Positively invariant Sets for ODEs and numerical Integration

Peter Giesl [1][a], Sigurdur Hafstein[2][b], and Iman Mehrabinezhad[2][c]
[1]*Department of Mathematics, University of Sussex, Falmer, BN1 9QH, United Kingdom*
[2]*Science Institute, University of Iceland, Dunhagi 3, 107 Reykjavík, Iceland*
P.A.Giesl@sussex.ac.uk, shafstein@hi.is, i.mehrabinezhad@gmail.com

Keywords: Positively invariant sets, Ordinary Differential Equations, Numerical Integration

Abstract: We show that for an ordinary differential equation (ODE) with an exponentially stable equilibrium and any compact subset of its basin of attraction, we can find a larger compact set that is positively invariant for both the dynamics of the system and a numerical method to approximate its solution trajectories. We establish this for both one-step numerical integrators and multi-step integrators using sufficiently small time-steps. Further, we show how to localize such sets using continuously differentiable Lyapunov-like functions and numerically computed continuous, piecewise affine (CPA) Lyapunov-like functions.

1 Introduction

In this paper we study the ordinary differential equation (ODE)

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \quad \mathbf{f} \in C^s(\mathbb{R}^n;\mathbb{R}^n), \quad s \geq 1, \tag{1}$$

and numerical methods to approximate its solution for given initial values. The solution $\mathbf{x}(t)$ to the initial value problem (1) with $\mathbf{x}(0) = \boldsymbol{\xi}$ is denoted by $\boldsymbol{\phi}(t,\boldsymbol{\xi})$.

Our ultimate goal is to characterize the long term behavior of system (1) by computing Lyapunov functions and contraction metrics. As many real-world systems are modelled by (1), such a characterization has numerous practical uses in engineering and science. In this paper we assume that (1) possesses an exponentially stable equilibrium at $\mathbf{x}_0 \in \mathbb{R}^n$, i.e. the Jacobian $D\mathbf{f}(\mathbf{x}_0) \in \mathbb{R}^{n\times n}$ of $\mathbf{f}$ at $\mathbf{x}_0$ is Hurwitz (the real part of all its eigenvalues is negative). We denote the equilibrium's basin of attraction by

$$\mathcal{A}(\mathbf{x}_0) := \{\mathbf{x} \in \mathbb{R}^n \colon \lim_{t\to\infty} \boldsymbol{\phi}(t,\mathbf{x}) = \mathbf{x}_0\}.$$

The Lyapunov stability theory is a generalization of the concept of dissipative energy in physics. It is the centerpiece of practical and theoretical stability analysis and is treated in various detail in virtually all textbooks and monographs on ODE systems, cf. e.g. (Hahn, 1967; Yoshizawa, 1966; Zubov, 1964; Khalil, 2002; Sastry, 1999; Vidyasagar, 2002).

For a general ODE there is no obvious candidate for a Lyapunov function and there are no analytical methods to compute a Lyapunov function. Hence, various methods for the numerical computation of Lyapunov functions have been developed. To name a few, in (Valmorbida and Anderson, 2017; Vannelli and Vidyasagar, 1985) the computation of rational Lyapunov functions was studied, in (Anderson and Papachristodoulou, 2015; Parrilo, 2000) sum-of-squared (SOS) polynomial Lyapunov functions were computed using semi-definite optimization (SOS method), see also (Ratschan and She, 2010; Kamyar and Peet, 2015) for other approaches using polynomials, and in (Giesl, 2007) a Zubov type PDE was numerically solved using radial basis functions (RBF method). For an overview of more methods see, e.g., the review paper (Giesl and Hafstein, 2015b). In (Julian et al., 1999; Marinósson, 2002) linear programming was used to compute continuous and piecewise affine (CPA) Lyapunov functions; this approach is referred to as the CPA method. In the CPA method the domain, where the Lyapunov function is to be computed, is triangulated, i.e. subdivided into simplices, and a feasibility problem is derived, such that its solution can be used to define a CPA Lyapunov function for the system. In (Giesl and Hafstein, 2014; Hafstein, 2004; Hafstein, 2005) it was shown that the CPA method always succeeds in computing a Lyapunov function for an ODE with an asymptotically stable equilibrium, if the simplices in the triangulation are sufficiently small and non-degenerate in a certain sense.

[a] https://orcid.org/0000-0003-1421-6980
[b] https://orcid.org/0000-0003-0073-2765
[c] https://orcid.org/0000-0002-6346-9901

In (Giesl and Hafstein, 2015a), the CPA and the RBF method were combined to deliver a method that is as fast as the RBF method and delivers a verified Lyapunov function as the CPA method. This was achieved by solving a system of linear equations in the RBF method rather than a linear optimization problem as in the CPA method. At the same time, the obtained function is verified to be a Lyapunov function by checking that it satisfies the constraints of the feasibility problem. Further, the authors proved that this approach is constructive and one is always able to compute a verified Lyapunov function in any compact subset of an exponentially stable equilibrium's basin of attraction. A similar approach uses numerical integration of solution trajectories of the ODE to generate values for the variables of the feasibility problem of the CPA method and then verifies the constraints, see (Björnsson et al., 2014; Björnsson et al., 2015; Björnsson and Hafstein, 2017; Doban, 2016; Doban and Lazar, 2016; Hafstein et al., 2014a; Hafstein et al., 2014b; Hafstein et al., 2015; Hafstein and Valfells, 2017; Hafstein and Valfells, 2019; Li et al., 2015) and also (Gudmundsson and Hafstein, 2015; Hafstein, 2019a) for more implementation oriented papers. This technique works well in practice and in (Giesl and Hafstein, 2023) it is proved that it always works, assuming that one can find a compact set that is positively invariant for both the ODE and for a numerical scheme to approximate its solution trajectories. The main contribution of this paper is to prove the existence of such a positively invariant set.

The results in this paper also make an important contribution in the context of numerical methods for contraction metrics, see (Giesl et al., 2023c), where the results of this paper are used to derive a uniform error estimate on compact sets. Contraction metrics are Riemannian metrics such that the corresponding distance of adjacent solution trajectories decreases exponentially over time. They have received considerable attention in the literature (Lewis, 1949; Lewis, 1951; Demidovič, 1961; Krasovskiĭ, 1963; Borg, 1960; Hartman, 1961; Hartman, 1964; Lohmiller and Slotine, 1998; Aminzare and Sontag, 2014; Simpson-Porco and Bullo, 2014; Forni and Sepulchre, 2014; Giesl, 2015). The analytical computation of a contraction metric for an ODE is notoriously difficult, even more difficult than the computation of a Lyapunov function, as it requires the computation of a matrix-valued function. Numerical methods for the construction of contraction metrics include (Aylward et al., 2008; Giesl and Hafstein, 2013; Giesl, 2019; Giesl et al., 2023a), see also the recent review (Giesl et al., 2023b).

Let us give an overview of the paper: In Section 2 we recall some facts about numerical integration methods of ODEs and prove a theorem about approximations of solutions with multi-step methods. In Section 3 we establish the existence of positively invariant sets, both for the dynamics of the system (1) and a numerical integration scheme to approximate the solution trajectories in the basin of attraction of an exponentially stable equilibrium; the main result is Theorem 3.5. Such positively invariant sets are very useful, in fact necessary, to prove that Lyapunov functions and contraction metrics can be approximated arbitrarily close on compact subsets of basins of attraction, using numerical integration with subsequent numerical quadrature. We prove our results using the fourth-order Adams-Bashforth (AB4) multi-step scheme initialized with fourth-order Runge-Kutta (RK4), but we discuss how the results can be extended to AB-RK numerical schemes of arbitrary order. Finally we conclude the paper in Section 4.

Notation: We define $\mathbb{N}_0 := \{0,1,2,\ldots\}$ as the set of the natural numbers and $\mathbb{N}_+ := \mathbb{N}_0 \setminus \{0\}$ as the set of the positive natural numbers. We denote the usual p-norms on $\mathbb{R}^n$ and the corresponding induced matrix norms by $\|\cdot\|_p$, $1 \le p < \infty$. For both vectors in $\mathbb{R}^n$ and matrices in $\mathbb{R}^{n\times n}$ we write $\|\cdot\|_{\max}$ for the maximum absolute value norm, i.e. $\|\mathbf{x}\|_{\max} := \max_{i=1,2,\ldots,n}|x_i|$ for a vector $\mathbf{x}\in\mathbb{R}^{n\times n}$ and $\|A\|_{\max} := \max_{i,j=1,2,\ldots,n}|a_{ij}|$ for a matrix $A = (a_{ij}) \in \mathbb{R}^{n\times n}$. Apart from the usual equivalence estimates for the p-norms on $\mathbb{R}^n$, recall the norm equivalence $\|A\|_{\max} \le \|A\|_2 \le n\|A\|_{\max}$ for a matrix $A\in\mathbb{R}^{n\times n}$ and that $\|\cdot\|_{\max}$ is not sub-multiplicative, but $\|Ab\|_{\max} \le n\|A\|_{\max}\|b\|_{\max}$ for $b := B \in \mathbb{R}^{n\times n}$ or $b := \mathbf{b} \in \mathbb{R}^n$. We denote the closure of a set $U \subset \mathbb{R}^n$ by $\overline{U}$ and its boundary by ∂U.

