gauss, 1809/1857, p. 253-254)

Gauss goes on to define the errors as *v=V-M, v'=V'-M', v''=V''-M'',* etc. and uses calculus and a probabilistic argument to claim that the sum squares of the errors *vv+v'v'+v''v''*+etc., "*will be the most probable system of values of the unknown quantities* p,q,r,s, *etc., in which the sum of the squares of the differences between the observed and computed values of the functions* V, V', V'', *etc. is a minimum,* if the same degree of accuracy is to be presumed in all the observations" (Gauss, 1809/1857, p. 260).

Gauss explains how to use this principle to determine values for the unknowns *p,q,r,s,* etc. when the functions *V, V', V'',* etc. are linear. He expresses the set of errors to be minimized as a system of linear equations in terms of the unknowns *p,q,r,s,* etc.:

$$
\begin{aligned}
&V - M = v = n + ap + bq + cr + ds + etc. \\
&V' - M' = v' = n' + a'p + b'q + c'r + d's + etc. \\
&V'' - M'' = v'' = n'' + a''p + b''q + c''r + d''s + etc. \\
&etc.
\end{aligned}
\qquad (1)
$$

The only way in which the above equation differs from Gauss's original inscription is that I have used positive *n'*s instead of the *-m'*s that Gauss used to represent the constant values in these expressions. The reason for this is that later in my explanation, it will be helpful to not have to deal with the negative sign, but rather to just acknowledge that *n* may assume a negative value. In order to minimize the sum of squares of the errors *vv+v'v'+v''v''*+etc, he creates a system of equations by taking the partial derivative of this sum with respect to each unknown and requiring that each of those partials be zero. We could write this as

$$
\begin{cases}
\frac{\partial}{\partial p}(vv + v'v' + v''v'' + etc) = 0 \\
\frac{\partial}{\partial q}(vv + v'v' + v''v'' + etc) = 0 \\
\frac{\partial}{\partial r}(vv + v'v' + v''v'' + etc) = 0 \\
etc
\end{cases}
\qquad (2)
$$

It is easy to see that this process yields Gauss's new system of linear equations to be solved:

$$\begin{aligned} &av + a'v' + a''v'' + etc = 0 \\ &bv + b'v' + b''v'' + etc = 0 \\ &cv + c'v' + c''v'' + etc = 0 \\ &etc \end{aligned} \qquad (3)$$

Gauss does not provide details of the process of elimination to be used to solve this new system in his 1809 work (although he does illuminate some of the details of the back-substitution in his discussion of measuring the precision of his method). I will continue with the explanation he offers in his 1811 publication, retaining Gauss's notation and changing only the choice of letters so as to be consistent with those used above in this explanation.

Gauss defines the following abbreviations:

$$\begin{aligned} &an + a'n' + a''n'' + \ldots = (an) \\ &a^2 + a'^2 + a''^2 + \ldots = (aa) \\ &ab + a'b' + a''b'' + \ldots = (ab) \\ &etc \end{aligned} \qquad (4)$$

We can express (3) in terms of unknowns *p, q, r, s,* etc by rearranging terms and substituting the abbreviations shown above in (4), yielding the system:

$$\begin{aligned} &(an) + (aa)p + (ab)q + (ac)r + \ldots = 0 \\ &(bn) + (ab)p + (bb)q + (bc)r + \ldots = 0 \\ &etc \end{aligned} \qquad (5)$$

If we name the sum of the squares Ω so that we have $\Omega = vv + v'v' + v''v''$+etc,, when we can write Ω in terms of the abbreviations from (4). If we then name the left hand side of the first equation from (5), say $A=(an)+(aa)p+(ab)q+(ac)r+\ldots$ then "All the terms of $\frac{A^2}{(aa)}$ which contain the factor *p*, are found in the expression Ω" (Gauss, 1811/1957, p. 3). Then by computing $\Omega - \frac{A^2}{(aa)}$, Gauss explains that *p* has been eliminated. He goes on to describe how one can continue eliminating one variable at a time until only one remains, at which point one determines the value of the single unknown quantity and performs back substitution to determine the values of the other unknowns.

