A certificate-based approach to formally verified approximations
Florent Bréhard, Assia Mahboubi, Damien Pous

To cite this version:
Florent Bréhard, Assia Mahboubi, Damien Pous. A certificate-based approach to formally verified approximations. 2019. hal-02088529

HAL Id: hal-02088529
https://hal.laas.fr/hal-02088529
Submitted on 3 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
A certificate-based approach to formally verified approximations

Florent Bréhard, Assia Mahboubi, Damien Pous

LIP, LAAS, CNRS, Inria

Abstract

We present a library to verify rigorous approximations of univariate functions on real numbers, with the Coq proof assistant. Based on interval arithmetic, this library also implements a technique of validation a posteriori based on the Banach fixed-point theorem. We illustrate this technique on the case of operations of division and square root. This library features a collection of abstract structures that organise the specification of rigorous approximations, and modularise the related proofs. Finally, we provide an implementation of verified Chebyshev approximations, and we discuss a few examples of computations.

2012 ACM Subject Classification General and reference → General literature; General and reference

Keywords and phrases approximation theory, Chebyshev polynomials, Banach fixed-point theorem, interval arithmetic, Coq

1 Introduction

While numerical analysis offers sophisticated computational methods to solve various function space problems, the numerical errors caused by floating-point computations, discretisations or finite iterations, are a major concern in domains like safety-critical engineering or computer assisted proofs in mathematics. To address these issues, rigorous numerics [34] provides algorithms to compute validated enclosures of the exact solution. However, their correctness is ensured by pen-and-paper mathematical proofs. In particular, there is no guarantee concerning their implementations.

In this regard, formal proof offers the highest level of confidence. Several noteworthy works use formally proved rigorous numerics to completely formalise highly nontrivial mathematical results, like the Flyspeck project [16] for the Kepler conjecture or the formal verification [19] of the computer-aided proof of the Lorenz attractor [33]. However, those methods often require intensive computations, which rapidly becomes restrictive inside proof assistants. In the context of formal verification, certificate-based methods is an appealing strategy [1]. It consists in discharging part of the computation work load to external oracles, while correctness remains guaranteed via a posteriori validation steps performed inside the proof assistant. This approach has mostly been used for the purpose of verifying symbolic computations, e.g. primality proofs [15], but we illustrate here how it can also be used in the context of rigorous numerical analysis.

Interval arithmetic. Invented in the 60s by Moore [29] and significantly developed in the 80s by Kulisch et al., interval arithmetic is an essential building block of rigorous numerics. The key idea consists in using real intervals with representable endpoints (e.g., floating-point numbers) as rigorous enclosures of real numbers, and providing operations preserving correctness. For example, from $\pi \in [3.1415, 3.1416]$ and $e \in [2.7182, 2.7183]$, one obtains $\pi + e \in [3.1415, 3.1416] \oplus [2.7182, 2.7183] = [5.8597, 5.8599]$. Efficient implementations are available, as MPFI [30], INTLAB [31], C-XSC [24], ARB [20]. The CoqInterval library [28] moreover provides a fully verified implementation inside the Coq proof assistant.

Rigorous Chebyshev approximations. Interval arithmetic is however not a panacea, and replacing all operations on real numbers by interval ones should always be considered with caution: the dependency phenomenon may lead to disastrous over-approximations. In
such cases, higher order methods such as rigorous polynomial approximations (RPAs) are preferable. A pioneer work is that of Berz and Makino on Taylor models [4]. Those provide not only a polynomial, but also a remainder s.t. the latter contains the difference between the former and the represented function. Since then, efforts were made to clarify the definition of RPAs and extend them to other bases, in particular the Chebyshev basis [10, 22], due to their far better approximation properties than Taylor expansions [32].

On the formal proof side, the CoqInterval library includes an implementation of Taylor models called CoqApprox [27], allowing in particular for an automated rigorous evaluation procedure of definite integrals inside Coq [26]. Unfortunately, an equally accomplished equivalent with Chebyshev approximations does not exist now. Our contribution is a first step towards a formally proved counterpart of the popular Chebfun package [13] for MATLAB.

**Fixed-point based a posteriori validation.** Some operations in function spaces admit straightforward self-validating algorithms by replacing all operations in \( \mathbb{R} \) by interval ones. Unfortunately, more complicated operations (e.g., division, square root, differential equations) face several obstructions: the intervals may fail to give sufficiently tight enclosures, bounds for the remainders may be unknown, or only asymptotic, or depend on noneffective quantities.

In such cases, a posteriori validation techniques are an attractive alternative, widely used in rigorous numerics. They consist in reconstructing afterwards an error bound for a candidate approximation. Dating back from the works of Kantorovich about Newton’s method, they gained prominence with the rise of modern computers and were applied to numerous functional analysis problems [23, 36, 35, 25]. Even more recently, those methods were used to compute RPAs for solutions of linear ODEs [2, 8]. Broadly speaking, the function of interest is characterised as a fixed-point of a contracting operator, from which an error bound is recovered thanks to the Banach fixed-point theorem [3, Thm. 2.1]. Such techniques are of special interest for formal verification, for they allow one to rely on efficient but untrusted external tools while keeping the trusted codebase small: it suffices to formalise the theory about contracting operators and provide means of computing with those operators.

**Contributions and outline.** We present a Coq library that makes it possible to compute rigorous Chebyshev approximations of functions on reals. We support basic operations like multiplication or integration in the standard way. For more complex operations like division and square root, we resort to a posteriori validation techniques, thus making a first step towards a potential cooperation between external numerical tools and Coq.

We use the interval arithmetic provided by CoqInterval, but we design our abstractions for RPAs from scratch: this allows us to experiment with different design choices, with more flexibility. We first describe the main lines of the hierarchy (Section 2): we rely on canonical structures to abstract over the concrete implementation details of interval arithmetic, and we use them to denote both real valued functions and their rigorous approximations. We also abstract away from the concrete basis for approximations, to work in the future with different bases, even non polynomial ones (e.g., Bessel functions). We provide instances for the monomial and Chebyshev bases, the latter being described in Section 3.

The main theorem we need to perform a posteriori validation is the Banach fixed-point theorem, whose formalisation is described in Section 4. We show in Section 5 how to apply this theorem to compute rigorous approximations for division and square root using Newton-like operators. We finally discuss the benefits of our approach on two examples (Section 6): RPAs for the absolute value function, and verified computation of integrals related to the second part of Hilbert’s 16th problem.
Approximating real numbers and functions

Numerical errors come from the estimation of both real numbers, e.g. using floating-point numbers, and real functions, e.g. using polynomials. Rigorous estimations must take all these uncertainties into account. For this purpose, interval arithmetic provides an explicit enclosure and rigorous polynomial approximations attach an interval to a polynomial approximant, which bounds the method error on a given domain. Note that the coefficients of polynomial approximations are usually themselves obtained from evaluations of the function or of its derivatives, and therefore also subject to numerical errors. A formal library about rigorous approximation thus implements several variants of each operations, on real numbers, floats, intervals, mathematical functions, approximants, etc., whose relationships are made precise in the various layers of specifications. Our library features a small hierarchy of structures which formalises and organises the dependencies between these variants.

2.1 Reals and Intervals

At the bottom of the hierarchy, structure Ops0 collects the operations available on reals, floats, intervals, but also on polynomials and rigorous approximations. It provides the signature of a ring structure, with symbols \(+\), \(-\), \(*\), \(1\) and \(0\) shared by all instances thanks to COQ’s system of canonical structures. Yet the ring equational theory is a priori only available for real numbers. These operations are also those trivially self-validating. A super-structure Ops1 collects other operations required on data-structures used for scalars: reals, floats, interval endpoints, intervals, etc. They are not meant to be implemented on polynomial approximations.

As much as possible, we will work with polymorphic functions like the following one:

Definition f (C: Ops1) (x: C): C := 1 / (1 + sqrt x).

First of all, this allows us to define at once a function on real numbers (here, \(x \mapsto \frac{1}{1 + \sqrt{x}}\)) and a function on intervals, whatever the implementation of intervals. Second, and even more importantly, the corresponding approximation correctness theorem will always hold—by a parametricity meta-result, such a function \(f\) will always satisfy the following lemma:
A certificate-based approach to formally verified approximations

**Lemma** rf: \( \forall C D (T: Rel1 C D), \forall x y, T x y \to T (f x) (f y). \)

This is only a meta-result: we need to provide a proof for each function \( f \); but the proof is always trivial, and we automatise it.

