Constructing the real numbers in HOL

John Harrison

University of Cambridge Computer Laboratory, New Museums Site, Pembroke Street, Cambridge, CB2 3QG, England.

Abstract

We describe a construction of the real numbers in the HOL theorem-prover by strictly definitional means using a version of Dedekind's method. We also outline the theory of mathematical analysis that has been built on top of it, and discuss current and potential applications in verification and computer algebra.

Keyword Codes: F.4.1; I.2.3
Keywords: Mathematical Logic; Deduction and Theorem Proving

1 Constructing the real numbers

We do not address any philosophical issues, but merely take a formal view that the real numbers are an algebraic structure obeying the following axioms:

1. $1 \neq 0$
2. $\forall x\ y.\ x + y = y + x$
3. $\forall x\ y\ z.\ x + (y + z) = (x + y) + z$
4. $\forall x.\ 0 + x = x$
5. $\forall x.\ (-x) + x = 0$
6. $\forall x\ y.\ xy = yx$
7. $\forall x\ y\ z.\ x(yz) = (xy)z$
8. $\forall x.\ 1x = x$
9. $\forall x.\ (x \neq 0) \Longrightarrow (x^{-1}x = 1)$
10. $\forall x\ y\ z.\ x(y + z) = xy + xz$

11. $\forall x\ y.\ (x = y) \vee x < y \vee y < x$

12. $\forall x\ y\ z.\ x < y \wedge y < z \Longrightarrow x < z$

13. $\forall x.\ x \not< x$

14. $\forall y\ z.\ y < z \Longrightarrow \forall x.\ x + y < x + z$

15. $\forall x\ y.\ 0 < x \wedge 0 < y \Longrightarrow 0 < xy$

16. $\forall S.\ (\exists x.\ x \in S)\ \wedge (\exists M.\ \forall x \in S.\ x \leq M)$
    $\Longrightarrow \exists m.\ (\forall x \in S.\ x \leq m)\ \wedge (\forall m' < m.\ \exists x \in S.\ x > m'))$

In the above, we use the normal notation for ordered fields without prejudice as to what 0, say, actually denotes. Note that juxtaposition represents 'multiplication' here; context and spacing should disambiguate it from function application later on.

The axiom that sets the reals apart from other ordered fields is the last of the above, the so-called *supremum property* which states that every nonempty set which has an upper bound has a *least upper bound* (LUB) or *supremum*. For example, the two sets $\{x \mid x^2 < 2\}$ and $\{x \mid x^2 \leq 2\}$ *both* have a supremum of $\sqrt{2}$, although one of the sets contains $\sqrt{2}$ and the other does not.

1.1 *The* real numbers?

An analagous axiomatization to the above in first-order logic cannot incorporate the necessary quantification (over sets or properties of real numbers) to express the last axiom properly; this can be got around by using a schema of axioms, but this still admits of nonstandard models.

However in higher-order logic, the above axioms do determine the reals uniquely up to isomorphism, *provided* the inverse is defined only on the subset of nonzero reals. (For a proof, see [4], [6] or [16].) Since HOL functions are total, this is not convenient in our formalization. We therefore have an extra 'degree of freedom' that does not exist traditionally: we can define 0^{-1} to be whatever we like. To avoid surprises we keep it 'undefined' (effectively $\varepsilon x.\ F$), but it is important to realize that this is not the same as true undefinedness. For example we can prove $\forall x.\ 0x = 0$ and so in particular, assuming we define $x/y = xy^{-1}$ (as we do),

$$0/0 = 0$$

Conversely, a theorem true classically which is *not* true in our framework is:

$$\forall x \in \mathbb{R}.\ (tan(x) = 0) \Longrightarrow \exists n \in \mathbb{Z}.\ x = n\pi$$

because we cannot exclude the possibility that $cos(x)^{-1}$ is zero at odd multiples of $\pi/2$, in other words that $0^{-1} = 0$. There is even something to be said for *defining* $0^{-1} = 0$ since this makes the inverse into a bijection, so things like the following are true universally:

$$\forall x.\ (x^{-1})^{-1} = x$$

$$\forall x.\ 0 < x \equiv 0 < x^{-1}$$

We feel these issues are unlikely to present problems in practice, because division by zero is normally treated as a special case anyway, but one should be aware of them.

1.2 The classical construction

There are a variety of different ways of constructing the real numbers from a simple basis. One interesting recent method is the use of Conway Games explained in [7]. Others favour approaches using positional representations — see for example [1], [4] or [8].

However the two classical methods both start from the rational numbers. In their turn, these are normally constructed from the integers, which are themselves constructed from the naturals. We look at each of these steps in more detail.

1.2.1 From naturals ($\mathbb{N}$) to integers ($\mathbb{Z}$)

There are various possible representations of the integers in terms of the natural numbers, such as:

- A pair consisting of a boolean 'sign bit' and a natural number. For example $(true, 1)$ might represent $+1$ and $(false, 2)$ represent -2.
- A pair of natural numbers, where one imagines (m, n) standing for $m - n$ in the integers. Thus $(1, 0)$ represents $+1$ and $(5, 7)$ represents -2.

The main problem with both the above is non-uniqueness. Manifestly, every number has an infinity of representatives in the second case; $+1$ could equally well be represented by $(2, 1)$, $(3, 2)$, $(4, 3)$ and so on. Less egregious is the first case, but there are still two representations of zero, $(false, 0)$ and $(true, 0)$. There are two natural ways round this problem:

- Consider only a minimal set of representatives, which are in some sense canonical. For example one might in the first case exclude $(false, 0)$, and in the second insist that one or both numbers of the pair be zero.
- Define an equivalence relation expressing the effective identity of sets of terms, and use the equivalence classes under this relation, rather than the representatives themselves, to construct the new type.

