Reprinted from the MATHEMATICS MAGAZINE
Vol. 58, No. 4, September 1985

The Mathematics of Double Entry Bookkeeping

DAVID P. ELLERMAN
Boston College
Chestnut Hill, MA 02167

For several centuries, the double entry system has been the bookkeeping system used by most sizable business enterprises throughout the world. It is little known in mathematics and it is virtually unknown in accounting that the double entry system is based on a well-known mathematical construction of undergraduate algebra, the group of differences, in which the integers are represented as equivalence classes of ordered pairs of natural numbers.

The T-accounts of double entry bookkeeping are precisely the ordered pairs of the group of differences construction. With the one exception of a paragraph by D. E. Littlewood (see below), the author has not found a single mathematics book, elementary or advanced, popular or esoteric, which notes that this construction is the theoretical basis of a mathematical technique applied, day in and day out, in the mundane world of business for over five centuries. And even though the construction is standard fare in an undergraduate modern algebra course, the connection with double entry bookkeeping is totally absent in the accounting literature (see References; **[4]** gives a review of the literature in mathematical accounting).

The encounters between mathematics and double entry bookkeeping have been so sparse that the highlights can be easily specified. A description of double entry bookkeeping was first published by the Italian mathematician Luca Pacioli in 1494 [**9**]. The system had been developed in Italy during the fourteenth century. Although Pacioli's system was governed by precise rules, his presentation was in a practical and nonmathematical form.

As an abstract mathematical construction, the group of differences seems to have been first published by Sir William Rowan Hamilton in 1837 [**6**]. Hamilton presented the ordered-pairs construction of the integers as a prelude to his ordered-pairs treatment of the complex numbers. He made no mention of bookkeeping although accountants had, at that time, been using the intuitive algebra of the ordered pairs called "*T*-accounts" for about four centuries.

Arthur Cayley (1821–1895) was one of the few later mathematicians who wrote about double entry bookkeeping. In the year before his death, he published a small pamphlet entitled *The Principles of Book-keeping by Double Entry* in which he wrote:

> The Principles of Book-keeping by Double Entry constitute a theory which is mathematically by no means uninteresting: it is in fact like Euclid's theory of ratios an absolutely perfect one, and it is only its extreme simplicity which prevents it from being as interesting as it would otherwise be. [**1**, Preface]

In the pamphlet, Cayley did not present a mathematical formulation, but only described double entry bookkeeping in the practical informal terms familiar to Cayley from his fourteen years of work as a lawyer. However, in his presidential address to the British Association for Advancement of Science, Cayley hinted that the "notion of a negative magnitude" is "used in a very refined manner in bookkeeping by double entry" [**2**, p. 434].

Another brief but insightful observation was made in a semipopular work by D. E. Littlewood in which he noted that the ordered pairs in the group of differences construction function like the debit and credit balances in a bank account.

> The bank associates two totals with each customer's account, the total of moneys credited and the total of moneys withdrawn. The net balance is then regarded as the same if, for example, the credit amounts of £102 and the debit £100, as if the credit were £52 and the debit £50. If the debit exceeds the credit the balance is negative.
>
> This model is adopted in the definition of signed integers. Consider pairs of cardinal numbers (a, b) in which the first number corresponds to the debit, and the second to the credit. A definition of equality is adopted such that
>
> $$(a, b) = (c, d)$$
>
> if and only if $a + d = b + c$. [**8**, p. 18]

Some modern accounting theorists believe that the mathematical treatment of double entry bookkeeping must involve transaction matrices. The presentation of transactions involving scalars can be facilitated using a square array or table of scalars usually called a "transactions matrix." These transactions tables were first used by the English mathematician Augustus DeMorgan [3], and have been popularized a century later by the American mathematician John Kemeny and his colleagues in an influential text [7].

Transactions tables have, however, retarded the development of a mathematical formulation of double entry bookkeeping. As will be seen below, double entry bookkeeping lives in group theory, *not* in matrix algebra. When double entry bookkeeping is mathematically formulated using the group of differences, it can be generalized to new systems of accounting such as vector accounting (or even "fraction accounting" where multiplication replaces addition). But the relatively superficial use of matrix algebra involved in "transactions matrices" does not generalize to these new domains [4, Chapter 12, section 2, "Transactions Matrices"]. Hence transactions tables will not be used here.

