Numerical Analysis of Ordinary Differential Equations in Isabelle/HOL

Fabian Immler and Johannes Hölzl*

www.in.tum.de/~immler www.in.tum.de/~hoelzl
Institut für Informatik, Technische Universität München

Abstract. Many ordinary differential equations (ODEs) do not have a closed solution, therefore approximating them is an important problem in numerical analysis. This work formalizes a method to approximate solutions of ODEs in Isabelle/HOL.
We formalize initial value problems (IVPs) of ODEs and prove the existence of a unique solution, i.e. the Picard-Lindelöf theorem. We introduce generic one-step methods for numerical approximation of the solution and provide an analysis regarding the local and global error of one-step methods.
We give an executable specification of the Euler method as an instance of one-step methods. With user-supplied proofs for bounds of the differential equation we can prove an explicit bound for the global error. We use arbitrary-precision floating-point numbers and also handle rounding errors when we truncate the numbers for efficiency reasons.

Keywords: Formalization of Mathematics, Ordinary differential equation, Numerical Analysis, One-Step method, Euler method, Isabelle/HOL

1 Introduction

Ordinary differential equations (ODEs) have a lot of important applications. They are for example used to describe motion or oscillation in Newtonian mechanics, the evolution or growth of organisms in biology, or the speed of chemical reactions.

The Picard-Lindelöf theorem states the existence of a unique solution (under certain conditions) but unfortunately, many problems do not allow an explicit closed formula as solution (e.g. the seemingly simple ODE $\dot{x} = x^2 - t$ for initial values $x(t_0) = x_0$). In such cases, one has to content oneself with numerical methods that give approximations to the solution.

In order to evaluate the quality of an approximate solution (which depends very much on the concrete problem) you need to choose the parameters of your numerical method (i.e. step size, precision) wisely. This is where the use of an interactive theorem prover might be useful: We formalize initial value problems (IVPs) of ODEs and prove the existence of a unique solution in Isabelle/HOL. We give an executable specification of the Euler method – a basic numerical algorithm – and prove the error bound of the approximation.

* Supported by the DFG Graduiertenkolleg 1480 (PUMA)

2 Related Work

When an ODE has a solution representable in a closed form, an approximation method for this closed form can be used. Muñoz and Lester [12] use rational interval arithmetic in PVS to efficiently approximate real valued functions in theorem provers. In addition to basic arithmetic operations they also support trigonometric functions. Melquiond [11] implements a similar method in Coq. He also implements interval arithmetic, but uses floating-point numbers and sophisticated methods to avoid loss of correlation.

An alternative to interval arithmetic is to approximate real numbers by a sequence of rational numbers. Each element in this sequence has a defined distance to the exact result. Harrison [5] uses this approach to compute the logarithm. O'Connor [15] approximates real numbers by organizing their completion of rational numbers in a monad. O'Connor and Spitters [14] use this monadic construction in order to implement arbitrary approximations of the Riemann integral. Krebbers and Spitters [9,10] extend this work to use arbitrary-precision floating-point numbers. Similar to the proof of the existence of the unique solution of an IVP, one can iterate an integral operation in order to approximate the solution (as suggested in [17]).

Boldo *et al.* [1] formalize partial differential equations stemming from acoustic wave equations in Coq. As they analyse partial differential equations they can not show a general existence or uniqueness theorem. Their particular problem admits an analytical solution and they simply assume that the solution is unique. They also show consistency and stability and that in their case convergence follows. However, they needed to find an analytical formula for the rounding errors.

3 Preliminaries

3.1 Isabelle/HOL

The formalizations presented in this paper are done in the Isabelle/HOL theorem prover. In this section we give an overview of our syntactic conventions.

The term syntax follows the λ-calculus, i.e. function application is juxtaposition as in $f\ t$ and function abstraction is written as $\lambda x.\ t$. The notation $t :: \tau$ means that t has type τ. Types are built from base types like $\mathbb{N}$ (natural numbers), $\mathbb{R}$ (real numbers), $\mathbb{R}^n$ (Euclidean spaces of dimension n), type variables (α, β, etc.), functions $\alpha \to \beta$, sets $\mathcal{P}(\alpha)$, and pairs $\alpha \times \beta$.

Our work builds on the `Multivariate_Analysis` library which was ported from Harrison's Euclidean spaces for HOL-Light [6]. In our formalization the Euclidean space $\mathbb{R}^n$ is not just the function space $n \to \mathbb{R}$; it is a type class denoting a Euclidean space. $\mathbb{R}^n \times \mathbb{R}^m$, $\mathbb{R}$, and $n \to \mathbb{R}$ are in this type class.

We write $(a, b) :: \alpha \times \beta$ for pairs, $A \times B := \{(a, b) \mid a \in A \wedge b \in B\} :: \mathcal{P}(\alpha \times \beta)$ for the Cartesian product of A and B (do not confuse with the product type), $\|x\|$ for the norm of x, $\mathcal{B}_r(x) := \{y \mid \|x - y\| \leq r\} :: \mathcal{P}(\mathbb{R}^n)$ for the closed ball around x with radius $r :: \mathbb{R}$, and $\mathit{sup}\ A$ and $\mathit{inf}\ A$ for the supremum and infimum of A.

With $\dot{x}(t) = y$ we denote that $x :: \mathbb{R} \to \mathbb{R}^n$ has at t the derivative $y :: \mathbb{R}^n$, and with $\int_a^b f\, x\, dx$ we denote the integral of $f :: \mathbb{R}^n \to \mathbb{R}^m$ over $[a; b]$. Hilbert choice is $(\varepsilon\, x.\ P\ x)$, i.e. $(\exists x.\ P\ x) \Rightarrow P(\varepsilon\, x.\ P\ x)$ holds. In this section x_i is the i-th projection of the vector x. We write $[a; b] := \{x \mid \forall i.\ a_i \leq x_i \wedge x_i \leq b_i\} :: \mathcal{P}(\mathbb{R}^n)$ for hyperrectangles on Euclidean spaces (which are closed intervals on $\mathbb{R}$), and $\mathcal{R}_r(x) := \{y \mid \forall i.\ y_i \in [x_i - r; x_i + r]\}$ for hypercubes around x. The predicate $\mathit{is\text{-}interval}\ S := (\forall a, b \in S.\ \forall x.\ (\forall i.\ x_i \in [a_i; b_i] \cup [b_i; a_i]) \Rightarrow x \in S)$ accepts intervals in general, mixtures of open and closed, and infinite ones.

3.2 Arbitrary-Precision Floating-Point Numbers ($\mathbb{F}$)

The previous formalization of arbitrary-precision floating-point numbers in Isabelle/HOL [13,7] used pairs of integer exponents e and mantissas m, representing the real numbers $m \cdot 2^e$. Unfortunately this results in a type which has multiple representations for the same number, e.g. zero is represented by $0 \cdot 2^e$ for every e. Therefore the resulting type does not support any interesting type class, like linear order, commutative groups for addition, etc.

