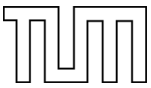

Technische Universität München

Department of Mathematics

Bachelor's Thesis

Theory and Implementation of the Adaptive Explicit Midpoint Rule Including Order and Stepsize Control

Konstantin Althaus

Supervisor: Prof. Dr. Folkmar Bornemann

Submission Date: December 18, 2018

I assure the single handed composition of this bachelor's thesis only supported by declared resources.

Garching, December 18, 2018

Abstract

In this Bachelor's thesis we develop the theory of extrapolating explicit methods for initial value problems. The extrapolation of the explicit midpoint rule is discussed in detail. Furthermore we present the derivation and implementation of an algorithm with order and stepsize control. Our implementation is written in the programming language Julia.

In dieser Bachelorarbeit entwickeln wir die Theorie der Extrapolation expliziter Verfahren für Anfangswertprobleme. Detailliert gehen wir auf die Extrapolation der expliziten Mittelpunktsregel ein. Des Weiteren werden wir Herleitung und Implementierung eines Algorithmus mit Ordnungs- und Schrittweitensteuerung präsentieren. Unsere Implementierung ist in der Programmiersprache Julia verfasst.

Contents

1 Introduction

An implementation of the algorithm whose theoretical background is content of this thesis can be found here:

`https://github.com/AlthausKonstantin/Extrapolation`

The code is written in the programming language Julia.
In this chapter we discuss the mathematical prerequisites that are necessary for the rest of the thesis. We cite or proof some basic theoretic results in our own words to be able to present the thesis' content in a consistent and precise manner. Finally note that we will mostly follow the notation and nomenclature used in [3].

Notation The notation in this thesis, that is not common or used in [3], is given here.

- $[1:N] := \{1, 2, \dots, N-1, N\}$ for $N \in \mathbb{N}$.
- $[x_1 \cdots x_N] = [x_n]_{n=1:N} \in \mathbb{R}^{d \times N}$. is the horizontal concatenation of the vectors $(x_n)_{n=1:N} \subset \mathbb{R}^d$. Furthermore $[x_1; \dots; x_N] \in \mathbb{R}^{N \cdot d}$ denotes a vertical concatenation.
- $|x| := [|x_1|; \dots; |x_d|]$ for a vector $x \in \mathbb{R}^d$.
- $\mathbb{1}_d = [1; \dots; 1] \in \mathbb{R}^d$.

1.1 Initial Value Problems

The underlying mathematical problem we are dealing with is approximating the solution of an *ordinary differential equation*

$$x'(t) = f(t, x(t)).$$

More specific we consider an *initial value problem* (IVP)

$$x' = f(t, x), \quad x(t_0) = x_0, \tag{1.1}$$

Here it is common practice to suppress the variable t of x. At first let us clarify the notation and nomenclature.

Definition 1.1. $f : I \times \Omega_0 \to \mathbb{R}^d$ is called the *right side*. It is defined on the *augmented state space* $\Omega := I \times \Omega_0$. We assume the interval $I \subseteq \mathbb{R}$ and the *state space* $\Omega_0 \subseteq \mathbb{R}^d$ to be open and nonempty. In several applications the scalar t is interpreted as *time*.
Furthermore we assume $(t_0, x_0) \in \Omega$; x_0 is referred to as the *initial value*.
A differentiable function $x : J \to \Omega_0$ with $t_0 \in J \subseteq I$ is called a *solution* of Equation 1.1 on the interval J if $x'(t) = f(t, x(t)) \quad \forall\, t \in J$ and $x(t_0) = x_0$.

Now the question arises if Equation 1.1 has a solution and if so whether it is unique. We cite a well known result:

Theorem 1.2 (Existence and Uniqueness)**.** *If $f \in \mathcal{C}^q(\Omega, \mathbb{R}^d)$ with $q \in \mathbb{N}$ there is for any initial data $(t_0, x_0) \in \Omega$ a unique solution $x \in \mathcal{C}^{q+1}(I_{\max}, \Omega_0)$ of Equation 1.1. $I_{\max}$ is an open interval (t_-, t_+) defined as*

$$I_{\max} := I_{\max}(t_0, x_0) := \bigcup \left\{ (a, b) \subseteq I : \ x \in \mathcal{C}^{q+1}\left((a, b), \Omega_0\right) \text{ solves Equation 1.1} \right\}.$$

Proof. One can find a proof of the existence of a solution in [9, Theorem 1.9], uniqueness in [9, Theorem 1.10] and existence of *the* maximal solution in [9, Theorem 1.9]. □

Remark 1.3. Existence and uniqueness can be obtained on weaker assumptions. Namely continuity and local Lipschitz continuity with respect to the state variable are sufficient conditions for the right side f – cf. [9, Theorem 1.10].
But in the next chapter it will become apparent that for our endeavor only IVPs with a higher regularity are of interest. Thus from now we assume that f is at least continuous differentiable.

Now we introduce a useful concept of thinking about the solution x of Equation 1.1, which reflects the deterministic dependency on the initial data (t_0, x_0).

Definition 1.4 (Continuous Evolution)**.** The *continuous evolution* $\Phi : \mathbb{R}^d \to \mathbb{R}^d$ is a (in general non linear) operator with two scalar parameters. It is defined pointwise for $(t_0, x_0) \in \Omega$ and $t \in I_{\max(t_0, x_0)}$ by the solution x of Equation 1.1:

$$\Phi^{t, t_0} x_0 := x(t)$$

The evolution Φ has some benign properties which we will exploit throughout the thesis.

Lemma 1.5. *The evolution $(s, t, x) \mapsto \Phi^{s,t} x$ is continuous on its domain $D_\Phi := \{(s, t, x) \in I \times \Omega : \ s \in I_{\max}(t, x)\}$.*

Proof. The claim is proven in [9, Theorem 1.23]. □

We conclude this section by proving an even stronger property of the continuous evolution Φ.

Theorem 1.6. *Assume $f \in \mathcal{C}^q(\Omega, \mathbb{R}^d)$ for some $q \in \mathbb{N}$. Then also the partial derivatives $\partial_s^n \Phi^{s,t} x$ for $n \in [1 : q + 1]$ are continuous on D_Φ.*

Proof. We show by induction over n: For $n \in [1 : q + 1]$ there are functions $d_n \in \mathbb{C}^{q+1-n}(\Omega, \mathbb{R}^d)$ such that $d_n(s, \Phi^{s,t} x) = \partial_s^n \Phi^{s,t} x$ on D_Φ. The base case is provided by $d_1 \equiv f$. Now let $1 \leq n \leq q$. The chain rule implies:

$$\partial_s^{n+1} \Phi^{s,t} x = \partial_s d_n \left(s, \Phi^{s,t} x\right) = \partial_\sigma d_n \left(\sigma, \Phi^{s,t} x\right) \big|_{\sigma = s} + \partial_\xi d_n \left(s, \xi\right) \big|_{\xi = \Phi^{s,t} x} f(s, \Phi^{s,t} x).$$

Thus we set $d_{n+1} := \partial_t d_n + (\partial_x d_n) f$. Applying the induction hypothesis we note $d_{n+1} \in \mathcal{C}^k(\Omega, \mathbb{R}^d)$ with $k = \min(q, (q + 1) - n - 1) = (q + 1) - (n + 1)$.
Now it is immediate from Lemma 1.5 that $d_n(s, \Phi^{s,t} x) = \partial_s^n \Phi^{s,t} x$ is continuous on D_Φ for $n \in [1 : q + 1]$. □

1.2 Discretizing Φ Leads to Ψ

After establishing the existence and uniqueness of the solution of Equation 1.1, we "solve" the IVP numerically on a given interval $[t_0, T] \subset I_{\max}$. Here we actually compute a discrete *grid function* $x_\Delta : \Delta \to \Omega$, defined on a *grid* $\Delta := (t_n : \ n \in [0 : N_\Delta]) \subset [t_0, T]$, to approximate the solution x on Δ: $x_\Delta \approx x\big|_\Delta$.
That process can be interpreted as discretizing $[t_0, T]$ and Φ. The continuous interval $[t_0, T]$ becomes a discrete $N_\Delta + 1$ tuple of grid points which we assume to be sorted in *strictly ascending* order. Analogously the infinite family of operators $(\Phi^{t,s})_{t,s\in[t_0,T]}$ turns into $(\Psi^{t,s})_{t,s\in\Delta}$. Naturally Ψ is called *discrete evolution*. The approximation x_Δ is constructed recursively with the operator Ψ:

$$\begin{aligned} x_\Delta(t_0) &:= x_0 && \text{(1.2a)} \\ x_\Delta(t_n) &:= \Psi^{t_n,t_{n-1}} x_\Delta(t_{n-1}) \quad n \in [1 : N_\Delta]. && \text{(1.2b)} \end{aligned}$$

We denote the *stepsize* of the nth *time step* by $\tau_n := t_n - t_{n-1}$ and the *maximal stepsize* attained in the grid Δ by τ_Δ. Note that also $T < t_0$ is possible. The nodes in Δ must then be ordered strictly descending, τ_n becomes negative and τ_Δ is the maximum of $(|\tau_n|)_{n=1:N_\Delta}$. For simplicity though we will mostly confine ourself to the case $t_0 < T$.
Since the the computation of each time step is solely relying on information of the previous one, such algorithms are subsumed under the label *one step methods*.
Now we present the reasoning behind actually constructing the operator Ψ. As the solution's domain is one–dimensional, it is an intuitive approach to discretize Φ in the t by a Taylor expansion:

$$\Phi^{t+\tau,t} x = \sum_{i=0}^{p} a_i \tau^i + \mathcal{O}(\tau^{p+1}) \stackrel{!}{=} \Psi^{t+\tau,t} x + \mathcal{O}(\tau^{p+1}) \quad (\tau \to 0).$$

This course of attack turns out to be a fruitful one. In particular because information of the derivatives $x^{(i)}$ is (more or less) readily available by mere evaluations of the right side f.
On a machine one usually implements the evaluation of $\Psi^{t+\tau,t} x$ by computing the *increment function* ψ, which depends on $(t, x) \in \Omega$ and the stepsize τ. It is defined by

$$\Psi^{t+\tau,t} x = x + \tau\psi(t, x, \tau).$$

Note that the here outlined approach inevitably leads to operators Ψ which are in general not well defined for an arbitrary τ because the Taylor polynomial is a priori a local approximation.
Based on that observation we denote by $\tau^*(t, x) > 0$ the supremum such that $\Psi^{t+\tau,t} x$ and $\Phi^{t+\tau,t} x$ are well defined in Ω_0 for $|\tau| < \tau^*(t, x)$. Without loss of generality we can assume that $\tau^*(t, x)$ is continuous on Ω. Based on the the definition of τ^* we also define the domain of the increment function ψ:

$$\Sigma := \{(t, x, \tau) \in \Omega \times \mathbb{R} : \ |\tau| < \tau^*(t, x)\}.$$

.

1.3 Consistency and Convergence

In this section we want to quantify how well x_Δ approximates $x\big|_\Delta$. For this we distinguish between a local and a global error. The *local error* measures how much the discrete evolution differs after one time step from its continuous counterpart *if starting in the same point*:

$$\epsilon(t,x,\tau) := \Phi^{t+\tau,t}x - \Psi^{t+\tau,t}x \quad (t,x,\tau) \in \Sigma.$$

Since x_Δ is computed recursively the local errors sum up and are *propagated into the future*. The resulting error is called the *global* or *discretization error*

$$\epsilon_\Delta := x\big|_\Delta - x_\Delta.$$

A quantification of the local error is given by the *consistency* of Ψ.

Definition 1.7 (Consistency)**.** The discretization Ψ is *consistent of order* $p \in \mathbb{N}$ if firstly its increment function ψ *inherits the regularity of the right side*, i.e. $f \in \mathcal{C}^p(\Omega, \mathbb{R}^d)$ implies that ψ is also p–times continuously differentiable. And secondly if the local error is *uniformly of order p+1*:

$$\forall\, K \Subset \Omega \quad \exists\, \hat\tau, C > 0 : \ \|\epsilon(t,x,\tau)\| \le C|\tau|^{p+1} \quad \forall\,(t,x) \in K, |\tau| \le \hat\tau.$$

The following theorem is a concretization of the ansatz mentioned in Section 1.2.

Theorem 1.8 (Local Error Representation)**.** *Consider the IVP Equation 1.1 with a right side $f \in \mathcal{C}^q(\Omega, \mathbb{R}^d)$. Let Ψ be a discrete evolution with its increment function $\psi \in \mathcal{C}^r(\Sigma, \mathbb{R}^d)$ and set $N := \min(q,r) \in \mathbb{N}$. Then there are unique* coefficient functions

$$\epsilon_n \in \mathcal{C}^{N+1-n}(I_{\max}(t_0,x_0), \mathbb{R}^d) \quad n \in [1:N]$$

and a remainder

$$\rho \in \mathcal{C}(D_\rho, \mathbb{R}^d) \text{ with } D_\rho := \{(t,\tau) \in I_{\max}(t_0,x_0) \times \mathbb{R} : \ |\tau| < \tau^*(t,x(t))\}$$

such that

$$\Phi^{t+\tau,t}x(t) - \Psi^{t+\tau,t}x(t) = \sum_{n=1}^{N} \epsilon_n(t)\tau^n + \rho(t,\tau)\tau^{N+1} \tag{1.3}$$

for $(t,\tau) \in D_\rho$. Furthermore Ψ has consistency order $p \le N$ if and only if

$$\forall\, n \in [1:p], (t_0,x_0) \in \Omega : \ \epsilon_n(t_0) = 0.$$

Proof. The coefficient functions and the remainder are obtained by Taylor expansion. Let $(t,x,\tau) \in \Sigma$ be arbitrary. Firstly we have

$$\begin{aligned}
\Phi^{t+\tau,t}x &= \sum_{n=0}^{N} \frac{\partial_s^n \Phi^{s,t}x\big|_{s=t}}{n!}\tau^n + \int_0^\tau \frac{\partial_s^{N+1}\Phi^{s,t}x\big|_{s=\lambda}}{N!}(\tau-\lambda)^N d\lambda \\
&= x + \sum_{n=1}^{N} \frac{\partial_s^n \Phi^{s,t}x\big|_{s=t}}{n!}\tau^n + \tau^{N+1}\int_0^1 \frac{\partial_s^{N+1}\Phi^{s,t}x\big|_{s=\tau-\lambda\tau}}{N!}\lambda^N d\lambda
\end{aligned}$$

Secondly we get

$$\begin{aligned}\Psi^{t+\tau,t}x &= x+\tau\left(\sum_{n=0}^{N-1}\frac{\partial_\sigma^n\psi(t,x,\sigma)\big|_{\sigma=0}}{n!}\tau^n+\int_0^\tau\frac{\partial_\sigma^N\psi(t,x,\sigma)\big|_{\sigma=\lambda}}{(N-1)!}(\tau-\lambda)^{N-1}d\lambda\right)\\ &= x+\sum_{n=1}^{N}\frac{\partial_\sigma^{n-1}\psi(t,x,\sigma)\big|_{\sigma=0}}{(n-1)!}\tau^n+\tau^{N+1}\int_0^1\frac{\partial_\sigma^N\psi(t,x,\sigma)\big|_{\sigma=\tau-\lambda\tau}}{(N-1)!}\lambda^{N-1}d\lambda\end{aligned}$$

Subtracting the second expansion from the first yields

$$\begin{aligned}\epsilon_n(t,x) &:= \frac{\partial_s^n\Phi^{s,t}x\big|_{s=t}}{n!}-\frac{\partial_\sigma^{n-1}\psi(t,x,\sigma)\big|_{\sigma=0}}{(n-1)!}\quad n\in[1:N],\\ \rho(t,x,\tau) &:= \int_0^1\frac{\partial_s^{N+1}\Phi^{s,t}x\big|_{s=\tau-\lambda\tau}}{N!}\lambda^N-\frac{\partial_\sigma^N\psi(t,x,\sigma)\big|_{\sigma=\tau-\lambda\tau}}{(N-1)!}\lambda^{N-1}d\lambda.\end{aligned}$$

