TRANSFINITE INDUCTION
AND THE REAL LINE

A. E. McCluskey & T. B. M. McMaster

Abstract This is a largely expository article, aimed at introducing the technique of *transfinite* induction to those who are familiar only with the 'positive integers' aspect of the concept. It targets the real continuum as the most intuitively accessible non-denumerable field of exploration for this, reviewing a handful of results from both ends of the present century. A version of this article was presented at the Eleventh September Meeting of the IMS at the University of Ulster at Coleraine, 7-8 September 1998.

1. Introduction

Those of us who learned induction in the context of a first–year undergraduate mathematics course usually emerged knowing two things about it: its main purpose was to prove things that we could already see were correct, and it only worked for positive integers. Fortunately, both of these "well-known facts" turn out to be remote from the truth. Inductive arguments are at least as useful when employed as engines of construction, they will serve to build up the profoundly unobvious as well as to consolidate the self-evident and–far from being limited to "plodding up the endless staircase of the natural numbers"–they are capable of ranging across the equally familiar, if far more mysterious, continuum of the real line. These are the general points at which this essay is directed.

Perhaps ironically, the first-year computer science undergraduate is generally more aware of the first of these issues than his/her mathematics contemporary through greater exposure to recursive definitions, such as those of the Fibonacci numbers:

Example 1

$$\begin{aligned} f_1 &= f_2 = 1 \\ f_n &= f_{n-1} + f_{n-2} \text{ whenever } n \geq 3 \end{aligned}$$

or of the factorials:

Example 2

$$\begin{aligned} 0! &= 1 \\ (n+1)! &= (n+1) \cdot n! \text{ whenever } n \geq 0. \end{aligned}$$

It is this 'constructive' face of induction, in which functions or other artefacts are built up layer by layer from preceding layers, that we wish to exploit. The first step in this exploitation is to free the process from its apparent dependency on the positive (or non-negative) integers, and the following two examples indicate how to do this.

Example 3

We define a function on the positive rational numbers. Begin by listing these numbers in some methodical fashion, such as the 'diagonal counting' that harks back to Cantor:

$$\mathbb{Q}^+ = \{\frac{1}{1}, \quad \frac{2}{1}, \quad \frac{1}{2}, \quad \frac{3}{1}, \quad \frac{1}{3}, \quad \frac{4}{1}, \quad \frac{3}{2}, \quad \frac{2}{3}, \quad \frac{1}{4}, \quad \frac{5}{1}, \quad \ldots\}.$$

Now, whenever x is such a number, define

$$f(x) = \begin{cases} x^{-1}, & \text{if } x \text{ is an integer;} \\ \sum f(y) & \text{where } y \text{ ranges over all the predecessors} \\ & \text{of } x \text{ in the list, if not.} \end{cases}$$

Example 4

Consider the set of numbers

$$E = \{n - 2^{1-m} : n, m \text{ are positive integers}\},$$

which can (in some sense) be listed in the natural order, thus:

$$\{0, \frac{1}{2}, \frac{3}{4}, \frac{7}{8}, \frac{15}{16}, \ldots, 1, 1\frac{1}{2}, 1\frac{3}{4}, 1\frac{7}{8}, \ldots, 2, 2\frac{1}{2}, 2\frac{3}{4}, 2\frac{7}{8}, \ldots\}.$$

For each number x in this list, define

$$g(x) = \begin{cases} x!, & \text{if } x \text{ is an integer;} \\ (g(\text{the predecessor of } x))^2, & \text{if not.} \end{cases}$$

In all of these examples the definitions are valid: intuitively this is because, given time and patience, we could build up a numerical calculation to tell us the value of the function at any particular permitted input number. However, the more precise reason for their validity is that the domain of each proposed function–$\mathbb{N}$, $\mathbb{N} \cup \{0\}$, $\mathbb{Q}^+$ and E in the above illustrations–had been given a *well-order*: that is, its elements had been ordered in such a way that *every non-empty sub-collection contained a least element.*

It is important to appreciate that this simple requirement, that the underlying set be well-ordered in this sense, is the *only* thing that is needed to power the induction machine. The exact nature or composition of the underlying set is irrelevant, as is its 'size', as is its 'natural order' if any. It is not necessary that each of its elements shall have an immediate predecessor in the ordering, nor that each element can be reached in a finite number of 'moves' from the very first one. Well-ordering is the sole rocket fuel required to propel this Explorer to the stars and beyond!