In what follows, we give an elementary introduction to the modern mathematical formulation of double entry bookkeeping. The basic concepts of double entry bookkeeping are introduced in their natural mathematical context.

The construction of the Pacioli group

Double entry bookkeeping is based on the construction of the integers (positive and negative) as equivalence classes of ordered pairs of natural numbers. The ordered pairs of this construction correspond to the T-accounts of double entry bookkeeping. The left-hand entry in the ordered pair corresponds to the **debit** side of the T-account, and the right-hand entry to the **credit** side. The notation $[d//c]$ for a T-account is drawn from Pacioli himself.

> At the beginning of each entry, we always provide "per", because, first, the debtor must be given, and immediately after the creditor, the one separated from the other by two little slanting parallels (virgolette), thus, //.... [9, p. 43]

Thus a general T-account with a debit entry of d and a credit entry of c will be represented as

$$[d//c] = \frac{\text{Debits} \quad \text{Credits}}{d \;\;\big|\;\; c}.$$

Since the label "T-account" will be used later in a specific accounting context, the general ordered pairs $[d//c]$ will be called ***T*-terms**. The additive group of integers is obtained by defining addition on T-terms and then imposing an equivalence relation compatible with the addition operation. Given T-terms $[w//x]$ and $[y//z]$, their **sum** is defined as the T-term obtained by adding debit to debit and credit to credit, i.e.,

$$[w//x] + [y//z] = [w + y//x + z].$$

The **zero *T*-term** $[0//0]$ is the T-term which 'acts' like zero in the sense that adding it to any T-term makes no difference, i.e.,

$$[d//c] + [0//0] = [d//c].$$

Given two T-terms $[w//x]$ and $[y//z]$, the **cross-sums** are the two numbers obtained by adding the debit in one to the credit in the other:

$$\overbrace{[w//x] \quad [y//z]}^{x+y} \quad \underbrace{}_{w+z}$$

Cross-sums: $x + y \qquad w + z.$

Two T-terms are set **equal** if their cross-sums are equal, i.e.,

$$[w//x] = [y//z] \quad \text{if and only if} \quad x + y = w + z.$$

This equivalence relation is compatible with addition in the sense that if

$$[w//x] = [w'//x'] \quad \text{and} \quad [y//z] = [y'//z']$$

then

$$[w + y//x + z] = [w' + y'//x' + z'].$$

Thus addition is well-defined on equivalence classes independently of their representatives. The notation $[d//c]$ will henceforth be used to represent the equivalence class of the ordered pair.

The numbers occurring in a T-term can never be negative, but we can still define the negative of a T-term without negative numbers. The negative of a T-term $[y//z]$ is another T-term such that when added to $[y//z]$ the sum is the zero T-term. It suffices to reverse the debit and credit entries. Hence we define the negative or **inverse** of a T-term $[y//z]$ as its 'reverse' $[z//y]$, since

$$[y//z] + [z//y] = [y + z//y + z] = [0//0].$$

This completes the definition of the ordered-pairs construction of the integers from the natural numbers. In view of the connection with double entry bookkeeping, we will call it the **Pacioli group**.

The double entry method of bookkeeping

The double entry method uses the Pacioli group to perform additive algebraic operations on equations. First we must translate or **encode** equations into the Pacioli group. A T-term equal to the zero T-term $[0//0]$ will be called a **zero-term**. For example, $[x//x]$ is a zero-term even when x is nonzero. The translation of equations into the Pacioli group is very simple: equations between nonnegative numbers correspond to zero-terms. That is, for any nonnegative numbers w and y,

$$w = y$$

if and only if

$$[w//0] + [0//y] = [w//y] = [0//0].$$

In more general terms, given any equation where all numbers are nonnegative such as $w + \cdots + x = y + \cdots + z$, we encode each left-hand-side number as a **debit-balance** T-term, such as $[w//0]$, and we encode each right-hand-side number as a **credit-balance** T-term, such as $[0//y]$. Then the original equation holds if and only if the sum of the encoded T-terms is a zero-term:

$$w + \cdots + x = y + \cdots + z$$

if and only if

$$[w//0] + \cdots + [x//0] + [0//y] + \cdots + [0//z]$$

is a zero-term.