By virtue of Theorem 1.6 we see that $\epsilon_n\in\mathcal{C}^{N+1-n}(\Omega,\mathbb{R}^d)$ is well defined. Furthermore the theorem of parameter dependent integrals [5, Theorem 5.6] shows that ρ is continuous on Σ.
Now we show that the functions $(\epsilon_n)_{n=1:p}$ vanish if and only if Ψ is consistent of order p. First assume $\epsilon_n\equiv 0$ on Ω for $n\in[1:p]$ and choose an arbitrary subset $K\Subset\Omega$. We need to prove the existence of some $C,\hat\tau>0$ such that $\|\epsilon(t,x,\tau)\|$ is bounded by $C|\tau|^{p+1}$ on $K\times[-\hat\tau,\hat\tau]$. Since τ^* is continuous on the compact set $\overline{K}$, we can choose a $\hat\tau\in\mathbb{R}^+$ such that

$$\hat\tau<\inf\{\tau^*(t,x):\ (t,x)\in K\}.$$

Now also $K\times[-\hat\tau,\hat\tau]$ is compactly embedded in Σ and there is $C>0$ such that

$$\|\epsilon(t,x,\tau)\|=\left\|\sum_{n=p+1}^{N}\epsilon_n(t,x)\tau^{n-p-1}+\rho(t,x,\tau)\tau^{N-p}\right\||\tau|^{p+1}$$

since the local error is continuous on $K\times[-\hat\tau,\hat\tau]$. To prove the reversed implication we assume Ψ to be consistent of order p and let $(t,x)\in\Omega$ be arbitrary. $\|\epsilon(t,x,\tau)\|\le C|\tau|^{p+1}$ for $|\tau|\le\hat\tau$ then implies

$$\left\|\sum_{n=1}^{N}\epsilon_n(t,x)\tau^{p-q}+\rho(t,x,\tau)\tau^{N+1-q}\right\|\le C\tau^{p+1-q}\quad\forall\,\tau\in(0,\hat\tau],q\in[2:p+1].\tag{1.4}$$

Taking the limit $\tau\downarrow 0$ on both sides successively for each q shows $\epsilon_n(t,x)=0$ for $n\in[1:p]$. Finally the functions of the local error representation Equation 1.3 are defined as $\epsilon_n(t):=\epsilon_n(t,x(t))$ and $\rho(t,\tau):=\rho(t,x(t),\tau)$ where $t\mapsto x(t)\in\mathcal{C}^{N+1}(I_{\max}(t_0,x_0),\mathbb{R}^d)$ is the solution of Equation 1.1. □

Note that the proof of the preceding theorem shows that the local error is uniformly of of order $p+1$ if and only if it is pointwise of order $p+1$ - let us record this observation.

Corollary 1.9. *A discretization Ψ is already consistent of order $p\in\mathbb{N}$ if its increment function ψ* inherits the regularity of the right side *and the local error is* pointwise of order $p+1$*:*

$$\forall\,(t,x)\in\Omega\quad\exists\,\hat\tau,C>0:\ \|\epsilon(t,x,\tau)\|\le C|\tau|^{p+1}\quad\forall\,|\tau|\le\hat\tau.$$

Proof. Given the regularity of f and ψ we obtain an expansion of the local error in (t_0, x_0). Now as shown in the last proof the pointwise estimate suffices to conclude by equating the coefficients that the coefficient functions $(\epsilon_n)_{n=1:p}$ vanish in t_0. □

That makes it easy to decide whether a given one step method is consistent, i.e. consistent of any order p. One simply has to verify that ψ is continuously differentiable if f itself is and that $\psi(t, x, 0) = f(t, x)$ on Ω.
Next we build a notion of convergence, i.e. the behavior of the global error as the grid Δ is refined.

Theorem 1.10 (Global Error Estimate)**.** *Let x be the solution of Equation 1.1 restricted to the interval $[t_0, T] \subset I_{\max}$.*
We approximate x with a discretization Ψ that is consistent of order p along the graph of *x, i.e.*

$$\exists\, C, \hat{\tau} > 0 \quad \forall\, t \in [t_0, T] :\ \|\epsilon(t, x(t), \tau)\| \leq C\tau^{p+1} \quad \forall\, |\tau| \leq \hat{\tau}.$$

If the right side f and the increment function ψ are continuously differentiable, there are constants $C_1, C_2 > 0$ such that for all grid functions x_Δ, which are produced by Ψ according to Equation 1.2 on a grid $\Delta \subset [t_0, T]$ that satisfies $\tau_\Delta \leq \hat{\tau}$, the global error can be estimated:

$$\left\|x\big|_\Delta(t) - x_\Delta(t)\right\| = \|\epsilon_\Delta(t)\| \leq C_1 \left(e^{C_2(t-t_0)} - 1\right) \tau_\Delta^p \quad \forall\, t \in \Delta. \tag{1.5}$$

Proof. See [3, Theorem 4.10]. □

1.4 Runge–Kutta Methods

The most famous one–step methods are the *Runge–Kutta methods*. Any Runge–Kutta method can be described by its *Butcher array*

$$\begin{array}{c|c} c & \mathcal{A} \\ \hline & b' \end{array}.$$

with $c = [c_1; \ldots; c_s], b = [b_1; \ldots; b_s] \in \mathbb{R}^s$ and $\mathcal{A} = [a_{i,j}]_{i,j=1:s} \in \mathbb{R}^{s\times s}$.
The method Ψ belonging to the array above is then defined as follows:

Definition 1.11 (Runge–Kutta methods)**.** The Runge–Kutta method associated with the triple $(\mathcal{A}, b, c)$ is given by

$$k_i(t, x, \tau) \ := \ f\left(t + c_i\tau, x + \tau \sum_{j=1}^{s} a_{i,j}k_j\right) \quad i \in [1 : s], \tag{1.6a}$$

$$\Psi^{t+\tau,t}x \ := \ x + \tau \sum_{i=1}^{s} b_i k_i. \tag{1.6b}$$

The k_i's are the *stages* and s is the *stage number* of Ψ. Ψ is an *explicit Runge–Kutta method* if $a_{i,j} = 0$ for $j \leq i$. Furthermore the method is *invariant with respect to autonomisation* (see [3, Lemma 4.16]) if $c = \mathcal{A}\mathbb{1}_s$.

2 Extrapolation

In this chapter we develop the framework of extrapolation. We start with its heuristic idea, give a precise definition of extrapolation methods and prove their relevant properties. Again we will follow the line of thought presented in [3].

2.1 The Idea of Extrapolation

Suppose we apply a consistent discretization Ψ of order p to the IVP of Equation 1.1 on the interval $[t_0, T]$. Assume that f is smooth and an equidistant grid Δ_n with stepsize $\tau_n = (T - t_0)/n$ is used. In the last chapter we have shown that

$$x_{\Delta_n}(T) \to x(T) \quad (n \to \infty).$$

Our plan is now to compute $x_{\Delta_n}(T)$ for some values of$n$, say for $n \in [0 : N]$, and then compute, just on basis of these values, an improved approximation of $x(T)$.
In order to make an educated guess about the limit of a sequence just based on its first $(N + 1)$ entries $x_{\Delta_0}(T), \dots x_{\Delta_N}(T)$ we need some regularity which we can exploit.
The very regularity needed here is proven in Section 2.4. There we will see that there exists a sequence of smooth functions $(e_n)_{n\in\mathbb{N}}$ such that for any equidistant grid Δ with a sufficiently small stepsize τ

$$\underbrace{x_\Delta(T)}_{X(\tau)} = \underbrace{x(T) - \sum_{n=0}^{N-1} e_{p+n}(T)\tau^{p+n}}_{P(\tau)} - \underbrace{r_\Delta^N(T)\tau^{p+N}}_{R(\tau)}. \tag{2.1}$$

This formula is called the *asymptotic expansion of the global error*, since

$$\epsilon_\Delta(t) = \sum_{n=0}^{N-1} e_{p+n}(t)\tau^{p+n} + r_\Delta^N(t)\tau^{p+N} \quad t \in \Delta.$$

For now we can interpret the first part on the right side of Equation 2.1 as a polynomial P in τ and the second part as a non-polynomial remainder $R(\tau)$.
Given the values $(X(\tau_n))_{n=0:N} = (x_{\Delta_n}(T))_{n=0:N}$ it is natural to interpolate the function X in the nodes $(\tau_n)_{n=0:N}$ to obtain the interpolant π and use $\pi(0)$ as a new approximation of $x(T)$

Remark 2.1. Evaluating the interpolant of the values $(x_i)_{i=0:n}$ corresponding to the nodes $t_0 < t_1 < \ldots < t_n$ outside (Latin: *extra*) the interval $[t_0, t_n]$ is denominated extrapolation while interpolation implies an evaluation inside (Latin: *inter*) the interval. Nonetheless we will refer to the task of *computing* π as the *interpolation problem*.

2.2 A special Interpolation Problem

In Section 2.1 we formulated a non–standard interpolation problem. Rather than computing

$$\pi \in \mathbb{P}_{N-1}^d := \left\{ \tau \mapsto \sum_{n=0}^{N-1} a_n \tau^n : \ a_n \in \mathbb{R}^d \right\},$$

we look for

$$\pi \in \left\{ \tau \mapsto a_0 + \sum_{n=0}^{N-1} a_{p+n} \tau^{p+n} : \ a_n \in \mathbb{R}^d \right\}$$

such that $\pi(\tau_n) = x_{\Delta_n}(T)$ for $n \in [0:N]$. So before we can proceed we make sure that this special interpolation problem (and its generalisation) is well–posed.

Lemma 2.2. *Let integers $p, \omega \geq 1$ and $N \geq 0$, nodes $\Gamma = (\tau_n)_{n=0:N} \subset \mathbb{R}^+$ and values $(\tau_n)_{n=0:N} \subset \mathbb{R}^d$ be given. Then there is a unique polynomial*

$$\pi \in \Pi_N := \left\{ \tau \mapsto a_0 + \sum_{n=0}^{N-1} a_{p+n} \tau^{p+n\omega} : \ a_n \in \mathbb{R}^d \right\}$$

such that $\pi(\tau_n) = x_n$ for $n \in [0:N]$.

Proof (cf. [3, Lemma 4.33]). Since vector valued interpolation is executed componentwise it suffices to consider $d = 1$. Furthermore let N be positive since the case of $N = 0$ is trivially true. The claim is equivalent to the following statement:

$$\exists!\ \tilde{\pi} \in \mathbb{P}_{N-1}, a_0 \in \mathbb{R} : \ a_0 + \tau_n^p \, \tilde{\pi}(\tau_n^\omega) = x_n \quad n \in [0:N]. \tag{2.2}$$

This in turn can be rephrased by introducing the linear map

$$P : \mathbb{P}_{N-1} \to \mathbb{R}^{N+1}, \tilde{\pi} \mapsto (\tau_n^p \, \tilde{\pi}(\tau_n))_{n=0:N}.$$

Equation 2.2 is then equivalent to $\operatorname{span}_{\mathbb{R}}(\mathbb{1}_{N+1}) \oplus \operatorname{im} P = \mathbb{R}^{N+1}$, which we will actually prove.
First notice that P is injective. For $\tilde{\pi} \in \ker P$ we have $\tau_n^p \, \tilde{\pi}(\tau_n^\omega) = 0$ for all n. Since all τ_n are nonzero, we can divide each component by τ_n^p and get $\tilde{\pi}(\tau_n^\omega) = 0$ for $n = [0:N]$. This already shows $\tilde{\pi} \equiv 0$ as $\tilde{\pi}$ has at most a degree of $N-1$ but $N+1$ distinct roots.
That entails $\operatorname{im} P = \mathbb{R}^N$ and in order to prove $\mathbb{1}_{N+1} \oplus \operatorname{im} P = \mathbb{R}^{N+1}$ we will now show that $\mathbb{1}_{N+1} \notin \operatorname{im} P$.
Assume the opposite, i.e. that there is a $\tilde{\pi} \in \mathbb{P}_{N-1}$ such that $\tau_n^p \, \tilde{\pi}(\tau_n^\omega) = 1$ for all n. Again we use that all nodes are positive and get

$$\tilde{\pi}(\tau_n) = \frac{1}{\tau_n^{p/\omega}} \quad n \in [0:N].$$

But this means $\tilde{\pi}$ is the unique interpolant of $f : \tau \mapsto \tau^{-p/\omega} \in \mathcal{C}^\infty(\mathbb{R}^+, \mathbb{R})$ in the nodes $\tau_0, \ldots, \tau_{N-1}$. Thus we can make use of a well known representation of the interpolation error (e.g. stated in [4, Theorem 7.16]): For $\tau > 0$ there is a $\sigma > 0$ such that

$$f(\tau) - \tilde{\pi}(\tau) = \frac{f^{(N)}(\sigma)}{N!} \prod_{n=0}^{N-1} (\tau - \tau_k).$$

Since no derivative of f has any roots in $\mathbb{R}^+$, plugging τ_N^ω into the last equality leads to the sought contradiction and concludes the proof. □

Thus our interpolation problem is indeed well posed:

Theorem 2.3 (Interpolation and extrapolation operators)**.** *In the setting of Lemma 2.2 we define the* interpolation operator

$$\mathcal{P}_\Gamma : \mathbb{R}^{d\times(N+1)} \to \Pi_N,\ [x_n]_{n=0:N} \mapsto \pi$$

with $\pi(\tau_n) = x_n$ *for* $n \in [0:N]$ *and the* extrapolation operator

$$\mathcal{E}_\Gamma : \mathbb{R}^{d\times(N+1)} \to \mathbb{R}^d,\ [x_n]_{n=0:N} \mapsto (\mathcal{P}_\Gamma[x_n]_{n=0:N})\,(0).$$

Both are linear and continuous. Furthermore $\mathcal{E}_\Gamma[x_n]_{n=0:N}$ *is invariant under (nontrivial) scaling of* Γ*.*

Proof. In Lemma 2.2 we proved that $\mathcal{P}_\Gamma$ is well–defined. It is linear by construction and since it is a map between finite dimensional vector spaces it is also continuous, i.e. we have assured well posedness.
Since the evaluation of continuous functions $E : f \mapsto f(0)$ itself is linear and continuous all properties can be immediately transferred to $\mathcal{E}_\Gamma = E \circ \mathcal{P}_\Gamma$.
Now let $\lambda \neq 0$ be an arbitrary real number and rescales Γ to $\widetilde{\Gamma} := (\lambda\tau_n)_{n=0:N}$. Note that $\mathcal{P}_{\widetilde{\Gamma}}[x_n]_{n=0:N} = \pi(\lambda^{-1}\tau)$ if $\pi \in \Pi_N$ is the interpolant of $(x_n)_{n=0:N}$ in $(\tau_n)_{n=0:N}$. Thus

$$\mathcal{E}_{\widetilde{\Gamma}}[x_n]_{n=0:N} = \pi(\lambda^{-1}0) = \mathcal{E}_\Gamma[x_n]_{n=0:N}.$$

□

2.3 Formal Definition of Extrapolation Methods

Now that we have convinced ourselves that we actually *can* interpolate the asymptotic expansion stated in Equation 2.1, let us cast the idea outlined before in a pseudocode.
In Section 2.1 we described the modus operandi for the interval $[t_0, T]$. The extrapolation method we define below executes those computations for each time step, i.e. the interval becomes $[t, t+\tau]$.

