UNIVERSITÀ
DEGLI STUDI
DI UDINE
hic sunt futura

www.uniud.it

CORSO DI DOTTORATO DI RICERCA IN

INFORMATICA E SCIENZE MATEMATICHE E FISICHE

CICLO XXXII

COLLOCATION METHODS FOR COMPLEX DELAY MODELS OF STRUCTURED POPULATIONS

DOTTORANDO
ALESSIA ANDÒ

SUPERVISORE
DIMITRI BREDA

2020

Alessia Andò

Collocation methods for complex delay models of structured populations

COLOPHON

This document was written with LaTeX using Lorenzo Pantieri's ArsClassica, a reworking of the ClassicThesis style designed by André Miede, inspired to the masterpiece *The Elements of Typographic Style* by Robert Bringhurst.

CONTACT

☞ ando.alessia@spes.uniud.it

CODES

☞ http://cdlab.uniud.it/software

ABSTRACT

This thesis presents the main results obtained during the author's PhD studies, devoted to numerical methods for investigating the behavior of dynamical systems generated by certain classes of delay equations, that is, functional equations where the present value of a function, or that of its derivative, is defined in terms of its past values.

Such dynamical systems are infinite-dimensional, and that represents the main reason why, when it comes to actually compute their solutions or other relevant quantities, it is impossible to attain any results analytically, and thus necessary to resort to some numerical tools.

More specifically, the equations of interest are the Delay Differential Equations (DDEs), the Renewal Equations (REs) and the coupled systems defined by both DDEs and REs. These equations appear in a wide range of applications, from engineering to natural sciences, and coupled systems are especially present in population dynamics models. This is what motivates all the work done in order to produce this thesis.

The problems that have been tackled concern the computation of equilibria and periodic solutions. The complexity of such problems lies in part in that of the realistic models. Indeed, these models are often formulated through complicated right-hand sides, as in the case of the consumer-resource model for the *Daphnia magna* water flea feeding on algae [Diekmann, Gyllenberg, Metz, Nakaoka and de Roos, J. Math. Biol. 61 (2010)]. Moreover, realistic models depend on several (varying or uncertain) model parameters. In fact, all the methods presented are intended to be used within a parameter continuation framework.

It is fair to claim that collocation methods constitute the leitmotif of the thesis. Indeed, they are the fundament of the standard approach used to discretize the relevant dynamical systems into finite-dimensional ones - an operation which is at the basis of the application of numerical methods.

The methods presented in this thesis show how collocation can also be used to improve existing methods for the computation of equilibria - notably badly-performing in many realistic cases -, as well as to compute periodic solutions.

Concerning the latter, it is worth observing that the current literature on REs is not yet as developed as the one on DDEs. In particular, the well-known method presented in [Engelborghs, Luzyanina, in 't Hout and Roose, SIAM J. Comput. 22 (2001)] for computing periodic solutions of DDEs was never extended to REs previously.

The main *theoretical* contribution of the thesis is the proof of the convergence of (a variant of) the method for DDEs, which is highly based on the abstract approach in [Maset, Numer. Math. 133 (2016)]. To the best of the author's knowledge, such a proof has never been obtained before, but for some specific forms of DDEs.

To sum up, the original contributions of this thesis concern the numerical study of certain delay equations under various aspects. In particular, the description of the new method to compute (parameter-dependent) equilibria is accompanied by some numerical test which show its outperformance of existing methods. Similarly, the validation of the method to compute periodic solutions of delay equations is also supported by some numerical tests contained herein.

SOMMARIO

Questa tesi raccoglie i risultati principali ottenuti dalla ricerca svolta durante il percorso di dottorato dell'autrice, dedicata a metodi numerici per lo studio del comportamento di sistemi dinamici generati da certe classi di equazioni con ritardo, ovvero equazioni funzionali in cui il valore di una funzione al tempo attuale, o quello della sua derivata, è definito tramite i valori nel passato.

Tali sistemi dinamici sono infinito-dimensionali, il che rappresenta la ragione principale per cui, quando si tratta di calcolare effettivamente le loro soluzioni o altre quantità correlate, non è possibile ottenere dei risultati analicamente, e dunque necessario ricorrere a metodi numerici.

Nello specifico, le equazioni oggetto di questa tesi sono le equazioni differenziali con ritardo (DDE), le equazioni di rinnovo (RE) e i sistemi definiti da DDE ed RE accoppiate. Tali equazioni appaiono in un'ampia gamma di applicazioni, dall'ingegneria alle scienze naturali, e i sistemi accoppiati sono particolarmente presenti nei modelli di dinamica delle popolazioni. Questo è ciò che motiva tutto il lavoro svolto al fine di produrre questa tesi.

I problemi affrontati riguardano il calcolo degli equilibri e delle soluzioni periodiche. la complessità di questi problemi è in parte conseguenza di quella degli stessi modelli che appaiono effettivamente nelle applicazioni. In particolare, questi modelli sono spesso formulati in termini di membri destri molto complessi, come nel caso del modello consumatore-risorsa della pulce d'acqua *Daphnia magna* che si ciba di alghe [Diekmann, Gyllenberg, Metz, Nakaoka and de Roos, J. Math. Biol. 61 (2010)]. Inoltre, i modelli realistici dipendono da diversi parametri (variabili o dai valori incerti). In effetti, tutti gli approcci presentati sono pensati per un utilizzo all'interno di un metodo di continuazione dei parametri.

È legittimo affermare che i metodi di collocazione costituiscono il filo conduttore della tesi. Nella fattispecie, sono alla base del metodo standard per discretizzare i sistemi dinamici di interesse al fine di trasformarli in sistemi finito-dimensionali - un processo che è alla base dell'applicazione di qualsiasi metodo numerico.

I metodi presentati nella tesi mostrano anche come la collocazione può essere usata per migliorare i metodi già sviluppati per la continuazione degli equilibri - noti per essere poco efficienti in molti casi realistici -, nonché per il calcolo delle soluzioni periodiche.

Riguardo a queste ultime, è opportuno menzionare che la letteratura attuale sulle RE è meno sviluppata di quella sulle DDE. In particolare, il noto metodo presentato nel lavoro [Engelborghs, Luzyanina, in 't Hout and Roose, SIAM J. Comput. 22 (2001)] per il calcolo delle soluzioni periodiche di DDE non era mai stato esteso alle RE in precedenza.

Il principale contributo *teorico* della tesi è la dimostrazione di convergenza di una variante del metodo per le DDE, ampiamente basato sull'approccio

astratto in [Maset, Numer. Math. 133 (2016)]. A conoscenza dell'autrice, tale dimostrazione non era mai stata ottenuta in precedenza, se non per qualche forma specifica di DDE.

Riassumendo, i contributi originali della tesi riguardano lo studio numerico di alcune equazioni con ritardo sotto diversi aspetti. In particolare, la descrizione del nuovo metodo per il calcolo di equilibri (dipendenti da parametri) è accompagnata da alcune simulazioni numeriche che mostrano la sua superiorità rispetto ai metodi standard in termini di efficienza. Allo stesso modo, la validità del metodo per il calcolo delle soluzioni periodiche di equazioni con ritardo è supportata da alcune simulazioni qui presentate.

CONTENTS

1 INTRODUCTION: MODELING DELAYED EFFECTS

Ordinary Differential Equations (ODEs) describe the rate of change of a quantity at a specific time as a function of the value of the quantity itself at the same time. Their use in mathematical modeling dates back to the eighteenth century, when Malthus formulated his exponential growth law in one of the overall first works in the field of population dynamics [76]. Progress has been made since then: ODEs are nowadays a widespread tool in all areas of social and natural sciences, and can be solved numerically with a high accuracy. However, in many social and natural phenomena, there is a *delay* between a cause and the corresponding effect. In other words, the present evolution of the relevant quantities is also influenced by their past values, and ODEs are often not able to capture this dependence on the past adequately.

