SAND2018-3923J

BIT manuscript No.
(will be inserted by the editor)

A Lyapunov and Sacker-Sell spectral stability theory for one-step methods

Andrew J. Steyer · **Erik S. Van Vleck**

Received: date / Accepted: date

Abstract Approximation theory for Lyapunov and Sacker-Sell spectra based upon QR techniques is used to analyze the stability of a one-step method solving a time-dependent (nonautonomous) linear ordinary differential equation (ODE) initial value problem in terms of the local error. Integral separation is used to characterize the conditioning of stability spectra calculations. The stability of the numerical solution by a one-step method of a nonautonomous linear ODE using real-valued, scalar, nonautonomous linear test equations is justified. This analysis is used to approximate exponential growth/decay rates on finite and infinite time intervals and establish global error bounds for one-step methods approximating uniformly, exponentially stable trajectories of nonautonomous and nonlinear ODEs. A time-dependent stiffness indicator and a one-step method that switches between explicit and implicit Runge-Kutta methods based upon time-dependent stiffness are developed based upon the theoretical results.

Keywords one-step methods · stiffness · Lyapunov exponents · Sacker-Sell spectrum · nonautonomous differential equations

Mathematics Subject Classification (2010) 65L04 · 65L05 · 65P40 · 34D08 · 34D09

1 Introduction

Stability plays a central role in determining the time asymptotic behavior of dynamical systems. In the seminal works of Lyapunov [33] and Dahlquist [15, 16, 17], sta-

This research was supported in part by NSF grant DMS-1419047.

Andrew J. Steyer
Sandia National Laboratories; P.O. Box 5800, MS 1320; Albuquerque, NM 87185-1320, USA
E-mail: asteyer@sandia.gov

Erik S. Van Vleck
Department of Mathematics - University of Kansas; 1460 Jayhawk Blvd; Lawrence, KS 66045; USA
E-mail: erikvv@ku.edu

bility theories for ordinary differential equation (ODE) initial value problems (IVPs) and methods for their numerical solution were respectively established. The stability of time-dependent (nonautonomous) solutions to ODEs can be determined using a variety of techniques, but does not in general reduce to a time-dependent eigenvalue problem (see the third example on page 24 of [30] or the example at the bottom of page 3 of [14]). Understanding the stability of numerical methods approximating time-dependent solutions to ODE IVPs is important for preventing spurious computational modes, detecting and quantifying stiffness, and controlling the global error. The complementary dynamical systems viewpoint is that the dynamics of numerical solutions should mimic the dynamics of differential equations. In this paper we embrace both of these points of view and use Lyapunov and Sacker-Sell spectral theory to develop a time-dependent stability theory for one-step methods approximating solutions of ODE IVPs.

Our contribution is to establish a Lyapunov stability theory for variable step-size one-step methods approximating time-dependent solutions of ODE IVPs that can fail to satisfy the hypotheses of AN- and B-stability theories (see Equation 1.1 below for an example of such an ODE). Henceforth in this paper, whenever we use the word stability we are referring to Lyapunov stability in either continuous or discrete time. Our main results, Theorems 3.3 and 3.4, characterize the Lyapunov and Sacker-Sell spectra of one-step methods approximating the solution of nonautonomous linear ODEs. We use integral separation, the time-dependent analog of gaps between eigenvalues, to characterize the conditioning of the Lyapunov and Sacker-Sell spectra and related quantities. A time-dependent and orthogonal change of variables is employed to transform to a linear ODE with an upper triangular coefficient matrix, from which spectral endpoints and integral separation properties can be determined from the diagonal. Theorem 3.3 concludes that if the coefficient matrix of a linear ODE is bounded and sufficiently smooth, then the Sacker-Sell spectrum of the numerical solution approximates that of the ODE. Theorem 3.4 concludes that if the ODE has an integral separation structure, then the Lyapunov and Sacker-Sell spectrum of its numerical solution accurately approximate the spectra of the ODE in terms of the local truncation error. Additionally, the endpoints of the spectra of the numerical solution can be estimated from the diagonal entries of the transformed upper triangular coefficient matrix of the linear difference equation it defines.

Theorems 3.3 and 3.4 together with Lemma 3.1 justify characterizing the stability of a one-step method solving a nonautonomous linear ODE of dimension d with d scalar, real-valued, nonautonomous linear test equations. In Theorem 3.5 we demonstrate the necessity of using a step-size restriction to control the time-dependent stability of Runge-Kutta methods (even those that are implicit and A- or AN-stable) solving real- or complex-valued scalar, nonautonomous linear test equations. After this we prove Theorem 3.6 showing that the stability of a Runge-Kutta method solving a real- or complex-valued scalar, nonautonomous linear test equation can be characterized by when the time-averages of the coefficient function of the test equation lie in the linear stability region of the method.

The linear stability results are applied to prove two theorems (Theorems 4.1 and 4.2) on the numerical solution by a one-step method of a uniformly exponentially stable solution of a nonlinear and nonautonomous ODE. Theorem 4.1 shows that as time limits to infinity the error of the numerical solution by a one-step method of a uniformly exponentially stable trajectory of a nonlinear IVP remains accurate in terms of the order of the truncation error of the method. Theorem 4.2 shows that the numerical approximation by a one-step method of a uniformly exponentially stable trajectory of a nonlinear ODE is uniformly exponentially attracted to the exact solution with decay rates estimated by the Sacker-Sell spectrum of the linear variational equation. The nonlinear results, which draw on the spirit of the one-step approximation theory developed in [5], [27], and [29], show that the Lyapunov stability of the numerical solution of a nonlinear ODE IVP by a one-step method can be characterized and quantified in terms of the spectral stability of the numerical solution of the associated linear variational equation.

The linear and nonlinear theoretical results are applied in Section 5. In Section 5.2 we develop an efficient time-dependent stiffness indicator and in Section 5.3 we develop a one-step method, referred to as a QR-IMEX-RK method, that switches between using implicit and explicit Runge-Kutta methods. Our stiffness indicator is computed using Steklov averages approximated from the discrete QR method for computing Lyapunov exponents [22]. This indicator is in general more efficient to compute than methods such as that proposed in Definition 4.1 of [11] that require approximating logarithmic norms or time-dependent eigenvalues and additionally our indicator is able to detect stiffness in IVPs with non-normal Jacobians where logarithmic norms and time-dependent eigenvalues can fail to indicate stiffness. Being able to detect stiffness efficiently and robustly is necessary in the context of our QR-IMEX-RK methods where we switch between using an implicit or explicit Runge-Kutta method based on where approximate Steklov averages are at each time-step in relation to the linear stability regions of the explicit and implicit methods.

The stability of numerical solutions of ODE IVPs is a classic topic in numerical analysis dating back at least to the PhD thesis of Dahlquist (published as [16]) and also [15, 17] where concepts such as A-stability were first introduced. Other stability theories for the numerical solution of nonautonomous and nonlinear ODE IVPs, such as B-stability [9] or algebraic stability and AN-stability [8] provide an analysis for various classes of ODEs that are monotonically contracting. The equivalences amongst these nonlinear and nonautonomous stability theories are investigated in [10]. In the case of Runge-Kutta methods the analysis in AN-, B-, and algebraic stability requires that the methods be implicit and at least A-stable while our analysis holds so long as the method is convergent.

The theory developed in this work is based on the time-dependent spectral stability theories of the Lyapunov and Sacker-Sell spectra. We refer to the monographs [1] by Adrianova and [14] by Coppel as general references on time-dependent stability and related topics such as integral separation and exponential dichotomies. The theory of Lyapunov exponents and the associated Lyapunov spectrum arose from the thesis of Lyapunov [33]. The Sacker-Sell spectrum first appears in the literature in the the fundamental 1978 paper [39] of Sacker and Sell. The Lyapunov spectrum characterizes the exponential stability while the Sacker-Sell spectrum characterizes the uniform exponential stability of a nonautonomous linear ODE or difference equation.

In this paper we apply the QR approximation theory for Lyapunov and Sacker-Sell spectra (see e.g. [18, 19, 21, 23, 25], [26], [42], and [4]). QR approximation theory constructs the orthogonal factor in a QR factorization of a fundamental matrix solution (in continuous or discrete time) to transform a linear system to one with an upper triangular coefficient matrix. Then, assuming either that the system has an integral separation structure or a bounded and continuous coefficient matrix, the end-points of the Lyapunov or Sacker-Sell spectrum respectively can be approximated from the diagonal entries of the transformed upper triangular matrix.

The development of our theory is motivated by the following nonautonomous linear ODE whose coefficient matrix has time-dependent normality:

$$\dot{x} = A(t)x, \quad A(t) = L(t)C(t)L(t)^T, \quad t > 0 \tag{1.1}$$

$$C(t) = \begin{bmatrix} \lambda_1 & \beta(t) \\ 0 & \lambda_2 \end{bmatrix}, \quad L(t) = \begin{bmatrix} \cos(\omega(t)) & -\sin(\omega(t)) \\ \sin(\omega(t)) & \cos(\omega(t)) \end{bmatrix},$$

where $\lambda_1 > 0 > \lambda_2$ with $\lambda_1 + \lambda_2 < 0$, $\beta(t) = \beta_0 + \beta_1(1 + \cos(a_1 t)/(1 + \beta_2 t^2))$, $\omega(t) = a_2 t$, $\sigma := (\lambda_1 + \lambda_2)^2 - 4(a_1(a_1 + \beta_0) + \lambda_1\lambda_2) \geq 0$ and both $\frac{1}{2}(\lambda_1 + \lambda_2 \pm \sqrt{\sigma}) < 0$ for constants $a_1, a_2, \beta_1, \beta_2 \geq 0$ and $\beta_0 \in \mathbb{R}$. The ODE (1.1) does not satisfy the hypotheses of B-stability theory since there exists $v, w \in \mathbb{R}^2$ so that $(v-w)^T A(0)^T (v-w) > 0$ nor AN-stability since $\lambda_1 > 0$ is one of the time-dependent eigenvalues of $A(t)$. However, by using the change of variables $x = L(t)y$ and Theorem 4.3.2 of [1], it follows that zero is an asymptotically stable equilibrium of (1.1).

If we solve (1.1) using the implicit Euler method with step-size $h_0 > 0$ and initial condition $(0,0)^T \neq x_0 \in \mathbb{R}^2$, then the numerical solution $\{x_n\}_{n=0}^{\infty}$ satisfies the following linear difference equation:

$$x_{n+1} = [I - h_0 A(t_{n+1})]^{-1} x_n, \quad n \geq 0. \tag{1.2}$$

If $a_1 = a_2 = 2\pi$, $h_0 = 1$, and $\lambda_1 \in (0,1)$, then the solution of (1.2) with $x_0 \neq (0,0)^T$ is such that $\|x_n\| \to \infty$ as $n \to \infty$ at a rate of $(1-\lambda_1)^n$ where $\|\cdot\|$ is any norm on $\mathbb{R}^2$ despite the fact that the implicit Euler method is AN-stable. In Section 3.1 we prove that there is an $h^* > 0$ so that if $h_0 \in (0, h^*)$, then all solutions of (1.2) decay to zero.

The rest of this paper is organized as follows. In Section 2 we introduce some definitions, notation, and necessary background material. In Section 3.1 we state Theorems 3.3 and 3.4 which are subsequently proved in Section 3.3. We prove Theorems 3.5 and 3.6 in Section 3.2 which is dedicated to the thorough analysis of a scalar, nonautonomous linear test equation. The nonlinear stability results, Theorems 4.1 and 4.2, are stated and proved in Section 4. In Section 5 we develop and test a time-dependent stiffness indicator and an algorithm for switching between implicit and explicit Runge-Kutta methods based on time-dependent stiffness. Concluding remarks are given in Section 6.

2 Preliminaries

2.1 Stability of initial value problems

Consider the following nonautonomous and nonlinear ODE:

$$\dot{x} = f(x,t) \tag{2.1}$$

where $f : \mathbb{R}^d \times (\tau_0,\infty) \to \mathbb{R}^d$ for some positive integer d and $\tau_0 \geq -\infty$. We assume that $f(x,\cdot)$ is bounded for each fixed $x \in \mathbb{R}^d$ and $f \in C^1$ is sufficiently regular so that each IVP

$$\begin{cases} \dot{x} = f(x,t) \\ x(t_0) = x_0 \end{cases} \tag{2.2}$$

has a unique and globally defined solution $x(t;x_0,t_0)$ for all initial conditions $x_0 \in \mathbb{R}^d$ and initial times $t_0 > \tau_0$.

Fix an arbitrary norm $\|\cdot\|$ on $\mathbb{R}^d$ and use the same symbol $\|\cdot\|$ to denote the induced matrix norm on $\mathbb{R}^{d\times d}$. For each positive integer r let I_r denote the $r\times r$ identity matrix. Assume that $x(t;x_0,t_0)$ is a bounded solution of (2.2) and consider the linear variational equation:

$$\dot{x} = A(t)x, \quad t > t_0, \quad A(t) = Df(x(t;x_0,t_0),t), \quad D := \partial/\partial x. \tag{2.3}$$

Since $x(t;x_0,t_0)$ is bounded and $f \in C^1$ it follows that $A(\cdot)$ is bounded and continuous. A fundamental matrix solution of (2.3) is a matrix solution $X : (t_0,\infty) \to \mathbb{R}^{d\times d}$ such that $X(t)$ is invertible for all $t \in (t_0,\infty)$.

Definition 2.1 We say that (2.3) is exponentially stable if for any fundamental matrix solution X of (2.3) there exists $\gamma > 0$ and $K > 0$ so that

$$\|X(t)\| \leq Ke^{-\gamma(t-t_0)}\|X(t_0)\|, \quad t > t_0.$$

(2.3) is said to be uniformly exponentially stable if for any fundamental matrix solution X of (2.3) there exists $\gamma > 0$ and $K > 0$ so that

$$\|X(t)\| \leq Ke^{-\gamma(t-s)}\|X(s)\|, \quad t \geq s > t_0.$$

We characterize exponential and uniform exponential stability using Lyapunov and Sacker-Sell spectra which we define below (see [21] for a review of the definitions and properties of these spectra). The Lyapunov spectrum is defined in terms of characteristic exponents of fundamental matrix solutions of (2.3).

Definition 2.2 Let $\{e_1,\ldots,e_d\}$ denote the standard basis of $\mathbb{R}^d$. For a given fundamental matrix solution $X(t)$ of (2.3) the upper characteristic exponents $\mathscr{U}_1,\ldots,\mathscr{U}_d$ are defined as

$$\mathscr{U}_i = \limsup_{t_0<t\to\infty} \frac{1}{t}\ln\|X(t)e_i\|, \quad i = 1,\ldots,d.$$

The upper Lyapunov exponents $\mu_1,\ldots,\mu_d$ of (2.3) are the upper characteristic exponents whose sum is minimized over all fundamental matrix solutions of (2.3). The lower Lyapunov exponents $\eta_1,\ldots,\eta_d$ of (2.3) are the upper Lyapunov exponents of the opposite adjoint equation $\dot{x} = -A(t)^T x$. The Lyapunov spectrum of (2.3) is $\Sigma_L := \cup_{i=1}^d [\eta_i,\mu_i]$.

The Sacker-Sell spectrum is defined in terms of exponential dichotomies.

Definition 2.3 A linear system of the form (2.3) is said to have exponential dichotomy if there exists a fundamental matrix solution X, a projection P, and constants $K,L \geq 1$ and $\alpha,\beta > 0$ so that

$$\begin{aligned} \|X(t)PX(s)^{-1}\| &\leq Ke^{-\alpha(t-s)}, & t \geq s > t_0, \\ \|X(t)(I_d - P)X(s)^{-1}\| &\leq Le^{\beta(t-s)}, & t_0 < t \leq s. \end{aligned} \tag{2.4}$$

The Sacker-Sell spectrum Σ_{ED} is the set of all $\lambda \in \mathbb{R}$ such that the shifted variational equation $\dot{x} = [A(t) - \lambda I_d]x$ does not have exponential dichotomy. The Sacker-Sell spectrum Σ_{ED} can be expressed as a union of at most d disjoint closed intervals $\Sigma_{ED} = \cup_{i=1}^{d}[\alpha_i^A, \beta_i^A]$ (see Theorem 2 of [39]).