2. The nature of induction

Now let us formulate what, essentially, induction is:

(i) Lay out a well-ordered 'scale' suitable to the problem.

(ii) Identify a 'desirable circumstance' concerning a typical scale point.

(iii) Establish it as a consequence of its validity at earlier scale points (if any).

Then the circumstance (e.g. that a statement be true, that a function be assigned a value, that a construct be defined) must prevail everywhere: because *if not*, the well-ordering tells us that there would be an earliest point at which it fails; yet then it would be valid at all earlier points, so (iii) implies its validity also at the first failure point: *a contradiction*. (It is implicit in (iii), and very often stated explicitly also, that validity at the least point of the whole scale has to be established unconditionally.)

Disappointingly few of the canonical scales and number systems come pre-equipped with a natural well-ordering, but once again this is not really the point. For instance, the positive rational system $\mathbb{Q}^+$ has a built-in ordering by magnitude under which it is *not* well-ordered, but as Example 3 illustrates, we can still run induction over $\mathbb{Q}^+$ by imposing a different order that *is* a well-ordering. The question we should, therefore, be asking at this stage is 'which sets *can* be given a well-ordering?' And the reply provided by standard set theories is more encouraging: *they all can*. (As the reader may well be aware, the assertion just made is far from being a simple truth. Its status is more like that of the 'parallel postulate' in Euclidean geometry: a working assumption built into the structure of the discipline, whose logical consequences are explored therein. But it is no intention of the present article to agonize over such issues, critically important though they be. Most practitioners in this area, for most of the time, take it that every set can be well-ordered; we shall go along with that.)

In particular, the real number system $\mathbb{R}$ can be well-ordered, and this is what validates the use of induction as an exploratory tool for investigating it. Before employing it, we need to sketch in a few details concerning the *size of infinite sets*.

(i) Sets A and B *have the same size (the same number of elements)* if there is a way of pairing off all of the elements of A with all of the elements of B in a one-to-one fashion.

(ii) A *is smaller than (has fewer elements than)* B if (i) is not the case *and* A is the same size as some part or subset of B.

(iii) B *is bigger than (has more elements than)* A is a re-wording of (ii).

(iv) A *is countably infinite* if it has the same size as $\mathbb{N}$, the set of positive integers.

(v) A *is uncountable* if it is bigger than $\mathbb{N}$.

(vi) Well-order $\mathbb{R}$ so that it has a bottom element r_o and a top element r_t. Use interval notation in this ordered set with the obvious meaning, e.g.

$$[r_o, x) = \{z : r_o \leq z < x \text{ in the particular well–ordering}\}.$$

Choose the *earliest* point y in $\mathbb{R}$ for which $[r_o, y)$ has the same size as $\mathbb{R}$. Then the well-ordered set $[r_o, y)$ is denoted by $\mathbf{c}$; effectively, it is the 'smallest' well–ordered set that has as many points as $\mathbb{R}$. In summary:

(vii) $\mathbf{c}$ has the same number of elements as $\mathbb{R}$, but for any x belonging to $\mathbf{c}$, $[r_o, x)$ has *fewer* points than $\mathbb{R}$.

3. Applications

We begin with an old (1908) result, due essentially to Bernstein [1] which is still one of the clearest and cleanest illustrations extant of transfinite induction in action. Recall that a subset of the real line is *bounded* if it all lies within some fixed distance from the origin, and *closed* if it contains all of its boundary points.