One of the first works concerning the theory of systems which take delay effects into account dates back to the beginning of the twentieth century [103]. Some time later, delay models started to become more popular, as a result of the spread of automatic control systems in engineering. A popular example in this area is the hot shower problem [68], where a *measurement delay* arises. Indeed, the change of the temperature of the water can only be perceived by the relevant device after a certain amount of time.

More natural examples of models involving delays can be found in several areas of biology (see, e.g., [71] or [95] for a general reference). A first example in population dynamics is the delayed logistic equation, introduced in [63] to improve the logistic equation proposed by Verhulst in [102], which, in turn, aimed at including competition within the species in the model of Malthus. Other examples are the Physiologically Structured Population Models (PSPM) [37, 38, 39, 40, 43], where delays appear as a result of the presence of different life stages (e.g., juveniles and adults). In epidemiology, delays can be used to model latent periods (when individuals are infected but not yet contagious) or incubation periods (when individuals are infected but do not yet show any symptoms). In behavioral epidemiology, delays have been introduced to model, e.g., the vaccination coverage at birth as a function of some *memory* variable, which contains information about past values of the relevant quantities (see [48] as one of the first works on the subject, and [47] as a general reference on behavioral epidemiology).

The class of models of interest in this thesis can be expressed in the general form

$$\begin{cases} x(t) = F(x_t, y_t) \\ y'(t) = G(x_t, y_t), \end{cases} \tag{1.1}$$

where

$$F : \mathtt{X} \times \mathtt{Y} \to \mathbb{R}^{d_X}, \qquad G : \mathtt{X} \times \mathtt{Y} \to \mathbb{R}^{d_Y}$$

are autonomous, in general nonlinear functions for positive integers d_X and d_Y, and the *state spaces* $\mathtt{X}$ and $\mathtt{Y}$ are Banach spaces classically (see, e.g., [40]) defined respectively as

$$\mathtt{X} := L^1([-\tau, 0], \mathbb{R}^{d_X}), \qquad \mathtt{Y} := C([-\tau, 0], \mathbb{R}^{d_Y})$$

for some $\tau > 0$, called the *delay*. The x-component of the *state* of the dynamical system on the state space $\mathtt{X} \times \mathtt{Y}$ at time t associated to (1.1) is denoted by x_t, defined as

$$x_t(\theta) := x(t + \theta), \quad \theta \in [-\tau, 0], \tag{1.2}$$

and the same notation holds for the y-component.

The first equation of (1.1) is called a Renewal Equation (RE) and the second is a Delay Differential Equation (DDE). In its most general form, (1.1) is called *coupled* or *delay* equation or system.

Coupled systems have appeared more and more frequently in modeling of structured populations [21, 41, 42, 43, 44, 64, 65, 82, 91]. In these realistic cases, the function F defining the right hand side of the RE has a *smoothing* effect, i.e., is integral in the x-component. Examples are

$$F(\phi) = \int_{-\tau}^{0} H(\theta, \phi(\theta))\, \mathrm{d}\theta \tag{1.3}$$

for some nonlinear function H, or

$$F(\phi) = g\left(\int_{-\tau}^{0} H(\theta)\phi(\theta)\, \mathrm{d}\theta\right) \tag{1.4}$$

for some nonlinear function g. Given the interest towards this class of models, many works have also recently appeared concerning numerical approaches for their stability and bifurcation analyses [22, 23, 24, 25, 26, 38, 90].

In the following sections of this introductory chapter, several aspects of delay models will be introduced. In particular, Section 1.1 will give a general mathematical background, while Sections 1.2 and 1.3 will give an overview on the numerical issues concerning delay equations and on related software, respectively. Section 1.4 will introduce an emblematic complex PSPM described by a coupled system.

1.1 DYNAMICAL SYSTEMS (FROM DELAY EQUATIONS)

A (continuous-time) dynamical system is a triple $\{\mathbb{T}, X, \{T(t)\}_{t\in\mathbb{T}}\}$ where the *set of times* $\mathbb{T} \subseteq \mathbb{R}$ contains 0 and is closed under addition, the state space X is the set of possible values of the evolving quantities, and $\{T(t)\}_{t\in\mathbb{T}}$ are *evolution operators* on X, i.e., they satisfy the following properties:

- $T(0) = I$;
- $T(t + s) = T(t) \circ T(s), \quad t, s \in \mathbb{T}$.

In particular, the operator $T(t) : X \to X$ is defined as

$$T(t)x_0 = x_t,$$

if x_0 is the initial state and x_t is the state at time t.

Many real-life phenomena in science and engineering are defined by Initial Value Problems (IVPs). Those for autonomous ODEs, DDEs and REs define dynamical systems, as long as they are well-posed, i.e., their solutions exist and are unique. For example, an IVP for the DDE

$$y'(t) = G(y_t), \tag{1.5}$$

with state space $\mathtt{Y} \subseteq \mathbb{F}([-\tau, 0], \mathbb{R}^d) := \{f : [\tau, 0] \to \mathbb{R}^d\}$, is defined by an initial function $\phi \in \mathtt{Y}$ and some $M > 0$, and consists of finding a solution of (1.5) on $[-\tau, M]$ which satisfies the initial condition $y_0 = \phi$, i.e., a solution of

$$\begin{cases} y'(t) = G(y_t), & t \in [0, M], \\ y_0 = \phi, & t \in [-\tau, 0]. \end{cases} \tag{1.6}$$

In this context, $y'(t)$ stands for the right-hand derivative $y'(t)^+$. A sufficient condition for the well-posedeness of the problem is given by the following theorem.

Theorem 1.1 ([68, Theorem 2.1, Chapter 3]). *Given an open set $\Omega \subseteq \mathtt{Y}$, if G is locally Lipschitz continuous in Ω, then for all $\phi \in \Omega$ there exists $M > 0$ such that (1.6) has a unique solution in $[-\tau, M]$.*

Results concerning the well-posedeness of IVPs for RE defined by a general right hand side and in particular of the form (1.3) can be found in [59, Sections 12.1 and 12.2].

The evolution operator $T(t)$ of a dynamical system as the one above is also called *solution operator*. In this context, an invariant set for all solution operators is simply called *invariant set*. These include *equilibria*, i.e., constant solutions, or *cycles*, i.e., periodic solutions.

In many applications, there is a strong interest in determining the *stability* properties of invariant sets. An invariant set $S \subseteq X$ is *stable* if for all neighborhoods $U \supset S$ there exists a neighborhood $V \supset S$ such that $T(t)$ maps V to U for all $t \geq 0$. It is *unstable* otherwise. If S is stable and there exists a neighborhood $V_S \supset S$ such that $T(t)V_S \to S$ as $t \to +\infty$, then S is *asymptotically stable*.

A nonlinear system can be *linearized* around an equilibrium or a periodic solution; for instance, in the case of (1.5), linearization around a solution y^* would read

$$y'(t) = DG(y_t^*)y_t. \tag{1.7}$$

If the relevant solution is an equilibrium, then the linearized system will be autonomous, while if it is a periodic solution, then the linearized system will have periodic coefficients. The linearization procedure turns out to be crucial upon studying the stability properties of an invariant set (see, e.g., [75]). Indeed, in general the local stability properties of a solution are strongly related to those of the null solution of the relevant linearized system (e.g., [45, Section 5 of Chapter VII]). In the case of linear autonomous DDEs and REs, they are in turn determined by the spectrum of the semigroup of solution operators of the linearized problem, or that of its infinitesimal generator (see [51] as a general reference).

