Taking Place Value Seriously:
Arithmetic, Estimation and Algebra

0. Introduction and Summary

Arithmetic, first of whole numbers, then of decimal and common fractions, and later of rational expressions and functions, is a central theme in school mathematics. This essay attempts to point out ways to make the study of arithmetic more unified and more conceptual through systematic emphasis of place value structure in the decimal number system.

The essay is divided into five sections, with two appendices and a set of related exercises and problems.

The first section reviews the basic principles of decimal notation. It points out the sophisticated structure which underlies the amazing efficiency of decimal notation, and it introduces the terminology of *digit, denomination,* and *decimal component,* for the basic constituents out of which decimal numbers are formed. To enable discussion of the sizes of decimal numbers, it introduces the idea of *order of magnitude* of decimal components and numbers.

The second section discusses the how decimal notation permits efficient algorithms for the basic operations of addition and multiplication for whole numbers. The main theme is that the form of decimal numbers, as sums of the very special numbers called decimal components, together with the Rules of Arithmetic,[1] determines the main outlines of the procedures for adding and multiplying. The inverse operations, of subtraction and division, are discussed in appendices.

Section 3 discusses ordering, estimation and approximation of numbers. For comparing numbers, a crucial idea is that of *relative* place value. The corresponding idea for approximation is *relative error.* Both relative and absolute error can be controlled in terms of decimal expansions. A key concept is *significant digit.* It is pointed out that relative accuracy of approximation improves rapidly with the number of significant digits. It is usually unreasonable to expect to know a "real-life" number (meaning the result of a measurement) to more than three or four significant digits, and often one must settle for, and can live with, much less. Failure to appreciate the limits of accuracy seems to be one of the most pervasive forms of innumeracy: it affects many people who are for the most part quite comfortable

[1]By the "Rules of Arithmetic", we mean what mathematicians refer to as the Field Axioms, and what are often mentioned in mathematics education literature as "number properties", or just "properties". (They are, in fact, not properties of numbers, but properties of the operations.) They are nine in number: four (Commutative, Associative, Identity and Inverse Rules) for addition, four parallel ones for multiplication, and the Distributive Rule to connect addition and multiplication. Other rules, such as "Invert and multiply", or the formula for adding fractions, or the rules of signs for dealing with negative numbers, can be deduced from these nine basic rules.

with numbers. Scientific notation, which focuses attention on the size of numbers and the accuracy to which they are known, is discussed in an appendix.

The fourth section combines the ideas from sections 2 and 3 in a short discussion of accuracy and estimation in arithmetic computation. Focusing on decimal components permits a highly effective treatment of this issue.

The last section explores the connections of decimal notation with algebra. It is suggested that decimal numbers can be profitably thought of as "polynomials in 10", and the parallels between decimal computation and computation with polynomials are illustrated. Moreover, the structural understanding that would come with a study of arithmetic based on decimal components should promote comfort with algebraic manipulation. Further parallels between decimal arithmetic and algebra are explored in the accompanying problems.

This essay addresses an issue which the author hopes will be of interest to a broad spectrum of mathematics educators (a term we use inclusively, to comprise mathematics teachers of all levels, as well as educators with professional interest in mathematics). It probably will meet with the same problem which bedevils every teacher of mathematics – variations in sophistication and background among its intended audience. It is hoped that many readers can appreciate the broad message, that place value can serve as an organizing and unifying principle across a surprising span of the elementary (and beyond!) mathematics curriculum; and that individual readers will be patient with particular parts which may seem either over- or under-elaborated, keeping in mind that other readers may have the opposite view. The earlier sections tend to provide more details. Later sections, and especially the appendices, are terser.

Acknowledgements: The author is grateful for comments on earlier drafts of this essay from the MSSG committee members, from Johnny Lott and his associates in NCTM and ASSM, and from Scott Baldridge, Thomas Roby, and Kristin Ulmann. Thanks to Mel Delvecchio and W. Barker for vital production help.

1. The decimal system: place value, expanded form and decimal components.

Our common decimal system is a highly sophisticated method for writing whole numbers efficiently. It uses only 10 symbols (the *digits*: 1 to 9, together with 0), arranged in carefully structured groups, to express any whole number. Further, it does so with impressive economy. To express the total human population of the world would require only a ten-digit number, needing just a few seconds to write.

The efficiency of the decimal system is possible because of its systematic use of mathematical structure – essentially, it enlists algebra in support of counting.

Making this structure explicit may help to make mathematics instruction more effective, by increasing conceptual understanding and computational flexibility. It should promote numeracy by making students more sensitive to order of magnitude, and to the estimation capabilities of decimals. It should also bring out the parallels between arithmetic and algebra, thus making arithmetic a preparation for algebra, rather than the impediment that it now sometimes is [KSF, Ch.8].

We will call a whole number expressed in decimal form a *decimal number.* The first thing to understand about decimal notation is that the decimal expression of a number implicitly breaks the number up into a sum of numbers of a very special type. Thus

$$7,452 = 7,000 + 400 + 50 + 2.$$

This is usually called *expanded form.* It is mentioned in many state standards documents. This essay explores what would be entailed in making it central to arithmetic instruction.

We will call the summands in the expanded form of a number the *decimal components* of the number. Each decimal component is a digit (possibly 0) times a power of 10:

$$7,000 = 7 \times 1000 = 7 \times (10 \times 10 \times 10) = 7 \times 103.$$

$$400 = 4 \times 100 = 4 \times (10 \times 10) = 4 \times 102.$$

$$50 = 5 \times 10 = 5 \times 101.$$

$$2 = 2 \times 1 = 2 \times 100.$$

We will call the powers of 10 *denominations.* (In this, we follow [BP].) Thus, a decimal component is a digit times a denomination.

We will call the exponent involved in a denomination the *order of magnitude* of the unit. The order of magnitude may also be described as the number of zero digits used to write the denomination, or one less than the total number of digits used. By the order of magnitude of a (non-zero) decimal component, we mean the order of magnitude of its denomination. Finally, the order of magnitude of a decimal number is defined to be the order of magnitude of its largest (non-zero) decimal component. Thus, 7,452 has order of magnitude 3, the same as its largest decimal component, 7,000, and the 400, the 50 and the 2 have orders of magnitude 2, 1 and 0, respectively.

For this essay, "order of magnitude" replaces the usual naming of places. We want to replace the standard terminology involving specific place names with this order-of-magnitude terminology in order to be able to better discuss the relationship between different orders of magnitude. It will often be convenient to abbreviate "order of magnitude" to just "magnitude".

A key point about the expanded form of a number is that no two decimal components have the same order of magnitude. This condition in fact characterizes the expanded forms: a sum of decimal components is the expanded form of a number exactly when it involves at most one component of any given order of magnitude. The non-zero digits of the decimal expression of the number are then just the digits coming from the decimal components, each in its corresponding place. Besides these, for orders of magnitude smaller than the magnitude of the number, if the decimal component of that magnitude is "not there", i.e., is zero, then one puts a zero in the corresponding place of the number, to indicate the absence of any multiple of that power of 10, so that the orders of magnitude corresponding to the non-zero digits can be read correctly. This is the principle of place value. Thus

$$
\begin{aligned}
82 &= 80 + 2 = 8 \times 10 + 2 \times 1. \\
802 &= 800 + 2 = 8 \times 100 + 0 \times 10 + 2 \times 1. \\
80020 &= 80,000 + 20 = 8 \times 10,000 + 0 \times 1,000 + 0 \times 100 + 2 \times 10 + 0 \times 1.
\end{aligned}
$$

2. Decimal components and the algorithms of arithmetic.

The fact that decimal numbers are sums has a pervasive influence on the methods for computing with them. In fact, the procedures for carrying out the four arithmetic operations with decimal numbers are largely determined by the fact that numbers are sums of their decimal components, together with the Rules of Arithmetic (see footnote, page 1).