1.2.2 From integers ($\mathbb{Z}$) to rationals ($\mathbb{Q}$)

This stage is a particular case of a well-known construction in abstract algebra, constructing the *field of fractions of an integral domain*, an integral domain being a nontrivial commutative ring with the property that

$$\forall x\ y.\ (xy = 0) \Longrightarrow (x = 0)\ \vee\ (y = 0)$$

The procedure consists of considering pairs of integers, which one thinks of as the numerator and denominator of a fraction; it is necessary to exclude 0 from the possible denominators. Then one uses equivalence classes of this subset of pairs under the obvious 'cross multiplication' equivalence relation

$$(x, y) \sim (x', y') \equiv (xy' = x'y)$$

As with the path from $\mathbb{N}$ to $\mathbb{Z}$, we have the option of eschewing equivalence classes in favour of choosing canonical elements. The natural choice of canonical form would be to insist that the pair of elements be coprime, i.e. represent a cancelled fraction (though this is not available in a general integral domain).

1.2.3 From rationals ($\mathbb{Q}$) to reals ($\mathbb{R}$)

This is the most difficult stage. There are two well-established classical methods, which were published independently by Cantor and Dedekind, both in 1872. (Cantor's method was largely anticipated by Meray in 1869, and one can find precursors of Dedekind's method as far back as Eudoxus with his theory of proportion.)

Cantor's Method This method identifies a real number with the set of all rational sequences which converge to it. To say that a sequence (s_n) converges to s, written $s_n \to s$ means:

$$\forall \epsilon > 0.\ \exists N.\ \forall n \geq N.\ |s_n - s| < \epsilon$$

This is no good as a definition, because it contains the limit itself, which may not be rational. However the following variant is equivalent (we prove this later after constructing our real numbers):

$$\forall \epsilon > 0.\ \exists N.\ \forall m \geq N, n \geq N.\ |s_m - s_n| < \epsilon$$

(It does not matter that we will restrict ϵ to rational values, since $\mathbb{Q}$ is dense in $\mathbb{R}$, i.e. between any two distinct reals there is a rational.) A sequence which satisfies this property is called a *Cauchy sequence*.

The fact that two series (s_n) and (t_n) converge to the same limit can also be expressed without using the limit itself:

$$\forall \epsilon > 0.\ \exists N.\ \forall n \geq N.\ |s_n - t_n| < \epsilon$$

It is easy to see that the above defines an equivalence relation on Cauchy sequences, and the real numbers can be defined as its equivalence classes. The arithmetic operations can be inherited from those of the rationals in a fairly natural way, although the supremum presents slightly more difficulty. A complete treatment is given by Thurston [17].

Dedekind's Method Dedekind's method identifies a real number with the set of all rational numbers less than it. Once again this is not immediately satisfactory as a definition, but it is possible to give an equivalent definition of such a set without referring to the bounding real number. We shall call such a set a *cut*. The four properties required of a set C for it to be a cut are as follows:

1. $\exists x.\ x \in C$
2. $\exists x.\ x \notin C$
3. $\forall x \in C.\ \forall y < x.\ y \in C$
4. $\forall x \in C.\ \exists y > x.\ y \in C$

These state respectively that a cut is not empty, is not $\mathbb{Q}$ in its entirety, is a 'down set', and has no greatest element. Again the arithmetic operations can be inherited from $\mathbb{Q}$ in a natural way, and the supremum of a set of cuts is simply the union of that set.

1.3 The lattice of number systems

It is harder than it might appear at first sight to make the above rigorous. Defining the integers as sign/magnitude pairs excluding $(false, 0)$ means that even proving the associative law of addition for integers represents a considerable amount of work, because there are eight different cases according to the sign of each of the three integers. (Ostensibly at any rate; avoiding the case split requires careful exploitation of symmetry and the distributivity of negation over addition.)

Defining the rationals as pairs of integers is even worse. Using cancelled fractions seems a bad idea because the proofs of the elementary theorems require a fair few lemmas about coprimality and divisibility. It seems that the use of equivalence classes is better: essentially we would have to prove similar theorems anyway in the case of cancelled fractions, with the added complexity of canonicalization. However the use of equivalence classes is slightly harder than intuition would suggest, particularly since the equivalence relation is *not* an equivalence relation over the whole type, but only on the subset with nonzero denominators. To see this, observe that

$$(1, 2) \sim (0, 0) \ \wedge \ (0, 0) \sim (1, 3)$$

But clearly $(1, 2) \not\sim (1, 3)$, so transitivity fails globally. Therefore many of the proofs are hedged with conditions about membership of the admissible subset, which renders them more complicated. (An alternative which is slightly cleaner, but still tedious, is to define the 'subtype' of nonzero integers and prove the closure of the arithmetic operators under certain conditions.) Finally, when it comes to constructing the reals from the rationals, neither of the classical methods is very appealing.

- Cauchy's method requires the generation of a significant body of 'analytical' lemmas about the rational numbers. Moreover, the problem of defining equivalence relations over subsets of types would have to be faced, which as noted above is tiresome.

- When proving the axioms for the structure, Dedekind's method requires several case-splits according to the sign of variables. This is mainly because the product of two negative rationals is positive, so the natural definition of multiplication on cuts

  $$XY = \{xy \mid x \in X \ \wedge \ y \in Y\}$$

  does not work. The two cuts X and Y extend to $-\infty$, so there will exist products of these large and negative numbers which are arbitrarily large and positive. Therefore the set on the right is not a cut.