Hence, we introduce a new type $\mathbb{F}$ as the dyadic rationals, i.e. all numbers x which are representable as $m \cdot 2^e$. We have an injective function $(\cdot)_{\mathbb{R}} :: \mathbb{F} \to \mathbb{R}$ and its partially specified inverse $(\cdot)_{\mathbb{F}} :: \mathbb{R} \to \mathbb{F}$. As (non-injective) constructor we introduce $\mathit{Float}\ m\ e = (m \cdot 2^e)_{\mathbb{F}}$ and declared it as a datatype constructor for code generation [3]. We lift the arithmetic constants 0, 1, $+$, $-$, $\cdot$, $<$, $\leq$ from the reals and provide executable equations, e.g. for multiplication:

$$(\mathit{Float}\ m_1\ e_1) \cdot (\mathit{Float}\ m_2\ e_2) = \mathit{Float}\ (m_1 \cdot m_2)\ (e_1 + e_2).$$

3.3 Bounded Continuous Functions

The proof for the existence of a unique solution to an IVP is based on an application of the Banach fixed point theorem, which guarantees the existence of a unique fixed point of a contraction mapping on metric spaces. The textbook-proof of Walter [18] defines a metric space on continuous functions with a compact domain. As functions in Isabelle/HOL are required to be total, one cannot simply restrict the domain, hence a slightly different approach is necessary: We define a type $\overline{\mathcal{C}}$ carrying bounded continuous functions, i.e. functions which are continuous everywhere and whose values are bounded by a constant:

$$\overline{\mathcal{C}} = \{f :: \mathbb{R}^n \to \mathbb{R}^m \mid f \text{ continuous on } \mathbb{R}^n \wedge (\exists B.\ \forall t.\ \|f\ t\| \leq B)\}$$

The morphisms $\mathit{Rep}_{\overline{\mathcal{C}}} : \overline{\mathcal{C}} \to (\mathbb{R}^n \to \mathbb{R}^m)$ and $\mathit{Abs}_{\overline{\mathcal{C}}} : (\mathbb{R}^n \to \mathbb{R}^m) \to \overline{\mathcal{C}}$ allow to use an element of type $\overline{\mathcal{C}}$ as function and to define elements of type $\overline{\mathcal{C}}$ in terms of a function. We define a norm on $\overline{\mathcal{C}}$ as the supremum of the range and operations $+$, $-$, $\cdot$ pointwise.

$$\begin{aligned} \|f\| &:= \mathit{sup}\ \{\|\mathit{Rep}_{\overline{\mathcal{C}}}\ f\ x\| \mid x \in \mathbb{R}^n\} \\ f + g &:= \mathit{Abs}_{\overline{\mathcal{C}}}(\lambda x.\ f\ x + g\ x) \\ f - g &:= \mathit{Abs}_{\overline{\mathcal{C}}}(\lambda x.\ f\ x - g\ x) \\ a \cdot f &:= \mathit{Abs}_{\overline{\mathcal{C}}}(\lambda x.\ a \cdot f\ x) \end{aligned}$$

We prove that $\overline{\mathcal{C}}$ is a normed vector space, hence also a metric space. In order to be able to apply the Banach fixed point theorem we need to show that $\overline{\mathcal{C}}$ is a complete space, meaning that every Cauchy sequence converges. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ that is continuous on a compact interval $[a; b]$ is converted to $\overline{\mathcal{C}}$ by extending the function continuously outside the domain with the help of *clamp*:

$$(\mathit{clamp}_{[a;b]}\ x)_i := \textbf{\textit{if}}\ x_i \leq a_i\ \textbf{\textit{then}}\ a_i\ \textbf{\textit{else}}\ (\textbf{\textit{if}}\ x_i \geq b_i\ \textbf{\textit{then}}\ b_i\ \textbf{\textit{else}}\ x_i)$$
$$\mathit{ext\text{-}cont}_{[a;b]}\ f := \mathit{Abs}_{\overline{\mathcal{C}}}(\lambda x.\ f\ (\mathit{clamp}_{[a;b]}\ x))$$

The key property we use is that an extended function is continuous everywhere when it was continuous on the interval. Inside the interval the resulting function takes the same values as the original function:

$$f \text{ continuous on } [a; b] \Rightarrow \mathit{Rep}_{\overline{\mathcal{C}}}(\mathit{ext\text{-}cont}_{[a;b]} f) \text{ continuous on } \mathbb{R}^n$$
$$x \in [a; b] \Rightarrow \mathit{Rep}_{\overline{\mathcal{C}}}(\mathit{ext\text{-}cont}_{[a;b]} f)\ x = f\ x$$

4 Initial Value Problems

4.1 Definition

An equation in which an unknown function $u :: \mathbb{R} \to \mathbb{R}^n$ and derivatives of this function occur is an ODE. ODEs with derivatives of higher order can be reduced to a first order system, which is why we handle ODEs of first order only. An ODE together with values t_0, x_0 for an initial condition $u\ t_0 = x_0$ is an IVP and can always be written in terms of a right-hand side f which is supposed to be defined on a domain $I \times D$ (Compare with Figure 1):

$$\dot{u}\ t = f(t, u\ t) \qquad u\ t_0 = x_0 \in D \qquad t_0 \in I$$

We define IVPs in Isabelle as a record, which is a named tuple of elements

$$f :: \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n, \quad t_0 :: \mathbb{R}, \quad x_0 :: \mathbb{R}^n, \quad I :: \mathcal{P}(\mathbb{R}), \quad D :: \mathcal{P}(\mathbb{R}^n)$$
$$\mathit{ivp} = (f, t_0, x_0, I, D).$$

For the rest of the paper, we assume an IVP *ivp* with $(t_0, x_0) \in I \times D$ and use the symbols f, t_0, x_0, I, and D without further notice as part of this IVP.[1] We will notate modified IVPs in subscripts, for example $\mathit{ivp}_{f:=g}$ for the IVP *ivp* with the right-hand side g instead of f. Other components are updated analogously. Later definitions implicitly depend on *ivp* and may be updated in a similar fashion.

