CAMBRIDGE

Cambridge University Press
978-0-521-31205-9 - An Introduction to the Theory of Surreal Numbers
Harry Gonshor
Excerpt
More information

1 INTRODUCTION

The surreal numbers were discovered by J.H. Conway. He was mainly interested in games for which he built up a formalism for generalizing the classical theory of impartial games. Numbers were obtained as special cases of games. Donald E. Knuth began a study of these numbers in a little book [2] in the form of a novel in which the characters are trying to use their creative talents to discover proofs. Conway goes into much more depth in his classic book On Numbers and Games [1].

I was introduced to this subject in a talk by M.D. Kruskal at the A.M.S. meeting in St. Louis in January 1977. Since then I have developed the subject from a somewhat different foundation from Conway, and carried it further in several directions. I define the surreal numbers as objects which are rather concrete to most mathematicians, as compared to Conway's, which are equivalence classes of inductively defined objects.

The surreal numbers form a proper class which contains the real numbers and the ordinals among other things. For example, in this system $\omega-1$, $\sqrt{\omega}$, etc. make sense and, in fact, arise naturally. I believe that this system is of sufficient interest to be worthy of being placed alongside the other systems that are being studied by mathematicians. First, as we shall see, we obtain a nice way of building up the real number system. Instead of being compelled to create new entities at each stage and make new definitions, we have unified definitions at the beginning and obtain the reals as a subsystem of what we already have. Secondly, and more important than obtaining a new way of building up a familiar set such as the real numbers, is the enrichment of mathematics by the inclusion of a new structure with interesting properties.

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-31205-9 - An Introduction to the Theory of Surreal Numbers
Harry Gonshor
Excerpt
More information

In fact, it is because the system seems to be so natural to the author that the first sentence contains the word "discovered" rather than "constructed" or "created." Thus the fact that the system was discovered so recently is somewhat surprising. Be that as it may, the pioneering nature of the subject gives any potential reader the opportunity of getting in on the ground floor. That is, there are practically no prerequisites for reading this book other than a little mathematical maturity. Thus the reader has the opportunity which is all too rare nowadays of getting to the surface and tackling interesting original problems without having to learn a huge amount of material in advance.

The only prerequisite worthy of mention is a minimal intuitive knowledge of ordinals, for example familiarity with the distinction between non-limit and limit ordinals. For a fuller understanding it is useful to be familiar with the basic operations of addition, multiplication, and exponentiation.

The results and some of the proofs in the earlier chapters are essentially the same as those in [1] but the theory begins with a different foundation. The later chapters tend to be more original. The ideas in Chapters 6 and 7 are new as far as I know. [1] contains several remarks related to chapter 9 where the ideas are studied in detail. Part of the material in chapter 10 was done independently by Kruskal. At present, his work is unpublished. I would like to give credit to Kruskal for pointing out to me that exponentiation can be defined in a natural way for the surreal numbers. Using his hints I developed the theory independently. Although naturally there is an overlap at the beginning, it appears from private conversations that Kruskal did not pursue the topics in sections C and D.

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-31205-9 - An Introduction to the Theory of Surreal Numbers
Harry Gonshor
Excerpt
More information

2 DEFINITION AND FUNDAMENTAL EXISTENCE THEOREM

A DEFINITION

Definition. A surreal number is a function from an initial segment of the ordinals into the set {+,-}, i.e. informally, an ordinal sequence consisting of pluses and minuses which terminate. The empty sequence is included as a possibility.

Examples. One example is the function f defined as f(0) = +, f(1) = - and f(2) = + which is informally written as (+-+). An example of infinite length is the sequence of ω pluses followed by ω minuses.

The length $\ell(a)$ of a surreal number is the least ordinal α for which it is undefined. (Since an ordinal is the set of all its predecessors this is the same as the domain of a, but I prefer to avoid this point of view.) An initial segment of a is a surreal number b such that $\ell(b) \leq \ell(a)$ and $b(\alpha) = a(\alpha)$ for all α where $b(\alpha)$ is defined. The tail of b in a is the surreal number c of length $\ell(a)-\ell(b)$ satisfying $c(\alpha) = a[\ell(b)+\alpha]$. Informally, this is the sequence obtained from a by chopping off b from the beginning. a may be regarded as the juxtaposition of b and c written bc.