The important, central aspect of this process is the sequential use of substitutions performed in such a way that that one variable is removed with each step of the process until only one variable remains. This value can then be determined from its single equations, and is the substituted into the previous equation with two unknowns to solve, and so and so forth until all unknown values have been found (Althoen & Mclaughlin, 1987).

SOLVING LINEAR SYSTEMS IN ANCIENT CHINA

The *Nine Chapters on the Mathematical Art* is an ancient Chinese text comprised of 246 problems and solution methods dating before 221 BC. This book is actually a reconstruction of an earlier text that was burned during the reign of Emperor Ch'in Shih Huang, a controversially tyrannical ruler credited with the unification of China as well as construction of the Great Wall of China (van der Waerden, 1983). The problems in this text arise from contexts such as field measurement (which gives rise to the development of geometry, fractions, and square and cube roots); trade, commerce, and taxation (which give rise to the development of ratios, proportions, and systems of equations); and distance-rate-time problems.

China's Mathematical Toolbox ~200 BC

In order to contextualize the mathematics that appears in the *Nine Chapters*, it is important to consider the mathematical tools and ideas that the Chinese had at their disposal at the time it was written. One of the most prominent mathematical tools in common use around 200 BC in China was the counting board, on which counting rods made of bamboo or ivory were arranged in rectangular arrays so that various calculations could be performed (Shen, Crossley, and Lun, 1999). Counting rod arithmetic was the central method of calculation in Chinese mathematics beginning around 500 BC and continuing until it was gradually replaced by the abacus between 1368 and 1644 AD (Shen, Crossley, and Lun, 1999). Common calculations included addition, subtraction, multiplication, and division. Standard algorithms for these calculations leveraged the structure of the base-10 system much like common algorithms of today, although the procedures looked rather different than today's standard column arithmetic.

Interestingly, the base-10 number system came into use in China around the same time as the counting boards. It is likely a combination of increasing needs for calculation (due to a cultural factors such as increase in trade and commerce) accompanied by the availability of a new calculation tool (the counting board) that helped drive forward the development and popularization of the base-10 number system, which replaced an earlier system whose structure was both additive and multiplicative in nature. (One might think of the earlier system as being somewhat analogous to the Roman Numeral system, but with different characters and some variation in conventions. For instance, there was no rule analogous to the rule for IV or IX where the "I" being written to the left means 'one less than.')

Counting rods could be arranged horizontally or vertically on the counting boards to represent digits. Mikami's (1974) illustration of how the rods were arranged to represent the digits is reproduced below in Figure 5.

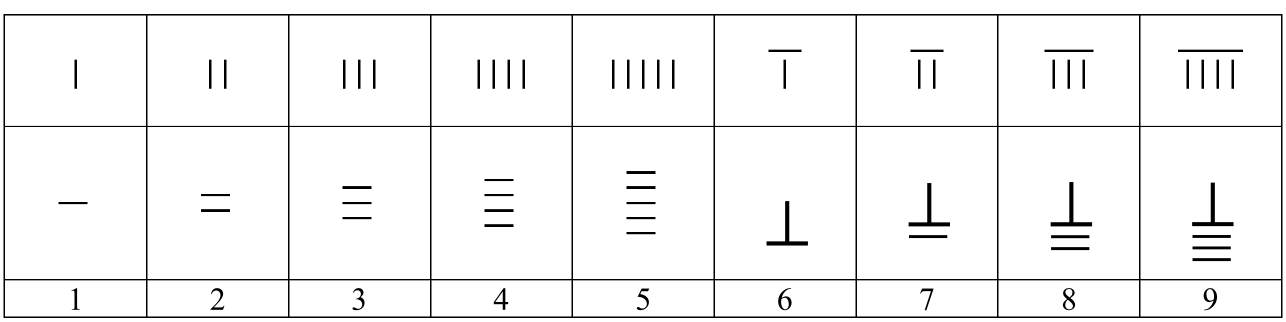


Figure 5: Two accepted ways of representing digits 1-9 with counting rods.