4.2 Solutions

We capture the notion of a solution to an IVP in the predicate *is-solution*: A solution needs to satisfy the initial condition and the derivative has to be given by f. Apart from that, we need to state explicitly that the potential solution must not leave the codomain. Mathematicians usually do that implicitly when declaring f as a function with domain $I \times D$.

$$\mathit{is\text{-}solution}\ u := u\ t_0 = x_0 \wedge (\forall t \in I.\ \dot{u}\ t = f(t, u\ t) \wedge u\ t \in D)$$

[1] In Isabelle/HOL: *ivp* is fixed as a locale parameter

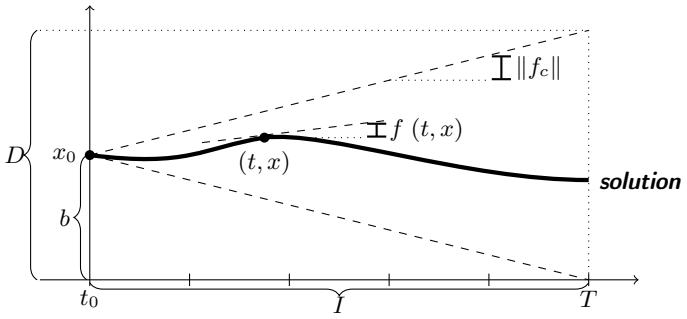

Fig. 1. The solution of an IVP on a rectangular domain and related variables

unique-solution The Picard-Lindelöf theorem states the existence of a unique solution. We formalize the notion of a unique solution as follows: If two functions are solutions, they must attain the same values on I. We use Hilbert choice to obtain *solution* and relate all upcoming facts about solutions to IVPs to it.

$$\begin{aligned}\textit{has-solution} &:= (\exists u.\ \textit{is-solution}\ u)\\ \textit{solution} &:= (\varepsilon u.\ \textit{is-solution}\ u)\\ \textit{unique-solution} &:= \textit{has-solution} \wedge (\forall v.\ \textit{is-solution}\ v \Rightarrow (\forall t \in I.\ v\ t = \textit{solution}\ t))\end{aligned}$$

4.3 Combining Initial Value Problems

Working with IVPs in a structured way helped us to implement the proofs in a maintainable fashion. An important operation is to be able to "connect" two solutions at a common point. We therefore assume two IVPs $\textit{ivp}_1$ and $\textit{ivp}_2$ that we want to combine to an IVP $\textit{ivp}_c$:

$$\begin{aligned}\textit{ivp}_1 &= (f_1, t_0, x_0, [t_0; t_1], D)\\ \textit{ivp}_2 &= (f_2, t_1, \textit{solution}_{\textit{ivp}_1}\ t_1, [t_1; t_2], D)\\ f_c &:= (\lambda(t,x).\ \textbf{\textit{if}}\ t \le t_1\ \textbf{\textit{then}}\ f_1(t,x)\ \textbf{\textit{else}}\ f_2(t,x))\\ \textit{ivp}_c &:= (f_c, t_0, x_0, [t_0; t_2], D)\end{aligned}$$

Assuming unique solutions for $\textit{ivp}_1$ and $\textit{ivp}_2$, we prove a unique solution for $\textit{ivp}_c$:

$$\begin{aligned}&f_1(t_1, \textit{solution}_{\textit{ivp}_1}\ t_1) = f_2(t_1, \textit{solution}_{\textit{ivp}_1}\ t_1) \Rightarrow\\ &\quad \textit{unique-solution}_{\textit{ivp}_1} \Rightarrow \textit{unique-solution}_{\textit{ivp}_2} \Rightarrow \textit{unique-solution}_{\textit{ivp}_c}\end{aligned}$$

4.4 Quantitative Picard-Lindelöf

In this section, we show that certain sets of assumptions (*bnd-strip*, *strip*, *rect*) imply *unique-solution*, i.e. the existence of a unique solution and therefore several variants of the Picard-Lindelöf theorem[2]. We will present key parts of the proofs

[2] In Isabelle/HOL: We show that e.g. *rect* is a sublocale of *unique-solution*

and how they have been implemented in Isabelle, especially to show how our choice of formalization helped to structure the proofs. The proofs in this section are inspired by Walter [18] and follow closely the structure which is given there. All of the proofs provide concrete results that we use later in the numerical approximation of IVPs.

Lipschitz continuity is a basic assumption for the Picard-Lindelöf theorem. It is stronger than continuity: it limits how fast a function can change. A function g is (globally) Lipschitz continuous on a set S if the slope of the line connecting any two points on the graph of g is bounded by a constant:

$$\mathit{lipschitz}\ S\ g\ L := (\forall x, y \in S.\ \|g\ x - g\ y\| \leq L \cdot \|x - y\|)$$

bnd-strip If we choose $I = [t_0; T]$ and $D = \mathbb{R}^n$ for the domain of a function f that is continuous on $I \times D$ and assume a global Lipschitz constant L on D, we can show the existence of a unique solution, provided that $(T - t_0) \cdot L < 1$ holds. We call this set of assumptions *bnd-strip*:

$$\begin{aligned}\mathit{bnd\text{-}strip}\ T\ L := {} & f \text{ continuous on } I \times D \wedge \\ & (\forall t \in I.\ \mathit{lipschitz}\ D\ (\lambda x.\ f(t, x))\ L) \wedge \\ & I = [t_0; T] \wedge D = \mathbb{R}^n \wedge (T - t_0) \cdot L < 1\end{aligned}$$

Using the fundamental theorem of calculus, any solution to the equalities of the IVP must also satisfy $u\ t - x_0 = \int_{t_0}^{t} f(\tau, u\ \tau)\mathsf{d}\tau$ for all t in I. This equality can be seen as an iteration of functions, known as *Picard iteration* which we conduct in the space of bounded continuous functions (moving explicitly between functions $\mathbb{R} \to \mathbb{R}^n$ and $\overline{\mathcal{C}}$ is the most prominent difference to the textbook proof):

$$P\ x := \mathit{ext\text{-}cont}_{[t_0;T]} \left(\lambda t.\ x_0 + \int_{t_0}^{t} f\big(\tau, \mathit{Rep}_{\overline{\mathcal{C}}}\ x\ \tau\big)\mathsf{d}\tau\right); \qquad P :: \overline{\mathcal{C}} \to \overline{\mathcal{C}}$$

In this metric space, we show that P is Lipschitz continuous for the constant $(T - t_0) \cdot L$. The Lipschitz constant is less than 1 (i.e. $(T - t_0) \cdot L < 1$). This is a necessary assumption – together with the completeness of $\overline{\mathcal{C}}$ – to apply the Banach fixed point theorem (from the `Multivariate_Analysis` library). It guarantees the existence of a unique fixed point x^* for the mapping P.

Together with the fundamental theorem of calculus we show that the fixed point x^* of P is a solution. Moreover every (continuously extended) solution u is a fixed point of P

$$\mathit{is\text{-}solution}\ (\mathit{Rep}_{\overline{\mathcal{C}}}\ x^*)$$

$$\mathit{is\text{-}solution}\ u \Rightarrow P(\mathit{ext\text{-}cont}_{[t_0;T]}\ u) = \mathit{ext\text{-}cont}_{[t_0;T]}\ u$$

from which we conclude the existence of a unique solution:

Theorem 1 (Picard-Lindelöf). *If f is continuous and Lipschitz continuous in its second variable and the interval $I = [t_0; T]$ is small enough, then there exists a unique solution:*

$$\mathit{bnd\text{-}strip}\ T\ L \Rightarrow \mathit{unique\text{-}solution}.$$

strip According to *bnd-strip*, the size of the interval $[t_0; T]$ in which we have proven the existence of a unique solution depends on the Lipschitz-constant L.