Addition:

The basic strategy for adding two decimal numbers is to break each number (summand) into its decimal components, and, for each order of magnitude, to add the components of that magnitude. These sums are then recombined into a decimal number. The details of this last step give rise to the more finicky parts of the addition algorithm.

One can justify this process of decomposing, adding components, and then recombining by means of the two most basic rules of addition: the Commutative Rule and the Associative Rule. We recall them.

The Commutative Rule: The value of a sum does not depend on the order of the summands: for any two whole numbers, a and b,

$$a + b = b + a.$$

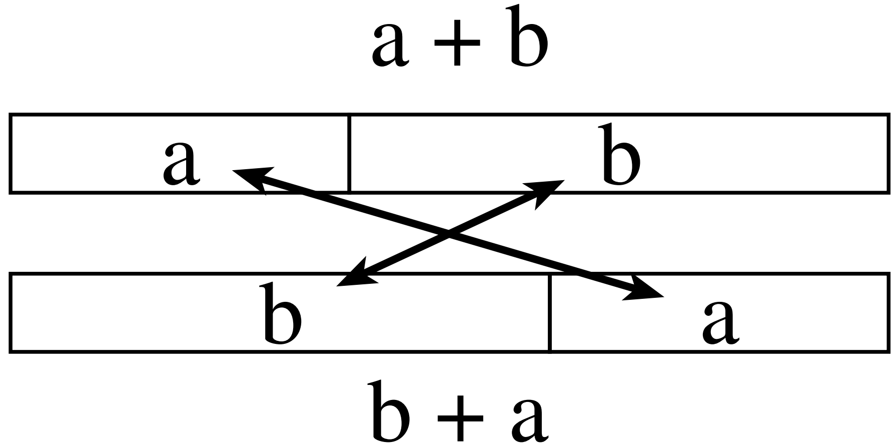
The Associative Rule: When adding three numbers, the value of the sum does not depend on the way the numbers are combined into pairwise sums: more precisely, for any three whole numbers a, b and c,

$$(a + b) + c = a + (b + c).$$

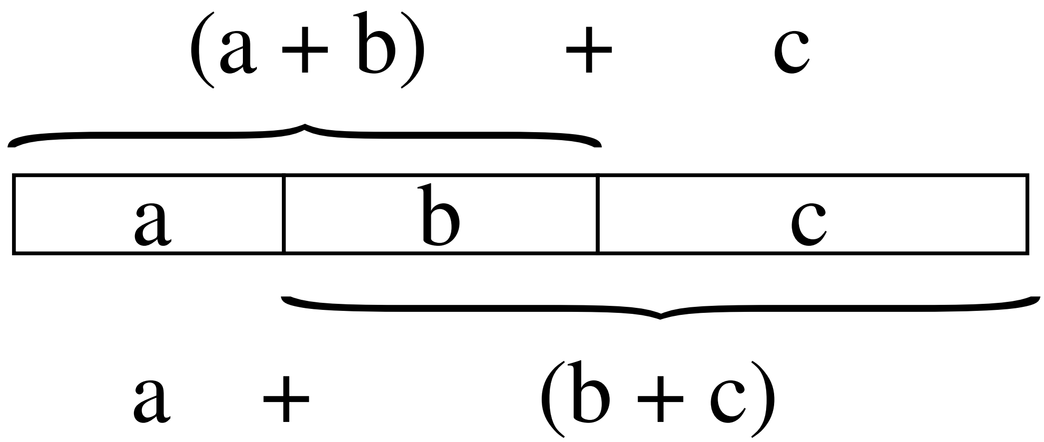
These rules are easy to justify in terms of common models of addition. If we model addition by, say, concatenation of lengths (sticking bars end to end), the Commutative Rule can be demonstrated by combining bars of lengths a and b into a bar of length $a + b$, and then rotating the bar around its midpoint to reverse the order of the summands. Associativity is even simpler: one just takes a bar composed of subbars of lengths a, b and c, and observes that it can be thought of as being made of subbars of length $a + b$ and c, or, equally easily, of subbars of lengths a and $b + c$. It is all in the point of view.

Additon Rules Diagrams

Commutativity



Associativity



Although the Commutative and Associative Rules are basic principles of arithmetic, one rarely uses them in isolation. Most practical manipulations call for more or less complicated combinations of both the Commutative and Associative Rules. Therefore, it is advisable to discuss explicitly the logical consequence of these rules, which is that, when a list of numbers is to be added, it does not matter what sequence of pairwise sums we do, or the order of the summands in any of these intermediate sums: the final result will always be the same. We call this Any Which Way Rule, a name which emphasizes the freedom we have in performing additions. (In [BP], it is called the Any Order Property.) Although a formal demonstration of the Any Which Way Rule is probably not appropriate at the level that multi-digit addition is first discussed, nevertheless an explicit discussion of the rule, with examples, particularly examples showing how it can simplify calculations and aid mental math, does seem feasible and advisable.

In any case, the Any Which Way Rule justifies the strategy of adding two decimal numbers by breaking each summand into its decimal components, adding pairs of components of the same order of magnitude, then adding all the results. The question next is, what does this do for us? That is, why is it an effective method?

The first part of the answer to this question comes from looking at the sum of two decimal components of the same order of magnitude. These addition problems simply amount to applying the "basic addition facts" (i.e., how to add two single digit numbers), plus keeping track of the order of magnitude. Thus, for example,

$$7,000 + 2,000 = 7 \times 1000 + 2 \times 1000 = (7 + 2) \times 1000 = 9 \times 1000 = 9,000.$$

It is always the same: to add two decimal components of the same order of magnitude, simply add their digits, then append the order of magnitude number of zeros on the right. The formal justification for this uses the Distributive Rule, which we will discuss later on. It can also be thought of as keeping track of units: in the sum above, we are adding 7 thousands to 2 thousands, and the result is (7 + 2) thousands, or 9 thousands.

This observation is sufficient by itself for adding many pairs of numbers. For example

$$7,452 + 1,326 = (7,000 + 1,000) + (400 + 300) + (50 + 20) + (2 + 6)$$

$$= 8000 + 700 + 70 + 8 = 8,778.$$

However, addition will be this simple only when, for every order of magnitude, the pair of decimal components of that magnitude have digits summing to less than 10. When this happens, all sums will again be decimal components, of the same magnitude, and may be directly put back together to form the sum of the original numbers. Whenever a digit sum is more than 10, finding the sum gets more complicated. For example,

$$7,000 + 6,000 = 13,000 = 10,000 + 3000.$$

Again, this is a repetition on a larger scale of a basic addition fact, which is why those facts are so important. The new problem presented by this sum is that it has two decimal components, one of the same magnitude as the components being added, and the other with magnitude one larger. When the sum of two decimal components has a component of larger order of magnitude, this must be added to the components of the summands of that magnitude. This is "carrying" or "regrouping" or "renaming". An example of addition with regrouping is

$$7,453 + 1,729 = (7000 + 1000) + (400 + 700) + (50 + 20) + (3 + 9)$$

$$= 8,000 + 1,100 + 70 + 12$$

$$= 8,000 + (1,000 + 100) + 70 + (10 + 2)$$

$$= (8,000 + 1,000) + 100 + (70 + 10) + 2$$

$$= 9,000 + 100 + 80 + 2 = 9,182.$$

Since (when adding two numbers) the sum of components of a given magnitude contributes directly at most to this magnitude and to the next larger one, an efficient way of finding any sum would be to start with order of magnitude zero (the ones place), find the sum of those components, combine the 10 (if it occurs) with the magnitude one components, find the sum of those components, and so forth, progressing systematically to larger orders of magnitude, one at a time. This is exactly what one does in the standard addition algorithm, although it may not be expressed in these terms, and the relevance of the expanded form may be suppressed in a strictly procedural approach. The procedure may also be adapted readily to addition of many numbers ("adding the columns").