Theorem. *There is a set which meets every uncountable bounded closed set in* $\mathbb{R}$ *but does not contain any such set.*

Construction

(a) How many uncountable bounded closed sets are there?
It is fairly easy to show that there are exactly $\mathbf{c}$–many of them. That is, there are as many as there are point of $\mathbf{c}$.

(b) How many points are there in each?
By first confirming that every such set contains a copy of the "Cantor middle-thirds set", it is then routine to show that each uncountable bounded closed set has exactly $\mathbf{c}$-many points.

(c) So $\mathbf{c}$ is the 'right' well–ordered scale for this problem. Use its elements to label the collection of all the uncountable bounded closed sets as

$$\{K_\alpha : \alpha \text{ belonging to } \mathbf{c}\}.$$

(d) In each K_α pick two points–let us call them red_α and $green_\alpha$– *different from all previous choices.*

Now this is the heart of the construction. The labelling by points from $\mathbf{c}$ has effectively strung out all the K_α in a transfinite list within which (compare Example 4) each of them has a clearly defined collection of predecessors (the "previous K_α" in the list) but many of them have *infinitely many predecessors*. The special characteristic of $\mathbf{c}$, pointed out in (vii), is that each K_α has *fewer than* $\mathbf{c}$ *predecessors*, and this is what makes the selection of red_α and $green_\alpha$ possible: because the previous choices amount to fewer than $\mathbf{c}$–many pairs of points and, since the K_α currently in our sights has $\mathbf{c}$–many points, there is still room to choose from it two points that were not picked in any previous selection.

(e) By induction, the choice runs for all α in $\mathbf{c}$.

(f) The set G consisting of all the '*green*' points meets the typical K_α in at least the one point $green_\alpha$, but also excludes from K_α the point labelled red_α. It is therefore a set with the desired characteristic that was enunciated in the statement of this theorem: whose demonstration is thus concluded.

Sets that have this peculiar property of 'sampling, but not consuming' each and every uncountable bounded closed set in $\mathbb{R}$ are called *Bernstein sets*. They feature largely in our second application, which concerns 'incomparable' sets in $\mathbb{R}$, and which needs some preliminary discussion of 'comparable' ones. If A and B are two subsets of the real line, and if it is possible to find a real function from A onto B that is one-to-one (for which reason it will have an inverse from B onto A) and such that both the function and its inverse are continuous (at each point of A and of B respectively) we shall call B a *copy* of A. For instance, $\mathbb{R}$ itself and $(-\frac{\pi}{2}, \frac{\pi}{2})$ are copies of one another since tan() and arctan() map continuously between them. We call sets C and D *comparable* if one of them contains a copy of the other, and *incomparable* if this is not the case. Thus (0,1) and $\mathbb{Q}$ are comparable because, with some effort, one may show that the set $\mathbb{Q} \cap (0,1)$ of those rationals that lie inside (0,1) is a copy of the whole of $\mathbb{Q}$. Again, [0,1] and (0,1) are comparable in a double sense since the first contains not merely a copy but the original of the second and the second contains $[\frac{1}{3}, \frac{2}{3}]$ which is easily seen to be a copy of the first.

Now it is a common experience that if two subsets of $\mathbb{R}$ are chosen 'at random' without a good deal of malice aforethought, they are virtually always comparable. Nevertheless, incomparable pairs of sets do exist in profusion, and the Bernstein sets will help us to detect them *in this sense at least*: for any given Bernstein set A, we can construct another one, B, that is incomparable with it.