For stylistic reasons I shall occasionally say that $a(\alpha) = 0$ if a is undefined at α. This should be regarded as an abuse of notation since we do not want the domain of a to be the proper class of all ordinals.

Definition. If a and b are surreal numbers we define an order as follows:

a < b if $a(\alpha) < b(\alpha)$ where α is the first place where a and b differ, with the convention that - < 0 < +, e.g. (+-) < (+) < (++).

© Cambridge University Press www.cambridge.org

Cambridge University Press
978-0-521-31205-9 - An Introduction to the Theory of Surreal Numbers
Harry Gonshor
Excerpt
More information

It is clear that this is a linear order. In fact, this is essentially a lexicographical order.

B FUNDAMENTAL EXISTENCE THEOREM

Theorem 2.1. Let F and G be two sets of surreal numbers such that $a \in F$ and $b \in G \Rightarrow a < b$. Then there exists a unique c of minimal length such that $a \in F \Rightarrow a < c$ and $b \in G \Rightarrow c < b$. Furthermore c is an initial segment of any surreal number strictly between F and G. (Note that F or G may be empty.)

Note. Henceforth I use the natural convention that if F and G are sets then "$F < G$" means "$a \in F$ and $b \in G \Rightarrow a < b$," "$F < c$" means "$a \in F \Rightarrow a < c$" and "$c < G$" means "$b \in G \Rightarrow c < b$." Thus we may write the hypothesis as $F < G$.

Example. Let F consist of all finite sequences of pluses and G be the unit set whose only member is the sequence of ω pluses. Then $F < G$. It is trivial to verify directly that c consists of ω pluses followed by a minus, i.e., $F < c < G$ and that any sequence d satisfying $F < d < G$ begins with c.

This theorem makes an alternative approach to the one in [1] possible. In [1] the author regards pairs (F,G) as abstract objects where the elements in F and G have been previously defined by the same method, as pairs of sets. (It is possible to start this induction by letting F and G both be the null set.) Since different pairs can give rise to the same number, the author needs an inductively defined equivalence relation. Theorem 2.1 gives us a definite number corresponding to the pair (F,G) so that we dispense with abstract pairs.

Proof. Clearly, it suffices to prove the initial segment property.

Case 1. If F and G are empty, then clearly the empty sequence works.

Case 2. G is empty but F is nonempty.

Let α be the least ordinal such that there does not exist $a \in F$ such that $a(\beta) = +$ for all $\beta < \alpha$. Thus α cannot equal zero,

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-31205-9 - An Introduction to the Theory of Surreal Numbers
Harry Gonshor
Excerpt
More information

since any a vacuously satisfies the condition $a(\beta) = +$ for all $\beta < 0$.

Subcase 1. α is a limit ordinal. I claim that the desired c is the sequence of α pluses, i.e., $\ell(c) = \alpha$ and $c(\beta) = +$ if $\beta < \alpha$.

Since, by choice of α, no element a of F exists such that $a(\beta) = +$ for all $\beta < \alpha$, every element of F is less than c.

Now let d be any surreal number such that $F < d$.

Suppose $\gamma < \alpha$. Then $\gamma + 1 < \alpha$, since α is a limit ordinal. Hence, by choice of α, there exists $a \in F$ such that $a(\beta) = +$ for all $\beta < \gamma+1$, i.e. $\beta \leq \gamma$. Since $a < d$, $d(\beta) = +$ for all $\beta \leq \gamma$. In particular, $d(\gamma) = +$. Thus c is an initial segment of d.

Subcase 2. α is a non-limit ordinal, $\gamma + 1$. In this case there exists an $a \in F$ such that $a(\beta) = +$ for all $\beta < \gamma$ but there is no $a \in F$ such that $a(\beta) = +$ for all $\beta \leq \gamma$. Hence any $a \in F$ satisfying: $(\beta < \gamma \Rightarrow a(\beta) = +)$ must satisfy: $(a(\gamma) = -$ or $0)$. If all such a satisfy $a(\gamma) = -$ then it is easy to see that the sequence of γ pluses works for c. If there exist such an $a \in F$ such that $a(\gamma) = 0$, i.e. the sequence of γ pluses belongs to F, then the sequence of $(\gamma+1)$ pluses works for c.