Theorem 1.2 (Principle of linearized stability for equilibria of DDEs, [45, Theorem 6.8, Chapter VII]). *Let $G \in C^1(\Upsilon, \mathbb{R}^d)$ and let y^* be an equilibrium of (1.5). If all the eigenvalues of the infinitesimal generator of the semigroup of solution operators of (1.7) have negative real part, then y^* is asymptotically stable. If at least one has positive real part, then y^* is unstable.*

Similar results for equilibria of REs can be found in [40, Section 3.4]. In the case of linear periodic DDEs, which can arise upon linearization around a periodic solution, the stability properties of the null solution are determined by the eigenvalues of the *monodromy operator*, i.e., the evolution operator which shifts a state of one period. Such eigenvalues are called the *Floquet multipliers* or *characteristic multipliers*. As stated in, e.g., [68, Theorem 1.6, Chapter 4], if all of them, except the multiplier 1 whenever it is simple, are inside the unit circle then the null solution is asymptotically stable. If at least one of them is outside of the unit circle, then the null solution is unstable. Similar results for REs have only recently been obtained (see [27]).

Note that, if the equations are not autonomous, then their evolution vary with time. This is expressed by defining solution operators $\{T(t,s)\}_{t>s\in\mathbb{T}}$ depending on two time parameters, and satisfying the properties

- $T(t,t) = I, \quad t \in \mathbb{T}$;
- $T(t,s) = T(t,r) \circ T(r,s), \quad t > r > s \in \mathbb{T}$.

1.2 THE COMPLEXITY OF DELAY MODELS

The main difficulties arising from the inclusion of delay terms are attributable to the consequent enlargement of the state space, in particular from finite- to infinite-dimensional. Indeed, as mentioned in Section 1.1, for an IVP to be well-posed, a single initial value is no longer sufficient. Rather, a history map defined in $[-\tau, 0]$ is necessary. Moreover, this leads to infinite dimensionality of the evolution operators defining delay dynamical systems and, potentially, an infinite amount of eigenvalues to detect in order to determine the stability of a certain invariant set.

Clearly, such a theoretical investigation is not feasible in practice. However, analyzing the stability of an equilibrium or periodic solution, as well as detecting bifurcations, are all typical targets in this context. This brings the need for some kind of *discretization*, i.e., transformation of the problem into a finite-dimensional one which preserves the relevant properties. To this aim, several works proposing and developing a pseudospectral discretization technique (based on collocation, see Section 3.3) have appeared in the past decade. In particular, the stability of linear DDEs was first tackled in [29]. [26] is the first systematic attempt to attack the stability of periodic solutions of REs, which was extended to coupled systems in [27]. Eventually, in [22] the method is extended to nonlinear equations in a way such that linearization and discretization commute (see Section 4.2 for more details).

The relevance of the latter work is explained by the wide availability of software tools for ODEs which are able to investigate stability and bifurcations, taking only the (discretized) right-hand side as an input (e.g., MAT-

CONT [4], AUTO [1]). However, this approach has a downside: it becomes very computationally demanding when applied to realistic models defined by a complicated right-hand side. For instance, it is often the case - especially in PSPM, see Section 1.4 - that the right-hand side is not given by an exact expression, but rather through solutions of external ODEs. Indeed, those ODEs are solved within the function defining the right-hand side, and not directly by MATCONT, which means that they have to be solved from scratch for every value of the continuation parameter.

In other words, it is impossibile to exploit the intermediate results obtained during previous continuation steps (e.g., the solutions of external ODEs for other values of the continuation parameter). It is fair to assume that this plays a role in determining the overall computational cost, and to conjecture that the inclusion of the (resolution of) the external problems into the continuation framework would help lighten the computational burden.

This alternative *internal* continuation approach is proposed and described in full detail in Chapter 4. Experimental proof of its validity - and superiority over the standard *external* approach - for the continuation of equilibria is also provided therein, by means of numerical simulations on PSPM-like models of growing complexity. The plan for the immediate future is to extend the approach in order to be able to compute more general solutions, in particular periodic ones.

The computation of periodic solutions, indeed, is often a topic of interest in applications of delay equations. A numerical method for DDEs was proposed in [53], but was never extended to REs (or coupled systems), for which such methods still lack. An attempt to extend the approach in [53] to REs will be presented in Chapter 5. It will be also shown, through some numerical simulations, that the orders of convergence obtained in [53] also hold for REs and coupled systems.

From a more theoretical point of view, references lack even in the case of DDEs. To the best of the author's knowledge, a full analysis of the error and the relevant convergence of the method has never been performed for the general case (1.5), but rather only in some specific cases (see Sections 6.1 for more details). Chapters 6 and 7 will describe an attempt to bridge this gap, based on the abstract framework proposed in [79] for computing solutions of Boundary Value Problems (BVPs) for delay equations. It will be shown, in particular, that the reformulation of the problem as a BVP requires an infinite-dimensional boundary condition in order to fit into the class of BVPs considered in [79]. Eventually, convergence of the finite element method (see Subsection 3.3.1) will be proved and accompanied by a detailed convergence analysis.

Thus, the continuation of equilibria and the computation of periodic solutions of delay equations constitute the main contributions of this work, and both relate to the use of collocation methods as a discretization tool.

1.3 SOFTWARE TOOLS: STATE OF ART

The growing interest in dynamical systems has naturally led to a growing number of numerical packages for stability and bifurcation analyses. As mentioned in Section 1.2, most of them are only suited for finite dimensional systems (i.e., ODEs), e.g., MATCONT [4] and AUTO [1].

Some classes of delay equations can still be partially studied without a (manual) prior discretization. In particular, several software packages are available for DDEs defined by *discrete delays* only, i.e., with right-hand side of the form

$$G(y(t), y(t-\tau_1), \dots, y(t-\tau_n)).$$

Examples are DDE-BIFTOOL [2], XPP-AUT [6], KNUT [3] and PYDSTOOL [5].

Other software is available for the study of PSPM described by a coupled system, as will be described in Section 1.4. Examples are EBTTOOL [36] and PSPMANALYSIS [35]. The work [90], which will be mentioned in Chapter 4, is based on the same ideas. As general references on numerical methods for DDEs and related software, one can consider [17] and [30].

1.4 DAPHNIA

Coupled RE/DDE equations often arise in PSPM. In these models, the vital rates of the individuals (e.g., rates of reproduction, death or consumption of some resources) depend on one or more real variables that constitute the *structure* of the individuals (usually their size) which, in turn, vary piecewise continuously with their age. The potential discontinuity points are due to the different life stages of individuals (e.g., juvenile and mature), that are reached as soon as the structuring variables pass certain treshold values. In the majority of cases, the model includes some feedback from the environment (e.g., when they involve predator-prey interactions). A famous example that can be used to illustrate these concepts is the one of the *Daphnia magna*, or simply *Daphnia* [43]. The following description is mainly inspired by [25].

Daphnia feeds on a resource which has concentration $S(t)$ at time t. In the absence of consumers, this concentration evolves according to an IVP for an ODE of the form

$$\begin{cases} S'(t) = f(S(t)), & t \geq 0, \\ S(0) = S_0 \end{cases} \tag{1.8}$$

for given $f : [0, +\infty) \to \mathbb{R}$ and $S_0 > 0$, the latter condition in order to be biologically meaningful. In the classical version of the model, the resource history S_t is defined on a closed interval $[-a_{\max}, 0]$, meaning that there is a maximum age $a_{\max}$ that individuals can reach.

The structuring variable of the model is the size $\xi(a, S_t)$, which depends on both the age and the experienced resource history. The size $\bar{\xi}(\alpha) := \bar{\xi}(\alpha; a, \psi)$

at age $\alpha \in [0, a]$ of an individual that at age a has experienced a resource history ψ is defined by means of the solution of the IVP

$$\begin{cases} \bar{\xi}'(\alpha) = g(\bar{\xi}(\alpha), \psi(-a+\alpha)), & \alpha \in [0, a], \\ \bar{\xi}(0) = \xi_b, \end{cases} \tag{1.9}$$

for a given growth rate $g : [\xi_b, +\infty) \times [0, +\infty) \to (0, +\infty)$ and size at birth $\xi_b > 0$. Then $\xi(a, \psi) := \bar{\xi}(a; a, \psi)$.