2 Numerical Integration Methods

For computing Lyapunov functions or contraction metrics for the ODE (1), initial value problems can be solved numerically for all vertices of a triangulation of a relevant compact subset of $\mathbb{R}^n$, as discussed in the last section. As the number of vertices can be very high, it is advantageous to use multi-step methods rather than single-step methods, as these are considerably faster for the same degree of precision. Additionally, the examples in (Hafstein, 2019b) indicate that the *corrector* step in the Adams-Bashforth-Moulton predictor-corrector methods does not deliver better results than the Adams-Bashforth method without the corrector step. Therefore we will concentrate in this paper on the Adams-Bashforth method, and in order to be concrete, we will concentrate on the

Adams-Bashforth method of order four (AB4) initialized with the usual Runge-Kutta method of the same order (RK4). However, it is straight-forward to adapt the proofs to the Adams-Bashforth method of any order initialized with the Runge-Kutta of the same order and we will elaborate at the appropriate places.

The true solution to the ODE (1) with initial-value $\boldsymbol{\xi}$ is denoted by $t \mapsto \boldsymbol{\phi}(t,\boldsymbol{\xi})$ and its numerical approximations at $t_i = hi$, $i \in \mathbb{N}_0$, by

$$\widetilde{\boldsymbol{\phi}}_i(\boldsymbol{\xi}) \quad \text{or } \widetilde{\boldsymbol{\phi}}_i, \quad \text{i.e. } \widetilde{\boldsymbol{\phi}}_i := \widetilde{\boldsymbol{\phi}}_i(\boldsymbol{\xi}),$$

dependent on whether the initial-value $\boldsymbol{\xi}$ is clear from the context or not. The time-step $h > 0$ is a parameter of the numerical method. Thus, we set $\widetilde{\boldsymbol{\phi}}_0 = \boldsymbol{\xi}$ and for $i = 0, 1, 2$ we use the RK4 formulas

$$\begin{aligned} \mathbf{k}_1 &= h\mathbf{f}(\widetilde{\boldsymbol{\phi}}_i) \\ \mathbf{k}_2 &= h\mathbf{f}(\widetilde{\boldsymbol{\phi}}_i + \mathbf{k}_1/2) \\ \mathbf{k}_3 &= h\mathbf{f}(\widetilde{\boldsymbol{\phi}}_i + \mathbf{k}_2/2) \\ \mathbf{k}_4 &= h\mathbf{f}(\widetilde{\boldsymbol{\phi}}_i + \mathbf{k}_3) \\ \widetilde{\boldsymbol{\phi}}_{i+1} &= \widetilde{\boldsymbol{\phi}}_i + \frac{1}{6}(\mathbf{k}_1 + 2\mathbf{k}_2 + 2\mathbf{k}_3 + \mathbf{k}_4). \end{aligned} \tag{2}$$

For $i \geq 3$ we use the AB4 formula and set

$$\begin{aligned} \widetilde{\boldsymbol{\phi}}_{i+1} &= \widetilde{\boldsymbol{\phi}}_i \\ &+ \frac{h}{24}\left(55\mathbf{f}(\widetilde{\boldsymbol{\phi}}_i) - 59\mathbf{f}(\widetilde{\boldsymbol{\phi}}_{i-1}) + 37\mathbf{f}(\widetilde{\boldsymbol{\phi}}_{i-2}) - 9\mathbf{f}(\widetilde{\boldsymbol{\phi}}_{i-3})\right). \end{aligned} \tag{3}$$

Let $S \subset \mathbb{R}^n$ be a compact set and $\mathbf{f} \in C^4(\mathbb{R}^n;\mathbb{R}^n)$. It is well known that one can find a constant C_{RK4} such that for every $\boldsymbol{\xi} \in S$ we have

$$\|\widetilde{\boldsymbol{\phi}}_1(\boldsymbol{\xi}) - \boldsymbol{\phi}(h,\boldsymbol{\xi})\|_{\max} \leq C_{\text{RK4}} h^5, \tag{4}$$

where the constant C_{RK4} depends on the derivatives of $\mathbf{f}$, up to and including the fourth order, in a compact, convex set $\widetilde{S} \supset S$, fulfilling $\widetilde{\boldsymbol{\phi}}_1(\boldsymbol{\xi}), \boldsymbol{\phi}(h,\boldsymbol{\xi}) \in \widetilde{S}$ for all $\boldsymbol{\xi} \in S$.

The existence of such a set $\widetilde{S}$ is clear from the formulas (2). The reason for this is simply that the error estimate (2) is derived by using Taylor polynomials for $t \mapsto \boldsymbol{\phi}(t,\boldsymbol{\xi})$ and using the fact that $\dot{\boldsymbol{\phi}}(t,\boldsymbol{\xi}) = \mathbf{f}(\boldsymbol{\phi}(t,\boldsymbol{\xi}))$. From this it is obvious, that for $\tau^* > 0$ there exists a constant C such that (4) holds true with $C_{\text{RK4}} = C$ and all $0 < h \leq \tau^*$ and all $\boldsymbol{\xi} \in S$. This is the property we need for the assumptions in Theorem 3.5.

For multi-step methods like AB4 the error estimate is usually formulated differently. Namely that there exists a constant $C_{\text{AB4}} > 0$ such that

$$\|\widetilde{\boldsymbol{\phi}}_{i+1}(\boldsymbol{\xi}) - \boldsymbol{\phi}((i+1)h,\boldsymbol{\xi})\|_{\max} \leq C_{\text{AB4}} h^5$$

if $\widetilde{\boldsymbol{\phi}}_j = \boldsymbol{\phi}(jh,\boldsymbol{\xi})$ for $j = i, i-1, i-2, i-3$ in formula (3), i.e. if the previous approximations $\widetilde{\boldsymbol{\phi}}_j$ are exact. Again, the constant C_{AB4} depends on the derivatives of $\mathbf{f}$, up to and including the fourth order, in a compact, convex set $\widetilde{S} \supset S$, fulfilling $\widetilde{\boldsymbol{\phi}}_{i+1}(\boldsymbol{\xi}), \boldsymbol{\phi}(jh,\boldsymbol{\xi}) \in \widetilde{S}$ for $j = i, i-1, i-2, i-3$ for all $\boldsymbol{\xi} \in S$.

In this form the error estimate is not useful for our application of computing Lyapunov functions and contraction metrics. Therefore we prove the following theorem.

Theorem 2.1. *(Error estimate for AB4-RK4 method) Consider the system* (1) *and assume that* $\mathbf{f} \in C^4(\mathbb{R}^n;\mathbb{R}^n)$. *Then AB4 initialized with RK4 fulfills the property:*

For any compact set $S \subset \mathbb{R}^n$ *there exist constants* $C, h^* > 0$ *such that for all step-sizes* $0 < h \leq h^*$ *we have for any* $i \in \mathbb{N}_0$ *that*

$$\|\widetilde{\boldsymbol{\phi}}_{i+1}(\boldsymbol{\xi}) - \boldsymbol{\phi}(h,\widetilde{\boldsymbol{\phi}}_i(\boldsymbol{\xi}))\|_2 \leq Ch^5$$

whenever

$$\widetilde{\boldsymbol{\phi}}_0(\boldsymbol{\xi}), \widetilde{\boldsymbol{\phi}}_1(\boldsymbol{\xi}), \ldots, \widetilde{\boldsymbol{\phi}}_i(\boldsymbol{\xi}) \in S.$$

Proof. Fix a constant $0 < h_1 \leq 1$ and a compact, convex set $S' \supset S$ such that $\boldsymbol{\phi}(h,\boldsymbol{\xi}) \in S'$ for all $\boldsymbol{\xi} \in S$ and $\widetilde{\boldsymbol{\phi}}_{i+1}(\boldsymbol{\xi}) \in S'$, whenever $\widetilde{\boldsymbol{\phi}}_j(\boldsymbol{\xi}) \in S$, for $j = 0, 1, 2, \ldots, i$, $i \in \mathbb{N}_0$, and when using step-size $0 < h \leq h_1$. Furthermore, let $0 < h_2 \leq h_1$ be a constant and $\widetilde{S}$ be a compact, convex set such that $\phi([-3h_2, 0], S') \subset \widetilde{S}$. Fix an arbitrary, but constant $\boldsymbol{\xi} \in S$ for the rest of the proof.