There are however operations which cannot be implemented at this level of abstraction, even if we were to add some operations to the record \( Ops1 \). This is typically the case for division and square root of rigorous approximations, which require operations on intervals that do not make sense on real numbers (e.g., computing the range of a function and checking that it is bounded). In order to define those operations while remaining rather agnostic about the choice of interval implementation, we setup an intermediate layer of abstraction using the structure \( NBH \) (for neighbourhood):

\[
\text{Record } NBH := \{
\begin{align*}
  \text{II:} & \text{> Ops1; } (* \text{abstract intervals} *) \\
  \text{contains:} & \text{ Rel1 II ROps1; } (* \text{containment relation; ROps1 is the Ops1 instance on R} *) \\
  \text{convex:} & \forall X x y, \text{contains } X x \to \text{contains } X y \to \forall z, x \leq z \leq y \to \text{contains } X z \\
  \text{bnd:} & \text{ II } \to \text{ II } \to \text{ II; } (* \text{directed convex hull} *) \\
  \text{is_lt:} & \text{ II } \to \text{ II } \to \text{ bool; } (* \text{strict above test} *) \\
  \text{min, max:} & \text{ II } \to \text{ option II; } (* \text{min, max, if any} *) \\
  \text{bot:} & \text{ II; } (* \text{uninformative, contains all reals} *) \\
  \text{bndE:} & \forall X x, \text{contains } X x \to \forall Y y, \text{contains } Y y \to \forall z, x \leq z \leq y \to \text{contains } (\text{bnd } X Y ) z \\
  \text{is_ltE:} & \forall X Y, \text{wreflect } (\forall x y, \text{contains } X x \to \text{contains } Y y \to x < y) (\text{is_lt } X Y) \\
  \text{minE, maxE, botE:} & \ldots 
\end{align*}
\]

We will also make use of the two following derived operations:

**Definition** mag \((N: NBH) (X: \text{II}): \text{option II := max } (\text{abs X}).\)

**Definition** sym \((N: NBH) (X: \text{II}): \text{II := let X := abs X in bnd } (-X) X.\)

The first one approximates the magnitude as an interval, if possible; the second one returns an interval centered in 0 that contains the argument. Note that we assume that intervals are convex. We provide an instance of this structure using the \texttt{CoqInterval} library, using intervals of floating point numbers from the \texttt{Flocq} library \[6\]. It is actually a family of instances indexed by the desired precision.

### 2.2 Abstract functions

The structure \texttt{FunOps} describes inductively the catalogue of expressions that the library can approximate.

\[
\text{Record FunOps (C: Type) := \{ }
\begin{align*}
  \text{funcar:>} & \text{ Ops0; } (* \text{abstract type for functions, and pointwise basic operations} *) \\
  \text{id:} & \text{funcar; } \\
  \text{cst:} & \text{C } \to \text{funcar; } \\
  \text{eval:} & \text{funcar } \to \text{C } \to \text{C; } \\
  \text{integrate:} & \text{funcar } \to \text{C } \to \text{C; } \\
  \text{div':} & \text{nat } \to \text{funcar } \to \text{funcar } \to \text{funcar; } \\
  \text{sqrt':} & \text{nat } \to \text{funcar } \to \text{funcar } \\
\end{align*}
\]

It is parameterised by a type \( \text{C} \) of ground values (typically, reals or intervals); it packages a set of basic operations on some abstract type for functions (pointwise addition, multiplication...), together with operations specific to functions: identity, constant function, evaluation, integration. It also asks for division and square root operations; those have an additional argument which is used to pass parameters to the oracles used in the implementation of those operations (for now, the degree of the interpolants).
When \( C = \mathbb{R} \) this structure is instantiated with the standard operations on \( \mathbb{R} \rightarrow \mathbb{R} \) (ignoring the extra parameters for division and square root); our main goal is to provide instances with intervals for \( C \), with which it is possible to perform computations. Like for ground values, the structure \texttt{FunOps} makes it possible to write polymorphic functions like:

```plaintext
Definition g (C: Ops1) (F: FunOps C): F :=
let f: F := div' 33 1 (1 + sqrt' 33 id ) in
let a: C := integrate f 0 1 in
pi + id * cst a
```

Such a declaration defines at the same time a function on reals \((x \mapsto \pi + x \int_0^1 \frac{dt}{1+\sqrt{t}})\) and approximations of it, which will be obvious to prove correct whenever the chosen instance \( F \) satisfies appropriate properties. Those instances are obtained using rigorous approximations.

### 2.3 Rigorous Approximations

Approximating a function usually consists in projecting this function onto a finite dimension vector space, by expansion on a basis with appropriate properties. For instance, so-called Taylor models [4], are an instance of rigorous polynomial approximation. They attach an interval bounding the remainder to a certain polynomial, in this case represented in monomial basis, so as to describe a set of functions containing the one to be approximated. More generally in this section, a rigorous approximations refers to a linear combination of elements in a suitable basis, packaged with an interval remainder. In the code, we will also use the shorter term models, by analogy with Taylor models.

A basis is described by a family of functions, non necessarily polynomials, indexed by natural numbers, that is a term \( T: \text{nat} \rightarrow \mathbb{R} \rightarrow \mathbb{R} \). The structure \texttt{BasisOps_on} below describes the signature required on a basis \( T \). It is parameterised by the type \( C \) of coefficients; sequences of such coefficients (\( \text{seq} C \)) represent linear combinations of elements of \( T \). Linear operations \((+, -, 0)\) need not be provided since they can be implemented independently from the basis.

The range operation is important: its role is to bound the range on the given domain; it should be as accurate as possible since it is used at many places to compute error bounds in rigorous approximations (e.g., for multiplication and a posteriori validation). We define \texttt{BasisOps} to be a polymorphic function so that we capture with a single object the idealised operations on reals and their concrete implementation with intervals.

#### Record BasisOps_on (C: Type) :=
- \( \text{lo}, \text{hi}: C; \) (* bounds for the domain *)
- \( \text{beval}: \text{seq C} \rightarrow C \rightarrow C; \) (* (efficient) evaluation *)
- \( \text{bmul}: \text{seq C} \rightarrow \text{seq C} \rightarrow \text{seq C}; \) (* multiplication *)
- \( \text{bone}, \text{bid}: \text{seq C}; \) (* constant to 1, identity *)
- \( \text{bprim}: \text{seq C} \rightarrow \text{seq C}; \) (* primitive *)
- \( \text{brange}: \text{seq C} \rightarrow C*C; \) (* range *)

#### Definition BasisOps := ∀ C: Ops1, BasisOps_on C.

Given such operations, we equip \( \text{seq} C \) with the basic operations in \texttt{Ops0}. Then we can define rigorous approximations:

#### Record Model C :=
- \{ \text{pol}: \text{seq C}; \text{rem}: C \}.

Like with \( \text{seq} C \), we equip \texttt{Model C} with the basic operations in \texttt{Ops0}, and then with those from \texttt{FunOps}. For instance, addition, evaluation and integration are defined as follows:

#### Definition madd (C: Ops1) (M N: Model T C): Model T C :=
- \{ \text{pol} := \text{pol M} + \text{pol N}; \text{rem} := \text{rem M} + \text{rem N} \}.