The underlying idea of this construction is very similar to that of the previous one. We shall inspect the list of uncountable bounded closed sets K_α and, within each, select a point to *include* in B and a point to *exclude* from B: all these choices to be distinct. In parallel with this, *we should like* to scan through a full list of all the mappings π that 'might' copy A into B or B into A, and sabotage each of them: for example, by excluding from B a point of the form $\pi(a)$ for some a in A, and including in B a point b for which $\pi(b)$ lies outside A; or perhaps by including in B two points at which the values of π were equal (thus preventing π from being one–to–one on B), or some combination of such actions. Unfortunately there are too many mappings for this strategy to succeed unmodified: the set of potential copying maps is bigger than $\mathbf{c}$, so we should run out of K_α to choose within, long before we had spiked every possible π. Luckily, a continuous real function that is defined on a set which contains the rationals $\mathbb{Q}$ is fully determined by what it does on $\mathbb{Q}$; and there are only $\mathbf{c}$–many functions from $\mathbb{Q}$ to $\mathbb{R}$. These two observations allow us to use the continuous one–to–one maps from $\mathbb{Q}$ to $\mathbb{R}$ in place of the exhaustive list of copying maps, secure in the knowledge that neither the list of K_α nor the shortened list of maps will terminate before the other does.

Here now is a detailed look at the *first* step in the transfinite sabotaging process.

Lemma. *Suppose that A is a Bernstein set containing $\mathbb{Q}$, and π is a continuous one-to-one function from $\mathbb{Q}$ to $\mathbb{R}$. There are two disjoint sets In_o, Out_o, containing no more than two points each, such that if B is any set containing In_o and disjoint from Out_o, then* no *extension of π can copy A into B nor B into A.*

Proof: Exactly one of the following statements must be true:

(i) there is a number x at which π does not have a limit;

or

(ii) there are two distinct numbers x_1 and x_2 at which π has the same limit;

or

(iii) we can continuously extend π to a one–to–one mapping π^* on the whole of $\mathbb{R}$.

In cases (i) and (ii), put $In_o = \{x\}$ and $\{x_1, x_2\}$ respectively, and this will already prevent any continuous, one–to–one extension of π from copying into A a set that contains In_o. In case (iii), choose any uncountable bounded closed set K, observe that $\pi^*(K)$ is another such set and therefore cannot be entirely contained in A, so choose k in K with $\pi^*(k)$ not in A and declare In_o to be $\{k\}$. This again blocks π from extending to a copying map that could take a superset of In_o to A.

Continuing, exactly one of the next two statements is true:

(iv) the limit of π *at each point of* A *at which such a limit exists* belongs to In_o;

or

(v) there is a point of A at which π has a limit y lying outside In_o.

Now in case (iv), either the limit of π at some point of A fails to exist or, bearing in mind the comparative sizes of A and In_o! there are two points of A at which π has the same limit. In either eventuality, no continuous one–to–one extension of π over A can exist and we can take Out_o to be empty. In case (v) let $Out_o = \{y\}$ and we see that a continuous extension of π over A could not map into a set disjoint from Out_o since it will be compelled to take y as one of its values. This completes the proof.

The next step requires the iteration of this lemma so as to sabotage not merely one mapping π but a transfinite sequence of such mappings. As this iteration runs, the '*In*' sets arising from each stage are accumulated, and eventually become infinite. However, as long as no stage is reached before which we shall have already examined $\mathbf{c}$–many maps, the accumulating sets will each contain fewer than $\mathbf{c}$–many numbers and will therefore represent only a 'small proportion' of $\mathbb{R}$ itself. The revised version of the lemma thus necessitated says:

Lemma. *Given A and π as before, and given two disjoint sets In and Out, each containing fewer than* $\mathbf{c}$*–many points, it is possible to augment them by at most two points each to create new disjoint sets In(+), Out(+) such that no extension of π can copy A into a set that contains In(+) and is disjoint from Out(+), nor vice versa.* (The proof, though identical in strategy with the preceding, is more technical in detail and will be omitted. The interested reader is referred to [6].)

Now we are ready to demonstrate the existence of incomparable pairs of Bernstein sets (see [2], [3] for classical, ground–breaking explorations of this idea, and [5], [8] for associated contemporary developments).

