Writing Numbers in Base 3, the Hard Way

GARY E. MICHALEK
La Salle University
Philadelphia, PA 19141-1199

Introduction  An important feature of our system of writing numbers is the significance attached to the position of the numerals. (The Romans had it almost, but not quite, right.) When we write 563, for example, we really mean

$$563 = 5(10^2) + 6(10^1) + 3(10^0),$$

where each numeral is the coefficient of the power of 10 corresponding to its position.
This is the base 10 system of expressing numbers. Similarly, in base 2, nonnegative integers are expressed using the numbers 0 and 1 as coefficients of powers of 2. For example, 27 is written as 11011, since

$$27 = 1(2^4) + 1(2^3) + 0(2^2) + 1(2^1) + 1(2^0).$$

Negative integers can be expressed in a similar fashion, if we allow the use of $-1$ as a coefficient. That is, every integer can be expressed in base 2 using the set $\{-1, 0, 1\}$ as a coefficient system. Do we need to use the set $\{-1, 0, 1\}$, or might another set of numbers do as well? This question was considered in a more complicated context by the Hungarian number theorist Imre Kátai, and most of the following results are implicit in his work [1]. The question is interesting in its own right, and the answers provide nice illustrations of modular arithmetic and suggest some interesting exercises in computer programming.
We consider base 3, which provides richer examples than base 2. Let \( A \) denote a set of integers. If an integer \( m \) can be written in the form

\[
m = c_n(3^n) + c_{n-1}(3^{n-1}) + \cdots + c_1(3^1) + c_0(3^0)
\]

for some nonnegative integer \( n \), with each coefficient \( c_i \) in \( A \), we say that \( m \) is expressible in \( A \). The set \( A \) is a coefficient system if every integer is expressible in \( A \).

There are some obvious restrictions on the sets that can constitute coefficient systems. For instance, if \( m \) is expressed in the form above, \( m \) and \( c_0 \) must be congruent modulo 3. Therefore, for every integer to be expressible, \( A \) must contain at least one representative from each congruence class. For the moment, consider sets of exactly three elements, say \( A = \{a_0, a_1, a_2\} \), where \( a_i \equiv i \pmod{3} \). A coefficient system of minimal size (three elements, for base 3) is called a number system. For simplicity, in this note we will assume a number system has \( a_0 = 0 \). The following observations shorten the list of possible base three number systems:

(i) \( a_1 \) and \( a_2 \) must have opposite signs so that both positive and negative numbers are expressible;

(ii) If \( a_1 \) and \( a_2 \) have a common divisor \( k \), then every number expressible in \( A \) is divisible by \( k \). Hence \( \gcd(a_1, a_2) = 1 \) for a number system.

The remaining candidates for number systems are sets of the form \( \{0, a_1, a_2\} \), where \( a_1 \) and \( a_2 \) are relatively prime, of opposite signs, and congruent to 1 and 2 modulo 3, respectively. To gain further insight into which triples form number systems, we consider how to find the expression for a number (if such an expression exists).

**Expressing a number** Given \( A = \{a_0 = 0, a_1, a_2\} \), define a function \( F_A : \mathbb{Z} \to \mathbb{Z} \) by \( F_A(m) = (m - a_i)/3 \), where \( a_i \) is the element of \( A \) congruent to \( m \) modulo 3. Consider, for instance, the expression for 10 in \( A = \{0, -11, 2\} \):

\[
10 = 2(3^3) + 0(3^2) + -11(3^1) + -11
\]

Repeated application of \( F_A \) gives

\[
F_A(10) = \frac{10 - 11}{3} = 7; \quad F_A^2(10) = \frac{7 - 11}{3} = 6
\]

\[
F_A^3(10) = \frac{6 - 0}{3} = 2; \quad F_A^4(10) = \frac{2 - 2}{3} = 0.
\]

At each stage, \( F_A \) subtracts the appropriate member of \( A \) and divides by 3. These appropriate numbers are, of course, precisely the coefficients in the expression for 10. If \( m \) is expressible in \( A \), then clearly \( F_A^n(m) = 0 \) for some positive integer \( n \). Conversely, if \( F_A^n(m) = 0 \), we can find the expression for \( m \) by keeping track of which element of \( A \) is subtracted at each stage. In short, \( m \) is expressible in \( A \) if and only if \( F_A^n(m) = 0 \) for some positive integer \( n \).

**Number system or not?** How can we check whether all numbers are expressible, as we must do to decide whether \( A \) is a number system? The key is a fortuitous property of the function \( F_A \), which is easily established by checking the inequalities involved. As \( F_A \) is applied repeatedly, the results decrease in absolute value, and eventually the result is in the interval \( I_A = [-M, M] \), where \( M \) is the integer part of \( \max\{|a_1|/2, |a_2|/2\} \). With \( A = \{0, -11, 2\} \), where \( I_A = [-5, 5] \), we have

\[
F_A(100) = 37, \quad F_A(37) = 16, \quad F_A(16) = 9, \quad F_A(9) = 3.
\]
Furthermore, $F_A(I_A) \subset I_A$, so one can think of the interval as a black hole. Once inside the interval, only two things can happen as you continue to apply $F_A$: Either 0 is eventually reached (in which case the number you began with is expressible) or numbers begin to repeat (in which case 0 will never be reached and the number is not expressible). In the case above, one can continue to apply $F_A$ to see that 100 is not expressible. These repeating nonzero numbers (those $m$ for which $F^k(m) = m$ for some positive $k$) are called periodic numbers. The existence of periodic numbers makes clear that $A$ cannot be a number system.