The vital rates involved are the survival probability and the birth and consumption rates. The former, just like the size, is defined by means of the solution of an IVP. The survival probability $\bar{\mathcal{F}}(\alpha) := \bar{\mathcal{F}}(\alpha; a, \psi)$ at age $\alpha \in [0, a]$ of an individual that at age a has experienced a resource history ψ is given by

$$\begin{cases} \bar{\mathcal{F}}'(\alpha) = -\mu(\bar{\xi}(\alpha), \psi(-a+\alpha))\bar{\mathcal{F}}(\alpha), & \alpha \in [0, a], \\ \bar{\mathcal{F}}(0) = 1 \end{cases} \tag{1.10}$$

for a given mortality rate $\mu : [\xi_b, +\infty) \times [0, +\infty) \to (0, +\infty)$. Then $\mathcal{F}(a, \psi) := \bar{\mathcal{F}}(a; a, \psi)$. The other vital rates are defined through given functions $\beta, \gamma : [\xi_b, +\infty) \times [0, +\infty) \to \mathbb{R}$. In particular, the reproduction and consumption rates of a consumer individual that at time t has age a and size $\xi(a, S_t)$ are denoted respectively $\beta(\xi(a, S_t), S(t))$ and $\gamma(\xi(a, S_t), S(t))$.

The lifespan of the individual consists of two stages, juveniles and adults, the former being unable to reproduce. The adult stage is reached when the size passes the maturation size ξ_A ($> \xi_b$), which is the only possible breaking point of the vital rates (with respect to their first argument on $[\xi_b, +\infty)$), as well as of the growth rate. The maturation age a_A is implicitly given through the maturation condition

$$\xi(a_A, \psi) = \xi_A. \tag{1.11}$$

The model consists, therefore, of a coupled RE/DDE system. The RE describes the consumer birth rate $b(t)$, which is obtained by integrating with respect to the age the contribution of the individuals who are in the adult stage at the present time t, i.e.,

$$b(t) = \int_{a_A(S_t)}^{a_{\max}} \beta(\xi(a, S_t), S(t))\mathcal{F}(a, S_t)b(t-a)\,\mathrm{d}a.$$

The length of the age interval involved is therefore not fixed, but rather *state-dependent*, i.e., depending on the resource history (through (1.11)).
The DDE describes the dynamics of the resource by subtracting the total consumption from the right-hand side of the ODE in (1.8), i.e.,

$$S'(t) = f(S(t)) - \int_0^{a_{\max}} \gamma(\xi(a, S_t), S(t))\mathcal{F}(a, S_t)b(t-a)\,\mathrm{d}a. \tag{1.12}$$

The presence of all the complications above, in particular external ODEs and the implicit equation defining the maturation age, make Daphnia a good representative of the class of problems of interest in this thesis.

1.5 ORGANIZATION OF THE THESIS

The rest of the thesis will be organized as follows. The current Chapter will be concluded with Section 1.6, which introduces the relevant notation that will be later used in the rest of the thesis. Chapters 2 and 3 will summarize some well-known results which are used in the rest of the thesis (while being explicitly referenced), and may, therefore, be skipped by the reader who has already some experience with the relevant topics. In particular, Chapter 2 will introduce the theoretical preliminaries. These include general properties of Banach spaces, as well as some theory concerning DDEs and REs. Chapter 3 will present, on the other hand, the numerical preliminaries. Being most of the work for the thesis focused on (piecewise) polynomial collocation methods, some attention will be devoted to polynomial interpolation and how its properties affect the relevant collocation framework. Moreover, general results on parameter-dependent numerical continuation will be provided. Chapter 4 will focus on the continuation of equilibria in the context of parameter-dependent models, particularly on the proposed internal continuation. Chapters 5, 6 and 7 will concern the computation of periodic solutions. The former, from a more experimental point of view: in particular, the method proposed in [53] for DDEs is extended to REs and tested to an example of RE from the work [23]. Chapter 6 will describe the work done in view of a theoretical convergence analysis of the aforementioned method for DDEs, based on the abstract approach in [79]. Chapter 7 will play the role of appendix of Chapter 6, by containing the proof of some results that are considered rather technical. Finally, Chapter 8 will provide some conclusions, as well as possible perspectives for future work.

1.6 NOTATION AND CONVENTIONS

- $|\cdot|$ denotes any norm in $\mathbb{R}^n$, for any $n \in \mathbb{N}$.
- $\|\cdot\|_{\mathbb{B}}$ denotes the norm of an infinite-dimensional Banach space $\mathbb{B}$.
- $\mathbb{F}$ denotes the set of all the functions, that is

$$\mathbb{F}(A, B) = \{f : A \to B\}.$$

- $C = C^0$ denotes the set of all the continuous functions, and is a Banach space when equipped with the *uniform* norm. For functions in $C(A, B)$, this is defined as

$$\|f\|_\infty = \sup_{x \in A} \|f(x)\|_B.$$

  C^k denotes the set of all the functions having continuous k-th derivative, and is a Banach space when equipped with the norm defined by

$$\|f\|_{C^k} = \sum_{i=0}^{k} \|f^{(i)}\|_\infty.$$

  C^∞ is defined as the intersection $\bigcap_{k \in \mathbb{N}} C^k$.

- B^∞ denotes the measurable and bounded functions, intended as functions which are defined pointwise, and not classes of functions which are equal almost everywhere. $B^{1,\infty}$ denotes the continuous functions having derivative in B^∞. They are Banach spaces when equipped with the norms defined respectively as

$$\|f\|_{B^\infty} := \|f\|_\infty$$

and

$$\|f\|_{B^{1,\infty}} := \|f\|_\infty + \|f'\|_\infty. \tag{1.13}$$

- Given $G : \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}^n$ and $(\bar{x}, \bar{y}) \in \mathbb{R}^m \times \mathbb{R}^p$, $G_x(\bar{x}, \bar{y}) \in \mathbb{R}^{n\times m}$ denotes the matrix

$$\begin{pmatrix} \frac{\partial G_1}{\partial x_1}(\bar{x}, \bar{y}) & \cdots & \frac{\partial G_1}{\partial x_m}(\bar{x}, \bar{y}) \\ \vdots & \ddots & \vdots \\ \frac{\partial G_n}{\partial x_1}(\bar{x}, \bar{y}) & \cdots & \frac{\partial G_n}{\partial x_m}(\bar{x}, \bar{y}) \end{pmatrix}$$

- $I_{\mathbb{B}} : \mathbb{B} \to \mathbb{B}$ denotes the identity operator on a Banach space $\mathbb{B}$. I without a subscript is used in the place of $I_{\mathbb{B}}$ whenever there is no ambiguity concerning the relevant Banach space.

- With reference to a Banach space $\mathbb{B}$, the notation $\mathcal{B}$ refers to the closed set

$$\mathcal{B}(b, r) := \{b' \in \mathbb{B} \mid \|b - b'\|_{\mathbb{B}} \leq r\}, \qquad b \in \mathbb{B}.$$

- Π_n denotes the space of $\mathbb{R}^d$-valued (for any d) polynomials of degree at most n.

- Given U_1, U_2 normed spaces, unless otherwise specified, the norm of the space $U := U_1 \times U_2$ is

$$\|\cdot\|_U = \max\{\|\cdot\|_{U_1}, \|\cdot\|_{U_2}\}, \tag{1.14}$$

and makes U a Banach space whenever U_1, U_2 are Banach spaces.

- $\mathcal{L}(U, V)$ is the set of linear bounded operators $U \to V$, equipped with the induced norm

$$\|A\|_{V\leftarrow U} = \sup_{u\in U\setminus\{0\}} \frac{\|Au\|_V}{\|u\|_U}. \tag{1.15}$$

2 THEORETICAL PRELIMINARIES

2.1 LINEAR (PERIODIC) DDES

This section collects some of the results of [61, Chapters 6, 8 and 9] and is devoted to the representation of the solution of linear inhomogeneous DDEs of the form

$$y'(t) = L(t, y_t) + h(t), \tag{2.1}$$

as well as the relevant theory of adjoint equations. Most of the content of this section is relatively advanced, when compared to other topics included in this Chapter. The reason why it is nevertheless presented in this preliminary Chapter is to improve the flow of reading in Subsection 6.3.1, which constitutes one of the main contributions of this thesis and is heavily based, indeed, on the theory of adjoint equations.