Theorem. *Given a Bernstein set A containing the set $\mathbb{Q}$ of rationals, there is another that is incomparable with it.*

Construction
(a) There are $\mathbf{c}$–many uncountable bounded closed sets (each of which contains $\mathbf{c}$–many elements). There are also $\mathbf{c}$–many continuous one–to–one mappings from $\mathbb{Q}$ to $\mathbb{R}$. It follows that there are $\mathbf{c}$–many *pairs*

$$(K, \pi)$$

that combine an uncountable bounded closed set (K) with a continuous one–to–one map (π) from $\mathbb{Q}$ to $\mathbb{R}$. Therefore the collection of all such pairs can be listed and labelled as

$$\{(K_\alpha, \pi_\alpha) : \alpha \text{ belonging to } \mathbf{c}\}.$$

Clearly, $\mathbf{c}$ is the 'right' well–ordered scale for the problem.

(b) Now consider the 'desirable circumstance' about a typical scale point β in $\mathbf{c}$, that:

- (i) there should be defined, for each $\gamma < \beta$ in $\mathbf{c}$, two disjoint sets In_γ, Out_γ of real numbers each containing fewer than $\mathbf{c}$–many points;
- (ii) as γ increases, both In_γ and Out_γ expand;
- (iii) K_γ has at least one point in common with In_γ, and at least one point in common with Out_γ;

(iv) no extension of π_γ can copy A into any set which contains In_γ and is disjoint from Out_γ, nor *vice versa.*

(c) Kick–start the process by putting all the rational numbers into the '*In*-box' before beginning.

(d) At stage β, if everything has been implemented at all stages γ before β, then the unions of the accumulated sets In_γ, Out_γ still have fewer than **c**–many points each. Use the 'enhanced' Lemma to augment these unions in such a way as to block π_β from having an extension that could copy between A and any set that contained the augmented '*In*' set but avoided the augmented '*Out*' set. Also select from K_β any two points that have not been so far chosen, and put one into *In* and the other into *Out*. Declare In_β and Out_β to mean the newly augmented '*In*' and '*Out*' sets. Observe that the 'desirable circumstance' now also prevails at the stage immediately following β.

(e) By induction, the process runs for all β in **c**.

(f) The union B of all the sets In_β (for β in **c**), according to (iii), both chooses and rejects a point from each and every K_β: that is, it is a Bernstein set. On the other hand, (iv) shows that none of the maps π_β can extend to copy A into B, nor B into A; bearing in mind that A and B both contain $\mathbb{Q}$, and that therefore every continuous one–to–one map defined on A or B is an extension of one of these maps, that shows that A and B are incomparable in the present sense of the word. Thus, the construction is complete and the theorem established.

Notes (i) Purely routine modifications of this argument will show that, given a finite list of Bernstein sets (each, for convenience, containing $\mathbb{Q}$), we can build another that is incomparable with each of them.

(ii) With a little more care, the construction can be tweaked to show that, given a list of **c**–many Bernstein sets (or fewer), another can be constructed that is incomparable with every one in the list.

(iii) This observation, in its turn, can be used as the key step in another induction process that will generate a collection of *more* than **c**–many Bernstein sets, every two of which are incomparable.

(iv) An argument similar in spirit (though different in detail) to this discourse will produce a collection of more than **c**–many Bernsteins such that, for each two chosen from the collection, one *can* be copied into the other but not *vice versa.*

4. Towards the frontier

The end–products of the above discussion (the results in (iii) and (iv) of the last Note) were known to the great Polish mathematicians Banach, Kuratowski and Sierpiński by the mid 1920's (see for example, [2] and [3]) although they derived them by somewhat different methods. So also was the style of argument we have outlined here. Let us conclude the present article by referring to some more recent developments in the same vein.

An (abstract) *ordered set* is a collection of 'nodes' between pairs of which a 'precedence' may be specified subject to two rules:

- every node precedes itself
- if node 1 precedes node 2, and node 2 precedes node 3, then node 1 precedes node 3.