Definition 2.4 (Extrapolation methods)**.** Let Ψ be a consistent one step method of order $p \geq 1$ which has for smooth f and equidistant grids Δ the asymptotic expansion

$$x_\Delta(T) = x(T) - \sum_{n=0}^{N-1} e_{p+n\omega}(T)\tau_\Delta^{p+n\omega} - r_\Delta^N(T)\tau_\Delta^{p+N\omega} \quad \omega \in \{1,2\}.$$

Furthermore choose a *subdividing sequence* $(\nu_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{N}$ that is strictly increasing and an *order of extrapolation* $N \in \mathbb{N}_0$. Then Ψ_N given by Algorithm 1 is the *N–th extrapolation of* Ψ.

input : initial data (t, x), step length τ
output: $\Psi_N^{t+\tau,t} x$

for $n = 0 : N$ **do**
  $\tau_n := \frac{\tau}{\nu_n}$
  $\Delta_n := \left(t + \nu\tau_n \middle| \nu = 0 : \nu_n\right)$
  use Ψ to compute x_{Δ_n} starting in $x_{\Delta_n}(t) := x$
end
$\Gamma := (\tau_n)_{n=0:N}$
$\Psi_N^{t+\tau,t} x := \mathcal{E}_\Gamma[x_{\Delta_n}(t+\tau)]_{n=0:N}$

Algorithm 1: The Nth extrapolation of Ψ

Now we show that the method Ψ_N is indeed a discretization of higher consistency order than Ψ.

Theorem 2.5. *Ψ_N is a one step method. It is consistent of order $p + \omega N$ if the right side f and the increment function of Ψ are smooth.*

In preparation of proving Theorem 2.5 we provide a lemma.

Lemma 2.6. *Let $(\Psi_N)_{n=0:N}$ be one step methods with increment functions $(\psi_n)_{n=0:N} \subset \mathcal{C}^k(\Sigma, \mathbb{R}^d)$ for some $k, N \in \mathbb{N}$. Then also*

$$\begin{aligned} \Psi_c^{t+\tau,t} x &:= \Psi_1^{t+\tau,t+\lambda\tau} \Psi_0^{t+\lambda\tau,t} x \text{ with } \lambda \in [0,1] \text{ and} \\ \Psi_l^{t+\tau,t} x &:= \sum_{n=0}^{N} \lambda_n \Psi_n^{t+\tau,t} x \text{ with } (\lambda_n)_{n=0:N} \subset \mathbb{R} \end{aligned}$$

are one step methods with k–times continuously differentiable increment functions ψ_c and ψ_l.

Proof. We write down the respective increment functions for some $(t, x, \tau) \in \Sigma$:

$$\begin{aligned} \psi_c(t,x,\tau) &= \lambda\psi_0(t,x,\lambda\tau) + (1-\lambda)\psi_1(t+\lambda\tau, x+\lambda\tau\psi_0(t,x,\lambda\tau), (1-\lambda)\tau), \\ \psi_l(t,x,\tau) &= \sum_{n=0}^{N} \lambda_n \psi_n(t,x,\tau). \end{aligned}$$

Since $(t+\lambda\tau, x+\lambda\tau\psi_0(t,x,\lambda\tau)) \in \Omega$, the function ψ_c is a well defined element of $\mathcal{C}^k(\Sigma, \mathbb{R}^d)$. Obviously this is also true for ψ_l. Note that, while $\Psi_c^{t+\tau,t} x \in \Omega_0$ for any $(t, x, \tau) \in \Sigma$, we may need to restrict the domain of ψ_l in order to guarantee $\Psi_l^{t+\tau,t} x \in \Omega_0$ as well. For instance by demanding for $(t, x) \in \Omega$:

$$|\tau| < \tau_l^*(t,x) := \frac{1}{2} \min_{|\tau| \le \frac{\tau^*(t,x)}{2}} \left(\tau^*(t,x), \frac{\operatorname{dist}(x, \partial\Omega_0)}{\|\psi_l(t,x,\tau)\|} \right).$$

□

Now we can present the

Proof of Theorem 2.5 . We show that Ψ_N is a consistent one step method. Let Σ denote the domain of the increment function of Ψ and fix a $(t, x, \tau) \in \Sigma$.
The recursive application of Lemma 2.6 shows that $\Psi_{\Delta_n}^{t+\tau,t} x := x_{\Delta_n}(t+\tau)$ induces a one step method for each $n \in [0:N]$. Since $\mathcal{E}_\Gamma$ is linear there are $(\lambda_n)_{n=0:N} \subset \mathbb{R}$ such that

$$\Psi_N^{t+\tau,t} x = \mathcal{E}_\Gamma[x_{\Delta_n}(t+\tau)]_{n=0:N} = \sum_{n=0}^{N} \lambda_n \Psi_{\Delta_n}^{t+\tau,t} x.$$

Thus we make again use of Lemma 2.6 and see Ψ_N is indeed a one step method. Without loss of generality we can assume that its increment function is well defined on Σ. Furthermore the increment function ψ_N of Ψ_N inherits the regularity of ψ.
Now we show that Ψ_N has consistency order $p + N\omega$. Since the right side f and the increment function ψ_N are smooth, we expand its local error $\epsilon(t, x, \tau)$:

$$\epsilon(t, x, \tau) = \sum_{n=1}^{p+N\omega} \epsilon_n(t)\tau^n + \rho(t, \tau)\tau^{p+N\omega+1}.$$

On the other hand we can estimate the local error by making use of Theorem 2.7 ($\omega = 1$) and Theorem 2.12 ($\omega = 2$). If the right side f and the increment function ψ_N are smooth, those affirm the existence of some constants $C_N, \tau_N > 0$ such that for $n \in [0:N]$ and $|\tau| \leq \hat{\tau}_N$

$$x_{\Delta_n}(t+\tau) = x(t+\tau) - \sum_{k=0}^{N-1} e_{p+k\omega}(t+\tau)\tau_n^{p+k\omega} - r_{\Delta_n}^N(t+\tau)\tau^{p+N\omega} \text{ and } \|r_{\Delta_n}^N(t+\tau)\| \leq C_N|\tau|.$$

Now we use the auxiliary notation of Equation 2.1 and estimate:

$$\begin{aligned}
\|\epsilon(t, x, \tau)\| &= \|x(t+\tau) - \Psi_N^{t+\tau,t} x\| = \|P(0) - \mathcal{E}_\Gamma[\Psi_{\Delta_n}^{t+\tau,t} x]_{n=0:N}\| \\
&= \|\mathcal{E}_\Gamma[P(\tau_n)]_{n=0:N} - \mathcal{E}_\Gamma[P(\tau_n) - R(\tau_n)]_{n=0:N}\| = \|\mathcal{E}_\Gamma[R(\tau_n)]_{n=0:N}\| \\
&\leq \sum_{n=0}^{N} |\lambda_n| \|r_{\Delta_n}^N(t+\tau)\nu_n^{-(p+N\omega)}\| |\tau|^{p+N\omega} \leq \sum_{n=0}^{N} \frac{|\lambda_n| C_N}{\nu_n^{p+N\omega}} |\tau|^{p+N\omega+1}.
\end{aligned}$$

Finally Corollary 1.9 yields that Ψ_N is indeed consistent of order $p + N\omega$. □

2.4 Asymptotic Expansion of the Global Error

In this section we proof the theoretical foundation of the ideas outlined in the previous one. We will show that the global error of any consistent one step method has an asymptotic expansion. Furthermore there are certain methods that allow for an expansion in even powers of τ. Without loss of generality we assume $t_0 < T$.

Theorem 2.7 (Asymptotic Expansion, $\omega = 1$)**.** *Let x be the solution of Equation 1.1 restricted to the interval $[t_0, T] \subset I_{\max}$. We discretize x with a one step method Ψ that is consistent of order p along the graph of x. If we furthermore assume that f and ψ*

are smooth, the global error has an asymptotic expansion. There is a unique sequence of coefficient functions $(e_{p+n})_{n\in\mathbb{N}_0} \subset \mathcal{C}^\infty(I_{\max}, \mathbb{R}^d)$ *that satisfies:*

$$\begin{aligned}&\forall\, N \in \mathbb{N} \quad \exists\, \hat{\tau}_N, c_N > 0 \quad \forall\, t \in \Delta := (t_0 + k\tau :\ k \in [0:K]) \subset [t_0, T] \text{ with } 0 < \tau \leq \hat{\tau}_N : \\ &x_\Delta(t) = x(t) - \sum_{n=0}^{N-1} e_{p+n}(t)\tau^{p+n} - r_\Delta^N(t)\tau^{p+N} \text{ and } \|r_\Delta^N(t)\| \leq c_N(t - t_0). \end{aligned} \tag{2.3}$$

Again we first provide a helpful lemma.

Lemma 2.8. *Let* Ψ *be a consistent discretization and* $f \in \mathcal{C}^2(\Omega, \mathbb{R}^d)$ *the right side in Equation 1.1. Furthermore let* $\delta \in \mathcal{C}(D_\delta, \mathbb{R}^d)$ *with an open domain* $D_\delta \subseteq \mathbb{R}^2$ *and* $n \in \mathbb{N}$ *be given, such that* $(t, x + \delta(t,\tau)\tau^n, \tau) \in \Sigma$ *if* $(t, x, \tau) \in \Sigma$ *in the first place. Then there is a remainder* $\rho \in \mathcal{C}(\Sigma, \mathbb{R}^d)$ *such that*

$$\psi(t, x + \delta(t,\tau)\tau^n, \tau) = \psi(t, x, \tau) + \partial_x f(t, x)\delta(t,\tau)\tau^n + \rho(t, x, \tau)\tau^{n+1}$$

on Σ.

(cf. [3, Lemma 4.38]). Let $(t, x, \tau) \in \Sigma$ be arbitrary. We note that, since Ψ is consistent, its increment function ψ is twice continuously differentiable. Thus we can successively expand it in x and τ:

$$\begin{aligned}\psi(t, x + \delta(t,\tau)\tau^n, \tau) \;&=\; \psi(t, x, \tau) + \partial_x\psi(t, x, \tau)\delta(t,\tau)\tau^n + \\ &\quad \tau^{n+1} \underbrace{\int_0^1 [\partial_x^2\psi(t, x + \lambda\delta(t,\tau)\tau^n, \tau)](\delta(t,\tau), \delta(t,\tau))(1-\lambda)\tau^{n-1}d\lambda}_{\rho_1(t,x,\tau)\in\mathcal{C}(\Sigma,\mathbb{R}^d)}\end{aligned}$$

and

$$\partial_x\psi(t, x, \tau) = \partial_x\psi(t, x, 0) + \tau \underbrace{\int_0^1 [\partial_\tau\partial_x\psi(t, x, \lambda\tau)]d\lambda}_{:=\rho_2(t,x,\tau)\in\mathcal{C}(\Sigma,\mathbb{R}^d)}.$$

If we combine both expansions and set $\rho := \rho_1 + \delta\rho_2$ we get:

$$\psi(t, x + \delta(t,\tau)\tau^n, \tau) = \psi(t, x, \tau) + \partial_x\psi(t, x, \tau)\delta(t,\tau)\tau^n + \rho(t, x, \tau)\tau^{n+1}.$$

The continuity of ρ_1 and ρ_2 is again obtained by [5, Theorem 5.6].
Secondly we show $\partial_x f(t, x) = \partial_x\psi(t, x, 0)$. Define the auxiliary function $h : \tau \mapsto f(t, x) - \psi(t, x, \tau) \in \mathcal{C}^2([-\hat{\tau}, \hat{\tau}], \mathbb{R}^d)$ for some suitable $\hat{\tau} > 0$. Since Ψ is consistent $h(0) = 0$ (cf. Theorem 1.8), we have:

$$f(t, x) - \psi(t, x, \tau) = h(\tau) - h(0) = \int_0^\tau \partial_\tau\psi(t, x, \lambda)d\lambda.$$

Now we can differentiate both sides with respect to x and swap the order of differentiation and integration (the latter by virtue of [5, Theorem 5.7]):

$$\partial_x f(t, x) - \partial_x\psi(t, x, \tau) = \int_0^\tau \partial_{x,\tau}\psi(t, x, \lambda)d\lambda.$$

Since both sides of the last equality are continuous on $[-\hat{\tau}, \hat{\tau}]$, the right side of the vanishes for $\tau \to 0$. □

Now we can develop the Proof of Theorem 2.7. The version we are about to present is due to Hairer and Lubich from 1984 (cf. [6, Theorem 1]).

Proof. In order to prove existence and uniqueness of the functions e_{p+n} we pursue the following strategy: Let us assume those functions exist, derive some of their properties and hopefully find a more manageable characterization.
If we set

$$x_\Delta^0 := x_\Delta \quad \text{and} \quad x_\Delta^{n+1} := x_\Delta^n + e_{p+n}\big|_\Delta \tau^{p+n} \quad n \in \mathbb{N}_0$$

and denote the respective discretizations which produced x_Δ^n with Ψ_n. Equation 2.3 implies

$$x\big|_\Delta - x_\Delta^n = \mathcal{O}(\tau^{p+n}) \quad (\tau \to 0).$$

That means x_Δ^n *converges of order* $p+n$. Thus we regard the asymptotic expansion as a grid function and conjecture that the discretization Ψ_n, which would produce x_Δ^n, is consistent of order $p+n+1$.
In order to prove this we first derive a recursion for the increment functions ψ_n. For $t \in \Delta, t \neq T$ and $n \in \mathbb{N}_0$ we compute:

$$\begin{aligned}
\Psi_{n+1}^{t+\tau,t} x_\Delta^{n+1}(t) &= \Psi_n^{t+\tau,t} x_\Delta^n(t) + e_{p+n}(t+\tau)\tau^{p+n} \\
&= \Psi_n^{t+\tau,t}[x_\Delta^{n+1}(t) - e_{p+n}(t)\tau^{p+n}] + e_{p+n}(t+\tau)\tau^{p+n} \\
&= [x_\Delta^{n+1}(t) - e_{p+n}(t)\tau^{p+n}] + \tau\psi_n(t, [x_\Delta^{n+1}(t) - e_{p+n}(t)\tau^{p+n}], \tau) + e_{p+n}(t+\tau)\tau^{p+n} \\
&= x_\Delta^{n+1}(t) + \tau\left[\psi_n(t, x_\Delta^{n+1}(t) - e_{p+n}(t)\tau^{p+n}, \tau) + (e_{p+n}(t+\tau) - e_{p+n}(t))\,\tau^{p+n-1}\right].
\end{aligned}$$

This leads to the recursion

$$\begin{aligned}
\psi_0(t,x,\tau) &:= \psi(t,x,\tau), \\
\psi_{n+1}(t,x,\tau) &:= \psi_n(t, x - e_{p+n}(t)\tau^{p+n}, \tau) + (e_{p+n}(t+\tau) - e_{p+n}(t))\,\tau^{p+n-1} \quad n \in \mathbb{N}_0.
\end{aligned}$$

We denote the domain of ψ_n with Σ_n which can be obtained by appropriate scaling of Σ_{n-1} in τ–direction. We show now inductively that there exists exactly one $e_{p+n} \in \mathcal{C}^\infty(I_{\max}, \mathbb{R}^d)$ such that Ψ_{n+1} is consistent of order $p+n+1$ along the graph of x.
Let $n \geq 0$ and choose $(t,x,\tau) \in \Sigma_{n+1}$ such that $t \in I_{\max}$ and $x = \Phi^{t,t_0}x_0$. Then we have:

$$\begin{aligned}
& \Phi^{t+\tau,t}x - \Psi_{n+1}^{t+\tau,t}x \\
= \;& \Phi^{t+\tau,t}x - x - \tau\left[\psi_n(t, x - e_{p+n}(t)\tau^{p+n}, \tau) + (e_{p+n}(t+\tau) - e_{p+n}(t))\,\tau^{p+n-1}\right] \\
\overset{(a)}{=} \;& \Phi^{t+\tau,t}x - x - \tau\left[\psi_n(t,x,\tau) - \partial_x f(t,x) e_{p+n}(t)\tau^{p+n} + \rho(t,x,\tau)\tau^{p+n+1}\right] \\
& - [e_{p+n}(t+\tau) - e_{p+n}(t)]\,\tau^{p+n} \\
\overset{(b)}{=} \;& \epsilon_{p+n+1}^n(t)\tau^{p+n+1} + \rho^n(t,\tau)\tau^{p+n+2} + \partial_x f(t,x)e_{p+n}(t)\tau^{p+n+1} - \rho(t,x,\tau)\tau^{p+n+2} \\
& - \left[e_{p+n}(t) + e'_{p+n}(t)\tau + \rho_e(t,\tau)\tau^2 - e_{p+n}(t)\right]\tau^{p+n} \\
= \;& \underbrace{\left[\epsilon_{p+n+1}^n(t) + \partial_x f(t,x(t))e_{p+n}(t) - e'_{p+n}(t)\right]}_{=\epsilon_{p+n+1}^{n+1}(t)}\tau^{p+n+1} + \underbrace{\left[\rho^n(t,\tau) - \rho(t,x(t),\tau) - \rho_e(t,\tau)\right]}_{=\rho^{n+1}(t,\tau)}\tau^{p+n+2}.
\end{aligned}$$

