Applied Mathematical Sciences Vol. 8, 2014, no. 133, 6689 - 6697
HIKARI Ltd, www.m-hikari.com
http://dx.doi.org/10.12988/ams.2014.49683

Old and New Proofs of Cramer's Rule

Maurizio Brunetti

Dipartimento di Matematica e Applicazioni "Renato Caccioppoli"
Università degli Studi di Napoli "Federico II" — P.le Tecchio 80, 80125 Napoli

Abstract

In spite of its high computational cost, Cramer's Rule for solving systems of linear equations is of historical and theoretical importance. In this paper we list six different proofs of it, the last of which has not apparently been published elsewhere. A discussion on their educational value and the tools involved is also included.

Mathematics Subject Classification: 15A06, 97H60, 15A15

Keywords: Cramer's Rule, Determinants, Linear equations

1 History, notations and tools

As described in any first course in linear algebra, the problem of solving a system of linear equations with n equations and n unknowns

$$\begin{array}{ccccc} a_{11}x_1 & + \; a_{12}x_2 & + \cdots + & a_{1n}x_n = b_1 \\ \vdots & & & \vdots \\ a_{n1}x_1 & + \; a_{n2}x_2 & + \cdots + & a_{nn}x_n = b_n \end{array} \tag{1}$$

is equivalent to solve a matrix equation of the form $AX = B$, where

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}, \quad X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad \text{and} \quad B = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \tag{2}$$

are the *matrix of coefficients*, the *column of unknowns*, and the *column of constants* respectively. The matrix obtained from A by adding on the right the column of the constants is often called the *augmented matrix* associated to (1).

Examples and exercises proposed in an undergraduate course usually involve real numbers, nevertheless the provided algorithms to find solutions work when coefficients and constants belong to any field $\mathbb{F}$. One of those techniques consists in applying, when possible, the following rule.

Theorem 1.1. (CRAMER'S RULE) *If the matrix of coefficients A is nonsingular, then the unique solution $(\bar{x}_1, \dots, \bar{x}_n)$ to the system (1) is given by*

$$\bar{x}_i = (\det A)^{-1} \det(A \overset{i}{\leftsquigarrow} B) \tag{3}$$

where $A \overset{i}{\leftsquigarrow} B$ is the matrix obtained from A by replacing the i-th column of A by the column of constants B.

Cramer's Rule holds even when coefficients and constants are taken in a commutative ring K. In such context, a matrix A is said to be nonsingular if its determinant is invertible in K.

The Swiss mathematician Gabriel Cramer (1704-1752) published the rule which would come to bear his name in Appendix I of its celebrated *Introduction à l'analyse des lignes courbes algébriques* (see [6], pp. 657-659). Theorem 1.1 often appears in lists of misnamed theorems, i.e. those well known results in mathematics which are not named for the originator. In fact, C. B. Boyer, B. A. Hedman and others claimed that Colin Maclaurin (1698-1746) knew Cramer's Rule as early as 1729, and incorporated it in his posthumous *Treatise of Algebra* published in 1748 (see [1] and [10]).

As a matter of fact, both Cramer and Maclaurin wrote down the solution of a system of 3 linear equations with 3 unknowns, as ratios of two quantities, each of which sum of 6 summands. After that, without giving any proof, they both explained how to build formulæ for more general cases. Note that neither of them could rely on the notion of determinant as a closed-form function, introduced only in 1771 by Alexandre-Théophile Vandermonde (1735-1796) [23]. Unfortunately, as noted in [13], the rule given by Maclaurin to choose the appropriate sign for each summand is wrong; on the contrary, Cramer's idea to count the number of transpositions (*dérangements*) in the permutation attached to a given term flawlessly reproduces the right one. It seems, therefore, that Cramer's Rule is genuinely due to Cramer.

We now list the classical tools employed in proofs of Section 2, starting with the formula named in honor of Gottried Wilhelm von Leibniz (1646-1716) for the determinant of $n \times n$ matrix A:

$$\det A = \sum_{\sigma \in \Sigma_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} \dots a_{\sigma(n)n}, \tag{4}$$

where Σ_n denotes the symmetric group on n letters, and $\operatorname{sgn}(\sigma) = (-1)^{N(\sigma)}$ is the *sign of permutation* σ, depending on the number $N(\sigma)$ of its inversions.