In order to avoid confusion regarding adjacent digits (e.g. when ‘|’ and ‘||’ are juxtaposed, they could easily look like ‘|||’ which could easily be mistaken as 3 rather than the 12 that was intended), the Chinese adopted the convention of alternating the orientation of adjacent digits. So, for example, the number 214 could be represented as shown below in Figure 6.

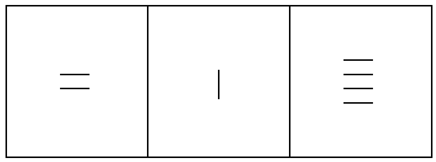


Figure 6: Representation of the number 214.

The Chinese did not have a digit for zero, but this was not a problem with the use of the counting boards. The slot on the counting board where we would place the digit was simply left blank. For example, the number 3,702 could be represented as shown below in Figure 7.

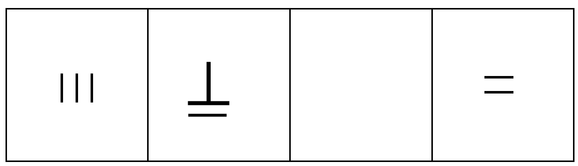


Figure 7: Representation of the number 3,702.

In addition to the use of counting boards and a base-ten number system, the Chinese also had positive and negative integers, as well as fractions. They did not use literal symbols to represent unknown or unspecified quantities, nor did they have a system of axiomatic deductive logic. Euclid (~300 BC, Greece) is generally considered to be the first to apply axiomatic deductive logic to mathematics in his *Elements*. While this work was greatly influential in Western mathematics for centuries to come, there was no Chinese translation of the book until 1606 AD (Shen, Crossley, and Lun, 1999).

Linear Systems in the Nine Chapters

Shen, Crossley, and Lun (1999) identify as many as seven different ways of solving linear equations that appear in ancient China. There is some amount of overlap among these methods, and several of them only work in systems with one or two equations. This paper will focus on the more general method for solving systems of linear equations discussed in the problems in Chapter 8, whose title can be translated to read "Rectangular Arrays."

Chapter 8 contains 18 problems, all of which are linear systems with between 2 and 6 unknown quantities. With one exception, all of the problems have a unique solution with the number of constraints (equations) equaling the number of unknown quantities. The only exception to this was a problem that had one free variable, which the author dealt with by adding a reasonable assumption to the context. This indicates that the Chinese had a strong intuition that the number of (independent) constraints must equal the number of unknowns in order to ensure a unique solution – however, because the original text only includes problem statements with descriptions of how to solve them, it is unknown what rationale or justification supported this intuition for mathematicians of the time.

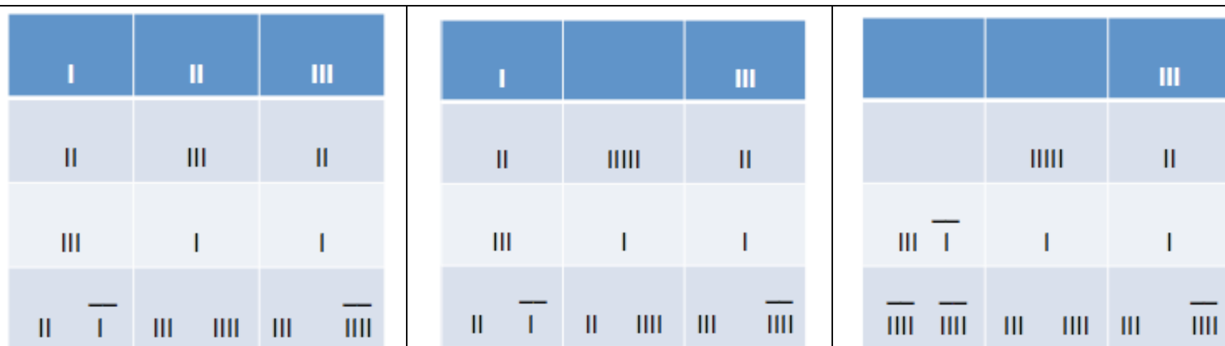
Solving Linear Systems with Rectangular Arrays: An example