For a fixed step-size $0 < h \leq h_2$ denote $\boldsymbol{\phi}_i(\mathbf{y}) := \boldsymbol{\phi}(ih,\mathbf{y})$, $i \in \mathbb{Z}$, and by $\widetilde{\boldsymbol{\phi}}_i$, $i \in \mathbb{N}_0$, the approximation to $\boldsymbol{\phi}_i(\boldsymbol{\xi})$ generated by the numerical method, i.e. AB4 initialized with RK4. There exists a constant $C_{\text{RK4}} > 0$, independent of $\boldsymbol{\xi} \in S$, such that for $0 < h \leq h_2$ we have

$$\|\widetilde{\boldsymbol{\phi}}_{i+1} - \boldsymbol{\phi}_1(\widetilde{\boldsymbol{\phi}}_i)\|_{\max} \leq C_{\text{RK4}} h^5, \tag{5}$$

as long as $0 < h \leq h_2$ and $\widetilde{\boldsymbol{\phi}}_j \in S$, $j = 0, 1, \ldots, i$, for $i = 0, 1, 2$; this is just the classical local truncation error estimate for RK4.

For the steps with the AB4 method it is advantageous to define

$$\begin{aligned} &\text{AB4}_{j=i}(\mathbf{x}_j) := \\ &\mathbf{x}_i + \frac{h}{24}[55\mathbf{f}(\mathbf{x}_i) - 59\mathbf{f}(\mathbf{x}_{i-1}) + 37\mathbf{f}(\mathbf{x}_{i-2}) - 9\mathbf{f}(\mathbf{x}_{i-3})] \end{aligned}$$

where either $\mathbf{x}_j = \widetilde{\boldsymbol{\phi}}_j$ or $\mathbf{x}_j = \boldsymbol{\phi}_j(\mathbf{y})$. Note that $\mathbf{x}_j$ refers to the sequence $(\mathbf{x}_j)_j$, and j in $j = i$ refers to the index j of the sequence $(\mathbf{x}_j)_j$. i is the value of the last index used in the sequence when $\text{AB4}_{j=i}(\mathbf{x}_j)$ is computed from $\mathbf{x}_i, \mathbf{x}_{i-1}, \mathbf{x}_{i-2}, \mathbf{x}_{i-3}$. In the case $\mathbf{x}_j = \widetilde{\boldsymbol{\phi}}_j$ we have

$$\text{AB4}_{j=i}(\widetilde{\boldsymbol{\phi}}_j) = \widetilde{\boldsymbol{\phi}}_{i+1} \quad \text{if } i \geq 3$$

and in the case $\mathbf{x}_j = \boldsymbol{\phi}_j(\mathbf{y})$ there exist a constant $C_{\text{AB4}} > 0$, such that

$$\|\boldsymbol{\phi}_{i+1}(\mathbf{y}) - \text{AB4}_{j=i}(\boldsymbol{\phi}_j(\mathbf{y}))\|_{\max} \leq C_{\text{AB4}} h^5 \tag{6}$$

as long as $0 < h \leq h_2$ and $\boldsymbol{\phi}_j(\mathbf{y}) \in \widetilde{S}$ for $j = i, i-1, i-2, i-3$, in particular for all $\mathbf{y} \in S'$. We now prove the theorem by induction.

Denote by (I) the proposition:

There exist constants $C, C^*, h^* > 0$, such that for every time-step $0 < h \leq h^*$ we have for any $i \in \mathbb{N}_0$ that

$$\|\widetilde{\boldsymbol{\phi}}_i - \boldsymbol{\phi}_1(\widetilde{\boldsymbol{\phi}}_{i-1})\|_{\max} \leq Ch^5, \tag{7}$$

whenever $\widetilde{\boldsymbol{\phi}}_k \in S$ for $k = 0, 1, \ldots, i-1$, and additionally we have for $i \geq 3$ and with $j = 0, 1, 2, 3$, that

$$\|\widetilde{\boldsymbol{\phi}}_{i-j} - \boldsymbol{\phi}_{-j}(\widetilde{\boldsymbol{\phi}}_i)\|_{\max} \leq C^* h^5. \tag{8}$$

The assertions of the theorem clearly follow from (I).

To prove (I) let us first fix the constants. Let $L > 0$ be a Lipschitz constant for $\mathbf{f}$ on $\widetilde{S}$ and set

$$C := 2 \max\{C_{\mathrm{RK4}}, C_{\mathrm{AB4}}\}, \tag{9}$$

$$C^* := e^L(1 + e^L + e^{2L})C, \tag{10}$$

and

$$h^* := \min\left\{h_2, \frac{C - C_{\mathrm{AB4}}}{5C^* L}\right\}. \tag{11}$$

We now show that (I) holds true for $i = 0, 1, 2, 3$. Indeed, (7) follows immediately from (5), and (8) follows immediately from the standard error estimate for an explicit numerical integrator with local truncation error Ch^5 since $0 < h \leq h^*$, see e.g. (Sauer, 2012).

Now assume that (I) holds true for all natural numbers up to and including some $i \geq 3$. We assume that $\widetilde{\boldsymbol{\phi}}_i \in S$ and show that (I) also holds true for $i+1$.

Let us first consider (7) with i replaced by $i+1$. Observe that

$$\begin{aligned}\|\widetilde{\boldsymbol{\phi}}_{i+1} - \boldsymbol{\phi}_1(\widetilde{\boldsymbol{\phi}}_i)\|_{\max} &= \|\,\mathrm{AB4}_{j=i}(\widetilde{\boldsymbol{\phi}}_j) - \boldsymbol{\phi}_1(\widetilde{\boldsymbol{\phi}}_i)\|_{\max} \\ &\leq \|\,\mathrm{AB4}_{j=i}(\widetilde{\boldsymbol{\phi}}_j) - \mathrm{AB4}_{j=0}(\boldsymbol{\phi}_j(\widetilde{\boldsymbol{\phi}}_i))\|_{\max} \\ &\quad + \|\,\mathrm{AB4}_{j=0}(\boldsymbol{\phi}_j(\widetilde{\boldsymbol{\phi}}_i)) - \boldsymbol{\phi}_1(\widetilde{\boldsymbol{\phi}}_i)\|_{\max} \end{aligned} \tag{12}$$

and for the second term on the right-hand-side we have the bound

$$\|\,\mathrm{AB4}_{j=0}(\boldsymbol{\phi}_j(\widetilde{\boldsymbol{\phi}}_i)) - \boldsymbol{\phi}_1(\widetilde{\boldsymbol{\phi}}_i)\|_{\max} \leq C_{\mathrm{AB4}} h^5 \tag{13}$$

by (6).

To bound the first term on the right-hand-side of (12) we use the formula for AB4, that $\boldsymbol{\phi}_0(\widetilde{\boldsymbol{\phi}}_i) = \widetilde{\boldsymbol{\phi}}_i$, the Lipschitz condition on $\mathbf{f}$ on $\widetilde{S}$, and induction hypothesis (8), and we get