In *strip* we drop the restriction $T - t_0 < \frac{1}{L}$ to the size of the interval. This can be done by splitting the desired interval $[t_0; T]$ into n sub-intervals, such that $\frac{T-t_0}{n}$ is small enough to satisfy the assumptions of *bnd-strip*. In an inductive (on $n \geq 1$) proof, one has (as hypothesis) the existence of a *unique-solution* on $[t_0; T - \frac{1}{n}(T - t_0)]$. The interval $[T - \frac{1}{n}(T - t_0); T]$ satisfies the assumptions for *bnd-strip* – consequently we have a *unique-solution* there, too. The respective solutions can then be combined. The argumentation in the textbook proof relies on geometric intuition when one combines solutions – doing this formally requires more efforts, but section 4.3 helped in retaining structure in the proofs.

rect In *strip*, there is the assumption $D = \mathbb{R}^n$ for the codomain of the solution. One might want to restrict this part – e.g. if there is no Lipschitz constant on the whole codomain – to $D = \mathcal{R}_b(x_0)$. In this case (which we will call *rect*) we continuously extend the right-hand side f outside the rectangle to f_c:

$$f_c := \textit{ext-cont}_{[t_0;T]\times\mathcal{R}_b(x_0)}\ f; \qquad f_c :: \overline{\mathcal{C}}$$

The textbook also works with a continuous extended function, but we do so more explicitly with the utilization of *ext-cont*. We used *ext-cont* to obtain f_c, hence $\textit{Rep}_{\overline{\mathcal{C}}}\ f_c$ is continuous on the whole domain $I \times \mathbb{R}^n$. We apply Theorem 1 to obtain the existence of a unique solution for $\textit{Rep}_{\overline{\mathcal{C}}}\ f_c$. We show that the solution does not leave the codomain D – to ensure that $f = \textit{Rep}_{\overline{\mathcal{C}}}\ f_c$. For this, one has to choose a small enough upper bound T of the existence interval $[t_0; T]$. This depends on the maximum slope of the solution which is bounded by $\|f_c\|$, see Fig. 1. The formal proof centers around an application of the mean value theorem, this is tedious compared to the geometric intuition given in the textbook.

$$\begin{aligned}\textit{rect}\ T\ b\ L := f \text{ continuous on } I \times D \wedge (\forall t \in I.\ \textit{lipschitz}\ D\ (\lambda x.\ f(t,x))\ L) \wedge \\ I = [t_0; T] \wedge D = \mathcal{R}_b(x_0) \wedge b \geq 0 \wedge T \leq t_0 + b/\|f_c\|\end{aligned}$$

Under these assumptions, we show that any solution to *ivp* cannot leave D.

$$\textit{rect}\ T\ b\ L \Rightarrow \textit{is-solution}\ u \Rightarrow (\forall t \in I.\ u\ t \in D)$$

Having this, we can show that $\textit{solution}_{f:=f_c}$ is a solution to *ivp* and that every other solution to *ivp* is also a solution to $\textit{ivp}_{f:=f_c}$. Consequently:

Theorem 2 (Picard-Lindelöf on a restricted domain).

$$\textit{rect}\ T\ b\ L \Rightarrow \textit{unique-solution}$$

4.5 Qualitative Picard-Lindelöf

In this section, we present a variant of the Picard-Lindelöf theorem (following the textbook proof of Walter [18] closely), which is mainly of mathematical interest: One does not get explicit values that could be used to estimate the error in a numerical approximation – which is what we need in the upcoming sections.

local-lipschitz Many functions do not have a global Lipschitz constant L (e.g. $f(t,x) = x^2$ on $\mathbb{R}$). The weaker assumption of *local Lipschitz continuity* allows to prove the existence of a solution in a neighborhood of the initial value. A function f is locally Lipschitz continuous in its second variable if for every point (t,x) of the domain, there exists a neighborhood $\mathcal{B}_\epsilon(t,x)$ inside which there exists a Lipschitz constant L:

$$\begin{aligned}\mathit{local\text{-}lipschitz} := \forall (t,x) \in I \times D.\ \exists \epsilon > 0.\ \exists L.\\ \forall u \in \mathcal{B}_\epsilon(t) \cap I.\ \mathit{lipschitz}\ (\lambda x. f(u,x))\ (\mathcal{B}_\epsilon(x) \cap D)\ L\end{aligned}$$

open-domain Together with the notion of local Lipschitz continuity, we get a very general result for the existence of a unique solution if we assume an open domain. We will use the set of assumptions *open-domain* to prove the existence of a unique solution on a maximal existence interval.

$$\mathit{open\text{-}domain} := \mathit{local\text{-}lipschitz} \wedge \mathit{open}\ I \wedge \mathit{open}\ D$$

Under these assumptions, we construct a small enough rectangle inside a neighborhood of the initial values that is inside the domain and possesses a Lipschitz-constant. From this we can conclude

$$\exists T > t_0.\ [t_0;T] \subseteq I \wedge \mathit{unique\text{-}solution}_{I:=[t_0;T]}.$$

We define Φ (similar to the textbook, but more explicit) to be the set of all solutions to *ivp* and upper bounds of their existence intervals starting from t_0:

$$\Phi := \{(u,T) \mid\ t_0 < T \wedge [t_0;T] \subseteq I \wedge \mathit{is\text{-}solution}_{I:=[t_0;T]}\ u\}$$

For this set, we can show that all solutions u, v in Φ take the same values on the intersection of their existence intervals. We do so by falsely assuming that they differ at a point t_1 ($u(t_1) \neq v(t_1)$) and showing that there has to be a maximal point t_m at which u and v are equal. Then, however, one can use the previous theorem about the existence of a unique solution in a neighborhood of t_m, to show that the two solutions have to be equal at larger points than t_m, contradicting its maximality.

One can then define a solution on the interval $J := \bigcup_{(u,T)\in\Phi}[t_0;T]$ for which $\mathit{unique\text{-}solution}_{I:=J}$ holds. Additionally, for every other interval K for which there exists a solution, K is a subset of J and the solution is only a restriction. From a mathematical point of view this is an important result, stating the existence of a maximal existence interval for the unique solution:

Theorem 3 (Picard-Lindelöf on an open domain, maximal existence interval).

$$\begin{aligned}&\mathit{unique\text{-}solution}_{I:=J} \wedge\\ &\forall K \subseteq I.\ \mathit{is\text{-}interval}\ K \Rightarrow \mathit{inf}\ K = t_0 \Rightarrow \mathit{has\text{-}solution}_{I:=K} \Rightarrow\\ &\qquad (K \subseteq J \wedge (\forall t \in K.\ \mathit{solution}_{I:=K}\ t = \mathit{solution}_{I:=J}\ t))\end{aligned}$$

5 One-step methods

The aim of this paper is to approximate solutions of IVPs with the Euler method. The Euler method is a one-step method: it approximates a function (the solution) in discrete steps, each step operating exclusively on the results of one previous step. For one-step methods in general, one can give assumptions under which the method works correctly – where the error of the approximation goes to zero with the step size.