If the Lyapunov spectrum of (2.3) is contained in $(-\infty,0)$, then (2.3) is exponentially stable. A sufficient condition for uniform exponential stability of zero is that the Sacker-Sell spectrum of (2.3) is contained in $(-\infty,0)$. The linear concepts of exponential stability have the following analogous definitions in the nonlinear setting.

Definition 2.4 A trajectory $x(t;x_0,t_0)$ of (2.1) is exponentially stable if there exist constants $\gamma, K, \delta > 0$ so that if $\|u_0 - x_0\| < \delta$ and $t > t_0$, then $\|x(t;u_0,t_0) - x(t;x_0,t_0)\| \leq Ke^{-\gamma(t-t_0)}\|u_0 - x_0\|$. We say that $x(t;x_0,t_0)$ is uniformly exponentially stable if there exist constants $\gamma, K, \delta > 0$ so that if $\|u_s - x(s;x_0,t_0)\| < \delta$ and $t \geq s > t_0$, then $\|x(t;u_s,s) - x(t;x_0,t_0)\| \leq Ke^{-\gamma(t-s)}\|u_s - x(s;x_0,t_0)\|$.

If the linear variational equation (2.3) of $x(t;x_0,t_0)$ is uniformly exponentially stable and $f \in C^2$, then $x(t;x_0,t_0)$ is a uniformly exponentially stable trajectory of (2.1). However, if the linear variational equation of $x(t;x_0,t_0)$ is exponentially stable, but not uniformly exponentially stable, then we cannot even guarantee that $x(t;x_0,t_0)$ is stable (see [36] or Equation 14 in [32] for an example) unless additional hypotheses are placed on (2.3).

2.2 One-step methods

A one-step method is an approximation to solutions of ODE IVPs (2.2) of the following form:

$$x_{n+1} = \varphi(x_n, t_n; f, h) \tag{2.5}$$

where $x_n \approx x(t_n;x_0,t_0)$, $f = f(x,t)$ is the right-hand side function of (2.1), h is a sequence of step-sizes $h = \{h_n\}_{n=0}^{\infty}$ which we always assume is such that $0 < \inf_{n\geq 0} h_n \leq \sup_{n\geq 0} h_n < \infty$, and $t_{n+1} = t_n + h_n$ for all $n \geq 0$. Note that for such sequences there exists $\delta_h \geq 1$ so that $(\sup_{n\geq 0} h_n)/(\inf_{n\geq 0} h_n) \leq \delta_h$. The quantity δ_h provides a bound on the variability of the step-size and is used to quantify the nonlinear stability estimates in Section 4. We let $\|\cdot\|_\infty$ denote the l^∞ norm for sequences with $\|h\|_\infty = \sup_{n\geq 0} h_n$. We say that the one-step method (2.5) has local truncation error of order $p \in \mathbb{N}$ if there exists $h^* > 0$ so that if $f \in \mathbb{C}^{p+1}$ and $\|h\|_\infty \in (0, h^*)$, then the Taylor expansion of any solution $x : (t_0, \infty) \to \mathbb{R}^d$ of (2.1) takes the following form:

$$x(t_{n+1}) - \varphi(x(t_n), t_n; f(x(t_n), t_n), h) = K_n h_n^{p+1}, \quad n \geq 0.$$

where $K_n = K(t_n)$ defines some sequence depending on $x(t)$ and its derivatives. Of special interest is the form of one-step methods approximating the solution of a linear ODE of the form (2.3). We henceforth only consider one-step methods for which, when applied to approximate the numerical solution of a linear ODE of the form (2.3), there exists an $h^* > 0$ so that if $h = \{h_n\}_{n=0}^{\infty}$ is such that $\|h\|_\infty \in (0, h^*)$, then the one-step map (2.5) takes the form $x_{n+1} = \Phi^A(n; h) x_n$ where each $\Phi^A(n; h) \in \mathbb{R}^{d \times d}$ is independent of x_n. While this is true for Runge-Kutta methods and many other well-known one-step methods, one can modify any one-step method to not satisfy this assumption by, for example, adding at each time-step a term of the form $C h_n^{p+1}$ to the one-step map where $0 \neq C \in \mathbb{R}^d$.

2.3 Spectral theory for continuous time systems

Consider the following d dimensional nonautonomous linear ODE:

$$\dot{x} = A(t)x, \quad t > t_0 \tag{2.6}$$

where $A : (t_0, \infty) \to \mathbb{R}^{d \times d}$ is bounded and continuous. The continuous QR method for transforming (2.6) to upper triangular form is as follows. Consider the following ODE (Equations 3.7-8 of [25]):

$$\dot{Q}(t) = Q(t) S(Q(t), A(t)), \quad S(Q, A)_{ij} = \begin{cases} (Q^T A Q)_{i,j}, & i > j \\ 0, & i = j \\ -(Q^T A Q)_{j,i}, & i < j \end{cases}. \tag{2.7}$$

Each orthogonal matrix solution $Q(t) \in \mathbb{R}^{d \times d}$ of (2.7) defines a linear system

$$\dot{y} = B(t)y, \quad B(t) = Q^T(t) A(t) Q(t) - Q^T(t) \dot{Q}(t), \quad t > t_0 \tag{2.8}$$

where $B(t)$ is upper triangular since the definitions of $S(Q, A)$ and $B(t)$ imply that $B_{i,j}(t) = 0$ when $i > j$. We refer to (2.8) as a corresponding upper triangular system (or ODE) to (2.6). Since $x = Q(t)y$ is a Lyapunov transformation the Lyapunov and Sacker-Sell spectral intervals of (2.6) coincide with those of any corresponding upper triangular system.

Theorem 2.1 (Theorems 2.8, 5.5, and 6.1 of [21]) *Let $B : (t_0, \infty) \to \mathbb{R}^{d \times d}$ be bounded, continuous, and upper triangular and let $\Sigma_{ED} = \cup_{i=1}^d [\alpha_i, \beta_i]$ denote the Sacker-Sell spectrum of the ODE $\dot{y} = B(t)y$. For $i = 1, \ldots, d$ we have:*

$$\alpha_i = \liminf_{0 < H \to \infty} \left(\inf_{t > t_0} \frac{1}{H} \int_t^{t+H} B_{i,i}(\tau) d\tau \right), \quad \beta_i = \limsup_{0 < H \to \infty} \left(\sup_{t > t_0} \frac{1}{H} \int_t^{t+H} B_{i,i}(\tau) d\tau \right).$$

□

For a bounded and continuous $A(\cdot)$, the Sacker-Sell spectrum of (2.6) is continuous with respect to $L^\infty(t_0,\infty)$ perturbations of $A(t)$ (for a proof see Theorem 6 of [39] or Chapter 4 of [14]). For the Lyapunov spectrum to be continuous an additional hypothesis must be placed on (2.6).

Definition 2.2 *Suppose that $B:(t_0,\infty)\to\mathbb{R}^{d\times d}$ is bounded, continuous, and upper triangular and that for any $i<j$ one of the two following conditions hold:*

1. *$B_{i,i}$ and $B_{j,j}$ are integrally separated: there exists $a_{i,j}>0$ and $b_{i,,j}\in\mathbb{R}$ so that if $t\geq s>t_0$, then*
$$\int_s^t B_{i,i}(\tau)-B_{j,j}(\tau)d\tau \geq a_{i,j}(t-s)+b_{i,j}. \tag{2.9}$$
2. *For every $\varepsilon>0$ there exists $M_{i,j}(\varepsilon)>0$ so that if $t\geq s>t_0$, then*
$$\left|\int_s^t B_{i,i}(\tau)-B_{j,j}(\tau)d\tau\right| \leq M_{i,j}+\varepsilon(t-s). \tag{2.10}$$

Then we say that $\dot{y}=B(t)y$ and $B(t)$ have an integral separation structure. If the first condition is satisfied for all $i<j$, then we say that $B(t)$ and $\dot{y}=B(t)y$ are integrally separated. If the system (2.6) has a corresponding upper triangular system that has an integral separation structure, then we say that (2.6) and $A(t)$ have an integral separation structure and if the corresponding upper triangular system is integrally separated, then we say that (2.6) and $A(t)$ are integrally separated.

Integral separation is a generic property (see page 21 of [35]) for linear equations of the form (2.6) with respect to the sup-norm topology. This, together with the following theorem, show why it is natural to assume that a linear equation (2.6) has an integral separation structure when approximating Lyapunov spectral intervals.

Theorem 2.3 (Theorem 5.1 in [21]) *Assume that $B:(t_0,\infty)\to\mathbb{R}^{d\times d}$ has an integral separation structure and let $\Sigma_L=\cup_{i=1}^d[\eta_i,\mu_i]$ denote the Lyapunov spectrum of the ODE $\dot{y}=B(t)y$. Then the Lyapunov spectrum of $\dot{y}=B(t)y$ is continuous with respect to $L^\infty(t_0,\infty)$ perturbations of $B(t)$ and for $i=1,\dots,d$ we have:*

$$\eta_i=\liminf_{0<t\to\infty}\frac{1}{t}\int_{t_0}^{t_0+t}B_{i,i}(\tau)d\tau,\quad \mu_i=\limsup_{0<t\to\infty}\frac{1}{t}\int_{t_0}^{t_0+t}B_{i,i}(\tau)d\tau.$$

□

If the system (2.6) does not have an integral separation structure, then the Lyapunov spectrum may be discontinuous with respect to $L^1(t_0,\infty)$ perturbations of the coefficient matrix (see Example 5.4.2 of [1]). Theorems 2.1 and 2.3 are the basis for the assumptions that we place on system (2.6) in Section 3.

Remark 2.1 In this work we never assume that the linear system (2.6) is regular, that is, that its Lyapunov spectrum Σ_L is a point spectrum: $\Sigma_L=\{\mu_1,\dots,\mu_d\}$. Regular systems may have Lyapunov spectra that are not continuous with respect to $L^\infty(t_0,\infty)$ perturbations of the coefficient matrix (see e.g. Example 2.17 of [20] or Example 4.4.1 of [1]) and hence are computationally ill-conditioned.

2.4 Spectral theory for discrete time systems

Consider a family of nonautonomous linear difference equations of the following form:

$$x_{n+1} = \Phi^A(n;h)x_n, \quad n \geq 0 \tag{2.11}$$

where $x_n \in \mathbb{R}^d$, $h = \{h_n\}_{n=0}^{\infty}$ is a sequence of step-sizes, and $\{\Phi^A(n;h)\}_{n=0}^{\infty} \subset \mathbb{R}^{d\times d}$ is bounded and each matrix $\Phi^A(n;h)$ is invertible. We remark that invertibility of $\Phi^A(n;h)$ is only needed to guarantee uniqueness, but not existence, of a discrete QR iteration defined as follows. Let $Q_0 \in \mathbb{R}^{d\times d}$ be an orthogonal matrix and fix some step-size sequence h. Since $\Phi^A(n;h)$ is invertible for all $n \geq 0$ we can form unique QR factorizations $\Phi^A(n;h)Q_n = Q_{n+1}R^A(n;h)$ where $Q_{n+1} \in \mathbb{R}^{d\times d}$ is orthogonal and $R^A(n;h) \in \mathbb{R}^{d\times d}$ is upper triangular with positive diagonal entries. This process is referred to as a discrete QR iteration. The system $u_{n+1} = R^A(n;h)u_n$ where $R^A(n;h) = Q_{n+1}^T \Phi^A(n;h)Q_n$ is referred to as a corresponding upper triangular system and its Lyapunov and Sacker-Sell spectra coincide with those of (2.11).

We shall always use the following product notation: $\prod_{k=n}^{m} C_k := C_n \cdot C_{n-1} \cdot \ldots \cdot C_m$ for sequences $\{C_k\}_{k=0}^{\infty} \subset \mathbb{R}^{d\times d}$ with the convention that $\prod_{k=n}^{m} C_k = I_d$ when $n < m$.

Theorem 2.4 (Section 5.1 of [7] or Corollary 3.25 of [37]) *Assume that the sequence $\{R^A(n;h)\}_{n=0}^{\infty}$ is bounded and that each $R^A(n;h)$ is invertible and upper triangular. Let $\Sigma_{ED}^A = \cup_{i=1}^{d}[\alpha_i^A, \beta_i^A]$ denote the Sacker-Sell spectrum of $u_{n+1} = R^A(n;h)u_n$. Then for $i = 1,\ldots,d$ we have*

$$\alpha_i^A = \liminf_{0<m\to\infty}\left(\inf_{n\geq 0}\frac{1}{t_{n+m}-t_n}\ln|\prod_{k=n+m}^{n} R_{i,i}^A(k;h)|\right), \quad \beta_i^A = \limsup_{0<m\to\infty}\left(\sup_{n\geq 0}\frac{1}{t_{n+m}-t_n}\ln|\prod_{k=n+m}^{n} R_{i,i}^A(k;h)|\right).$$

□

Theorem 4.1 of [38] implies that the Sacker-Sell spectrum of (2.11) is continuous with respect to $l^{\infty}(\mathbb{N})$ perturbations of the coefficient matrix. Discrete integral separation characterizes when the Lyapunov spectrum of (2.11) is continuous.

Definition 2.5 *Consider $u_{n+1} = R^A(n;h)u_n$ where each $R^A(n;h) \in \mathbb{R}^{d\times d}$ is invertible and upper triangular, the sequence $\{R^A(n;h)\}_{n=0}^{\infty}$ is bounded, and $\inf_{n\geq 0} R_{i,i}^A(n;h) > 0$ for $i = 1,\ldots,d$. Let $p \geq 1$ and suppose there exists an $h^* > 0$ so that if $\|h\|_\infty \in (0,h^*)$ and $i < j$, then one of the two following conditions hold:*

1. *$R_{i,i}^A(n;h)$ and $R_{j,j}^A(n;h)$ are discretely integrally separated: there exists $b_{i,j} \in \mathbb{R}$ and $a_{i,j} > 0$ so that if $n > m$, then*

$$\prod_{k=n-1}^{m} R_{i,i}^A(k;h)(R_{j,j}^A(k;h))^{-1} \geq \exp\left(a_{i,j}(t_n - t_m) + b_{i,j}\right). \tag{2.12}$$

2. *$R_{i,i}^A(n;h)$ and $R_{j,j}^A(n;h)$ satisfy that there exists $K_{i,j} > 0$ such that for each $\varepsilon > 0$ there exists $M_{i,j} > 0$ so that if $n > m$, then*

$$\left|\ln\left(\prod_{k=n-1}^{m} R_{i,i}^A(k;h)(R_{j,j}^A(k;h))^{-1}\right)\right| \leq M_{i,j} + (\varepsilon + K_{i,j}\|h\|_\infty^p)(t_n - t_m). \tag{2.13}$$

Then we say that $y_{n+1} = R^A(n;h)y_n$ *and* $R^A(n;h)$ *have p-approximate discrete integral separation structures. If the first condition is satisfied for all* $i < j$*, then we say that* $R^A(n;h)$ *and* $y_{n+1} = R^A(n;h)y_n$ *are discretely integrally separated. If* (2.11) *has a corresponding upper triangular system with a p-approximate discrete integral separation structure, then we say that* (2.11) *and* $\Phi^A(n;h)$ *have p-approximate discrete integral separation structures.*

The next theorem follows from Theorem 4.1 in [42] and Proposition 8.1 and Theorem 8.3 in [4]; see also Theorem 3.13 of [21] and Theorems 5.1-2 of [22].