  This difficulty is usually noted in sketch proofs given in books, but to carry through in detail the complicated case splits they gloss over is far from easy.

A rather more sophisticated approach is suggested by Conway. Although the novel method for constructing the reals explained in [7] did not seem to be possible to formalize in HOL, his incidental discussion of the classical methods is perceptive. He considers the following lattice of number systems:

$$\begin{tikzcd}
 & \mathbb{R} \arrow[dl, dash] \arrow[dr, dash] & & \\
\mathbb{R}^+ \arrow[dr, dash] & & \mathbb{Q} \arrow[dl, dash] \arrow[dr, dash] & \\
 & \mathbb{Q}^+ \arrow[dr, dash] & & \mathbb{Z} \arrow[dl, dash] \\
 & & \mathbb{N} &
\end{tikzcd}$$

Here $\mathbb{Q}^+$ represents the positive rationals and $\mathbb{R}^+$ the positive reals (we do not worry yet about whether these include 0). Firstly, Conway emphasizes the difficulty of constructing $\mathbb{R}$ from $\mathbb{Q}$ by Dedekind cuts:

> Nobody can seriously pretend that he has ever discussed even eight cases in such a theorem — yet I have seen a presentation in which one theorem actually had 64 cases ... Of course an elegant treatment will manage to discuss several cases at once, but one has to work very hard to find such a treatment.

He advocates instead following the path through $\mathbb{Q}^+$ and $\mathbb{R}^+$, at least if Dedekind's method is to be used. This avoids the case splits (otherwise it is essentially the same as the signed case presented above), and as we shall see, has other advantages as well.

One apparent drawback of using this path is that we lose the potentially useful intermediate types $\mathbb{Z}$ and $\mathbb{Q}$. However this is not really so, because the method used to construct $\mathbb{R}$ from $\mathbb{R}^+$ can be used almost unchanged (and this is where a computer theorem prover scores over a human) to construct $\mathbb{Z}$ and $\mathbb{Q}$ from their positive-only counterparts.

Landau's book [14] is one of the few to present Dedekind's construction in more or less full detail. This also, significantly, uses the path through $\mathbb{Q}^+$ and $\mathbb{R}^+$. Landau makes the step from $\mathbb{R}^+$ to $\mathbb{R}$ by introducing zero and negative numbers, which is similar to the sign/magnitude representation, and with the same problems of case splitting. Conway anticipates this and prefers instead the use of equivalence classes of pairs of numbers. In fact there is then quite a close analogy with the construction of the rationals, with addition taking the role of multiplication.

1.4 The theory of semirings

To journey along the path through $\mathbb{Q}^+$ and $\mathbb{R}^+$, we must know which algebraic laws for these structures we are going to need to get $\mathbb{R}$ out at the other end, and whether we are going to include 0. We will refer to structures like $\mathbb{N}$, $\mathbb{Q}^+$ and $\mathbb{R}^+$ as *semirings*, although to some authors [9] this implies that they contain a zero.

Semirings can be characterized by relatively few axioms. Ordering does not have to be primitive: we can *define* an ordering as follows, whether or not the structure contains a zero:

$$x < y \equiv (x \neq y) \wedge (\exists d.\ y = x + d)$$

Investigation reveals that the following set of axioms is sufficient to allow the derivation of a nontrivial ordered ring by the method of equivalence classes of pairs of numbers under $(x, y) \sim (x', y') \equiv (x + y' = x' + y)$. They are chosen to be true whether or not the semiring contains a zero, which explains why the penultimate one looks a bit peculiar.

1. $\forall x\ y.\ x + y = y + x$
2. $\forall x\ y\ z.\ x + (y + z) = (x + y) + z$
3. $\forall x\ y.\ xy = yx$
4. $\forall x\ y\ z.\ x(yz) = (xy)z$
5. $\forall x\ y\ z.\ x(y + z) = xy + xz$
6. $\forall x.\ 1x = x$
7. $\forall x\ y.\ (x = y)\ \vee (\exists d.\ x = y + d)\ \vee (\exists d.\ y = x + d)$
8. $1 + 1 \neq 1$
9. $\forall x\ y\ z.\ (x + (y + z) = x) \Longrightarrow (x + y = x)$
10. $\forall x\ y\ z.\ (x + y = x + z) \Longrightarrow (y = z)$

It is not hard to see that the last three axioms can all be derived from the single axiom

$$\forall x\ y.\ x + y \neq x$$

This is a strong argument for not including a zero in the structure: we will have to prove fewer axioms as primitive for the semirings we construct. Further, it allows certain theorems such as the field axiom and the Archimedean property (see below) to be written in a simpler form.

There is unfortunately a problem: the standard HOL theory of natural numbers does contain a zero. We could have defined a new type of nonzero natural numbers, but that seemed rather wasteful, so instead we wrote a procedure which works whether or not the semiring contains a zero. The procedure requires the full list of axioms above, although where possible it is easier to prove $\forall x\ y.\ x + y \neq x$ explicitly and derive the others from that.