Case 3. F is empty but G is nonempty. This case is similar to Case 2.

Case 4. Both F and G are nonempty.

Let α be the least ordinal such that there do not exist $a \in F$, $b \in G$ such that $a(\beta) = b(\beta)$ for all $\beta < \alpha$. Again $\alpha \neq 0$.

Subcase 1. α is a limit ordinal. Suppose $\gamma < \alpha$; then $\gamma+1 < \alpha$. Hence there exist $a \in F$, $b \in G$ such that $a(\beta) = b(\beta)$ for all $\beta \leq \gamma$. The value $a(\gamma)$ is well-defined in the following sense. If (a_1, b_1) is another pair satisfying the above properties then $a(\beta) = a_1(\beta)$ for all $\beta \leq \gamma$. Otherwise, suppose $\delta \leq \gamma$ is the least ordinal for which $a(\beta) \neq a_1(\beta)$. Without loss of generality assume $a(\delta) < a_1(\delta)$. Then by the lexicographical order $b < a_1$, which is a contradiction since $b \in G$ and $a_1 \in F$. Thus there exists a sequence of length α, such that for all $\gamma < \alpha$ there exist $a \in F$ and $b \in G$ such that $a(\beta) = d(\beta) = b(\beta)$ for $\beta \leq \gamma$.

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-31205-9 - An Introduction to the Theory of Surreal Numbers
Harry Gonshor
Excerpt
More information

By hypothesis on α, d cannot be an initial segment of an element in F as well as an element in G. Furthermore, an element of F which does not have d as an initial segment must be less than d. (Otherwise we obtain the same contradiction, as before.) Similarly an element of G which does not have d as an initial segment must be greater than d.

It follows that if d is neither an initial segment of an element of F nor an initial segment of an element of G then d works.

Now suppose F has elements with initial segment d. Then G does not have such elements. Let F' be the set of tails with respect to d of all such elements in F. Apply case 2 to F' and ϕ to obtain d'. Then the juxtaposition dd' works.

First, as before the required inequality is satisfied with respect to all elements in F or G which do not begin with d. Since $F' < d'$ it follows from the lexicographical order that dd' is larger than all elements in F beginning with d.

On the other hand, let e be any element satisfying $F < e < G$. Recall that for all $\gamma < \alpha$ there exist $a \in F$ and $b \in G$ such that $a(\beta) = d(\beta) = b(\beta)$ for $\beta \leq \gamma$. This implies by the lexicographical order that $e(\beta) = d(\beta)$ for $\beta < \alpha$. Thus d is an initial segment of e. Again using the lexicographical order the tail e must satisfy $F' < e'$. Hence d' is an initial segment of e'. Therefore dd' is an initial segment of e.

A similar argument applies if G has elements with initial segment d.

<u>Subcase 2</u>. α is a non-limit ordinal $\gamma+1$. Then there exist $a \in F$, $b \in G$ such that $a(\beta) = b(\beta)$ for all $\beta < \gamma$ but there do not exist $a \in F$, $b \in G$ which agree for all $\beta \leq \gamma$. As before, the values $a(\beta)$ are well-defined, and we obtain a sequence d of length γ. Again, as before, any element in F which does not have d as an initial segment must be less than d and an element in G which does not have d as an initial segment must be greater than d.

Let F' be the set of tails with respect to d of elements in F which begin with d and similarly for G'. Then as stated previously, there do not exist $a \in F'$, $b \in G'$ such that $a(0) = b(0)$. [Note that in contrast to subcase 1, F' and G' are non-empty although

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-31205-9 - An Introduction to the Theory of Surreal Numbers
Harry Gonshor
Excerpt
More information

one of these sets might contain the empty sequence as its only element.] Since $F' < G'$, it follows that $a(0) < b(0)$ for all $a \in F'$, $b \in G'$.

