MATHEMATICAL INDUCTION AND RECURSIVE DEFINITIONS

R. C. BUCK, University of Wisconsin and Institute of Defense Analyses

Many students first encounter mathematical induction during a beginning course in algebra, either in secondary school or in college. For some of these students, this can become their first introduction to mathematical ideas, turning their attention away from computational exercises to notions of structure and proof. Their initial distrust is gradually replaced by an appreciation for its power; with reference to dominoes or inductive sets, they will usually become convinced of its reasonableness. Some students will retain a cautious attitude toward certain types of applications of induction; in this paper, I wish to increase the number of these students by discussing some of these problems.

Suppose that we have agreed upon a workable definition of the notion of function. We will deal only with the set $I = \{0, 1, 2, \cdots\}$ of nonnegative integers, or with the set $I \times I$ of pairs of integers, or more generally, with the set $I^k = I \times I \times \cdots \times I$ for some specific k. A function will always be defined on a subset of such a set, and will take its values in I. If we identify a function f with its graph, then a function f on I to I will become a specific set of pairs $\langle n, f(n) \rangle$ for all $n \in I$. More generally, we would say that any nonempty subset S of $I \times I$ that is univalent is a function; its domain will be the projection of S into the first coordinate space.

We now present a student with the following definition of a function f on I to I:

$$
\begin{aligned}
f(0) &= 1 \\
f(1) &= 2 \\
f(n+1) &= f(n) + f(n-1) \qquad \text{all } n = 1, 2, 3, \cdots .
\end{aligned}
\tag{1}
$$

I think that it will be quite convincing to the student that there is such a function f, and that it is uniquely defined by formula (1). He has no difficulty seeing that $f(2) = 3$, $f(3) = 5$, $f(4) = 8$, and so on. In short, he believes in the existence of f. However, there often remains a certain unhappiness in the mind of the student; he may say that he wants a formula for f. If he is pressed to explain, it will be found that he feels that functions ought to be described by means of certain allowed operations such as addition and composition, and ought to be built up from simpler functions. Students are constructivists at heart.

In the present example, of course, a formula can be given and should be given. He will express astonishment that so much complexity is needed for such a simple appearing function.

$$
f(n) = \frac{5 + 3\sqrt{5}}{10}\left(\frac{1 + \sqrt{5}}{2}\right)^n + \frac{5 - 3\sqrt{5}}{10}\left(\frac{1 - \sqrt{5}}{2}\right)^n . \tag{2}
$$

Of course, you can convince him that this is correct, first by computing several values to check it, and then by using mathematical induction. We observe that

$$f(0) = \frac{5+3\sqrt{5}}{10} + \frac{5-3\sqrt{5}}{10} = \frac{10}{10} = 1,$$

and that

$$\begin{aligned} f(1) &= \frac{5+3\sqrt{5}}{10} \quad \frac{1+\sqrt{5}}{2} + \frac{5-3\sqrt{5}}{10}\ \frac{1-\sqrt{5}}{2} \\ &= \frac{5+8\sqrt{5}+15}{20} + \frac{5-8\sqrt{5}+15}{20} \\ &= 2. \end{aligned}$$

We have verified the correctness of (2) for $n=0$ and $n=1$. Suppose that we have verified it for $n=0, 1, 2, \cdots, k$, can we be sure that it holds for $n=k+1$? For simplicity, set $A = (5+3\sqrt{5})/10$, $B = (5-3\sqrt{5})/10$, $\alpha = (1+\sqrt{5})/2$ and $\beta=(1-\sqrt{5})/2$. Then, formula (2) can be written as

$$f(n) = A\alpha^n + B\beta^n. \tag{3}$$

By (1), we have the right to express $f(k+1)$ as $f(k)+f(k-1)$, so that the inductive hypothesis gives us

$$\begin{aligned} f(k+1) &= [A\alpha^k + B\beta^k] + [A\alpha^{k-1} + B\beta^{k-1}] \\ &= A\alpha^{k-1}[1+\alpha] + B\beta^{k-1}[1+\beta]. \end{aligned}$$

Observing that $\alpha^2=1+\alpha$, and $\beta^2=1+\beta$, we obtain

$$f(k+1) = A\alpha^{k+1} + B\beta^{k+1},$$

verifying (3) and thus (2). Or, if the instructor has been more formal in his presentation, he can point out that this argument has shown that the set $E\subset I$ of those integers n for which (2) holds is an inductive set; since it contains 0, it must, perhaps by an axiom rather than a theorem, be the whole set I.