There are a few extra axioms we need for particular semirings. To get from $\mathbb{R}^+$ to $\mathbb{R}$, we require a form of the supremum property for $\mathbb{R}^+$. And to get from $\mathbb{Q}^+$ to $\mathbb{R}^+$ we need to prove for $\mathbb{Q}^+$ both the field axiom

$$\forall x.\ x^{-1}x = 1$$

and also a form of the *Archimedean property*. This states that if we define a function

$$addn\ n\ x = x + \ldots + x$$

where there are n terms in the sum (of course if we are sloppy with types, $addn\ n\ x$ is just nx), then the following is true

$$\forall x\ y.\ \exists n.\ addn\ n\ x > y$$

Note that neither of the above would be true if $\mathbb{Q}^+$ contained a zero; they would both be consequent on $x \neq 0$.

1.5 Equivalence relations

Every step we are to take in the lattice, with the sole exception of the line from $\mathbb{Q}^+$ to $\mathbb{R}^+$, involves constructing a set of equivalence classes. To make this easier, we wrote a procedure to automate it, given:

- A name for the new type
- A theorem asserting that a (2-place curried) relation (say R) is an equivalence relation, in the following simple form:

  $$\forall x\ y.\ xRy \equiv (Rx = Ry)$$

  Supposing we are constructing the integers from pairs of natural numbers as explained above, the relation would be defined by

  $$(x_1, y_1)R(x_2, y_2) = (x_1 + y_2 = x_2 + y_1)$$

- A list of operations on the representatives together with the desired name of the corresponding operators over the equivalence classes. For example, we might give it an addition operation defined on pairs of numbers as follows:

$$(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$$

- A list of theorems asserting that the operations on representatives are all well-defined, in the sense that, taking addition as our example again:

$$(x\ R\ x') \ \wedge\ (y\ R\ y') \Longrightarrow (x + y)\ R\ (x' + y')$$

  When the relevant argument or result is not of the representing type, equality takes the place of R.

- A list of theorems about the operations on representatives, e.g. the associative law:

$$x + (y + z) = (x + y) + z$$

The procedure first constructs a type of equivalence classes. The characteristic predicate required to select the set $\{R\ x\}$ is formally:

$$\lambda C.\ \exists x.\ C = R\ x$$

Next, the appropriate operations on the new type are defined. For example $+$ gives rise to a new operator $+^*$ (we use the star consistently, but in fact the user specifies the name of the operator) on the equivalence relation as follows:

$$X +^* Y = R((\varepsilon x.\ x \in X) + (\varepsilon y.\ y \in Y))$$

In other words, pick, using the ε operator, arbitrary representatives of each equivalence class, operate on them and then take the equivalence class of the result. If arguments or result are not of the representing type, then we avoid picking representatives or applying R, respectively. For example an 'is positive' predicate would be elevated as follows:

$$ispos^* X = ispos(\varepsilon x.\ x \in X)$$

and the *addn* function mentioned above as

$$addn^*\ n\ X = R(addn\ n\ (\varepsilon x.\ x \in X))$$

Finally, the proof procedure tries to convert the theorems on representatives into theorems about the new type. This is the only part which requires R to be an equivalence relation and the operations to be well defined. In the case of the associative law, we get:

$$X +^* (Y +^* Z) = (X +^* Y) +^* Z$$

Our procedure does *not* work if we have to deal with a subset of the basic type. We could have generalized the procedure: one would also have to supply it with theorems expressing the closure of the operations with respect to this subset, and the other theorems may become conditional on the variables belonging to the subset. The extra complexity was not necessary for the task in hand; further, there are a few problems to be resolved. Consider a definition of multiplicative inverse for a field of fractions in an integral domain:

$$(x, y)^{-1} = (y, x)$$

Then one needs a condition $x \neq 0$ on the closure theorem. Dealing with things like this in a regular way seems quite awkward.

1.6 From $\mathbb{N}$ to $\mathbb{Q}^+$

This is reasonably straightforward, but we do have to deal with the problem that the natural numbers contain 0 when we would rather they did not. The solution chosen was to use (x, y) to represent $(x + 1)/(y + 1)$.

The use of $y + 1$ avoids zero denominators without using subsets, which would defeat the equivalence class procedure. Using $x+1$ avoids including zero in the rational semiring, which is what we want, and also makes the proofs more symmetrical and regular. The only drawback is that the definitions of the operations are somewhat more complicated. For example the addition of (x_1, y_1) and (x_2, y_2) is defined as

$$((x_1 + 1)(y_2 + 1) + (x_2 + 1)(y_1 + 1) - 1, (y_1 + 1)(y_2 + 1) - 1)$$

The apparent extra difficulty of the proofs can be overcome by a simple tactic which manages to eliminate a lot of the above complexity.

1.7 From $\mathbb{Q}^+$ to $\mathbb{R}^+$

This is the most difficult part of the whole procedure. Although we are dealing with semirings, the Dedekind cuts procedure is essentially identical to the full case, but includes none of the complicated case splits. We define the operations on cuts as follows (taking some liberties with the HOL notation):

- $sup\ S = \bigcup S$
- $X + Y = \{x + y \mid x \in X \ \wedge \ y \in Y\}$
- $XY = \{xy \mid x \in X \ \wedge \ y \in Y\}$
- $X^{-1} = \{w \mid \exists d < 1.\ \forall x \in X.\ wx < d\}$

Only the last of these is unobvious: the more natural definition is

$$X^{-1} = \{w \mid \forall x \in X.\ wx < 1\}$$

However this would mean that unless a cut denoted an irrational real, its inverse would not be a cut, for example:

$$\{x \in \mathbb{Q}^+ \mid x < 1\}^{-1} = \{x \in \mathbb{Q}^+ \mid x \le 1\}$$

The construction then consists of proving closure, i.e. that when applied to cuts, all the operations yield a cut (including the trivial instance of proving that the set representing real 1 is a cut), and that all the required axioms hold. These proofs, while mostly routine, are sometimes quite long, so it is not possible to discuss them all here. We sketch only the proof of the axiom

$$\forall X.\ X^{-1}X = 1$$

which is a fairly representative example. In the above and what follows, we assume X denotes a cut, without stating it explicitly; in fact in the HOL proofs X becomes *cut* X, where *cut* is the type bijection from the real number type to the set of rational cuts. This means that no explicit set constraint is necessary since *cut* X is *always* a cut. Firstly we need the following sequence of lemmas.