Now suppose $d \in F$ and $d \in G$. This means that neither F' nor G' contains the empty sequence, i.e. $a(0)$ and $b(0)$ are never undefined. Since $a(0) < b(0)$, we obtain: $a(0) = -$ and $b(0) = +$. It is then clear that d works.

Since F and G are disjoint, d belongs to at most one of F and G. Suppose that $d \in G$. A similar argument will apply if $d \in F$. Then every a in F' satisfies $a(0) = -$. Let F'' be the set of tails of F' with respect to this $-$. (Such an iterated tail is, clearly, the tail with respect to the sequence $(d-.)$ Apply case 2 to F'' and ϕ to obtain d'. Then the juxtaposition $c = d{-}d'$ works. We already know that c satisfies the required inequality with respect to those elements that do not begin with d. Since no $b \in G'$ has $b(0) = -$, this takes care of all of G. The choice of d' takes care of all elements in F beginning with d (the next term of which is necessarily $-$). On the other hand, any element e satisfying $F < e < G$ must begin with d. Since $d \in G$, the next term must be $-$. By choice of d', it must be an initial segment of the tail of e with respect to $d-$, i.e. e must begin with $d{-}d'$.

This completes the proof.

<u>Definition</u>. $F|G$ is the unique c of minimal length such that $F < c < G$.

<u>Remark</u>. A slightly easier but less constructive proof is possible. First one extracts what is needed from the above proof to obtain an element c such that $F < c < G$. Although this is all that is required for the conclusion, the proof does not simplify tremendously. Neverthe-less, it simplifies slightly since there is no concern about the initial segment property. Once a c is obtained, the well-ordering principle gives us a c of minimal length. At this stage it is useful to have a definition.

<u>Definition</u>. The <u>common initial segment</u> of a and b where $a \neq b$ is the element c whose length is the least α such that $a(\alpha) \neq b(\alpha)$ and such that $c(\beta) = a(\beta) = b(\beta)$ for $\beta < \alpha$. If $a = b$ then $c = a = b$.

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-31205-9 - An Introduction to the Theory of Surreal Numbers
Harry Gonshor
Excerpt
More information

If one of a or b is an initial segment of the other, then c is the shorter element. If neither is an initial segment of the other, then either $a(\gamma) = +$ and $b(\gamma) = -$ or $a(\gamma) = -$ and $b(\gamma) = +$. In either case c is strictly between a and b.

Now let c satisfy $F < c < G$ and be of minimal length. Suppose $F < d < G$. Let e be the common initial segment of c and d. Then $F < e < G$. Since c has minimal length and e is an initial segment of c, $e = c$. Hence $c = e$ is an initial segment of d.

C ORDER PROPERTIES

Theorem 2.2. If $G = \phi$ then $F|G$ consists solely of pluses.

Proof. This follows immediately from the construction in the proof of theorem 2.1. It can also be seen trivially as follows. Suppose c has minuses. Let d be the initial segment of c of length γ where γ is the least ordinal at which c has the value plus. Then clearly $F < d$ and d has shorter length than c. This contradicts the minimality of the length of c.

Theorem 2.2a. If $F = \phi$ then $F|G$ consists solely of minuses.

Proof. Similar to the above.

Note that the empty sequence consists solely of pluses and solely of minuses!

Theorem 2.3. $\ell(F|G) \leq$ the least α such that $\forall a[a \in F \cup G \Rightarrow \ell(a) < \alpha]$.

This is trivial because of the lexicographical order, since otherwise the initial segment b of $F|G$ of length α would also satisfy $F < b < G$ contradicting the minimality of $F|G$.

Note that $<$ cannot be replaced by $\leq$. For example, if $F = \{(+)\}$ and $G = \{(++)\}$, then $F|G = (++-)$. The result also follows from the construction in the proof of theorem 2.1. In fact, the construction gives the more detailed information that every proper initial segment of $F|G$ is an initial segment of an element of $F \cup G$. (An initial segment b of a is proper if $b \neq a$).

Theorem 2.3 has a kind of converse.

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-31205-9 - An Introduction to the Theory of Surreal Numbers
Harry Gonshor
Excerpt
More information

Theorem 2.4. Any a of length α can be expressed in the form $F|G$ where all elements of $F \cup G$ have length less than α.