Theorem 2.6 *Suppose* $u_{n+1} = R^A(n;h)u_n$ *is a system with a p-approximate discrete integral separation structure with Lyapunov spectrum* $\Sigma_L^A = \cup_{i=1}^d [\eta_i^A, \mu_i^A]$*. Then there exists* $h^* > 0$ *so that if* $\|h\|_\infty \in (0, h^*)$*, then for* $i = 1, \ldots, d$ *we have:*

$$\eta_i^A = \liminf_{n\to\infty} s_i^A(n) + E_i(n;h), \quad \mu_i^A = \limsup_{n\to\infty} s_i^A(n) + F_i(n;h)$$

where $\|E_i(n;h)\|, \|F_i(n;h)\| = \mathscr{O}(\|h\|_\infty^p)$ *and* $s_i^A(n) = \frac{\sum_{k=0}^n \ln(R_{i,i}^A(k;h))}{t_n - t_0}$ *for* $i = 1, \ldots, d$ *and if* $u_{n+1} = R^A(n;h)u_n$ *is integrally separated, then* $\|E_i(n;h)\|, \|F_i(n;h)\| = 0$ *for* $i = 1, \ldots, d$*.* □

Consider the perturbed system $z_{n+1} = (\Phi^A(n;h) + F_n)z_n$ and assume that $\Phi^A(n;h)$ and $\Phi^A(n;h) + F_n$ are bounded and invertible for all $n \geq 0$. Fix an initial orthogonal $Q_0 = \overline{Q}_0 \in \mathbb{R}^{d\times d}$ and inductively construct unique QR factorizations $\Phi^A(n;h)Q_n = Q_{n+1}R^A(n;h)$ and $(\Phi^A(n;h) + F_n)\overline{Q}_n = \overline{Q}_{n+1}\overline{R}^A(n;h)$ where Q_n and $\overline{Q}_n$ are orthogonal and $R^A(n;h)$ and $\overline{R}^A(n;h)$ are upper triangular with positive diagonal entries.

Theorem 2.7 **(Theorem 7.7 in [4] and Theorem 4.1 in [42])**

Suppose $\overline{R}^A(n;h)$ *has a p-approximate discrete integral separation structure and let* $E_n := -\overline{Q}_{n+1}^T F_n \overline{Q}_n$ *and* $G := sup_{n\geq 0}\|G_n\| = sup_{n\geq 0} max\{\|E_n\|, \|F_n\|\}$*. There exist constants* $h^*, \delta, K > 0$ *so that if* $\|h\|_\infty \in (0, h^*)$ *and* $\|G\| < \delta$ *such that*

$$\tilde{Q}_{n+1}R^A(n;h) = [\overline{R}^A(n;h) + E_n]\tilde{Q}_n, \quad \|\tilde{Q}_n - I\| \leq KG, \quad n \geq 0.$$ □

3 Main Results

3.1 Statement of the main results for linear ODEs

For the remainder of this section fix a one-step method $\mathscr{M}$ with local truncation error of order $p \geq 1$ and consider a linear system (2.6) with Sacker-Sell spectrum $\Sigma_{ED} = \cup_{i=1}^d [\alpha_i, \beta_i]$ and Lyapunov spectrum $\Sigma_L = \cup_{i=1}^d [\eta_i, \mu_i]$. We make use of the following assumptions to characterize the approximation properties of these two spectra.

Assumption 3.1 *The coefficient matrix* $A(t)$ *of* (2.6) *is bounded and at least* C^{p+1}.

Assumption 3.2 *The ODE* (2.6) *satisfies Assumption 3.1 and in addition there is a corresponding upper triangular ODE*

$$\dot{y}(t) = B(t)y(t), \quad B(t) = Q(t)^T A(t)Q(t) - Q(t)^T \dot{Q}(t) \tag{3.1}$$

that has an integral separation structure defined by the estimates in Definition 2.2.

Let $x_{n+1} = \Phi^A(n;h)x_n$ denote the numerical solution by $\mathscr{M}$ of (2.6) with initial condition $x(t_0) = x_0$ using the sequence of step-sizes $h = \{h_n\}_{n=0}^{\infty}$ and let $y_{n+1} = \Phi^B(n;h)y_n$ denote the numerical solution of (3.1) using $\mathscr{M}$ with the same sequence of step-sizes and initial condition $y_0 := Q(t_0)^T x_0$. We shall always assume that $\|h\|_\infty$ is so small that $\Phi^A(n;h)$ and $\Phi^B(n;h)$ are both bounded in n and invertible for all $n \geq 0$. The matrices $\Phi^B(n;h)$ are upper triangular since $B(t)$ is upper triangular and for $j = 1,\dots,d$ each diagonal entry $\Phi^B_{j,j}(n;h)$ is the numerical approximation by $\mathscr{M}$ at time t_{n+1} of the scalar equation $\dot{y}_j(t) = B_{j,j}(t)y_j(t)$ with $y_j(t_n) = 1$ using the step-size h_n. Since $\Phi^A(n;h)$ is invertible for all $n \geq 0$ we can inductively construct unique QR factorizations of $\Phi^A(n;h)Q_n$ as $\Phi^A(n;h)Q_n = Q_{n+1}R^A(n;h)$ for all $n \geq 0$ where each Q_n is orthogonal, $Q_0 = Q(t_0)$, and $R^A(n;h)$ is upper triangular with positive diagonal entries.

For the remainder of Section 3 we denote the Lyapunov and Sacker-Sell spectra of $x_{n+1} = \Phi^A(n;h)x_n$ by $\Sigma_L^A = \cup_{i=1}^d [\eta_i^A, \mu_i^A]$ and $\Sigma_{ED}^A = \cup_{i=1}^d [\alpha_i^A, \beta_i^A]$ respectively and those of $y_{n+1} = \Phi^B(n;h)y_n$ by $\Sigma_L^B = \cup_{i=1}^d [\eta_i^B, \mu_i^B]$ and $\Sigma_{ED}^B = \cup_{i=1}^d [\alpha_i^B, \beta_i^B]$ respectively. We do not explicitly express the dependence of the spectra of these discrete systems on h. The following two theorems are proved in Section 3.3.

Theorem 3.3 *Suppose (2.6) satisfies Assumption 3.1. Given $\varepsilon > 0$, there exists $h^* > 0$ so that if $\|h\|_\infty \in (0,h^*)$, then for $i = 1,\dots,d$ the following holds:*

$$|\alpha_i^A - \alpha_i| < \varepsilon, \quad |\beta_i^A - \beta_i| < \varepsilon, \quad \alpha_i^B = \alpha_i + \mathscr{O}(\|h\|_\infty^p), \quad \beta_i^B = \beta_i + \mathscr{O}(\|h\|_\infty^p). \quad \square$$

Theorem 3.4 *Suppose (2.6) satisfies Assumption 3.2. There exists $h^* > 0$ so that if $\|h\|_\infty \in (0,h^*)$, then the following three conclusions hold:*

1. *The systems $y_{n+1} = \Phi^B(n;h)y_n$ and $u_{n+1} = R^A(n;h)u_n$ have p-approximate discrete integral separation structures and $\|R^A(n;h) - \Phi^B(n;h)\| = \mathscr{O}(\|h\|_\infty^{p+1})$.*
2. *For $i = 1,\dots,d$ we have $\alpha_i^A = \alpha_i^B + \mathscr{O}(\|h\|_\infty^p) = \alpha_i + \mathscr{O}(\|h\|_\infty^p)$ and $\beta_i^A = \beta_i^B + \mathscr{O}(\|h\|_\infty^p) = \beta_i + \mathscr{O}(\|h\|_\infty^p)$.*
3. *For $i = 1,\dots,d$ if $s_i^A(n) := \frac{\sum_{k=0}^n \ln(R_{i,i}^A(k;h))}{t_n - t_0}$ and $s_i^B(n) := \frac{\sum_{k=0}^n \ln(\Phi_{i,i}^B(k;h))}{t_n - t_0}$, then*

$$|\eta_i^A - \liminf_{n\to\infty} s_i^A(n)| = \mathscr{O}(\|h\|_\infty^p), \quad |\mu_i^A - \limsup_{n\to\infty} s_i^A(n)| = \mathscr{O}(\|h\|_\infty^p),$$

$$|\eta_i^B - \liminf_{n\to\infty} s_i^B(n)| = \mathscr{O}(\|h\|_\infty^p), \quad |\mu_i^B - \limsup_{n\to\infty} s_i^B(n)| = \mathscr{O}(\|h\|_\infty^p),$$

$$|\eta_i^A - \eta_i^B|, |\eta_i^A - \eta_i| = \mathscr{O}(\|h\|_\infty^p), \quad |\mu_i^A - \mu_i^B|, |\mu_i^A - \mu_i| = \mathscr{O}(\|h\|_\infty^p). \quad \square$$

The following corollary is immediate from the conclusions of Theorems 3.3 and 3.4.

Corollary 3.1 *If (2.6) satisfies Assumption 3.1 and $\max_{1\leq i\leq d}\beta_i^A < 0$, then there exists $h^* > 0$ so that if $\|h\|_\infty \in (0,h^*)$, then $\max_{1\leq i\leq d}\beta_i^A < 0$ and zero is a uniformly exponentially stable equilibrium of $x_{n+1} = \Phi^A(n;h)x_n$. If (2.6) satisfies Assumption 3.2 and $\max_{1\leq i\leq d}\mu_i < 0$, then there exists $h^* > 0$ so that if $\|h\|_\infty \in (0,h^*)$, then $\max_{1\leq i\leq d}\mu_i^A < 0$ and zero is an exponentially stable equilibrium of $x_{n+1} = \Phi^A(n;h)x_n$.*
$\square$

Example 3.1 Consider the ODE (1.1). If we let $x = L(t)v$, then $\dot{v} = D(t)v \equiv [D_1 + D_2(t)]v$ where the matrix $D_1 = \begin{bmatrix} \lambda_1 & \beta_0 + a_1 \\ -a_1 & \lambda_2 \end{bmatrix}$ has two real and distinct eigenvalues with real parts $\frac{1}{2}(\lambda_1+\lambda_2) \pm \frac{1}{2}\sqrt{(\lambda_1+\lambda_2)^2 - 4(\lambda_1\lambda_2 + a_1(\beta_0+a_1))} < 0$ and $D_2(t) = \begin{bmatrix} 0 & \frac{\beta_1 \cos(a_1 t)}{1+\beta_2 t^2} \\ 0 & 0 \end{bmatrix}$. Since D_1 has real and distinct eigenvalues the system $\dot{u} = D_1 u$ is integrally separated. Since $D_2(t)$ is integrable and $\dot{u} = D_1 u$ is integrally separated, it follows that $\dot{v} = [D_1 + D_2(t)]v$ is integrally separated. Then the fact that $x = L(t)z$ is a Lyapunov transformation implies that (1.1) is integrally separated.

Once again consider the solution of (1.1) by the first order implicit Euler method with constant step-size $h_0 > 0$. Since $A(t) \in C^2$ is bounded and integrally separated, Theorem 3.4 implies that there exists $h^* > 0$ so that if $h_0 \in (0, h^*)$, then the endpoints of the Lyapunov and Sacker-Sell spectrum of the discrete system (1.2) agree with those of the continuous system (1.1) to $\mathscr{O}(h_0)$ accuracy. Corollary 3.1 implies that there exists $h^{**} \in (0, h^*)$ so that if $h_0 \in (0, h^{**})$, then the Lyapunov and Sacker-Sell spectrum of (1.2) are less than zero and the numerical solution is uniformly exponentially decaying for all sufficiently small $h_0 > 0$. □

We now discuss the approximate average exponential growth/decay rates of (2.6) on a finite length interval $(t, t+\Delta t)$.

Lemma 3.1 *Assume that the ODE (2.6) satisfies Assumption 3.2. Let X be a fundamental matrix solution of (2.6) and let $X(t) = Q(t)R(t)$ be a QR factorization where $Q(t) \in \mathbb{R}^{d\times d}$ is orthogonal and $R(t) \in \mathbb{R}^{d\times d}$ is upper triangular with positive diagonal entries. The approximate average exponential growth/decay rates of $X(t)$ on the interval $(t, t+\Delta t)$ where $t \geq t_0$ and $\Delta t > 0$ are given by the following Steklov averages:*

$$s_i(t,\Delta t) = \frac{1}{\Delta t}\int_t^{t+\Delta t} B_{i,i}(\tau)d\tau, \quad i = 1,\dots,d. \tag{3.2}$$

Proof Let $t > t_0$ and $\Delta t > 0$. Since $X(t) = Q(t)R(t)$ and $Q(t)$ is orthogonal the exponential growth/decay of $X(t)$ on $(t, t+\Delta t)$ is given by the exponential growth/decay of $R(t)$ on $(t, t+\Delta t)$. We express $R(t+\Delta t) = R(t+\Delta t, t)R(t)$ where $R(\tau, t)$ is the unique and upper triangular solution of the matrix ODE IVP (recall that I_d is the $d \times d$ identity matrix):

$$\dot{\Phi} = B(\tau)\Phi, \quad \tau > t, \quad \Phi(t,t) = I_d.$$

Since (2.6) satisfies Assumption 3.2, Theorem 5.2 of [21] implies that for each $\varepsilon > 0$ there exists $K > 0$ so that if $t > t_0$ and $\Delta t > 0$, then

$$R(\tau,t) = \operatorname{diag}\left(e^{\Delta t(\varepsilon + s_1(t,\Delta t))}, \dots, e^{\Delta t(\varepsilon + s_d(t,\Delta t))}\right)(I_d + N(t+\Delta t, t))$$

where N is upper triangular with $\|I_d + N(t+\Delta t, t)\| \leq K$. Hence the approximate average exponential growth/decay rates of $X(t)$ on $(t, t+\Delta t)$ are given by the quantities $s_i(t,\Delta t)$ for $i = 1,\dots,d$. □

We can prove a result analogous to Lemma 3.1 for discrete systems (2.11) with a p-approximate integral separation structure where exponentials of Steklov averages are replaced by the products of diagonal entries of the upper triangular factor $R^A(n;h)$ in a discrete QR iteration. Theorem 3.4 implies that $R^A_{i,i}(n;h) = e^{h_n s_i(t_n,h_n)} + \mathcal{O}(\|h\|^{p+1})$ and hence the approximate average exponential growth/decay rate of fundamental matrix solutions of (2.11) are approximately (up to a term of the form $\mathcal{O}(\|h\|_\infty^{p+1})$) given by the Steklov averages (3.2). It follows that for sufficiently small step-sizes the approximate average exponential growth/decay of a numerical solution of (2.6) from t_n to t_{n+k} for $k \geq 1$ is given by the average exponential growth/decay rate on the interval $[t_n, t_{n+k}]$ of the following d real-valued test equations:

$$\dot{y}_i = B_{i,i}(t) y_i, \quad i = 1, \ldots, d.$$

This local-in-time stability argument is important for applications since for a nonlinear ODE we can not form $A(t) = Df(x(t;x_0,t_0),t)$ exactly without knowing the exact solution. However, regardless of the global error of x_n from $x(t_n;x_0,t_0)$ we can still approximately quantify the average exponential growth/decay rates of the numerical solution on the next time interval (t_n, t_{n+1}) assuming that h_n is sufficiently small.

3.2 Stability of the test problem

In this section we consider the numerical stability of a linear scalar test equation

$$\dot{z} = \lambda(t) z, \quad t > t_0 \tag{3.3}$$

where $\lambda : (t_0, \infty) \to \mathbb{C}$ is C^{p+1} and bounded with $\sup_{t>t_0} |\lambda(t)| \leq M^\lambda$ for some $M^\lambda > 0$. For full generality we consider the complex-valued case rather than the real-valued case justified in Section 3.1. The numerical solution of (3.3) by $\mathcal{M}$ using a sequence of step-sizes $h = \{h_n\}_{n=0}^\infty$ takes the form $z_{n+1} = \Phi^\lambda(n;h) z_n$ where $\Phi^\lambda(n;h) \in \mathbb{C}$. The next theorem shows that no Runge-Kutta method can preserve the asymptotic decay of every ODE of the form (3.3) with Sacker-Sell spectrum contained in $(-\infty, 0)$ without restricting the maximal step-size.