#### Definition meval (C: Ops1) (M: Model T C) (X: C): C :=
- \text{beval (pol M)} X + \text{rem M}.

#### Definition mintegrate (C: Ops1) (M: Model T C) (a b: C): C :=
- \text{let N := bprim (pol M) in beval N b - beval N a + (b-a)*rem M}.
For those relatively simple operations, it suffices to have the basic operations \(\text{Ops1}\) on \(\mathbb{C}\). For other operations like the range of a model, we actually need the additional operations on intervals provided by the structure \(\text{NBH}\):

\[
\text{Definition mrange } (N: \text{NBH}) (M: \text{Model}) := \text{let } (a,b) := \text{brange } (\text{pol } M) \text{ in } \text{bnd } a b + \text{rem } M.
\]

This is also the case for division and square root, which we will discuss in Section 5. All in all, we obtain instances \(\text{FunOps}\) through a construction of the following type:

\[
\text{Canonical Structure } M\text{FunOps } (N: \text{NBH}) (B: \text{BasisOps}): \text{FunOps II}.
\]

It finally remains to show that those operations defined on rigorous approximations properly match the idealised operations on functions over reals. We fix in the sequel an instance \(N: \text{NBH}\) of neighbourhood and basis operations \(B: \text{BasisOps}\), and we write \(\text{Model}\) for \(\text{Model II}\). The central definition to establish this correspondence is the following one, where the function \(\text{eval}\) is the obvious evaluation function for linear combinations of elements of \(T\).

\[
\text{Definition mcontains } (F: \text{Model}) (f: \mathbb{R} \to \mathbb{R}) := \exists p: \text{seq } \mathbb{R}, \text{scontains } (\text{pol } F) p \land \forall x, \text{lo} \leq x \leq \text{hi} \to \text{contains } (\text{rem } F) (f x - \text{eval } T p x)
\]

Intuitively, a model contains a real-valued function \(f\) if it contains a generalised polynomial which is close enough to \(f\) on the domain of the basis. (The binary predicate \(\text{scontains}\) denotes the pointwise extension of the relation \(\text{contains}\) to sequences: in the definition, the real coefficients of \(p\) should be pointwise contained in the interval coefficients of \(\text{pol } F\).

Equipped with this definition, we prove lemmas like

\[
\text{Lemma rmmul: } \forall f g, \text{mcontains } F f \to \text{mcontains } G g \to \text{mcontains } (F \times G) (f \times g).
\]

\[
\text{Lemma rmdiv: } \forall n f g, \text{mcontains } F f \to \text{mcontains } G g \to \text{mcontains } (\text{div' } n F G) (\text{div' } n f g).
\]

\[
\text{Lemma rmintegrate: } \forall f a b, (\forall x, \text{lo} \leq x \leq \text{hi} \to \text{continuous_at } f x) \to \text{mcontains } F f \to \text{contains } A a \to \text{contains } B b \to \text{contains } (\text{integrate } F A B) (\text{integrate } f a b).
\]

Of course, we need assumptions on the basis operations in order to do so. Those assumptions are summarised in the following structure. Recall that \(B: \text{BasisOps}\) provides us with operations \(B R\text{Ops1}\) on reals and operations \(B \text{II}\) on intervals. The structure assumes: 1/ the expected properties on the operations on reals (i.e., efficient evaluation corresponds to evaluation with \(T\), multiplication indeed corresponds to pointwise multiplication under evaluation, etc.); and 2/ a relationship between the operations on reals and on intervals. This separation of concerns is very convenient: the latter containment lemmas are always proved in a trivial way (i.e., automatically), and the former properties do not involve intervals at all, but only real numbers and functions, for which usual mathematical intuitions apply.

\[
\text{Record ValidBasisOps } (N: \text{NBH}) (B: \text{BasisOps}) := (\text{properties of operations on reals } (B R\text{Ops1}) \star)
\]

\[
\text{lloi: } \text{lo} < \text{hi};
\]

\[
\text{bevalE: } \forall p x, \text{beval } p x = \text{eval } T p x;
\]

\[
\text{eval_cont: } \forall p x, \text{continuity_pt } (\text{eval } T p) x;
\]

\[
\text{eval_mul: } \forall p q x, \text{eval } T (\text{bmul } p q) x = \text{eval } T p x \times \text{eval } T q x;
\]

\[
\text{eval_prm: } \forall p a b, \text{eval } T (\text{bprim } p b) b - \text{eval } T (\text{bprim } p a) a = \text{RInt } (\text{eval } T p) a b;
\]

\[
\text{...}
\]

\[
\text{Record ValidBasisOps } (N: \text{NBH}) (B: \text{BasisOps}) := (\text{properties of operations on reals } (B R\text{Ops1}) \star)
\]

\[
\text{rbeval: } \forall p X x, \text{scontains } P p x \to \text{contains } (\text{beval } P X) (\text{beval } p x);
\]

\[
\text{rbeval: } \forall p q X x, \text{scontains } P p x \to \text{contains } Q q x \to \text{scontains } (\text{bmul } P Q) (\text{bmul } p q);
\]

\[
\text{rbeval: } \forall p X x, \text{scontains } P p x \to \text{scontains } (\text{bprim } P) (\text{bprim } p);
\]

\[
\text{...}
\]
3 Arithmetic on Chebyshev polynomials

In order to use the previously described rigorous approximations, it remains to provide implementation of operations (\texttt{BasisOps}) for certain families \( T \) of functions. We provide two instances of them: one for the standard monomial basis, where \( T_n(x) = x^n \), and one described in this section for Chebyshev basis, where \( T_n \) is the \( n \)-th Chebyshev polynomial.

Chebyshev polynomials are defined by the following recurrence, which immediately translates to a fixpoint definition in Coq.

\[
T_0 = 1 \quad T_1 = x \quad T_{n+2} = 2xT_{n+1} - T_n
\]

We can then prove simple properties of those polynomials, for instance:

\[
T_nT_m = (T_{n+m} + T_{m-n})/2 \quad (n \leq m)
\]

\[
T_0 = T_1' \quad T_1 = T_2'/4 \quad T_{n+3} = \frac{T_n'}{2(n+3)} - \frac{T_{n+1}'}{2(n+1)}
\]

\[
T_n(\cos t) = \cos(nt)
\]

Those are proved in a few lines using existing lemmas about derivation and cosine.

3.1 Clenshaw’s evaluation algorithm

The first operation we must implement for \texttt{BasisOps} is the evaluation function (\texttt{beval}). This operation should be polymorphic and as efficient as possible: it will be executed repeatedly when constructing and using rigorous approximations. We use Horner evaluation scheme for the monomial basis, and Clenshaw’s algorithm [14] for Chebyshev, which are both linear in the number of elementary operations. The latter is usually presented as a dynamic programming routine. We translate it into a recursive function with two accumulators:

\[
\text{Fixpoint Clenshaw (C: Ops1) b c (p: seq C) x :=}
\]

\[
\begin{align*}
&\text{match p with} \\
&\quad | [] => c - x*b \\
&\quad | a::q => \text{Clenshaw c (a + 2*x*c - b) q x} \\
&\end{align*}
\]

\[
\text{Definition beval (C: Ops1) (p: seq C) x := Clenshaw 0 0 (rev p) x.}
\]

This code might look mysterious; it is justified by the following invariant on real numbers:

\[
\text{Lemma ClenshawR b c p x: Clenshaw b c p x = eval T (catrev p [c - 2*x*b; b]) x.}
\]

In the right-hand side, \texttt{catrev} is the function that reverses its first argument and concatenate it with the second one. The proof is done by induction in just three lines, using the Coq tactic for ring equations. Correctness (i.e., field \texttt{bevalE} from structure \texttt{ValidBasisOps}) follows.

Note that while the definition of \texttt{beval} can be used with any \texttt{Ops1} structure, its correctness is proved only on reals: the lemma \texttt{ClenshawR} does not hold in every \texttt{Ops1} structure. The behaviour of \texttt{beval} on those structures is specified only through the fact that it respects containments (field \texttt{rbeval} from structure \texttt{ValidBasisOps}, which is proved automatically.)

3.2 Multiplication

Another important operation is multiplication. Again, this operation should be polymorphic, and efficient. A difficulty here is that due to Equation (1), the \( n \)-th coefficient of a multiplication potentially depends on all coefficients of its arguments, not only on the coefficient of smaller rank. We use the following definition, with two auxiliary recursive functions

\[
\begin{align*}
&\text{Fixpoint Clenshaw (C: Ops1) b c (p: seq C) x :=} \\
&\quad | [] => c - x*b \\
&\quad | a::q => \text{Clenshaw c (a + 2*x*c - b) q x} \\
&\end{align*}
\]

\[
\text{Definition beval (C: Ops1) (p: seq C) x := Clenshaw 0 0 (rev p) x.}
\]

This code might look mysterious; it is justified by the following invariant on real numbers:

\[
\text{Lemma ClenshawR b c p x: Clenshaw b c p x = eval T (catrev p [c - 2*x*b; b]) x.}
\]

In the right-hand side, \texttt{catrev} is the function that reverses its first argument and concatenate it with the second one. The proof is done by induction in just three lines, using the Coq tactic for ring equations. Correctness (i.e., field \texttt{bevalE} from structure \texttt{ValidBasisOps}) follows.