At this point, the line of development is sure to be interrupted by a clamor to know "where formula (2) came from." This is handled either by suppressive measures, or by embarking upon a brief discussion of difference equations with constant coefficients.

But all of this has, to a certain extent, detoured the real and valid question which was in the student's minds. Is it in fact legitimate to *define* a function by a dodge like that of formula (1)? This certainly does not describe f either as a mapping or as a class of pairs. In this particular example, we were lucky enough to have found a formula. Is this always the case? Moreover, an alert and cautious student may also raise the general problem of how one can tell whether a relation similar to (1) admits a solution. For example, is there a function f which satisfies this relation?

$$
(4)\qquad \begin{aligned} f(1) &= 1 \\ f(2n+1) &= n^2 - n + 1 \\ f(3n+1) &= 2n + f(2n+1) \end{aligned} \qquad n \geqq 1.
$$

[Ans. No; try computing $f(13)$; then, replace n^2-n+1 by $4n+1$ and see what happens.]

Perhaps the following example, which deserves to be better known, will bring the matter more clearly to a head. Suppose we wish to define a function F on the set $I \times I$, which we can for convenience picture as the first quadrant. Suppose we write down the relation:

$$
(5)\qquad F(m+1, n+1) = F(F(m, n+1), n) \qquad m, n = 0, 1, 2, \cdots.
$$

Suppose that we assign the values of F on the edges of the quadrant. Then, a little experimentation leads us to believe that there is such a function, that it is uniquely determined, and that we can calculate any desired value of F.

First, by assumption, we have available a complete knowledge of $F(0, n)$ for each n, and of $F(m, 0)$ for each m; we can calculate their values for any specific choice of m or n. Set $n=0$ in (5), obtaining a simple recursion akin to formula (1):

$$
(6)\qquad F(m+1, 1) = F(F(m, 1), 0) \qquad m = 0, 1, 2, \cdots.
$$

From this, and the knowledge of the boundary values of the supposed F, we can generate the value of $F(x, 1)$ for any desired x; for example, $F(1, 1) = F(F(0, 1), 0)$ which is computable since we know the number $F(0, 1)$, and can compute the specific value of $F(x, 0)$ which results from setting $x = F(0, 1)$. Proceeding, we next have

$$
F(2, 1) = F(F(1, 1), 0)
$$

which is now computable since we know $F(1, 1)$, and so on.

We can thus regard $F(x, 1)$ as known for any specific integer x. Now, put $n=1$ in (4), obtaining

$$
(7)\qquad F(m+1, 2) = F(F(m, 2), 1) \qquad m = 0, 1, 2, \cdots.
$$

Since we know the initial value $F(0, 2)$, we could proceed in the same fashion to compute $F(1, 2)$, $F(2, 2)$, and any later value of $F(x, 2)$. Notice in particular that at any stage, we have needed to know only a finite number of the values of F at *earlier* points; we did not need to know *all* the values of $F(x, 1)$ in order to compute $F(2, 2)$. In Figure 1, we have attempted to make this point clear by shading the region that might be needed in order to compute $F(4, 3)$.

We have thus reached the same spot with this example that we encountered with equations (1). Apparently, we can compute any desired value of F; intuitively, we are therefore convinced that relation (4) admits a solution which is a function defined on the entire set $I \times I$. However, the behavior illustrated by formula (4) may lead a student to seek assurance that the process outlined above leads to a consistent answer, that the value ascribed to $F(4, 3)$ does not depend upon his mode of procedure; again, he would be much happier if we were to exhibit F as an explicit class of ordered pairs, constructed from the relation (5) by standard set operations, for this would demonstrate existence in a much more satisfying way. We shall in fact do this at the end of this paper, and at the same time show that (5) cannot have two different solutions with the same assigned boundary values; the situation is analogous to the study of the Dirichlet problem in partial differential equations.