Thinking explicitly about the component-by-component aspect of decimal addition can give greater flexibility in doing addition. For example, in adding two-digit numbers mentally, it is usually easier to think of adding the 10s and adding the 1s separately (which is exactly using expanded form in this situation). Further, one usually adds the 10s first, since this gives the main part of the sum, and then adds the sum of the 1s to that. If there is a carry from the 1s, it just means increasing the result from the 10s addition by one more 10, which is easy to do when there are only two decimal places to worry about.

We will see later on that these same considerations lead to quick estimation procedures for sums, easy enough to be carried out mentally in many cases.

In summary: if addition is thought of in terms of expanded form, then the basic strategy in doing addition of two decimal numbers is this:

i a) For each order of magnitude, take the decimal component of that order of magnitude from each summand, and add them. The sum is the sum of the digits of the two components, times their common denomination.

ii a) In cases when the digit sums are all less than 10, the component sums will be the decimal components of the sum to be found.

ii b) If, however, a digit sum is 10 or more, the 10 times the relevant denomination produces the denomination of the next larger magnitude, which must be added to the components of that magnitude.

These ideas need not be introduced in a single large lump. They can be taught as the decimal system is now, in a gradual way, starting with two-digit numbers. This would simply involve recalling that a two digit number means so many 10s and so many 1s, e.g., 47 is four 10s and seven 1s. (Probably explicit attention to this would be helpful to many students in any case, since the irregularities of English two-digit number words somewhat impede children to think in term of place value [KSF, p.167].) Then the strategy of "adding the 10s and adding the 1s" should appear more or less natural, perhaps after discussion. First problems with no carries could be done to establish the basic principle, then there should be consideration of what to do when one or the other of the digit sums is greater than 10. This approach should extend easily to numbers with more digits.

In fact, the basic grouping principle of decimal notation is not something that all students take to readily. Even before general two-digit addition is considered, it can be valuable to emphasize the grouping principle involved in decimal notation by using the "make a ten" idea [KSF, p.189] in learning the addition facts with sums above ten. As explained in [Ma], Chinese teachers think of the addition facts as not simply as a list to be memorized, but as a place to begin teaching the structure of the decimal system, and they emphasize the process of forming ten ones into a 10 when a sum of digits is greater than 10, and of decomposing a 10 back into ones in the corresponding subtraction problems.

Multiplication:

Thinking in terms of decimal components also sheds light on multiplication and division. However, the development requires a firmer grasp of the rules of arithmetic than was needed for addition. We begin with a discussion of the relevant rules.

Multiplication satisfies the Commutative Rule and the Associative Rule, just as addition does, but the justification for these rules is more work. Probably the best approach is through the Array Model, which is in any case an important interpretation of multiplication. The Array Model also helps to prepare for understanding area.

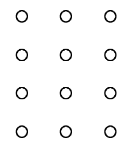
If we model a whole number as a horizontal row of some sort of object, say small circles, then multiplication involves replicating the row a certain number of times. Thus, if 3 is pictured as

$$\circ \quad \circ \quad \circ,$$

then 4×3 can be represented

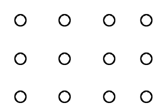
$$\circ \quad \circ \quad \circ \,|\, \circ \quad \circ \quad \circ \,|\, \circ \quad \circ \quad \circ \,|\, \circ \quad \circ \quad \circ .$$

For the Array Model, instead of arranging all the groups along the same line, we stack them one above the other, to make a rectangular array.



This is the Array Model for 4×3: four horizontal rows of three items each, stacked vertically above each other.

Using the Array Model, it is easy to justify the Commutative Rule for multiplication: one just flips an $m \times n$ array over its diagonal, turning it into an $n \times m$ array:



A justification similar in spirit, although more complicated, may also be given for the Associative Rule. Building on the Array Model, one thinks of $n \times (m \times \ell)$ as a rectangular 3-dimensional array of n layers of planar arrays of size $m \times \ell$. One can then re-slice this array by planes perpendicular to the rows of length ℓ, and exhibit the same array as an assemblage of ℓ layers of $n \times m$ planar arrays. Thus

$$n \times (m \times \ell) = (n \times m) \times \ell.$$

We d o not attempt the illustration here. However, students might well benefit from studying it carefully.

Once students firmly believe the Commutative Rule and the Associative Rule for multiplication, the same logic as used in the case of addition shows that the Any Which Way Rule holds also for multiplication. If this was not discussed very explicitly when developing addition, it might be well to consider examples of how the two rules combine to create more extensive rearrangements. Here is one. The equality

$$(a \times b) \times (c \times d) = (a \times c) \times (b \times d)$$

may be justified by the steps

$$(a \times b) \times (c \times d) = a \times (b \times (c \times d)) = a \times ((b \times c) \times d)$$

$$= a \times ((c \times b) \times d) = a \times (b \times (c \times d)) = (a \times c) \times (b \times d).$$

Here the first equation uses the Associative Rule to replace $(a \times b) \times (c \times d)$ with $a \times (b \times (c \times d))$. Note that the product structure of $c \times d$ is being ignored; $c \times d$ is being treated simply as a number. The second equation uses the Associative

Rule for $b \times (c \times d)$. The third equation uses the Commutative Rule for $b \times c$. The fourth and fifth equations reverse the steps of the first two equations, with b and c interchanged.

If the reader thinks that the above transformations amount to a lot of work to justify a simple exchange of factors, s/he would be right; but the general symbolic formulation given above may make the identity seem too transparent. Consider a numerical example:

$$24 \times 56 = (3 \times 8) \times (7 \times 8) = (3 \times 7) \times (8 \times 8) = 21 \times 64.$$

The equality of the first and last products may not be immediately obvious to all.

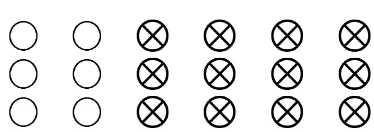
In addition to the Any Which Way Rule, multidigit multiplication makes heavy use of the Distributive Rule, which is the one rule of arithmetic involving both addition and multiplication. More precisely, it tells us how to multiply sums of numbers. The standard formulation is:

Distributive Rule: For any whole numbers a, b and c, the equation

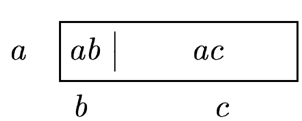
$$a \times (b + c) = a \times b + a \times c$$

holds.

The Distributive Rule may be also justified using the Array Model. The picture is as follows:

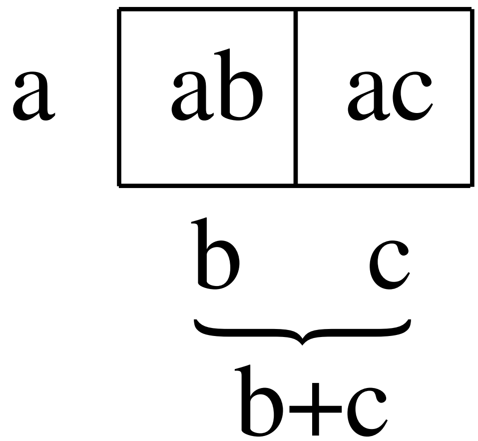


$$3 \times (2 + 4) = 3 \times 6$$



$$a \times (b + c) = a \times b + a \times c$$

Distributive Rule Diagram



Just as the Commutative and Associative Rules combine many times to produce the practical Any Which Way Rule, the Distributive Rule also has a more practical form, which we will designate (rather prosaically) as the

Extended Distributive Rule: If A and B are sums of several numbers, then the product AB may be computed by multiplying each summand of B by each summand of A, and adding all the resulting products.