This translation or encoding of equations into zero-terms works even if the original equation contains negative numbers since any equation can be transformed to one all of whose terms are positive by transferring the negative numbers to the other side.

In double entry bookkeeping, transactions must be recorded in such a way as to maintain the truth of an equation such as the balance sheet equation:

$$\text{Assets} = \text{Liabilities} + \text{Net Worth}.$$

That is, transactions must be recorded by valid algebraic operations which transform equations into equations. In the Pacioli group, an equation translates into a zero-term, so a valid algebraic operation would be an operation that transforms zero-terms (equations) into zero-terms (equations). But there is *only one such operation: add a zero-term*. Zero plus zero equals zero. Thus a transaction must be represented by a zero-term to be added to the zero-term representing the balance sheet equation.

In bookkeeping, the **double entry principle** is that each transaction must be recorded with equal debits and credits. The mathematical basis for this principle is simply that transactions are represented by zero-terms (so the debits must equal the credits in the transaction). In double entry bookkeeping, the zero-terms arising as the representations of equations (e.g., the balance sheet equation) and of transactions will be called respectively **equational zero-terms** and **transactional zero-terms**.

Any valid (additive) algebraic operation on an equation then boils down to one scheme:

$$\begin{array}{l} \text{original equational zero-term} \\ \underline{+\ \text{transactional zero-term}} \\ =\ \text{final equational zero-term.} \end{array}$$

In bookkeeping, there are many transactions to record, but each is still represented by a transactional zero-term. The result of adding many zero-terms to the original equational zero-term still yields another zero-term, the final equational zero-term.

It remains to specify how to reverse the translation process, how to **decode** zero-terms as equations. A zero-term, such as an equational zero-term, is a sum of T-terms. It is not itself an equation with a left- and right-hand side. Indeed, the T-terms can be shuffled around in any order. To decode a zero-term into an equation, one can use any criterion one wishes to divide the T-terms into two sets, L and R. Then construct an equation as follows: if a T-term $[d//c]$ is in the set L, decode it as the number $d-c$ on the left-hand side of the equation, and if $[d//c]$ is in the set R, decode it as $c-d$ on the right-hand side of the equation. This procedure will always yield a valid equation, given a zero-term. For example, consider the zero-term

$$[6//2]+[7//2]+[1//7]+[13//16].$$

Let the set L be, say, $[6//2]$ and $[1//7]$, and thus the set R contains the remaining T-terms, $[7//2]$ and $[13//16]$. The set L decodes as $6-2+1-7$ on the left-hand side, and R decodes as $2-7+16-13$ on the right-hand side, so we have the equation:

$$6-2+1-7=2-7+16-13.$$

In bookkeeping, the T-accounts in the final equational zero-term would be put in the sets L and R according to the side of the initial balance sheet equation from which the accounts were originally encoded.

A numerical example of double entry bookkeeping

Double entry bookkeeping is used to update the balance sheet equation

$$\text{Assets} = \text{Liabilities} + \text{Net Worth}$$

to show the effects of economic transactions. We develop a numerical example using very broad accounts (such as the three above). For simplicity, we will not use temporary or flow accounts such as Revenue or Expenses. The mathematical structure is the same regardless of the accounts used.

The initial equation must have all accounts expressed as positive numbers:

$$\begin{matrix} \text{Assets} & & \text{Liabilities} & & \text{Net Worth} \\ 15000 & = & 10000 & + & 5000. \end{matrix} \tag{1}$$

It is the position of the account in the all-positive equation that identifies the account as a left-hand side (LHS) or **debit-balance account**, or as a right-hand side (RHS) or **credit-balance account**.

The equation is encoded as an equational zero-term by encoding each debit-balance account as a debit-balance T-term $[d//0]$, and each credit-balance account as a credit-balance T-term $[0//c]$. Hence balance sheet equation (1) yields the equational zero-term

$$\begin{matrix} \text{Assets} & & \text{Liabilities} & & \text{Net Worth} \\ [15000//0] & + & [0//10000] & + & [0//5000]. \end{matrix}$$

These T-terms with accounting labels, like "Assets" and "Liabilities," attached to them may properly be called ***T*-accounts**. Since only plus signs appear between the T-terms, the plus signs may be left implicit and the T-terms may be reshuffled in any order. This yields what accountants call the

Ledger

$$\begin{matrix} \text{Assets} & \text{Liabilities} & \text{Net Worth} \\ [15000//0] & [0//10000] & [0//5000]. \end{matrix}$$

Thus a ledger is just an abbreviated form (without plus signs) of an equational zero-term, i.e.,

"Ledger" = "Equational Zero-term."