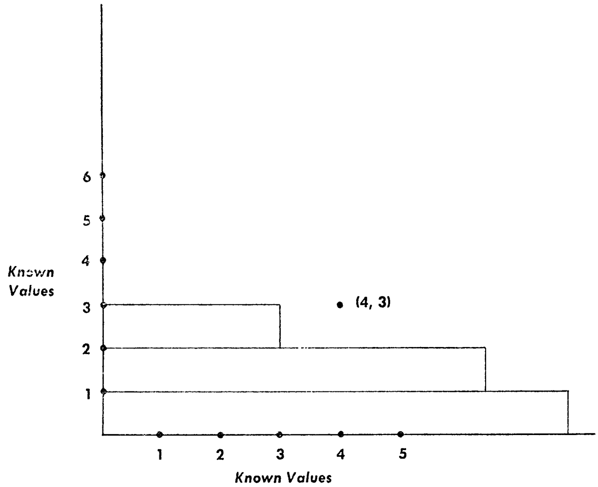

FIG. 1

Before doing this, however, we can gain some appreciation for the latent strength of the scheme (5) by examining the results of a specific choice of boundary values:

$$
(8) \qquad \begin{aligned}
F(m, 0) &= m + 1 & \qquad m &= 0, 1, \cdots. \\
F(0, 1) &= 2 \\
F(0, 2) &= 0 \\
F(0, n) &= 1 & \qquad n &= 3, 4, \cdots.
\end{aligned}
$$

Formula (6) becomes

$$
\begin{aligned}
F(m + 1, 1) &= F(F(m, 1), 0) & \qquad m &= 0, 1, 2, \cdots. \\
&= F(m, 1) + 1.
\end{aligned}
$$

With the initial value $F(0, 1)=2$ from (8), this is easily seen to have the solution

$$F(m, 1) = 2 + m \qquad m = 0, 1, \cdots. \tag{9}$$

In the same manner, (7) becomes

$$\begin{aligned} F(m + 1, 2) &= F(F(m, 2), 1) \\ &= 2 + F(m, 2) \qquad m = 0, 1, 2, \cdots \end{aligned}$$

with the initial condition $F(0, 2)=0$. From this, we obtain

$$\begin{aligned} F(1, 2) &= 2 + 0 = 2 \\ F(2, 2) &= 2 + 2 = 2(2) \\ F(3, 2) &= 2 + (2)(2) = 2(3) \end{aligned}$$

and in general,

$$F(m, 2) = 2m. \tag{10}$$

Continuing in the same way, we have

$$\begin{aligned} F(0, 3) &= 1 \\ F(m + 1, 3) &= 2F(m, 3) \end{aligned}$$

from which we deduce that $F(1, 3)=2$, $F(2, 3)=2^2$, $F(3, 3)=2^3$, and in general

$$F(m, 3) = 2^m.$$

What happens when we go to the next stage? Our simple recursion becomes

$$F(m + 1, 4) = 2^{F(m,4)}$$

with $F(0, 4)=1$. We are able to compute the values of F as before:

$$\begin{aligned} F(1, 4) &= 2 \\ F(2, 4) &= 2^2 \\ F(3, 4) &= 2^{2^2} = 16 \\ F(4, 4) &= 2^{2^{2^2}} = 2^{16} = 65536 \\ F(5, 4) &= 2^{2^{2^{2^2}}} = 2^{65536} \end{aligned}$$

and by using dots, we can fake a general formula

$$F(m, 4) = 2^{2^{\cdot^{\cdot^{2}}}} \qquad [\text{with } m \text{ two's}].$$

Let us try the next case; we have the simple recursion

$$\begin{aligned} F(0, 5) &= 1 \\ F(m + 1, 5) &= F(F(m, 5), 4) \qquad m = 0, 1, 2, \cdots \end{aligned}$$

and using our value for $F(x, 4)$, we write this as

$$F(m+1, 5) = 2^{2^{2^{\cdot^{\cdot^{2}}}}} \qquad [\text{with } F(m, 5) \text{ two's}].$$