A look at Shen, Crossley, and Lun's (1999) translation of the *Nine Chapters* offers some insight into the types of contexts that gave rise to systems of linear equations in ancient China. Thirteen of the eighteen problems in Chapter 8 draw on agricultural contexts, dealing with quantities of livestock or grain (by number, weight, volume, or cost). The others contexts range from practical (pulling forces of horses, amounts of water used by families sharing a communal well, and amounts of chicken eaten by people based on their social class) to more riddle-like (combinations of different types of coins, weights of sparrows and swallows). These evidence the existence of a common currency for trade, the domestication of horses for use as beasts of burden, and the class structure of Chinese society around 200 BC. The first problem in Chapter 8 reads:

> Now given 3 bundles of top grade paddy, 2 bundles of medium grade paddy, [and] 1 bundle of low grade paddy. Yield: 39 dou of grain. 2 bundles of top grade paddy, 3 bundles of medium grade paddy, [and] 1 bundle of low grade paddy, yield 34 dou. 1 bundle of top grade paddy, 2 bundles of medium grade paddy, [and] 3 bundles of low grade paddy, yield 26 dou. Tell: how much paddy does one bundle of each grade yield. (Shen, Crossley, & Lun, 1999, p. 399)

In order to fully understand the context of this problem, it is important that the reader understand that paddy is grain and that a dou is a unit used for measuring volume. If we were to rephrase the first sentence in a more contemporary way, it might read "A combination of 3 bundles of high-quality grain, 2 bundles of medium-quality grain, and 1 bundle of low-quality grain will yield 39 cups of flour."

The reader is instructed to begin laying down counting rods vertically in the far right column, as is shown below in Figure 8a. The steps to follow are remarkably similar to what you might find in an undergraduate textbook describing the steps of Gaussian elimination, albeit with the conventional use of rows and columns switched.



8a. Initial arrangement of rods	8b. "Multiply the middle column throughout by the superior grain on the right and continuously subtract it." (van der Waerden, 1983, p 47-48)	8c. Final arrangement of rods

Figure 8: Chinese method of solving linear systems using counting boards

Once the counting board is as shown in Figure 8c, the Array (Fangcheng) Rule states, "… the low grade paddy in the left column is the divisor, the entry below is the dividend. The quotient is the yield of low grade paddy" (Shen, Crossley, & Lun, 1999, p. 399). This gives that the yield (per bundle) of the low-grade paddy is 99/36 or 2 ¾ dou. The rule proceeds to describe how to use the yield per bundle of low-grade paddy with the middle column to obtain the yield per bundle of medium and high-grade paddy, respectively. Thus we see that the Chinese process was markedly similar to Gaussian elimination as it is now commonly taught, in that all entries above the diagonal of the coefficient matrix are eliminated, and then back substitution is performed to compute the values of the unknowns.

Discussion

It seems that the Chinese did not necessarily conceive of coefficients in the same way we do today. For instance, we would likely express the far right column as the equation $3x+2y+z=39$. From this point of view, the 3, 2, and 1 can be conceptualized in two ways. They are obviously the given numbers of bundles of each quality of grain, but they are also viewed as the weights on the unknowns x, y, and z that give an outcome yield of 39 for given values of yield per bundle of each type of grain.