Theorem 3.5 *Let $\mathcal{M}$ be an s-stage Runge-Kutta method with Butcher tableau* $\begin{array}{c|c} c & A \\ \hline & b^T \end{array}$ *with local truncation error of order $p \geq 1$. Given any $h_0 > 0$ and any $-\alpha < 0$ we can find $\lambda : \mathbb{R} \to \mathbb{R}$ so that the equation $\dot{x} = \lambda(t) x$ with $t > t_0$ has Sacker-Sell spectrum with right endpoint given by $-\alpha$ and the numerical solution of $\dot{x}(t) = \lambda(t) x(t)$ using $\mathcal{M}$ with fixed step-size $h_0 > 0$ and initial condition $x(t_0) = x_0 \neq 0$ grows at an exponential rate.*

Proof Let $h_0 > 0$ and $-\alpha < 0$ be given and express $c = (c_1, \ldots, c_s)^T$. Let $\xi(\cdot)$ be the stability function of $\mathcal{M}$. Since $\mathcal{M}$ has local truncation error of order $p \geq 1$ there exists $\delta > 0$ so that if $r \in (0, \delta)$, then $|\xi(1+r)| > 1$. Let $D > |\alpha|$ be such that $D - \alpha \in (0, \delta)$. There exists a function $g : \mathbb{R} \to \mathbb{R}$ as smooth as desired so that $g((n + c_j)h_0) = 1$ for $n \geq 0$ and $j = 1, \ldots, s$ and $|\int_t^{t+H} g(\tau) d\tau|$ is bounded by some constant for all $t \in \mathbb{R}$ and $H > 0$ (such a function $g(t)$ can be constructed using, for example, piecewise trigonometric interpolation). Let $\lambda(t) = Dg(t) - \alpha$ and note that the right endpoint of the Sacker-Sell spectrum of $\dot{x} = \lambda(t)x$ is $-\alpha$. Since by construction λ is equal to the constant $D - \alpha$ at every Runge-Kutta stage time $t = h_0(n + c_j)$ for $n \geq 0$ and $j = 1, \ldots, s$ it follows that the numerical solution of (3.3) with the method $\mathscr{M}$ using the fixed step-size h_0 is $x_{n+1} = \xi(h_0(D-\alpha))x_n$ and $|\xi(h_0(D-\alpha))| > 1$. It follows that $|x_n| \to \infty$ as $n \to \infty$ at a rate of $|\xi(h_0(D-\alpha))|^n$. □

The geometric intuition for Theorem 3.5 is that time-dependent oscillations of $h_0\lambda(t)$ into and out of the linear stability domain of a method can trigger instabilities in the numerical solution. The following proof of Theorems 3.3 and 3.4 for scalar ODEs of the form (3.3) shows how we can control the accuracy of the Lyapunov and Sacker-Sell spectrum of the numerical solution using bounds on the local truncation error to guarantee exponential decay.

Proof (Proof of Theorems 3.3 and 3.4 in one dimension) Suppose $\lambda(t) \in \mathbb{R}$ for $t > t_0$. Because the method $\mathscr{M}$ has local truncation error of order $p \geq 1$ and $\lambda \in C^{p+1}$ is bounded there exists $h_1^* > 0$ so that if $h = \{h_n\}_{n=0}^{\infty}$ is any sequence of step-sizes with $\|h\|_\infty \in (0, h_1^*)$ and $n \geq 0$, then

$$\Phi^\lambda(n;h) = \exp\left(\int_{t_n}^{t_{n+1}} \lambda(\tau)d\tau\right) + E^\lambda(n;h) \equiv I^\lambda(n;h) + E^\lambda(n;h)$$

where $E^\lambda(n;h) = K^\lambda(n;h)h_n^{p+1}$ and $\sup_{n\geq 0}|K^\lambda(n;h)| \leq K^\lambda$ for some $K^\lambda > 0$. If $n > m \geq 0$ and $\|h\|_\infty \in (0, h_1^*)$, then

$$\prod_{k=n-1}^{m} \Phi^\lambda(k;h) = \left(\prod_{k=n-1}^{m} (1 + E^\lambda(k;h)(I^\lambda(k;h))^{-1})\right) \exp\left(\int_{t_m}^{t_n} \lambda(\tau)d\tau\right). \quad (3.4)$$

Let $h_2 \in (0, h_1^*]$ be so small that if $h = \{h_n\}_{n=0}^{\infty}$ is any sequence of step-sizes with $\|h\|_\infty \in (0, h_2^*)$, then $\sup_{n\geq 0}\|E^\lambda(n;h)(I^\lambda(n;h))^{-1}\| \leq K^\lambda \|h\|_\infty^{p+1} e^{M^\lambda \|h\|_\infty} < 1/2$. If $\|h\|_\infty \in (0, h_2^*)$, and $n > m \geq 0$, then (3.4) implies that the following two inequalities hold:

$$\begin{aligned}
&\exp\left(\textstyle\int_{t_m}^{t_n} \lambda(\tau)d\tau - 2\sum_{k=m}^{n-1} |E^\lambda(k;h)(I^\lambda(k;h))^{-1}|\right) \leq \prod_{k=n-1}^{m} \|\Phi^\lambda(k;h)\|, \\
&\textstyle\prod_{k=n-1}^{m} \|\Phi^\lambda(k;h)\| \leq \exp\left(\int_{t_m}^{t_n} \lambda(\tau)d\tau + \sum_{k=m}^{n-1} E^\lambda(k;h)(I^\lambda(k;h))^{-1}\right).
\end{aligned} \quad (3.5)$$

For any sequence of step-sizes h with $\|h\|_\infty \in (0, h_2^*)$ we have

$$|\sum_{k=m}^{n-1} E^\lambda(k;h)(I^\lambda(k;h))^{-1}| \leq \sum_{k=m}^{n-1} K^\lambda h_k^{p+1} e^{M\|h\|_\infty^p} \leq K^\lambda e^{M\|h\|^p} \|h\|_\infty^p (t_n - t_m). \quad (3.6)$$

If $\|h\|_\infty \in (0, h_2^*)$, then the conclusions of Theorem 3.4 follow from inequalities (3.5) and (3.6). The conclusion of Theorem 3.3 follows by letting $\varepsilon > 0$ be given and then setting h^* to be so small that if $\|h\|_\infty \in (0, h^*)$, then $K^\lambda e^{M\|h\|_\infty^p} \|h\|_\infty^p < \varepsilon/2$. □

Certain subsets of A-stable Runge-Kutta methods, such as AN-stable methods, have superior stability properties compared to other classes of implicit and explicit methods Runge-Kutta methods. For an s-stage AN-stable Runge-Kutta method $\mathscr{M}' = \begin{array}{c|c} c & A \\ \hline & b^T \end{array}$ where $c = (c_1, \dots, c_s)^T$, if there exists $\tau_2 \geq \tau_1 > t_0$ such that $\mathrm{Re}(\lambda(t)) \leq 0$ for $t \in [\tau_1, \tau_2]$, then $|\Phi^\lambda(n;h)| \leq 1$ whenever h_n and t_n are such that $(t_n + c_j h_n) \in [\tau_1, \tau_2]$ for $j = 1, \dots, s$. We extend this type of analysis to Runge-Kutta methods that are not AN-stable. Fix a step-size sequence $h = \{h\}_{n=0}^\infty$ and a Runge-Kutta method $\mathscr{M}'$ and for each $n \geq 0$ consider the following associated mean autonomous ODE:

$$\dot{w} = \xi_n w, \quad \xi_n := \xi(\lambda; h_n) = \frac{1}{h_n} \int_{t_n}^{t_n + h_n} \lambda(\tau) d\tau. \tag{3.7}$$

Suppose that the approximate solution of (3.3) by $\mathscr{M}'$ at time t_n is given by z_n. Then the exact solutions of (3.3) and (3.7) with the initial condition $z(t_n) = z_n$ are the same:

$$w(t_n + h_n) = z(t_n + h_n) = \exp\left(\int_{t_n}^{t_n + h_n} \lambda(\tau) d\tau \right) z_n.$$

The solutions of (3.3) and (3.7) by $\mathscr{M}'$ using the step-size h_n are then given by

$$w(t_n + h_n) \approx w_{n+1} = \Psi(h_n \xi_n) w_n, \quad z(t_n + h_n) \approx z_{n+1} = \Phi^\lambda(n;h) z_n.$$

Since the exact solutions are equal, there exists $h^* > 0$ so that if $\|h\| \in (0, h^*)$, then

$$\Phi^\lambda(n;h) = \Psi(h_n \xi_n) + \mathscr{O}(h_n^{p+1}), \quad n \geq 0. \tag{3.8}$$

Equation (3.8) implies the following theorem.

Theorem 3.6 *Let S be the linear stability region of the Runge-Kutta method $\mathscr{M}'$ and let $h_n > 0$. For each $\varepsilon \in (0,1)$ define $S'(\varepsilon) = \{z \in S : |\Psi(z)| \leq 1 - \varepsilon\}$. If $\int_{t_n}^{t_n + h_n} \lambda(\tau) d\tau \in S'(\varepsilon)$ for some $\varepsilon \in (0,1)$ and $|\Phi^\lambda(n;h) - \Psi(h_n \xi_n)| \leq \varepsilon$, then $|\Phi^\lambda(n;h)| \leq 1$.* □

We close this section by remarking that we cannot extend equation (3.8) in a straightforward way to higher-dimensional problems since for $d \geq 2$ the matrix exponential function $\exp(\int_{t_n}^t A(\tau) d\tau)$ is not in general a solution of (2.6). It is necessary to employ a time-dependent change of variables to reduce the analysis of (2.6) to a scalar test problem of the form of (3.3).

3.3 Proof of the main results for linear ODEs

Let X be a fundamental matrix solution of (2.6) and let $X(t) = Q(t)R(t)$ be a QR factorization where $Q(t)$ is orthogonal and $R(t)$ is upper triangular with positive diagonal entries. Without loss of generality we assume that $Q(t)$ is the orthogonal matrix of the corresponding upper triangular ODE (3.1). For each $n \geq 0$ let the transition matrix $X(t, t_n)$ be the unique $d \times d$ matrix solution of the following matrix ODE IVP (recall that I_d is the $d \times d$ identity matrix):

$$\begin{cases} \dot{\Psi}(t) = A(t)\Psi(t) \\ \Psi(t_n) = I_d \end{cases}, \quad t > t_n, \quad \Psi(t) \in \mathbb{R}^{d \times d}. \tag{3.9}$$

For $n \geq 0$ we factor $X(t_n)$ as $X(t_n) = X(t_n,t_{n-1})\cdot\ldots\cdot X(t_1,t_0)X(t_0)$. Similarly for each $n \geq 0$ we let $R(t,t_n) \in \mathbb{R}^{d\times d}$ be the unique solution of the following matrix ODE IVP:

$$\begin{cases} \dot{\Phi}(t) = B(t)\Phi(t) \\ \Phi(t_n) = I_d \end{cases}, \quad t > t_n, \quad \Phi(t) \in \mathbb{R}^{d\times d}.$$

For $n \geq 0$ we then factor $R(t_n)$ as $R(t_n) = R(t_n,t_{n-1})\cdot\ldots\cdot R(t_1,t_0)R(t_0)$. Notice that we have $X(t,t_n) = Q(t)R(t,t_n)Q(t_n)^T$ for $n \geq 0$. The local error equations

$$\Phi^A(n;h) = X(t_{n+1},t_n) + E^A(n;h), \quad \Phi^B(n;h) = R(t_{n+1},t_n) + E^B(n;h)$$

and the definition $\overline{F}(h;n) := -E^B(n;h) + Q(t_{n+1})^T E^A(n;h)Q(t_n)$ imply that

$$\Phi^A(n;h) = Q(t_{n+1})[\Phi^B(n;h) + \overline{F}(n;h)]Q(t_n)^T, \tag{3.10}$$

where $\|\overline{F}(n;h)\| \leq L(\|E^A(n;h)\| + \|E^B(n;h)\|)$ for some $L > 0$ since Q_n and $Q(t_n)$ are orthogonal. Since we assume that $\|h\|_\infty$ is always such that $\Phi^A(n;h)$ is invertible for all $n \geq 0$ so we can let $Q_0 := Q(t_0)$ and inductively form QR factorizations

$$\Phi^A(n;h)Q_n = Q_{n+1}R^A(n;h), \quad n \geq 0 \tag{3.11}$$

where Q_n is orthogonal and $R^A(n;h)$ is upper triangular with positive diagonal entries for all $n \geq 0$. Combining (3.10) and (3.11) results in the equation

$$R^A(n;h) = Q_{n+1}^T Q(t_{n+1})[\Phi^B(n;h) + \overline{F}(n;h)]Q(t_n)^T Q_n.$$

The Lyapunov and Sacker-Sell spectra of $v_{n+1} = [\Phi^B(n;h) + \overline{F}(n;h)]v_n$ and $x_{n+1} = \Phi^A(n;h)x_n$ coincide since $x_n = Q(t_n)v_n$ and $x_n = Q_n w_n$ are discrete Lyapunov transformations.

Proof (Proof of Theorem 3.3) By the estimates (3.5) and (3.6) in the proof of Theorems 3.3 and 3.4 in one dimension, there exists $h_1^* > 0$ so small that if h is any sequence of step-sizes with $\|h\|_\infty \in (0,h_1^*)$, then

$$\beta_i^B = \beta_i + \mathscr{O}(\|h\|_\infty^p) \text{ and } \alpha_i^B = \alpha_i + \mathscr{O}(h_{\max}^p).$$

Let $\varepsilon > 0$ be given. By continuity of the Sacker-Sell spectrum there exists $\delta > 0$ so that if $\|\overline{F}(n;h)\| < \delta$, then the endpoints of the Sacker-Sell spectrum of $v_{n+1} = [\Phi^B(n;h) + \overline{F}(n;h)]v_n$ (and hence of $x_{n+1} = \Phi^A(n;h)x_n$) satisfy

$$|\alpha_i^A - \alpha_i| < \varepsilon \text{ and } |\beta_i^A - \beta_i| < \varepsilon, \quad i = 1,\ldots,d.$$

We can always bound $\|\overline{F}(n;h)\| < \delta$ as follows. Since $\mathscr{M}$ has local truncation error of order $p \geq 1$, we can choose $h_2^* \in (0,h_1^*]$ be so small that if $\|h\|_\infty < h_2^*$, then $\|\overline{F}(n;h)\| = \mathscr{O}(h_n^{p+1}) = \mathscr{O}(\|h\|_\infty^{p+1})$. Then we can choose $h^* \in (0,h_2^*]$ so small that $\|\overline{F}(n;h)\| < \delta$. □

We assume for the remainder of this section that (2.6) satisfies Assumption 3.2. The proof of Theorem 3.4 is accomplished using several technical lemmas.