An often crucial role is played by the *cofactor* (or *Laplace*) *expansion* of the determinant, named after Pierre-Simon, marquis de Laplace (1749-1827), involving the *ij-cofactor* of the matrix A, i.e. the scalar A_{ij} defined by

$$A_{ij} \stackrel{\text{def}}{=} \frac{\partial(\det A)}{\partial a_{ij}} = (-1)^{i+j} M_{ij}, \tag{5}$$

where M_{ij} is the determinant of the $(n-1) \times (n-1)$ matrix that results from deleting the i-th row and the j-th column of A.

Theorem 1.2. *Let h and k be integers in the set $\{1, 2, \ldots, n\}$.*

$$\sum_{i=1}^{n} a_{ih} A_{ik} = \sum_{j=1}^{n} a_{hj} A_{kj} = \begin{cases} \det A & \text{if } h = k; \ (\text{First Laplace Theorem}) \\ 0 & \text{if } h \neq k; \ (\text{Second Laplace Theorem}). \end{cases} \tag{6}$$

By definition, the *matrix of cofactors* $C(A)$ has A_{ij} as its (i,j)-th element. From Theorem 1.2 it follows quite easily that the inverse A^{-1} of a nonsingular matrix A has the following form:

$$A^{-1} = (\det A)^{-1} \operatorname{adj} A, \tag{7}$$

where $\operatorname{adj} A$, the *adjoint of* A, is the transpose of the matrix of cofactors.

The following Theorem summarizes the so-called *three elementary properties* of the determinant function.

Theorem 1.3. *(I) If the matrix A^* is obtained from a square matrix A by swapping two rows or two columns, then* $\det A^* = -detA$.

(II) If the matrix A^* is obtained by A multiplying the i-th row, or the j-th column by the scalar c, then $\det A^* = c \det A$.

(III) If the matrix A^* is obtained by A by replacing the k-th row A_k by $A_k + cA_i$, or the k-th column A^k by $A^k + cA^i$, with $i \neq k$, then $\det A^* = \det A$.

Another nice property of the determinant function is multiplicativity.

Theorem 1.4. *The determinant of a matrix product of $n \times n$ matrices A and A^* equals the product of their determinants:*

$$\det(AA^*) = \det(A) \det(A^*) \tag{8}$$

In Italy, Theorem 1.4 is known as *Binet Theorem* after Jacques Philippe Marie Binet (1786-1856). Theorem 1.4 is in fact a special case of the Cauchy-Binet formula concerning the determinant of the product of two rectangular matrices of transpose shapes (see, for instance, Proposition 3.4 in [21]).

We end this section by recalling the renowned condition for the existence of a solution to a system of linear equations.

Theorem 1.5. *A system of equations* (1) *has a solution if and only if the rank of the coefficient matrix is equal to the rank of the augmented matrix.*

Theorem 1.5 is known as *Rouché-Capelli Theorem* in Italy, *Kronecker-Capelli Theorem* in Russia, *Rouché-Fontené Theorem* in France, and *Rouché-Frobenius Theorem* in Spain and many countries in Latin America. Eugène Rouché (1832-1910) wrote the condition for consistency in 1875 in terms of the existence of non-zero minors of same order in A and in the augmented matrix [20]. Alfredo Capelli (1855-1910) has been probably the first who reworded the condition using the notion of rank, his correspondent Italian word for it being *"caratteristica"* (see [3]).

2 The six proofs

The first proof of Cramer's Rule we propose goes back to 1841 and appeared in a paper by Carl Gustav Jacob Jacobi (1804-1851) [12]. This is not the oldest proof ever published. In 1825, for instance, Heinrich Ferdinand Scherk (1798-1885) published a 17 pages long proof by induction on the number of unknowns sketched in [16]. Because of its poor didactic value, not to say its hefty length, it has not been included here.