Proof. Let $F = \{b: \ell(b) < \alpha$ and $b < a\}$ and $G = \{b: \ell(b) < \alpha$ and $b > a\}$. Then $F < a < G$ and every element of length less than α is, by definition, in F or G so that a satisfies the minimum length condition. Note that the argument is valid even if a is the empty sequence.

The last result is a step in the way of showing the connection between what is done here and the spirit of [1], since the result says that every element can be expressed in terms of elements of smaller length, thus every element can be obtained inductively by the methods of [1]. The next theorem shows that the ordering in [1] is equivalent to the one used here.

Theorem 2.5. Suppose $F|G = c$ and $F'|G' = d$. Then $c \leq d$ iff $c < G'$ and $F < d$.

Proof. We know that $F < c < G$ and $F' < d < G'$. Suppose $c \leq d$; then $c \leq d < G'$ and $F < c \leq d$. For the converse, assume $c < G'$ and $F < d$. We show that $d < c$ leads to a contradiction. This assumption yields $F < d < c < G$. Hence c is an initial segment of d. Also $F' < d < c < G'$ so d is an initial segment of c. Hence $c = d$ which contradicts $d < c$.

This last result is of minor interest for our purpose. Its main interest is that together with theorems 2.1 and 2.4 it shows that we are dealing with essentially the same objects as in [1] although here they are concretely defined. Since the present work is self-contained this is not of urgent importance, although it is worthy of noting.

Of fundamental importance here will be what I call the "cofinality theorems." They are analogous to classical results such as: In the ε,δ definition of a limit, it suffices to consider rational ε; and in the definition of a direct limit of objects with respect to a directed set, a cofinal subset gives rise to an isomorphic object.

Definition. (F',G') is cofinal in (F,G) if $(\forall a \varepsilon F)(\exists b \varepsilon F')(b \geq a) \wedge (\forall a \varepsilon G)(\exists b \varepsilon G')(b \leq a)$.

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-31205-9 - An Introduction to the Theory of Surreal Numbers
Harry Gonshor
Excerpt
More information

It is clear that (F,G) is cofinal in (F,G), and that (F",G") cofinal in (F',G') and (F',G') cofinal in (F,G) implies (F",G") cofinal in (F,G). Also if $F \subset F'$ and $G \subset G'$ then (F',G') is cofinal in (F,G).

The following theorems are important although they are trivial to prove.

Theorem 2.6 (the cofinality theorem). Suppose $F|G = a$, $F' < a < G'$, and (F',G') is cofinal in (F,G); then $F'|G' = a$.

Proof. Suppose $\ell(b) < \ell(a)$ and $F' < b < G'$. It follows immediately from cofinality that $F < b < G$, contradicting the minimality of $\ell(a)$. Hence $a = F'|G'$.

Theorem 2.7 (cofinality theorem b). Suppose (F,G) and (F',G') are mutually cofinal in each other. Then $F|G = F'|G'$.

Note that it is enough to assume that $F|G$ has meaning since $F < G \Rightarrow F' < G'$.

Proof. $\{x:F<x<G\} = \{x:F'<x<G'\}$. Hence the element of minimal length is the same.

Although the two above theorems are closely related they are not quite the same. Theorem 2.6 will be especially useful in the sequel. I emphasize that in spite of the simplicity of the proof it is more convenient to quote the term "cofinality" than to repeat the trivial argument every time it is used. Also it is convenient often with abuse of notation to say that H' is cofinal in H. However, this is unambiguous only if it is clearly understood whether H and H' appear on the left or right, i.e. we must consider whether we are comparing (H,G) with (H',G') or (F,H) with (F',H'). This is usually clear from the context.

Cofinality will be used to sharpen theorem 2.4 to obtain the canonical representation of a as $F|G$. Of course, the representation in theorem 2.4 itself may be regarded as the "canonical" representation. The choice is simply a matter of taste.

Theorem 2.8. Let a be a surreal number. Suppose that $F' = \{b: b < a$

© Cambridge University Press www.cambridge.org