$$\begin{aligned}&\|\,\mathrm{AB4}_{j=i}(\widetilde{\boldsymbol{\phi}}_j) - \mathrm{AB4}_{j=0}(\boldsymbol{\phi}_j(\widetilde{\boldsymbol{\phi}}_i))\|_{\max} \\ &= \frac{h}{24}\|-59\mathbf{f}(\widetilde{\boldsymbol{\phi}}_{i-1}) + 59\mathbf{f}(\boldsymbol{\phi}_{-1}(\widetilde{\boldsymbol{\phi}}_i)) \\ &\qquad + 37\mathbf{f}(\widetilde{\boldsymbol{\phi}}_{i-2}) - 37\mathbf{f}(\boldsymbol{\phi}_{-2}(\widetilde{\boldsymbol{\phi}}_i)) \\ &\qquad - 9\mathbf{f}(\widetilde{\boldsymbol{\phi}}_{i-3}) + 9\mathbf{f}(\boldsymbol{\phi}_{-3}(\widetilde{\boldsymbol{\phi}}_i))\|_{\max} \\ &\leq \frac{h}{24}\Big(59\|\mathbf{f}(\widetilde{\boldsymbol{\phi}}_{i-1}) - \mathbf{f}(\boldsymbol{\phi}_{-1}(\widetilde{\boldsymbol{\phi}}_i))\|_{\max} \\ &\qquad + 37\|\mathbf{f}(\widetilde{\boldsymbol{\phi}}_{i-2}) - \mathbf{f}(\boldsymbol{\phi}_{-2}(\widetilde{\boldsymbol{\phi}}_i))\|_{\max} \\ &\qquad + 9\|\mathbf{f}(\widetilde{\boldsymbol{\phi}}_{i-3}) - \mathbf{f}(\boldsymbol{\phi}_{-3}(\widetilde{\boldsymbol{\phi}}_i))\|_{\max}\Big) \\ &\leq \frac{hL}{24}\Big(59\|\widetilde{\boldsymbol{\phi}}_{i-1} - \boldsymbol{\phi}_{-1}(\widetilde{\boldsymbol{\phi}}_i)\|_{\max} \\ &\qquad + 37\|\widetilde{\boldsymbol{\phi}}_{i-2} - \boldsymbol{\phi}_{-2}(\widetilde{\boldsymbol{\phi}}_i)\|_{\max} \\ &\qquad + 9\|\widetilde{\boldsymbol{\phi}}_{i-3} - \boldsymbol{\phi}_{-3}(\widetilde{\boldsymbol{\phi}}_i)\|_{\max}\Big) \\ &\leq \frac{105hL}{24} C^* h^5 < 5C^* L h^6. \end{aligned} \tag{14}$$

Hence, (12), (13), and (14) deliver

$$\|\widetilde{\boldsymbol{\phi}}_{i+1} - \boldsymbol{\phi}_1(\widetilde{\boldsymbol{\phi}}_i)\|_{\max} \leq 5C^* L h^6 + C_{\mathrm{AB4}} h^5 \leq Ch^5, \tag{15}$$

because

$$5C^* Lh + C_{\mathrm{AB4}} \leq 5C^* Lh^* + C_{\mathrm{AB4}} \leq C$$

by (11).

Hence, the bound (7) in (I) holds true for i replaced by $i+1$.

Let us now consider the bound (8) in (I) for i replaced by $i+1$.

The case $j = 0$ is obvious and from

$$\begin{aligned}&\|\widetilde{\boldsymbol{\phi}}_{i+1-j} - \boldsymbol{\phi}_{-j}(\widetilde{\boldsymbol{\phi}}_{i+1})\|_{\max} \\ &= \|\boldsymbol{\phi}_{-j}(\boldsymbol{\phi}_j(\widetilde{\boldsymbol{\phi}}_{i+1-j})) - \boldsymbol{\phi}_{-j}(\widetilde{\boldsymbol{\phi}}_{i+1})\|_{\max} \\ &\leq e^{jLh}\|\boldsymbol{\phi}_j(\widetilde{\boldsymbol{\phi}}_{i+1-j}) - \widetilde{\boldsymbol{\phi}}_{i+1}\|_{\max}\end{aligned}$$

the case $j = 1$, i.e.

$$\|\widetilde{\boldsymbol{\phi}}_i - \boldsymbol{\phi}_{-1}(\widetilde{\boldsymbol{\phi}}_{i+1})\|_{\max} \leq e^{Lh}\|\boldsymbol{\phi}_1(\widetilde{\boldsymbol{\phi}}_i) - \widetilde{\boldsymbol{\phi}}_{i+1}\|_{\max} \leq C^* h^5$$

follows from (15) and $e^{Lh}C \leq C^*$. Here we used the well known

$$\|\boldsymbol{\phi}(t, \mathbf{a}) - \boldsymbol{\phi}(t, \mathbf{b})\|_{\max} \leq e^{L|t|}\|\mathbf{a} - \mathbf{b}\|_{\max}.$$

The cases $j = 2$ and $j = 3$ now follow similarly from (15) and the induction hypothesis (8). For $j = 2$ we have

$$\begin{aligned}&\|\widetilde{\boldsymbol{\phi}}_{i-1} - \boldsymbol{\phi}_{-2}(\widetilde{\boldsymbol{\phi}}_{i+1})\|_{\max} \\ &\leq \|\widetilde{\boldsymbol{\phi}}_{i-1} - \boldsymbol{\phi}_{-1}(\widetilde{\boldsymbol{\phi}}_i)\|_{\max} + \|\boldsymbol{\phi}_{-1}(\widetilde{\boldsymbol{\phi}}_i) - \boldsymbol{\phi}_{-2}(\widetilde{\boldsymbol{\phi}}_{i+1})\|_{\max} \\ &\leq e^{Lh}\|\boldsymbol{\phi}_1(\widetilde{\boldsymbol{\phi}}_{i-1}) - \widetilde{\boldsymbol{\phi}}_i\|_{\max} + e^{2Lh}\|\boldsymbol{\phi}_1(\widetilde{\boldsymbol{\phi}}_i) - \widetilde{\boldsymbol{\phi}}_{i+1}\|_{\max} \\ &\leq e^{Lh}(1 + e^{Lh})Ch^5\end{aligned}$$

and

$$e^{Lh}(1+e^{Lh})Ch^5 \leq C^*.$$

For $j=3$ we have

$$\begin{aligned}\|\widetilde{\boldsymbol{\phi}}_{i-2} - \boldsymbol{\phi}_{-3}(\widetilde{\boldsymbol{\phi}}_{i+1})\|_{\max} &\leq \|\widetilde{\boldsymbol{\phi}}_{i-2} - \boldsymbol{\phi}_{-1}(\widetilde{\boldsymbol{\phi}}_{i-1})\|_{\max} \\ &\quad + \|\boldsymbol{\phi}_{-1}(\widetilde{\boldsymbol{\phi}}_{i-1}) - \boldsymbol{\phi}_{-2}(\widetilde{\boldsymbol{\phi}}_{i})\|_{\max} \\ &\quad + \|\boldsymbol{\phi}_{-2}(\widetilde{\boldsymbol{\phi}}_{i}) - \boldsymbol{\phi}_{-3}(\widetilde{\boldsymbol{\phi}}_{i+1})\|_{\max} \\ &\leq e^{Lh}\|\boldsymbol{\phi}_1(\widetilde{\boldsymbol{\phi}}_{i-2}) - \widetilde{\boldsymbol{\phi}}_{i-1}\|_{\max} \\ &\quad + e^{2Lh}\|\boldsymbol{\phi}_1(\widetilde{\boldsymbol{\phi}}_{i-1}) - \widetilde{\boldsymbol{\phi}}_{i}\|_{\max} \\ &\quad + e^{3Lh}\|\boldsymbol{\phi}_1(\widetilde{\boldsymbol{\phi}}_{i}) - \widetilde{\boldsymbol{\phi}}_{i+1}\|_{\max} \\ &\leq e^{Lh}(1+e^{Lh}+e^{2Lh})Ch^5,\end{aligned}$$

and

$$e^{Lh}(1+e^{Lh}+e^{2Lh})C = C^* \leq C^*.$$

Thus, we have proved the bound (8) of (I) for i replaced by $i+1$, which concludes the proof.

□

Definition 2.2. *(Order of numerical methods)*
A numerical method to solve

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}), \quad \mathbf{f} \in C^p(\mathbb{R}^n;\mathbb{R}^n), \tag{16}$$

is said to be of order $p \in \mathbb{N}$, if for any compact set $S \subset \mathbb{R}^n$ there exist constants $C_{\boldsymbol{\phi}}, h^* > 0$ such that for all step-sizes $0 < h \leq h^*$ we have for any $i \in \mathbb{N}_0$ that

$$\|\widetilde{\boldsymbol{\phi}}_{i+1}(\boldsymbol{\xi}) - \boldsymbol{\phi}(h, \widetilde{\boldsymbol{\phi}}_i(\boldsymbol{\xi}))\|_2 \leq C_{\boldsymbol{\phi}} h^{p+1}$$

whenever

$$\widetilde{\boldsymbol{\phi}}_0(\boldsymbol{\xi}), \widetilde{\boldsymbol{\phi}}_1(\boldsymbol{\xi}), \ldots, \widetilde{\boldsymbol{\phi}}_i(\boldsymbol{\xi}) \in S.$$

With Definition 2.2, Theorem 2.1 can be formulated as: Assume $\mathbf{f} \in C^4(\mathbb{R}^n;\mathbb{R}^n)$ in (1). Then AB4 initialized with RK4 is of order 4 in the sense of Definition 2.2.

Remark 2.3. *It is straight forward to adapt the proof of Theorem 2.1, under the assumption that $\mathbf{f}$ in (1) is in $C^p(\mathbb{R}^n;\mathbb{R}^n)$, to the Adams-Bashforth method of order p initialized with Runge-Kutta of the same order. Hence, an AB-RK pair of order p is a numerical method of order p in the sense of Definition 2.2.*

We are now ready to study positively invariant sets for the ODE (1), that are also positively invariant for the numerical method.