Before introducing the latter, a couple of well-known results in the context of linear DDEs will be presented.

Theorem 2.1 (Variation of constants formula, [61, Theorem 2.1 and (2.7) in Chapter 6]). *Let $h : [0, +\infty) \to \mathbb{R}^d$ be a Lebesgue integrable function in each compact subset of its domain. Let $\mathtt{Y}$ be a subset of $\mathbb{F}([-\tau, 0], \mathbb{R}^d)$, and let $L : \mathbb{R} \times \mathtt{Y} \to \mathbb{R}^d$ be a function which is linear with respect to the second argument and can be written through the Riemann-Stieltjes integral*

$$L(t, \psi) = \int_{-\tau}^{0} \mathrm{d}_\sigma \mathtt{n}(t, \sigma)\psi(\sigma), \tag{2.2}$$

where $\mathtt{n} : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is normalized so that

$$\mathtt{n}(t, \theta) = 0,\ \theta \geq 0, \qquad \mathtt{n}(t, \theta) = \mathtt{n}(t, -\tau),\ \theta \leq -\tau,$$

for $t \in \mathbb{R}$, $\mathtt{n}(t, \cdot) : [-\tau, 0] \to \mathbb{R}^{d\times d}$ is of bounded variation and $\mathtt{n}(t, \cdot) : (-\tau, 0) \to \mathbb{R}^{d\times d}$ is continuous from the left. Moreover, assume that there is a Lebesgue integrable function $m : \mathbb{R} \to \mathbb{R}$ such that $|L(t, \psi)| \leq m(t)|\psi|$ for all $\psi \in \mathtt{Y}$. Then, the solution y^* of the IVP

$$\begin{cases} y'(t) = L(t, y_t) + h(t), & t \geq 0, \\ y_0 = \psi \end{cases}$$

satisfies the variation of constants formula

$$y^*_t = \psi + \int_0^t [T(t, s)X_0]h(s)\,\mathrm{d}s, \quad t \geq 0, \tag{2.3}$$

where $T(t, s)$ is the solution operator of the linear homogeneous part and

$$X_0(\theta) := \begin{cases} 0, & \theta \in [-\tau, 0), \\ I, & \theta = 0. \end{cases}$$

A general reference for functions of normalized bounded variation and Riemann-Stieltjes integral can be found in [45, Appendix I].

A sufficient condition for a linear functional to satisfy the hypotheses of Theorem 2.1 is given by the following result.

Theorem 2.2 ([45, Theorem 1.1, Chapter I]). *Let L be a continuous linear mapping from $\mathtt{Y}$ into $\mathbb{C}^d$. Then there exists a unique function of normalized bounded variation $\mathtt{n} : [-\tau, 0] \to \mathbb{C}^{d\times d}$ such that, for all $\psi \in \mathtt{Y}$,*

$$L\psi = \int_{-\tau}^{0} \mathrm{d}_\sigma \mathtt{n}(\sigma)\psi(\sigma),$$

where the integral is a vector whose i-th component reads

$$\sum_{j=1}^{d} \int_{-\tau}^{0} \psi_j(\sigma)\, \mathrm{d}_\sigma \mathtt{n}_{i,j}(\sigma).$$

The rest of the Section focuses on periodic DDEs, particularly concerning the theory of adjoint equations. (2.1) can be written, through (2.2), as

$$y'(t) = \int_{-\tau}^{0} \mathrm{d}_\sigma \mathtt{n}(t,\sigma)y(t+\sigma) + h(t). \tag{2.4}$$

Assuming it has a solution defined on the whole line, the corresponding *formal adjoint equation* reads

$$z(t) + \int_{t}^{\infty} z(\sigma)\mathtt{n}(\sigma, t-\sigma)\,\mathrm{d}\sigma = \text{constant}, \tag{2.5}$$

where $z(t)$ is intended as a row vector. An element of the corresponding state space is denoted by

$$z^t(\theta) := z(t+\theta), \quad \theta \in [0,\tau].$$

Note that ω-periodic equations are exactly the ones defined by some ω-periodic h and some $\mathtt{n}$ which is ω-periodic with respect to the first argument.

Theorem 2.3 ([61, Lemma 2.1, Chapter 8]). *Consider the bilinear form*

$$(\psi,\phi)_t := \psi(0)\phi(0) + \int_{-\tau}^{0} \mathrm{d}_\beta \left[\int_{0}^{\tau} \psi(\xi)\mathtt{n}(t+\xi,\beta-\xi)\,\mathrm{d}\xi\right]\phi(\beta). \tag{2.6}$$

If y is a solution of (2.4), and z is a solution of (2.5) for $t \geq \sigma$, then

$$\frac{d}{dt}(z^t, y_t)_t = y(t)h(t).$$

The following results concern homogeneous systems. Recall that the characteristic multipliers of an equation are the eigenvalues of the relevant monodromy operator (see Section 1.1). In particular, a solution is periodic if and only if its state at any time t is an eigenvector corresponding to the eigenvalue 1.

Lemma 2.4. *If h is the null function and $\mathtt{n}$ is periodic with respect to its first argument, then μ is a characteristic multiplier of (2.4) if and only if μ is a characteristic multiplier of (2.5).*

Proof. By [61, Section 8.2], the monodromy operators of (2.4) and of its adjoint (2.5) share the same spectrum (see [61, Section 6.4] for the relevant definitions and properties in the autonomous case). □

Corollary 2.5. *Under the hypotheses of Lemma 2.4, (2.4) has an ω-periodic solution if and only if (2.5) does.*

Proposition 2.6 ([61, page 200])**.** *Under the hypotheses of Lemma 2.4, if $\Lambda(t)$ is a basis of the generalized eigenspace corresponding to the multiplier λ of (2.4) and $\tilde{\Lambda}(t)$ is a basis of the generalized eigenspace corresponding to the same multiplier for (2.5), then the matrix $(\Lambda(t), \tilde{\Lambda}(t))$ defined by the bilinear form (2.6) is nonsingular.*

The last result concerns again inhomogeneous linear periodic systems.

Theorem 2.7 ([61, Theorem 1.2, Chapter 9])**.** *If h is ω-periodic and so is the linear term of (2.4), then the latter has a ω-periodic solution if and only if*

$$\int_0^\omega z(t)h(t)\,\mathrm{d}t = 0$$

for all ω-periodic solutions z of (2.5).

As a final remark, it is worth observing that the results of this section concerning the adjoint theory could be seen in the general framework of Fredholm theory, see, e.g., [70, Chapter 4]. Indeed, by choosing suitable (subspaces of the) state spaces of (2.4) and (2.5), the bilinear form (2.6) is nondegenerate, and the spaces constitute a *dual system*. Moreover, by [61, Equation (2.12), Section 8.2], their respective monodromy operators are adjoint. Thus, it follows from the first Fredholm theorem [70, Theorem 4.14] that (2.4) and its adjoint (2.5) have the same number of linearly independent 1-periodic solutions. The fact that they are not orthogonal ($c^* \neq 0$) would follow from the Fredholm alternative theorem [70, Theorem 4.17] and the decomposition (6.36).