Proof 1 (Jacobi) Fix an integer h in $\{1, 2, \ldots, n\}$. The i-th equation of (1) multiplied by the cofactor A_{ih} becomes

$$A_{ih}a_{i1}x_1 + \ \ A_{ih}a_{i2}x_2 \ + \cdots + A_{ih}a_{in}x_n = A_{ih}b_i. \tag{9}$$

Adding together the n different equalities of type (9), we get

$$\sum_i (A_{ih}a_{i1})\, x_1 + \cdots + \sum_i (A_{ih}a_{ih})\, x_h + \cdots + \sum_i (A_{ijh}a_{in})\, x_n = \sum_i (A_{ih}b_i). \tag{10}$$

By Theorem 1.2, the only non-zero coefficient on the first side of (10) is the h-th one, and is equal to $\det A$. On the right side we have the cofactor expansion of $\det(A \overset{h}{\leftsquigarrow} B)$ along the h-th column. Hence (10) reads

$$(\det A)x_h = \det(A \overset{h}{\leftsquigarrow} B) \tag{11}$$

q.e.d. $\square$

Jacobi's proof also appears in the first linear algebra textbook published in Italy [2], followed soon after by *Teoria de' determinanti e loro applicazioni* written by Nicola Trudi (1811-1884), professor of infinitesimal calculus at University of Naples. The book [22] exhibits two different proofs of Cramer's Rule. The first is Jacobi's; we now explain the other one.

Proof 2 (Trudi) Thinking the matrix of coefficients A portioned by columns, we write $A = (A^1 \ldots A^n)$, and recall that another equivalent way to write the system (1) is $B = x_1A^1 + \cdots + x_nA^n$.

Now, by Theorem 1.3 (II), we get

$$x_h(\det A) = \det(A^1 \dots x_h A^h \dots A^n).$$

As a consequence of Thm 1.3 (III), the addition of $x_j A^j$ to the h-th column of

$$(A^1 \dots x_h A^h \dots A^n) = (A \overset{h}{\leftsquigarrow} x_h A^h)$$

does not affect its determinant whenever $j \neq h$. Hence

$$x_h(\det A) = \det(A \overset{h}{\leftsquigarrow} x_h A^h) = \det(A \overset{h}{\leftsquigarrow} \underbrace{x_h A^h + \sum_{j \neq h} x_j A^j}_{=B}).$$

It follows that $x_h = (\det A)^{-1} \det(A \overset{h}{\leftsquigarrow} B)$ q.e.d. $\square$

Trudi's proof has been rediscovered in [24] and included in some modern widespread textbooks (e.g. [14] and [5]). It also appears on the Italian Wikipedia page devoted to Cramer's Rule [17]. Nevertheless most textbooks on linear algebra (we just mention the classic [7], [18], and the recently published Italian textbook [15]) choose to prove Cramer's Rule via the adjoint matrix, which is our Proof 3.

Proof 3 Since A is non singular, the matrix equation $AX = B$ is equivalent to $X = A^{-1}B$. Recalling (7), on the i-th row of $A^{-1}B$ we find the scalar $(\det A)^{-1}$ multiplied by the cofactor expansion of $\det(A \overset{i}{\leftsquigarrow} B)$ along the i-th column, q.e.d. $\square$

Next proof is dead simple. The paper [19] is the oldest source found in print. Such proof has been adopted in [11]. As noted in [4], it gives practice in the important skill of exploiting the structure of sparse matrices, hence it should be worthy of more consideration.

Proof 4 Let I be the identity matrix. Using Laplace expansion, we immediately see that

$$x_i = \det(I \overset{i}{\leftsquigarrow} X). \tag{12}$$

Consider now the product

$$A(I \overset{i}{\leftsquigarrow} X) = (A^1 A^2 \dots AX \dots A^n) = A \overset{i}{\leftsquigarrow} B, \tag{13}$$

and apply Theorem 1.4 to (13):

$$(\det A)\det(I \overset{i}{\leftsquigarrow} X) = \det(A \overset{i}{\leftsquigarrow} B). \tag{14}$$

By (12), Equation (14) reads

$$(\det A)x_i = \det(A \overset{i}{\leftsquigarrow} B),$$

proving Cramer's Rule. □

Not suitable for engineering students, we now show a highly conceptual, though short, proof that only requires familiarity with Gaussian elimination. It is due to Richard Ehrenborg [8]. Since it uses row reduction, it is the only proof in this section that really requires coefficients and constants belonging to a field.