In the equality marked with (a) we applied Lemma 2.8. In (b) we applied the induction hypothesis to make use of Theorem 1.8 – the superscripts of the coefficient functions

denote the discretization they belong to – and secondly expanded e_{p+n} in t:

$$e_{p+n}(t+\tau) = e_{p+n}(t) + e'_{p+n}(t)\tau + \tau^2 \underbrace{\int_0^1 e^{(2)}_{p+n}(t+\lambda\tau)(1-\lambda)d\lambda}_{:=\rho_e(t,\tau)}.$$

But now we have indeed derived a second characterization for e_{p+n}. The local error of Ψ_{n+1} is pointwise of order $p+n+2$ if and only if ϵ^{n+1}_{p+n+1} vanishes. The latter is equivalent to e_{p+n} being the solution of the following initial value problem:

$$e'_{p+n}(t) = \partial_x f(t, x(t))e_{p+n}(t) + \epsilon^n_{p+n+1}(t), \quad e_{p+n}(t_0) = 0. \tag{2.4}$$

The initial value is given by plugging t_0 into the definition of x^{n+1}_Δ:

$$x_0 = x^n_\Delta(t_0) + e_{p+n}(t_0)\tau^{p+n} = x_0 + e_{p+n}(t_0)\tau^{p+n}.$$

Since the right side of the linear inhomogeneous IVP in Equation 2.4 is smooth on $I_{\max}$, $e_{p+n} \in \mathcal{C}^\infty(I_{\max}, \mathbb{R}^d)$ does indeed exist and is unique (cf. [3, Example 2.11]).
By virtue of Theorem 1.8 the pointwise estimate of the local error we have derived above suffices to show that Ψ_{n+1} has the claimed consistency order along the graph of x. This completes the induction.
Now we prove the existence and the properties of the remainder. Fix a $N \in \mathbb{N}_0$. Since Ψ_N fulfills the prerequisites of Theorem 1.10, we know there is a $\hat{\tau}_N > 0$ such that for any equidistant grid $\Delta = (t_0, \dots, x_{N_\Delta}) \subset [t_0, T]$ with stepsize $0 < \tau \leq \hat{\tau}_N$ the global error $\epsilon^N_\Delta := x\big|_\Delta - x^N_\Delta$ is well defined. Thus we set $r^N_\Delta := \frac{\epsilon^N_\Delta}{\tau^{p+N}}$ and immediately get for any $t \in \Delta$:

$$\begin{aligned} \|r^N_\Delta(t)\| &\leq c^N_1 \left(e^{c^N_2(t-t_0)} - 1\right) = c^N_1 \int_0^{t-t_0} c^N_2 e^{x^N_2 \lambda} d\lambda = (t-t_0) \int_0^1 c^N_1 c^N_2 e^{c^N_2 \lambda(t-t_0)} d\lambda \\ &\leq c_N(t-t_0) \quad \text{for some } c_N > 0. \end{aligned}$$

In the last inequality we used that the integral the left side is continuous in t on the compact set $[t_0, T]$. □

If we change the prerequisites of Theorem 2.7 and assume that f and ψ are not smooth, one could ask how many functions $e_p, e_{p+1}, \dots$ can be constructed using the reasoning of the presented proof.

Corollary 2.9. *In the setup of Theorem 2.7 assume that f and ψ are only q times continuously differentiable for some integer $p \leq q$. Then the induction performed in the proof of Theorem 2.7 guarantees the existence of*

$$e_p, \dots, e_{p+N} \text{ with } 0 \leq N \leq \frac{1}{2}\left(-2p - 1\sqrt{4p^2 + 8q + 1 - 12p}\right).$$

Proof. We denote the regularity of ψ_n by q_n, i.e. $\psi_n \in \mathcal{C}^{q_n}(\Sigma_n, \mathbb{R}^d)$. Thus $e_{p+n} \in \mathcal{C}^{q_{n+1}}(I_{\max}, \mathbb{R}^d)$. We proof by induction

$$q_n = q - (p-1)n - \sum_{\nu=0}^{n-1} \nu.$$

The base case, $q_0 = q$, is confirmed by $\psi_0 = \psi$. Now let $n \geq 0$. Since e_{p+n} solves Equation 2.4, we have $q_{n+1} = \min(q-1, r_n)+1$ with $\epsilon^n_{p+n+1} \in \mathcal{C}^{r_n}(I_{\max}, \mathbb{R}^d)$. We compute:

$$\begin{aligned} r_n &\overset{\text{Theorem 1.8}}{=} \min(q, q_n) + 1 - (p+n+1) \overset{\text{ind. hyp}}{=} q_n - (p-1) - n - 1 \\ &= \quad q - (p-1)(n+1) - \sum_{\nu=0}^{n} \nu - 1 < q - 1. \end{aligned}$$

And thus $q_{n+1} = \min(q-1, r_n) + 1 = r_n + 1 = q - (p-1)(n+1) - \sum_{\nu=0}^{n} \nu$.
For the construction of e_{p+n} we assumed $e_{p+n} \in \mathcal{C}^2(I_{\max}, \mathbb{R}^d)$ and in order to show that Ψ_{n+1} is consistent of order $p+n+1$ we made use of Theorem 1.8. Therefore we have following restriction on n: $\min(q, q_{n+1}) = q_{n+1} \geq p+n+1 \geq 2$, i.e $q_n \geq p+n$. This implies

$$0 \leq n \leq \frac{1}{2}\left(-2p - 1\sqrt{4p^2 + 8q + 1 - 12p}\right).$$

□

We will see later on that this barrier does not pose a problem if we extrapolate Runge–Kutta methods (cf. Theorem 3.6).
But first we address the case of $\omega = 2$. We will elucidate that there are certain discretizations whose asymptotic expansion contains only coefficient function e_{p+n} with even index n. Equation 2.3 then becomes:

$$x_\Delta(t) = x(t) - \sum_{n=0}^{N-1} e_{p+2n}(t)\tau^{p+2n} - r^N_\Delta(t)\tau^{p+2N}. \tag{2.5}$$

Let us characterize the class of methods that produce expansions with $\omega = 2$. We let $0 < \tau \leq \hat{\tau}_N$ and set $E(t,\tau) := \epsilon_\Delta(t) = \sum_{n=0}^{N-1} e_{p+n}(t)\tau^{p+n} + r^N_\Delta(t)\tau^{p+N}$. Then the case $\omega = 2$ is equivalent to

$$E(t,\tau) = (-1)^p E(t,-\tau) \quad \forall\, t \in \Delta = (t_0 + k\tau :\ k \in [0:K]).$$

The grid function $E(t,-\tau)$ can be interpreted as the global error $\hat{\epsilon}_\Delta$ of the approximation $\hat{x}_\Delta$ that is produced by Ψ in *reversed* order, i.e.

$$\begin{aligned} \hat{x}_\Delta(t_0 + K\tau) &:= x_\Delta(t_0 + K\tau) \\ \hat{x}_\Delta(t_0 + (K-k)\tau) &:= \Psi^{t_0+(K-k)\tau, t_0+(K-k+1)\tau}\hat{x}_\Delta(t_0 + (K-k+1)\tau) \quad k \in [1:K]. \end{aligned}$$

Then $\epsilon_\Delta = (-1)^p\hat{\epsilon}_\Delta$ is equivalent to $x_\Delta = (-1)^p\hat{x}_\Delta$. We fix a $t \in \Delta\backslash\{t_0, t_0 + K\tau\}$ and derive a conditional equation for the increment function ψ:

$$\begin{aligned} x_\Delta(t+\tau) &= x_\Delta(t) + \tau\psi(t, x_\Delta(t), \tau) \\ &= (-1)^p\hat{x}_\Delta(t) + \tau\psi(t, x_\Delta(t), \tau) \\ &= (-1)^p[\hat{x}_\Delta(t+\tau) - \tau\psi(t+\tau, \hat{x}_\Delta(t+\tau), -\tau)] + \tau\psi(t, x_\Delta(t), \tau) \\ &= x_\Delta(t+\tau) - (-1)^p\tau\psi(t+\tau, (-1)^p x_\Delta(t+\tau), -\tau) + \tau\psi(t, x_\Delta(t), \tau) \\ \Leftrightarrow \psi(t, x_\Delta(t), \tau) &= (-1)^p\psi(t+\tau, (-1)^p x_\Delta(t+\tau), -\tau) \\ \Leftrightarrow \quad \psi(t, x, \tau) &= (-1)^p\psi(t+\tau, (-1)^p\Psi^{t+\tau,t}x, -\tau) \quad \forall\,(t,x,\tau) \in \Sigma. \end{aligned}$$

If we additionally demand that Ψ is consistent of order $p > 0$, that is by Theorem 1.8 $\psi(t, x, 0) = f(t, x)$ on Ω, we furthermore see that p must be even. An odd p leads to the contradiction

$$f(t, x) = -f(t, -x) \quad \forall\, f \in \mathcal{C}^\infty(\Omega, \mathbb{R}^d).$$

Thus a necessary condition for a an asymptotic expansion of the global error in even powers of τ is

$$\psi(t, x, \tau) = \psi(t + \tau, \Psi^{t+\tau,t}x, -\tau) \quad \Leftrightarrow \quad \Psi^{t,t+\tau}\Psi^{t+\tau,t}x = x \quad \forall\, (t, x, \tau) \in \Sigma. \tag{2.6}$$

Definition 2.10 (Reversible one step method)**.** A one step method Ψ is reversible if it satisfies Equation 2.6.

We will now show that reversibility is also a sufficient criterion for an even asymptotic expansion of ϵ_Δ. We start with a lemma about the consistency order of reversible methods.

Lemma 2.11. *Assume that the discrete evolution Ψ is consistent of order p, additionally let f and ψ be $p+1$ times continuously differentiable on their respective domains. If Ψ is reversible, p is even or Ψ has already consistency order $p + 1$.*

Proof. Let x denote the solution of Equation 1.1 and let $|\tau| < \tau^*(t_0, x_0)$. Using the reversibility and Theorem 1.8 we obtain

$$\begin{aligned} x_0 &= \Psi^{t_0,t_0+\tau}\Psi^{t_0+\tau,t_0}x_0 \\ x_0 &= \Psi^{t_0,t_0+\tau}[x(t_0 + \tau) - \epsilon_{p+1}(t_0)\tau^{p+1} - \rho(t_0, \tau)\tau^{p+2}] \\ &= x(t_0 + \tau) - \epsilon_{p+1}(t_0)\tau^{p+1} - \rho(t_0, \tau)\tau^{p+2} \\ &\quad -\tau\psi(t_0 + \tau, x(t_0 + \tau) + (-\epsilon_{p+1}(t_0) - \rho(t_0, \tau)\tau)\tau^{p+1}, -\tau). \end{aligned} \tag{2.7}$$

Now we set $\delta(\tau) := -\epsilon_{p+1}(t_0) - \rho(t_0, \tau)\tau$ and expand ψ:

$$\begin{aligned} \psi(t_0 + \tau, x(t_0 + \tau) + \delta(\tau)\tau^{p+1}, -\tau) = \psi(t_0 + \tau, x(t_0 + \tau), -\tau) \\ +\tau^{p+1} \underbrace{\int_0^1 \partial_x\psi(t_0 + \tau, x(t_0 + \tau) + \lambda\delta(\tau)\tau^{p+1}, -\tau)\delta(\tau)d\lambda}_{=:\rho_1(\tau)}. \end{aligned}$$

We carry on, Equation 2.7 becomes:

$$\begin{aligned} -\epsilon_{p+1}(t_0)\tau^{p+1} - [\rho(t_0, \tau) + \rho_1(\tau)]\tau^{p+2} &= x_0 - x(t_0 + \tau) + \tau\psi(t_0 + \tau, x(t_0 + \tau), -\tau) \\ &= x(t_0) - \Psi^{t_0,t_0+\tau}x(t_0 + \tau) \\ &= \epsilon_{p+1}(t_0 + \tau)(-\tau)^{p+1} + \rho(t_0 + \tau, -\tau)(-\tau)^{p+2}. \end{aligned}$$

Since ϵ_{p+1} is continuously differentiable yet another Taylor expansion yields

$$\epsilon_{p+1}(t_0 + \tau) = \epsilon_{p+1}(t_0) + \tau \underbrace{\int_0^1 \epsilon'_{p+1}(t_0 + \lambda\tau)d\lambda}_{=:\rho_2(\tau)}.$$

We plug this into the preceding equality and obtain:

$$[(-1)^{p+1} + 1]\epsilon_{p+1}(t_0)\tau^{p+1} = [\rho_2(\tau) - \rho(t_0 + \tau, -\tau)](-\tau)^{p+2} - [\rho(t_0, \tau) + \rho_1(\tau)]\tau^{p+2}.$$

Since the right side of the last equality is continuous in τ for $|\tau| < \tau^*(t_0, x_0)$ we can equate the coefficients and see that either p is even or $\epsilon_{p+1}(t_0) = 0$. As we the initial data (t_0, x_0) was arbitrary, Theorem 1.8 implies that either p is even or Ψ is already consistent of order $p + 1$. □

Now we can prove

Theorem 2.12 (Asymptotic Expansion, $\omega = 2$)**.** *Assume that additional to the prerequisites of Theorem 2.7 Ψ is reversible along the graph of $x\big|_{[t_0,T]}$ and its consistency order p is maximal in the sense that the coefficient function ϵ_{p+1} in Equation 1.3 does not vanish. Then*

$$e_{p+n} \equiv 0 \quad \forall\, n \in \mathbb{N}_0, 2 \nmid n$$

and the asymptotic expansion Equation 2.3 transforms for $t \in \Delta$ to

$$x_\Delta(t) = x(t) - \sum_{n=0}^{N-1} e_{p+2n}(t)\tau^{p+n} - r_\Delta^N(t)\tau^{p+2N} \text{ and } \|r_\Delta^N(t)\| \le c_N(t - t_0). \tag{2.8}$$

Proof. We use the objects and notation introduced in the proof of Theorem 2.7.
Since e_{p+n} is the solution of Equation 2.4, that coefficient function vanishes if and only if the inhomogeneous term ϵ_{p+n+1}^n does. Thus it suffices to show by induction that all Ψ_n are reversible and ϵ_{p+n+1}^n vanishes for odd n.
Let $n \in \mathbb{N}_0$ and $(t, x, \tau) = (t, \Phi^{t,t_0}x_0, \tau) \in \Sigma_{n+1}$ with $t \in I_{\max}$ be arbitrary. Using the definition and the induction hypothesis we compute

$$\begin{aligned}
\psi_{n+1}(t, x(t), \tau) &= \psi_n(t, x(t) - e_{p+n}(t)\tau^{p+n}, \tau) + [e_{p+n}(t+\tau) - e_{p+n}(t)]\tau^{p+n-1} \\
&= \psi_n(t+\tau, \Psi_n^{t+\tau,t}[x(t) - e_{p+n}(t)\tau^{p+n}], -\tau) + [e_{p+n}(t+\tau) - e_{p+n}(t)]\tau^{p+n-1} \\
&= \psi_n(t+\tau, \Psi_{n+1}^{t+\tau,t}x(t) - e_{p+n}(t+\tau)\tau^{p+n}, -\tau) + [e_{p+n}(t+\tau) - e_{p+n}(t)]\tau^{p+n-1} \\
&= (*)
\end{aligned}$$

Now we consider two cases. First let n be even which is by Lemma 2.11 equivalent to $p + n$ being even. We we carry on:

$$\begin{aligned}
(*) &= \psi_n(t+\tau, \Psi_{n+1}^{t+\tau,t}x(t) - e_{p+n}(t+\tau)(-\tau)^{p+n}, -\tau) + [e_{p+n}(t+\tau-\tau) - e_{p+n}(t+\tau)](-\tau)^{p+n-1} \\
&= \psi_{n+1}(t+\tau, \Psi_{n+1}^{t+\tau,t}x(t), -\tau).
\end{aligned}$$

Thus Ψ_{n+1} is indeed reversible. Since it has the odd consistency order $p+n+1$, Lemma 2.11 implies $\epsilon_{p+(n+1)+1}^{(n+1)} \equiv 0$ and the first case is completed.
Now let n be odd. By the induction hypothesis $\epsilon_{p+n+1}^n \equiv e_{p+n} \equiv 0$. This enables us to repeat the algebra of the first case to produce $(*) = \psi_{n+1}(t+\tau, \Psi_{n+1}^{t+\tau,t}x(t), -\tau)$ a second time and thereby complete the inductive step.
This shows for $N \in \mathbb{N}_0$ that Ψ_{2N} has consistency order $p + 2n$ along the graph of $x\big|_{[t_0,T]}$ and produces the grid function

$$x_\Delta^{2N} = x_\Delta + \sum_{n=0}^{2N-1} e_{p+n}\big|_\Delta \tau^{p+n} = x_\Delta + \sum_{n=0}^{N-1} e_{p+2n}\big|_\Delta \tau^{p+2n}$$

Now we can derive analogously to the proof of Theorem 2.7 the existence of the remainder r_Δ^N with all its properties. □

3 Development of the Algorithm

So far we have sketched the idea of extrapolation and presented the underlying theoretical framework to proof that Algorithm 1 does indeed work. Now we want to develop the details of its implementation.