If periodic numbers exist, they must lie in the “black hole interval” for $F_A$. We therefore have a way to check (in a finite amount of time!) whether $A$ is a number system: Display the action of $F_A$ on $I_A$ in a directed graph, where an integer $m$ is connected by an arrow to $F_A(m)$. The directed graph for $A = \{0, -11, 2\}$ is shown in Figure 1. The presence of the nonzero periodic numbers 1, 4, 5, and −1 demonstrates that this $A$ is not a number system.

![Figure 1](image1.png)

**Figure 1** Directed graph for $\{0, -11, 2\}$

For $A = \{0, -11, 2\}$, only −5 and 2 are expressible, as well as any numbers outside $I_A$ that connect to −5 and 2. Contrast this with the directed graph of an actual number system, $A = \{0, -11, 5\}$, shown in Figure 2. Here, all integers in $I_A = [-5, 5]$, and therefore all integers, connect to 0.

![Figure 2](image2.png)

**Figure 2** Directed graph for $\{0, -11, 5\}$

Note that there are no number systems in base 2: If $a \in A$, then $-a$ will not be expressible, since

$$F_A(-a) = \frac{-a - a}{2} = -a.$$  

**New criteria and a theorem** Our analysis reveals some additional requirements for a base three number system:

(iii) A number system cannot contain an even nonzero number, for if $a$ is even and $a \in A$, then $F_A(-a/2) = -a/2$, so $-a/2$ will not be expressible in $A$.

(iv) A number system cannot contain the number 13. If 13 $\in A$, we have $F_A^3(-5) = -5$, as the reader may verify. (Notice, in this computation the other nonzero number in $A$ doesn’t come into play. This phenomenon is similar to the even number problem. Even numbers and 13 carry their “own loops” with them and so cannot belong to a number system. Can you find other numbers that have their “own loops”?)

(v) The set $A$ will not be a number system if both nonzero numbers in $A$ lie outside $I_A$. If, say, $A = \{0, 13, -19\}$ with $I_A = [-9, 9]$, then the only expressible number in $I_A$ is 0: The only nonzero numbers $m$ with $F_A(m) = 0$ are 13 and −19.
However, the nonzero numbers in $I_A$ cannot connect to 13 and $-19$, and so will not connect to 0 in the directed graph.

Must a set that meets all the requirements (i)–(v) be a number system? The answer is no. Consider $A = \{0, 7, -25\}$, for which 4 is a periodic number. In general, no quick inspection of the members of $A$ will determine whether $A$ is a number system, and the directed graph must be drawn to determine whether periodic numbers exist.

It may also be interesting to examine a set that fails to meet the listed requirements in an extreme way. Consider $A = \{0, 7, -7\}$, where both nonzero numbers lie outside $I_A$, and are certainly not relatively prime. The directed graph here has an interesting shape. It is a good exercise in modular arithmetic to show that repeated application of $F_A$ connects an integer $m$ to an element of $I_A$ that is congruent to $m$ modulo 7.

One positive result of observations (i)–(v) relates to the earlier definition of a coefficient system. Here the set $A$ is not necessarily minimal in size, and, unlike the case for number systems, the expression for a number need not be unique. Consider two arbitrarily chosen positive integers $a$ and $b$, with $a < b$, where one of the numbers is congruent to 1 and the other is congruent to 2 modulo 3. We know $A = \{0, a, b\}$ is not a coefficient system, but we can expand the set in a natural way to make it so. If we consider instead the set $A - A$, consisting of all differences between elements of $A$, the obstacles presented by remarks (i) and (v) are removed. Since $A - A$ contains $-a$ and $-b$, we have the necessary mixture of signs. Furthermore, some number must lie in $I_A$, since either $b - a$ or $a$ is less than $b/2$. In fact, we have the following nice theorem, conjectured by Kátai after extensive testing on a computer. (It is an interesting exercise to write a computer program to check whether a given set is a coefficient system.) The proof of the theorem, rather long but using only simple algebraic techniques, appears in [2].

**Theorem.** Let $A = \{0, a, b\}$, where $a$ and $b$ are congruent to 1 and 2 modulo 3. Then $A - A = \{0, a, b, -a, -b, b - a, a - b\}$ is a coefficient system if and only if $\gcd(a, b) = 1$.

For example, $A = \{0, 7, -25\}$ is not a coefficient system, but the expanded set $A - A = \{0, 7, -25, -7, 25, 32, -32\}$ is, since $\gcd(7, -25) = 1$. A theorem recently proved by the author [3] generalizes this result to other bases:

**Theorem.** Let $A = \{a_0 = 0, a_1, a_2, \ldots, a_{n-1}\}$ be a set of representatives of the congruence classes modulo $n$, with $n > 1$. Then $A - A$ is a coefficient system in base $n$ if and only if $\gcd(a_1, a_2, \ldots, a_{n-1}) = 1$.

REFERENCES