The methodology is as follows (cf. Bornemann [2]): If the error in one step goes to zero with the step size, the one-step method is called **consistent**. One can show that every consistent one-step method is **convergent**: the global error – the error after a series of steps – goes to zero with the step size, too.

For efficiency reasons, we want to limit the precision of our calculations – which causes rounding errors. The effect of small errors in the execution of a one-step method is studied with the notion of **stability**: The error between the ideal and the perturbed one-step method goes to zero with the step size.

We first give a formal definition of one-step methods, formalize the notions of consistency, convergence and stability. We prove that consistent one-step methods are convergent and stable. We are going to use these definitions and results in the upcoming section to show that the Euler method is consistent and can therefore be used to approximate IVPs.

5.1 Definition

Following the textbook [2], we want to approximate the solution $u : \mathbb{R} \to \mathbb{R}^n$ at discrete values given by $\Delta :: \mathbb{N} \to \mathbb{R}$ with $\forall j.\ \Delta\ j \leq \Delta\ (j+1)$. We notate $\Delta_j := \Delta\ j$, denote by $h_j := \Delta_{j+1} - \Delta_j$ the step size, and by $h_{max} := \mathsf{max}_j\ h_j$ its maximum.

The approximation should be given by a one-step method (or ***g***rid ***f***unction) gf such that $gf\ j \approx u\ \Delta_j$. One-step methods use for the approximation at Δ_{j+1} only the information of the previous point at Δ_j. A one-step method gf on a grid Δ for a starting value x_0 can therefore be defined recursively. It is characterized by an increment function ψ which gives the slope of the connection (the so called discrete evolution Ψ) between two successive points (depending on the step size h and the position (t, x) of the previous point):

$$h, t :: \mathbb{R}; \quad x, x_0 :: \mathbb{R}^n; \quad \psi, \Psi_\psi :: \mathbb{R} \to \mathbb{R} \to \mathbb{R}^n \to \mathbb{R}^n$$

$$\begin{aligned} \Psi_\psi\ h\ t\ x &:= x + h \cdot \psi\ h\ t\ x \\ gf\ \Delta\ \psi\ x_0\ 0 &:= x_0 \\ gf\ \Delta\ \psi\ x_0\ (j+1) &:= \Psi_\psi\ h_j\ \Delta_j\ (gf\ \Delta\ \psi\ x_0\ j) \end{aligned}$$

5.2 Consistency implies Convergence

We now describe up to which extent one-step methods can be used to approximate an arbitrary function $u : \mathbb{R} \to \mathbb{R}^n$ on an interval $I := [\Delta_0; T]$. We first

formalize the notion of consistency (bounding the local error), then summarize a set of required assumptions in the predicate *convergent* from which we show that one-step methods converge.

The error in one step (the local error) is given by $\|u\ (t+h) - \Psi_\psi\ h\ t\ (u\ t)\|$ at a point $(t, u\ t)$ for a step size h. We (as well as the textbook) call a one-step method consistent with u of order p if the local error is in $\mathcal{O}(h^{p+1})$. This means that there exists a constant B such that the local error is less than $B \cdot h^{p+1}$:

$$\begin{aligned}&\mathit{consistent}\ u\ B\ p\ \psi :=\\ &\quad \big(\forall t \in [\Delta_0; T].\ \forall h \in [0; T-t].\ \|u\ (t+h) - \Psi_\psi\ h\ t\ (u\ t)\| \leq B \cdot h^{p+1}\big)\end{aligned}$$

convergent As in the proof of the Picard-Lindelöf theorem, we need the notion of Lipschitz continuity: The textbook defines a cylindrical neighborhood of radius r around u in which the increment function ψ needs to be Lipschitz continuous. Moreover the increment function is assumed to be consistent with u. The definition of *convergent* summarizes the assumptions required to show convergence.

$$\begin{aligned}\mathit{convergent}\ u\ B\ p\ \psi\ r\ L := \mathit{consistent}\ u\ B\ p\ \psi \wedge p > 0 \wedge B \geq 0 \wedge L \geq 0 \wedge\\ \big(\forall t \in [\Delta_0; T].\ \forall h \in [0; T-t].\\ \mathit{lipschitz}\ (\mathcal{B}_r(u\ t))\ (\psi\ h\ t)\ L\big)\end{aligned}$$

We need to give a constant C such that the global error is less than $C \cdot h^p$. This constant depends on B and L and the length S of the interval I. We need this constant as a bound in several upcoming proofs, hence we define it here as $\mathit{bound}_S\ B\ L$ for the sake of readability. We want to limit the step size depending on this constant, the order of consistency, and the radius r of the neighborhood with a Lipschitz constant, hence we introduce $\mathit{step\text{-}bnd}\ B\ L\ p\ r$.

$$\mathit{bound}_S\ B\ L := \tfrac{B}{L} \cdot \big(e^{L \cdot S + 1} - 1\big) \qquad \mathit{step\text{-}bnd}\ B\ L\ p\ r := \sqrt[p]{\frac{\|r\|}{\mathit{bound}_{T-\Delta_0}\ B\ L}}$$

Given a one-step method *gf* satisfying the assumptions of *convergent*, we show (inductively on the number of the step j) for a small enough step size that *gf* is convergent: the global error $\|u\ \Delta_j - \mathit{gf}\ \Delta\ \psi\ x_0\ j\ \|$ is in $\mathcal{O}(h^p)$.

Theorem 4 (Convergence of One-Step methods).

$$\begin{aligned}&\mathit{convergent}\ u\ B\ p\ \psi\ r\ L \Rightarrow h_{max} \leq \mathit{step\text{-}bnd}\ B\ L\ p\ r \Rightarrow\\ &\quad (\forall j.\ \Delta_j \leq T \Rightarrow \|u\ \Delta_j - \mathit{gf}\ \Delta\ \psi\ x_0\ j\ \| \leq \mathit{bound}_{T-\Delta_0}\ B\ L \cdot {h_{max}}^p)\end{aligned}$$

5.3 Stability

Since we want to limit the precision of our calculations for reasons of efficiency we need to take the sensitivity against (rounding) errors into account. This is

captured by the notion of stability. For a one-step method defined by ψ, we want to study the effect of small perturbations in every step.