3.1 The Algorithm of Fixed Order and Stepsize

3.1.1 The Choice of Ψ

Our final goal is an efficient and flexible algorithm with control of stepsize and order. The extrapolated one step method Ψ we are looking for should not only be explicit and reversible to guarantee efficiency but also of low consistency. The latter enables a precise approximation of the optimal consistency order for a given problem (cf. Section 3.2). With this in mind our first choice would be Runge's method, cf. [3, Table 4.1]:

$$\hat{x} = x + \tau \underbrace{f\left(t + \frac{\tau}{2}, x + \frac{\tau}{2} f(t,x)\right)}_{:=\psi(t,x,\tau)}. \tag{3.1}$$

This explicit method has consistency order 2 but unfortunately it is not reversible, in fact:

Lemma 3.1. *There are no reversible consistent explicit Runge–Kutta methods.*

Proof (cf. [3, Lemma 4.43]). Assume the consistent explicit Runge–Kutta method Ψ is reversible. If we integrate the test–equation

$$x' = x, \quad x(0) = 1.$$

It is immediate from Definition 1.11 that there is a polynomial $P \in \mathbb{P}$ such that $\Psi^{t+\tau,t}x = P(\tau)x$ for $t, x, \tau \in \mathbb{R}$. Furthermore P is not constant as equating the coefficients in local error estimate shows. There are some constants $c, \hat{\tau} > 0$ such that

$$|\Phi^{t,0}x(0) - \Psi^{t,0}x(0)| = |e^t - P(\tau)| \leq c\tau^{p+1} \leq c\tau^2 \quad \forall\, |\tau| \leq \min \hat{\tau}, 1.$$

Now reversibility leads to the following contradiction:

$$1 = \Psi^{0,\tau}\Psi^{\tau,0}1 \;\Rightarrow\; P(\tau) = \frac{1}{P(-\tau)} \notin \mathbb{P}.$$

□

This means we have to construct an appropriate method from scratch. Let us have a closer look why Runge's method is not reversible. We would need

$$\begin{aligned} & \psi(t,x,\tau) = \psi(t+\tau,\hat{x},-\tau) \\ \Leftrightarrow\ & f\left(t+\frac{\tau}{2}, x+\frac{\tau}{2}f(t,x)\right) = f\left(t+\frac{\tau}{2}, \hat{x}-\frac{\tau}{2}f(t+\tau,\hat{x})\right) \\ \Leftarrow\ & x+\frac{\tau}{2}f(t,x) = \hat{x}-\frac{\tau}{2}f(t+\tau,\hat{x}) \\ \Leftrightarrow\ & \hat{x} = x+\frac{\tau}{2}\left[f(t,x)+f(t+\tau,\hat{x})\right]. \end{aligned} \tag{3.2}$$

Now we construct a new method using an idea proposed by H.J. Stetter 1970 (cf. [10, Theorem 3.1]). The new method will be based on Equation 3.1 and Equation 3.2 in order to enforce reversibility. To accommodate both equalities we literally have to make some space and double the systems dimensions:

$$\text{Equation 1.1} \quad \longrightarrow \quad y' = \begin{bmatrix} y_1' \\ y_2' \end{bmatrix} = \begin{bmatrix} f(t,y_2) \\ f(t,y_1) \end{bmatrix}, \quad y(t_0) = \begin{bmatrix} x_0 \\ x_0 \end{bmatrix} \tag{3.3}$$

$$\text{Equation 3.1, Equation 3.2} \longrightarrow \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \tau \begin{bmatrix} f\left(t+\frac{\tau}{2}, y_2+\frac{\tau}{2}f(t,y_i)\right) \\ \frac{1}{2}\left[f(t,y_1)+f(t+\tau,\hat{y}_1)\right] \end{bmatrix} \tag{3.4}$$

$$=: y + \tau\theta(t,y,\tau) =: \Theta^{t+\tau,t}y$$

In Equation 3.3 we just duplicated the original initial value problem and interweaved the two block–components by swapping y_1 and y_2 in the definition of the new right side. According to Theorem 1.2 $y(t) = [x(t); x(t)]$ is *the* solution if x solves Equation 1.1. In Equation 3.4 we defined the new discretization. By demanding this method to be consistent and explicit only one degree of freedom in form of the index $i \in \{1,2\}$ in the first block–component of the increment function remained. It turns out that only the choice $i = 1$ leads to a reversible method:

Lemma 3.2. Θ *is reversible and has consistency order 2.*

Proof (cf. [3, Lemma 4.44]). Let Σ_θ denote the domain of θ and fix some $(t,y,\tau) \in \Sigma_\theta$. We check if θ is invariant under interchanging (t,y,τ) with $(t+\tau,\hat{y},-\tau)$. Obviously the increment function's second block–component satisfies this condition. And also first one does since

$$\hat{y}_2 - \frac{\tau}{2}f(t+\tau,\hat{y}_1) = y_2 + \frac{\tau}{2}f(t,y_1) + \frac{\tau}{2}f(t+\tau,\hat{y}_1) - \frac{\tau}{2}f(t+\tau,\hat{y}_1) = y_2 + \frac{\tau}{2}f(t,y_1)$$

implies $f\left(t+\frac{\tau}{2}, \hat{y}_2 - \frac{\tau}{2}f(t,\hat{y}_1)\right) = f\left(t+\frac{\tau}{2}, y_2+\frac{\tau}{2}f(t,y_1)\right)$.
Since $\theta(t,y,0) = [f(t,y_2); f(t,y_1)]$ Theorem 1.8 shows that Θ is at least consistent of order $p \geq 1$. But using the already proven reversibility Lemma 2.11 guarantees even $p \geq 2$. Finally $p = 2$ because the first block–component of $y(t_0+\tau) - \Theta^{t_0+\tau,t_0}y_0$ is identical to the local error of Runge's method which has only consistency order 2. □

It is not necessary to use Θ in an implementation because we can half the dimension again! Stetter showed that there is an *two step method* that can reproduce y^1_Δ (the first block–component of the grid function produced by Θ) on equidistant grids if it is applied two a second grid that has half the stepsize of the original one. Thus the transition from $2d$ to d is paid by a transition from N_Δ to $2N_\Delta$. In particular the two step method in question, the so called *explicit midpoint rule*, has a somewhat nicer recursion. In analogy to Equation 1.2 it is defined on an equidistant grid Δ by

$$\begin{aligned} x_\Delta(t_0) &:= x_0 && \text{(3.5a)}\\ x_\Delta(t_1) &:= x_0 + \tau_\Delta f(t_0, x_0) && \text{(3.5b)}\\ x_\Delta(t_n) &:= x_\Delta(t_{n-2}) + 2\tau_\Delta f(t_{n-1}, x_\Delta(t_{n-})) \quad n \in [2 : N_\Delta]. && \text{(3.5c)} \end{aligned}$$

This is the recursion we will actually use for the computation of of x_{Δ_n} in Algorithm 1. We explain the relation of Θ and the explicit midpoint rule and proof that the latter has an even asymptotic expansion.

Lemma 3.3. *Let $y_\Delta = [y^1_\Delta; y^2_\Delta]$ with $\Delta = (t_0 + n\tau : \ n \in [0 : N_\Delta])$ be produced by Θ. Set $\sigma := \frac{\tau}{2}$ and $\Gamma := (t_0 + n\sigma : \ n \in [0 : 2N_\Delta])$. If x_Γ is obtained by the explicit midpoint rule then $x_\Delta := x_\Gamma\big|_\Delta \equiv y^1_\Delta$.*

Proof. We set $s_n := t_0 + n\sigma$ and show by induction over n:

$$x_\Gamma(s_n) = \begin{cases} y^1_\Delta(s_n) & \text{if } 2 \mid n \\ y^2_\Delta(s_{n-1}) + \sigma f(s_{n-1}, y^1_\Delta(s_{n-1})) & \text{else} \end{cases} \quad n \in [0 : 2N_\Delta].$$

The cases $n \in \{0, 1\}$ are immediate from the respective definitions. Now let $2 \le n \le 2N_\Delta$. We consider two cases and first assume n to be even:

$$\begin{aligned} x_\Gamma(s_n) &= x_\Delta(s_{n-2}) + 2\sigma f(s_{n-1}, x_\Gamma(s_{n-1})) \\ &\overset{\text{ind. hyp.}}{=} y^1_\Delta(s_{n-2}) + \tau f\left(s_{n-2} + \frac{\tau}{2}, y^2_\Delta(s_{n-2}) + \frac{\tau}{2} f(s_{n-2}, y^1_\Delta(s_{n-2}))\right) \\ &= y^1_\Delta(s_n). \end{aligned}$$

For the second case we assume n to be odd – in particular we then know $n \ge 3$. Furthermore we see:

$$\begin{aligned} x_\Gamma(s_n) &= x_\Delta(s_{n-2}) + 2\sigma f(s_{n-1}, x_\Gamma(s_{n-1})) \\ &\overset{\text{ind. hyp.}}{=} y^2_\Delta(s_{n-3}) + \sigma f(s_{n-3}, y^1_\Delta(s_{n-3})) + 2\sigma f(s_{n-1}, y^1_\Delta(s_{n-1})) \\ &= y^2_\Delta(s_{n-1}) + \sigma f(s_{n-1}, y^1_\Delta(s_{n-1})). \end{aligned}$$

□

Now it is clear that the discretization error of the explicit midpoint rule has an asymptotic expansion in even powers of τ if the underlying equidistant grid has an even number of time steps:

Theorem 3.4. *Let x be the solution of Equation 1.1 restricted to the interval $[t_0, T] \subset I_{\max}$ and assume that the right side f is smooth.*

We discretize x with the explicit midpoint rule and an equidistant grid $\Gamma = (t_0, t_1, \ldots, t_{2K-1}, T)$. There is a unique sequence of coefficient functions $(e_{2n})_{n\in\mathbb{N}} \subset \mathcal{C}^\infty(I_{\max}, \mathbb{R}^d)$ that satisfies:

$$\begin{aligned}&\forall\, N \in \mathbb{N}_0 \quad \exists\, K_N, c_N > 0 \quad \forall\, t \in \Delta := (t_{2k} :\ k \in [0:K]) \subset \Gamma \text{ with } K \geq K_N :\\ &x_\Gamma(t) = x(t) - \sum_{n=1}^{N} e_{2n}(t)\tau_\Delta^{2n} - r_\Delta^N(t)\tau_\Delta^{2N+2} \text{ and } \|r_\Delta^N(t)\| \leq c_N(t - t_0). \end{aligned} \tag{3.6}$$

Proof. Let $N \in \mathbb{N}$ be arbitrary and assume $y_\Delta : \Delta \to \mathbb{R}^{2d}$ is produced by Θ starting in $y_\Delta(t_0) = [x_0; x_0]$. According to Theorem 2.12 there is for $\tau_\Delta \leq \hat{\tau}_N$, i.e. $K_N := \frac{T-t_0}{2\hat{\tau}_N} \leq K$, an expansion of the global error of the form

$$\begin{bmatrix} y_\Delta^1(t) \\ * \end{bmatrix} = \begin{bmatrix} \Phi^{t,t_0}x_0 \\ * \end{bmatrix} - \sum_{n=1}^{N} \begin{bmatrix} e_{2n}(t) \\ * \end{bmatrix} \tau^{2n} - \begin{bmatrix} r_\Delta^N(t) \\ * \end{bmatrix} \tau^{2N+2} \text{ and } \left\| \begin{bmatrix} r_\Delta^N(t) \\ * \end{bmatrix} \right\| \leq c_N(t - t_0).$$

Since by by Lemma 3.3 $y_\Delta^1(t) = x_\Gamma(t)$ and without loss of generality also $\|r_\Delta^N(t)\| \leq c_N(t - t_0)$ for $t \in \Delta$, this concludes the proof. □

3.1.2 The Choice of $(\nu_n)_{n\in\mathbb{N}_0}$

We already mentioned in Definition 2.4 that the subdividing sequence $(\nu_n)_{n\in\mathbb{N}_0}$ has to be a strictly increasing sequence of positive integers. This ensures that the stepsizes $\tau_n = \frac{\tau}{\nu_n}$ are well defined and distinct. In the previous section it has additionally become clear that all elements should be even as well since we plan to extrapolate the explicit midpoint rule. Thus we will use instead of $(\nu_n)_{n\in\mathbb{N}_0}$ the subdividing sequence $(2\nu_n)_{n\in\mathbb{N}_0}$. Popular choices for $(\nu_n)_{n\in\mathbb{N}_0}$ are (cf. [8] p. 285, 186):

1. The *harmonic sequence,* $\nu_n = n + 1$,
2. the *Romberg sequence*, $\nu_n = 2^n$, and
3. the *Bulirsch sequence*:

$$\nu_n = \begin{cases} 1 & \text{if } n = 0 \\ \sqrt{2}\ 2^{n/2} & \text{if } n \in \mathbb{N}, 2 \nmid n \\ 1.5\ 2^{n/2} & \text{else} \end{cases}.$$

3.1.3 The Implementation of $\mathcal{E}_\Gamma$

We will show how to compute the interpolation polynomial π from Section 2.2 and discuss how to implement the extrapolation operator.
Fix some $N \in \mathbb{N}_0$, choose a subdividing sequence $(2\nu_n)_{n=0:N}$ and a step size $\tau > 0$. We set $\tau_n := \frac{\tau}{2\nu_n}$ and denote $\mathcal{E}_\Gamma$ by $\mathcal{E}_N$ for $\Gamma = (\tau_n)_{n=0:N}$. In our case the polynomial space Π_N becomes

$$\Pi_N = \left\{ \tau \mapsto a_0 + \sum_{n=0}^{N-1} a_{2+n}\tau^{2+2n} :\ a_n \in \mathbb{R}^d \right\} = \left\{ \tau \mapsto \sum_{n=0}^{N} a_n\tau^{2n} :\ a_n \in \mathbb{R}^d \right\}.$$