Lemma 1

$$\forall X\ x\ y.\ x \in X \wedge y \notin X \Longrightarrow x < y$$

This follows easily from cut property 3.

Lemma 2

$$\forall X\ x\ y.\ x \notin X \wedge x < y \Longrightarrow y \notin X$$

Also a straightforward consequence of cut property 3.

Lemma 3 This states that we can get arbitrarily close to the "top" of a cut.

$$\forall X\ e.\ \exists x.\ x \in X \wedge x + e \notin X$$

To prove this, choose any $x_0 \in X$, and $x_1 \notin X$ (this is possible by virtue of cut properties 1 and 2). Then consider the sequence of rationals

$$\{x_0 + \mathit{addn}\ n\ e \mid n \in \mathbb{N}\}$$

It is a simple consequence of the Archimedean property that there is an n with

$$x_0 + \mathit{addn}\ n\ e > x_1$$

Then from Lemma 2 and the wellfoundedness of the naturals, we know there is a *least* n such that $x_0 + \mathit{addn}\ n\ e \notin X$, say k. Further, $k \neq 0$ because we know $x_0 \in X$. Consequently

$$x = x_0 + \mathit{addn}\ (k-1)\ e$$

has the required property. (Strictly, *addn* is not defined at 0 because $\mathbb{Q}^+$ has no zero, but the reasoning is essentially the same as the above.)

Lemma 4 This is just a multiplicative rather than additive version of Lemma 3.

$$\forall X\ u.\ u > 1 \Longrightarrow \exists x.\ x \in X \wedge ux \notin X$$

To prove this, choose any $x_0 \in X$, possible by cut property 1. If $ux_0 \notin X$, then we are finished. Otherwise let

$$e = x_0(u-1)$$

so by Lemma 3, we can pick an x with

$$x \in X \wedge x + e \notin X$$

But now we have

$$x_0 + x_0(u-1) \in X \wedge x + x_0(u-1) \notin X$$

Consequently, using Lemmas 1 and 2, together with simple properties of $<$, we deduce that $x_0 < x$, and therefore $x + x(u-1) \notin X$, as required.

Main theorem

$$\forall X.\ X^{-1}X = 1$$

Translated into cuts, we want to show

$$\{wx \mid x \in X \wedge \exists d < 1.\ \forall y \in X.\ wy < d\} = \{z \mid z < 1\}$$

Expressed formally in HOL, this means establishing the following logical equivalence

$$\forall z.\ (\exists w\ x.(z = wx) \wedge x \in X \wedge (\exists d < 1.\ \forall y \in X.\ wy < d)) = z < 1$$

This reduces to two implications. A little thought will show that the left to right implication is straightforward. For the other, suppose $z < 1$. Then it is a simple property of the rationals that we can find a d such that $z < d$ and $d < 1$ (for example $d = (z+1)/2$). Therefore $z^{-1}d > 1$. By Lemma 4, we can choose an x such that

$$x \in X \wedge z^{-1}dx \notin X$$

Now let $w = zx^{-1}$. Then w, x and d are going to be witnesses for the correspondingly named existentially quantified variables. It remains only to prove

$$\forall y \in X.\ wy < d$$

But if $y \in X$, then by Lemma 1, $y < z^{-1}dx$; and $z = wx$, so we have $y < w^{-1}x^{-1}dx$, i.e. $wy < d$, as required.

1.8 From $\mathbb{R}^+$ to $\mathbb{R}$

Most of this is the same as the construction of $\mathbb{Z}$ and $\mathbb{Q}$ from their half-counterparts. The extra theorem that has to be transferred across is the supremum property. The supremum property for the positive reals states that

> Every nonempty set of positive reals which is bounded above has a supremum.

The first step is to transfer this result to the type of reals. Although not vacuous (formally, the positive reals are a completely different type), this is straightforward because the type bijections define an isomorphism between the *type* of positive reals and the positive elements of the type of reals. The theorem now becomes

> Every nonempty set of real numbers which is bounded above, and all of whose elements are strictly positive, has a supremum.

We generalize this in two stages. Firstly it is simple to prove the following strengthening:

> Every nonempty set of real numbers which is bounded above, and which contains at least one strictly positive element, has a supremum.

(The property 'nonempty' is actually superfluous here, but we keep it in for regularity.) This follows because l is a supremum of the whole set if and only if it is a supremum of the strictly positive elements of it, since any positive number is greater than any negative number.

Finally we prove the lemma that for any d, positive or negative, l is a supremum of S if and only if $l + d$ is of $\{x + d \mid x \in S\}$. Now this can be used to reduce the case of any nonempty bounded set to the above, by choosing a d to 'translate' it such that it has at least one strictly positive element. We now have the full result:

> Every nonempty set of real numbers which is bounded above has a supremum.