For example, the collection of nonzero integers can be made into an interesting ordered set by declaring that m precedes n whenever m is an exact divisor of n. (Notice that neither of 6, 10 precedes the other. Also note that 3 precedes -3 *and* -3 precedes 3.)

Any collection of subsets of $\mathbb{R}$ can be turned into an ordered set by declaring that 'A precedes B' shall mean that B contains a copy of A. The question that can now be perceived as lying behind much of this discussion is: *which ordered sets arise like this*? That is, given an abstract ordered set E, how can we determine whether there exists a collection of subsets of $\mathbb{R}$ whose precedence relation under 'contains a copy of' precisely matches E? Such a collection is called a *realization* of E. In this language, the partial answers referred to in (iii) and (iv) above assert that a realization can be found for:

(iii*) a certain ordered set having more than **c**–many nodes and *no* precedence relations *other* than those that are forced by the requirement that 'every node precedes itself', and for

(iv*) a certain ordered set having more than **c**–many nodes, *every* two of which are related by a precedence.

’Trisha Matthews, as part of her doctoral thesis [4], was able to prove that every ordered set on $\mathbf{c}$–many nodes that satisfies an additional and natural rule (that two nodes, each of which precedes the other, are actually one and the same node) does have a realization [5]. The present authors have recently shown [6] that this conclusion remains valid *without* the additional rule. For ordered sets having *more* than $\mathbf{c}$–many nodes, the problem is much less tractable. It is known that the family of all subsets of $\mathbb{R}$, ordered by set–inclusion, does have a realization [8]; this is one of very few extant positive results (concerning more than $\mathbf{c}$–many nodes), excluding those closely related to (iii*) and (iv*) in which the precedence relations are either as sparse or as abundant as possible. On the negative side, apart from the obvious impossibility of realizing an ordered set that has more nodes than $\mathbb{R}$ has subsets, no unrealizible example has ever been exhibited: and this has in recent years fuelled the conjecture that none exist. However, evidence has been obtained in the last few months that this conjecture is unprovable: that is, there is a logical ‘model’ of set theory within which at least one ordered set (having no more nodes than $\mathbb{R}$ has subsets) possesses no realization. (Remarkably, it is also possible to devise, for any given ordered set E of this size, another set–theoretic model in which E does have a realization!) The fundamental problem of recognizing *which ones have* consequently remains open (and, moreover, has a major logical dimension in being model-sensitive).

References

[1] F.Bernstein, *Zur Theorie der trigonometrischen Reihe*, Bericht der Königliche Sächsische Gesellschaft der Wissenschaften zu Leipzig **60** (1908).

[2] C. Kuratowski, *Sur la puissance de l’ensemble des “nombres de dimension” au sens de M. Fréchet*, Fund. Math. **8** (1926), 201-208.

[3] C. Kuratowski and W. Sierpiński, *Sur un problème de M. Fréchet concernant les dimensions des ensembles linéaires*, Fund. Math. **8** (1926), 193-200.

[4] P. T. Matthews, *Quasi-order in relation to the total negation operator*, Ph. D. thesis (1994), Queen’s University, Belfast.

[5] P. T. Matthews and T. B. M. McMaster, *Families of spaces having prescribed embeddability order-type*, Rend. Ist. Mat. Univ. Trieste **25** (1993), 345-352.

[6] A. E. McCluskey and T. B. M. McMaster, *Realising quasi-ordered sets by subspaces of "continuum-like" spaces*, to appear in Order.

[7] A. E. McCluskey and T. B. M. McMaster, *Antichains in the powerset of R: realization through induction*, in preparation.

[8] A. E. McCluskey, T. B. M. McMaster and S. W. Watson, *Representing set–inclusion by embeddability among the subspaces of the real line*, to appear in Top. Appl.

A. E. McCluskey and T. B. M. McMaster,
Department of Pure Mathematics,
The Queen's University of Belfast,
Belfast BT7 1NN,
Northern Ireland.