Thus we can can obtain $\mathcal{E}_N[x_n]_{n=0:N}$ by classical interpolation. We interpolate the values $(x_n)_{n=0:N}$ in the nodes $(\tau_n^2)_{n=0}$, obtain the interpolant $\pi \in \mathbb{P}_N^d$ and evaluate it 0. Furthermore note that according to Theorem 2.3 we can rescale the nodes $(\tau_n^2)_{n=0:N}$ uniformly and hence it suffices to consider the nodes $(\nu_n^{-2})_{n=0:N}$.
Concerning the implementation of $\mathcal{E}$ we refrain from using the, otherwise widely employed, algorithm of Aitken–Neville (cf. [2, (1.3)]). The reason for this is that it computes $\pi(0)$ in $\mathcal{O}(N^2 d)$ operations - a price we can undercut.
We use the *Lagrange* and *barycentric formulas*. Let $\omega_n := \prod_{i=0:N, i\neq n} \left(\nu_n^{-2} - \nu_i^{-2}\right)^{-1}$ denote the *barycentric weights* and $\rho(\tau) := \prod_{n=0:N} (\tau - \nu_n^{-2})$ the *nodal polynomial*. Then we have

$$\begin{aligned}
\pi(\tau) &= \rho(\tau) \sum_{n=0}^{N} \frac{\omega_n}{\tau - \nu_n^{-2}} x_n && \text{(1. barycentric formula)} \\
&= \left(\sum_{n=0}^{N} \frac{\omega_n}{\tau - \nu_n^{-2}} x_n \right) \left(\sum_{n=0}^{N} \frac{\omega_n}{\tau - \nu_n^{-2}} \right)^{-1} && \text{(2. barycentric formula)} \\
&= \sum_{n=0}^{N} \frac{\rho(\tau)\omega_n}{\tau - \nu_n^{-2}} x_n. && \text{(Lagrange formula)}
\end{aligned}$$

All three are discussed in more detail in [1]. All of these representations of $\pi(0)$ consist of a weighted sum of the the values $(x_n)_{n=0:N}$ which itself is scaled. Since the weights are independent from the stepsize τ, we have to compute them once in exact arithmetic, i.e. as rationals, and tabulate them. The initial computation can be executed for all $N \in [0 : N_{\max}]$ in $\mathcal{O}(N_{\max}^2)$ operations (cf. [1, Section 3]) and the evaluation of $\pi(0)$ takes one matrix vector and one scalar vector multiplication, that are $\mathcal{O}(Nd)$ operations. Furthermore the tabulation of the exact weights is not disturbed by any numerical instabilities one has to deal with if a floating point representation is used.
Eventually we also conjecture that, as we use exact weights, the two barycentric formulas perform numerically identical (although in general the first one is preferred for extrapolation, cf. [11]). We test the three formulas above and the algorithm of Aitken–Neville in a numerical experiment. [1] The results are:

1. The two barycentric formulas perform indeed identical but suffer from under– and overflow. As suggested in [1, Section 7] we rescale the weights uniformly by the *capacity* of the interpolation interval $[\nu_N^{-2}, \nu_0^{-2}]$:

$$\rho(0),\ (\omega_n)_{n=0:N} \longrightarrow \frac{\rho(0)}{c^N},\ (c^N \omega_n)_{n=0:N} \qquad c := \frac{\nu_0^{-2} - \nu_N^{-2}}{4}.$$

   The effect for the Romberg and Bulirsch sequence is neglectable. Using the Lagrange formula on the other hand, which corresponds to rescaling by $c^N := \rho(0)$, seems to be the proper choice.

2. All four algorithms perform comparable (see Figure 3.1). For sufficiently large N each implementation $\widetilde{\mathcal{E}}_N$ satisfies

$$\|\tilde{f}(0) - \widetilde{\mathcal{E}}_N[\tilde{f}(\nu_n^{-2})]_{n=0:N}\|_\infty \leq \|\mathcal{E}_N\|_\infty \mathrm{eps}_{\mathbb{F}}.$$

[1] A Julia notebook of this experiment can be found here: `https://github.com/AlthausKonstantin/Extrapolation`

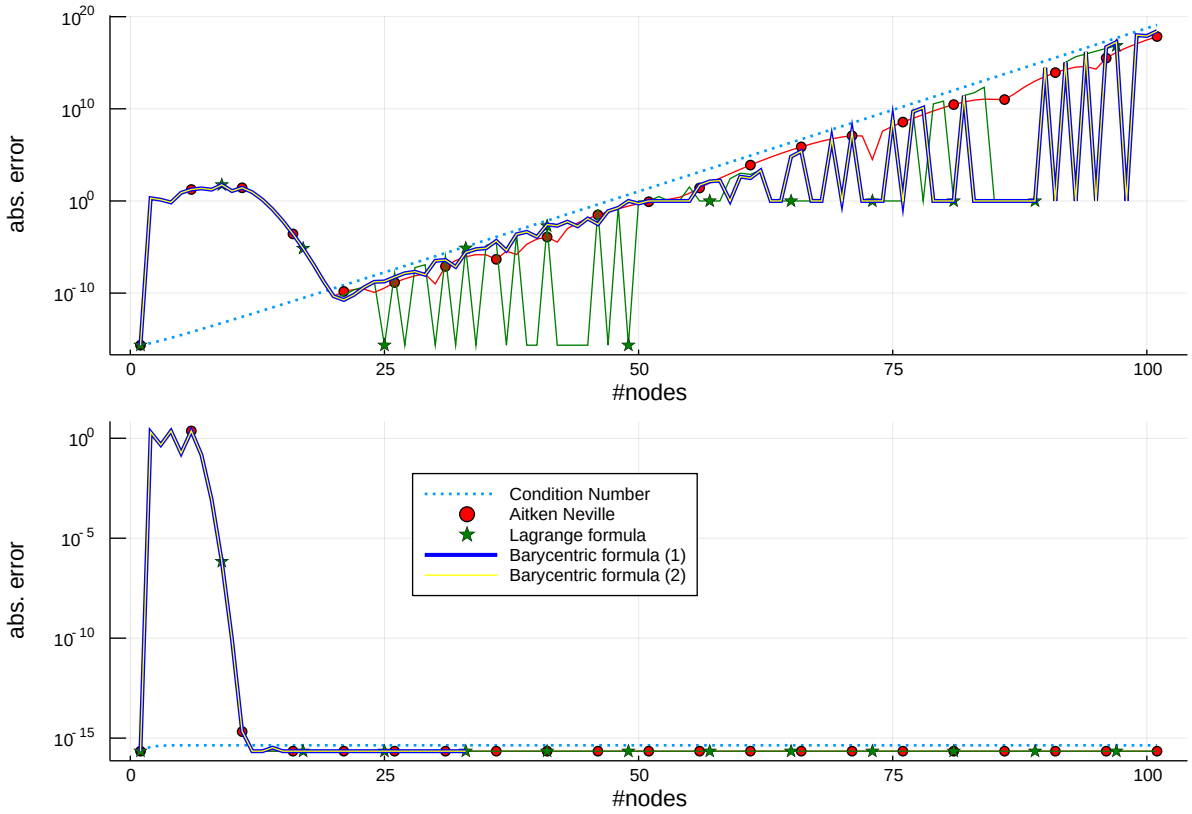

Figure 3.1: Extrapolation of T_{40}, the 40. Chebyshev polynomial, fitted on the interval $[0, 1]$. The condition number is rescaled by the machine precision of `Float64`.

Her $\tilde{f}$ denotes the implementation of the extrapolated test function and $\text{eps}_{\mathbb{F}}$ the machine precision of the chosen floating point numbers $\mathbb{F}$.
The absolute condition number of polynomial extrapolation, i.e. $\|\mathcal{E}_N\|_\infty$, does dramatically depend on the choice of the subdividing sequence. Using the harmonic sequence $N \mapsto \|\mathcal{E}_N\|_\infty$ grows exponentially but seems to be bounded for the Romberg and Bulirsch sequence. In Theorem 3.5 we confirm this observation at least for the harmonic and Romberg sequence.

For our implementation we will tabulate the following quantities for $0 \le m \le n \le N$:

$$\omega_{m,n} := -\nu_m^2 \prod_{\substack{l=0 \\ l\neq m}}^{n} \frac{1}{\nu_m^{-2} - \nu_l^{-2}}, \quad \rho_n := \prod_{l=0}^{n} -\nu_n^{-2}. \tag{3.7}$$

The user may then choose between the first barycentric or Lagrange formula:

$$\mathcal{E}_n[x_l]_{l=0:n} = \rho_n[x_l]_{l=0:n}[\omega_{0,n}; \ldots; \omega_{n,n}] = [x_l]_{l=0:n}[\rho_n\omega_{0,n}; \ldots; \rho_n\omega_{n,n}].$$

Theorem 3.5. *The supremum norm of the extrapolation operator $\mathcal{E}_N$ grows at least exponentially in N if the harmonic sequence is used. If on the other hand the Romberg sequence is used, $\|\mathcal{E}_N\|_\infty$ is bounded.*

Proof. We use the Lagrange polynomials of the nodes $(-\nu_k^{-2})_{k=0:n}$ to obtain

$$\|\mathcal{E}_N\|_\infty = \sum_{n=0}^{N} |L_n(0)| = \sum_{n=0}^{N} \prod_{\substack{k=0 \\ k\neq n}}^{N} \left|1 - \frac{\nu_k^2}{\nu_n^2}\right|^{-1}.$$

First let $\nu_n = n+1$. It suffices to show that $|L_n(0)|$ is unbounded. For this we use the fact that $1 - x \leq e^{-x}$, which is equivalent to $(1-x)^{-1} \geq e^x$ for $x < 1$. We compute:

$$\begin{aligned} |L_N(0)| &= \prod_{k=0}^{N-1} \left(1 - \left(\frac{k+1}{N+1}\right)^2\right)^{-1} \geq \exp\left(\sum_{k=0}^{N-1} \left(\frac{k+1}{N+1}\right)^2\right) \\ &= \exp\left(\frac{N(1+2N)}{6(1+N)}\right) = \mathcal{O}(e^{N/3}) \quad (N \to \infty). \end{aligned}$$

Now let $\nu_n = 2^n$. Again two elemental estimates are used:

$$-\ln(1-x) \leq -4\ln(3/4)x \quad \forall\, x \in [0, 1/4],$$

and

$$x \leq 4^x - 1 \quad \forall\, x \in \mathbb{R}_0^+.$$

Now fix a $n \in [0:N]$. The first estimate implies

$$\begin{aligned} &\ln\left(\prod_{k=0}^{n-1} \left(1 - 4^{k-n}\right)^{-1}\right) \leq -4\ln(3/4) \sum_{k=0}^{n-1} 4^{k-n} \leq \frac{-4\ln(3/4)}{1-1/4} \\ \Rightarrow &\prod_{k=0}^{n-1} \left(1 - 4^{k-n}\right)^{-1} \leq \exp\left(\frac{16\ln(4/3)}{3}\right). \end{aligned}$$

From the second estimate it is immediate that

$$\prod_{k=n+1}^{N} \left|1 - 4^{k-n}\right|^{-1} = \prod_{k=n+1}^{N} \left(4^{k-n} - 1\right)^{-1} \leq \prod_{k=n+1}^{N} \frac{1}{k-n} = \frac{1}{(N-n)!}.$$

We combine both estimates and obtain

$$\|\mathcal{E}_N\|_\infty = \sum_{n=0}^{N} \prod_{\substack{k=0 \\ k\neq n}}^{N} \left|1 - \frac{4^k}{4^n}\right|^{-1} \leq \exp\left(\frac{16\ln(4/3)}{3}\right) \sum_{n=0}^{N} \frac{1}{(N-n)!} \leq \exp\left(1 + \frac{16\ln(4/3)}{3}\right).$$

□

3.1.4 The Final Algorithm

We summarize our discussion on the development of the extrapolation algorithm by casting it a second time in pseudo code. In abuse of notation we specify Ψ_N to denote from

now on the extrapolation of the explicit midpoint rule.

input : initial data (t, x), step length τ
output: $\Psi_N^{t+\tau,t} x$

for $n = 0 : N$ **do**
  $\tau_n := \frac{\tau}{2\nu_n}$
  $x_n^0 := x$ and $x_n^1 := x + \tau_n f(t, x)$
  for $\nu = 2 : 2\nu_n$ **do**
    $x_n^\nu := x_n^{\nu-2} + 2\tau_n f(t + (\nu - 1)\tau_n, x_n^{\nu-1})$
  end
end
$\Psi_N^{t+\tau,t} x := \mathcal{E}_N[x_n^{2\nu_n}]_{n=0:N}$

Algorithm 2: The Nth extrapolation of the explicit midpoint rule

Theorem 3.6. *For $N \in \mathbb{N}$ Ψ_N from Algorithm 2 is an explicit Runge–Kutta method that is invariant under autonomisation and consistent of order $2(N + 1)$.*

As always the first part of the proof is given in a lemma.

Lemma 3.7. *Assume x_Δ defined on the grid $\Delta = (n\tau : n \in [0 : N])$ is a grid function produced by the explicit midpoint rule. Then $\Psi_n^{n\tau,0} x_0 := x_\Delta(n\tau)$ induces for each $n = 1 : N$ an explicit Runge–Kutta method that is invariant under autonomisation.*

Proof. We derive the respective the Butcher arrays $(\mathcal{A}_n, b_n, c_n)$. Note that an Runge–Kutta scheme is consistent if $1 = b'\mathbb{1}_s$ (cf. Theorem 1.8 and Definition 1.11). For $n \in \{1, 2\}$ the claim is true since Ψ_1 is the explicit Euler method while Ψ_2 is Runge's method. For $2 \le n \le N - 1$ we compute using the definitions:

$$\begin{aligned}\Psi_{n+1}^{(n+1)\tau,0} x_0 &= \Psi_{n-1}^{(n-1)\tau,0} x_0 + 2\tau f\left(n\tau, \Psi_n^{n\tau,0} x_0\right) \\ &= x_0 + (n+1)\tau \sum_{i=1}^{s_{n-1}} \frac{n-1}{n+1} b_{n-1}^i k_{n-1}^i \left(0, x_0, (n+1)\tau \frac{n-1}{n+1}\right) + (n+1)\tau \frac{2}{n+1} \cdot \\ &\quad \cdot f\left((n+1)\tau \frac{n}{n+1}, x_0 + (n+1)\tau \sum_{j=1}^{s_n} \frac{n}{n+1} b_n^j k_n^j \left(0, x_0, (n+1)\tau \frac{n}{n+1}\right)\right).\end{aligned}$$

Now we can write down $(\mathcal{A}_{n+1}, b_{n+1}, c_{n+1})$:

$$\begin{array}{c|ccc} \frac{n-1}{n+1} c_{n-1} & \frac{n-1}{n+1}\mathcal{A}_{n-1} & 0 & 0 \\ \frac{n}{n+1} c_n & 0 & \frac{n}{n+1}\mathcal{A}_n & 0 \\ \frac{n}{n+1} & 0 & \frac{n}{n+1} b_n' & 0 \\ \hline & \frac{n-1}{n+1} b_{n-1}' & 0 & \frac{2}{n+1} \end{array}.$$

The induction hypothesis implies that $\mathcal{A}_{n+1}$ and $\mathcal{A}_n$ are lower triangular matrices and thus Ψ_{n+1} is indeed an explicit method. The consistency and invariance under autonomisation is also immediate from the Butcher array. □