1.9 Interface issues

It is desirable to use the normal arithmetic symbols like $+$ both for the real numbers and the natural numbers (not to mention the integers, the rationals, and other uses such as binary summation in process calculi). At present, HOL does not have a mechanism to support overloading of constant names, but it is hoped to add something in the future. At present the interface map feature is used to allow the user to swap easily between using $+$ etc. for different purposes, but it is not possible to use the same symbols for different operations in the same term. Furthermore, a different notation is required for real number constants. At present the ampersand is used as an interface map for the 'inclusion' function $\iota : \mathbb{N} \to \mathbb{R}$, so the real constants can be written &0, &1 etc. This is reminiscent of programming languages like C and Standard ML where the floating-point constants are distinguished syntactically from integer ones by being written 1.0 or $1e1$ etc.

2 Building on the real number axioms

To make the theory useful as a library, we have built a fair amount of theory on top of the basic construction. This includes a large number of algebraic lemmas, and enough mathematical analysis to define and prove the main properties of the transcendental functions like exp and sin. We will outline the main results.

2.1 Topology and metric spaces

Recall that a metric on a set S is a function $\rho : S \times S \to \mathbb{R}$ with the following two properties:

1. $\forall x\ y.\ \rho(x, y) = 0 \equiv x = y$

2. $\forall x\ y\ z.\ \rho(y, z) \leq \rho(x, y) + \rho(x, z)$

The latter is usually called the *triangle law*; in the usual metric on $\mathbb{R}^2$ it states that one side of a triangle is no greater than the sum of the other two sides. From the above two properties, it follows that a metric is both nonnegative ($\rho(x, y) \geq 0$) and symmetric($\rho(x, y) = \rho(y, x)$).

A metric is a generalization of the notion of distance between points; in subsequent developments we invariably use the usual metric on $\mathbb{R}$, namely

$$\rho(x, y) = |x - y|$$

but many of the theorems we prove are true in a more general framework. (In fact some are proved in the even more general structure of a topological space.) Since on *any* set we can define the *discrete* metric:

$$\rho(x, y) = (x = y) \to 0 \mid 1$$

we can define a type operator $(\alpha)metric$. (HOL insists that all types be inhabited.)

2.2 Convergence nets

Nets generalize the notions of sequences and pointwise limits, and allow us to avoid proving twice many theorems like

$$x \to l_x \wedge y \to l_y \implies (\lambda n.\ x_n + y_n) \to (l_x + l_y)$$

Nets are simply functions out of a set with a directed partial order, i.e. one where for any x, y there is a z with $z \geq x$ and $z \geq y$. In the case of sequences, the order is just the usual order on the natural numbers. For more details see [13].

2.3 Sequences and series

This consists in part of specializing net theorems, but then some theorems specifically about sequences are added. These include:

- A bounded and monotonic sequence converges. Suppose the sequence is increasing, the other case being analogous. Consider the set $\{x_n | n \in \mathbb{N}\}$. This must have a supremum l such that for any $\epsilon > 0$, there exists an N with $|x_N - l| < \epsilon$. But because the sequence is increasing, this means that $\forall n \geq N.\ l - \epsilon < x_n \leq l$, so the sequence in fact converges to l.

- Every sequence has a monotonic subsequence. Call n a *terrace point* if we have $\forall m > n.\ x_m \leq x_n$. If there are infinitely many such terrace points, we can just form a decreasing sequence by successively picking them. If on the other hand there are only finitely many terrace points, then suppose N is the last one (or $N = 0$ if there are none). Now for *any* $n > N$, there is an m with $x_m > x_n$ (otherwise n would be a terrace point). Hence we can choose a (strictly) increasing subsequence by repeatedly making such choices.

- Every Cauchy sequence converges. A Cauchy sequence is bounded because for any $\epsilon > 0$, say $\epsilon = 1$, we can find an N such that $\forall n \geq N,\ m \geq N.\ |x_m - x_n| < \epsilon$, so

  $$max(x_0, \ldots, x_{N-1}, x_N + \epsilon)$$

  is an upper bound. Hence we can find a subsequence which is both bounded and monotonic, and hence convergent. But now because of the Cauchy criterion, the limit of the subsequence is in fact a limit for the sequence itself.

Then we move on to infinite series, defined as limits of a finite sums. Additional theorems proved are mainly tests for convergence such as the comparison test and ratio test.

2.4 Limits, continuity and differentiability

Part of this is specializing the net theorems again; next we define continuity and differentiability, proving the various combining theorems such as the derivative of a product and the chain rule. A valuable function we define here is a conversion which will automatically apply the chain rule and any particular differentiation theorems it is given in order to prove theorems about differentiating arbitrary expressions. We move on to prove various classical theorems, leading up to Taylor's Theorem:

- A function continuous on a closed interval is bounded. Suppose the function is f, and the interval is $\lambda x.\ a \leq x \leq b$. Consider the set

  $$\{x | a \leq x \leq b + 1 \wedge \exists K.\ \forall x'.\ a \leq x' \leq x \Longrightarrow |f(x')| < K\}$$

This is nonempty (it contains a) and bounded above (by $b+1$). It therefore has a supremum, s. It is clear that $s > b$, proving the theorem, because if $s \leq b$, continuity means we can pick some $\epsilon > 0$ so that for any x' with $x - \epsilon \leq x' \leq s + \epsilon$, f is bounded, contradicting the fact that s is a supremum. (We assume two-sided continuity at a and b, although this is not necessary.)