The sum of all the credit entries in a zero-term must equal the sum of the debit entries. In the case of the ledger, that summation is precisely what is called the **trial balance** in accounting.

According to the double entry principle, each transaction is recorded by an equal debit d and credit c so it would be represented in the Pacioli group by the zero-term $[d//0]+[0//c]$ obtained by encoding the equation $d=c$. **Debiting** any X to an account means adding the debit T-term $[X//0]$ to the T-account, and **crediting** X to an account means adding the credit T-term $[0//X]$ to the T-account. Hence the transaction represented by the zero-term $[d//0]+[0//c]$ would be recorded by adding the debit part $[d//0]$ and the credit part $[0//c]$ to the appropriate T-accounts.

It is a common mistake of nonaccountants to think that "debit" means "negative." But a debit-balance account like Assets does not have a negative balance. To debit an account does not necessarily mean to subtract from the balance in the account. That is only true for credit-balance accounts. Debiting a debit-balance account means *adding* to the account's balance.

Each transaction is recorded by adding the debit and credit parts of the transactional zero-term to the appropriate T-accounts. In accounting, the **journal** is the listing of the debit and credit parts of the transactional zero-terms and the T-accounts to which they are added. Thus

"Journal" = "List of Transactional Zero-terms."

Consider three simple transactions:

(1) $1200 is expended on inputs used in production,
(2) $1500 of outputs is produced and sold, and
(3) $800 of principal is paid on a loan.

Transaction (1) covers both the expenditure of $1200 to purchase inputs and the use of those inputs in production. Thus the debit-balance account of Assets is to be reduced by crediting it with $1200. The new assets produced are not recorded until transaction (2), so the other end of transaction (1) is a reduction in the credit-balance account of Net Worth by debiting it $1200. Transaction (2) records the production and sale of $1500 of new assets with an addition to the debit-balance account of Assets (a $1500 debit) and an addition to the credit-balance account of Net Worth (a $1500 credit). Transaction (3) records an $800 loan payment by equally reducing the Assets and Liabilities accounts, i.e., by crediting Assets and debiting Liabilities by $800.

Journal

Trans.	Accounts and Description	Debit Credit
1	Net Worth Assets $1200 expended on inputs	[1200 // 0] [0 // 1200]
2	Assets Net Worth $1500 of outputs produced and sold	[1500 // 0] [0 // 1500]
3	Liabilities Assets $800 principal payment on a loan	[800 // 0] [0 // 800]

The debit and credit parts of the transactional zero-terms can then be added or "**posted**" to the appropriate T-accounts in the ledger. Thus
"Posting to the Ledger" = "Adding the Transactional Zero-terms to the Equational Zero-term."

Assets	Liabilities	Net Worth
[15000// 0]	[0 //10000]	[0 //5000]
[0 //1200]		[1200// 0]
[1500 // 0]		[0 //1500]
[0 // 800]	[800// 0]	
[16500//2000]	[800//10000]	[1200//6500]

This yields the updated ledger or equational zero-term:

Assets	Liabilities	Net Worth
[16500//2000]	[800//10000]	[1200//6500].

Finding the balance in a T-account means subtracting as much as possible from each side so long as neither side becomes negative. The resulting T-term is said to be in **reduced form**. The quick way to put an account in reduced form is to take the minimum of the debit and credit entries, and subtract it from both sides. A T-account with a debit or credit entry of zero is in reduced form. The other possibly nonzero entry in the T-account in reduced form is called the **balance** of the account. Putting the above T-accounts in reduced form yields:

Assets	Liabilities	Net Worth
[14500//0]	[0//9200]	[0//5300].

It remains only to decode the equational zero-term (= ledger with plus signs between the T-accounts) to obtain the final balance sheet equation. That involves selecting the two sets L and R of T-terms to be decoded respectively on the LHS and RHS of the final equation. The selection is obvious; decode each T-account onto the side of the equation that it originally came from. The Assets account is in the set L, and Liabilities and Net Worth are in the set R. Hence the ledger decodes to yield the:

Final Balance Sheet Equation

Assets		Liabilities		Net Worth
14500	=	9200	+	5300.