And now we can present the

Proof of Theorem 3.6. In Lemma 3.7 we have shown that there are explicit Runge–Kutta schemes $(\widehat{\Psi}_n)_{n=0:N}$ that are and invariant under autonomisation and satisfy $\widehat{\Psi}_n^{t+\tau,t}x = x_n^{2\nu_n}$. Furthermore there is an affine combination $(\lambda_n)_{n=0:N}$ such that $\Psi_N^{t+\tau,t}x = \sum_{n=0}^N \lambda_n \widehat{\Psi}_n^{t+\tau,t}x$. Given the Lagrange polynomials $(L_n)_{n=0N}$ of the nodes $(\nu_n^{-2})_{n=0:N}$ we have $\sum_{n=0}^N L_n \equiv 1 \in \mathbb{P}_N$ and $L_n(0) = \lambda_n$ for all n.
Now we denote the Butcher array of $\widehat{\Psi}_n$ by $(\widehat{\mathcal{A}}_n, \hat{b}_n, \hat{c}_n)$ and are able to state the array of Ψ_N:

$$\begin{array}{c|ccc} \hat{c}_0 & \widehat{\mathcal{A}}_0 & & 0 \\ \vdots & & \ddots & \\ \hat{c}_N & 0 & & \widehat{\mathcal{A}}_N \\ \hline & \lambda_0\hat{b}_0 & \dots & \lambda_N\hat{b}_N \end{array}.$$

As claimed Ψ_N is an explicit method as $\mathcal{A}_N$ is a lower triangular matrix and it is invariant under autonomisation since $\mathcal{A}_N\mathbb{1} = c_N$. In Theorem 2.5 we have already proven that Ψ_N has consistency order $2(N+1)$ if the right side f is smooth. But as Ψ_N is ax Runge–Kutta method and invariant under autonomisation, this already implies that Ψ_N is consistent of order $2(N+1)$ in the sense of Definition 1.7, i.e. $f \in \mathcal{C}^{2(N+1)}(\Omega, \mathbb{R}^d)$ is already sufficient (cf. [3, Theorem 4.24]). □

Note that Theorem 3.6 proves the existence of Runge–Kutta methods of arbitrary order. A result that is anything but trivial in the light of [3, Chapter 4.2] where the Runge–Kutta methods are constructed as Butcher arrays, that are the solution of nonlinear systems. On the other hand extrapolation methods are comparatively more *expensive*, that is they have a high number of stages, see Table 3.1.
The number of stages equals the number of distinct evaluation points of the right side. Let us count the number of stages of Ψ_n:
Computing $x_n^{2\nu_n}$ requires $2\nu_n$ evaluations of the right side f and only the first evaluation $f(t,x)$ can be reused. Thus Ψ_N has

$$s_N := 1 + \sum_{n=0}^N (2\nu_n - 1) = 2\sum_{n=0}^N \nu_n - N$$

stages. It is immediate that choosing the harmonic sequence yields the smallest possible stage number.

consistency order p	2	4	6	8	10
minimal stage number for p	2	4	7	11	17
s_n	2	5	10	17	26

Table 3.1: Number of stages for the harmonic sequence. The minimal stage numbers relate to *explicit* Runge–Kutta methods and can be found in [8, Chapter II.5].

3.2 Making the Algorithm Adaptive

In this section we develop an *adaptive* algorithm that computes a grid function x_Δ to approximate the solution x on the interval $[t_0, T]$. The algorithm will be adaptive in the sense that it automatically controls at each time step its stepsize τ and order of extrapolation n such that a user specified accuracy is met, e.g. $\|\epsilon_n\| := \|\epsilon_n(t, x, \tau)\| := \|\Phi^{t+\tau,t}x - \Psi_n^{t+\tau,t}x\| \leq \text{tol}$, while the total number of f evaluations is as small as possible.[2] That is:

$$\min \sum_{i=1}^{N_\Delta} s_{n_i} \quad \text{with} \sum_{i=1}^{N_\Delta} \tau_i = T - t_0 \text{ and } \|\epsilon_{n_i}(t_{i-1}, x_\Delta(t_{i-1}), \tau_i)\| \leq \text{tol} \quad i \in [1 : N_\Delta]. \tag{3.8}$$

Here s_{n_i} denotes the number stage number of the discretization used to compute the ith time step, i.e. $x_\Delta(t_i) = \Psi_{n_i}^{t_{i-1}+\tau_i,t_{i-1}} x_\Delta(t_{i-1})$. Note that in particular the total number of grid points N_Δ is not known until the computation has been completed.
An efficient approach to this minimization problem is to choose for each time step the *optimal tuple* $(\hat{\tau}, \hat{n})$ from a given range, e.g. $[t_0, T] \times [1 : N]$, which minimizes the *work per unit step*

$$\omega(\tau, n) := \frac{s_n}{\tau}.$$

In the following we sketch two different successful approaches to simultaneous stepsize and order control . One is due to Deuflhard (cf. [2]) and is the basis of the integrator `DIFEX1` while the second one has been proposed by Hairer and Wanner (cf. [8, Chapter II.9]) and has resulted in the routine `ODEX`.
Both algorithms hinge on the idea of estimating the local error. The idea and implementation of obtaining an approximation $\hat{\epsilon}_n(t, x, \tau) \approx \epsilon_n(t, x, \tau)$ while computing $\Psi_n^{t+\tau,t}x$ will be shown in the next section. After that we will go even further and show how to make a prediction about the local error $\epsilon_m(t, x, \tau)$ of a higher extrapolation order ($m > n$) on basis of the estimate $\hat{\epsilon}_n(t, x, \tau)$. Equipped with these tools we can approximate the optimal tuple in an efficient manner. Namely both adaptive extrapolation methods named above implement the following ideas:
Assume we are given some initial data (t, x) and are provided with a suggestion (τ, n_{ref}) for the optimal tuple.

- First define an *order window*. For the given suggestion (τ, n_{ref}) we are willing to accept $x_m := \Psi_m^{t+\tau,t}x$ as a solution if the extrapolation order m is contained in the order window $W(n_{\text{ref}}) := \{n \in [1 : N] : \ |n - n_{\text{ref}}| \leq 1\}$ and $\|\hat{\epsilon}_m\| \leq \text{tol}$.

- If for some $n \in W(n_{\text{ref}})$ the approximation x_n, which we have just computed, is not accurate enough ($\|\hat{\epsilon}_m\| > \text{tol}$), we check if we can expect convergence for the next extrapolation order and carry on computing x_{n+1}. That prediction is made on basis of the *convergence monitor*.

- *Possible stepsize and order reduction*: If the convergence monitor is violated or $\|\hat{\epsilon}_n\| > \text{tol}$ for all $n \in W(n_{\text{ref}})$, we reduce the stepsize and order. We replace first guess (τ, n_{ref}) with the reduced tuple $(\tau_{\text{red}}, n_{\text{red}})$ and repeat the current step.

[2] Caveat: From now on ϵ_n denotes the local error and not a coefficient function.

- *Optimal stepsize and order*: After we have accepted a solution, say x_m, we compute the optimal tuple $(\hat{\tau}, \hat{n})$ for the *current* step. Those values are then used as the first guess for the next step. For the very first step the user may supply that tuple.

3.2.1 Error Estimates

We approximate the local error $\epsilon_n(t, x, \tau)$ by choosing a second discretization $\widehat{\Psi}_n$ of lower order to obtain the computationally available estimate

$$\hat{\epsilon}_n(t, x, \tau) = \Psi_n^{t+\tau,t} x - \widehat{\Psi}_n^{t+\tau,t} x \tag{3.9}$$

By choosing some $k \in [0 : N]$ we can obtain the second discretization without any additional f–evaluations. We simply set

$$\widehat{\Psi}_n^{t+\tau,t} x := \mathcal{E}_{\{\tau_0,\ldots,\tau_n\}\setminus\{\tau_k\}}[x_l]_{l=0:n,l\neq k}.$$

Theorem 2.5 assures that $\widehat{\Psi}_n$ is a one–step method of order $2n$. Usually one chooses $k = 0$ as $x_0 = x_0^{2\nu_0}$ is expected to be the least accurate approximation of $\Phi^{t+\tau,t}x$. I.e. by choosing $k = 0$ we dismiss the "smallest amount of information".
Analogous to $\mathcal{E}_n$ in Section 3.1.3 we denote $\mathcal{E}_{\{\tau_1,\ldots,\tau_n\}}$ by $\widehat{\mathcal{E}}_n$. For its implementation we will additionally tabulate the rationals

$$\hat{\omega}_{m,n} := -\nu_m^2 \prod_{\substack{l=1\\ l\neq m}}^{n} \frac{1}{\nu_m^{-2} - \nu_l^{-2}}, \quad \hat{\rho}_n := \prod_{l=1}^{n} -\nu_n^{-2} \quad 1 \le m \le n \le N.$$

Then $\widehat{\Psi}_n^{t+\tau,t} x$ is given by

$$\widehat{\Psi}_n^{t+\tau,t} x = \widehat{\mathcal{E}}_n[x_l]_{l=1:n} = \hat{\rho}_n[x_l]_{l=1:n}[\hat{\omega}_{1,n} : \ldots ; \hat{\omega}_{n,n}] = [x_l]_{l=1:n}[\hat{\rho}_n\hat{\omega}_{1,n}; \ldots ; \hat{\rho}_n\hat{\omega}_{n,n}].$$

Accuracy Criterion We specify the accuracy criterion by choosing a *scaled mean square norm*:

$$\|\hat{\epsilon}_n\|^2 := \|D^{-1}\hat{\epsilon}_n\|_{\mathrm{MSN}}^2 = \frac{1}{d}\sum_{i=1}^{d}\left(\frac{|\hat{\epsilon}_n|_i}{\sigma_i}\right)^2. \tag{3.10}$$

The positive definite diagonal matrix $D \in \mathbb{R}^{d\times d}$ is recalculated in each time step. We define its diagonal entries to be $\sigma_i := \max(\sigma_i^{\mathrm{abs}}, \sigma_i^{\mathrm{rel}}|x_n|_i)$. This definition is motivated by the following ideas:

- The user can specify the vectors σ^{abs} and σ^{rel} in $(\mathbb{R}_0^+)^d$ such that the computed approximation x_n is accepted if and only if

  $$|\epsilon_n|_i \approx |\hat{\epsilon}_n|_i \le \sigma_i^{\mathrm{abs}}\mathrm{tol} \text{ and } \frac{|\epsilon_n|_i}{|\Phi^{t+\tau,t}x|_i} \approx \frac{|\hat{\epsilon}_n|_i}{|x_n|_i} \le \sigma_i^{\mathrm{rel}}\mathrm{tol}.$$

  That means the user can control the (approximated) absolute and relative error componentwise (cf. [8, (4.10)]).

- The vector σ^{abs} is a numerical safeguard if a component $|x_n|_i$ becomes too small. Thus on usually chooses parameters that satisfy (componentwise) $0 < |\sigma^{\mathrm{abs}}| \ll |\sigma^{\mathrm{rel}}|$.

- Furthermore the error estimate in Equation 3.10 approximates the relative rather than the absolute error. That yields an estimate that is invariant under external rescaling (cf. [3, pp.195]).

Convergence Monitor The *convergence monitor* is a sequence $(\delta_n^{n_{\text{ref}}}(\text{tol}))_{n=1:N} \subset \mathbb{R}^+$ that reflects the ideal convergence behavior of the discretizations $(x_n)_{n=1:N}$:

$$\delta_n^{n_{\text{ref}}}(\text{tol}) \approx \|\epsilon_n(t,x,\tau)\| \text{ such that } \delta_{n_{\text{ref}}}^{n_{\text{ref}}}(\text{tol}) = \text{tol}.$$

As hinted above we will use the sequence to decide in case we computed an approximation x_n for some $n \in W(n_{\text{ref}})$ that does not meet our prescribed accuracy ($\|\hat{\epsilon}_n\| \leq \text{tol}$), if at least $x_{n_{\text{ref}}+1}$ will be accurate enough, i.e. if

$$\|\hat{\epsilon}_n\| \leq \delta_n^{n_{\text{ref}}+1}(\text{tol}). \tag{3.11}$$

If this relation is satisfied we carry on and compute x_{n+1}. Otherwise we will reduce the stepsize τ and order n_{ref}.

Deuflhard's Ansatz Deuflhard's derives the sequence $\delta_n^{n_{\text{ref}}}(\text{tol}))_{n=1:N}$ in [2] from the viewpoint of information theory. For simplicity we consider only the one dimensional case. Say our goal is $\epsilon_m^{\text{rel}} = \text{tol}$ for some $m \in [1:N]$. Then convergence monitor should satisfy

$$\delta_n^m(\text{tol}) \approx \epsilon_n^{\text{rel}} := \left| \frac{\Phi^{t+\tau,t}x - \Psi_n^{t+\tau,t}x}{\Phi^{t+\tau,t}x} \right| \quad n \in [1:m]. \tag{3.12}$$

Let us denote the s_n evaluations of the right side we used to compute $x_n = \Psi_n^{t+\tau,t}x$ by $(f_i)_{i=1:s_n}$. Then we can interpret Ψ_n as a *code* that maps the input *message* $M = f_1 \dots f_{s_n}$ to the output x_n. On a machine not only the *signals* f_i that make up the input but also the output are elements of the finite set $\mathbb{F}$. Since f is an arbitrary function on Ω that is evaluated at s_n points, we assume that M is uniformly distributed on $\mathbb{F}^{s_n}$.
Now we measure the information of the in– and output of the code Ψ_n. The amount of information in M is given by its *entropy*

$$I_{\text{in}} = H(M) = \mathbb{E}[-\log \mathbb{P}_M] = \sum_{m \in \mathbb{F}^{s_n}} \mathbb{P}_M(m) \log \frac{1}{\mathbb{P}_M(m)} = \sum_{m \in \mathbb{F}^{s_n}} \frac{1}{\#\mathbb{F}^{s_n}} \log \#\mathbb{F}^{s_n} = \alpha s_n$$

for some $\alpha > 0$.
As a measure for the output's information Deuflhard proposes to use the number of correct binary digits of x_n which are given by

$$I_{\text{out}} = \log_2 \frac{1}{\epsilon_n^{\text{rel}}} \approx \log_2 \frac{1}{\delta_n^m(\text{tol})}.$$

If we furthermore assume that there is a $\beta \in (0,1]$ such that

$$I_{\text{out}} = \beta I_{\text{in}},$$

Equation 3.12 yields with $c := \alpha\beta$:

$$-\log_2 \delta_n^m(\text{tol}) \approx c s_n \text{ and } -\log_2 \text{tol} = c s_m.$$

We solve for $\delta_n^m(\text{tol})$ and use the result as a definition:

$$\delta_n^m(\text{tol}) := \text{tol}^{\frac{s_n}{s_m}} \quad 1 \leq n \leq m \leq N. \tag{3.13}$$

Ansatz of Hairer and Wanner Hairer and Wanner advocate a more direct approach by simply assuming

$$\|\hat{\epsilon}_n\| \approx \left(\frac{\nu_0}{\nu_n}\right)^2 \|\hat{\epsilon}_{n-1}\|.$$

Based on that assumption we expect that $\|\hat{\epsilon}_m\| \leq \text{tol}$ if

$$\|\hat{\epsilon}_n\| \leq \gamma_n^m(\text{tol}) := \text{tol} \prod_{i=n+1}^{m} \left(\frac{\nu_i}{\nu_0}\right)^2 \quad 1 \leq n \leq m \leq N.$$

3.2.2 Optimal Stepsize and Order

Now that we can estimate and predict the local error we can use that information to compute the optimal stepsize and order.