Note that while the definition of \texttt{beval} can be used with any \texttt{Ops1} structure, its correctness is proved only on reals: the lemma \texttt{ClenshawR} does not hold in every \texttt{Ops1} structure. The behaviour of \texttt{beval} on those structures is specified only through the fact that it respects containments (field \texttt{rbeval} from structure \texttt{ValidBasisOps}, which is proved automatically.)

3.2 Multiplication

Another important operation is multiplication. Again, this operation should be polymorphic, and efficient. A difficulty here is that due to Equation (1), the \( n \)-th coefficient of a multiplication potentially depends on all coefficients of its arguments, not only on the coefficient of smaller rank. We use the following definition, with two auxiliary recursive functions

\[
\begin{align*}
&\text{Fixpoint Clenshaw (C: Ops1) b c (p: seq C) x :=} \\
&\quad | [] => c - x*b \\
&\quad | a::q => \text{Clenshaw c (a + 2*x*c - b) q x} \\
&\end{align*}
\]

\[
\text{Definition beval (C: Ops1) (p: seq C) x := Clenshaw 0 0 (rev p) x.}
\]

This code might look mysterious; it is justified by the following invariant on real numbers:

\[
\text{Lemma ClenshawR b c p x: Clenshaw b c p x = eval T (catrev p [c - 2*x*b; b]) x.}
\]

In the right-hand side, \texttt{catrev} is the function that reverses its first argument and concatenate it with the second one. The proof is done by induction in just three lines, using the Coq tactic for ring equations. Correctness (i.e., field \texttt{bevalE} from structure \texttt{ValidBasisOps}) follows.

Note that while the definition of \texttt{beval} can be used with any \texttt{Ops1} structure, its correctness is proved only on reals: the lemma \texttt{ClenshawR} does not hold in every \texttt{Ops1} structure. The behaviour of \texttt{beval} on those structures is specified only through the fact that it respects containments (field \texttt{rbeval} from structure \texttt{ValidBasisOps}, which is proved automatically.)

3.2 Multiplication

Another important operation is multiplication. Again, this operation should be polymorphic, and efficient. A difficulty here is that due to Equation (1), the \( n \)-th coefficient of a multiplication potentially depends on all coefficients of its arguments, not only on the coefficient of smaller rank. We use the following definition, with two auxiliary recursive functions
A certificate-based approach to formally verified approximations

Fixpoint mul_pls (C: Ops1) (p q: seq C): seq C :=
match p, q with
| [], _ | _, [] => []
| a::p', b::q' => sadd (a*b::(sadd (sscal a q') (sscal b p'))) (0::0::mul_pls p' q')
end.

Fixpoint mul_mns (C: Ops1) (p q: seq C): seq C :=
match p, q with
| [], _ | _, [] => []
| a::p', b::q' => sadd (a*b::(sadd (sscal a q') (sscal b p'))) (mul_mns p' q')
end.

Definition smul C (p q: seq C): seq C := sscal (1/2) (sadd (mul_mns p q) (mul_pls p q))

(sscal is multiplication of a polynomial by a scalar, and sadd is addition of polynomials—we cannot yet use the standard notation for this operation since we are in the process of defining an Ops0 structure on seq C.) The function mul_pls actually corresponds to multiplication in the monomial basis, it covers the first summand in the right-hand side of (1). The function mul_mns differs only in the fact that the recursive call is not pushed away using two 'cons' operations; it covers the second summand in the right-hand side of (1). Like previously, that smul preserves containments (field rbmul of structure ValidBasisOps) is obvious: this operation only performs a finite sequence of operations preserving containments. Proving that it behaves correctly on reals numbers is more interesting; the key invariant is the following one:

Lemma eval_mul_:
∀ (p q: seq R) n x,
eval_ n p x * eval_ n q x = (eval (mul_mns p q) x + eval (n+n) (mul_pls p q) x)/2.

Here, eval_ n p evaluates p padded with n zeros in front of it. Again, the difficulty is to find the lemma: it is proved in six lines using (1), and correctness of smul on reals immediately follows. Taking primitives in Chebyshev basis follows the same pattern (see Appendix A).

3.3 Range

As mentioned above, we need accurate estimations of the range of a given polynomial in order to be able to compute precise rigorous approximations. This range can always be estimated by evaluating the polynomial on the interval representing the domain (i.e., given p: seq C, compute beval p (bnd lo hi)). This technique is however not sufficient in practice: this tends to produce largely over-estimated bounds. With Chebyshev basis we can proceed differently: indeed, thanks to Equation (3), $T_n$ ranges over $[-1;1]$ on $[-1;1]$. Therefore, the range of a polynomial on $[-1;1]$ can be estimated by using the sum of the absolute values of the coefficients in Chebyshev basis (and actually, we do not need to take the absolute value of the first coefficient since $T_0 = 1$).

Definition range_ (C: Ops1): seq C → C := foldr (fun A X => abs A + X) 0.

Definition range (C: Ops1) (P: seq C): C*C :=
match p with
| [] => (0,0)
| A::Q => let R := range_ Q in (A-R, A+R)
end.

3.4 Rescaling

Putting everything together, we obtain the polymorphic operations chebyshev.basis: BasisOps, which can readily be used to construct rigorous approximations, with the instance MFunOps from Section 2.3. This basis however requires to work on the domain $[-1;1]$ (for estimating the range as explained in the previous section, but also to perform interpolation, see Section 5.1). In order to use it on other domains, we provide a rescaling function that takes a B: BasisOps
and rescales it to a given interval \([a; b]\) using the obvious affine function. We show that this operation preserves validity of basis operations, so that we can use it whenever needed.

4 Formalisation of Banach fixed-point theorem

Banach fixed-point theorem is the cornerstone of the method discussed here.

**Theorem 1 (Banach fixed-point).** Let \((X, \| \cdot \|)\) be a Banach space, an operator \(F : X \to X\), \(h^0 \in X\), and \(\mu, b, r \in \mathbb{R}_+\), satisfying the following conditions:

1. \(\|h^0 - F \cdot h^0\| \leq b\);
2. \(F\) is \(\mu\)-Lipschitz over the closed ball \(B(h^0, r) := \{h \in X \mid \|h - h^0\| \leq r\}\):
   \[
   \forall h_1, h_2 \in X, \ h_1 \in B(h^0, r) \land h_2 \in B(h^0, r) \Rightarrow \|F \cdot h_1 - F \cdot h_2\| \leq \mu \|h_1 - h_2\|;
   \]
3. \(\mu < 1\); 
4. \(b + \mu r \leq r\).

Then \(F\) admits a unique fixed-point \(h^*\) in \(B(h^0, r)\).

This classic result has been formalised in various flavours of logic and proof assistants. In particular, Boldo et al. have provided a formal proof of a version of this fixed-point theorem, based on the Coquelicot library, for the purpose of the formalisation of the Lax-Milgram theorem [5]. Using the same backbone library, we provide a different statement that is more suitable for our effective validation purposes. We describe below its formalisation.