- A function continuous on a closed interval attains its supremum and infimum. The following slick proof is taken from [2]. Suppose f does not attain its supremum M. Then the function defined by $\lambda x.\ (M - f(x))^{-1}$ is continuous on the interval (a previous theorem about continuity assures us of this because the denominator is never zero), and therefore *it* is bounded, by K say, which must be strictly positive. But this means that we have $M - f(x) \geq K^{-1}$, which is a contradiction because M is a *least* upper bound.

- Rolle's theorem: if f is continuous for $a \leq x \leq b$ and differentiable for $a < x < b$, and in addition $f(a) = f(b)$, then there is some $a < x < b$ with $f'(x) = 0$. We know that f attains its bounds, and in fact its derivative must be zero there, otherwise it would exceed its bounds on one side or the other.

- The Mean Value Theorem states that if f is continuous for $a \leq x \leq b$ and differentiable for $a < x < b$, then there is some $a < x < b$ with $f(b) - f(a) = (b-a)f'(x)$. A proof is easy by applying Rolle's theorem to the function:

$$\lambda x.\ f(x) - (f(b) - f(a))x/(b-a)$$

- A function whose derivative is zero on an interval is constant on that interval. This is an immediate corollary of the Mean Value Theorem.

2.5 Power series

Here we bring together the theories of infinite series and differentiability, proving a few results about power series, in particular that they are characterized by a 'circle of convergence' and within it we can differentiate them term-by-term. This latter result was in fact the most difficult proof in the whole undertaking. Had we been developing analysis for its own sake, we would have proved some general results about uniform convergence. As it is, we prove the result by direct manipulation of the definition of derivative, following Theorem 10.2 in [3]. The theorem requires both the first and second formal derivative series to converge within the radius of convergence. This does in fact follow in general, but we did not need to prove it because the set of power series we are concerned with differentiate to 'each other', so we already have convergence theorems.

We also prove Taylor's theorem in its full infinite series form. In fact it is no longer used in the subsequent development, but could be useful for giving error bounds when truncating infinite series.

2.6 The transcendental functions

The functions *exp*, *sin* and *cos* are defined by their power series expansions (we do not need Taylor's theorem to do this):

$$
\begin{array}{rcl}
exp(x) & = & 1 + x + x^2/2! + x^3/3! + \ldots \\
sin(x) & = & x - x^3/3! + x^5/5! - x^7/7! + \ldots \\
cos(x) & = & 1 - x^2/2! + x^4/4! - x^6/6! + \ldots
\end{array}
$$

We show using the ratio test that the series for *exp* converges, and by the comparison test that the other two do. Now by our theorem about differentiating infinite series, we can show that the derivative of *sin* is *cos*, and so on. Furthermore, a few properties like $cos(0) = 1$ are more or less immediate from the series. Now we are in a position to prove more interesting properties. As an example, to show that $exp(x+y) = exp(x)exp(y)$, consider the function

$$\lambda x.\ exp(x+y)exp(x)^{-1}$$

Our automatic conversion, with a little manual simplification, shows that this has a derivative which is 0 everywhere. Consequently, by our previous theorem, it is constant everywhere. But at $x = 0$ it is just $exp(y)$, so the result follows.

The addition formulas for *sin* and *cos* can also be proved in this way, but we need the slight trick of combining the two as follows to make the procedure work:

$$
\begin{array}{ll}
\lambda x. & (sin(x+y) - (sin(x)cos(y) + cos(x)sin(y)))^2 + \\
 & (cos(x+y) - (cos(x)cos(y) - sin(x)sin(y)))^2
\end{array}
$$

By the same method as above we can prove that this function is always zero, and the addition formulas follow easily. Other function like *ln* and *asn* are defined as the inverses of *exp* and *sin*, and their basic properties are derived. Periodicity of the trigonometric functions follows from the addition formulas and the fact that there is a least $x > 0$ with $cos(x) = 0$. This latter fact is proved by observing that $cos(0) > 0$ and $cos(2) < 0$. Since $sin(x)$ is positive for $0 < x < 2$, *cos* is strictly decreasing there, so by the Intermediate Value Theorem there must be a unique zero. (These proofs involve some fiddly manipulations of the first few terms of the series for *sin* and *cos*.) The zero is of course $\pi/2$, and this serves as our definition of π.

3 Applications

There seem to be several promising areas of application, which have only been partially investigated so far.

3.1 Verification of floating-point hardware

This seems an ideal area for theorem-proving; it is hard to see how one could verify by model-checking a circuit to calculate logarithms, for example. We have already done a verification of a toy floating-point square root circuit. (By toy we mean that it uses a simple floating point format rather than the full IEEE [10] standard with special cases and denormalized numbers. Also our circuit is probably inefficient compared with a commercial design.) It is hoped in the future to do verifications of more realistic circuits and/or circuits for more complicated functions like sin.

3.2 Numerical work

It would be quite easy to program HOL to produce mathematical tables with high assurance (for human consumption or insertion of constants into hardware or software). After all, this is what Babbage designed his Difference Engine to do! More generally, there are many areas of application in the error analysis of numerical methods [18].

3.3 Computer Algebra

Computer algebra systems are widely used by applied mathematicians and others. In view of their complexity it seems likely that they include bugs, or consciously implement rules of a theoretically dubious nature. It would be an interesting project to build a computer algebra system which works by proof on top of a programmable theorem-prover like HOL which offers a fairly high degree of security (no complex decision procedures built in). We have taken a step towards this goal with our expression-differentiation conversion; this is rather a trivial computer algebra system, but it represents a start!

3.4 Hybrid Systems

Various formalisms for dealing with hybrid systems have been proposed [15]. Mixing the discrete and the continuous may well require use of the real numbers.