One could argue that, from the Chinese viewpoint, each constraint is viewed as a unit that functions of as a sort of 'self contained truth' about the problem context. These 'truth units' can be replicated and/or traded with other 'truth units' so that the result is still a piece of information that is consistent with the problem context (i.e. the result is still a 'truth unit'). For instance, (3,2,1,39) is viewed as a 4-tuple corresponding to the truth that 3, 2, and 1 bundles of high, medium, and low grade grain yield 39 dou. The doubled unit (6,4,2,78) still represents the same truth about the problem situation in the sense that it is the same true proportion of values. Trading of non-proportional truth units is interesting in that it obviously does not yield a truth units that is proportional to either of the original truth units, yet it clearly still yields 4-tuples that are consistent with the problem situation. Overall, the process relies on the notion of "trading" in a clever way that leverages ideas from elementary number theory (replicating two columns until the top entry to be eliminated agrees, then replacing one of the columns with the difference of the two) as well as the fact that the quantities in the columns are integers. The counting board affords the continued systematic elimination of variables. It is clear that the Chinese understood that replacing a column with a linear combination of that column with another in this systematic way would eliminate extraneous information while preserving the information needed to arrive at a solution, but their rationale for this is unknown.

In this contextual framing of a linear system of equations, the central demand on the mathematical thinker is that he or she determines a solution to the system of equations. Mathematical activity is thus focused on the process of finding a solution. I argue that, in Sfard's framing, this is an example of a context that promotes a process view of a solution to a system of linear equations. Furthermore, it seems that the way in which systems were represented afforded the intuitive problem solving approach while also constraining the opportunity to view these linear systems or their solutions as mathematical objects in their own right.

DISCUSSION OF PEDAGOGICAL IMPLICATIONS

I discuss implications for instruction that emerge from my analysis, with a particular emphasis on (1) the contexts that give rise to a need for ideas relating to linear systems of equations, (2) the ways in which available tools and representations shaped the development of those ideas, and (3) the identification of central, underlying ideas and questions that drove the development of a coherent theory of systems of linear equations.

On Context

In China, the contexts that gave rise to a need for linear systems came largely from agriculture and trade. While solutions to these systems were often fractional values, the problems were posed exclusively with integer-valued constraints. Furthermore, no problems with more constraints than unknowns were posed, although the contexts used by the Chinese could have easily and sensibly been extended to include more constraints than unknowns. For instance, consider the context where the unknowns are the respective yields of each quality of grain. With information about three (independent) combinations of grain and their yields, one can uniquely determine the yield of each type of grain, but the yield would not necessarily provide a representative sample of the yield one could expect from each quality of grain over a broader harvesting area. Thus it is feasible that considering the yields of more combinations of grain could have been useful to the ancient Chinese, yet there is no evidence that they considered such systems.

Gauss, on the other hand, worked in contexts that required non-integer valued coefficients, and he explicitly discussed the numbers of possible solutions to linear systems and the conditions under which each would occur. He also discussed in detail the need for increased accuracy provided by using additional measurements, and the way in which this created a system with conflicting constraints for which a "best" solution was needed.

Gauss's work points to the importance of seriously considering meaningful contexts in which the number of constraints (equations) exceeds the number of unknowns – and the importance of not simply dismissing such systems as having no solution. In the case where a system of equations with no solution arises from a meaningful context, there is likely a need to find the best approximation to a solution.

In addition to my recommendation, per the work of Gauss, that indeterminate systems receive their due attention in realistic contexts, I suggest that instructors not immediately take it as immediate or obvious that the number of (independent) constrains needs to equal the number of variables in order to have a unique solution. It is not necessarily obvious why we need the number of (independent) equations to match the number of unknowns in order for a linear system to have a unique solution. In addition to the question "How do we solve linear systems of equations?" (which in its very phrasing, almost seems to presume a unique solution) it is important to address the following questions:

- How does the relationship between the number of unknowns and the number of constraints (equations) affect the solution to a system of linear equations?
- How can we detect and account for conflicting constraints and/or hidden redundancies when counting the constraints described in the previous question?
- How do we know, when we manipulate systems of equations or perform row/column operations, what information is changed, and what is left the same? (For instance, if we say two augmented matrices are row equivalent, what is equivalent about them, and how do we know that aspect of the system was unchanged by the row operations we performed?)

We owe it to our students to move beyond a procedural treatment of Gaussian elimination in introductory linear algebra courses. There is a rich history with a variety of contexts that have contributed to the development of the current theory of the nature of linear systems of equations and their solutions – and this set of ideas is a foundational part of a complete understanding of linear algebra.