This will suffice to compute some of the early values of F, so that we have for example

$$F(1, 5) = 2$$
$$F(2, 5) = 2^2 = 4$$
$$F(3, 5) = 2^{2^{2^{2}}} = 65536$$
$$F(4, 5) = 2^{2^{\cdot^{\cdot^{2}}}} \quad \text{with } 65536 \text{ two's}$$
$$F(5, 5) = 2^{2^{\cdot^{\cdot^{2}}}} \quad [\text{with } 2^{2^{\cdot^{\cdot^{2}}}} \; [\text{with } 65536 \text{ two's}] \text{ two's}].$$

However, I think that it is quite clear that we do not have a suitable way to write down any nonbogus general formula for $F(m, 5)$ within the notational schemes of the standard terminology.

Still less, then, will this be true for $F(m, 6)$, and manifestly more so for the function ψ of one variable which is now definable by the equation

$$\psi(x) = F(x, x) \qquad \text{for } x = 0, 1, 2, \cdots.$$

However, it is also clear that $\psi(x)$ can be computed for any specific value of x, granting the necessary time and paper—which undoubtedly exceeds both the estimated size of the universe and its duration. Indeed, $\psi(0)=1$, $\psi(1)=3$, $\psi(2)=4$, $\psi(3)=8$, $\psi(4)=65536$, and $\psi(5)=F(5, 5)$, which we have written down just above.

The existence of functions such as ψ yields an unexpected dividend. The following personal illustration may be amusing. I have found that most beginning analysis students seem to accept as plausible the conjecture that, given any increasing sequence of integers $\{c_n\}$, one could find an entire function f such that $f(n) > c_n$ for $n=1, 2, \cdots$. If you suggest $c_n = 2^n$, they counter with $f(z)=\exp(z)$. If you suggest $c_n = n!$, they suggest $f(z)=\exp(\exp(z))$. However, once they have been shown the construction of the special function ψ, and have come to appreciate its stupendous rate of growth, and the obvious possibility of creating functions which grow even more rapidly, their confidence in the conjecture seems to fade; analyticity is too delicate a phenomenon to match such catastrophic growth. Indeed, in one instance, the only student in the class who was able to overcome this feeling and find the simple general proof was one who had been absent the day before, and did not know about the function ψ. (Proof: Put $f(z) = \sum c_n (z/n)^{c_n}$, convergent for all z.)

The power of recursive definitions is now plain to the student; he will not find it hard to modify (5) for a function of three variables, generalizing Peano's recursion, so that:

$$F(x, y, 1) = x + y$$
$$F(x, y, 2) = xy$$
$$F(x, y, 3) = x^y$$

thus obtaining all the usual arithmetic operations at once. (This and the preceding example are slightly modified versions of examples given originally by Ackermann; see [3] or [4].) At this point, the student is also prepared to see the point of general theorems which deal with the more subtle aspects of existence, definability, and computability of functions.

As an illustration, let us re-examine the recursion schemes we have used, and prove that the solution of (5) is unique and can be exhibited as a set of ordered pairs. Let us start from the simple recursion relation:

$$\begin{aligned} f(0) &= a \\ f(m+1) &= g(f(m)) \qquad m = 0, 1, 2, \cdots, \end{aligned} \tag{11}$$

where a is an integer, and g is a previously defined function on I to I. Introduce a special mapping S of $I \times I$ into itself defined by:

$$\text{if } p = \langle u, v \rangle, \quad \text{then} \quad S(p) = \langle u+1, g(v) \rangle.$$

Let us say that a subset $A \subset I \times I$ is *admissible* if it obeys the pair of conditions

$$\begin{aligned} &\langle 0, a \rangle \in A \\ &\text{if } p \in A, \quad \text{then} \quad S(p) \in A. \end{aligned} \tag{12}$$

There are admissible sets, for example $I \times I$. More to the point, if there is a function f that obeys (11), then its graph is an admissible set.

