Efficient Computation with Dedekind Reals

Andrej Bauer[1]

Faculty of Mathematics and Physics
University of Ljubljana
Ljubljana, Slovenia

Abstract

Cauchy's construction of reals as sequences of rational approximations is the theoretical basis for a number of implementations of exact real numbers, while Dedekind's construction of reals as cuts has inspired fewer useful computational ideas. Nevertheless, we can see the computational content of Dedekind reals by constructing them within Abstract Stone Duality (ASD), a computationally meaningful calculus for topology. This provides the theoretical background for a novel way of computing with real numbers in the style of logic programming. Real numbers are defined in terms of (lower and upper) Dedekind cuts, while programs are expressed as statements about real numbers in the language of ASD. By adapting Newton's method to interval arithmetic we can make the computations as efficient as those based on Cauchy reals. The results reported in this talk are joint work with Paul Taylor.

Keywords: Dedekind real numbers, Abstract Stone Duality, exact real number computation.

1 Introduction

The material presented in this talk is joined work with Paul Taylor. We thank Danko Ilik, Matija Pretnar, and Chris Stone for taking part in fruitful discussions during our meeting in June 2008 in Ljubljana.

Implementations of (exact) real numbers are usually based on Cauchy's construction of real numbers as sequences of rational approximations. Typically, a real number is represented as stream of signed binary digits $-1, 0, 1$, or a sequence of intervals with dyadic[2] endpoints whose widths converge to zero. The preference for Cauchy reals is also visible in Bishop's constructive analysis [2] and Type Two Effectivity [9].

Dedekind's construction of reals as cuts is well known, but used less often in real number computation. This may be so because it is less natural to compute with sets than with sequences. Nevertheless, Dedekind's construction may be carried out in a constructive setting, such as type theory, constructive set theory, or topos theory.

[1] Email: Andrej.Bauer@andrej.com

[2] A *dyadic* rational is one of the form $m/2^k$.

This paper is electronically published in
Electronic Notes in Theoretical Computer Science
URL: www.elsevier.nl/locate/entcs

The question is what precisely the computational content of such a construction is, and how to use it to compute with real numbers.

Recall that a *Dedekind cut* $\langle L, U\rangle$ is a pair of subsets $L, U \subseteq \mathbb{Q}$, called respectively the *lower* and *upper* cut, which are

(i) *(lower and upper) rounded:* $q \in L \iff \exists r\,.\,(q < r \wedge r \in L)$ and $q \in U \iff \exists r\,.\,(r < q \wedge r \in U)$,

(ii) *bounded* (or inhabited): $\exists q\,.\,q \in L$ and $\exists q\,.\,q \in U$,

(iii) *disjoint:* $\neg(q \in U \wedge q \in L)$,

(iv) *located:* $q < r \implies q \in L \vee r \in U$.

The set/type/object of *Dedekind reals* $\mathbb{R}$ consists of all Dedekind cuts. Depending on the underlying foundational setting, $\mathbb{R}$ may enjoy all or some of the properties stated in the following *definition* of Dedekind reals.

Definition 1.1 A *Dedekind real line* is

(i) an overt Hausdorff space, which is

(ii) an ordered Archimedean field,

(iii) Dedekind complete, and such that

(iv) the closed interval $[0, 1]$ is compact.

Few would dispute that the reals are overt [3] and Hausdorff, or that they form an Archimedean ordered field. Dedekind completeness says that nothing is gained by iterating Dedekind's construction: a pair $\langle L, U\rangle$ of sets of *reals* $L, U \subseteq \mathbb{R}$ determines a unique real $x \in \mathbb{R}$ such that $\ell < x < u$ for all $\ell \in L$, $u \in U$. Compactness of the closed interval is not universally accepted, partly because there are several definitions of compactness, and partly because of difference of opinion in what constructive mathematics is. Anyhow, it is not our purpose here to debate this issue. We are taking Definition 1.1 as a starting point from which we want to develop a novel way of computing with real numbers.