The Coquelicot library formalises topological concepts using filters [7, 17], which we briefly recall here. A filter on a type \(T\) is a collection of collections of inhabitants of \(T\) which is non-empty, upward closed and stable under finite intersections:

```coq
Record Filter (T : Type) (F : (T Prop) Prop) := {
  filter_true : F (fun _ => True);
  filter_and : \forall P Q : T Prop, F P \land F Q \Rightarrow F (fun x => P x Q x);
  filter_imp : \forall P Q : T Prop, (\forall x, P x \Rightarrow Q x) \Rightarrow F P \Rightarrow F Q}.
```

While filters are used to formalise neighbourhoods, balls allow for expressing the relative closeness of points in the space. Balls are formalised using a ternary relation between two points in the carrier type, and a real number, with the following axioms:

```coq
ball : M \to R \to M \to Prop;
ax1 : \forall x (e > 0), ball x e x;
ax2 : \forall x y e, ball x e y \to ball y e x;
ax3 : \forall x y z e1 e2, ball x e1 y \to ball y e2 z \to ball x (e1 + e2) z
```

Two points are called \textit{close} when they cannot be separated by balls:

```coq
Definition close (x y : M) : Prop := \forall eps > 0, ball x eps y.
```

A filter is called a \textit{Cauchy filter} when it contains balls of arbitrary (small) radius:

```coq
Definition cauchy (T : UniformSpace) (F : (T Prop) Prop) := \forall eps > 0, \exists x, F (ball x eps).
```

Finally, a \textit{uniform space} is a type equipped with a ball relation and a \textit{complete space} moreover has a limit operation on filters, which ensures the convergence of Cauchy sequences via the following axioms (where ProperFilter F is equivalent to Filter F \land \forall P, F P \Rightarrow \exists x, P x):

```coq
lim : ((T Prop) Prop) T;
ax1 : \forall F, ProperFilter F \to cauchy F \to \forall eps > 0, F (ball (lim F) eps);
ax2 : \forall F1 F2, F1 \subseteq F2 \to F2 \subseteq F1 \to close (lim F1) (lim F2)
```
The above formal definition of balls does not enforce closedness nor openness. We thus introduced the relation associated with the closure of balls, so as to model closed neighbourhoods:

**Definition** \( \text{cball } x \ y := \forall \ e > 0, \text{ ball } x \ (r+e) \ y. \)

Equipped with this definition, hypothesis (1 ii) of Theorem 1 is formalised as follows:

**Definition** \( \text{lipschitz} \_\text{on} \ (F : U \rightarrow U) \ (mu : \mathbb{R}) \ (\ P : U \rightarrow \text{Prop}) := \forall \ x \ y : U, \forall \ r \geq 0, \ P \ x \rightarrow \ P \ y \rightarrow \text{cball } x \ y \rightarrow \text{cball } (F \ x) \ (\mu \ast r) \ (F \ y). \)

We now sketch our formalised proof, using mathematical notations. We consider a complete space \( X \) and we write \( y \in B(x, r) \) for the formal \((\text{ball } x \ r \ y)\), and \( y \in \overline{B}(x, r) \) for \((\text{cball } x \ y)\). The key notion is that of strongly stable ball:

**Definition 2** (Strongly stable ball). A ball \( \overline{B}(u, r) \) is \( \mu \)-strongly stable for \( F \) if \( F \) is \( \mu \)-Lipschitz on \( \overline{B}(u, r) \) and if there is a non-negative real number \( s \), called the offset, s.t.:

\[
F \cdot u \in \overline{B}(u, s) \quad \text{and} \quad \ s + \mu r \leq r.
\]

**Remark 3** (Stability). For any \( x \) in \( \overline{B}(u, r) \), a strongly stable ball for \( F \), \( F \cdot x \in \overline{B}(u, r) \).

**Remark 4** (Contracting case). When \( 0 \leq \mu < 1 \), for any \( \mu \)-strongly stable ball \( \overline{B}(v, \rho) \), with offset \( \sigma \), \( \overline{B}(F \cdot v, \mu \rho) \) is also a strongly stable ball, with offset \( \mu \sigma \). Moreover, \( \overline{B}(F \cdot v, \mu \rho) \) is included in \( \overline{B}(v, \rho) \).

Assume that \( F \) has a \( \mu \)-strongly stable ball \( \overline{B}(u, r) \) of offset \( s \), with \( \mu < 1 \). In particular, \( F \) is contracting on \( \overline{B}(u, r) \). Consider the sequence of balls defined as follows:

\[
\overline{B}_n = \overline{B}(u_n, r_n) \quad \text{with} \quad u_n = F^n \cdot u \quad \text{and} \quad r_n = r \mu^n
\]

where \( F^n \cdot u \) denotes the iterated images of \( u \) under \( F \). By Remark 4, \((\overline{B}_n)_{n \in \mathbb{N}}\) is a nested sequence of \( \mu \)-strongly stable ball for \( F \), with offset \( s \mu^n \). Let \( \mathcal{F} \) be the family of collections of points in \( U \) defined as:

\[
\mathcal{F} = \{ P \subseteq U \mid \exists \ n, \overline{B}_n \subseteq P \}.
\]

It is a proper filter: \( \mathcal{F} \) contains \( U \), it is obviously upward closed, and for \( P, Q \in \mathcal{F} \), \( P \cap Q \) is also in \( \mathcal{F} \) because \((\overline{B}_n)_{n \in \mathbb{N}}\) is decreasing for inclusion. Thus \( \mathcal{F} \) has a limit \( w \), such that for any \( \varepsilon > 0 \), balls \( \overline{B}_n \) are eventually included in \( B(w, \varepsilon) \). We provide a formal proof of Theorem 5, a reformulation of Theorem 1 using the vocabulary of the COQUELICOT library.

**Theorem 5.** The limit \( w \) of the filter \( \mathcal{F} \) is in \( \overline{B}_0 \), and \( w \) is a fixed point of \( F \). Moreover, \( w \) is close to every other fixed point of \( F \) in \( \overline{B}_0 \).

**Proof.** In this statement “\( w \) is a fixed point of \( F^n \)” means “\( w \) is close to \( F \cdot w \)”. First, \( w \in \overline{B}_n \) for all \( n \). Indeed, for any \( \varepsilon > 0 \), there is an \( m \geq n \) s.t. \( \overline{B}_m \subseteq B(w, \varepsilon) \), and since \( \overline{B}_m \subseteq \overline{B}_n \), \( u_m \in \overline{B}_n \cap B(w, \varepsilon) \). In particular, \( w \in \overline{B}_0 \). It is also clear by stability that \( F \cdot w \in \overline{B}_n \) for all \( n \). Moreover, \( w \) is close to any point \( v \) s.t. \( v \in \overline{B}_n \) for all \( n \) (for any \( \varepsilon > 0 \), choose \( n \) s.t. \( 2\mu^n < \varepsilon \)). Taking \( v := F \cdot w \) proves that \( w \) is a fixed point of \( F \).

Finally, if \( w' \in \overline{B}_0 \) is another fixed point of \( F \), then it follows from an easy induction that \( w' \in \overline{B}_n \) for all \( n \). Hence, the foregoing shows that \( w \) is close to \( w' \).
Strongly stable balls model the requirements set on the untrusted data to be formally verified. They can also be seen as balls centered at the initial point, and large enough to include all its successive iterates, i.e., as instances of the locus at stake in classical presentations of the proof. The version proved by Boldo et al. has a slightly more technical wording, which seems to be made necessary by its further usage in the verification of the Lax-Milgram theorem. Our proof script is significantly shorter, partly because we automate proofs of positivity conditions (for radii of balls) using canonical structures for manifestly positive expressions. But the key ingredient for concision is to make most of the filter device in the proof, and to refrain from resorting to low-level properties of geometric sequences. To the best of our knowledge, the other libraries of formalised analysis featuring a proof of this result, notably Isabelle/HOL and HOL-Light, are based on variant of proof strategy closer to the approach of Boldo et al. than to ours.

5 Newton-like validation operators

The purpose of this section is twofold. We first present the general principle of fixed-point based \textit{a posteriori} validation methods, and more particularly, the use of Newton-like validation operators. Then we apply it to the division and square root of models.

Throughout this section, let \((X, \|\cdot\|)\) denote a Banach space, and \(h^*\) the exact solution of an equation in \(X\). In this article, \(X\) stands for the space \(\mathcal{C}(I)\) of real-valued continuous functions defined over a compact segment \(I = [a, b]\), with the uniform norm \(\|h\| := \sup_{x \in I} |h(x)|\). The division and square root of functions are simple examples of solutions of equations in \(\mathcal{C}(I)\), but there are also differential equations, integral equations, delay equation, etc. The general scheme for Banach fixed-point based \textit{a posteriori} validation methods follows two steps:

1. **Approximation step.** A numerical approximation \(h^c \in X\) of \(h^*\) is obtained by an oracle, which may resort to any approximation method. In particular, this step requires no mathematical assumption and can be executed purely numerically outside the proof assistant, good approximation properties being only desirable for efficiency. In our setting with \(X = \mathcal{C}(I)\), interpolation at Chebyshev nodes (Section 5.1) is an efficient and accurate oracle for a wide range of function space problems.