3 Positively Invariant Sets

A positively invariant set for system (1), i.e. a set $P \subset \mathbb{R}^n$ such that $\boldsymbol{\phi}(t,\mathbf{x}) \in P$ for all $t \geq 0$ whenever $\mathbf{x} \in P$, is not necessarily positively invariant for a numerical procedure to approximate its solution trajectories. The following example is quite revealing for the general situation; we show for a simple system and Euler's Method that

- no matter how small the time-step $h > 0$ is, the discrete semi-dynamical system defined by Euler's Method does not necessarily have the same basins of attraction as the original system, and
- if we restrict Euler's Method to certain compact and positively invariant subsets of the basins of attraction of the system, then these sets are also positively invariant for the discrete semi-dynamical system defined by Euler's Method for sufficiently small step sizes.

Recall that semi-dynamical systems are dynamical systems, with the exception that solution trajectories are not defined for negative times.

Example 3.1. *Consider the system $\dot{\theta} = 1$ and $\dot{r} = -r(1-r^2)$ in polar coordinates; the origin is an asymptotically stable equilibrium and the circular disc $B_1 := \{\mathbf{x} \in \mathbb{R}^2 \colon \|\mathbf{x}\|_2 < 1\}$ is its basin of attraction. In Cartesian coordinates the equations of motion are*

$$\begin{pmatrix}\dot{x}\\ \dot{y}\end{pmatrix} = \begin{pmatrix}-x(1-x^2-y^2)-y\\ -y(1-x^2-y^2)+x\end{pmatrix} =: \mathbf{f}(x,y). \tag{17}$$

Using Euler's method at $\mathbf{z}_0 := (0,y)$ and with step-size $h > 0$ delivers the approximation

$$\mathbf{z} := \begin{pmatrix}0\\ y\end{pmatrix} + h\begin{pmatrix}-y\\ -y(1-y^2)\end{pmatrix}$$

at time h. Now, for $y^2 < 1$, we have

$$\|\mathbf{z}\|_2^2 = h^2y^2 + (y - hy(1-y^2))^2 > 1$$

if

$$\begin{aligned}0 &< g(y,h) \\ &:= \underbrace{y^2(1+(1-y^2)^2)}_{=:a>0}h^2 + \underbrace{2y^2(y^2-1)}_{=:b<0}h + \underbrace{y^2-1}_{=:c<0},\end{aligned}$$

i.e. for

$$h > \frac{-b+\sqrt{b^2-4ac}}{2a} > 0.$$

Because g is continuous and $\lim_{y\to\pm 1} g(y,h) = h^2$, for a fixed $h > 0$ we can always find y close enough to 1 (or -1), such that $\|\mathbf{z}\|_2 > 1$. Hence, B_1 is not positively invariant for this system when Euler's method is used, no matter how small $h > 0$ is.

Now fix $0 < r < 1$ and consider the compact set $\overline{B_r} := \{\mathbf{x} \in \mathbb{R}^2 \colon \|\mathbf{x}\|_2 \leq r\}$. Note that $\overline{B_r}$ is positively *invariant for the dynamical system defined by* (17). *If* $h > 0$ *satisfies*

$$\begin{aligned}(1+(1-r^2)^2)h+2(r^2-1) &\leq 0 \qquad (18)\\ i.e.\ h &\leq \frac{2(1-r^2)}{1+(1-r^2)^2},\end{aligned}$$

then for all $\mathbf{z}_0 = (x,y)$ *with* $\|\mathbf{z}_0\|_2 \leq r$ *one has*

$$\|\mathbf{z}\|_2^2 = \|\mathbf{z}_0 + h\mathbf{f}(\mathbf{z}_0)\|_2^2 \leq \|\mathbf{z}_0\|_2^2,$$

i.e. $\overline{B_r}$ *is positively invariant for Euler's method with h fufilling* (18). *Note that condition* (18) *can be derived from considering the special case* $\mathbf{z}_0 = (0,y)$.

We now show in Theorem 3.2 and Corollary 3.5 that for general systems, sublevel-sets $S \subset \mathbb{R}^n$ of certain Lyapunov-like functions V are positively invariant for both the system (1) and numerical methods of order p in the sense of Definition 2.2, $p \in \mathbb{N}_+$ and $1 \leq p \leq s$, to approximate its solution trajectories in S and sufficiently small step size $h > 0$.

Theorem 3.2. *(Positively invariant sets) Consider the system* (1)*, let* $V \in C^1(\mathbb{R}^n;\mathbb{R})$ *and assume S is a compact connected component of* $\{\mathbf{x} \in \mathbb{R}^n \colon V(\mathbf{x}) \leq m\}$, $m \in \mathbb{R}$. *Further assume that* $\nabla V(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}) < 0$ *and that* $\nabla V(\mathbf{x})$ *points out of S for every* $\mathbf{x} \in \partial S$. *Then S is positively invariant for* (1).

Further, assume that $\mathbf{f} \in C^p(\mathbb{R}^n;\mathbb{R}^n)$ *and that we have a numerical method of order p in the sense of Definition 2.2. Then there is an* $h' > 0$ *such that if the time-step h of the numerical method fulfills* $0 < h \leq h'$, *then* $\widetilde{\boldsymbol{\phi}}_{i+1}(\boldsymbol{\xi}) \in S$, *whenever* $\widetilde{\boldsymbol{\phi}}_k(\boldsymbol{\xi}) \in S$ *for* $k = 0,1,2,\ldots,i$, $i \in \mathbb{N}_0$.

Proof. Define

$$\delta := -\frac{1}{2}\max_{\mathbf{x}\in\partial S} \nabla V(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}) > 0$$

and let $\varepsilon > 0$ be such that $\nabla V(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}) \leq -\delta < 0$ for all

$$\mathbf{x} \in \mathcal{W} := \{\mathbf{x} \in \mathbb{R}^n \colon d(\mathbf{x},\partial S) \leq 2\varepsilon\},$$

where

$$d(\mathbf{x},K) := \min_{\mathbf{y}\in K}\|\mathbf{x}-\mathbf{y}\|_2$$

for a compact set K.

To see that S is positively invariant for (1) consider that if it is not, then some solution trajectory starting in S must intersect ∂S at some point $\mathbf{x}$ and then leave S, that is, there exists an $\mathbf{x} \in \partial S$ and an $\tau^* > 0$ such that $\boldsymbol{\phi}(\tau,\mathbf{x}) \in \mathcal{W} \setminus S$ for all $0 < \tau \leq \tau^*$. Then

$$\begin{aligned}m < V(\boldsymbol{\phi}(\tau^*,\mathbf{x})) &= V(\mathbf{x}) + \int_0^{\tau^*} \frac{d}{d\tau}V(\boldsymbol{\phi}(\tau,\mathbf{x}))d\tau\\ &= m + \int_0^{\tau^*} \nabla V(\boldsymbol{\phi}(\tau,\mathbf{x})) \cdot \frac{d}{d\tau}\boldsymbol{\phi}(\tau,\mathbf{x})d\tau \leq m - \delta\tau^*,\end{aligned}$$

a contradiction.

In order to prove the desired property for the numerical method, set

$$\begin{aligned}&\mathcal{V} := \{\mathbf{x} \in \mathbb{R}^n \colon d(\mathbf{x},\partial S) \leq \varepsilon\} \subset \mathcal{W},\\ &F := \max\{\max_{\mathbf{x}\in S}\|\mathbf{f}(\mathbf{x})\|_2, 1\}, \quad \text{and}\\ &h' := \min\{\varepsilon/(2\max\{F,C_{\boldsymbol{\phi}}\}), 1, \tau^*, h^*\}.\end{aligned}$$

Here and later in the proof $C_{\boldsymbol{\phi}}, h^* > 0$ are the constants for the numerical method from Definition 2.2.