Proof 5 (Ehrenborg) Solving a linear system through Gaussian elimination consists in performing a finite sequence of row operations: (i) exchange two rows; (ii) multiply a row with a non-zero scalar; (iii) add one row to another row. The key point is that the quotient

$$\frac{\det(A \overset{i}{\leftsquigarrow} B)}{\det A}$$

is invariant under operation (i), (ii) and (iii) by Theorem 1.3. Once the nonsingular matrix A is row-reduced to the identity matrix, the system will be replaced by an equivalent system of the following form:

$$IX = \begin{pmatrix} \bar{x}_1 \\ \vdots \\ \bar{x}_n \end{pmatrix} \overset{\text{def}}{=} \bar{X}, \tag{15}$$

where $(\bar{x}_1, \ldots, \bar{x}_n)$ is obviously a solution of both (1) and (15). Thus, we get a sequence of equalities

$$\frac{\det(A \overset{i}{\leftsquigarrow} B)}{\det A} = \cdots = \frac{\det(I \overset{i}{\leftsquigarrow} \bar{X})}{\det I} = \frac{\bar{x}_i}{1} = \bar{x}_i,$$

that proves Cramer's Rule. □

Our parade of proofs ends with an item not picked up from literature.

Proof 6 Since the matrix A of coefficients is nonsingular, its rank and the rank of the augmented matrix are both equal to n. By Theorem 1.5, the system (1) has at least one solution. Let $(\bar{x}_1, \ldots \bar{x}_n)$ be one of them. If we write the equation

$$x_i = \bar{x}_i$$

below the others in (1), then the resulting system $\mathcal{S}$ with $n+1$ equations and n unknowns is still compatible. As a consequence of Theorem 1.5, the determinant of

$$\begin{pmatrix} A^1 & \cdots & A^i & \cdots & A^n & B \\ 0 & \cdots & 1 & \cdots & 0 & \bar{x}_i \end{pmatrix},$$

the augmented matrix associated to $\mathcal{S}$, vanishes. By performing such determinant through cofactor expansion along the $(n+1)$-th row, we get

$$0 = (-1)^{(n+1+i)} \det(A^1 \ldots A^{i-1} A^{i+1} \ldots A^n \; B) + \bar{x}_i \det A. \tag{16}$$

The matrix $(A \overset{i}{\leftsquigarrow} B)$ can be obtained from $(A^1 \dots A^{i-1} A^{i+1} \dots A^n\ B)$ through $(n-i)$ swappings of columns. Hence, by Theorem 1.3 (I), Equation (16) becomes

$$0 = (-1)^{(n+1+i)}(-1)^{(n-i)}(A \overset{i}{\leftsquigarrow} B) + \bar{x}_i \det A,$$

which proves Cramer's Rule, and uniqueness of solution in particular, since

$$(-1)^{(n+1+i)}(-1)^{(n-i)} = (-1)^{2n+1} = -1.$$

3 Pedagogical implications

On the educational side, Cramer's Rule has its fervent opponents, who always emphasize its computational inefficiency when the number n of equations becomes large, say $n \geq 4$. Indeed, when determinants are calculated via minors, Cramer's Rule time complexity is $\mathcal{O}(n!)$, which makes it useless for any practical application when compared to Gaussian elimination, whose complexity stands at $\mathcal{O}(n^3)$.

Such argument recently lost much of its appeal. Through condensation methods, Cramer's Rule time complexity can be also reduced to $\mathcal{O}(n^3)$ (see [9]).

Nevertheless, no argument of that sort will convince the majority of students to abandon their most beloved technique for solving linear system: substitution, which is also, incidentally, one of the most reliable source of errors in written tests.

After listing six proofs of Cramer's Rule, it is natural to ask whether one of them "should" replace Proof 3 in its present role of standard arguing procedure in a first-year linear algebra course.

The answer very much depends on how the lecturer organizes material in his classroom. Not all proofs use, in fact, the same tools, and the "best" one should be chosen taking into account the theoretical contents of previous lectures. Next table recaps the background devices on which the various proofs rely.

The checkmark in parentheses means that the 2nd Laplace Theorem is implicitly used in Proof 3, since you need it to prove the formula for the inverse of a matrix A in terms of its determinant and adjoint.