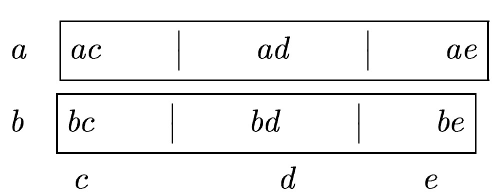
For example, if $A = a + b$ and $B = c + d + e$, then

$$AB \;=\; (a+b)(c+d+e) \;=\; ac+ad+ae$$

$$+bc+bd+be.$$

(Here, and in what follows, we use the standard convention to indicate multiplication simply by juxtaposition, with no operation sign written explicitly.)

The Extended Distributive Rule can be pictured using arrays almost as easily as the basic version:



Extended Distributive Rule Diagram

a	ac	ad	ae
b	bc	bd	be
	c	d	e

These rules, plus of course the Any Which Way Rule for addition, provide excellent guidance for multiplying decimal numbers together. Since each decimal number is the sum of its decimal components, the Extended Distributive Rule tells us that the product of two decimal numbers may be found by multiplying each decimal component of one factor by each decimal component of the other, and then summing all the products. The Any Which Way Rule for addition says that we have great latitude how we sum them. Differences in multiplication algorithms come mainly from choosing different summation procedures.

Before getting into details of summation schemes, we should have a good grasp of multiplying decimal components. Here we find a situation analogous to that of addition. When we multiply two decimal components, the result is equal to the product of the digits, times the appropriate order of magnitude, which is just the sum of the orders of magnitude of the factors, since $10^a \times 10^b = 10^{a+b}$. Examples:

$$20 \times 40 = 800 \qquad 300 \times 6,000 = 1,800,000.$$

Thus, the basic multiplication facts, combined with a grasp of place value, allow us to multiply any two decimal components.

We should perhaps point out that the formal correctness of a formula like $300 \times 6,000 = 1,800,000$ depends on the Any Which Way Rule for multiplication. In carrying out the detailed manipulations, we would write

$$300 \times 6,000 = (3 \times 100) \times (6 \times 1000) = (3 \times 6) \times (100 \times 1000) = 18 \times 100,000$$

$$= 1,800,000.$$

(Note that the crucial middle equality is a case of the equation $(a \times b) \times (c \times d) = (a \times c) \times (b \times d)$ discussed above.) Even the multiplication of decimal units, such as $100 \times 1000 = 100,000$, which is essentially an instance of the Law of Exponents, also relies on the Associative Rule for multiplication for formal justification.

Just as in one-digit multiplication, the product of two decimal components can itself have one or two components in its expanded form. Their magnitudes will be the sums of the magnitudes of the factors, and possibly one more. Perhaps the trickiest case is when one digit is even, and the other is 5. Then the product of digits will be a multiple of 10, and so the full product will have only one decimal component, but its order of magnitude is one larger than the sum of the orders of magnitude of the factors. This case, however, follows the same rules as all other cases: the extra order of magnitude is supplied by the product of the digits. An example:

$$50 \times 800 = (5 \times 10) \times (8 \times 100) = (5 \times 8) \times 1,000 = 40 \times 1,000 = 40,000.$$

Let us combine these observations with the Extended Distributive Rule to do multidigit multiplication. For example, consider

$$\begin{aligned} 21 \times 437 &= (20 + 1) \times (400 + 30 + 7) \\ &= \ 20 \times 400 + 20 \times \ 30 + 20 \times 7 \\ &\quad\ + 1 \times 400 + \ 1 \times \ 30 + \ 1 \times \ 7 \\ &= \ \ 8,000 + 600 + 140 \\ &\qquad + 400 + \ \ 30 + \ \ \ 7 \\ &= \ \ 9,177 \end{aligned}$$

In this computation, we have given no clue as to how the addition of the array of six numbers was performed. One method might be to sum the terms along each row of the array, and then sum the results. This would yield the intermediate sum $8,740 + 437 = 9,177$. We can recognize this sum as the one which appears in the standard calculation

$$\begin{array}{r} 437 \\ \times 21 \\ \hline 437 \\ 8740 \\ \hline 9177 \end{array}$$

(Here we have filled in the often omitted zero in the 8740 in order to make clear the connection with the row sums.)

On the other hand, if we sum down the columns, then we obtain as an intermediate step $8400 + 630 + 147 = 9177$. We can recognize these summands as those appearing in the standard procedure with the factors taken on the other order:

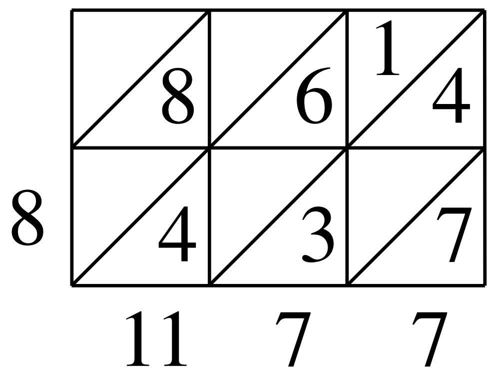
$$\begin{array}{r} 21 \\ \times 437 \\ \hline 147 \\ 630 \\ 8400 \\ \hline 9177 \end{array}$$

Thus from the array of products of decimal components, we can produce the standard method for computing products, in either order, by summing first along rows of the array. or along the columns. Interpreting these procedures in terms of the array of products of components could help students to understand why both give the same answer (as they must if multiplication is to be commutative), although the intermediate steps are so different.

However, we have seen that in some sense the most natural way to do an addition is to break each summand into its decimal components, and to add the components of a given magnitude. Neither of the standard algorithms does this. However, this procedure can be carried out in quite an elegant way, giving an algorithm known as the *array method,* or *lattice method* , or *Napier's Bones.* It refines the observation that, if we write the products of components in an array, as in the example above, then the order of magnitude of the products decreases from left to right along rows, and from top to bottom along columns, and products of the same order of magnitude are lined up in diagonals running from upper right to lower left.

In more detail, to multiply a d digit number and an e digit number, we construct a $d \times e$ array of boxes, with each box divided by a diagonal slanting from lower left to upper right. We arrange the digits of the first factor, from left to right above the boxes of the top row of the array, and we arrange the digits of the second factor, from top to bottom, beside the boxes at the right side of the array. In each box, we record the product of the digits in its row and column, with the 10s digit of the product above the diagonal, and the 1s digit below. Then to find the total coefficient of each power of 10, we sum the digits along the diagonals. The lower rightmost diagonal (which has only 1 box) corresponds to the 1s place, and moving left one diagonal increases order of magnitude by 1. If the sum of the digits along a diagonal exceeds 10, the 10s digits should be carried to the next place (the diagonal to the left). Napier's Bones is a pleasant algorithm for multiplication which better displays the structure of the process than does the standard algorithm. We illustrate it below for our example of 437×21.

Napier's Bones Diagram



$$437 \times 21 = 7+70+1100+8000=9277$$

We remark that the process of summing along the diagonals in Napier's Bones is parallel to shifting the partial products to the left in the standard method.