Then, for $\mathbf{x} \in S \setminus \mathcal{V}$ and $0 \leq h \leq h'$ we have

$$\|\boldsymbol{\phi}(h,\mathbf{x}) - \mathbf{x}\|_2 \leq \int_0^h \|\mathbf{f}(\boldsymbol{\phi}(s,\mathbf{x}))\|_2 ds \leq hF$$

and it follows that

$$d(\boldsymbol{\phi}(h,\mathbf{x}), S\setminus\mathcal{V}) \leq \varepsilon/2, \quad \forall \mathbf{x} \in S\setminus\mathcal{V}. \qquad (19)$$

Note that from (19), Definition 2.2 and for $\widetilde{\boldsymbol{\phi}}_k(\boldsymbol{\xi}) \in S$ for $k = 0,1,2,\ldots,i$, $\mathbf{x} := \widetilde{\boldsymbol{\phi}}_i(\boldsymbol{\xi})$, we have

$$\begin{aligned}&d(\widetilde{\boldsymbol{\phi}}_{i+1}(\boldsymbol{\xi}), S\setminus\mathcal{V}) \leq d(\boldsymbol{\phi}(h,\mathbf{x}), S\setminus\mathcal{V})\\ &\quad + \|\boldsymbol{\phi}(h,\mathbf{x}) - \widetilde{\boldsymbol{\phi}}_{i+1}(\boldsymbol{\xi})\|_2 \leq \varepsilon/2 + \varepsilon/2 = \varepsilon.\end{aligned}$$

Hence, $\widetilde{\boldsymbol{\phi}}_{i+1}(\mathbf{x}) \in S$ if the time-step of the numerical method fulfills $0 \leq h \leq h'$, whenever $\widetilde{\boldsymbol{\phi}}_k(\boldsymbol{\xi}) \in S$ for $k = 0,1,2,\ldots,i$ and $\widetilde{\boldsymbol{\phi}}_i(\boldsymbol{\xi}) = \mathbf{x} \in S \setminus \mathcal{V}$.

To finish the proof we need to show the statement in the case $\widetilde{\boldsymbol{\phi}}_k(\boldsymbol{\xi}) \in S$ for $k = 0,1,2,\ldots,i$ and

$$\mathbf{x} = \widetilde{\boldsymbol{\phi}}_i(\boldsymbol{\xi}) \in S \cap \mathcal{V}.$$

We assume on the contrary that there are sequences $\boldsymbol{\xi}_j \in S$ and $0 < h_j \leq h'$, $h_j \to 0$ as $j \to \infty$, such that

$$\widetilde{\boldsymbol{\phi}}^j_{i_j+1}(\boldsymbol{\xi}_j) \notin S$$

for all j, although

$$\widetilde{\boldsymbol{\phi}}^j_0(\boldsymbol{\xi}_j), \widetilde{\boldsymbol{\phi}}^j_1(\boldsymbol{\xi}_j), \ldots, \widetilde{\boldsymbol{\phi}}^j_{i_j}(\boldsymbol{\xi}_j) \in S, \;\; \mathbf{x}_j := \widetilde{\boldsymbol{\phi}}^j_{i_j}(\boldsymbol{\xi}_j) \in S \cap \mathcal{V}.$$

Here

$$\widetilde{\boldsymbol{\phi}}^j_0(\boldsymbol{\xi}_j), \widetilde{\boldsymbol{\phi}}^j_1(\boldsymbol{\xi}_j), \ldots, \widetilde{\boldsymbol{\phi}}^j_{i_j}(\boldsymbol{\xi}_j), \widetilde{\boldsymbol{\phi}}^j_{i_j+1}(\boldsymbol{\xi}_j)$$

is the sequence generated by the numerical method with initial value $\boldsymbol{\xi}_j \in S$ and step-size h_j.

Note that since S is positively invariant for (1) we have $\boldsymbol{\phi}(h_j,\mathbf{x}_j) \in S$ and therefore $V(\mathbf{x}_j) \leq m$ and $V(\boldsymbol{\phi}(h_j,\mathbf{x}_j)) \leq m$ for all j. Further, there exists $I \in \mathbb{N}_+$ such that for all $j \geq I$ we have $\boldsymbol{\phi}(\theta h_j,\mathbf{x}_j) \in \mathcal{W} \cap S$ for all $\theta \in [0,1]$ and $V(\widetilde{\boldsymbol{\phi}}^j_{i_j+1}(\boldsymbol{\xi}_j)) > m$. Moreover, there is a convex and compact set $\widetilde{S} \supset S$ such that $\widetilde{\boldsymbol{\phi}}^j_{i_j+1}(\boldsymbol{\xi}_j) \in \widetilde{S}$ for all j. Let L_V be a Lipschitz constant for V on $\widetilde{S}$ and recall that by Definition 2.2 we have

$$\|\widetilde{\boldsymbol{\phi}}^j_{i_j+1}(\boldsymbol{\xi}_j) - \boldsymbol{\phi}(h_j,\mathbf{x}_j)\|_2 \leq C_{\boldsymbol{\phi}} h_j^{p+1}.$$

Now

$$\left|\frac{V(\widetilde{\boldsymbol{\phi}}^j_{i_j+1}(\boldsymbol{\xi}_j)) - V(\boldsymbol{\phi}(h_j,\mathbf{x}_j))}{h_j}\right|$$
$$\leq \frac{L_V\|\widetilde{\boldsymbol{\phi}}^j_{i_j+1}(\boldsymbol{\xi}_j) - \boldsymbol{\phi}(h_j,\mathbf{x}_j)\|_2}{h_j}$$
$$\leq \frac{L_V C_{\boldsymbol{\phi}} h_j^{p+1}}{h_j} = L_V C_{\boldsymbol{\phi}} h_j^p,$$

$$\frac{V(\widetilde{\boldsymbol{\phi}}^j_{i_j+1}(\boldsymbol{\xi}_j)) - V(\mathbf{x}_j)}{h_j} > \frac{m-m}{h_j} = 0,$$

and

$$\frac{V(\widetilde{\boldsymbol{\phi}}^j_{i_j+1}(\boldsymbol{\xi}_j)) - V(\mathbf{x}_j)}{h_j} = \frac{V(\widetilde{\boldsymbol{\phi}}^j_{i_j+1}(\boldsymbol{\xi}_j)) - V(\boldsymbol{\phi}(h_j,\mathbf{x}_j))}{h_j} + \frac{V(\boldsymbol{\phi}(h_j,\mathbf{x}_j)) - V(\mathbf{x}_j)}{h_j} \quad (20)$$

for all $j \geq I$.

Define $g_j(t) := V(\boldsymbol{\phi}(t,\mathbf{x}_j))$. By the Mean-Value theorem there exist $\theta_j \in (0,1)$ such that

$$g_j(h_j) - g_j(0) = g'(\theta_j h_j)h_j \leq -\delta h_j$$

holds for all $j \geq I$. From (20) it follows that

$$0 \leq \limsup_{j\to\infty} \frac{V(\widetilde{\boldsymbol{\phi}}^j_{i_j+1}(\boldsymbol{\xi}_j)) - V(\mathbf{x}_j)}{h_j}$$
$$\leq \limsup_{j\to\infty} \frac{V(\widetilde{\boldsymbol{\phi}}^j_{i_j+1}(\boldsymbol{\xi}_j)) - V(\boldsymbol{\phi}(h_j,\mathbf{x}_j))}{h_j} + \limsup_{j\to\infty} \frac{V(\boldsymbol{\phi}(h_j,\mathbf{x}_j)) - V(\mathbf{x}_j)}{h_j}$$
$$\leq 0 - \delta < 0,$$

which is a contradiction and thus the theorem is proved. □

Remark 3.3. *Note that $V(x,y) = x^2 + y^2$ is a Lyapunov function for system* (17) *on B_1 and for $0 < r < 1$ the set $V^{-1}([0,r^2]) = \{\mathbf{x} \in \mathbb{R}^2 \colon \|\mathbf{x}\|_2 \leq r\} \subset B_1$ is a compact sublevel set that is positively invariant.*

Further,

$$\nabla V(x,y) \cdot \mathbf{f}(x,y) = -2r^2(1-r^2) < 0$$

for $x^2 + y^2 = r^2$, i.e. $\nabla V(x,y) \cdot \mathbf{f}(x,y)$ is less than a negative constant at the boundary of $V^{-1}([0,r^2])$. The last theorem then tells us that there must be an $h' > 0$ so small that $V^{-1}([0,r^2])$ is positively invariant for Euler's Method with step size $0 < h \leq h'$ too.

In applications it can be more convenient to have the following version of Theorem 3.2, which can be used for Lyapunov functions computed by first approximately solving the Zubov's PDE

$$\nabla V(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}) = -\sqrt{\delta^2 + \|\mathbf{f}(\mathbf{x})\|_2^2}$$

using generalized interpolation in reproducing kernel Hilbert spaces and then interpolating the values over the simplices of a triangulation $\mathcal{T}$, see (Giesl and Hafstein, 2015a; Giesl et al., 2021) for more details. Note that in the following theorem, CPA$[\mathcal{T}]$ denotes the set of these interpolating functions, called continuous piecewise affine (CPA) functions, which are affine on each simplex of the triangulation $\mathcal{T}$ and continuous overall. Further, $\nabla V_\nu \in \mathbb{R}^n$ denotes the constant gradient of V in the interior of a simplex $\mathfrak{S}_\nu \in \mathcal{T}$.