2 Dedekind Reals in Abstract Stone Duality

Our foundational setting is Abstract Stone Duality (ASD), which is a calculus for topology developed by Paul Taylor [7]. It gives topological notions and constructions a computational meaning. For background on ASD, the construction of Dedekind reals in ASD, and verification that they satisfy Definition 1.1 we refer to [1]. Here we summarize those points which are relevant for the task at hand. Furthermore, in order to simplify and focus the presentation we limit our attention to a "first-order" part of ASD and the of space of reals in the style of [8]. [4]

The terms of our language are arithmetical expressions involving rational numbers, variables ranging over $\mathbb{R}$, and Dedekind cuts as described below. The rationals may be replaced by any dense Archimedean subring of $\mathbb{R}$ with decidable order, and

[3] You may never have heard of the topological property of overtness because classically all spaces are overt.

[4] In particular, we are omitting higher-order exponentials $\Sigma^{\Sigma^{\cdots}}$, formation of subspaces, and general recursion on natural numbers.

in fact our prototype implementation uses dyadic rationals. If (the full) ASD calculus proves the existence of a term $t(x)$ with a free variable x, we may adjoin a term constructor t. This allows us to introduce elementary functions, such as roots, exponentials and trigonometric functions.

The (topo)logical formulae are built from constants for falsehood $\bot$ and truth $\top$, strict inequalities $e_1 < e_2$, conjunctions $\phi_1 \wedge \phi_2$, disjunctions $\phi_1 \vee \phi_2$, existential quantifications of the form $\exists\, x \in \mathbb{R} \,.\, \phi$ and $\exists\, x \in [a, b] \,.\, \phi$, and universal quantifications $\forall\, x \in [a, b] \,.\, \phi$, where a and b are rational constants. These operations are monotone and preserve open sets, whereas equality $=$, inequality $\leq$, negation, implication, and universal quantification $\forall\, x \in \mathbb{R} \,.\, \phi$ are excluded because they do not. Thus the logical formulae have a double reading, as truth values and as open sets. We introduce the shorthand $e_1 \neq e_2$ for $e_1 < e_2 \vee e_2 < e_1$, and $e_1 =_\epsilon e_2$ for $-\epsilon < e_1 - e_2 < \epsilon$.

If we prove that a pair of predicates δ, υ with a free variable x form a Dedekind cut,[5] then we may introduce the real number $\mathsf{cut}(x, \delta, \upsilon)$. For instance, $\sqrt{2}$ is expressed as

$$\mathsf{cut}(x, (x < 0 \vee x^2 < 2), (0 < x \wedge 2 < x^2)).$$

More generally, if an expression $f(x)$ is strictly increasing in the interval $[a, b]$ and $f(a) < 0 < f(b)$ then we may define the unique root of f on $[a, b]$ as $\mathsf{cut}(x, \delta(x), \upsilon(x))$ where

$$\delta(x) \;\equiv\; x < a \vee (x < b \wedge f(x) < 0)$$

and

$$\upsilon(x) \;\equiv\; b < x \vee (a < x \wedge 0 < f(x)).$$

A cut may always be eliminated because $\phi(\mathsf{cut}(x, \delta, \upsilon))$ is equivalent to[6]

$$\exists\, d, u \in \mathbb{R} \,.\, (\delta(d) \wedge d < u \wedge \upsilon(u) \wedge \forall\, x \in [d, u] \,.\, \phi(x)) \,.$$

The topological reading of this equivalence is that $\mathbb{R}$ is locally compact.

Even though the language is rather restricted we can still express interesting facts about real numbers, such as local compactness. We refer to [8] for elaborate examples of real analysis in ASD, including the intermediate value theorem, connectedness of the real line, and maximum of a function on a closed interval.

3 Computing with Dedekind Reals

In ASD the Dedekind reals are constructed in several steps. First the *ascending* $\underline{\mathbb{R}}$ and *descending* reals $\overline{\mathbb{R}}$ are defined, the former being just the lower and the latter just the upper cuts, and they need not be bounded or inhabited so that $-\infty$ and $+\infty$ are included as elements, too. Then we define the *interval lattice* $L = \underline{\mathbb{R}} \times \overline{\mathbb{R}}$ whose elements can be thought of as intervals $[a, b]$, except that the endpoints may be infinite or even back to front, i.e., $b < a$. After several more steps of construction the reals $\mathbb{R}$ are formed as a subspace of L. Consequently, the *monadic principle* of ASD guarantees that open subsets of $\mathbb{R}$ may be canonically[7] extended to open subsets

[5] By this we mean that the *extensions* $L = \{x \in \mathbb{R} \mid \delta\}$ and $U = \{x \in \mathbb{R} \mid \upsilon\}$ form a cut.