The above discussion has tried to make the case that thinking explicitly in terms of decimal components yields benefits for developing understanding of arithmetic. This approach is extended to subtraction and division in Appendices 1 and 2. We will next attempt to show that doing decimal arithmetic with emphasis on the role of decimal components has advantages beyond computation. It also makes estimation a relatively simple matter, and it brings out the connections between decimal arithmetic and algebra.

3. Decimal components, estimation and error.

Besides the arithmetic operations, the other main structure on the whole numbers is order: we know when one whole number is larger than another. (In fact, order can be expressed in terms of addition: whole number a is greater than whole number b if $a = b + c$, where c is another (non-zero) whole number. However, the role order plays is quite different from the role of the operations.)

The decimal system is highly compatible with the ordering of the whole numbers, and makes it easy to compare numbers. Recall that, for any whole number, we have defined its order of magnitude to be the order of magnitude of its largest non-zero decimal component. This is one less than the number of digits in its decimal expression. A basic fact is that order of magnitude sorts numbers according to size: <u>any</u> number of a given order of magnitude is larger than <u>any</u> number of a smaller order of magnitude. For example, the largest number of magnitude 2 is 999, and the smallest number of magnitude 3 is 1,000 = 999+1. This simple principle has several important consequences. The first is a recipe for comparing any two decimal numbers.

Ordering Algorithm: To decide which of two decimal numbers a and b is larger, compare their decimal components. Find the largest order of magnitude for for which the decimal components of a and b are different. Then the number with the larger component of that magnitude is larger.

In particular, if the order of magnitude of a is larger than the order of magnitude of b, then a is larger than b. If they have the same order of magnitude, but the leading (i.e., largest) component of a is larger than the leading component of b, then a is larger. If the leading components agree, proceed downward in orders of magnitude, until you first find a magnitude at which the components of a and b differ. Then the larger number is the one with the larger component.

Relative Place Value:

Often, we want to do more than say whether one number is larger than another, we want to say how much larger it is, or that it is very much larger. We also want to say when two numbers are close to one another. Decimal components also make these things easy to do. The main point here is that multiplying by 10 just increases order of magnitude by 1. This is true no matter what decimal place we are talking about. Thus, given any decimal place, the place just to the left represents numbers 10 times as large as the given place, and the place just to the right represents numbers only $\frac{1}{10}$ the size of the given place. Two places to the left, you find numbers 100 times as large as where you are, and two places to the right, the numbers are only $\frac{1}{100}$ of the place where you are. For facility with estimation, it is important that students understand not only the values of each place, but also these relative values of the places. The grasp of relative place value also can support the introduction of decimal fractions: if we make a place to the right of the ones place, relative place value would predict that it should represent $\frac{1}{10}$s; and the next place should represent $\frac{(\frac{1}{10})}{10} = \frac{1}{100}$, which, of course, it does. And so forth. Understanding of relative place value might also serve as a precursor to proportional reasoning in general.

Absolute and Relative Error:

Relative place value ideas help us understand that, in representing numbers, it is the largest few decimal components (the leftmost few decimal places) that account for most of the size of a number. This is true no matter what the magnitude of the number. These ideas are particularly important in dealing with error, which is unavoidable whenever we must deal with "real-life" numbers, meaning numbers which come from measurement. These are always known only approximately, and the question of how accurately they are known is fundamental. Here the key idea is *relative error:* the size of the error involved in approximating a given number, in comparison to the number itself. If you only have a few dollars, you care quite a bit whether it is \$4 or \$6. However, if you have around a thousand dollars, probably you don't care too much if it is \$1004 or \$1006. And if you have a million dollars, even a thousand dollars more or less, let alone \$2, will probably not keep you up at night.

Suppose you need to work with a number V, but you use a possibly different value v for it. (This could be for a variety of reasons: you may not know the precise value of V, or it may simply be more convenient to use v (or both).) For example, the number $\frac{22}{7}$ is frequently used in place of the number π. Call v an *approximation* to V. The (absolute value of) the difference, $|V - v| = e$ is the *absolute error* of approximating V by v. For many purposes, the actual size of the error e is not as important as how it compares to the actual value V. To capture this idea, we define the *relative error* to be the ratio $\frac{e}{V}$. The relative error can be thought of as measuring the error in the most relevant units, namely units of the correct amount.

Both the absolute and relative error are controlled easily in terms of decimal components. The main observation is that, if two numbers are close to each other, their decimal expressions should tend to be similar. (There is an exception to this, related to the rollover phenomenon discussed in Appendix 1, but we will ignore this in the present discussion.) In any case, the converse is definitely true: the more alike the decimal expressions (read from the left) of two numbers are, the closer together the numbers are. To be precise about this, suppose that we have two numbers V and v, both of order of magnitude m, and suppose that their largest decimal components are equal, or the two largest ones, or perhaps more. There are two ways of formulating this: we can specify the number of decimal components which agree, or we can specify the order of magnitude of the largest decimal components which differ. It is easy to convert one type of information into the other, but we distinguish them because they reflect the two points of view - relative versus ansolute error. In any case, the following simple statement holds.

Basic Decimal Estimation Theorem (BDET): Suppose that V and v are two numbers of magnitude m, and suppose their decimal components of magnitude larger than ℓ are equal. (That is, the largest $m - \ell$ decimal components of V are equal to the corresponding components of v.) Then:

i) The error $e = |V - v|$ has order of magnitude at most ℓ, and hence is less than $10^{\ell+1}$.

ii) The relative error $\frac{e}{V}$ is less than $\frac{1}{10^{m-\ell-1}}$.

For example, take $V = 1,081$ and $v = 1,919$. These numbers have order of magnitude 3, and both have 1,000 for the magnitude 3 decimal component. Thus, in this example $m = 3$ and $\ell = 2$. Their difference is $e = |V - v| = 838$. We see that it has magnitude 2, so that e is less than 1000, as predicted by the BDET. Since both numbers are at least equal to 1000, the relative error $\frac{e}{V}$ must be less than 1. (The exact value is $\frac{838}{1081}$, which is between .7 and .8.)

Here is a larger example. Take $V = 4,825,619$ and $v = 4,863,781$. In this case, $m = 6$ and $\ell = 4$. We calculate that $e = |V - v| = 38,162$. This is less than 100,000 $= 10^{\ell+1}$, while V and v are both larger than $1,000,000$, so the relative error $\frac{e}{V}$ is less than $\frac{100,000}{1,000.000} = \frac{1}{10}$. (In fact, in this example, $\frac{e}{V}$ is less than $\frac{1}{100}$. The estimate in the BDET will always be larger than the actual error, since it must allow for the worst case. Note that the statement allows the possibility that $V = v$, in which case both the absolute and relative error will be zero.)

The BDET expresses in a concise way the sense in which knowing the beginning (the largest several decimal components) of the decimal expression of a number tells us most of what we want to know about the number as a magnitude. If we are number theorists, it may be important to know all the decimal components of a number, in order to tell, for example, if it is a perfect square, or if it is divisible by 7, or if it is prime. But if it is a number coming from a measurement, then all we need to know, or can expect to know, is an approximate value, and the largest decimal components supply this with great efficiency. For example, if we know the largest four decimal components of a number, we know it with a relative error of less than $\frac{1}{1000}$.