Lemma 3.2 *There exists $h^* > 0$ so that if $h = \{h_n\}_{n=0}^{\infty}$ is any sequence of step-sizes with $\|h\|_\infty \in (0, h^*)$, then the system $y_{n+1} = \Phi^B(n;h)y_n$ has a p-approximate discrete integral separation structure.*

Proof For $l = 1, \ldots, d$ the diagonal entries $\Phi^B_{l,l}(n;h)$ are such that $y^l_{n+1} = \Phi^B_{l,l}(n;h)y^l_n$ are approximations to the scalar ODE $\dot{y}_l = B_{l,l}(t)y_l$ with $y_l(t_0) = y^l_0$ using the method $\mathscr{M}$. Because $\mathscr{M}$ has local truncation of order p and $B(t)$ is bounded and C^{p+1}, there exists $h^*_1 > 0$ so that if h is such that $\|h\|_\infty < h^*_1$, $n \geq 0$, and $l = 1, \ldots, d$, then

$$\Phi^B_{l,l}(n;h) = \exp\left(\int_{t_n}^{t_{n+1}} B_{l,l}(\tau)d\tau\right) + E^B_l(n;h) \equiv I^B_l(n;h) + E^B_l(n;h).$$

where $E^B_{l,l}(n;h) = K^B_l(n;h)h_n^{p+1}$ and $\sup_{n\geq 0} K^B_l(n;h) \leq K^B_l < \infty$ for $l = 1, \ldots, d$. There exists $h^*_2 > 0$ with $h^*_2 \in (0, h^*_1]$ so that if h is such that $\|h\|_\infty \in (0, h^*_2)$, then we have $\inf_{n\geq 0} \Phi^B_{l,l}(n;h) > 0$ for $l = 1, \ldots, d$ and therefore if $n > m \geq 0$ and $i > j$, then

$$\prod_{k=n-1}^{m} \frac{\Phi^B_{i,i}(k;h)}{\Phi^B_{j,j}(k;h)} = e^{\int_{t_m}^{t_n} B_{i,i}(\tau) - B_{j,j}(\tau)d\tau} \left[\prod_{j=n-1}^{m} \frac{1 + E^B_i(k;h)(I^B_i(k;h))^{-1}}{1 + E^B_j(k;h)(I^B_j(k;h))^{-1}}\right]. \tag{3.12}$$

Note that since $\inf_{n\geq 0} \Phi^B_{l,l}(n;h) > 0$ for $l = 1, \ldots, d$ and $\|h\| \in (0, h^*_2)$ it follows that $\{\Phi^B_{l,l}(n;h)^{-1}\}_{n=0}^{\infty}$ is uniformly bounded for $l = 1, \ldots, d$. Since $B(t)$ is bounded, for $l = 1, \ldots, d$ there exists $M^B_i > 0$ so that $\sup_{t\geq t_0} |B_{l,l}(t)| \leq M^B_l$. Therefore, there exists $h^*_3 \in (0, h^*_2]$ so that if $\|h\|_\infty \in (0, h^*_3)$ and $n \geq 0$, then

$$\sup_{n\geq 0} |E^B_l(n;h)I^B_l(n;h)^{-1}| \leq K^B_l \|h\|_\infty^{p+1} e^{\|h\|_\infty M^B_l} < 1/2, \quad l = 1, \ldots, d. \tag{3.13}$$

Assumption 3.2 implies that if $i > j$, then $B_{i,i}$ and $B_{j,j}$ satisfy either (2.9) or (2.10). Let *IS* be the set of all pairs of integers (i, j) with $1 \leq i, j \leq d$ and $i > j$ so that $B_{i,i}$ and $B_{j,j}$ satisfy (2.9). If $(i, j) \in IS$, then (2.9), (3.12), (3.13), and $\|h\|_\infty \in (0, h^*_3)$ imply that if $n > m \geq 0$, then

$$\begin{aligned}\prod_{k=n-1}^{m} \frac{\Phi^B_{i,i}(k;h)}{\Phi^B_{j,j}(k;h)} &\geq \exp\left(a_{i,j}(t_n - t_m) + b_{i,j} - \sum_{k=m}^{n-1} (2K^B_i e^{\|h\|_\infty M^B_i} + K^B_j e^{\|h\|_\infty M^B_j})h_k^{p+1}\right) \\ &\geq \exp\left((a_{i,j} - (2K^B_i e^{\|h\|_\infty M^B_i} + K^B_j e^{\|h\|_\infty M^B_j})\|h\|^p_\infty)(t_n - t_m) + b_{i,j}\right).\end{aligned}$$

Let $h^* \in (0, h^*_3]$ be such that if $\|h\|_\infty \in (0, h^*)$, then

$$a_{i,j} - (2K^B_i e^{\|h\|_\infty M^B_i} + K^B_j e^{\|h\|_\infty M^B_j})\|h\|^p_\infty > 0 \tag{3.14}$$

for all $(i, j) \in IS$. It then follows that if $(i, j) \in IS$ and $\|h\|_\infty \in (0, h^*)$, then $\Phi^B_{i,i}(n;h)$ and $\Phi^B_{j,j}(n;h)$ satisfy satisfy an inequality of the form (2.12).

If $(i, j) \notin IS$ so that $B_{i,i}$ and $B_{j,j}$ satisfy (2.10), then (3.12), (3.13), and $\|h\|_\infty \in (0, h^*_3)$ imply that given $\varepsilon > 0$, there exists $M_{i,j}(\varepsilon)$ so that if $n > m \geq 0$, then

$$\begin{aligned}\prod_{k=n-1}^{m} \Phi^B_{i,i}(k;h)(\Phi^B_{j,j}(k;h))^{-1} &\leq \exp\left(M_{i,j} + \varepsilon(t_n - t_m) + \sum_{k=m}^{n-1} (K^B_i + 2K^B_j)h_k^{p+1}\right) \\ &\leq \exp\left(M_{i,j} + (\varepsilon + (K^B_i + 2K^B_j)\|h\|^p_\infty)(t_n - t_m)\right).\end{aligned}$$

Similarly, if $\|h\|_\infty \in (0,h^*)$, then

$$\prod_{k=n-1}^{m} \Phi^B_{i,i}(k;h)(\Phi^B_{j,j}(k;h))^{-1} \geq \exp\left(-M_{i,j} - (\varepsilon + (2K^B_i + K^B_j)\|h\|^p_\infty)(t_n - t_m)\right)$$

and it then follows that (2.13) is satisfied whenever $\|h\|_\infty \in (0,h^*)$. Therefore, if $\|h\|_\infty < h^*$, then $\inf_{n\geq 0}\Phi^B_{i,i}(n;h) > 0$ for $n \geq 0$ and $i = 1,\ldots,d$ and conditions (2.12) and (2.13) are satisfied. It follows that $y_{n+1} = \Phi^B(n;h)y_n$ has a p-approximate discrete integral separation structure. □

The size that $h^* > 0$ must be taken in Lemma 3.2 depends on the integral separation through the inequality (3.14). Weaker integral separation between diagonal elements of $B(t)$ (i.e. smaller values of $a_{i,j}$) require the smaller step-sizes to ensure the discrete system inherits these properties.

Lemma 3.3 *There exists $h^* > 0$ so that if $\|h\|_\infty \in (0,h^*)$, then $J(n;h) := \|R^A(n;h) - \Phi^B(n;h)\| = \mathcal{O}(\|h\|^{p+1}_\infty)$.*

Proof Using Lemma 3.2 we can find $h^*_1 > 0$ such that if $\|h\|_\infty < h^*_1$, then $\Phi^B(n;h)$ has a p-approximate discrete integral separation structure and so that $F(n;h) := -Q(t_{n+1})\overline{F}(n;h)Q(t_n)$ with $F(n;h) = \mathcal{O}(h^{p+1}_n)$. Theorem 2.7 implies that there exists an $h^* \in (0,h^*_1]$, $K > 0$, and a sequence $\{\tilde{Q}_n\}^\infty_{n=0}$ with each $\tilde{Q}_n \in \mathbb{R}^{d\times d}$ orthogonal so that if $\|h\|_\infty < h^*$, then

$$\tilde{Q}_{n+1}R^A(n;h) = (\Phi^B(n;h) + E(n;h))\tilde{Q}_n, \quad \|\tilde{Q}_n - I\| \leq KG$$

where $E(n;h) = Q^T_{n+1}F(n;h)Q_n$ and $G = \sup_{n\geq 0}\{\|F(n;h)\|, \|E(n;h)\|\} = \mathcal{O}(h^{p+1}_n)$. It follows that if $\|h\|_\infty < h^*$, then $J(n;h) = R^A(n;h) - \Phi^B(n;h) = \mathcal{O}(\|h\|^{p+1}_\infty)$. □

Lemma 3.4 *Suppose that the ODE (2.6) satisfies Assumption 3.2. Then, there exists $h^* > 0$ so that if $\|h\|_\infty \in (0,h^*)$, then $u_{n+1} = R^A(n;h)u_n$ has a p-approximate integral separation structure.*

Proof Combine Lemma 3.3 and the method used to prove Lemma 3.2. □

We now complete the proof of Theorem 3.4. Let $h^* > 0$ be so small that if $\|h\|_\infty \in (0,h^*)$, then the conclusions of Lemmas 3.2, 3.3, and 3.4 and Theorem 2.6 hold. The conclusions of Theorem 3.4 are proved by combining the conclusions of Lemmas 3.2, 3.3, 3.4 and the conclusions of Theorems 2.4 and 2.6. □

4 Nonlinear stability

In this section we consider the stability of numerical solutions of nonlinear ODE IVPs by one-step methods. The results of this section justify using the linear stability theory developed in Section 3 to characterize nonlinear stability and develop the applications in Section 5. The results of this section require restrictions on the ratio $\|h\|_\infty/h_{\min} \geq 1$ in addition to restrictions on the the maximal step-size $\|h\|_\infty$. A small enough bound $\delta_h > 0$ such that $\|h\|_\infty / h_{\min} < \delta_h$ prevents oscillations in the step-size from destabilizing an approximation to an exponentially stable solution.

Consider the ODE (2.1) and assume that $f \in C^l$ for some integer $l \geq 2$. Recall that $x(t;u,s)$ denotes the unique solution of (2.1) with initial condition $u \in \mathbb{R}^d$ at initial time $s > \tau_0$. For the remainder of this section we assume that $x(t;x_0,t_0)$ is a bounded solution of (2.1) with initial condition $x(t_0) = x_0$ and initial time $t_0 > \tau_0$ and also that the right end-point of the Sacker-Sell spectrum of $\dot{u} = Df(x(t;x_0,t_0),t)u \equiv A(t)u$ is $-\alpha < 0$ so that $x(t;x_0,t_0)$ is uniformly exponentially stable.

Fix a one-step method $\mathscr{M}$ with local truncation error of order $p \geq 1$. We assume that there exist constants $h_1^*, \delta_1, K^T > 0$ so that the local truncation error of a single step of $\mathscr{M}$ applied to solve (2.1) with any initial condition $u \in \mathbb{R}^d$ at any initial time $s \geq t_0$ with $\|u - x(s;x_0,t_0)\| < \delta_1$ and step-size $h_1 < h_1^*$ takes the form $T(u,s)h_1^{p+1}$ with $\|T(u,s)\| \leq K^T$. In addition we assume there is a constant $L^T > 0$ so that if v is also such that $\|v - x(s;x_0,t_0)\| < \delta_1$, then $\|T(u,s) - T(v,s)\| \leq L^T \|u - v\|$.

Fix any $\gamma \in (0,\alpha)$ for the remainder of this section. Theorem 3.3 implies that there exist constants $h_2^*, K^A, E^A > 0$ so that if $\|h\|_\infty < h_2^*$, then $X(t_{n+1})X(t_n)^{-1} = \Phi^A(n;h) + E_n^A h_n^{p+1}$ where $\sup_{n\geq 0}\|E_n^A\| \leq E^A$ and so that if $n > m \geq 0$, then $\|\prod_{k=n-1}^{m} \Phi^A(n;h)\| \leq K^A e^{-\gamma(t_n - t_m)}$. Let $K^M > 0$ be such that $\|X(t_{n+1})X(\tau)^{-1}\| \leq K^M$ for all $\tau \in (t_n, t_{n+1})$ and $n \geq 0$ whenever $\|h\|_\infty < h_2^*$.

Fix any $u_0 \in \mathbb{R}^d$ and $t \geq s > t_0$, define $u(t) := x(t;u_0,s)$ and $x(t) := x(t;x_0,t_0)$, and let $l(t) = \{u(t) + (1-\sigma)x(t) : \sigma \in [0,1]\}$ be the line segment from $u(t)$ to $x(t)$. Taylor expanding $f(u,t) = f(x + (u-x),t)$ at $x(t)$ implies that

$$\dot{u} = f(x,t) + A(t)(u-x) + R(u,x,t) \equiv A(t)u + b(t) + R(u,x,t)$$

where $R(u,x,t)$ satisfies the estimate

$$\|R(u,x,t)\| \leq \frac{1}{2}\|u(t) - x(t)\|^2 \cdot \sup\{\|D^2 f(\xi,t)\| : \xi \in l(t)\}$$

where $D^2 f(v,t)$ denotes the Hessian of $f(v,t)$ with respect to v. the variation of parameters formula then implies that:

$$u(t) = X(t)X^{-1}(s)u_0 + \int_s^t X(t)X(\tau)^{-1}(b(\tau) + R(u_0,s,\tau))d\tau. \tag{4.1}$$

Boundedness and uniform exponential stability of $x(t;x_0,t_0)$ and the fact that $f \in C^2$ imply that there exist constants $\tilde{\gamma} \in (0,\alpha)$ and $\tilde{K}, K^H, \delta_2 > 0$ so that if $t \geq s > t_0$, $\|u_0 - x(s;x_0,t_0)\| < \delta_2$, and $\|v_0 - x(s;x_0,t_0)\| < \delta_2$, then

$$\begin{aligned} &\sup_{t\geq t_0}\sup\{\|D^2 f(\xi,t)\| : \xi \in l(t)\} \leq 2K^H, \\ &\|x(t;u_0,s) - x(t;x_0,t_0)\| \leq \tilde{K}e^{-\tilde{\gamma}(t-s)}\|u_0 - x(s;x_0,t_0)\|. \end{aligned} \tag{4.2}$$

Note that if $\|u_0 - x(s;x_0,t_0)\| < \delta_2$, then the following inequality holds:

$$\|R(u_0,t,s)\| \leq K^H\|x(t;u_0,s) - x(t;x_0,t_0)\|^2 \leq K^H\tilde{K}^2 e^{-2\tilde{\gamma}(t-s)}\|u_0 - x(s;x_0,t_0)\|.$$

We make use of the following discrete Gronwall inequalities.

Lemma 4.1 *[Gronwall inequalities] Let* $\{a_n\}_{n=0}^{\infty}$, $\{b_n\}_{n=0}^{\infty}$, *and* $\{c_n\}_{n=0}^{\infty}$ *be non-negative sequences,* $\sigma \in (0,1)$ *and* $C > 0$.

1. *If* $a_n \le C + \sum_{i=0}^{n-1} b_i a_i$, *then* $a_n \le C \exp\left(\sum_{i=0}^{n} b_i\right)$.
2. *If* $a_0 \le c_0$ *and* $a_n \le c_n + \sum_{i=0}^{n-1} \sigma a_i$ *for* $n \ge 1$, *then* $a_n \le c_n \exp(\sigma n)$ *for* $n \ge 0$.

Proof The first conclusion follows from induction. The proof of the second conclusion can be found in [43].

Let $k := \min\{p+1, l\} \ge 2$ and define three positive constants $K_1 := (1+K^A)(E^A + K^M K^H \tilde{K}^2)$, $K_2 := (1+K^A)K^T$, and $C_1 := (E^A + K^M K^H \tilde{K}^2 + L^T)(1+K^A)$. The following theorem provides a global error bound on the approximation to $x(t;x_0,t_0)$ by $\mathscr{M}$ by restricting the maximal step-size $\|h\|_\infty$ and the ratio $\|h\|_\infty / h_{\min}$.

Theorem 4.1 *Let* $k \ge 2$ *and if* $k = 2$ *assume that* $\frac{K_1+K_2}{\gamma} < 1$. *Consider the numerical solution* $\{u_n\}_{n=0}^{\infty}$ *generated by approximating* $x(t;x_0,t_0)$ *with the method* $\mathscr{M}$ *using the initial condition* x_0 *at initial time* t_0 *and let* $x_n := x(t_n;x_0,t_0)$ *for* $n \ge 0$. *Then there exists* $D, h^* > 0$ *and* $\delta_h^* > 1$ *so that if* $\delta_h \in [1, \delta_h^*)$ *and h is any sequence of step-sizes with* $\|h\|_\infty \in (0, h^*)$ *and* $\|h\|_\infty < \delta_h h_{min}$ *, then* $sup_{n\ge 0}\|u_n - x_n\| \le D\|h\|_\infty^{k-1}$.