Why the double entry system?

In conclusion, let us consider the distinctive features of the double entry system of bookkeeping. It is often thought that the characteristic feature of the system is the double aspect of the entries, the fact that at least two accounts are affected by each transaction. But the double entry system is a system of *recording* transactions. The fact that two or more accounts are affected is a characteristic of the transaction itself, not of the recording method. The same two or more accounts would be affected by any method of recording transactions which is based on updating a *complete* accounting equation. This is the general mathematical fact, having nothing in particular to do with accounting, that two or more terms must be changed when changes are made in an equation. Given an equation

$$a + b + \cdots + c = x + y + \cdots + z$$

it is impossible to change only one term, such as b, and still maintain the validity of the equation.

Another method of recording transactions, without using the double entry machinery of debits and credits, would be to directly perform additions and subtractions of positive numbers to the accounts in the balance sheet equation. But then there is no quick check on the plausibility of the transaction used to record a business event. Suppose that an event was formulated as the 'transaction' of adding $100 to both Liabilities and Net Worth. Is that possible? Some thought is required to see that this formulation of the business event could not possibly be correct. Much more thought would be necessary for a multiple (two or more) entry transaction in an accounting system with hundreds of accounts. Yet the check is immediate in the double entry system. Liabilities and Net Worth are both credit-balance accounts so the proposed 'transaction' is a double credit in violation of the double entry principle.

A third method of recording transactions, which does involve a quick check, is obtained by moving all the terms in the balance sheet equation to one side:

$$-\text{Assets} + \text{Liabilities} + \text{Net Worth} = 0.$$

Then each bona fide transaction would be recorded by adding equal positive and negative amounts to the accounts (some of which would have a negative balance). If the positive and negative amounts did not sum to zero, it could not represent a bona fide transaction. The problem with this accounting method is the counterintuitiveness of having negative balances in some accounts, e.g., Assets. An increase in such accounts would be represented by adding a negative number to the (negative) balance in the accounts, and a decrease would be represented by adding a positive number to the (negative) balance.

All three of these methods are logically sound, and all three record a transaction with "double entries" affecting two (or more) accounts. The popularity of the double entry method down through the centuries seems to be based on the fact that it provides a quick check on the plausibility of a transaction (the double entry principle of equal debits and credits) and that it uses the symmetric T-accounts (the ordered pairs of the group of differences construction) to represent the balance in each account as a positive number.

References

[1] A. Cayley, The Principles of Book-keeping by Double Entry, Cambridge University Press, Cambridge, 1894.

[2] ______, Presidential address to the British Association for the Advancement of Science, The Collected Mathematical Papers of Arthur Cayley, Vol. XI, Cambridge University Press, Cambridge, 1896.

[3] Augustus DeMorgan, On the main principle of book-keeping, Elements of Arithmetic, James Walton, London, 1869.

[4] David Ellerman, Economics, Accounting, and Property Theory, D. C. Heath, Lexington, Mass. 1982.

[5] John B. Geijsbeek, Ancient Double-Entry Bookkeeping, Scholars Book Company, Houston, 1914.

[6] Sir William Rowan Hamilton, Theory of conjugate functions, or algebraic couples; with a preliminary and elementary essay on algebra as the science of pure time, Trans. Royal Irish Academy, XVII (1837) 293–422. Also reprinted in The Mathematical Papers of Sir William Rowan Hamilton, III Algebra, H. Halberstam and R. E. Ingram (eds), Cambridge University Press, Cambridge, 1967.

[7] J. Kemeny, A. Schleifer, J. L. Snell, and G. Thompson, Finite Mathematics with Business Applications, Prentice-Hall, Englewood Cliffs, N.J., 1962.

[8] D. E. Littlewood, The Skeleton Key of Mathematics, Harper Torchbooks, New York, 1960 (orig. published in 1949).

[9] L. Pacioli, Summa de Arithmetica, Geometrica, Proporcioni et Proporcionalita, trans. by J. B. Geijsbeek as Ancient Double-Entry Bookkeeping, reprinted by Scholar Book Company (orig. published in 1494).