For this, we introduce (as in Reinhardt [16]) an error function s and an initial error s_0 and study the perturbed one-step method defined by ψ_s

$$\psi_s\ h\ t\ x := \psi\ h\ t\ x + s\ h\ t\ x$$

stable Small perturbations do not affect the results of a convergent one-step method too much if we assume a convergent ideal one-step method defined by ψ, a sufficiently small step size, and errors in the order of the step size (the textbook states the theorem for 'sufficiently small' errors, to obtain an explicit bound we basically make the perturbations part of the error we allow for consistency). We summarize this set of assumptions in the definition of *stable*:

$$\begin{aligned}\mathit{stable}\ u\ B\ p\ \psi\ r\ L\ s\ s_0 := \mathit{convergent}\ u\ B\ p\ \psi\ r\ L \wedge \\ h_{max} \leq \mathit{step\text{-}bnd}\ B\ L\ p\ \frac{r}{2} \wedge s_0 \leq \mathit{bound}_0\ B\ L \cdot {h_{max}}^p \wedge \\ (\forall j.\ \|s\ h_j\ \Delta_j\ \ (\mathit{gf}\ \Delta\ \psi_s\ (x_0 + s_0)\ j)\| \leq B \cdot {h_j}^p)\end{aligned}$$

Under these assumptions, we can show that the error between the distorted $\mathit{gf}\ \Delta\ \psi_s\ (x_0 + s_0)$ and the ideal one-step method $\mathit{gf}\ \Delta\ \psi\ x_0$ is in $\mathcal{O}(h^p)$:

Theorem 5 (Stability of one-step methods).

$$\begin{aligned}&\mathit{stable}\ u\ B\ p\ \psi\ r\ L\ s\ s_0 \Rightarrow \\ &\forall j.\ \Delta_j \leq T \Rightarrow \|\mathit{gf}\ \Delta\ \psi\ x_0\ j - \mathit{gf}\ \Delta\ \psi_s\ (x_0 + s_0)\ j\| \leq \mathit{bound}_{T-\Delta_0}\ B\ L \cdot {h_{max}}^p\end{aligned}$$

The textbook proof contains an induction quite similar to the one for the proof of convergence, and in fact, we managed to generalize the common part (the accumulation of an error) which we could re-use in both proofs.

6 Euler method

In this section, we define a simple one-step method, namely the Euler method. We show that the Euler method applied to an IVP is consistent and therefore convergent. For reasons of efficiency, we introduce an approximate implementation of the Euler method on floating point numbers and show that it is stable. We conclude that the approximate Euler method works correctly.

6.1 Definitions

We now assume an IVP *ivp* with domain $I \times D := [\Delta_0; T] \times \mathcal{R}_{b+r}(x_0)$ and take the assumptions from $\mathit{rect}\ T\ (b + r)\ L$. We want to approximate the solution of this IVP with the Euler method on a grid Δ. The Euler method connects a point (t, x) with its subsequent point by a line with the slope given by $f(t, x)$ – independently of the step size. Notice that in contrast to the previous sections,

we restrict ourselves and the Euler-method to the univariate case ($f :: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, *solution* $:: \mathbb{R} \to \mathbb{R}$): The proof requires Taylor series expansion, which – in Isabelle/HOL – is only available for the one-dimensional case.

$$\psi^f_{\mathsf{euler}}\ h\ t\ x := f(t,x)$$
$$\textit{euler}^f\ \Delta\ x_0\ j := \textit{gf}\ \Delta\ \psi^f_{\mathsf{euler}}\ x_0\ j; \qquad \textit{euler}^f :: (\mathbb{N} \to \mathbb{R}) \to \mathbb{R} \to \mathbb{N} \to \mathbb{R}$$

6.2 Euler on $\mathbb{R}$

Recall Theorem 4 about the convergence of one step methods and its set of assumptions *convergent*. The Euler method is clearly a one-step method. In order to prove convergence for the Euler method, we need to show that it is Lipschitz continuous and consistent with the solution.

In *rect* $T\ (b + r)\ L$ we have the assumption that f is Lipschitz continuous, hence the Euler method is Lipschitz continuous, too.

We show consistency with the solution with a Taylor series expansion of the *solution* around t, which requires explicit bounds for the derivative of f. Recall that $\|f_c\|$ (as defined in the assumptions of *rect*) is a bound for the values of f on the domain $I \times D$. In *deriv-bnd* we summarize the fact that f' is the (total) derivative of f and that at every point in $I \times D$, the derivative in every possible direction the solution might take (bounded by $\|f_c\|$) is bounded by B. It follows that under these assumptions the Euler method is consistent with the *solution*.

$$\textit{deriv-bnd}\ f'\ B := (\forall y \in I \times D.\ \forall dx.\ \|dx\| \leq \|f_c\| \Rightarrow f'\ y\ (1, dx) \leq 2 \cdot B)$$
$$\textit{deriv-bnd}\ f'\ B \Rightarrow \textit{consistent solution}\ B\ 1\ \psi^f_{\mathsf{euler}}$$

The Euler method being consistent and Lipschitz continuous, we can conclude with Theorem 4 that the Euler method converges:

Theorem 6 (Convergence of Euler). *When Picard-Lindelöf guarantees a unique solution on a rectangular domain (*rect $T\ (b + r)\ L$*) and with explicit bounds on the derivative of f (*deriv-bnd $f'\ B$*), the Euler method converges for a small enough step size ($h_{max} \leq$ step-bnd $B\ L\ 1\ r$) against the solution:*

$$\textit{convergent solution}\ B\ 1\ \psi_{\mathsf{euler}}\ r\ L$$

6.3 Euler on $\mathbb{F}$

We decided to add an implementation of the Euler method on arbitrary-precision floats for efficiency reasons. We define the approximate Euler method $\widetilde{\textit{euler}}$ as a one-step method operating on floating point numbers. As an increment function, we work with an approximation $\widetilde{f}$ of f in the sense that $(\widetilde{f}\ (\widetilde{t}, \widetilde{x}))_{\mathbb{R}}$ approximates $f((\widetilde{t})_{\mathbb{R}}, (\widetilde{x})_{\mathbb{R}})$ for $\widetilde{t}, \widetilde{x} \in \mathbb{F}$. The error of the approximation may stem from truncating the floating point numbers for reasons of efficiency.

We show that the approximate Euler method works correctly as follows: From Theorem 6, we have a bound on the error between the result of the ideal Euler

method and the solution (convergence). We apply Theorem 5 to obtain a bound on the error between the ideal and the approximate Euler method (stability). We summarize the required assumptions in *euler-rounded*: We need *rect* and *deriv-bnd* to show convergence. In addition to that, we need bounds on the error of the approximations $\widetilde{x_0}$ and $\widetilde{f}$ to obtain stability.

$$\begin{aligned}&\textit{euler-rounded}\ b\ r\ L\ f'\ B := \ \textit{rect}\ T\ (b+r)\ L \wedge \textit{deriv-bnd}\ f'\ B \wedge\\ &\quad t_0 = (\widetilde{\Delta_0})_{\mathbb{R}} \wedge \|x_0 - (\widetilde{x_0})_{\mathbb{R}}\| \leq \textit{bound}_0\ B\ L \cdot (\widetilde{h_0})_{\mathbb{R}} \wedge\\ &\quad (\forall j\ \widetilde{x}.\ \|f((\widetilde{\Delta_j})_{\mathbb{R}}, (\widetilde{x})_{\mathbb{R}}) - (\widetilde{f}(\widetilde{\Delta_j}, \widetilde{x}))_{\mathbb{R}}\| \leq B \cdot (\widetilde{h_j})_{\mathbb{R}}\end{aligned}$$

One subtle point is the fact that Theorem 5 applies only to one-step methods on real numbers. We therefore need to instantiate the theorem with the perturbed increment function $f_s(t,x) := (\widetilde{f}\,((t)_{\mathbb{F}}, (x)_{\mathbb{F}}))_{\mathbb{R}}$ and show that the result of $\textit{euler}^{f_s}$ equals $(\widetilde{\textit{euler}^{\widetilde{f}}})_{\mathbb{R}}$, which is easy since $\textit{euler}^{f_s}$ operates exclusively on real numbers representable as floating point numbers.