Let A_0 be the *smallest* admissible set, e.g. the intersection of all the admissible sets. We show that A_0 is the graph of a function. Observe first that if A is any admissible set, and q is any point in A other than $\langle 0, a \rangle$, and if q is not of the form $S(p)$ for any $p \in A$, then we can remove q from A and still have an admissible set, since neither condition of (12) will be violated. Consequently, since A_0 is minimal, every point in it, except $\langle 0, a \rangle$, is of the form $S(p)$ for some $p \in A_0$. Let π be the projection of $I \times I$ onto the first factor, sending $\langle u, v \rangle$ into u; it is then immediate that π maps A_0 onto I one-to-one, so that A_0 is a function with domain I, and the desired solution of the recursion (11).

We have therefore produced a new function Φ of two variables, one an integer and the other a function, whose values are functions, and which is described by saying that $\Phi(a, g) = f$, where f is the (unique) solution of (11). If we let $\mathfrak{F}$ denote the class of all functions on I to I, then Φ is a mapping on $I \times \mathfrak{F}$ to $\mathfrak{F}$. Suppose now that α and β are in $\mathfrak{F}$, and let us attempt to define a sequence of functions $F_n \in \mathfrak{F}$ by the format

$$F_0 = \alpha \tag{13}$$
$$F_{n+1} = \Phi(\beta(n), F_n) \qquad n = 0, 1, 2, \cdots.$$

The first line means that $F_0(m) = \alpha(m)$ for $m = 0, 1, \cdots$. The second line is harder to interpret; if we set $f = \Phi(\beta(n), F_n)$ then, by (11) which describes Φ,

$$f(0) = \beta(n)$$
$$f(m + 1) = F_n(f(m)) \qquad m = 0, 1, 2, \cdots.$$

Since (13) identifies f with F_{n+1}, these conditions amount to asking that

$$F_{n+1}(0) = \beta(n)$$
$$F_{n+1}(m + 1) = F_n(F_{n+1}(m)) \qquad m = 0, 1, 2, \cdots.$$

If we now write $F(x, y)$ for $F_y(x)$, we see that all together, we have recaptured the form of the double recursion (5) exactly:

$$F(m, 0) = \alpha(m) \qquad m = 0, 1, 2, \cdots$$
$$F(0, n) = \beta(n - 1) \qquad n = 1, 2, \cdots$$
$$F(m + 1, n + 1) = F(F(m, n + 1), n) \qquad m, n = 0, 1, 2, \cdots.$$

Thus, we have shown that a multiple recursion of the complicated type which we used to create the function F, and then ψ, can in fact be reduced to a primitive recursion format, provided we allow function valued functions. Does (13) have a solution? If so, it will be a sequence of functions F_n, that is, a function F on I to $\mathfrak{F}$; can we show its existence by exhibiting it as a subset of $I \times \mathfrak{F}$? The pattern used earlier can be repeated exactly. Introduce a special mapping S of $I \times \mathfrak{F}$ into itself by:

$$\text{if } p = \langle u, \gamma \rangle, \quad \text{then} \quad S(p) = \langle u + 1, \Phi(\beta(u), \gamma) \rangle.$$

Again, say that $A \subset I \times \mathfrak{F}$ is admissible if A contains the point $\langle 0, \alpha \rangle$ and is mapped into itself by S. Then, in exactly the same fashion, the unique minimal admissible set turns out to be the graph of the sought-for function F.

It is clear how this can be continued, basing the study of multiple recursions on that of primitive recursions with more elaborate function valued functions. At this point, we are within range of the concept of a general recursive function and the related notions of computability and constructability. The reader is herewith referred to the bibliography.

References

1. Martin Davis, Computability and Unsolvability, McGraw-Hill, 1958.
2. Leon Henkin, On mathematical induction, this MONTHLY 67 (1960), 323–337.
3. S. C. Kleene, Introduction to Metamathematics, Van Nostrand, 1952.
4. Rózsa Péter, Rekursive Funktionen, Akad. Kiado, Budapest, 1951.
5. J. B. Rosser, Logic for Mathematicians, McGraw-Hill, 1953.