

Encoding sequences of numbers as a single number

Incompleteness
Spring, 2008
Kent Johnson

By *number*, I will always mean natural numbers (i.e., elements of ω). An extremely fundamental and powerful fact about the natural numbers is that *each finite sequence of numbers can be (effectively) associated with a unique natural number*. Now, the reverse direction is obvious. The natural numbers themselves are lurking in the set of finite sequences of natural numbers: we can effectively associate each $x \in \omega$ with $\langle x \rangle \in \omega^*$.¹ But now we're going to show that we can effectively associate each $x \in \omega^*$ with a unique $n \in \omega$. This is significant, because it allows us to take any finite sequence of numbers, e.g.,

(*) $\langle 11, 66, 896, 57, 0, 0 \rangle$.

and (effectively) transform it into a single number. Thus, we can use individual numbers as “codes” for finite sequences of numbers, and vice-versa. This fact will later prove key to proving one of the most remarkable theorems of mathematics. Independently of this course, it is an interesting and highly useful fact.

There are a variety of ways to accomplish the goal stated above. Our method is the most natural in some sense, and it displays a very elementary feature of the natural numbers. Most of our proof lies in establishing the “Fundamental Theorem of Arithmetic”.

The Fundamental Theorem of Arithmetic (FTA): For any $a > 1$, there exists a unique (ordered) sequence p_1, \dots, p_n of prime numbers such that $a = p_1 \cdot \dots \cdot p_n$, where $p_i \leq p_j$ whenever $i \leq j$.

Proof. Let a be any number. We first show that a is the product of *at least* one ordered sequence of primes. We do this by a simple (strong) induction. *Base.* Let $a = 2$. then a is the product of the unary sequence $\langle 2 \rangle$, since 2 is a prime number. *Induction.* Suppose that the theorem holds for every number less than a . Now notice that a is either prime or composite. If a is prime, then it is the product of $\langle a \rangle$. If a is composite, then there are numbers b and c (both smaller than a) such

¹ Pop quiz #1: Thus, it follows by _____ (from set theory), that $|\omega^*| = \omega$.

that $a = bc$. By the induction hypothesis, $b = p_1 \cdots p_n$ and $c = q_1 \cdots q_m$, where these two new terms are ordered sequences of primes. Thus, $a = s_1 \cdots s_{n+m}$, where this last term is created by organizing the p s and q s into an ordered sequence. Thus, the theorem holds for a , and by induction, this half of the theorem is proved.

The second half of the proof employs some preliminaries.

Definition. We say that a *divides* b iff $b = ac$, for some number c .

Definition. a and b are *relatively prime* iff 1 is the only number that divides both a and b .

Definition. a is a *divisor* of b iff a divides b and $a \neq 1$.

Observation. If a and b are relatively prime and $a \neq 1$, then a does not divide b .

Proof. Assume $1 < a < b$ (otherwise the claim trivially holds). Since a divides itself, if a also divides b , then a and b are not relatively prime. ◻

Lemma 1. If a_1, \dots, a_n are each relatively prime to b , then the product $a_1 \cdots a_n$ is also relatively prime to b .

Proof. By hypothesis, no divisor of any of the a_i s (which will also be a divisor of the product $a_1 \cdots a_n$) can be a divisor of b . The only other divisors of $a_1 \cdots a_n$ are products $c_1 \cdots c_k$ of sets of divisors of the various a_i s. But if $c_1 \cdots c_k$ divides b , we would have $b = c_1 \cdots c_k \cdot d = c_1(c_2 \cdots c_k \cdot d)$, and so c_1 would divide b , which is impossible. ◻

Lemma 2. Let p be some prime number. Then if p divides the product $a_1 \cdots a_n$, then p divides a_i for some $1 \leq i \leq n$.

Proof. We will prove the contrapositive form. Suppose that p is prime but doesn't divide any a_i . Since p is prime, it has no divisors (besides itself). Thus, p is relatively prime to all of a_1, \dots, a_n . So by lemma 1, it follows that the product $a_1 \cdots a_n$ is also relatively prime to p . Since these two are relatively prime, by the observation, p does not divide the product $a_1 \cdots a_n$. ◻

Question. Could either of these lemmas be strengthened to “if and only if” statements?

We now show that every $a \in \omega$ is the product of *at most* one ordered sequence of primes.

Suppose that $a = p_1 \cdot \dots \cdot p_n = q_1 \cdot \dots \cdot q_m$, where the latter are both ordered sequences of primes.