Having convergence of the ideal Euler method (depending on the step size) and stability of the approximate Euler method (depending on the rounding error), we can approximate IVPs: The execution of $\widetilde{\textit{euler}^{\widetilde{f}}}$ on a grid $\widetilde{\Delta}$ results in an error compared to *solution* that can be made arbitrarily small by choosing smaller step sizes.

Theorem 7 (Convergence of the approximate Euler method on $\mathbb{F}$). *When Picard-Lindelöf guarantees a unique solution on a rectangular domain (*$\textit{rect}\ T\ (b+r)\ L$*) and with explicit bounds on the derivative of f and appropriate error bounds for the approximations $(\widetilde{x_0})_{\mathbb{R}}$ and $(\widetilde{f})_{\mathbb{R}}$ (*$\textit{euler-rounded}\ b\ r\ L\ f'\ B$*), the approximate Euler method converges for a small enough step size ($\widetilde{h_{max}} \leq \textit{step-bnd}\ B\ L\ 1\ \frac{r}{2}$) against the solution (for every j with $(\widetilde{\Delta_j})_{\mathbb{R}} \leq T$):*

$$\left\|\textit{solution}\ (\widetilde{\Delta_j})_{\mathbb{R}} - (\widetilde{\textit{euler}^{\widetilde{f}}}\ \widetilde{\Delta}\ \widetilde{x_0}\ j)_{\mathbb{R}}\right\| \leq 2 \cdot \textit{bound}_{T-t_0}\ B\ L \cdot (\widetilde{h_{max}})_{\mathbb{R}}$$

7 Example: $\dot{u}\ t = u^2 - t$

As a simple case-study, we chose the ODE $\dot{u}\ t = (u\ t)^2 - t$ which does not admit a closed formula as solution. In this section we show how we compute $u\ \frac{1}{2}$. First we introduce an IVP depending on the user supplied values $\widetilde{t_0}, \widetilde{x_0}, \widetilde{T}, b, r$:

$$\begin{aligned}f\ (t,x) &:= x^2 - t\\ I \times D &:= [(\widetilde{t_0})_{\mathbb{R}}; (\widetilde{T})_{\mathbb{R}}] \times \mathcal{B}_{b+r}((\widetilde{x_0})_{\mathbb{R}})\\ (t_0, x_0) &:= ((\widetilde{t_0})_{\mathbb{R}}, (\widetilde{x_0})_{\mathbb{R}})\end{aligned}$$

We then analyze this IVP: We provide the Lipschitz-constant L and the boundary B. We prove a bound of $\|f_c\|$, and the (total) derivative $\dot{f}$ of f:

$$\begin{aligned}L &:= 2 \cdot \mathsf{max}\ \|(\widetilde{x_0})_{\mathbb{R}} - (b+r)\|\ \|(\widetilde{x_0})_{\mathbb{R}} + (b+r)\|\\ B &:= 2 \cdot \mathsf{max}\ \|(\widetilde{x_0})_{\mathbb{R}} - b\|\ \|(\widetilde{x_0})_{\mathbb{R}} + b\| + \tfrac{1}{2}\\ \|f_c\| &\leq (\mathsf{max}\ \|(\widetilde{x_0})_{\mathbb{R}} - b\|\ \|(\widetilde{x_0})_{\mathbb{R}} + b\|)^2 + \mathsf{max}\ \|(\widetilde{t_0})_{\mathbb{R}}\|\ \|(\widetilde{T})_{\mathbb{R}}\|\\ \dot{f}\ (t,x)\ (dt,dx) &= 2 \cdot x \cdot dx - dt\end{aligned}$$

The Euler method $\widetilde{\mathit{euler}^{\widetilde{f}}}$ is defined with $\widetilde{f}(\widetilde{t},\widetilde{x}) := \mathit{round}_p(\widetilde{x}^2 - \widetilde{t})$ on an equidistant grid $\widetilde{\Delta}_j := \widetilde{t_0} + j \cdot \widetilde{h}$. For a fast computation we use the rounding operator $\mathit{round}_p\ x$ which reduces the precision of x, i.e. $\|\mathit{round}_p\ x - x\| \leq 2^{-p}$.

We now set the parameters to

$$\widetilde{t_0} := 0, \quad \widetilde{x_0} := 0, \quad b := 1, \quad r := 2^{-8}, \quad T := 1, \quad \widetilde{h} := 2^{-14}, \quad \text{and} \quad p := 14.$$

The error to the solution is bounded by 0.001 due to Theorem 7. We discharge all assumptions with the approximation method [7], as all parameters are fixed.

Theorem 8 (Approximation of the solution to $\dot{u}\ t = u^2 - t$).

$$\forall j \leq 2^{13}.\ \left\| \mathit{solution}\ (\widetilde{\Delta_j})_{\mathbb{R}} - \widetilde{\mathit{euler}^{\widetilde{f}}}\ \widetilde{\Delta}\ \widetilde{x_0}\ j \right\| \leq 0.001$$

The execution of $\widetilde{\mathit{euler}^{\widetilde{f}}}\ \widetilde{\Delta}\ \widetilde{x_0}\ 2^{13}$ in the target language SML (where we represent floating point numbers as pairs of integers) returns $-33140952 \cdot 2^{-28} \approx -0.123\ldots$ and takes about 2 seconds on a Core$^{\text{TM}}$2 Duo (E8335) and 2.93 GHz. We put everything together and obtain a result that is correct for two digits:

$$u\ \frac{1}{2} = \mathit{solution}\ (\widetilde{\Delta}_{2^{13}})_{\mathbb{R}} \in [-0.124\ldots; -0.122\ldots]$$

This proposition bypassed the LCF kernel of Isabelle since we trust code generation for the approximation method and the evaluation of $\widetilde{\mathit{euler}^{\widetilde{f}}}$, but it could (at least in principle) be proved by Isabelle's simplifier.

In order to evaluate the overestimations that stem from the proofs, it is worth noticing that one gets a result inside the same bounds with the choice of 2^{-11} instead of 2^{-14} for stepsize and rounding error. In an experiment with $\dot{u}\ t = u$ (i.e. $u\ t := e^t$), the actual error is more than a factor 22 smaller than the estimated error. Notice also that for our choice of parameters, $\frac{1}{2}$ is the maximum argument where our theorems guarantee correctness.