2. **Validation step.** The initial problem is rephrased in such a way that \(h^*\) is a fixed point of a (locally) contracting operator \(F : X \to X\). An \textit{a posteriori} error bound on \(\|h^c - h^*\|\) is deduced from the Banach fixed-point theorem (Theorem 1).

We thus need to find a contracting operator \(F\) of which \(h^*\) is a fixed point. To this end, we use Newton-like validation methods, which transform an equation \(M \cdot h = 0\) into an equivalent fixed-point equation \(F \cdot h = h\) with \(F\) contracting. More specifically, suppose that \(M : X \to Y\) is differentiable; we use a Newton-like operator \(F : X \to X\) defined as:

\[
F \cdot h = h - A \cdot M \cdot h,
\]

with \(A : Y \to X\) an \textit{injective bounded linear} operator, intended to be close to \((D M_{h^*})^{-1}\).

The operator \(A\) may be given by an oracle and does not need to be this exact inverse (which anyway might be non representable on computers exactly). The mean value theorem yields a Lipschitz ratio \(\mu\) for \(F\) over any convex subset \(S\) of \(X\):

\[
\forall h_1, h_2 \in S, \|F \cdot h_1 - F \cdot h_2\| \leq \mu \|h_1 - h_2\|, \quad \text{with} \quad \mu = \sup_{h \in S} \|DF_h\| = \sup_{h \in S} \|1_{X} - A \cdot D M_h\|,
\]

which is expected to be small over some neighborhood of \(h^*\).

Concretely, in order to apply Theorem 1, one needs to compute the following quantities:
A certificate-based approach to formally verified approximations

a bound \( b \geq \|A \cdot M \cdot h^5\| = \|h^5 - F \cdot h^5\| \);
a bound \( \mu_0 \geq \|1_X - A \cdot DM_h\| = \|DF_{h^5}\| \);
a bound \( \mu'(r) \geq \|A \cdot (DM_h - DM_{h^5})\| = \|DF_h - DF_{h^5}\| \) valid for any \( h \in B(h^5, r) \), and parameterised by a radius \( r \in \mathbb{R}_+ \).

If we are able to find a radius \( r \in \mathbb{R}_+ \) satisfying:

\[
\mu(r) := \mu_0 + \mu'(r) < 1, \quad \text{and} \quad b + r\mu(r) \leq r, \tag{4}
\]

then Theorem 1 guarantees the existence and uniqueness of a root \( h^* \) of \( M \) in \( B(h^5, r) \).

\[\textbf{Remark 6.} \] Finding an \( r \) as small as possible while satisfying (4) may be an nontrivial task for automated validation procedures. For many problems, \( \mu'(r) \) is polynomial, hence conditions (4) are polynomial inequalities over \( r \) (this is called the \textit{radii polynomial approach} \[18\] in rigorous numerics). In our case, division (resp. square root) induces an affine (resp. quadratic) equation, which admits closed form solutions.

5.1 Approximation step: interpolation

Since they are certified a posteriori, (non-rigorous) approximations for division and square root of given models can be obtained using arbitrary numerical techniques. We use interpolation at Chebyshev nodes of the second kind for its efficiency and excellent approximation properties [32].

Ideally, we would implement this operation outside of the proof assistant, in order not to pay the price of an interpreted language. This would however require a lot of work in order to design a proper interface between Coq values and external values (e.g., converting Coq representation of floating points numbers into machine level floating points, and back). Instead, and for now, we implement the oracles inside Coq, as unspecified functions. To this end, we add a field to the structure \texttt{BasisOps_on}, to compute interpolants of a given degree:

\[
\text{interpolate} : \text{nat} \rightarrow (\mathbb{C} \rightarrow \mathbb{C}) \rightarrow \text{seq } \mathbb{C};
\]

We implement this operation for Chebyshev basis using the discrete orthogonality relations on Chebyshev polynomials.

To reduce the price of staying inside Coq for those computations, we exploit the polymorphism built in our framework to perform those computations on floating-point numbers rather than intervals. To this end, we add the following fields to the structure \texttt{NBH}:

\[
\text{FF: Ops1}; \quad (* \text{abstract type for floating points and their operations} *)
\text{I2F: II} \rightarrow \text{FF}; \quad (* \text{conversion from floating points to intervals (midpoint)} *)
\text{F2I: FF} \rightarrow \text{II}; \quad (* \text{conversion from intervals to floating points (singleton)} *)
\]

Equipped with these operations, we can define conversion operations between models (on intervals) and polynomials with floating point coefficients:

\[
\text{Definition mcf (M: Model): seq FF := map I2F (pol M)}.
\text{Definition mfc (p: seq FF): Model := \{ | pol := map F2I p; rem := 0 | \}.}
\]

The field \texttt{FF}, of type \texttt{Ops1}, will moreover make it possible to call the functions \texttt{interpolate} and \texttt{beval} from the basis with \texttt{C=FF}, i.e., to let them operate on floating point numbers. By doing so we do not have to reimplement Clenshaw’s evaluation scheme on floating point numbers.
5.2 Validation step for division

For \( f, g \in C(I) \) with \( g \) nonvanishing over \( I \), the quotient \( f/g \) is the unique root of \( M : h \mapsto gh - f \). Let \( h^\circ \) be a candidate approximation given by the approximation step. Constructing the Newton-like operator \( F \) requires an approximation \( A \) of \( (DM_h^\circ)^{-1} : k \mapsto k/g \). For that purpose, suppose \( w \approx 1/g \in C(I) \) is also given by an oracle, and define:

\[
F : h \mapsto h - w(gh - f).
\]  

(5)

The next proposition computes an upper bound for \( \|h^\circ - f/g\| \); it is implemented in div.newton.

\[ \textbf{Proposition 7.}\ Let \( f, g, h^\circ, w \in C(I) \), and \( \mu, b \in \mathbb{R}_+ \) such that:

\[
\begin{align*}
(7\text{i}) \ |w(gh^\circ - f)| & \leq b, \\
(7\text{ii}) \ |1 - wg| & \leq \mu, \\
(7\text{iii}) \ \mu < 1.
\end{align*}
\]

Then \( g \) does not vanish over \( I \) and \( \|h^\circ - f/g\| \leq b/(1 - \mu) \).

\[ \text{Proof.}\ \text{Conditions (7\text{ii}) and (7\text{iii}) imply that } F \ \text{(Equation (5)) is contracting over } C(I) \ \text{with ratio } \mu. \ \text{The radius } r := \frac{b}{1 - \mu} \ \text{makes the ball } B(h^\circ, r) \ \text{strongly stable with offset } b \ \text{(7\text{i})}, \ \text{since} \ b + \mu r = r. \ \text{Therefore, } h^\ast \ \text{is the (global) unique root of } M, \ \text{and } \|h^\circ - h^\ast\| \leq r. \ \text{Finally, } w \ \text{and } g \ \text{do not vanish because } |1 - wg| \leq \mu < 1. \ \text{Hence, } h^\ast = f/g \ \text{over } I. \]  

The concrete division of models is implemented as follows:

\[
\begin{align*}
\text{Definition } \text{mdiv} \ \text{aux}(F \ G \ H \ W : \text{Model}): \text{Model} := \\
& \quad \text{let } K1 := 1 - \text{W} \ G \ \text{in} \\
& \quad \text{let } K2 := \text{W}(G \ H - F) \ \text{in} \\
& \quad \text{match } \text{mag}(\text{mrange} K1), \ \text{mag}(\text{mrange} K2) \ \text{with} \\
& \quad \quad \mid \text{Some } \text{mu}, \text{Some } b \ \text{when } \text{is} \_ \text{Lt } \mu \ 	ext{H} \Rightarrow \{ | \text{pol} := \text{pol} \ H; \rem := \text{rem} \ H + \text{sym}(b/(1-\text{mu})) \} \\
& \quad \quad \mid \_ \Rightarrow \text{mbot} \\
& \quad \text{end.} \\
\text{Definition } \text{mdiv} \ n(F \ G : \text{Model}): \text{Model} := \\
& \quad \text{let } p, q := \text{mcf} F, \text{mcf} G \ \text{in} \\
& \quad \text{mdiv} \ \text{aux} F G (\text{mfc}(\text{interpolate} n (\text{fun} x \Rightarrow \text{beval} p x / \text{beval} q x))) \\
& \quad \quad \text{mfc}(\text{interpolate} n (\text{fun} x \Rightarrow 1 / \text{beval} q x))).
\end{align*}
\]