[6] We assume that δ and υ do not contain any variables which are bound in ϕ.

[7] The extension procedure itself is continuous, and effective.

of L. Similarly, the arithmetical operations are extended from $\mathbb{R}$ to L using Moore's *interval arithmetic* [6] and *Kaucher multiplication* [4]. It therefore makes sense to apply a formula $\phi(x)$ to an interval $[a, b]$. When $x \in [a, b]$ the value of $\phi([a, b])$ approximates that of $\phi(x)$, which allows us to replace real numbers with intervals (with dyadic endpoints). This is the basis for a computational procedure, which we describe briefly. For simplicity we limit attention to the bounded existential quantifier $\exists\, x \in [a, b] \,.\, \phi$, and omit $\exists\, x \in \mathbb{R} \,.\, \phi$ from further discussion.

Given an ASD formula ϕ, we compute *lower* and *upper approximants* ϕ^- and ϕ^+. These are simple formulae whose truth values can be easily determined and such that we have the logical entailments

$$\phi^- \implies \phi \implies \phi^+.$$

If ϕ^- is true or ϕ^+ is false then the logical value of ϕ is known and the computation stops. The approximants are computed as follows:

$$\begin{aligned}
\top^- &= \top & \top^+ &= \top \\
\bot^- &= \bot & \bot^+ &= \bot \\
(\phi_1 \wedge \phi_2)^- &= \phi_1^- \wedge \phi_2^- & (\phi_1 \wedge \phi_2)^+ &= \phi_1^- \wedge \phi_2^+ \\
(\phi_1 \vee \phi_2)^- &= \phi_1^- \vee \phi_2^- & (\phi_1 \vee \phi_2)^+ &= \phi_1^+ \vee \phi_2^+ \\
(\forall\, x \in [a, b] \,.\, \phi)^- &= \phi([a, b])^- & (\forall\, x \in [a, b] \,.\, \phi)^+ &= \phi(\tfrac{a+b}{2})^+ \\
(\exists\, x \in [a, b] \,.\, \phi)^- &= \phi(\tfrac{a+b}{2})^- & (\exists\, x \in [a, b] \,.\, \phi)^+ &= \phi([b, a])^+.
\end{aligned}$$

To compute the lower and upper approximant of $e_1 < e_2$ we evaluate both sides of inequality using interval arithmetic. This gives us an interval inequality $[a_1, b_1] < [a_2, b_2]$. If $b_1 < a_2$ then the lower approximant is $\top$, if $b_2 < a_1$ then the upper approximant is $\bot$, otherwise both approximants are just $e_1 < e_2$.

Notice how we substitute an interval $[a, b]$ for a real variable x in the lower approximant for universal quantification, and even a *back to front* interval $[b, a]$ in the upper approximant for existential quantification.

If the lower approximant of $\exists\, x \in [a, b] \,.\, \phi$ is true, then we record $\frac{a+b}{2}$ as a *witness* of truth. Dually, when the upper approximant of $\forall\, x \in [a, b] \,.\, \phi$ is false, $\frac{a+b}{2}$ is a witness of falsehood. The computational procedure collects the witnesses and outputs them if so desired. For instance, a witness for $\exists\, x \in [a, b] \,.\, f(x) =_\epsilon 0$ is a number $q \in [a, b]$ such that $-\epsilon < f(q) < \epsilon$. This way we can handle general equation solving, and even equations with parameters ranging over closed intervals, e.g., $\forall\, x \in [a, b] \,.\, \exists\, y \in [c, d] \,.\, f(x, y) =_\epsilon 0$ would produce a list of witnesses, showing how the solution y depends on the parameter x.