Notice that q_1 divides $q_1 \cdot \dots \cdot q_m$, and so it also divides $p_1 \cdot \dots \cdot p_n$. Since q_1 is prime, by lemma 2, it divides some p_i . But p_i is prime, so it must be that $q_1 = p_i$. Now divide q_1 out of each term, and we have: $p_1 \cdot \dots \cdot p_{(n-1)} = q_2 \cdot \dots \cdot q_m$. (There are hash marks by the indexes of the p_i s, because we may have had to re-order them.) Now argue inductively in a precisely parallel fashion for q_2 , and then for the rest of q_3, \dots, q_m . After we've handled q_m , we will have identified each q_i with some p_j . Moreover, notice that there cannot be any p_j s remaining that weren't identified with some q_i , because that would make $p_1 \cdot \dots \cdot p_n$ larger than $q_1 \cdot \dots \cdot q_m$, a contradiction. \checkmark

FTA allows us to (effectively) turn any sequence of numbers into a number. Take, e.g., our example sequence from above:

(*) $\langle 11, 66, 896, 57, 0, 0 \rangle$.

This sequence has a length of 6 – it contains six elements. So take the sequence of the first six prime numbers: $\langle 2, 3, 5, 7, 11, 13 \rangle$ and consider the product:

(**) $2^{(11+1)} \cdot 3^{(66+1)} \cdot 5^{(896+1)} \cdot 7^{(57+1)} \cdot 11^{(0+1)} \cdot 13^{(0+1)}$

(We add one to each number to allow for the case where zero is an element of the sequence.

Since $n^0 = 1$, we want to represent 0 as 1 (and every other number k as $(k+1)$ so as to keep track of where the zeros are; cf. below.) Clearly this is a *very* large number – it is about 5.33×10^{713} .

But mathematically, all that matters is that it is a number. Notice that by FTA we can identify the product in (**) as an ordered sequence of primes. That is, the relevant sequence starts off with 12 2s, followed by 67 3s, then 897 5s, then... then 1 13s. In fact, (**) describes how we would uniquely pick out such an ordered sequence. So, within our chosen coding scheme (based on FTA), we can regard the product in (**) as uniquely representing (*).

Of course, there was nothing special about the sequence in (*) – a similar strategy would work for any sequence whatsoever. Thus, we have a method for encoding any and all finite sequences as single natural numbers.² A bit more generally: To transform a sequence a_1, \dots, a_n into a unique number b , first determine the first n prime numbers p_1, \dots, p_n , and then calculate the

² Pop quiz #2: Is the restriction to finite sequences of numbers important? Why or why not?

product: $b = p_1^{a_1+1} \cdot p_2^{a_2+1} \cdots p_n^{a_n+1}$. To retrieve the sequence a_1, \dots, a_n from b , simply find the unique ordered sequence of primes whose product equals b . Then collect identical members of this sequence into powers of primes, and subtract 1 from each of these n exponents. The resulting sequence of exponents is a_1, \dots, a_n .

Definition. For any set A , A^* is the set of all finite sequence of elements of A (repeats allowed). That is, $A^* = \{ \langle a_1, \dots, a_k \rangle : k \in \omega \}$.

Does the above method yield a total computable bijection from ω^* to ω ? No. First, the empty sequence $\langle \rangle$ is not assigned to any number, and second, the procedure only yields a 1-1 function – e.g., all numbers of the form $p_1^{a_1+1} \cdot p_2^{a_2+1} \cdots p_n^{a_n+1}$ will be even (because $p_1 = 2$). However, it does nicely order most of ω^* , and so a total computable bijection can readily be had. First we define an injection $g_{FTA} : \omega^* \rightarrow \omega$ as follows: $g_{FTA}(\langle \rangle) = 0$, and $g_{FTA}(\langle a_1, \dots, a_n \rangle) = p_1^{a_1+1} \cdot p_2^{a_2+1} \cdots p_n^{a_n+1}$, where p_i is the i th prime number. The function g_{FTA} effectively orders ω^* :

$$\langle a_1, \dots, a_n \rangle < \langle b_1, \dots, b_m \rangle \text{ iff } g_{FTA}(\langle a_1, \dots, a_n \rangle) < g_{FTA}(\langle b_1, \dots, b_m \rangle).$$

Moreover, every sequence occurs exactly once in this ordering. Thus, we can define $f_{FTA}(\langle a_1, \dots, a_n \rangle) = k$ iff $\langle a_1, \dots, a_n \rangle$ occupies the k th position in the order $<$ which is determined by g_{FTA} .

Corollary. If A is effectively enumerable, then so is A^* .

Proof. Suppose g effectively enumerates A . Then G is a computable surjection from ω^* onto A^* , where $G(\langle n_1, \dots, n_k \rangle) = \langle g(n_1), \dots, g(n_k) \rangle$. An effective enumeration of A^* is then given by:

$$G \circ f_{FTA}^{-1}(k), k \in \omega. \quad \delta$$

Exercise. Determine $f_{FTA}^{-1}(k)$, for $0 \leq k \leq 8$.

Exercise. Show/explain why $f_{FTA}^{-1}(\cdot)$ is computable.

Exercise. In the proof of lemma 2, we read that “the only other divisors of $a_1 \cdots a_n$ are products $c_1 \cdots c_k$ of sets of divisors of the various a_i ”. Verify this claim.