Note that we use the trivial model mbot=\{pol:=[],rem:=bot\} as a default value, when the concrete computations fail to validate the guess of the oracle (either because this guess is just wrong, or because of over-approximations in the computations). The correctness lemmas use the properties of operations on models to prove the assumptions of div.newton.

\[
\begin{align*}
\text{Lemma } \text{rmdiv} \ \text{aux}(F \ f \ G \ g \ H \ h \ W w): \\
& \quad \text{mcontains} F \ f \rightarrow \text{mcontains} G \ g \rightarrow \text{mcontains} H \ h \rightarrow \text{mcontains} W \ w \\
& \quad \text{mcontains}(\text{mdiv} \ \text{aux} F G H W)\ (f/g). \\
\text{Lemma } \text{rmdiv} \ n(F \ f \ G \ g): \\
& \quad \text{mcontains} F \ f \rightarrow \text{mcontains} G \ g \rightarrow \text{mcontains}(\text{mdiv} n F G)\ \ (f/g).
\end{align*}
\]

5.3 Validation step for square root

Let \( f \in C(I) \) be strictly positive over \( I \). The square root \( \sqrt{f} \) is one of the two roots of the quadratic equation \( M : h := h^2 - f = 0 \) (the other being \( -\sqrt{f} \)). Let \( h^\circ \) be a candidate approximation. Since \( \text{D}\text{M}_h : k \mapsto 2hk \), one also needs an approximation \( w \approx 1/(2h^\circ) \approx 1/(2\sqrt{f}) \in C(I) \) in order to define \( A : k \mapsto wk \), approximating \( (\text{D}\text{M}_h^\circ)^{-1} \). Then:

\[
F : h \mapsto h - w(h^2 - f).
\]  

(6)

The next proposition (implemented by sqrt.newton), computes an upper bound for \( \|h^\circ - \sqrt{f}\| \).
Proposition 8. Let $f, h^\circ, w \in C(I), \mu_0, \mu_1, b \in \mathbb{R}_+$ and $t_0 \in I$ such that:

\[(8 \text{i}) \| w(h^2) - f \| \leq b, \quad (8 \text{ii}) \| 1 - 2wh^\circ \| \leq \mu_0, \quad (8 \text{iii}) \| w \| \leq \mu_1, \quad (8 \text{iv}) \mu_0 < 1, \quad (8 \text{v}) (1 - \mu_0)^2 - 8b\mu_1 \geq 0, \quad (8 \text{vi}) w(t_0) > 0.\]

Then $f > 0$ over $I$ and $\| h^\circ - \sqrt{f} \| \leq r^*$ where $r^* := (1 - \mu_0 - \sqrt{(1 - \mu_0)^2 - 8b\mu_1}) / 4\mu_1$.

Proof. First, since $\| 1 - 2wh^\circ \| \leq \mu_0 < 1$ by (8ii) and (8iv) and $w(t_0) > 0$ (8vi), $w$ and $h^\circ$ are strictly positive over $I$, by continuity. Using (8iii), $\mu_1 > 0$.

If $b = 0$, then $r^* = 0$ and $h^\circ = \sqrt{f}$ over $I$, because $w(h^2 - f) = 0$ (8i) and $w, h^2 > 0$.

Hence the conclusion holds.

From now on, we assume $b > 0$. $F$ is Lipschitz of ratio $\mu(r) := \mu_0 + 2\mu_1 r$ over $B(h^\circ, r)$ for any $r \in \mathbb{R}_+$, because:

\[F \cdot h_1 - F \cdot h_2 = (h_1 - h_2) - w(h^2_1 - h^2_2) = [(1 - 2wh^\circ) + w(h^\circ - h_1) + w(h^\circ - h_2)](h_1 - h_2).\]

Therefore, satisfying $b + \mu(r)r \leq r$ is equivalent to the quadratic inequality:

\[2\mu_1 r^2 + (\mu_0 - 1)r + b \leq 0.\]

Condition (8v) implies that (7) admits solutions, and $r^*$ is the smallest one. Moreover, since $b, \mu_1 > 0$, we get $r^* > 0$, so that $b + \mu(r)r^* = r^*$ also implies $\mu(r^*) < 1$.

Now, all the assumptions of Theorem 1 are fulfilled. Hence, $F$ has a unique fixed point $h^*$ in $B(h^\circ, r^*)$. To obtain $h^* = \sqrt{f}$ over $I$, it remains to show that $h^* > 0$. This follows from $w > 0$ and:

\[\| 1 - 2wh^* \| \leq \| 1 - 2wh^\circ \| + \| 2w(h^* - h^\circ) \| \leq \mu_0 + 2\mu_1 r^* = \mu(r^*) < 1.\]

Remark 9. Contrary to the case of division where continuity was not needed at all, it is here used for $w$. Therefore, sqrt

Coq code for the corresponding operations on models msqrt_aux and msqrt, together with the statements of their correctness lemmas, are given in Appendix B.

6 Examples

6.1 Playing with approximations of the absolute value function

Consider the function $f_\varepsilon : x \mapsto \sqrt{\varepsilon + x^2}$ over $[-1, 1]$, with $\varepsilon > 0$. When $\varepsilon \to 0$, $f_\varepsilon$ converges uniformly to the absolute value function $x \mapsto |x|$ (which is not analytic at 0), with:

\[|f(x) - |x|| = \left| \sqrt{\varepsilon + x^2} - \sqrt{x^2} \right| = \left| \frac{\varepsilon}{\sqrt{\varepsilon + x^2} + \sqrt{x^2}} \right| \leq \sqrt{\varepsilon}.\]

Rigorous uniform approximations. Approximating $f_\varepsilon$ with polynomials becomes harder for small $\varepsilon$, due to the complex singularities $\pm i\sqrt{\varepsilon}$ getting closer to the interval $[-1, 1]$.

Nevertheless, Chebyshev interpolation still works and our implementation computes rigorous approximations as accurate as desired (see Figure 2b), of exponential convergence with ratio determined by $\varepsilon$. Note that for too small degree, the computed approximation of the square root is too far from the solution, and the a posteriori validation returns an infinite remainder.

In order to provide a comparison with COQAPPROX’s Taylor models, we used the tactic interval with (i_depth 1, i_bisect_taylor x N, i_prec p) to build a Taylor model of
degree $N$ with precision $p$. Timings given in Table 2c reveal a significant advantage of our implementation (there we use $\varepsilon = 2$ to avoid convergence issues of Taylor models).

Concerning accuracy, our experiments tend to show that when $\varepsilon \leq 1$, CoqApprox fails to compute converging Taylor models. Indeed, even with large $L$, a goal like:

\[
\Box x : \mathbb{R}, \; -1 \leq x \leq 1 \rightarrow \sqrt{\frac{1}{100+x^2}} \leq L
\]

is not solved when the degree $N$ becomes too large, probably indicating that the Taylor models diverge due to complex singularities inside the unit disk. (Note that the interval tactic can solve this goal, but only by resorting to subdivision techniques.)

**Error bounding.** We want to bound $|f_z(x) - |x||$ for $x \in [-1, 1]$ without making use of any symbolic manipulation like (8). At first glance, one can choose to use the rigorous approximations over $[-1, 1]$ obtained previously, and evaluate $f_z(x) - x$ (resp. $f_z(x) + x$) over $[0, 1]$ (resp. $[-1, 0]$) using Clenshaw’s algorithm. However, even if the approximations are quite good, this evaluation strategy gives huge overestimations because $[0, 1]$ and $[-1, 0]$ are not small intervals. Instead, we compute separately two approximations for $f_z$: one over $[0, 1]$ and one over $[-1, 0]$, and we evaluate $f_z(x) - x$ (resp. $f_z(x) + x$) over $[1, 0]$ (resp. $[-1, 0]$) using the Chebyshev range function. This approach yields bounds that are rather close to the optimal $\sqrt{\varepsilon}$ (see Figure 24). However, this does not allow for arbitrary accuracy: a subdivision procedure would be necessary here.