	1	2	3	4	5	6
1st Laplace Theorem	✓		✓	✓	✓	✓
2nd Laplace Theorem	✓		(✓)			
Elementary Property I					✓	✓
Elementary Property II		✓			✓	
Elementary Property III		✓			✓	
Matrix products			✓	✓	✓	
Inverse matrix			✓			
Binet Theorem				✓		
Gaussian elimination					✓	
Rouché-Capelli Theorem						✓

Acknowledgements. The author is indebted to Piero Brunetti, electrotechnical engineer and an amateur mathematician in his own right, for the germinal idea behind Proof 6.

References

[1] C. B. Boyer, *Colin Maclaurin and Cramers Rule*, Scripta Math. **27** (1966), 377–379.

[2] F. Brioschi, *Teorica dei determinanti e le sue applicazioni*, Pavia: Bizzoni (1854).

[3] A. Capelli, *Sopra la compatibilità o incompatibilità di pi equazioni di primo grado fra più incognite*, Rivista di Matematica **2** (1892), 54–58.

[4] D. Carlson, C. R. Johnson, D. Lay and A. Duane Porter, *Gems of Exposition in Elementary Linear Algebra*, College Math. J. **23**, no. 4 (1992), 299–303.

[5] D. Cohen, T. B. Lee and D. Sklar, *Precalculus: A Problems-Oriented Approach, Enhanced Edition*, 6th ed., Belmont: Cengage Learning (2010).

[6] G. Cramer, *Introduction à l'analyse des lignes courbes algébriques*, Geneva: Freres Cramer & Cl. Philbert (1750).

[7] C. W. Curtis, *Linear Algebra*, Boston: Allyn and Bacon (1963).

[8] R. Ehrenborg, *A Conceptual Proof of Cramer's Rule*, Math. Mag. **77** no. 4 (2004), 308.

[9] K. Habgood and I. Arel, *A condensation-based application of Cramer's Rule for solving large-scale linear systems*, J. Discrete Algorithms, **10** (2012), 88-109.

[10] B. A. Hedman, *An earlier date for 'Cramer's rule"*, Historia Math. **26** no. 4 (1999), 365–368.

[11] R. A. Horn and C. A. Johnson, *Matrix Analysis*, New York: Cambridge University Press (1985).

[12] C. G. J. Jacobi, *De Formatione et Proprietatibus Determinantium*, in Idem, *Gesammelte Werke* vol. III, Berlin: G. Reimer (1884), 355-392.

[13] A. A. Kosinsky, *Cramer's Rule is due to Cramer*, Math. Mag. **74** no. 4 (2001), 310–312.

[14] S. Lang, *Linear Algebra*, 3rd ed., New York: Springer-Verlag (2004).

[15] L. A. Lomonaco, *Geometria e algebra. Vettori, equazioni e curve elementari*, Roma: Aracne (2013).

[16] T. Muir, *The theory of determinants in the historical order of development*, vol. 1, New York: Dover Publications (1960).

[17] *Regola di Cramer*, in Wikipedia. L'Enciclopedia libera, http://it.wikipedia.org/wiki/Regola_di_Cramer, accessed February 8, 2014.

[18] D. J. S. Robinson, *A Course in Linear Algebra with Applications*, New Jersey: World Scientific (2006).

[19] S. M. Robinson, *A Short Proof of Cramer's Rule*, Math. Mag. **43** no. 2 (1970), 44–45.

[20] E. Rouché, *Sur la discussion des équations du premier degrè*, C. R. Math. Acad. Sci. Paris **LXXXI** (1875), 1050–1052.

[21] D. Serre, *Matrices. Theory and Applications*, 2nd ed., GTM 216, New York: Springer-Verlag (2010).

[22] N. Trudi, *Teoria de determinanti e loro applicazioni*, Napoli: Pellerano (1862).

[23] A-T. Vandermonde, *Mémoire sur I'élimination*, 2 partie, Hist. de l'Acad. Roy. des Sciences, Paris (1772), 516–532.

[24] D.E. Whitford and M. S. Klamkin, *On an elementary derivation of Cramer's rule*, Amer. Math. Monthly **60** (1953), 186–187.

Received: May 5, 2014