Proof If $k = 2$, then define $h^* := \min\{3/4, h_1^*, h_2^*, \frac{3}{4\gamma}, \delta_1, \delta_2\}$ and take $\delta_h^* = \frac{3}{4h^*}$ and if $k > 2$, then define $h^* := \min\{1, h_1^*, h_2^*, \frac{3}{2\gamma}, \delta_1, \delta_2, \frac{3\gamma}{4(K_1+K_2)}\}$ and take $\delta_h^* = \frac{3\gamma}{4(K_1+K_2)h^*}$. In either case ($k = 2$ and $k > 2$) we have that $\delta_h^* \ge 1$ is well-defined. Let $\delta_h \in (0, \delta_h^*)$ and let h be any sequence of step-sizes with $\|h\|_\infty < \delta_h h_{\min}$ and $\|h\|_\infty < h^*$. The fact that $\|u_0 - x_0\| = 0 < \|h\|_\infty < \min\{\delta_1, \delta_2, h_1^*\}$ together with (4.1) and $k \ge 2$ implies that there exists $N \ge 0$ such that $\|u_n - x_n\| < \|h\|_\infty$ and the equation

$$
\begin{aligned}
u_{n+1} - x_{n+1} &= X(t_{n+1})X(t_n)^{-1}(u_n - x_n) \\
&\quad + \int_{t_n}^{t_{n+1}} X(t_{n+1})X(\tau)^{-1} R(u_n, t_n, \tau)\, d\tau + T(u_n, t_n) h_n^{p+1}
\end{aligned}
\tag{4.3}
$$

holds for all $n \le N$. Let N_0 be the maximum of the set of all N such that $\|u_n - x_n\| < \|h\|_\infty$ holds for all $n \le N$. We show by way of contradiction that $N_0 = \infty$. Let $y_n := u_n - x_n$ and suppose that $N_0 < \infty$. Then $\{y_n\}_{n=0}^{N_0}$ satisfies a difference equation of the form $y_{n+1} = a_n y_n + b_n$ where $a_n = \Phi^A(n;h)$ and b_n is defined as the remainder of the right-hand side of (4.3). The discrete variation of parameters formula and $y_0 = 0$ imply that

$$
y_n = \sum_{i=0}^{n-1} \left[\prod_{j=n-1}^{i+1} a_j \right] b_i, \quad n = 1, \ldots, N_0 + 1 \tag{4.4}
$$

The fact that $\|y_n\| < \|h\|_\infty < \delta_2$ for $n = 0, \ldots, N_0$ means the inequalities of (4.2) hold with $t = t_{n+1}$ and $s = t_n$ for $n = 0, \ldots, N_0$ which together with $\|h\|_\infty < \min\{h_1^*, h_2^*\}$ implies that

$$
\begin{aligned}
\|b_n\| &\le E^A h_n^k \|y_n\| + K^M K^H \tilde{K}^2 \|y_n\|^2 \int_{t_n}^{t_{n+1}} e^{-2\tilde{\gamma}(t_{n+1}-\tau)} d\tau + K^T h_n^k \\
&= E^A h_n^k \|y_n\| + \frac{K^M K^H \tilde{K}^2}{2\tilde{\gamma}} \|y_n\|^2 (1 - e^{-2\tilde{\gamma} h_n}) + K^T h_n^k, \quad n = 0, \ldots, N_0.
\end{aligned}
$$

From this and the facts that $k \geq 2$, $\|y_n\| < \|h\|_\infty < 1$, and $1 - e^{-2\tilde{\gamma} h_n} \leq 2\tilde{\gamma} h_n$ for $n = 0, \ldots, N_0$ it then follows that

$$\|b_n\| \leq \left(E^A \|h\|_\infty^{k-2} + K^M K^H \tilde{K}^2 \right) \|y_n\| \|h\|_\infty^2 + K^T \|h\|_\infty^k, \quad n = 0, \ldots, N_0. \tag{4.5}$$

The choice that $h^* < h_2^*$ implies that

$$\| \prod_{j=n-1}^{i+1} a_j \| \leq (1 + K^A) e^{-\gamma(t_n - t_{i+1})}, \quad n - 1 \geq i + 1. \tag{4.6}$$

The choice that $h^* < \frac{1}{2\gamma}$ implies that

$$\sum_{i=0}^{N_0} e^{-\gamma h_{\min}(N_0 - i)} = \sum_{j=0}^{N_0} e^{-\gamma h_{\min} j} \leq \frac{1}{1 - \exp(\gamma h_{\min})} \leq \frac{4}{3\gamma h_{\min}} \tag{4.7}$$

Equations (4.4), (4.5), (4.6), $k \geq 2$, and $\|h\|_\infty < 1$ imply that for $n = 1, \ldots, N_0 + 1$:

$$\|y_n\| \leq K_2 \|h\|_\infty^k \sum_{j=0}^{n-1} e^{-\gamma(t_n - t_{i+1})} + K_1 \|h\|_\infty^2 \sum_{i=0}^{n-1} e^{-\gamma(t_n - t_{i+1})} \|y_i\|.$$

Combining this with Equation (4.7) and $\max_{n=0,\ldots,N_0} \|y_n\| < \|h\|_\infty$ then implies that the following holds for $n = 1, \ldots, N_0 + 1$:

$$\begin{aligned} \|y_n\| &< \frac{4K_2 \|h\|_\infty^{k-1} \delta_h}{3\gamma} + K_1 \|h\|_\infty^2 \sum_{i=0}^{n-1} e^{-\gamma(t_n - t_{i+1})} \|y_i\| \\ &< \frac{4K_2 \|h\|_\infty^{k-1} \delta_h}{3\gamma} + \frac{4K_1 \|h\|_\infty^2 \delta_h}{3\gamma} \leq \frac{4\delta_h}{3\gamma} \left(K_1 \|h\|_\infty^{k-2} + K_2 \|h\|_\infty \right) \|h\|_\infty \end{aligned} \tag{4.8}$$

If $k > 2$ the assumption that $\delta_h \in [1, \frac{3\gamma}{4h^*(K_1+K_2)})$ then implies that $\|y_{N_0+1}\| < \|h\|_\infty$. If $k = 2$, then $(K_1 + K_2)/\gamma < 1$ and $\delta_h < \frac{3}{4h^*}$ imply that $\|y_{N_0+1}\| < \|h\|_\infty$. Either case contradicts the maximality of N_0 and we therefore conclude that $N_0 = \infty$. Therefore the inequality (4.8) holds for all $n \geq 0$ and the first conclusion of the discrete Gronwall lemma (Lemma 4.1) implies that the following holds for $n \geq 0$:

$$\|y_n\| \leq \frac{4K_2 \|h\|_\infty^{k-1} \delta_h}{3\gamma} \exp\left(K_1 \|h\|_\infty^2 \sum_{i=0}^{n-1} e^{-\gamma(t_{n-1} - t_{i+1})} \right) \leq \frac{4K_2 \|h\|_\infty^{k-1} \delta_h e^{4K_1 \|h\|_\infty \delta_h/(3\gamma)}}{3\gamma}.$$

Therefore $\|y_n\| \leq D \|h\|_\infty^{k-1} \delta_h$ for all $n \geq 0$ where $D = 4K_2 e^{K_1/\gamma}/(3\gamma)$ if $k = 2$ and $D = 4K_2 e^{K_1/(K_1+K_2)}/(3\gamma)$ if $k > 2$. □

The following theorem shows that when the maximal step-size and the ratio $\|h\|_\infty / h_{\min}$ are properly restricted all numerical solutions with initial conditions sufficiently close to x_0 are uniformly exponentially attracted to $x(t; x_0, t_0)$.

Theorem 4.2 *Assume that $k \geq 2$ and $x_n := x(t_n; x_0, t_0)$ for $n \geq 0$. Given $\overline{\gamma} \in (0, \gamma)$ there exists $D, h^* > 0$ and $\delta_h^* > 1$ so that if $\delta_h \in [1, \delta_h^*)$ and h is any sequence of step-sizes with $\|h\|_\infty < h^*$ and $\|h\|_\infty < \delta_h h_{min}$, then there exists $\delta > 0$ so that if $\|u_0 - x_0\| < \delta$, then the numerical approximation $\{u_n\}_{n=0}^\infty$ of $x(t; u_0, t_0)$ by the method $\mathscr{M}$ with the sequence of step-sizes $h = \{h_n\}_{n=0}^\infty$ satisfies the following uniform exponential stability estimate $\|u_n - x_n\| \leq De^{-\overline{\gamma}(t_{n-1}-t_m)}\|u_m - x_m\|$ for all $n > m \geq 0$.*

Proof Let $\overline{\gamma} \in (0, \gamma)$. Let $h^* = \min\{1, \delta_1, \delta_2/2, h_1^*, h_2^*, \frac{\gamma-\overline{\gamma}}{C_1}, 1/K^A, 1/C_1\}$ and fix $\delta_h^* = \frac{\gamma-\overline{\gamma}}{C_1 h^*}$. Note that $\delta_h^* > 1$ is well-defined since $h^* < (\gamma - \overline{\gamma})/C_1$. Take any sequence of step-sizes h with $\|h\|_\infty < h^*$, $\delta_h \in [1, \delta_h^*)$, and $\|h\|_\infty < \delta_h h_{\min}$ and let $\delta > 0$ be such $\delta < \min\{\|h\|_\infty / K^A, \|h\|_\infty\}$. The fact that $\|u_0 - x_0\| < \delta < \|h\|_\infty$ implies that there exists $N \geq 0$ so that $\|u_n - x_n\| < \|h\|_\infty$ and therefore

$$u_{n+1} - x_{n+1} = X(t_{n+1})X(t_n)^{-1}(u_n - x_n) + \int_{t_n}^{t_{n+1}} X(t_{n+1})X(\tau)^{-1}R(u_n, t_n, \tau)d\tau$$
$$+ (T(u_n, t_n) - T(v_n, t_n))h_n^{p+1}. \tag{4.9}$$

holds for $n = 0, \ldots, N$. Let N_0 be the maximal N such that $\|u_n - v_n\| < \|h\|_\infty$ and suppose for contradiction that $N_0 < \infty$. Then as in the proof of Theorem 4.1 we let $y_n := u_n - v_n$ so that $\{y_n\}_{n=0}^{N_0}$ satisfies a difference equation of the form $y_{n+1} = a_n y_n + b_n$ where $a_n = \Phi^A(n; h)$ and b_n is defined as the remainder of the right-hand side of (4.9). The variation of parameters formula implies that if $0 \leq m < n \leq N_0 + 1$, then

$$y_n = \left[\prod_{j=n-1}^{m} \Phi^A(n; h)\right] y_m + \sum_{i=m}^{n} \left[\prod_{j=n-1}^{i} \Phi^A(j; h)\right] b_i$$

As in the proof of Theorem 4.1, since $\|y_n\| < \|h\|_\infty \leq \min\{1, h_1^*, h_2^*, \delta_1, \delta_2\}$ and $k \geq 2$ we obtain the following bound for $n = 0, \ldots, N_0$:

$$\|b_n\| \leq E^A \|h\|_\infty^k \|y_n\| + 2\|h\|_\infty^2 K^M K^H \tilde{K}^2 \|y_n\| + L^T \|h\|_\infty^k \|y_n\|$$
$$\leq \left(E^A + 2K^M K^H \tilde{K}^2 + L^T\right) \|h\|_\infty^2 \|y_n\| = C_1 \|h\|_\infty^2 \|y_n\|.$$

Since $k \geq 2$ it then follows that for $n = m, \ldots, N_0 + 1$ we have

$$\|y_n\| \leq K^A e^{-\gamma(t_{n-1}-t_m)} \|y_m\| + \sum_{i=m}^{n-1} (1 + K^A) e^{-\gamma(t_{n-1}-t_{i+1})} \|b_i\|$$
$$\leq K^A e^{-\gamma(t_{n-1}-t_m)} \|y_m\| + \sum_{i=m}^{n-1} C_1 e^{-\gamma(t_{n-1}-t_{i+1})} \|y_i\|$$

Since $h^* \leq \min\{1, 1/C_1, 1/K^A\}$, the second conclusion of the discrete Gronwall lemma (Lemma 4.1) implies that for $n = m, \ldots, N_0 + 1$ we have

$$\|y_n\| \leq K^A e^{-\gamma(t_{n-1}-t_m)} \|y_m\| e^{C_1 \|h\|_\infty^2 (n-m)} \leq K^A \|y_m\| e^{(-\gamma + C_1 \|h\|_\infty \delta_h)(t_{n-1}-t_m)}. \tag{4.10}$$

In the case that $m = 0$, $\|y_0\| < \delta < \|h\|_\infty / K^A$ and $C_1 \|h\|_\infty \delta_h < \gamma - \overline{\gamma}$ imply that $\|y_{N_0+1}\| < \|h\|_\infty$ which contradicts the maximality of N_0. It therefore follows that

$N_0 = \infty$ and therefore (4.10) holds for all $n > m \geq 0$. It follows that if h is any sequence of step-sizes with $\|h\|_\infty < h^*$, $\delta_h \geq 1$ is such that $\delta_h < \frac{\gamma-\overline{\gamma}}{C_1\|h\|_\infty}$, and $\delta <$ $\min\{\|h\|_\infty/K^A, \|h\|_\infty\}$, then $\|y_n\| \leq De^{-\overline{\gamma}(t_{n-1}-t_m)}\|y_m\|$ where $D = K^A$. □

Uniform exponential stability is a strong assumption to place on an IVP that excludes many interesting and important types of problems. We close this section with a discussion how to extend the ideas discussed herein to IVPs that are not uniformly exponentially stable. Suppose $\tau_0 = -\infty$ and $x(t;x_0,t_0)$ is bounded for all $t \in \mathbb{R}$ so that $A(t)$ is defined and bounded for all $t \in \mathbb{R}$. We now consider the case of full-line exponential dichotomies where the estimates of (2.4) hold for all $s,t \in \mathbb{R}$ rather than all $s,t \in (t_0,\infty)$. Suppose that $x(t;x_0,t_0)$ is a trajectory with a spectral gap at zero; the Sacker-Sell spectrum Σ_{ED} (defined by the full-line dichotomy estimates) of $\dot{u} = A(t)u$ is contained in $(-\infty,-\alpha)\cup(\alpha,\infty)$ where $\alpha > 0$. Additionally, suppose that there are p spectral intervals contained in (α,∞). Let $\dot{v} = B(t)v$ be a corresponding upper triangular system to $\dot{u} = A(t)u$ with $u = Q(t)v$ where $Q(t)$ is orthogonal. Then the under the change of variables $x = Q(t)y$ the equation (2.1) is transformed to the following:

$$\begin{bmatrix}\dot{y}_1\\ \dot{y}_2\end{bmatrix} = \begin{bmatrix}B_{1,1}(t) & B_{1,2}(t)\\ & B_{2,2}(t)\end{bmatrix}\begin{bmatrix}y_1\\ y_2\end{bmatrix} + \begin{bmatrix}R_1(y_1,y_2,t)\\ R_2(y_1,y_2,t)\end{bmatrix}, \quad B(t) = \begin{bmatrix}B_{1,1} & B_{1,2}\\ & B_{2,2}\end{bmatrix}$$

where $R = (R_1,R_2)$ is assumed to be Lipschitz in (y_1,y_2), $B_{1,1} \in \mathbb{R}^{p\times p}$, $B_{2,2}(t) \in \mathbb{R}^{(d-p)\times(d-p)}$, and $B_{1,2}(t) \in \mathbb{R}^{p\times(d-p)}$. Typically (see e.g. Section 2 of [24]), the diagonal entries of $B(t)$ will be ordered so that $\dot{v}_1 = B_{1,1}(t)v_1$ has Sacker-Sell spectrum contained in (α,∞) and $\dot{v}_2 = B_{2,2}(t)v_2$ has Sacker-Sell spectrum contained in $(-\infty,\alpha)$. If $\alpha > 0$ is large enough with respect to the Lipschitz constant of R, then by Theorem 2.1 of [3] there is a unique decoupling transformation φ so that y_2 satisfies the following differential equation that is independent of y_1:

$$\dot{y}_2 = B_{2,2}(t)y_2 + R_2(\varphi(y_2,t),y_2,t) \tag{4.11}$$

Solutions of (4.11) are uniformly exponentially stable and satisfy the hypotheses of Theorems 4.1 and 4.2. Therefore, if $\|h\|_\infty$ and $\delta_h - 1 \geq 0$ are sufficiently small, then the numerical error on the stable manifold defined by φ remains small and decays uniformly and exponentially as $t \to \infty$.