### 6.2 Evaluating an Abelian integral

Abelian integrals naturally appear when computing the number of limit cycles bifurcating from a Hamiltonian polynomial vector field in the plane. Indeed, the number of sign alternations of those contour integrals (parameterised by the energy level of the potential function) gives a lower bound on the number of limit cycles of the perturbed system, which is a hard question related to Hilbert’s 16th problem.

In [21], the author claims to prove the existence of 26 limit cycles for a well-constructed quartic system, whereas the previous record for degree 4 was 22 [12]. However, the implementation with which the Abelian integrals were “rigorously” computed was erroneous, which led to apparently more sign alternations than in reality. By tuning the coefficients and computing the integrals with another rigorous numerics library, the authors of the ongoing work [9] obtain 24 limit cycles, which, if not 26, is still greater than the current record 22.

To conclude this article, we rigorously evaluate some of these integrals inside Coq to show how our implementation behaves on non-crafted examples. Below are the formulas defining a family of integral $\mathcal{I}_{ij}(r)$ which need to be computed precisely for several values of $r$. Table 1 summarises the results of our computational experiments. In each line, we chose parameters that were enough to obtain the desired precision. These encouraging results give us hope that it will be possible to fully verify the critical computations involved in recent work of the first author [9].

$$
\mathcal{I}_{ij}(r) = \int_{x_-}^{x_+} x^i (y^+(x)^{j-1} - y^-(x)^{j-1}) \, dx + \int_{y_-}^{y_+} (x^-(y)^{j-1} + x^+(y)^{j-1}) y \, dy - \frac{y_0^2 - y_0}{\delta_x(y)}
$$

$$
x_0 = \frac{9}{10}, \quad x_\pm = \sqrt{x_0 \pm \sqrt{2}}, \quad \delta_y(x) = \sqrt{r^2 - (x^2 - x_0)^2}, \quad x^\pm(y) = \sqrt{x_0 \pm \delta_x(y)}
$$

$$
y_0 = \frac{11}{10}, \quad y_\pm = \sqrt{y_0 \pm \sqrt{2}}, \quad \delta_x(y) = \sqrt{r^2 - (y^2 - y_0)^2}, \quad y^\pm(x) = \sqrt{y_0 \pm \delta_y(x)}
$$
A certificate-based approach to formally verified approximations

\[ \varepsilon = 0.1 \]
\[ \varepsilon = 0.01 \]
\[ \varepsilon = 0.001 \]

\[ |x| - 1.0 \to 0.5 \to 1.0 \]

(a) Functions \( f_\varepsilon \) and \( x \mapsto |x| \) over \([-1, 1] \).

(b) Magnitude of the remainders of degree-\( N \) Chebyshev models approximating \( f_\varepsilon \) over \([-1, 1] \).

(c) Timings of degree-\( N \) models for \( f_2 \).

(d) Overapproximation ratio of the remainder of the degree-\( N \) Chebyshev model for \( f_\varepsilon - 1 \) over \([0, 1] \), compared to the optimal bound \( \sqrt{\varepsilon} \).

Figure 2 Approximating functions \( f_\varepsilon \) and \( x \mapsto |x| \) with Chebyshev models.

<table>
<thead>
<tr>
<th>( N )</th>
<th>( \text{time (in seconds)} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{Chebyshev} )</td>
<td>( \text{COQAPPROX} )</td>
</tr>
<tr>
<td>10</td>
<td>0.11</td>
</tr>
<tr>
<td>20</td>
<td>0.16</td>
</tr>
<tr>
<td>30</td>
<td>0.22</td>
</tr>
<tr>
<td>40</td>
<td>0.31</td>
</tr>
<tr>
<td>50</td>
<td>0.42</td>
</tr>
<tr>
<td>60</td>
<td>0.56</td>
</tr>
<tr>
<td>70</td>
<td>0.71</td>
</tr>
<tr>
<td>80</td>
<td>0.89</td>
</tr>
<tr>
<td>90</td>
<td>1.08</td>
</tr>
<tr>
<td>100</td>
<td>1.31</td>
</tr>
<tr>
<td>120</td>
<td>1.80</td>
</tr>
<tr>
<td>140</td>
<td>2.43</td>
</tr>
<tr>
<td>160</td>
<td>2.98</td>
</tr>
<tr>
<td>180</td>
<td>3.86</td>
</tr>
<tr>
<td>200</td>
<td>4.75</td>
</tr>
</tbody>
</table>

Table 1 Reached precision for \( \beta_{ij}(r) \) for different values of \( r \), computed with degree-\( N \) Chebyshev models and floating-point precision \( p \) (in each cell we display the width of the computed enclosure).

\[ r \quad N \quad p \quad \text{time (s)} \quad \beta_{00} \quad \beta_{20} \quad \beta_{22} \quad \beta_{40} \quad \beta_{04} \]
| 0.5 | 13 | 32 | 0.38 | 2.4e-05 | 2.9e-05 | 4.1e-05 | 3.0e-05 | 4.8e-05 |
| 0.78 | 15 | 32 | 0.47 | 4.6e-05 | 2.0e-05 | 2.7e-05 | 2.4e-05 | 1.1e-04 |
| 0.88 | 65 | 128 | 17.34 | 2.5e-08 | 5.0e-11 | 8.5e-11 | 5.3e-11 | 6.3e-08 |
| 0.89 | 95 | 128 | 35.13 | 2.0e-08 | 1.8e-11 | 2.9e-11 | 2.0e-11 | 5.1e-08 |
| 0.895 | 135 | 300 | 173.23 | 2.6e-08 | 1.7e-11 | 1.8e-11 | 1.3e-11 | 6.7e-08 |

7 Conclusion and future work

The COQ development is available online [11]. It consists of around 1300 lines of specifications and 1500 lines of proofs. We leave several directions for future work: integrate it with CoqInterval to benefit from its automatic subdivision techniques; interface the library with external tools for the approximation steps; implement other bases. Applying this approach to verify solutions of linear ODEs in a systematic way [2, 8] is also a longer-term perspective.
References


### A  Coq code for primitive in Chebyshev basis

```coq
Fixpoint prim_ (C : Ops1) (n : nat) (p : seq C) : seq C :=
  match P, n with
  | [], _ => []
  | a::q, 0 => sadd [0; a] (prim_ 1 q)
  | a::q, 1 => 0 :: (sadd [0; a/4] (prim_ 2 q))
  | a::q, _ => sadd [-a/(2*(n-1)); 0; a/(2*(n+1))] (0 :: prim_ (n+1) Q)
  end.

Definition prim (C : Ops1) (P : seq C) := prim_ 0 P.
```

The key lemma is the following one:

```coq
Lemma eval_prim_ n (p : seq R) x : Derive (eval_ T (n-1) (prim_ n P)) x = eval_ T n P x.
```

### B Coq code for the square root of a model

```coq
Let msqrt_aux (F H W : Model) (x : II) : Model :=
  let Wx := meval W x in
  if (is_lt lo x && is_lt x hi && is_lt 0 Wx) then mbot else
  let K1 := 1 - 2*W*H in
  let K2 := W*(H*H-F) in
  match mag (mrange K1), mag (mrange W), mag (mrange K2) with
  | Some mu0, Some mu1, Some b =>
    let delta := (1 - mu0)^2 - 8*b*mu1 in
    let rmin := (1 - mu0 - sqrt delta)/(4*mu1) in
    let mu := mu0 + 2*mu1*rmin in
    if is_lt mu0 1 && is_lt 0 delta && is_lt mu' 1 then
      { | pol := pol H; rem := rem H + sym rmin' |}
    else mbot
  | _ => mbot
  end.

Let msqrt n (F : Model) : Model :=
  let h := seq FF := mcf F in
  let h := seq FF := interpolate n (fun x => sqrt (beval p x)) in
  msqrt_aux M (mfc h) (mfc (interpolate n (fun x => 1/(2*beval h x)))) ((lo+hi)/2).

Lemma rmsqrt_aux (F H W : Model) (X : II) (f h w : R → R) (x : R):
  mcontains F f → mcontains H h → mcontains W w → contains X x → lo ≤ x ≤ hi →
  ∀ y, lo ≤ y ≤ hi → continuity_pt w x →
  mcontains (msqrt_aux F H W X) (sqrt f).

Lemma rmsqrt n F f : mcontains F f → mcontains (msqrt' n F) (sqrt f).
```