On Tools and Representations

Mathematical tools impact the way in which ideas are notated, represented, and conceptualized. For instance, the use of counting boards in China facilitated a shift to the base-10 numbering system, and they affected the way in which systems of equations were represented and manipulated so as to find solutions. Gauss's use of literal symbols, differentiated to specify unknown quantities from unspecified but known quantities, lent itself to the use of repeated substitutions – a crucial element Gauss's 1811 description of his method for solving linear systems. Differential calculus also served as an important mathematical tool in Gauss's development of his method of least squares, where his need for Gaussian elimination arose.

Another example that supports this claim is seen in Seki's 'matrix-like' arrangement of terms to develop determinant as compared with Cramer's clever use of indexing to leverage combinatorial description of determinant.

Pedagogically, this suggests that we must be mindful not only of the way in which we pose tasks and frame questions to our students, but we must also be sensitive to the notations, representations, and tools used in the posing of those tasks. For instance, consider the way in which the Chinese used of counting boards to represent and solve linear systems of equations. The representation is tightly tied to the quantities given in the real world context, and the manipulations of the columns are easily and intuitively justified in a way that ties directly to the problem context. Such coherence between problem context and representational tools available for problem solving is pedagogically desirable. On the other hand, representations that are so tightly connected to specific contexts and needed manipulations may lend themselves more readily to process views as argued previously in this paper. Process views are developmentally important, but when there is a need to shift to an object view, a shift in notation may help facilitate this change in perspective.

Moving Beyond Context & Representations: Central Ideas & Driving Questions

The goal of identifying the driving questions that led to the development of central ideas in linear algebra is in a way unrealistic, as the historical development of ideas that contributed to linear algebra was not driven by a set of cleanly articulated questions. Rather, a variety of contexts (both real world and mathematical) led to the development of a set of mathematical tools that have been abstracted, generalized, and formalized into a coherent mathematical theory we call linear algebra. Restating a mathematical idea in a way that 'transcends' particular choices of context and notation fundamentally changes the idea to be learned. If we change the goals, tools, notation system, and framing of a question, it fundamentally changes the underlying question – and from that point of view, we can think of generalizations as ideas that arise from comparing the similarities and differences of contexts and our problem-solving methods within those contexts. If we say that two different contexts have the same underlying mathematical idea, then the heart of that idea lies in the identification of the structural similarities the contexts share. For instance, if we say that Gauss's astronomy and Chinese rice paddies have something in common, being able to articulate what these situations have in common that makes them mathematically equivalent (or similar) is an important part of developing an understanding what a linear system is.

In this way, I argue that it is important to have explicit classroom discussions aimed at identifying the structural similarities between the two situations that cause us to say the situations are 'the same.' The identification of these similarities is the generalization by which we create the more general mathematical idea unifying the two contexts. This suggests that it is pedagogically important to provide students with opportunities to work within specific contexts and to articulate the relationships between those contexts as a means of developing a more generalized understanding of the ideas that crosscut the contexts.

I am not, however, suggesting that specific-to-general is the only appropriate instructional sequencing. Indeed, the act of 'noticing' is historically a crucial part of mathematical development. For instance, Euler was dealing with a more general, non-linear context when he made his observation about the potential for inclusive dependence in systems of linear equations. In this case, he began with a more general context, and made an observation within a more simple and specific context – and that observation became quite important the development of the theory of that simpler, more specific context. Thus we see that 'noticing' is an important aspect of development. This suggests that students may in fact also make important progress when shifting from general to specific.

Final Remarks

The contexts and resources for our two developments of Gaussian elimination are quite different from one another. The Chinese texts were pragmatically driven and relied on prototypical examples and inductive logic. Gauss's was more theoretical in nature, and his work drew on the Greek tradition of deductive logic. These contrasting views resonate in linear algebra students today. Sierpinska (2000) identifies the tension between practically-minded students and theoretically-focused learning objectives as a source of difficulty for teachers and students alike. While this is in one sense an echo of a long historic tradition, looking to the framings of the past that lead to progress holds great promise for helping students develop practical and theoretical understandings of the key ideas in linear algebra.