2.2 PERIODIC BOUNDARY VALUE PROBLEMS

A periodic solution of (1.1) with period $\omega > 0$ can be obtained by solving a BVP of the form

$$\begin{cases} x(t) = F(x_t, y_t), & t \in [0, \omega], \\ y'(t) = G(x_t, y_t), & t \in [0, \omega], \\ (x_0, y_0) = (x_\omega, y_\omega), \\ p(x|_{[0,\omega]}, y|_{[0,\omega]}) = 0 \end{cases} \tag{2.7}$$

where $p : \mathbb{R}^{d_X} \times \mathbb{R}^{d_Y} \to \mathbb{R}$ is a (usually linear) function defining the *phase* condition, which is necessary in order to remove translational invariance. An example of phase condition is the *trivial* one, of the form

$$x_k(0) = \hat{x} \quad \text{or} \quad y_k(0) = \hat{y}, \tag{2.8}$$

for some $k \in \{1, \dots, d_X\}$ or $k \in \{1, \dots, d_Y\}$ (respectively) where $\hat{x}$ and $\hat{y}$ are fixed. Otherwise, an *integral* phase condition is of the form

$$\int_0^\omega \langle x_k(t), \tilde{x}_k'(t) \rangle\,\mathrm{d}t = 0 \quad \text{or} \quad \int_0^\omega \langle y_k(t), \tilde{y}_k'(t) \rangle\,\mathrm{d}t = 0, \tag{2.9}$$

for some $k \in \{1, \dots, d_X\}$ or $k \in \{1, \dots, d_Y\}$ (respectively) where $\tilde{x}$ and $\tilde{y}$ are given reference solutions [46]. The solutions of (2.7) are intended as functions defined in $[-\omega, \omega]$.

Since the period ω is unknown, it is numerically convenient (see, e.g., [53]) to reformulate (2.7) through the map $s_\omega : \mathbb{R} \to \mathbb{R}$ defined by

$$s_\omega(t) := \frac{t}{\omega}. \tag{2.10}$$

(2.7) is thus equivalent to

$$\begin{cases} x(t) = F(x_t \circ s_\omega, y_t \circ s_\omega), & t \in [0,1], \\ y'(t) = \omega G(x_t \circ s_\omega, y_t \circ s_\omega), & t \in [0,1], \\ (x_0, y_0) = (x_1, y_1) \\ p(x|_{[0,1]}, y|_{[0,1]}) = 0, \end{cases} \tag{2.11}$$

the solution of which is intended to be defined in $[-1,1]$ and represents a 1-periodic function, while the corresponding natural state spaces are Banach spaces of functions defined in $[-\frac{\tau}{\omega}, 0]$. However, note that one could choose spaces of functions defined in $[-r, 0]$ for any $r \geq \frac{\tau}{\omega}$. This can be, in fact, necessary in cases when ω might vary while the spaces need to be fixed, such as, e.g., in Chapter 6. Note that, in the context of periodic solutions, one can always consider $\tau \leq \omega$ without loss of generality, since a solution with period ω is also a solution with period $k\omega$ for any positive integer k. Indeed, such assumption will be made throughout the thesis.

Alternatively to (2.11), a BVP for the original equation (1.1) can also be formulated by considering the solutions over just one period (namely, in $[0,1]$) and imposing the periodicity to the solution values at the extrema of the period, rather than to the whole state. Note, however, that this requires to evaluate x and y at points that fall off the interval $[0,1]$, due to the delay. In order to deal with this issue, one can exploit the assumed periodicity to bring back the evaluation to the domain $[0,1]$. Formally, this means defining a *periodic state* $\overline{x}_t \in \mathtt{X}$ (the same holds for $\overline{y}_t \in \mathtt{Y}$) as

$$\overline{x}_t(\theta) = \begin{cases} x(t+\theta), & t+\theta \in [0,1], \\ x(t+\theta+1), & t+\theta \in [-1,0), \end{cases} \tag{2.12}$$

and rewrite (2.11), equivalently, as

$$\begin{cases} x(t) = F(\overline{x}_t \circ s_\omega, \overline{y}_t \circ s_\omega), & t \in [0,1], \\ y'(t) = \omega G(\overline{x}_t \circ s_\omega, \overline{y}_t \circ s_\omega), & t \in [0,1], \\ (x(0), y(0)) = (x(1), y(1)) \\ p(x, y) = 0. \end{cases} \tag{2.13}$$

Both formulations (2.11) and (2.13) are present in the literature on periodic BVPs. However, the latter is much more common (see Section 6.1 for more details).

2.3 FRÉCHET DERIVATIVES AND DIFFERENTIALS

All the results in Chapter 6 concern normed spaces and operators between them. This section is devoted to *Fréchet-differentiability*, which is a regularity property of operators between Banach spaces, for which a good general reference is [8, Chapter 1].

Definition 2.8 ([8, Definition 1.1.1])**.** Let U and V be normed spaces and $u_0 \in U$. A function $F : U \to V$ is *Fréchet-differentiable* at u_0 if there is $L \in \mathcal{L}(U, V)$ such that

$$\|F(u_0 + u) - F(u_0) - L(u)\|_V = o(\|u\|_U), \quad u \in U.$$

L is called *Fréchet differential of F at u_0*.

Note that the Fréchet differential is well defined. Let $L_1, L_2 \in \mathcal{L}(U, V)$ satisfy Definition 2.8 for some $u_0 \in U$. It follows that

$$\|L_1(u) - L_2(u)\|_V = o(\|u\|_U).$$

Assume for a contradiction that $L_1 \neq L_2$, and let $\overline{u} \in U$ such that

$$l := \|L_1(\overline{u}) - L_2(\overline{u})\|_V \neq 0.$$

Then, for $h \in \mathbb{R}$,

$$h \cdot o(\|\overline{u}\|_U) = \|o(\|h\overline{u}\|_U)\|L_1(h\overline{u}) - L_2(h\overline{u})\|_V = |h|\|L_1(\overline{u}) - L_2(\overline{u})\|_V = hl,$$

which gives $o(\|\overline{u}\|_U) = l$, a positive constant, and that is a contradiction.

Definition 2.9 ([8, Definition 1.1.5])**.** If $W \subseteq X$ is the set of points at which F is Fréchet-differentiable, $DF \in \mathcal{L}(W, V)$ denotes the *Fréchet derivative* of F, which maps each $u_0 \in W$ to the relevant Fréchet differential.

$\mathcal{C}^1(U, V)$ denotes the set of maps $F : U \to V$ which are continuously Fréchet-differentiable in U, i.e., such that $DF : U \to \mathcal{L}(U, V)$ is continuous.

The following theorem summarizes the main properties of the Fréchet differential.

Theorem 2.10 ([8, Proposition 1.1.4])**.**

(i) (Linearity) If U, V are normed spaces, $F, G : U \to V$ are Fréchet-differentiable at $u_0 \in U$ and $a, b \in \mathbb{R}$, then $aF + bG : U \to V$ is Fréchet-differentiable at $u_0 \in U$ and

$$D(aF + bG)(u_0) = aDF(u_0) + bDG(u_0).$$

(ii) (Chain rule) If U, V, Z are normed spaces, $F : U \to V$ is Fréchet-differentiable at $u_0 \in U$ and $G : F(U) \to Z$ is Fréchet-differentiable at $F(u_0) \in F(U)$, then the composite map $G \circ F : U \to Z$ is Fréchet-differentiable at u_0 and

$$D(G \circ F)(u_0) = DG(F(u_0))DF(u_0).$$

3 NUMERICAL PRELIMINARIES

3.1 APPROXIMATION OF FUNCTIONS

Approximation of functions, particularly through polynomials, plays a major role in this thesis. Collocation methods (see Section 3.3) aim at discretizing infinite-dimensional problems to finite-dimensional ones, where the solution lies in, e.g., polynomial or piecewise polynomial spaces (see, e.g., [98] as a general reference). Thus, these techniques are strongly reliant on Lagrange interpolation, which will be presented in Subsection 3.1.2 after summarizing, in Subsection 3.1.1, the key ideas behind polynomial best approximation (e.g., [87]).