Theorem 3.4 (CPA version of Thm. 3.2)**.** *Let $V \in$ CPA$[\mathcal{T}]$ and assume there is a compact connected component S of $\{\mathbf{x} \in \mathbb{R}^n \colon V(\mathbf{x}) \leq m\}$, $m \in \mathbb{R}$. Assume that ∇V_ν points out of S at $\mathbf{x}$ for every $\mathbf{x} \in \partial S \cap \mathfrak{S}_\nu$ and every $\mathfrak{S}_\nu \in \mathcal{T}$, and that there is a constant $c > 0$ such that $\nabla V_\nu \cdot \mathbf{f}(\mathbf{x}) \leq -c$ for every $\mathbf{x}$ in a neighbourhood of ∂S and every ν such that $\mathbf{x} \in \mathfrak{S}_\nu$. Then S is positively invariant for* (1)*.*

Further, assume that $\mathbf{f} \in C^p(\mathbb{R}^n;\mathbb{R}^n)$ and that we have a numerical method of order p in the sense of Definition 2.2. Then there is an $h' > 0$ such that if the time-step h of the numerical method fulfills $0 < h \leq h'$, then $\widetilde{\boldsymbol{\phi}}_{i+1}(\boldsymbol{\xi}) \in S$, whenever $\widetilde{\boldsymbol{\phi}}_k(\boldsymbol{\xi}) \in S$ for $k = 0,1,2,\ldots,i$.

Proof. Essentially, the proof is the same as the proof of Theorem 3.2; the existence of $\delta = c/2$ and $\varepsilon > 0$ now follow directly from the assumptions. The only reasoning that needs modification is why

$$g_j(h_j) - g_j(0) \leq -\delta h_j.$$

For $V \in$ CPA$[\mathcal{T}]$ it follows because for

$$D^+V(\mathbf{x}) := \limsup_{h\to 0+} \frac{V(\boldsymbol{\phi}(h,\mathbf{x})) - V(\mathbf{x})}{h}$$

we have

$$D^+V(\mathbf{x}) \leq \max_{\nu\colon \mathbf{x}\in\mathfrak{S}_\nu} \nabla V_\nu \cdot \mathbf{f}(\mathbf{x}) \leq -\delta,$$

see e.g. (Hafstein, 2020, Lem. 2.2), and by a generalized Mean Value Theorem, see (Scheeffer, 1884) or (Walter, 2004, Thm. 12.24). □

Finally, we can state and prove the main result of this paper.

Theorem 3.5. *(Positively invariant sets for the system and the numerical method) Let $\mathbf{x}_0$ be an exponentially stable equilibrium of* (1)*, where $\mathbf{f} \in C^p(\mathbb{R}^n;\mathbb{R}^n)$ with*

$p \in \mathbb{N}$, and let $K \subset \mathcal{A}(\mathbf{x}_0)$ be compact. Then there exists a compact and connected set S, $K \subset S \subset \mathcal{A}(\mathbf{x}_0)$, with the following property:

Assume we have a numerical method of order p in the sense of Definition 2.2. Then there exists a constant $h' > 0$, such that S is positively invariant both for the original flow $\boldsymbol{\phi}(0,\boldsymbol{\xi}) = \boldsymbol{\xi} \in S$, $t \mapsto \boldsymbol{\phi}(t,\boldsymbol{\xi})$, induced by (1), and for the sequences $\widetilde{\boldsymbol{\phi}}_i(\boldsymbol{\xi})$, $i \in \mathbb{N}_0$, generated by the numerical method with step-size h, $0 < h \leq h'$, for the initial-values $\boldsymbol{\xi} \in S$. In other words, $\widetilde{\boldsymbol{\phi}}_i(\boldsymbol{\xi}) \in S$ for all $\boldsymbol{\xi} \in S$ and all $i \in \mathbb{N}_0$.

Proof. By (Giesl, 2007, Thm. 2.46) there exists a Lyapunov function V for the system (1) fulfilling

$$\nabla V(\mathbf{x}) \cdot \mathbf{f}(\mathbf{x}) = -\|\mathbf{x} - \mathbf{x}_0\|_2 \sqrt{1 + \|\mathbf{f}(\mathbf{x})\|_2^2}$$

for all $x \in \mathcal{A}(\mathbf{x}_0)$. We set $r := \max_{\mathbf{x} \in K} V(\mathbf{x})$ and $S := V^{-1}([0,r])$. Since V is also a Lyapunov function for the system

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})(1 + \|\mathbf{f}(\mathbf{x})\|_2^2)^{-1/2},$$

with a bounded right-hand-side, the set $S \subset \mathcal{A}(\mathbf{x}_0)$ is compact. Since $V(\boldsymbol{\phi}(t,\boldsymbol{\xi})) \leq r$ for all $t \geq 0$ and

$$\mathbf{x}_0 \in \overline{\boldsymbol{\phi}([0,\infty),\boldsymbol{\xi})} \subset V^{-1}([0,r]) = S$$

for all $\boldsymbol{\xi} \in S$, the set S is also connected. Using this Lyapunov function and Theorem 3.2 for the numerical method, the existence of $h' > 0$ with the claimed properties follows. ☐

Remark 3.6. *By Theorem 2.1 the AB4 method initialized by RK4 is a numerical method of order 4 in the sense of Definition 2.2, and thus, fulfills the assumptions in Theorem 3.5. By Remark 2.3 the same applies to the Adams-Bashforth method of any order, initialized with the Runge-Kutta method of the same order. These results are used in (Giesl and Hafstein, 2023) and (Giesl et al., 2023c) to prove that Lyapunov functions and contraction metrics can be approximated arbitrarily close on compact sets using numerical integration and quadrature.*

4 Conclusions

We have shown that for an ODE with an exponentially stable equilibrium $\mathbf{x}_0$ and any compact subset K of its basin of attraction $\mathcal{A}(\mathbf{x}_0)$, we can find a compact and connected set S, $K \subset S \subset \mathcal{A}(\mathbf{x}_0)$, that is positively invariant, both for the ODE and its numerical approximation. We considered the concrete case of using the Adams-Bashforth method of fourth order, initialized with the usual Runge-Kutta method of fourth order, but we also discussed obvious extensions to an arbitrary order. Finally, we demonstrated how such positively invariant sets can be computed in practice.

Acknowledgement: The research done for this paper was partially supported by the Icelandic Research Fund in the grant 228725-051, *Linear Programming as an Alternative to LMIs for Stability Analysis of Switched Systems*, which is gratefully acknowledged.

REFERENCES

Aminzare, Z. and Sontag, E. (2014). Contraction methods for nonlinear systems: A brief introduction and some open problems. In *Proceedings of the 53rd IEEE Conference on Decision and Control*, pages 3835–3847.

Anderson, J. and Papachristodoulou, A. (2015). Advances in computational Lyapunov analysis using sum-of-squares programming. *Discrete Contin. Dyn. Syst. Ser. B*, 20(8):2361–2381.

Aylward, E., Parrilo, P., and Slotine, J.-J. (2008). Stability and robustness analysis of nonlinear systems via contraction metrics and SOS programming. *Automatica*, 44(8):2163–2170.

Björnsson, J., Giesl, P., Hafstein, S., Kellett, C., and Li, H. (2014). Computation of continuous and piecewise affine Lyapunov functions by numerical approximations of the Massera construction. In *Proceedings of the CDC, 53rd IEEE Conference on Decision and Control*, pages 5506–5511, Los Angeles (CA), USA.

Björnsson, J., Giesl, P., Hafstein, S., Kellett, C., and Li, H. (2015). Computation of Lyapunov functions for systems with multiple attractors. *Discrete Contin. Dyn. Syst. Ser. A*, 35(9):4019–4039.

Björnsson, J. and Hafstein, S. (2017). Efficient Lyapunov function computation for systems with multiple exponentially stable equilibria. *Procedia Computer Science*, 108:655–664. Proceedings of the International Conference on Computational Science (ICCS), Zurich, Switzerland, 2017.

Borg, G. (1960). *A condition for the existence of orbitally stable solutions of dynamical systems*. Kungl. Tekn. Högsk. Handl. 153.

Demidovič, B. (1961). On the dissipativity of a certain non-linear system of differential equations. I. *Vestnik Moskov. Univ. Ser. I Mat. Meh.*, 1961(6):19–27.

Doban, A. (2016). *Stability domains computation and stabilization of nonlinear systems: implications for biological systems*. PhD thesis: Eindhoven University of Technology.