References

Althoen, S.C. & McLaughlin, R. (1987). Gauss-Jordan Reduction: A Brief History. *The American Mathematical Monthly 94*(2), 130-142.

Avital, S. (1995). History of Mathematics Can Help Improve Instruction and Learning. In F. Swetz, J. Fauvel, O. Bekken, B. Johansson, & V. Katz (Eds.), *Learn from the Masters* (pp. 3-12). Washington, DC: Mathematical Association of America.

Boyer, Carl. *A History of Mathematics*. Princeton, NJ: Princeton University Press, 1985.

Dorier, J.-L. (2000). Epistemological Analysis of the Genesis of the Theory of Vector Spaces. In J.-L. (Ed.), *On the Teaching of Linear Algebra* (pp. 3-81). Dordrecht, The Netherlands: Kluwer.

Fearnley-Sander, D. (1979). Hermann Grassmann and the Creation of Linear Algebra. *The American Mathematical Monthly, 86(10),* 809-817.

Freudenthal, H. (1991). Revisiting mathematics education. Dordrecht: Kluwer Academic Publishers.

Gauss, K. (1857). *Theory of the motion of the heavenly bodies moving about the sun in conic sections* (C.H. Davis, Trans.). Boston: Little, Brown and Company. (Original work published 1809)

Gauss, K. (1957). *Work (1803--1826) on the Theory of Least Squares* (H.F. Trotter, Trans.). Technical Report No.5, Statistical Techniques Research Group, Princeton, NJ: Princeton University. (Original work published 1811)

Jones, P.S. (1995). The Role in the History of Mathematics of Algorithms and Analogies. In F. Swetz, J. Fauvel, O. Bekken, B. Johansson, & V. Katz (Eds.), *Learn from the Masters* (pp. 13-23). Washington, DC: Mathematical Association of America.

Katz, V.J. (1995) Historical Ideas in teaching Linear Algebra. In F. Swetz, J. Fauvel, O. Bekken, B. Johansson, & V. Katz (Eds.), *Learn from the Masters* (pp. 189-206). Washington, DC: Mathematical Association of America.

Kleiner, I. (2007). *A History of Abstract Algebra*. Boston: Birkhauser.

Lesh, R., & Doerr, H. (2003). Foundations of a Models and Modeling Perspective on Mathematics Teaching, Learning, and Problem Solving. In R. Lesh & H. Doerr (Eds.), *Beyond Constructivism: Models and Modeling Perspectives on Mathematics Problem Solving, Learning, and Teaching* (pp. 3-33). Mahwah, NJ: Lawrence Erlbaum Associates.

Mikami, Y. (1914). On the Japanese Theory of Determinants. *Isis 2*(1), 9-36.

Mikami, Y. (1974). *The Development of Mathematics in China and Japan* (2nd ed.). New York, NY: Chelsea Publishing Company.

Rasmussen, C. (2002). Instructional materials for a first course in differential equations. Unpublished document. Purdue Calumet University.

Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational Studies in Mathematics*, *22*(1), 1-36.

Shen, K., Crossley, J.N., Lun, A. W.-C. (1999). *The Nine Chapters on the Mathematical Art: Companion and Commentary*. New York: Oxford University Press.

Sierpinska, A. (2000). On some aspects of students' thinking in linear algebra. In J.-L. Dorier (Ed.), *On the Teaching of Linear Algebra* (pp. 209-246). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Swetz, F.J. (1995). Using Problems from the History of Mathematics in Classroom Instruction.

In F. Swetz, J. Fauvel, O. Bekken, B. Johansson, & V. Katz (Eds.), *Learn from the Masters* (pp. 25-38). Washington, DC: Mathematical Association of America.

van der Waerden, B.L. (1983). *Geometry and Algebra in Ancient Civilizations.* Berlin: Springer-Verlag.