Pure Stepsize Control From Equation 3.8 it is immediate that for some *fixed* extrapolation order $n \in [1 : N]$ the optimal stepsize is given by

$$\tau_{\text{opt}} = \sup\{\tau \in \mathbb{R}^+ : \ \|\epsilon_n(t, x, \tau)\| = \text{tol}\}$$

Furthermore we know that the qualitative behavior of the local error as the stepsize approaches 0 is determined by the smallest order of τ in the local error representation from Equation 1.3. That is there is a $c > 0$ such that

$$\|\epsilon_n(t, x, \tau)\| = c\tau^{2n+3} + \mathcal{O}(\tau^{2n+4}) \quad (\tau \to 0).$$

We combine this with $\text{tol} \approx c|\tau_{\text{opt}}|^{2n+3}$, solve for τ_{opt} and obtain the following approximation:

$$\tau_{\text{opt}} \approx \left(\frac{\text{tol}}{\|\epsilon_n(t, x, \tau)\|}\right)^{\frac{1}{2n+3}} \tau + \mathcal{O}(\tau) \quad (\tau \to 0).$$

But our algorithm has only access the approximation $\|\hat{\epsilon}_n\|$ of the local error. Since we use a discretization of lower order to compute that estimate, it displays a different asymptotic behavior:

$$\|D^{-1}\hat{\epsilon}_n(t, x, \tau)\|_{\text{MSN}} = \hat{c}\tau^{2n+1} + \mathcal{O}(\tau^{2n+2}) \quad (\tau \to 0).$$

This motivates computing the approximation of τ_{opt} based on the computationally available error estimate $\hat{\epsilon}_n$ as

$$\hat{\tau}_n := \left(\frac{\text{tol}}{\|\hat{\epsilon}_n(t, x, \tau)\|}\right)^{\frac{1}{2n+1}} \tau. \tag{3.14}$$

One can interpret this choice of the exponent as follows. Assume we start in the first place with the approximation $\widehat{\Psi}_n$ to compute x_Δ and estimate the local error by choosing a second discretization of *higher* order, namely Ψ_n. Following this choice the local error and its estimate are of the same order, i.e. $\epsilon_n, \hat{\epsilon}_n = \mathcal{O}(\tau^{2n+1})$ for $\tau \to 0$. But instead of using the results produced $\widehat{\Psi}$ to define the grid function x_Δ we rather use the approximations obtained by Ψ_n. By assuming $\epsilon_n \approx \hat{\epsilon}_n$ we deem the latter to be *more accurate* than the former and there is no point in "wasting" the better result. For a more elaborate discussion see [3, Chapter 5.3].

Order and Stepsize Control Assume we have accepted at some point the solution x_m with $m \in W(n_{\text{ref}})$. Since we have computed $(x_n^{2\nu_n})_{n=0:m}$, we easily obtain $(\hat{\epsilon}_n)_{n=1:m}$ and $(\hat{\tau}_n)_{n=1:m}$ from Equation 3.9 and Equation 3.14. Then also the estimated work per unite step $\hat{\omega}_n := \omega_n(\hat{\tau}_n, n)$ is computationally available for $n \in [1:m]$. Now we use this data to compute the optimal order $\hat{n}$.
We start by specifying a subset $R \subset [1:N]$ from which we will choose $\hat{n}$.

Deuflhard's Approach Deuflhard restricts $\hat{n}$ to the set $R = [1 : n_{\text{ref}} + 1] \cap [1 : N]$. Then we determine the smallest optimal order $\hat{n} \in [1:m]$ such that

$$\hat{\omega}_{\hat{n}} \leq \hat{\omega}_n \quad n \in [1:m].$$

Now we would also like to predict if it may be even more efficient to choose a $\hat{n} > m$. In his paper Deuflhard suggests to proceed as follows: Only in the case of $\hat{n} = m < \min(n_{\text{ref}} + 1, N)$ we are willing to investigate if the order $\hat{n} = m + 1$ may be even more advantageous. This approach is conservative as we are not only demanding that the tuple $(m + 1, \hat{\tau}_{m+1})$ minimizes the work per unit step for $n \in [1 : m + 1]$ but that it is also preceded by a downward trend:

$$\hat{\omega}_{m-1} > \hat{\omega}_m > \omega(m + 1, \hat{\tau}_{m+1}).$$

Since we have no estimate of ϵ_{m+1} we use the already computed estimate $\hat{\epsilon}_m$ and the convergence monitor to compute $\hat{\tau}_{m+1}$. To this end we determine the optimal stepsize $\hat{\tau}_m$ in two different ways.
One time we use the measured error $\|\hat{\epsilon}_m\|$ and the second time the theoretically derived value $\delta_m^{m+1}(\text{tol})$. The latter is the local error we expect if we computed x_m using the stepsize $\hat{\tau}_{m+1}$ rather than τ:

$$\hat{\tau}_m = \left(\frac{\text{tol}}{\|\hat{\epsilon}_m\|}\right)^{\frac{1}{2m+1}} \tau \text{ and } \hat{\tau}_m \approx \left(\frac{\text{tol}}{\|\delta_m^{m+1}(\text{tol})\|}\right)^{\frac{1}{2m+1}} \hat{\tau}_{m+1} \Rightarrow \hat{\tau}_{m+1} \approx \left(\frac{\text{tol}^{\frac{s_m}{s_{m+1}}}}{\|\hat{\epsilon}_m\|}\right)^{\frac{1}{2m+1}} \tau.$$

Thus we can summarize the stepsize control presented in [2] by introducing the function

$$\lambda_n^m : \mathbb{R}^+ \to \mathbb{R}^+, \epsilon \mapsto \left(\frac{\text{tol}^{\frac{s_m}{s_n}}}{\epsilon}\right)^{\frac{1}{2m+1}} \qquad 1 \leq m \leq n \leq N$$

The optimal stepsize for order n based on the error measured at order m using the stepsize τ is then given by

$$\hat{\tau}_n = \tau \lambda_n^m(\|\hat{\epsilon}_m(t, x, \tau)\|)$$

The Approach of Hairer and Wanner Hairer and Wanner suggest to determine the optimal order $\hat{n} \in R = [2 : N - 1] \cap W(n_{\text{ref}})$ by monitoring the trend of $n \mapsto \hat{\omega}_n$ in the vicinity of m.
In general we are willing to increase (decrease) n_{ref}, i.e. setting $\hat{n} = n_{\text{ref}} \pm 1$, if the work

per unit step decreased (increased) significantly from the order $n = m - 1$ to $n = m$. That means we choose some $0 < \sigma \leq 1$ and define an auxiliary function:

$$h(m) := \begin{cases} m-1 & \text{if } \hat{\omega}_{m-1} < \sigma\hat{\omega}_m \\ m+1 & \text{if } \hat{\omega}_m < \sigma\hat{\omega}_{m-1} \\ m & \text{else} \end{cases}.$$

In the case $m \leq n_{\text{ref}}$ we use the function h to compute $\hat{n}$:

$$\hat{n} = (\text{pr}_R \circ h)(m).$$

Only in the case of $m = n_{\text{ref}}+1$ we also want to use the information $\hat{\omega}_n$ of all $n \in W(n_{\text{ref}})$. Thus Hairer and Wanner suggest first to compute $\hat{n} = (\text{pr}_R \circ h)(m-1)$ and then replace $\hat{n}$ by $n_{\text{ref}} + 1$ if this order is even more profitable:

$$\hat{n} \to n_{\text{ref}} + 1 \quad \text{if } \hat{\omega}_m < \sigma\hat{\omega}_{\hat{n}}$$

After accepting some $\hat{n}$ we compute $\hat{\tau}_{\hat{n}}$. Naturally we use Equation 3.14 if $\hat{n} \leq m$. If on the other hand $\hat{n} = m+1$ we assume that $\omega(\hat{\tau}_{\hat{n}}, \hat{n}) \approx \omega(\hat{\tau}_{\hat{n}-1}, \hat{n}-1) = \omega(\hat{\tau}_m, m)$ and solve for $\hat{\tau}_{m+1}$. This yields:

$$\hat{\tau}_{m+1} := \tau \frac{s_{m+1}}{s_m} \left(\frac{\text{tol}}{\|\hat{\epsilon}_m(t,x,\tau)\|} \right)^{\frac{1}{2m+1}}$$

Analogous to λ_n^m we describe also the stepsize selection according to [8] by a function:

$$\mu_n^m : \mathbb{R}^+ \to \mathbb{R}^+, \epsilon \mapsto \frac{s_n}{s_m} \left(\frac{\text{tol}}{\epsilon} \right)^{\frac{1}{2m+1}} \qquad 1 \leq m \leq n \leq N.$$

Stepsize Reduction In case that we decide at some index $m \in W(n_{\text{ref}})$ to reduce the suggested stepsize and order, we can do so by using the scaling functions λ and μ. Let $(\tau_{\text{red}}, n_{\text{red}})$ denote the adjustment of (τ, n_{ref}). Hairer and Wanner opt to reduce n_{ref} if $m = n_{\text{ref}} - 1$ and otherwise to adhere to the original value. Since we have in any case the estimate $\hat{\epsilon}_{n_{\text{red}}}$ at our disposal we set

$$n_{\text{red}} := \min(n_{\text{ref}}, m) \text{ and } \tau_{\text{red}} := \tau\mu_{n_{\text{red}}}^{n_{\text{red}}}(\|\hat{\epsilon}_{n_{\text{red}}}\|).$$

Deuflhard on the other hand is holding on to the suggested order n_{ref} and is only reducing the stepsize. For the latter we use the error estimate $\hat{\epsilon}_{n_{\text{ref}}-1}$ or if available $\hat{\epsilon}_{n_{\text{ref}}}$, i.e.

$$n_{\text{red}} := n_{\text{ref}} \text{ and } \tau_{\text{red}} := \tau\lambda_{n_{\text{ref}}}^{\min(n_{\text{ref}},m)}(\|\hat{\epsilon}_{\min(n_{\text{ref}},m)}\|).$$

3.2.3 Remarks on the Implementation

Safety Features It is common practice to add some safety factors and boundaries to the control loop of an adaptive algorithm. We make the following replacements.

$$\text{tol} \to \sigma\text{tol}$$

for some $\sigma < 1$. The default value in `ODEX` is $\sigma = 0.65$ (cf. [7, line 150]), Deuflhard proposes $\sigma = 0.25$ instead (cf. [2, (4.5)]).
Additionally we add an upper and lower bound to the scalar functions λ and μ (cf. [8, (4.13)]).

$$\lambda_n^m \to \mathrm{pr}_S \circ \lambda_n^m, \quad S = [\lambda_{\min}, \lambda_{\max}] \subset \mathbb{R}^+$$

For instance we may set $S = [0.02, 4]$. The function μ is then treated in the same way.
In [2] Deuflhard also points out that in some cases it is prudent to restrict the stepsize.

$$\hat{\tau} \to \min(\hat{\tau}, \tau_{\max})$$

Finally we avoid being trapped in an infinite loop by aborting the computation of x_Δ if the the total number of accepted steps N_Δ exceeds $N_{\max}$ or the stepsize has to be reduced too many times for the computation of a single step.

Backward Integration Although we have so far only considered grids with nodes given in ascending order, the developed theory does also cover the case of descending order. One merely has to interchange τ and $|\tau|$ in some lines of the code.

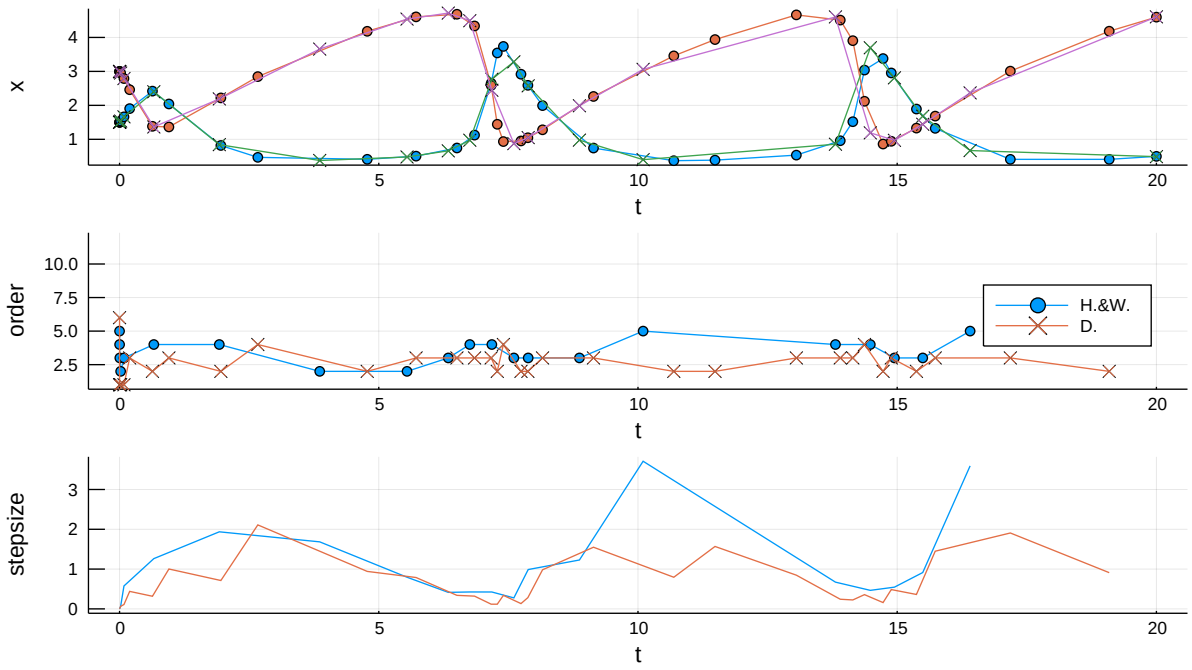

Figure 3.2: Integration of the Brusselator: $\mathrm{tol} = 10^{-3}, \sigma^{\mathrm{rel}} = 1, \sigma^{\mathrm{abs}} = 10^{-3}$. The Algorithm based on Hairer and Wanner's ideas used 21 steps while Deuflhard's approach needed 33 steps. Each rejected in total 6 approximations.

Bibliography

[1] J.-P. Berrut and L. N. Trefethen. "Barycentric Lagrange Interpolation". In: *SIAM Review* 46.3 (2004), pp. 501–517. DOI: 10.1137/S0036144502417715.

[2] P. Deuflhard. "Order and stepsize control in extrapolation methods". In: *Numerische Mathematik* 41.3 (1983), pp. 399–422. DOI: 10.1007/BF01418332.

[3] P. Deuflhard and F. Bornemann. *Scientific Computing with Ordinary Differential Equations*. New York: Springer, 2002. ISBN: 9780387215822.

[4] P. Deuflhard and A. Hohmann. *Numerical Analysis in Modern Scientific Computing: An Introduction*. 2nd ed. New York: Springer, 2003. ISBN: 9780387215846.

[5] J. Elstrodt. *Maß- und Integrationstheorie*. 7th ed. Berlin Heidelberg: Springer–Verlag, 2011. ISBN: 978-3-642-17905-1.

[6] E. Hairer and C. Lubich. "Asymptotic expansions of the global error of fixed-stepsize methods". In: *Numerische Mathematik* 45.3 (1984), pp. 345–360. DOI: 10.1007/BF01391413.

[7] E. Hairer and G. Wanner. *ODEX (source code)*. 2015. URL: http://www.unige.ch/~hairer/prog/nonstiff/odex.f (visited on 12/09/2018).

[8] E. Hairer, G. Wanner, and S. Nørsett. *Solving Ordinary Differential Equations I: Nonstiff Problems*. Vol. 8. Springer Series in Computational Mathematics. Berlin, Heidelberg: Springer, 1993. ISBN: 978-3-540-56670-0.

[9] N. G. Markley. *Principles of Differential Equations*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2011. ISBN: 9780471649564.

[10] H. J. Stetter. "Symmetric two–step algorithms for ordinary differential equations". In: *Computing* 5.3 (1970), pp. 267–280. DOI: 10.1007/BF02248027.

[11] M. Webb, L. N. Trefethen, and P. Gonnet. "Stability of Barycentric Interpolation Formulas for Extrapolation". In: *SIAM Journal on Scientific Computing* 34.6 (2012), pp. 3009–3015. DOI: 10.1137/110848797.