Doban, A. and Lazar, M. (2016). Computation of Lyapunov functions for nonlinear differential equations via a Yoshizawa-type construction. *IFAC-PapersOnLine*, 49(18):29 – 34.

Forni, F. and Sepulchre, R. (2014). A differential Lyapunov framework for Contraction Analysis. *IEEE Transactions on Automatic Control*, 59(3):614–628.

Giesl, P. (2007). *Construction of Global Lyapunov Functions Using Radial Basis Functions*. Lecture Notes in Math. 1904, Springer.

Giesl, P. (2015). Converse theorems on contraction metrics for an equilibrium. *J. Math. Anal. Appl.*, (424):1380–1403.

Giesl, P. (2019). Computation of a contraction metric for a periodic orbit using meshfree collocation. *SIAM J. Appl. Dyn. Syst.*, 18(3):1536–1564.

Giesl, P. and Hafstein, S. (2013). Construction of a CPA contraction metric for periodic orbits using semidefinite optimization. *Nonlinear Anal.*, 86:114–134.

Giesl, P. and Hafstein, S. (2014). Revised CPA method to compute Lyapunov functions for nonlinear systems. *J. Math. Anal. Appl.*, 410:292–306.

Giesl, P. and Hafstein, S. (2015a). Computation and verification of Lyapunov functions. *SIAM J. Appl. Dyn. Syst.*, 14(4):1663–1698.

Giesl, P. and Hafstein, S. (2015b). Review of computational methods for Lyapunov functions. *Discrete Contin. Dyn. Syst. Ser. B*, 20(8):2291–2331.

Giesl, P. and Hafstein, S. (2023). *Dynamical Systems in Theoretical Perspective*, chapter Lyapunov Functions by Interpolating Numerical Quadrature: Proof of Convergence, page (to appear). Springer. Springer Proceedings in Mathematics and Statistics.

Giesl, P., Hafstein, S., Haraldsdottir, M., Thorsteinsson, D., and Kawan, C. (2023a). Subgradient algorithm for computing contraction metrics for equilibria. *J. Comput. Dynamics*, 10(2):281–303.

Giesl, P., Hafstein, S., and Kawan, C. (2023b). Review on contraction analysis and computation of contraction metrics. *J. Comput. Dynamics*, 10(1):1–47.

Giesl, P., Hafstein, S., and Mehrabinezhad, I. (2021). Computation and verification of contraction metrics for periodic orbits. *J. Math. Anal. Appl.*, 503(2):Paper No. 125309, 32.

Giesl, P., Hafstein, S., and Mehrabinezhad, I. (2023c). Contraction metrics by numerical integration and quadrature: Uniform error estimate. In *Proceedings of the 20th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, page (submitted).

Gudmundsson, S. and Hafstein, S. (2015). Lyapunov function verification: MATLAB implementation. In *Proceedings of the 1st Conference on Modelling, Identification and Control of Nonlinear Systems (MICNON)*, number 0235, pages 806–811.

Hafstein, S. (2004). A constructive converse Lyapunov theorem on exponential stability. *Discrete Contin. Dyn. Syst. Ser. A*, 10(3):657–678.

Hafstein, S. (2005). A constructive converse Lyapunov theorem on asymptotic stability for nonlinear autonomous ordinary differential equations. *Dynamical Systems: An International Journal*, 20(3):281–299.

Hafstein, S. (2019a). *Computational Science - ICCS 2019: 19th International Conference, Faro, Portugal, June 12-14, 2019, Proceedings, Part V*, chapter Numerical Analysis Project in ODEs for Undergraduate Students, pages 412–434. Springer.

Hafstein, S. (2019b). Numerical ODE solvers and integration methods in the computation of CPA Lyapunov functions. In *Proceedings of the 18th European Control Conference (ECC)*, pages 1136–1141.

Hafstein, S. (2020). CPA Lyapunov functions: Switched systems vs. differential inclusions. In *Proceedings of the 17th International Conference on Informatics in Control, Automation and Robotics (ICINCO)*, pages 745–753.

Hafstein, S., Kellett, C., and Li, H. (2014a). Computation of Lyapunov functions for discrete-time systems using the Yoshizawa construction. In *Proceedings of 53rd IEEE Conference on Decision and Control (CDC)*.

Hafstein, S., Kellett, C., and Li, H. (2014b). Continuous and piecewise affine Lyapunov functions using the Yoshizawa construction. In *Proceedings of the 2014 American Control Conference (ACC)*, pages 548–553 (no. 0170), Portland (OR), USA.

Hafstein, S., Kellett, C., and Li, H. (2015). Computing continuous and piecewise affine Lyapunov functions for nonlinear systems. *J. Comput. Dyn*, 2(2):227 – 246.

Hafstein, S. and Valfells, A. (2017). Study of dynamical systems by fast numerical computation of Lyapunov functions. In *Proceedings of the 14th International Conference on Dynamical Systems: Theory and Applications (DSTA)*, volume Mathematical and Numerical Aspects of Dynamical System Analysis, pages 220–240.

Hafstein, S. and Valfells, A. (2019). Efficient computation of Lyapunov functions for nonlinear systems by integrating numerical solutions. *Nonlinear Dynamics*, 97(3):1895–1910.

Hahn, W. (1967). *Stability of Motion*. Springer, Berlin.

Hartman, P. (1961). On stability in the large for systems of ordinary differential equations. *Canadian J. Math.*, 13:480–492.

Hartman, P. (1964). *Ordinary Differential Equations*. Wiley, New York.

Julian, P., Guivant, J., and Desages, A. (1999). A parametrization of piecewise linear Lyapunov functions via linear programming. *Int. J. Control*, 72(7-8):702–715.

Kamyar, R. and Peet, M. (2015). Polynomial optimization with applications to stability analysis and control – an alternative to sum of squares. *Discrete Contin. Dyn. Syst. Ser. B*, 20(8):2383–2417.

Khalil, H. (2002). *Nonlinear Systems*. Pearson, 3. edition.

Krasovskiĭ, N. N. (1963). *Problems of the Theory of Stability of Motion*. Mir, Moskow, 1959. English translation by Stanford University Press.

Lewis, D. (1951). Differential equations referred to a variable metric. *Amer. J. Math.*, 73:48–58.

Lewis, D. C. (1949). Metric properties of differential equations. *Amer. J. Math.*, 71:294–312.

Li, H., Hafstein, S., and Kellett, C. (2015). Computation of continuous and piecewise affine Lyapunov functions for discrete-time systems. *J. Difference Equ. Appl.*, 21(6):486–511.

Lohmiller, W. and Slotine, J.-J. (1998). On Contraction Analysis for Non-linear Systems. *Automatica*, 34:683–696.

Marinósson, S. (2002). Lyapunov function construction for ordinary differential equations with linear programming. *Dynamical Systems: An International Journal*, 17:137–150.

Parrilo, P. (2000). *Structured Semidefinite Programs and Semialgebraic Geometry Methods in Robustness and Optimization*. PhD thesis: California Institute of Technology Pasadena, California.

Ratschan, S. and She, Z. (2010). Providing a basin of attraction to a target region of polynomial systems by computation of Lyapunov-like functions. *SIAM J. Control Optim.*, 48(7):4377–4394.

Sastry, S. (1999). *Nonlinear Systems: Analysis, Stability, and Control*. Springer.

Sauer, T. (2012). *Numerical Analysis*. Pearson, 2nd edition.

Scheeffer, L. (1884). Zur Theorie der stetigen Funktionen einer reellen Veränderlichen. *Acta Math.*, 5(1):183–194, 279–296.

Simpson-Porco, J. and Bullo, F. (2014). Contraction theory on Riemannian manifolds. *Systems Control Lett.*, 65:74–80.

Valmorbida, G. and Anderson, J. (2017). Region of attraction estimation using invariant sets and rational Lyapunov functions. *Automatica*, 75:37–45.

Vannelli, A. and Vidyasagar, M. (1985). Maximal Lyapunov functions and domains of attraction for autonomous nonlinear systems. *Automatica*, 21(1):69–80.

Vidyasagar, M. (2002). *Nonlinear System Analysis*. Classics in Applied Mathematics. SIAM, 2. edition.

Walter, W. (2004). *Analysis 1*. Springer, 7th edition.

Yoshizawa, T. (1966). *Stability theory by Liapunov's second method*. Publications of the Mathematical Society of Japan, No. 9. The Mathematical Society of Japan, Tokyo.

Zubov, V. I. (1964). *Methods of A. M. Lyapunov and their application*. Translation prepared under the auspices of the United States Atomic Energy Commission; edited by Leo F. Boron. P. Noordhoff Ltd, Groningen.