Moreover, continuation (see Section 3.4) of problems defined by integrals require quadrature formulas to approximate the latter. This is the case of, e.g., the Daphnia model presented in Section 1.4. Subsections 3.1.3 and 3.1.4 will describe the main features of two different sets of nodes that will be referred to in this thesis, and are used as collocation or quadrature nodes. Finally, Subsection 3.1.5 features an example of interpolatory quadrature formula.

3.1.1 Best approximation

This Subsection lists some basic results on best polynomial approximation, which plays an important role in estimating the interpolation error, as it will be clear from the following Subsections, starting from Theorem 3.6.

Definition 3.1 ([87, Example I.1])**.** For a given $f \in C([a,b],\mathbb{R}^d)$ and $n \in \mathbb{N}$, the *best approximation* of f in $[a,b]$ of degree n is a polynomial $p_n^* \in \Pi_n$ which satisfies

$$\|f - p_n^*\|_\infty \leq \|f - p_n\|_\infty$$

for all $p_n \in \Pi_n$.

A preliminary bound on the best approximation error $E_n(f) := \|f - p_n^*\|_\infty$ can be obtained by the following theorem.

Theorem 3.2 (Taylor's theorem, [96, Theorem 1 at page 409])**.** *For a given degree n, let $f \in C^{n-1}([a,b],\mathbb{R}^d)$ be such that $f^{(n)}$ is defined in (a,b). Let $\alpha \in (a,b)$ and consider the polynomial*

$$p(x) = \sum_{k=0}^{n-1} \frac{f^{(k)}(\alpha)}{k!}(x-\alpha)^k.$$

Then

$$\lim_{x\to\alpha} \frac{f(x) - p(x)}{(x-\alpha)^n} = 0.$$

Recall the definition of *modulus of continuity* ω of f on $[a,b]$, given by

$$\omega(\delta; f) := \sup_{\substack{x_1,x_2\in[a,b]\\ |x_1-x_2|\leq\delta}} |f(x_1) - f(x_2)|.$$

The main result on best approximation of continuous functions is the following Jackson's theorem.

Theorem 3.3 ([87, Corollary 1.4.1], Jackson's theorem)**.** *If $f \in C([a,b],\mathbb{R}^d)$, then*

$$E_n(f) \leq 6\omega\left(\frac{b-a}{2n}; f\right).$$

It follows from Jackson's theorem that if $f \in C([a,b],\mathbb{R}^d)$ is Lipschitz continuous, then

$$E_n(f) \leq \frac{C}{n},$$

for some constant C independent of n. The following result is a generalization.

Corollary 3.4 ([87, Theorem 1.5])**.** *If $f \in C^k([a,b],\mathbb{R}^d)$ and $n > k$, then*

$$E_n(f) \leq \frac{6}{n(n-1)\cdots(n-k+1)}\omega\left(\frac{b-a}{2(n-k)}; f^{(k)}\right).$$

Moreover, if $f^{(k)}$ is Lipschitz continuous, then

$$E_n(f) \leq \frac{C_k}{n^{k+1}},$$

for some constant C_k independent of n.

3.1.2 Lagrange interpolation

Consider a set of points $\{x_i\}_{0\leq i\leq n}$ in some interval $[a,b] \in \mathbb{R}$, and a set of corresponding values $\{y_i\}_{0\leq i\leq n} \subset \mathbb{R}^d$. The unique polynomial p of degree up to n interpolating the values $\{y_i\}_{0\leq i\leq n}$ at the nodes $\{x_i\}_{0\leq i\leq n}$, i.e., such that $p(x_i) = y_i$ for $i = 0,\ldots,n$, can be expressed as a linear combination of the *Lagrange coefficients*

$$\ell_i(x) := \prod_{\substack{j=0\\ j\neq i}}^{n} \frac{x - x_j}{x_i - x_j}, \qquad 0 \leq i \leq n,$$

defined so that $\ell_i(x_j) = 1$ if and only if $i = j$, and $\ell_i(x_j) = 0$ otherwise. Thus, from the uniqueness of p, and the fact that $p(x_i) = y_i$ for all $i = 0,\ldots,n$, it is straightforward to check that the Lagrange form of p is given by

$$p(x) = \sum_{i=0}^{n} y_i \ell_i(x). \tag{3.1}$$

A continuous function defined in $[a,b]$ can be discretized by a polynomial as follows. To the set of nodes $\{x_i\}_{0\leq i\leq n}$ one can associate the *Lebesgue interpolation operator* $\mathcal{L}_n : C([a,b],\mathbb{R}^d) \to \Pi_n$ defined by

$$(\mathcal{L}_n f)(x) := \sum_{i=0}^{n} f(x_i)\ell_i(x).$$

The uniform norm of such operator is bounded by the *Lebesgue constant*

$$\Lambda_n := \max_{x\in[a,b]} \sum_{i=0}^{n} |\ell_i(x)|,$$

to which the following theorem gives a lower bound.

Theorem 3.5 ([87, Theorem 4.2]). *For any choice of interpolation nodes,*

$$\Lambda_n > \frac{4}{\pi^2}\log(n+1) - 1.$$

The following theorem, on the other hand, provides an upper bound on the interpolation error and underlines the roles of both $E_n(f)$, which depends exclusively on the relugarity of f, and Λ_n, which depends exclusively on the choice of the interpolation nodes. Note that the former goes to 0 as n goes to infinity, by Theorem 3.3, but the latter does not, by Theorem 3.5.

Theorem 3.6 ([87, Theorem 4.1]). *If $f \in C([a,b],\mathbb{R}^d)$, then*

$$\|f - \mathcal{L}_n f\|_\infty \leq (1+\Lambda_n)E_n(f)$$

for any choice of interpolation nodes.

Another theorem that provides such an upper bound without making use of the best approximation, but at the price of more regularity, is the following.

Theorem 3.7 (Cauchy interpolation remainder, [67, Section 6.1, Theorem 2]). *Let $f \in C^{n+1}([a,b],\mathbb{R}^d)$, and p be the polynomial interpolating f at $n+1$ distinct points $x_0,\dots,x_n$ in the interval $[a,b]$. Then, for each $x \in [a,b]$ there exists $\xi_x \in (a,b)$ such that*

$$f(x) - p(x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi_x)\pi(x).$$

where $\pi(x) = \prod_{i=0}^{n}(x - x_i)$ is the so-called nodal polynomial.

The ultimate negative result on polynomial interpolation is given by the following theorem, stating that continuity is not enough to get convergence of the interpolation process.

Theorem 3.8 (Faber's Theorem, [56]). *For any interpolation scheme, there exists $f \in C([a,b])$ for which the interpolation process is not convergent.*

Moving to the numerical aspects of the Lagrange polynomial interpolation, the standard method to implement it is given by the barycentric formula [19].

Consider the nodal polynomial

$$\pi(x) := \prod_{i=0}^{n}(x - x_i).$$

The *barycentric weights*, defined for $i = 0,\dots,n$, are given by

$$w_i := \frac{1}{\prod_{j\neq i}(x_i - x_j)}$$

and allow to rewrite p as

$$p(x) = \pi(x) \sum_{i=0}^{n} \frac{w_i}{x - x_i} y_i. \tag{3.2}$$

Manipulating the interpolating polynomials through the first form of the barycentric interpolation formula (3.2) is more efficient than using (3.1) directly, since the barycentric weights can be computed once and for all - for a given set of nodes - and do not depend on either the values $\{y_i\}_{0 \leq i \leq n}$ or the point at which p will have to be evaluated. Their computation costs $O(n^2)$, and subsequent evaluations of p cost $O(n)$ each. Moreover, adding a new interpolation node to the old ones also costs $O(n)$.