5 Applications

In this section we apply the theoretical results from Sections 3 and 4 to develop a time-dependent stiffness indicator and a one-step method that switches between implicit and explicit Runge-Kutta methods based on time-dependent stiffness. We remark that by stiffness we mean parabolic stiffness related to a strongly attractive mode rather than hyperbolic stiffness arising in highly oscillatory problems. We denote Runge-Kutta methods by RK(ν-p-$\hat{p}$) where RK is an identifying string, ν is the number of stages, p is the order of the method, and $\hat{p}$ is the order of the embedded method. The following methods are used: the third order Bogacki-Shampine method BS(4-3-2) (Equation 2.6 of [6]), ESDIRK(4-3-2) (second table on page 175 of [13]), SDIRK(3-3-2) (Equation 5.4 of [34]), and SDIRK(4-3-2) (Equation 16 in [12]).

All the experiments in this section were conducted using a solver *odeqr* implemented in MATLAB. This solver forms an approximate solution using a Runge-Kutta method with the capability of switching between different methods at each step. In *odeqr* the step-size is either constant or adaptive where an initial step-size guess is reduced by increments of 25% until a tolerance is satisfied. For an implicit method *odeqr* solves the nonlinear stage equations using Newton's method with an option for using exact and inexact Jacobians using the previous solution step as initial guess and an error tolerance of 10^{-12}.

5.1 Test ODEs

In this section we discuss the three ODEs used in our experiments in Sections 5.2 and 5.3. The first ODE we consider is Equation (1.1) with $\lambda_1 = 0.1$, $\lambda_2 = -0.2$, $\beta_1 = 10^3$, $\beta_2 = 10^{-4}$, $\beta_0 = (\frac{1}{4}(\lambda_1+\lambda_2)^2 - \lambda_1\lambda_2)/a_1 - a_1 - 10^{-4}$, $a_1 = a_2 = 2\pi$, and initial condition $x(0) = (1,-1)^T$.

The second equation we consider is the forced Van der Pol equation [41] expressed as a first order ODE in two dimensional phase space:

$$\begin{cases} \dot{x_1} = \mu(1-x_1^2)x_2 + x_1 - A\sin(\omega t) \\ \dot{x_2} = x_1 \end{cases} \tag{5.1}$$

We use the initial condition $(x_1(0), x_2(0))^T = (0,2)^T$, $\mu = 100$, and $\omega = A = 1$. For our third example we first consider (see [2, 28, 31]) the one-dimensional Fitzhugh-Nagumo partial differential equation (PDE):

$$\begin{cases} u_t = \phi(u) - v + \alpha\frac{\partial^2 u}{\partial x^2} \\ v_t = \varepsilon(u - \delta v) \end{cases}, \quad u = u(x,t), v = v(x,t) \in \mathbb{R}, \quad t > 0, x \in (0,1) \tag{5.2}$$

with Neumann-type boundary conditions $u_x(0,t) = 0 = u_x(1,t)$ and $v_x(0,t) = v_x(1,t)$ and with ϕ given by $\phi(r) = -2r^3 + 6r$ and $\delta, \alpha, \varepsilon > 0$. We construct a system of ODEs by taking a uniform spatial discretization of (5.2) with $x_j = j/J \equiv j\Delta x$ for $j = 0,\dots,J$ and the following finite difference approximation to $\Delta u(x_j,t)$ which takes into account the boundary conditions:

$$\Delta u(x_j,t) \approx D(u_j) := \begin{cases} (u_1 - u_0)/(\Delta x)^2, & j = 0 \\ (u_{J-1} - u_J)/(\Delta x)^2, & j = J \\ (u_{j+1} - u_{j-1} - 2u_j)/(\Delta x)^2, & \text{otherwise} \end{cases}$$

where $u_j(t) \approx u(x_j,t)$ for $j = 0,\dots,J$ and $\Delta x = 1/J$. This leads to our third ODE which is the following $(2J+2)$-dimensional Fitzhugh-Nagumo system:

$$\begin{cases} \dot{u}_j = \phi(u_i) - v_i + \alpha D(u_j) \\ \dot{v}_j = \varepsilon(u_j - \delta v_j) \end{cases}, \quad j = 0,\dots,J \tag{5.3}$$

For parameter values we take $\varepsilon = 0.1$, $\alpha = 0.3$, $\delta = 0.01$, and $J = 14$ and for the initial condition we use $u_j(0) = \sin(0.5\pi j\Delta x)$ and $v_j(0) = \cos(0.5\pi j\Delta x)$ for $j = 0,\dots,J$.

5.2 Nonautonomous stiffness detection

In this section we develop a method for stiffness detection based on approximating Steklov averages as defined in Equation (3.2). Assume that (2.6) satisfies Assumption 3.2. The conclusion of Lemma 3.1 implies that the Steklov averages (3.2) of a corresponding upper triangular ODE measure average exponential growth/decay rates of solutions of (2.6) on the interval $(t, t+\Delta t)$. For a randomly chosen orthogonal $Q(t_0) = Q_0 \in \mathbb{R}^{d\times d}$ the Steklov averages of the corresponding upper triangular ODE $\dot{y} = B(t)y$ where $B = Q^T AQ - Q^T \dot{Q}$ tend to order themselves so that $s_1(t,\Delta t)$ corresponds to the right-most spectral interval and $s_d(t,\Delta t)$ corresponds to the left-most spectral interval (see e.g. Section 2 of [24]). This motivates using the following as a stiffness indicator:

$$S(t,\Delta t) = s_1(t,\Delta t) - s_d(t,\Delta t).$$

If $S(t,\Delta t)$ is large in absolute value, then we expect that the problem is stiff and if $S(t,\Delta t)$ is near zero, then we expect that the problem is nonstiff. We remark that in general $s_1(t,\Delta t) > s_d(t,\Delta t)$ holds on average, but does not hold point-wise, since for sufficiently large Δt the quantities $s_1(t,\Delta t)$ and $s_d(t,\Delta t)$ become approximations to respectively the right and left end-points of the Lyapunov and Sacker-Sell spectra.

We now discuss how to approximate $S(t,\Delta t)$. Consider the numerical solution $x_{n+1} = \Phi^A(n;h)x_n$ of (2.6) using a one-step method $\mathscr{M}$ with local truncation error of order $p \geq 1$. We first approximate $s_1(t_n,h_n)$ as follows. Given an initial $q_0 \in \mathbb{R}^d$ with $\|q_0\|_2 = 1$ ($\|\cdot\|_2$ is the Euclidean 2-norm) we inductively form $v_n := \Phi^A(n;h)q_n$ and $R^A_{1,1}(n;h) := \|v_n\|_2$ followed by normalization: $q_{n+1} := v_n/\|v_n\|_2$. We approximate $s_1(t_n,h_n)$ by $\xi_1(n) := \ln(R^A_{1,1}(n;h))/h_n$ which is justified since Theorem 3.4 implies that $s_1(t_n,h_n) = \xi_1(n) + \mathscr{O}(\|h\|^p_\infty)$ for sufficiently small $\|h\|_\infty$. We approximate $s_d(t_n,h_n)$ by applying the same method used to approximate $s_1(t_n,h_n)$ to the opposite adjoint equation $\dot{x} = -A(t)^T x$ to obtain $\xi_d(n) \approx s_d(t_n,h_n)$. This is justified since the right end-points of the Lyapunov and Sacker-Sell spectra of the opposite adjoint equation are the left end-points of the Lyapunov and Sacker-Sell spectra of (2.6).

Our approximation of $S(t,\Delta t)$ along a sequence of time-steps $\{h_n\}^\infty_{n=0}$ using window length $w \geq 0$ and $n \geq 0$ is defined as as

$$SI(n,w) = \frac{1}{t_{n+w+1} - t_{n-w}} \sum_{k=0}^{2w} (\xi_1(n-w+k) - \xi_d(n-w+k))/h_{n-w+k}.$$

For IVPs of nonlinear ODEs we compute $SI(n,w)$ by approximating the coefficient matrix $A(t) := Df(x(t;x_0,t_0),t)$ which is then used to form an approximate $\Phi^A(n;h)$. Approximating $\Phi^A(n;h)$ to high order may require approximating values of $A(t)$ for $t \in (t_n,t_{n+1})$. We use a piecewise cubic Hermite interpolating polynomial to obtain an $\mathscr{O}(\|h\|^4_\infty)$ order approximation to $A(t)$ for $t \in (t_n,t_{n+1})$ making use of the approximate solutions x_n and x_{n+1} and their approximate derivatives $f(x_n,t_n)$ and $f(x_{n+1},t_{n+1})$. This is sufficient for us to obtain order p approximations to $\Phi^A(n;h)$ for $p \leq 3$, a constraint satisfied by all the methods used in our examples. In general higher-order piecewise Hermite interpolants would be needed for higher order approximations of $\Phi^A(n;h)$. If the method is explicit we approximate $\Phi^A(n;h)$ by applying the method to compute a single step of the numerical solution of $\dot{x} = A(t)x$ starting from the identity using the cubic Hermite interpolating polynomial to approximate the necessary values of $A(t)$ for $t \in (t_n, t_{n+1})$. For an implicit method we avoid solving a linear system of equations to approximate $\Phi^A(n;h)$ by instead forming $\overline{\Phi}^A(n;h) = \Phi^A(n;h) + \mathcal{O}(\|h\|_\infty^p)$ where $\overline{\Phi}^A(n;h)$ is formed using an explicit method with truncation error of the same order.

In addition to our Steklov average based method we implement the stiffness indicator, denoted as $\sigma[A(t)]$, that was introduced in Definition 4.1 of [11] that is formulated in terms of the logarithmic norm of the Hermitian part of $A(t)$: $\mathrm{He}[A(t)] := (A(t)+A(t)^T)/2$. To simplify the computation of $\sigma[A(t)]$ we assume that we are using the Euclidean 2-norm. As noted in [11] this implies that $\sigma[A(t)]$ equals the smallest eigenvalue of $\mathrm{He}[A(t)]$ subtracted from the largest eigenvalue of $\mathrm{He}[A(t)]$.

In general we cannot expect any relation between $SI(n,w)$ and $\sigma[A(t_n))]$ as exemplified in Figure 5.1. However, we can characterize when these two indicators should be close to equal. Assume $w = 0$ and note that for any bounded and continuous $A(t)$ we have $X(t_{n+1},t_n) = \exp\left(\int_{t_n}^{t_{n+1}} A(\tau)d\tau\right) + \mathcal{O}(\|h\|_\infty^2)$ for all sufficiently small $\|h\|_\infty$ where $X(t;t_n)$ is the solution of (3.9). Hence, if $X(t_{n+1};t_n)$ is well-conditioned for eigenvalue computations (such as when $A(t_n)$ is normal and h_n is small), then the logarithms of the real parts of the eigenvalues of $\Phi^A(n;h)$ divided by h_n should be approximately equal to the average of the eigenvalues of $\mathrm{He}[A(t)]$ for $t \in (t_n, t_{n+1})$. Forming $\xi_1(n)$ and $\xi_d(n)$ is equivalent to performing one step of power iteration to approximate the real parts of the eigenvalues of $\Phi^A(n;h)$ and the associated opposite adjoint coefficient matrix followed by taking logarithms and division by h_n. If the largest (in terms of absolute value) eigenvalue of $He[A(t)]$ is significantly larger than the next, then power iteration converges rapidly implying that a single step of power iteration applied to $\mathrm{He}[A(t_n)]$ should be approximately the logarithm of a single step of power iteration applied to $\Phi^A(n;h)$ divided by h_n. The same statement holds for the adjoint coefficient matrix and the smallest eigenvalue of $He[A(t)]$. It follows that $SI(n,w)$ and $\sigma[A(t_n))]$ should be close when $\|h\|_\infty$ is small, $w \approx 0$, $A(t_n)A(t_n)^T - A(t_n)^T A(t_n) \approx 0$, and the largest eigenvalues of $A(t_n)$ and $-A(t_n)^T$ dominate over the next largest.

We now highlight the advantages of computing $SI(n,w)$ over $\sigma[A(t_n)]$. We first note that approximating $\sigma[A(t_n)]$ is norm dependent (see Section 4 of [11]) while $SI(n,w)$ is not. Accurately approximating $SI(n,w)$ depends on integral separation which is expected to be strong in a stiff IVP and does not require that $A(t_n)$ or $\mathrm{He}[A(t_n)]$ be normal or well-conditioned for eigenvalue computations. Forming the quantity $SI(n,w)$ is generally less expensive than $\sigma[A(t_n)]$ since forming $SI(n,w)$ essentially requires only a single step of power iteration applied to $\Phi^A(n;h)$ and the associated adjoint coefficient matrix followed by taking logarithms and a linear combination of $2w$ terms, whereas forming $\sigma[A(t_n)]$ requires at least one step of power iteration or some other method for approximating eigenvalues. This cost advantage is important in the next section where fast and accurate approximations to $\xi_1(n)$ and $\xi_d(n)$ are needed at each step.

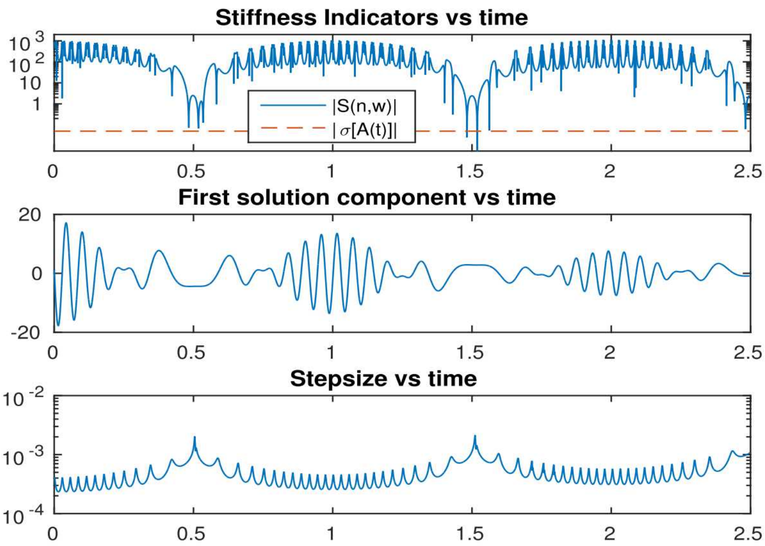

Fig. 5.1 Plots of approximate $|SI(n,w)|$ with $w = 2$ and $|\sigma[A(t_n)]|$, first solution component, and step-size versus time for the numerical solution of the 2D linear ODE (1.1) using the parameters specified in Section 5.1 solved with BS(4-3-2) using a relative and absolute error tolerance of 10^{-6}.

We compare the performance of $SI(n,w)$ with $\sigma[A(t_n)]$ with the linear ODE (1.1) and the forced Van der Pol equation in Figures 5.1 and 5.2. Figure 5.2 shows that $SI(n,w)$ and $\sigma[A(t_n)]$ produce qualitatively similar results when applied to the Van der Pol equation. However, as evidenced in Figure 5.1, our approximation to $SI(n,w)$ is more sensitive to changes in the step-size even over intervals where the solution is nonstiff. The 2D linear ODE (1.1) provides a clear example where the performance of $SI(n,w)$ is superior to that of $\sigma[A(t_n)]$, with $SI(n,w)$ detecting intervals over which the solver takes smaller and larger time-steps where $A(t)$ is respectively more or less non-normal, while $\sigma[A(t_n)]$ is approximately constant at all time-steps. The values $|SI(n,w)|$ and $|\sigma[A(t_n)]|$ are plotted since it is absolute values rather than sign that indicate stiffness.