4 Conclusion and related work

As far as we are aware, the only previous construction of the classical reals in a computer theorem prover was by Jutting [12], who translated Landau's book [14] into Automath.

The definition of the constructive reals has been done in NuPRL, with a proof of their completeness, i.e. that every Cauchy sequence converges [5]. Much of the construction has also been done in the LEGO prover, as well as some work on completing a general metric space [11].

The full construction described here, from $\mathbb{N}$ to $\mathbb{R}$, took about two weeks, but it would have taken much longer without careful selection of strategy. In particular, it seems that a quotient procedure tends to be much easier that picking canonical elements. Previous constructions of the integers from the naturals have been made by others using canonical representations, and their greater complexity seems to bear out this point.

The additional work on mathematical analysis took several months, on and off. Analytical proofs tend to have quite a lot of minor details which need to be filled in, particularly tedious bits of arithmetic reasoning. In the near future, derived decision procedures will become available which will greatly ease this sort of task without compromising the security of the system.

The fact that the HOL system can be programmed easily (the presence of ML rather than just an ad-hoc macro language, together with the simplicity of the underlying term structures) is a major advantage. Implementing procedures like the quotient types function and the differentiation conversion would otherwise be very difficult. Furthermore, even though some proofs are long and tedious, one can always get there with a little patience because of the system's great flexibility.

Here are some indicators of the 'size' of the proof. The complete theories described here generated 167608 primitive inferences, and took 92 minutes to build on a 48Mb SPARCserver. The total ML source is 10080 lines, including comments. The parts leading just to the real number "axioms" generated 49017 primitive inferences, took 14 minutes to build, and consisted of 2098 lines of ML.

The nature of the underlying logic has some impact on the formalization. We have already discussed the effect of total functions on the division operation. They also mean that certain traditional notations are less useful; for example we cannot infer from $lim(x_n) = l$ that the sequence (x_n) actually tends to a limit; the lim function is *always* defined. Instead, we tend to use relational notations like $x \to l$ in preference. In most cases this is no handicap; indeed it is often clearer.

If analysis were to be taken further, some extensions to the logic would be convenient. For example, it is difficult to reason in a clean way about arbitrary n-ary Cartesian products without some simple form of dependent types. Furthermore, subtypes would allow more transparent embedding of one number system in another. It is perhaps difficult to know how difficult some parts of mathematics are to formalize without actually trying. Devices like adjoining infinities to the real line are easily waffled over, but perhaps not so easily formalized.

Formalization itself can be clarifying. One is forced to be less sloppy about things like variable binding (for example, what does $f'(x) = g(x)$ mean?). To avoid proving almost identical theorems twice, we saw the need for some common framework for limiting processes before we were actually aware that such frameworks (nets and filters) already existed. This constitutes an example of how abstraction can be driven by mundane considerations of economy, rather than of beauty.

Applications have only just got off the ground, but it seems to be fertile territory for computer-aided verification by theorem proving.

5 Acknowledgements

I would like to thank Mike Gordon for his help and encouragement, including his patient debugging when I was learning HOL, and for drawing my attention to Kelley's discussion of convergence nets [13]. Thomas Forster and Larry Paulson gave me some useful advice about constructing the reals. Everyone in the group at Cambridge has contributed to making such a pleasant and stimulating work environment, and many of them, too numerous to mention here, have helped my work in one way or another. I am specially grateful to Richard Boulton, Jim Grundy, Monica Nesi and Eike Ritter, who read an early version of this paper and made many helpful comments. Thanks are also due to the Science and Engineering Research Council for financial support.

References

[1] N. G. de Bruijn, *Defining Reals Without the Use of Rationals*, Indagationes Mathematicae, vol. 38, no. 2, 1976.

[2] J. C. Burkill, *A First Course in Mathematical Analysis*, Cambridge 1962.

[3] J. C. Burkill, H. Burkill, *A Second Course in Mathematical Analysis*, Cambridge 1970.

[4] C. W. Burrill, *Foundations of Real Numbers*, McGraw-Hill 1967.

[5] J. Chirimar, D. J. Howe, *Implementing Constructive Real Analysis*, preprint 1992.

[6] L. W. Cohen, G. Ehrlich, *Structure of the real number system*, Van Nostrand 1963.

[7] J. H. Conway, *On Numbers and Games*, Academic Press 1976.

[8] F. Faltin, N. Metropolis, B. Ross, G.-C. Rota, *The Real Numbers as a Wreath Product*, Advances in Mathematics, vol. 16, pp. 278-304, 1975.

[9] J. S. Golan, *The Theory of Semirings with Applications in Mathematics and Computer Science*, Longman 1992.

[10] IEEE, *Standard for Binary Floating Point Arithmetic*, ANSI/IEEE Standard 754-1985.

[11] C. Jones, *Completing the Rationals and Metric Spaces in LEGO*, Proceedings of the 2nd Workshop on Logical Frameworks, Edinburgh 1991.

[12] L. S. Jutting, *Checking Landau's "Grundlagen" in the Automath System*, PhD thesis, Eindhoven University of Technology 1977.

[13] J. L. Kelley, *General Topology*, Van Nostrand 1955.

[14] E. Landau, *Foundations of analysis*, Chelsea 1951.

[15] A. Pnueli, *From Timed To Hybrid Systems*, preprint 1992.

[16] R. R. Stoll, *Set theory and logic*, Dover 1979.

[17] H. A. Thurston, *The number system*, Blackie 1956.

[18] P. J . L. Wallis, *Improving floating-point programming*, Wiley 1990.