When the truth value of ϕ cannot be determined from its approximants, ϕ is *refined* to an equivalent formula ϕ', and the procedure repeats with ϕ'. Refinement performs a number of basic optimizations such as $\phi \wedge \bot = \bot$ and $\phi \vee \bot = \phi$. The most interesting bit is refinement of quantifiers. An existential $\exists\, x \in [a, b] \,.\, \phi$ is refined to

$$(\exists\, x \in [a, \tfrac{a+b}{2}] \,.\, \phi) \vee (\exists\, x \in [\tfrac{a+b}{2}, b] \,.\, \phi)$$

and a universal $\forall x \in [a, b] . \phi$ is refined to

$$(\forall x \in [a, \tfrac{a+b}{2}] . \phi) \wedge (\forall x \in [\tfrac{a+b}{2}, b] . \phi).$$

Splitting intervals like this amounts to a search procedure akin to the familiar idea of solving equations with *bisection*. Unfortunately, this is just about the slowest officially recognized numerical method, because it only adds one bit of precision in each iteration. To counter this, we use Newton's interval method, which may roughly *double* the number of precise bits in each iteration. The basic idea is as follows.

Given an inequality $f(x) < 0$ on the interval $[a, b]$ we would like to estimate the region in which the inequality definitely holds and definitely fails. Let d_1 and d_2 be lower and upper Lipschiz constants for f so that we have, for all $x \in [a, b]$,

$$d_1(x - a) \leq f(x) - f(a) \leq d_2(x - a).$$

The constants d_1 and d_2 are computed by symbolically differentiating f and evaluating the derivative at the interval $[a, b]$. We may approximate $f(x) < 0$ and $0 \leq f(x)$ by linear inequalities $d_2(x - a) + f(a) < 0$ and $0 \geq f(x)$ by $0 \leq d_1(x - a) + f(a)$, respectively. From these we may easily compute subintervals I_1 and I_2 of $[a, b]$ on which the original inequality $f(x) < 0$ definitely holds and fails, respectively. If for example, the inequality appears inside an existential quantifier, $\exists x \in [a, b] . f(x) < 0$, and I_1 is non-empty then the search may terminated. If I_1 is empty then the search proceeds on the complement of I_2.

When we combine Newton's interval method with interval splitting, the result is a robust procedure which makes progress by halving intervals until it focuses on sufficiently small interval for Newton's method to become efficient.

It should be pointed out that the computation may diverge in "borderline" cases such as $\forall x \in [a, b] . 0 < f(x)$ with f positive everywhere except in a single point $x_0 \in [a, b]$. In the neighborhood of x_0 the procedure keeps refining the universal quantifier. In any case, we cannot expect a full decision procedure (certainly not with elementary functions and general recursion added to the language) neither do we want one. Our objective is to provide a high-level programming language for real number computation.

Our prototype implementation is written in Objective Caml [5] and uses the MPFR library for fast dyadic computations [3]. The performance is promising, but many issues remain for future work. In particular, we intend to investigate Newton's interval method in several variables, and take advantage of the obvious opportunities for parallel computation.

References

[1] Bauer, A. and P. Taylor, *The Dedekind reals in abstract Stone duality*, Mathematical Structures in Computer Science (2008), accepted for publication.

[2] Bishop, E. and D. Bridges, "Constructive Analysis," Grundlehren der math. Wissenschaften **279**, Springer-Verlag, 1985.

[3] Hanrot, G., V. Lefèvre, P. Pélissier and P. Zimmermann, "The MPFR Library," INRIA, `http://www.mpfr.org/`.

[4] Kaucher, E., *Interval analysis in the extended interval space IR*, in: G. Alefeld and R. Grigorieff, editors, *Fundamentals of Numerical Computation*, Computing. Supplementum **2** (1980), pp. 33–49.

[5] Leroy, X., D. Doligez, J. Garrigue, D. Rémy and J. Vouillon, *The Objective Caml system, documentation and user's manual - release 3.08*, Technical report, INRIA (2004).

[6] Moore, R., "Interval Analysis," Automatic Computation, Prentice Hall, 1966.

[7] Taylor, P., *Subspaces in abstract Stone duality*, Theory and Applications of Categories **13** (2002), pp. 300–366.

[8] Taylor, P., *A lambda calculus for real analysis*, in: *Computability and Complexity in Analysis*, 2005.

[9] Weihrauch, K., "Computable Analysis," Springer, Berlin, 2000.