In fact, it is rather rare to know a measured number with such accuracy. Take the example of the "radius" of Earth. It is approximately 4,000 miles, and sometimes you will see figures such as 3,928 miles. However, the last digit does not have a clear meaning. In speaking of the radius of Earth, we are pretending that Earth is a perfect sphere. However, it is not. It deviates from being a perfect sphere in three ways: first, because of its rotation, it is slightly oblate – flattened at the poles, and thickened around the equator; second, it has bumps and dimples - Mount Everest and the Challenger Deep; third, because of the motion of its liquid interior, it is slightly deformed, with a bulge in the north Pacific. These imperfections mean that it does not make sense to speak of a "radius" of Earth more accurately than about 10 miles. Thus, the radius of the Earth is a number defined only to 3 significant figures. Most other "real-life" numbers have similar limitations. The results of polls are typically accurate only to about $\pm 3\%$, which is slightly better than one significant figure. Although the U.S. Census reports state populations as exact numbers of people, in the millions (6 significant figures) or the 10s of millions (7 significant figures), it is lucky if these numbers are accurate to 3 figures.

In summary, the notion of significant figures successfully captures many of the key notions of error and approximation for decimal numbers. Especially, it provides an easy way to deal with relative error. For many purposes, it suffices to know a number to one significant figure. For most purposes, two significant figures are enough, and it is rare to know a "real-life" number to more than 3 significant figures. (Financial transactions can be exceptions to these rules; but it is also debatable how "real-life" the numbers involved in them are.)

Scientific notation:

In the realm of measurement, especially scientific measurement, the decimal components (starting from the largest) which are definitely known are called the *significant digits,* and the number $m-\ell+1$ in BDET is frequently called the *number of significant digits.* (Sometimes, a slightly looser notion of significant digits is used.) Scientists and others who use numbers resulting from measurements are mainly interested, first, in their size, and second, in the accuracy with which they are known. The way of writing numbers known as *scientific notation* is designed to exhibit these two points very clearly.

Formally, scientific notation is simply a variant way of writing decimal fractions. Instead of writing a decimal fraction d in the conventional way, with a certain number of places to the left of the decimal point, and a certain number to the right, scientific notation rewrites it in the form

$$d = (\frac{d}{10^m}) \times 10^m,$$

where m is the order of magnitude of d, so that $\frac{d}{10^m}$ is between 1 and 10. (A slight variant, used in IBM computers, writes $d = (\frac{d}{10^{m+1}}) \times 10^{m+1}$.)

In this form, scientific notation separates out a key feature of a number, namely its magnitude, and displays it prominently. A further convention frequently used by scientists (although it is at odds with standard mathematical notation) is to report only significant digits when recording a number in scientific notation, in order to be precise about the accuracy to which the number is known. By this convention, 2.3×104 would means some number known to be between 22,500 and 23,500, while 2.30×104 means some number known to be between 22950 and 23050. With this understanding, scientific notation uses exponents and the idea of significant digits to express in a very direct way the key aspects of a measured number. The BDET tells us the maximum relative error possible in a number with a given number of significant digits. A refinement of BDET tells the maximum possible relative error in a number with given significant digits. See the Exercises for a statement of this sharper result.

4. Decimal components and estimation in arithmetic.

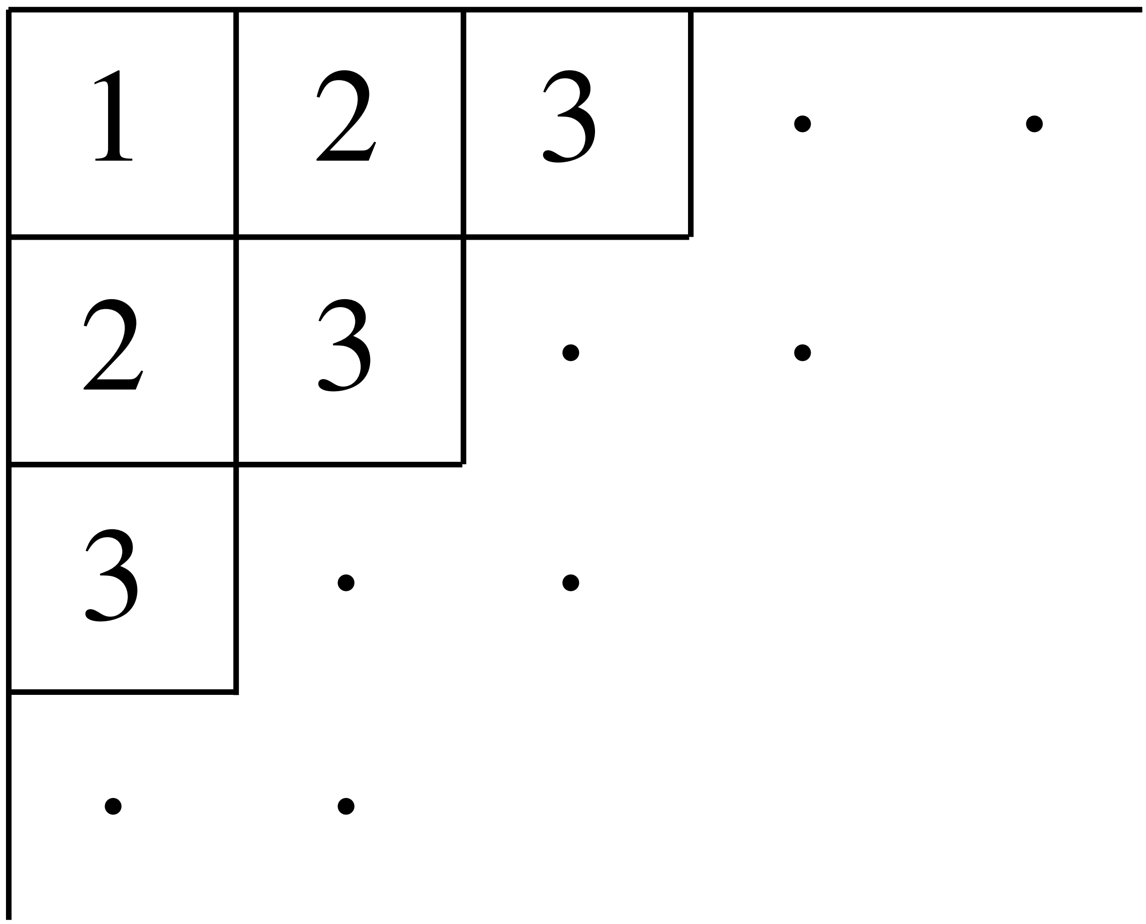
Thinking in terms of relative place value also provides efficient ways of estimating sums and products. We will avoid exact details here, to keep the discussion short, but the principle is clear: when estimating, focus on the largest decimal components. With our understanding of how decimal components enter into arithmetic, we can describe what to do to get an estimate with desired accuracy for a sum or product. For a sum, simply retain the sums of the largest decimal components. The three largest components of the larger summand, and the components of the same magnitudes for the other summand, will give an approximation to the sum with relative error small enough for most purposes. For example,

$$83,244 + 5,293 \simeq 83,200 + 5,200 \simeq 88,400.$$

Here the exact sum is 88, 537, so the relative error is $\frac{|88{,}537-88{,}400|}{88{,}537} = \frac{137}{88{,}537} < \frac{1}{600}$.

For multiplication, we recall that a product of decimal numbers is the sum of all the products, of any component of one factor times any component of the other factor. Of these summands, the largest will be the product of the largest decimal components of each factor. The next largest will be the two terms gotten by taking the product of the second largest component of one factor, with the largest component of the other. Thus, in the product 437×21, the largest contribution is the $8{,}000 = 400\times20$. The two next largest summands are $30\times20 = 600$ and $400\times1 = 400$.

The diagram shows the location of these terms in the array of the Napier's Bones algorithm. The largest term is labeled with a 1, the next two with 2s. The next largest terms, three in number, are labeled with 3s, the next (four) terms with 4s. The pattern should be evident. The sum of the largest six terms will give an estimate sufficiently accurate for many purposes.