5.3 QR implicit-explicit Runge-Kutta methods

Consider an explicit Runge-Kutta method RKex(ν-p-$\hat{p}$) and an implicit Runge-Kutta method RKim($\hat{\nu}$-p-$\hat{p}$). We construct a one-step method with local truncation error of order p, denoted as RKex(ν-p-$\hat{p}$)- RKim($\hat{\nu}$-p-$\hat{p}$)), that switches between using the implicit and explicit Runge-Kutta methods as follows. At time-step t_n we form $\xi_1(n)$ and $\xi_d(n)$ as described in Section 5.2. If $\xi_d(n)$ is too small and negative or if $\xi_1(n)$ is too large and positive, then we use the implicit method, otherwise we use the explicit method. More precisely we use the explicit method if $\xi_d(n) \geq d_2/H_0$ and $\xi_1(n) \leq d_1/H_0$ where d_1 and d_2 are chosen according to the right and left endpoints of the

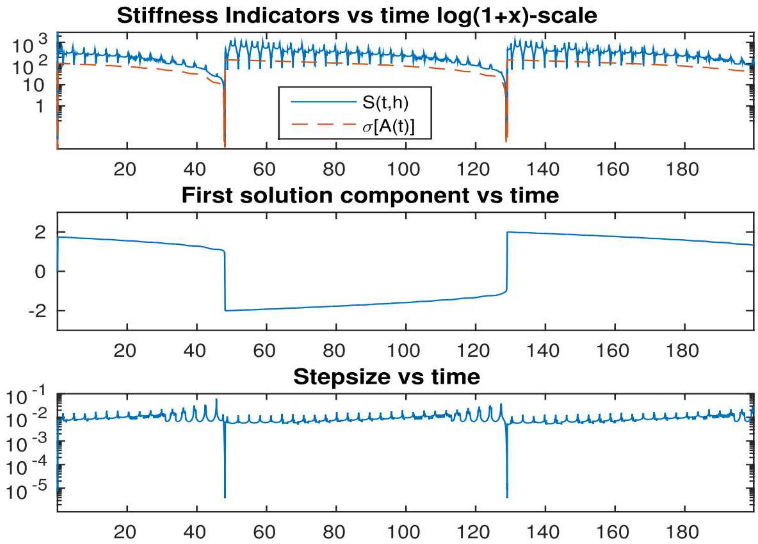

Fig. 5.2 Plots of approximate $|SI(n,w)|$ with $w = 10$ and $|\sigma[A(t_n)]|$, first solution component, and step-size versus time for the numerical solution of the Van der Pol ODE (5.1) solved with BS(4-3-2) using a relative and absolute error tolerance of 10^{-6}.

real parts of the linear stability regions of RKex(ν-p-$\hat{p}$) and RKim($\hat{\nu}$-p-$\hat{p}$) and the quantity H_0 is a parameter specifying the minimum allowable step-size restriction due to time-dependent stability that will be tolerated. We refer to such implicit-explicit switching methods as QR-IMEX-RK methods and implement them with *odeqr*.

We approximate the parameter H_0 as follows. Pick an interval over which the approximate solution is non-stiff. Over this interval compute the approximate mean step-size h_{mean} and set $H_0 = h_{\text{mean}}\alpha$ where $\alpha > 0$ is a factor quantifying the tolerance for the stability step-size restriction relative to the mean non-stiff step-size.

We now discuss the results (Table 5.1) of various QR-IMEX-RK methods applied to solve an IVP of the discretized Fitzhugh-Nagumo PDE (5.3). The stiffness increases as the error tolerance decreases leading to proportionally more uses of the implicit method by the QR-IMEX-RK methods. The results in Table 5.1 show that at the lowest tolerance TOL $= 10^{-5}$ the explicit method Mfhn1 (see the caption of Table 5.1 for descriptions of the methods) and the QR-IMEX-RK methods Mfhn2, Mfhn3, and Mfhn4 have about the same mean step-size and few implicit steps are taken by the QR-IMEX-RK methods. When tighter tolerances are used (TOL $= 10^{-6}, 10^{-7}, 10^{-8}$) the problem is stiffer and the QR-IMEX-RK solvers Mfhn2, Mfhn3, and Mfhn4 are able to take larger time-steps on average than the explicit method Mfhn1 at a cost of using more right-hand-side calls when TOL $= 10^{-5}, 10^{-6}, 10^{-7}$, fewer right-hand-side calls when TOL $= 10^{-8}$, and more Jacobian calls at all tested tolerances than Mfhn1. Notice that although Mfhn2, Mfhn3, and Mfhn4 are able to take larger step-sizes on average than Mfhn1 the additional implicit time-steps cost more in terms of right-hand-side and Jacobian evaluations, linear solves, and the overhead associated with forming $\xi_1(n)$ and $\xi_d(n)$ at each time-step.

Table 5.1 Table of results for experiments on the spatially discretized Fitzhugh-Nagumo PDE (5.3) solved on the time interval $[0,100]$ using $J = 14$. TOL is the absolute and relative error tolerance (always taken to be equal), nexp and nimp are the number of explicit and implicit steps taken, Feval is the number of evaluations by the ODE right-hand-side function $f(x,t)$, Jaceval is the number of evaluations of the Jacobian $A(t)$, Lsol is the number of linear solves, and NA is short for not applicable. The methods are the explicit method Mfhn1 = BS(4-3-2) and the QR-IMEX-RK methods Mfhn2 = BS(4-3-2)-ESDIRK(4-3-2), Mfhn3 = BS(4-3-2)-SDIRK(4-3-2) , and Mfhn4 = BS(4-3-2)-SDIRK(3-3-2). The QR-IMEX-RK parameters were $d_2 = -3.5$ and $d_1 = 10.0$ and H_0 was computed by taking the approximate mean step-size of Mfhn2 (around $1E-2$ for all tolerances) on the interval $[5,20]$ and using $\alpha = 1.5$. Jacobians were formed exactly and the initial step-size was $h_0 = 0.05$.

Method	TOL	h_{mean}	nexp	nimp	Feval	Jaceval	Lsol
Mfhn1	1E-5	1.164E-2	8595	NA	40845	NA	NA
Mfhn2	1E-5	1.186E-2	8333	129	54283	36860	328
Mfhn3	1E-5	1.183E-2	8317	124	50330	36796	309
Mfhn4	1E-5	1.182E-2	8595	143	49357	36507	373
Mfhn1	1E-6	9.501E-3	10525	NA	47893	NA	NA
Mfhn2	1E-6	1.045E-2	9247	323	61611	45476	738
Mfhn3	1E-6	1.043E-2	9304	283	62108	44832	669
Mfhn4	1E-6	1.037E-2	9237	399	61228	44817	846
Mfhn1	1E-7	5.804E-3	12834	NA	77929	NA	NA
Mfhn2	1E-7	7.439E-3	12846	609	81060	66384	1272
Mfhn3	1E-7	7.465E-3	12887	550	80773	65024	1155
Mfhn4	1E-7	7.271E-3	17229	867	80957	68339	1787
Mfhn1	1E-8	2.348E-3	42587	NA	170449	NA	NA
Mfhn2	1E-8	3.319E-3	28902	1228	161587	145536	2513
Mfhn3	1E-8	3.296E-3	29324	1018	163283	142216	2097
Mfhn4	1E-8	3.248E-3	28898	1891	163871	151845	3836

6 Afterword

We have used QR approximation theory for Lyapunov and Sacker-Sell spectra to develop a time-dependent stability theory for one-step methods approximating time-dependent solutions to nonlinear and nonautonomous ODE IVPs. This theory was used to justify characterizing the stability of a one-step method solving an ODE IVP with real-valued, scalar, nonautonomous linear test equations. In the companion paper [40] we use invariant manifold theory for nonautonomous difference equations to prove the existence of an underlying one-step method for general linear methods solving time-dependent ODE IVPs. This is then used to extend our analysis of one-step methods to general linear methods. It should also be possible to extend the theory developed in this paper to infinite dimensional IVPs (using the infinite dimensional QR approximation theory developed in [4]) arising from PDEs where the step-size restriction will become a time-dependent CFL condition. By using $Q(t)$ is should also be possible to use our techniques to measure oscillatory or hyperbolic stiffness in addition to parabolic stiffness.

Detecting, quantifying, and understanding stiffness has the been a major research focus of the time discretization community for the past 60 years. The methods we have developed in this work can be advantageous for problems with e,g. non-normal Jacobians where standard stiffness detection techniques, such as those using logarithmic norms or time-dependent eigenvalues, can potentially fail. Our techniques are justifiable in terms of Lyapunov and Sacker-Sell spectral theory and at a low computational cost produce qualitatively the same information in situations where existing methods are effective and meaningful information where existing methods are ineffective. Additionally, the QR and Steklov average based approach can be used to estimate Lyapunov exponents and Sacker-Sell spectral end-points which are useful in characterizing the dynamics of the differential equation whose IVPs are being approximated.

References

1. Adrianova, L. *Introduction to Linear Systems of Differential Equations*, volume 146. AMS, Providence, R.I., 1995.
2. Arimoto, S. and Nagumo, J. and Yoshizawa, S. An active pulse transmission line simulating nerve axon. *Proc. Inst. Radio Eng.*, 50:2060–2070, 1964.
3. Aulbach, B. and Wanner, T. Invariant foliations for Carathéodory type differential equations in Banach spaces. In A. Martynyuk, editor, *Advances in Stability Theory at the End of the 20th Century, Stability Control Theory Methods Appl. 13*. Taylor and Francis Inc, New York, NY, 1999.
4. Badawy, M. and Van Vleck, E.S. Perturbation theory for the approximation of stability spectra by QR methods for sequences of linear operators on a Hilbert space. *Linear Algebra Appl.*, 437(1):37–59, 2012.
5. Beyn, W.-J. On invariant close curves for one-step methods. *Numer. Math.*, 51: 103–122, 1987.
6. Bogacki, P. and Shampine, L. A 3(2) pair of Runge-Kutta formulas. *Appl. Math. Lett.*, 2(4):321–325, 1989.
7. Breda, D. and Van Vleck, E.S. Approximating Lyapunov exponents and Sacker-Sell spectrum for retarded functional differential equations. *Numer. Math.*, 126 (2):225–257, 2014.
8. Burrage, K. and Butcher, J.C. Stability criteria for implicit Runge-Kutta methods. *SIAM J. Numer. Anal.*, 16(1):46–57, 1979.
9. Butcher, J.C. A stability property of implicit Runge-Kutta methods. *BIT*, pages vol. 27, 358–361., 1975.
10. Butcher, J.C. The equivalence of algebraic stability and AN-stability. *BIT*, 27 (2):510–533, 1987.
11. Calvo, M., Jay, L., and Söderlind, G. Stiffness 1952-2012: Sixty years in search of a definition. *BIT*, 55(2):531–558, 2015.
12. Cameron, F., Palmroth, M., and Piché, R. Quasi stage order conditions for SDIRK methods. *Appl. Numer. Math.*, 42(1-3):61–75, 2002.
13. Carpenter, M.H. and Kennedy, C.A. Additive Runge-Kutta schemes for convection-diffusion-reaction equations. *Appl. Numer. Math.*, 44:139–181, 2003.

14. Coppel, W.A. *Lecture Notes in Mathematics # 629: Dichotomies in Stability Theory*, volume 629. Springer-Verlag, Berlin, 1978.
15. Dahlquist, G. Convergence and stability in the numerical integration of ordinary differential equations. *Math. Scan.*, 4:33–53, 1956.
16. Dahlquist, G. Stability and error bounds in the numerical integration of ordinary differential equations. *Trans. Royal Inst. Technol., Stockholm, Sweden*, Nr. 130: 87, 1959.
17. Dahlquist, G. A special stability problem for linear multistep methods. *BIT*, 3: 27–43, 1963.
18. Dieci, L. and Van Vleck, E.S. Unitary integrators and applications to continuous orthonormalization techniques. *SIAM J. Numer. Anal.*, 310(1):261–281, 1994.
19. Dieci, L. and Van Vleck, E.S. Lyapunov spectral intervals: Theory and computation. *SIAM J. Numer. Anal.*, 40(2):516–542, 2003.
20. Dieci, L. and Van Vleck, E.S. Lyapunov and other spectra: A survey. *Collected Lectures on the Preservation of Stability under Discretization, A Volume Published by SIAM*, pages 197–218, 2002.
21. Dieci, L. and Van Vleck, E.S. Lyapunov and Sacker-Sell spectral intervals. *J. Dynam. Differential Equations*, 19(2):265–293, 2007.
22. Dieci, L. and Van Vleck, E.S. On the error in computing Lyapunov exponents by QR methods. *Numer. Math.*, 101(4):619–642, 2005.
23. Dieci, L. and Van Vleck, E.S. Computation of orthonormal factors for fundamental matrix solutions. *Numer. Math.*, 83(4):599–620, 1999.
24. Dieci, L. and Van Vleck, E.S. Computation of a few Lyapunov exponents for continuous and discrete dynamical systems. *Appl. Numer. Math.*, 17(3):275–291, 1995.
25. Dieci, L., Russell, R., and Van Vleck E.S. On the computation of Lyapunov exponents for continuous dynamical systems. *SIAM J. Numer. Anal.*, 34(1):402–423, 1997.
26. Dieci, L., Elia, C., and Van Vleck E.S. Detecting exponential dichotomy on the real line: SVD and QR algorithms. *BIT*, 248:555–579, 2011.
27. Eirola, T. Invariant curves of one-step methods. *BIT*, 28(1):113–122, 1988.
28. Fitzhugh, R. Impulses and physiological states in theoretical models of nerve membrane. *Biophys. J.*, 1(6):445–466, 1961.
29. Kloeden, P. and Lorenz, J. Stable attracting sets in dynamical systems and in their one-step discretizations. *SIAM J. Numer. Anal.*, 23(5):986–995, 1986.
30. Kreiss, H.-O. Difference methods for stiff ordinary differential equations. *SIAM J. Numer. Anal.*, 15(1):21–58, 1978.
31. Krupa, M., Sandstede, B., and Szmolyan, P. Fast and slow waves in the Fitzhugh-Nagumo equation. *J. Differential Equations*, 133(1):49–97, 1997.
32. Leonov, G.A. and Kuznetsov, N.V. Time-varying linearization and the Perron effects. *Internat. J. Bifur. Chaos Appl. Sci. Engrg.*, 37(4):1079–1107, 2007.
33. Lyapunov, A. Problém géneral de la stabilité du mouvement. *Internat. J. Control*, 53(3):531–773, 1992.
34. Nørsett, S.P. and Thomsen, P.G. Embedded SDIRK-methods of basic order three. *BIT*, 24(4):634–646, 1984.
35. Palmer, K. The structurally stable systems on the half-line are those with expo-

nential dichotomy. *J. Differential Equations*, 33(1):16–25, 1979.
36. Perron, O. Die stabilitätsfrage bei Differentialgleichungen. *Math. Z.*, 32(1):703–728, 1930.
37. Pötzsche, C. Fine structure of the dichotomy spectrum. *Integral Equations and Operator Theory*, 73(1):107–151, 2012.
38. Pötzsche, C. Dichotomy spectra of triangular equations. *Discrete Contin. Dyn. Syst. Ser. A*, 36(1):423–450, 2013.
39. Sacker, R. and Sell, G. A spectral theory for linear differential systems. *J. Differential Equations.*, 27(3):320–358, 1978.
40. Steyer, A. and Van Vleck, E.S. Underlying one-step methods and nonautonomous stability of general linear methods. *Discrete Contin. Dyn. Syst. Ser. B*, To appear in print, 2017.
41. Van der Pol, B. A theory of the amplitude of free and forced triode vibration. *Radio Review*, 1:701–710, 1920.
42. Van Vleck, E.S. On the error in the product QR decomposition. *SIAM J. Matrix Anal. Appl.*, 31(4):1775–1791, 2010.
43. Zenisek, A. *Nonlinear Elliptic and Evolution Problems and their Finite Element Approximations*. Academic Press, 1990.