8 Conclusion and Discussion

We formalized the Picard-Lindelöf theorem to show the existence of a unique *solution* for IVPs in the multivariate case ($\mathbb{R} \to \mathbb{R}^n$). We conducted an analysis of the numerical errors of one-step methods that approximate arbitrary functions in $\mathbb{R} \to \mathbb{R}^n$, which we could use to show that an ideal Euler method euler^f (without rounding errors) approximates the solution (but only in the univariate case, since a multivariate version of Taylor series expansion has not been formalized in Isabelle/HOL yet). Analysis of stability for one-step methods yielded stability for the Euler method: small errors $f - \widetilde{f}$ do not affect the global behaviour of an approximate Euler method $\widetilde{\mathit{euler}^{\widetilde{f}}}$. See these relations summarized in Fig. 2.

Most of the theorems we presented require a large set of assumptions, where the use of locales [4] helped us structuring the theories (compare Fig. 3): We presented the basic Picard-Lindelöf theorem under assumptions with restrictions

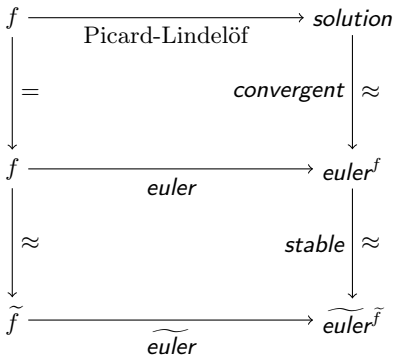

Fig. 2. Relationship between the differential f and the different approximations

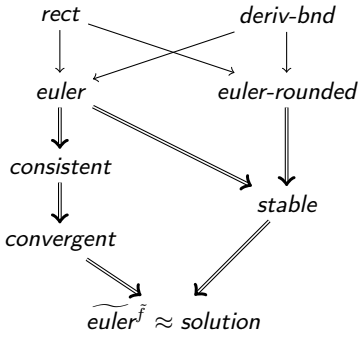

Fig. 3. Context hierarchy

on the size of the interval *bnd-strip*, then dropped this restriction in *strip*. More realistic applications require restricting the codomain of the solution in *rect* and a variant of the Picard-Lindelöf in the context of open domains of *open-domain* is of mathematical interest. We showed that consistent one step methods (*consistent*) are convergent (*convergent*) and (with additional assumptions) stable (*stable*) and showed these properties for the Euler method. We could conclude that an approximate Euler method converges against the solution.

The Euler method is rarely used in real applications but was relatively easy to analyze. However, the results from one-step methods apply to the widely used Runge-Kutta methods, therefore one can profit from our developments when one implements Runge-Kutta methods of higher order (e.g. the method of Heun or the "classical" Runge-Kutta method) where one only needs to show consistency in order to obtain results about convergence.

In order to obtain explicit bounds for the error of the Euler method in a concrete application, we rely on code generation. The user needs to provide proofs that the concrete application satisfies the assumptions of the contexts in which the error bounds hold. To some extent, an analysis of the differential equation is also necessary when one wants to evaluate the quality of the approximation obtained by some arbitrary numerical method. It might still be desirable to provide more automatization e.g. computing the bounds of the derivative or for deriving a minimum step size automatically.

Our development is available in the AFP [8] and consists of a total of 5020 lines, 1715 for $\overline{\mathcal{C}}$ and the floating-point numbers, 1696 for IVPs, 1217 for one-step methods, and 392 for the example.

References

1. Boldo, S., Clément, F., Filliâtre, J.C., Mayero, M., Melquiond, G., Weis, P.: Formal proof of a wave equation resolution scheme: The method error. In: Kaufmann, M.,

Paulson, L.C. (eds.) ITP 2010. LNCS, vol. 6172, pp. 147–162 (2010)
2. Bornemann, V., Deuflhard, P.: Scientific computing with ordinary differential equations (2002)
3. Haftmann, F., Nipkow, T.: Code generation via higher-order rewrite systems. In: Blume, M., Kobayashi, N., Vidal, G. (eds.) Functional and Logic Programming (FLOPS 2010). LNCS, vol. 6009, pp. 103–117 (2010)
4. Haftmann, F., Wenzel, M.: Local theory specifications in Isabelle/Isar. In: Berardi, S., Damiani, F., de'Liguoro, U. (eds.) TYPES 2008, LNCS, vol. 5497 (2009)
5. Harrison, J.: Theorem Proving with the Real Numbers. Ph.D. thesis (1996)
6. Harrison, J.: A HOL theory of Euclidean space. In: Hurd, J., Melham, T. (eds.) Theorem Proving in Higher Order Logics, 18th International Conference, TPHOLs 2005. Lecture Notes in Computer Science, vol. 3603, pp. 114–129 (2005)
7. Hölzl, J.: Proving inequalities over reals with computation in Isabelle/HOL. In: Reis, G.D., Théry, L. (eds.) Programming Languages for Mechanized Mathematics Systems (ACM SIGSAM PLMMS'09). pp. 38–45 (2009)
8. Immler, F., Hölzl, J.: Ordinary Differential Equations. Archive of Formal Proofs (Apr 2012), http://afp.sf.net/entries/Ordinary_Differential_Equations.shtml, Formal proof development
9. Krebbers, R., Spitters, B.: Computer certified efficient exact reals in Coq. In: Conference on Intelligent Computer Mathematics (CICM 2011). LNAI, vol. 6824, pp. 90–103 (2011)
10. Krebbers, R., Spitters, B.: Type classes for efficient exact real arithmetic in Coq. CoRR abs/1106.3448 (2011)
11. Melquiond, G.: Proving bounds on real-valued functions with computations. In: Armando, A., Baumgartner, P., Dowek, G. (eds.) International Joint Conference on Automated Reasoning (IJCAR'08). LNAI, vol. 5195, pp. 2–17 (2008)
12. Muñoz, C., Lester, D.: Real number calculations and theorem proving. In: Hurd, J., Melham, T. (eds.) Theorem Proving in Higher Order Logics (TPHOLs 2005). LNCS, vol. 3603, pp. 195–210 (2005)
13. Obua, S.: Flyspeck II: The Basic Linear Programs. Ph.D. thesis, München (2008)
14. O'Connor, R., Spitters, B.: A computer verified, monadic, functional implementation of the integral. Theoretical Computer Science 411(37), 3386–3402 (2010)
15. O'Connor, R.: Certified exact transcendental real number computation in Coq. In: Mohamed, O., Muñoz, C., Tahar, S. (eds.) Theorem Proving in Higher Order Logics (TPHOLs'08), LNCS, vol. 5170, pp. 246–261 (2008)
16. Reinhardt, H.J.: Numerik gewöhnlicher Differentialgleichungen. de Gruyter (2008)
17. Spitters, B.: Numerical integration in Coq (Nov 2010), Mathematics, Algorithms, and Proofs (MAP'2010), www.unirioja.es/dptos/dmc/MAP2010/Slides/Slides/talkSpittersMAP2010.pdf
18. Walter, W.: Ordinary Differential Equations. Springer, 1 edn. (1998)