When more accuracy is needed, a straightforward continuation of the above procedure prescribes the next terms to add. The upshot is a method to get a (relatively) very accurate approximation to a multidigit multiplication with fairly little work.

Caveat: because of the rollover phenomenon (see Appendix 1), we cannot guarantee that the approximate answers we compute by the above methods will agree in <u>any</u> decimal places with the true answer. However, we can show that they do approximate well the true answer, in the sense that they produce a small relative error. Usually, in a sense that can be made precise, they will have the same leading decimal components as the exact answer.

5. Decimal components and algebra.

Discussing arithmetic in terms of decimal components should help prepare students for algebra in several ways. One way is that it would create increased awareness of and facility with the Rules of Arithmetic, which are also the rules for manipulating algebraic expressions. (And essentially, they are the <u>only</u> rules, so they provide a remarkably compact summary of what constitutes legitimate algebraic manipulation.) Also, the Rules themselves are most succinctly expressed via symbolic equations, so they would give students exposure to the ideas of literal symbolism and variables. In these ways, this approach should promote increased confidence and facility with algebraic manipulation.

Another way in which working with decimal components would prepare students for algebra is, it would bring out the deep analogy between decimal arithmetic and algebra. As we remarked at the outset, decimal notation exploits algebra in the service of arithmetic. Decimal numbers can be usefully thought of as "polynomials in 10". Emphasizing this connection should shed light on arithmetic, and also make algebra more familiar and learnable.

Here are some examples of the analogy. Consider the numbers 21 and 13. We can compute

$$21 + 13 = 34, \quad \text{and} \quad 21 \times 13 = 273.$$

Now consider the expressions $2x + 1$ and $x + 3$. We can also compute, using the same rules as we did for regular arithmetic, that

$$(2x + 1) + (x + 3) = 3x + 4, \quad \text{and} \quad (2x + 1) \times (x + 3) = 2x2 + 7x + 3.$$

If students do some calculations like these, they should see parallels between the calculations with decimal numbers and the calculations with polynomials. Discussion could bring out the key to the analogy: that if we set $x = 10$, then the polynomial computations become the arithmetic computations. Further investigation will show that the analogy is imperfect. For example $(2x + 4) + (3x + 7) = 5x + 11$, whereas $24 + 37 = 61$. Nevertheless, if we plug $x = 10$ into $5x + 11$, we do get $50 + 11 = 61$ as the value. Students who are used to thinking in terms of decimal components should recognize $50 + 11$ as the intermediate result we obtain in adding $24 + 37$, before we combine ten 1s into a 10 to get the standard decimal form of $50 + 11$. In other words, polynomial arithmetic is decimal arithmetic without the regrouping process (and hence, with arbitrarily large coefficients).

In this analogy, the order of magnitude of decimal numbers translates into the *degree* of a polynomial: the largest power of x which appears in the polynomial with a non-zero coefficient. Because of the absence of carrying in polynomial arithmetic, degree is in several ways better behaved than order of magnitude. For example, the degree of the sum of two polynomials is always less than or equal to the larger of their degrees. The degree of a product of two non-zero polynomials is (exactly) the sum of the degrees of the two factors.

Just as order of magnitude guides us in dividing decimal numbers, degree is what we use to find the quotient of two polynomials. The recursive procedure for finding the quotient of two polynomials is quite analogous to the one (described in Appendix 2) for long division of decimals. Given two polynomials $a(x)$ and $b(x)$, the highest degree term of the quotient $\frac{a(x)}{b(x)}$ is the monomial $c_1 x^m$ such that

$a(x) - c_1 x^m b(x) = a_1(x)$ has degree lower than $a(x)$. This term is quite easy to find: it is exactly the quotient of the highest degree terms of $a(x)$ and $b(x)$. To find the next term in the quotient, one uses the same procedure for $a_1(x)$ and $b(x)$. This continues until one gets a difference of degree less than the degree of $b(x)$; this is then the remainder.

The connection between polynomial arithmetic and ordinary arithmetic can be developed and exploited in other ways. For example, the famous identity $x2 - y2 = (x + y)(x - y)$ has many applications in arithmetic. One very serious contemporary application is its use as a basis for algorithms to factor large numbers. On a more mundane level, it can be used to do mental math. Take the product 43×47. This can be computed mentally by thinking of $43 = 45 - 2$ and $47 = 45 + 2$. Hence $43 \times 47 = (45 + 2) \times (45$ - $2) = 452 - 22$. To compute 452, we can use the same trick: Write $40 = 45 - 5$ and $50 = 45 + 5$. Therefore, $452 = 40 \times 50 + 52 = 2000 + 25$, and so, $43 \times 47 = 40 \times 50 + 25 - 4 = 2021$. Similar tricks can be used to develop mental math skills, and to find factorizations of fairly large numbers whose prime factors also are fairly large (i.e., no factors less than 20). In order not to further lengthen this essay, we will curtail further examples. However, a variety of calculations revealing parallels between decimal arithmetic and algebra are given in the Exercises.

Regarding decimal numbers as "polynomials in 10" amounts to giving the variable x the value 10. This is called *specialization.* Specialization can be used to interpret other number systems in terms of polynomials. The specialization which turns polynomial arithmetic into decimal arithmetic amounts to requiring x to satisfy the equation $x - 10 = 0$. Polynomials can be made to mirror other number systems by positing that x satisfies other equations. For example, requiring $x2 - 2 = 0$ would make polynomials act like numbers of the form $a + b\sqrt{2}$ – in other words, x is imitating the irrational number $\sqrt{2}$. Requiring x to satisfy $x2 + 1 = 0$ will essentially create a copy of the complex numbers.

Here we are getting into fairly advanced ideas, drawing close to abstract algebra. It is questionable whether the constructions of the last paragraph could be productively inserted into the K-12 curriculum. However, it does seem desirable that high school teachers should be conversant with these issues. Whether or not they get used in school mathematics, these possibilities demonstrate the purchase that place value has throughout the K-12 curriculum, and point to the desirability of treating it more explicitly, and making it one of the "big ideas" that help students arrange the mathematics they learn into a coherent whole.

Appendix 1: Subtraction

Thinking in terms of components can also help with subtraction.

A key principle in dealing with subtraction is to link it with addition at every opportunity, until students think of it as the undoing of addition. This is important conceptually, to prepare for the submersion of subtraction into addition (i.e., as addition of the additive inverse) when negative numbers are introduced. However, it also has practical value for making sense of subtraction procedures, by linking them to the corresponding addition procedures. This linkage also brings more insight into addition.

Fact families should not be limited to addition of one-digit numbers. Rather, every subtraction problem $c - b = a$ should be linked to the corresponding addition problem $a + b = c$. In fact, the details of the computations involved in the two problems correspond closely. The subtraction $c - b = a$ will involve regrouping (borrowing) exactly when the addition $a + b = c$ involves regrouping (carrying). Indeed, the same orders of magnitude will be affected in both computations.

For example, in the problems

$$\begin{array}{r} 970,957 \\ +\quad 50,829 \\ \hline 1,021,786 \end{array} \qquad\qquad \begin{array}{r} 1,021,786 \\ -\quad 50,829 \\ \hline 970,957 \end{array}$$

we see that the addition involves carrying from the 1s to the 10s, and from the 100s to the 1000s, and that a carry from the 10,000s to the 100,000s causes a further carry to the 1,000,000s. Correspondingly, in the subtraction, there is a borrowing from the 10s to the 1s, another borrowing from the 1,000s to the 100s, and a "borrowing across zero", from the 1,000,000s to the 100,000s and 10,000s.

As this example indicates, study of the most troublesome aspect of subtraction, namely "borrowing across a zero", reveals an interesting parallel feature of addition that may go unnoticed in a purely mechanical approach.

In adding two decimal numbers, when the sum of the digits of a given order of magnitude add to 9, that place would not produce a carry. However, if the next smaller place does produce a carry, then the 1 from that, added to the 9 already there, will produce a carry. If the digit sum for the next larger order of magnitude is also 9, then the carry caused by the carry from the lower place will in turn cause a carry at the next larger place. This could continue for several places, producing a sum in which there are one or more zeroes in a row. We call this phenomenon *rollover*. Its extreme form occurs in adding 1 to a number like 99 or 999 or 9999, with all digits equal to 9. Then the result, of course, is the next larger decimal unit. This is what happened on car odometers when they were mechanical, and one could see the change of the 1s place producing a change in the 10s place, which produced a change in the 100s place, and so forth, until many or even all the digits on the odometer had rolled over.

Borrowing past a zero is the parallel for subtraction to rollover in addition: when an addition $a + b = c$ involves rollover, the corresponding subtraction $c - b = a$ will require borrowing past a zero, and vice versa. If rollover is explicitly studied as an interesting and exceptional situation in doing addition, and if subtraction is consistently connected to addition, then borrowing past a zero may seem less mysterious to students. Here is an example of rollover in an addition, and the corresponding borrowing across zeroes in the associated subtraction:

$$\begin{array}{r} 35,375 \\ \underline{+24,648} \\ 60,023 \end{array} \qquad\qquad \begin{array}{r} 60,023 \\ \underline{-24,648} \\ 35,375 \end{array}$$

We see that the carry from 10s to 100s in the addition causes two places to roll over: it causes the 100s to carry to the 1000s, and that makes the 1000s carry to the 10,000s. In the subtraction, in order to do the subtraction at the 10s place, one must borrow from the 10,000s place, across both the 1000s and 100s.

Placing emphasis on decimal components also suggests alternative procedures to the standard one for subtraction. As is well-known, there are multiple interpretations for subtraction; take away, difference, and comparison. Comparison involves a process of matching elements in two sets, followed by counting of the unmatched elements of the larger set after all elements of the smaller set have been used up. Thinking in this way, one can see that, if the same amount is subtracted from two numbers, the difference of the results is the same as the difference of the original numbers. In algebraic terms, this is expressed by the identity

$$a - b = (a - c) - (b - c).$$

This principle may be applied to simplify subtraction problems as follows: Given a subtraction $a - b$ of decimal numbers, for each order of magnitude, one subtracts from both numbers the smaller of the components of a and b of that magnitude, leaving only one of the numbers with a non-zero component at that place. Doing this systematically leaves one with an equivalent but simpler problem. For example

$$1,021,786 \qquad \rightarrow \qquad 1,001,060$$

$$\underline{-\ 50{,}829} \qquad\qquad \underline{-\ 30{,}103}$$

$$970{,}957$$

and

$$\begin{array}{rcr} 60{,}\ 023 & \rightarrow & 40{,}000 \\ \underline{-\,24{,}648} & & \underline{-4{,}625} \\ & & 35{,}375 \end{array}$$

As one sees from these examples, the simplified problem decomposes into a disoint collection of "making change" problems, in which one is subtracting a given number from a decimal unit. This suggests the possibility of an approach to subtraction in which this kind of problem is a subject of focussed attention, and after it is well understood, it is applied to the general subtraction problem.

Appendix 2: Division

Decimal components also help one understand what is going on in division of decimal numbers. Division, in the strict sense of inverting a multiplication, cannot in general be carried out in the whole numbers, or even in the system of decimal fractions (finite decimals). This is, indeed, one of the least satisfactory aspects of decimal arithmetic, and is a major reason for using fractions, i.e., general rational numbers.

Instead of division as the inverse of multiplication, one deals with a somewhat looser substitute, *division-with-remainder.* This can always be carried out with any two whole numbers, but it does not yield a single number as a result – it yields a pair of numbers: the quotient, and the remainder. Division-with-remainder, both theoretically and practically, depends on the order properties of whole numbers, as discussed in section 4.

We will give a brief description of the basic ingredients in long division. Our treatment is somewhat abstract, and uses symbolism which might not be appropriate for students being introduced to long division. However, the principle governing the procedure should be both teachable and plausible, and examples should make it convincing. The theoretical justification might be delayed until an algebra course.

The usual recipe for division-with-remainder (often called the *division algorithm,* although it gives little guidance about how one would do division in practice) depends on the order structure of the whole numbers. It is contained in the formula

$$b = aq + r.$$

Here b is being divided by a. The number q is called the *quotient* of b by a. The number r is the *remainder.* The essential property of r is that $0 \leq r < a$. Since r is non-negative, it follows that $aq \leq b$. Secondly, it also follows that q is the <u>largest</u> number such that $aq \leq b$. For if we increase q by 1, we see that $b+(a-r) = (q+1)a$, and the inequality $r < a$ means that $a - r > 0$, so that $(q + 1)a > b$.

Thus, in "dividing" b by a, one is seeking the largest number q such that $aq \leq b$. This rather bare-bones description actually adapts rather directly to what one does to find the decimal description of q, when a and b are given as decimal numbers.

Basic Fact of Long Division (BFLD): Suppose that q is the quotient of b divided by a, so that $b = aq + r$, with $0 \leq r < a$. Then the largest decimal component of q is simply the largest decimal component q_1 such that $aq_1 \leq b$.

Here is a brief justification for this statement. The problem of long division is to find the decimal components of q. The BFLD says, simply look for the largest decimal component q_1 such that $aq_1 \leq b$, and this will turn out to be the largest component of q. (Thanks to the Ordering Algorithm given in section 4, finding

q_1 a fairly easy thing to do in practice.) Suppose that $q_1 = d \times u_1$, where u_1 is a denomination and d is a digit. The key fact needed here about the family of all decimal components is that the next decimal component larger than q_1 is $q_1 + u_1 = (d+1) \times u_1$. This is true even if $d = 9$. By choice of q_1, it follows that $a(q_1 + u_1) = aq_1 + au_1 > b$. In other words, $au_1 > b - aq_1$. Therefore, the quotient of $b - aq_1$ by a is less than u_1. This is the same as saying that it has order of magnitude less than the magnitude of q_1. Hence, if we do division-with-remainder for $b - aq_1$ and a, we will get $b - aq_1 = aq' + r$, with q' of smaller order of magnitude than q_1. Hence also, $b = aq_1 + aq' + r = a(q_1 + q') + r$. This says that the quotient of b by a is $q = q_1 + q'$. Since q' has order of magnitude less than q_1, we see that q_1 is the largest decimal component of q.

The full process of long division procedes by iteration, using the BFLD repeatedly. Having determined the largest decimal component of q, we can now work with the difference $b - aq_1$, and look for the largest decimal compnent q_2 such that $aq_2 \leq b - aq_1$. Continuing in this fashion, we will eventually find the complete decimal expansion of the quotient. This procedure is in fact essentially the standard long division algorithm. Thanks to the structure of the family of decimal components, the connection between the theoretical construct and the practical procedure is quite close.

References

[BP] S. Baldridge and T. Parker, *Elementary Mathematics for Teachers*, Septon Ash Publishers, 2004.

[KSF] J. Kilpatrick, J.Swafford and B.Findell, eds., *Adding It Up: Helping Children Learn Mathematics*, National Academy Press, Washington, DC, 2001.

[Ma] L.Ma, *Knowing and Teaching Elementary Mathematics*, Lawrence Erlbaum Associates, Mahwah, NJ, 1999.