The barycentric interpolation formula can be also expressed in another form, which improves the stability properties of the evaluation $p(x)$ whenever $x \approx x_i$ for some i. Based on the identity

$$1 = \pi(x) \sum_{i=0}^{n} \frac{w_i}{x - x_i},$$

which comes straightforward from the definition of the quantities w_i, the second form of the barycentric interpolation formula reads

$$p(x) = \frac{\displaystyle\sum_{i=0}^{n} \frac{w_i}{x - x_i} y_i}{\displaystyle\sum_{i=0}^{n} \frac{w_i}{x - x_i}}.$$

3.1.3 Chebyshev nodes

Chebyshev polynomials are a family of orthogonal polynomials, i.e., such that any pair of distinct polynomials in the sequence are orthogonal according to the inner product

$$(p, q)_w := \int_a^b w(x) p(x) q(x) \, \mathrm{d}x,$$

where $[a, b]$ is an interval and $w \in C([a, b], [0, +\infty))$ is a weight function. In particular, the family of Chebyshev polynomials is defined on $[-1, 1]$ as

$$T_n(x) := \cos(n \arccos(x)), \qquad n \in \mathbb{N}.$$

The Chebyshev zeros of order n, also known as Chebyshev nodes of the first kind, are the zeros of T_n, namely

$$x_i := \cos\left(\frac{(2i - 1)\pi}{2n}\right), \qquad i \in \{1, \ldots, n\}.$$

The Chebyshev extrema of order n are the extremal points of T_n, namely

$$y_i := \cos\left(\frac{i\pi}{n}\right), \qquad i \in \{0, \ldots, n\}.$$

Both sets of nodes correspond to the x coordinates of equally spaced points on a semicircle. It is straightforward to define Chebyshev nodes in any interval $[a, b]$ through a change of variable:

$$x \mapsto \frac{b-a}{2}x + \frac{a+b}{2} \in [a, b], \qquad x = x_1, \dots, x_n.$$

The following properties of Chebyshev nodes constitute the key results on the convergence of Lagrange interpolation.

Theorem 3.9 ([87, Theorem 4.5]). *Let Λ_n be the Lebesgue constant relative to the Chebyshev zeros. Then*

$$\Lambda_n < \frac{2}{\pi}\log(n+1) + 4.$$

Theorem 3.10 ([50, Satz 4]). *Let Λ_n be the Lebesgue constant relative to the Chebyshev extrema. Then*

$$\begin{cases} \Lambda_n < \frac{2}{\pi}\log n + 4, & n \text{ odd}, \\ \Lambda_n < \frac{2}{\pi}\log n + 4 - \alpha_n, & 0 < \alpha_n < \frac{1}{n^2}, n \text{ even}, \end{cases}$$

Finally, the following holds.

Theorem 3.11. *If $f \in C([a, b], \mathbb{R}^d)$ is Lipschitz continuous, then interpolation on Chebyshev nodes is convergent.*

Proof. From the corollary of Jackson's theorem (Theorem 3.3) and Theorem 3.9 (or Theorem 3.10), we obtain

$$\|f - \mathcal{L}_n f\|_\infty = O\left(\frac{\log n}{n}\right).$$

□

Using Corollary 3.4, the following generalization is straightforward.

Corollary 3.12. *If $f \in C^k([a, b], \mathbb{R}^d)$, and f^k is Lipschitz continuous, then*

$$\|f - \mathcal{L}_n f\|_\infty = O\left(\frac{\log n}{n^{k+1}}\right).$$

In particular, polynomial interpolation on Chebyshev nodes has infinite order of convergence on arbitrarily smooth functions.

The following theorem gives a bound on the derivative of the interpolation error.

Theorem 3.13. *Let $f \in C^k([-1, 1], \mathbb{R}^d)$. Then, for $i = 0, \dots, k$ and $q = i, \dots, k$, the interpolation on Chebyshev zeros satisfies*

$$\|f^{(i)} - (\mathcal{L}_n f)^{(i)}\|_\infty \leq c\frac{E_{n-q}(f^{(q)})}{n^{q-i}}\log(n+1).$$

Proof. It follows directly from [80, Theorem 4.2.11] with $\alpha = \beta = \frac{1}{2}, r = s = 1$. □

3.1.4 Gauss-Legendre nodes

Legendre polynomials constitute another family of orthogonal polynomials. They are defined on $[-1,1]$ from the constant weight function $w = 1$ and can be expressed by

$$P_n(x) := \frac{(-1)^n}{2^n n!}\frac{d^n}{dx^n}(1-x^2)^n, \qquad n \in \mathbb{N}\setminus\{0\}.$$

The Gauss-Legendre nodes of order n are the zeros of P_n, and are the nodes used in Gauss formulae, quadrature formulae that maximize the degree of precision, i.e., the maximum degree d such that the monomial x^d is integrated exactly.

Theorem 3.14 ([62, Theorem 1, pag. 327]). *The maximum degree of precision of a quadrature formula using $n+1$ nodes is $2n+1$. The nodes of such formula are the zeros of the $(n+1)$-th orthogonal polynomial with respect to the weight $w(x) = 1$.*

3.1.5 Clenshaw-Curtis quadrature

The Clenshaw-Curtis formula (described, e.g., in [98]) is an interpolatory quadrature formula, i.e., given by

$$S_{n+1}(f) := \sum_{i=0}^{n} w_i f(y_i),$$

such that

$$w_i = \int_a^b \ell_i(x)\,\mathrm{d}x, \quad i \in \{0,\dots,n\}$$

where $\{\ell_i\}_{0\le i\le n}$ is the Lagrange basis associated to the quadrature nodes $\{y_i\}_{0\le i\le n}$. In the case of Clenshaw-Curtis, $[a,b] = [-1,1]$, the quadrature nodes are the Chebyshev extrema, and the weights can be computed explicitly as described in [98, Chapter 12, program `clencurt.m`].

The Clenshaw-Curtis quadrature has an infinite order of convergence for smooth integrands [99, Theorem 4.5].

3.2 LINEAR OPERATORS

This section includes some essential results on linear operators on Banach spaces, particularly on their approximation. They will be needed in some of the convergence proofs in Chapter 6.

Theorem 3.15 (Banach perturbation lemma, [83, Theorem 2.1.1]). *Let $\mathbb{B}$ be a Banach space and $L, \hat{L}$ be linear operators on $\mathbb{B}$ such that L is invertible and $\|L^{-1}(L-\hat{L})\|_{\mathbb{B}\leftarrow\mathbb{B}} < 1$. Then $\hat{L}$ is invertible and*

$$\|\hat{L}^{-1}\|_{\mathbb{B}\leftarrow\mathbb{B}} \le \frac{\|L^{-1}\|_{\mathbb{B}\leftarrow\mathbb{B}}}{1-\|L^{-1}(\hat{L}-L)\|_{\mathbb{B}\leftarrow\mathbb{B}}}.$$

Theorem 3.15 is a powerful tool in a context where there is a sequence $\{L_n\}_{n\in\mathbb{N}}$ of linear operators such that

$$\|L_n - L\|_{\mathbb{B}\leftarrow\mathbb{B}} \to 0, \quad n\to\infty.$$

Indeed, the above implies that there exists $\overline{n} \in \mathbb{N}$ such that, for all $n > \overline{n}$, the hypothesis $\|(L_n - L)L^{-1}\|_{\mathbb{B}\leftarrow\mathbb{B}} < 1$ is satisfied. Thus, invertibility of L leads to invertibility of L_n for sufficiently large n.

The lemma below, also concerning converging sequences of linear operators, gives an estimate on the convergence order of the corresponding eigenvalues and eigenfunctions.

Lemma 3.16. *Let $\mathbb{B}$ be a Banach space, L a linear and bounded operator on $\mathbb{B}$, and $\{L_n\}_{n\in\mathbb{N}}$ a sequence of linear and bounded operators such that*

$$\|L_n - L\|_{\mathbb{B}\leftarrow\mathbb{B}} \to 0, \quad n \to \infty.$$

Assume that μ is an eigenvalue of L with finite algebraic multiplicity ν, ascent l and eigenfunction φ normalized as $\|\varphi\|_{\mathbb{B}} = 1$, and that there is $r > 0$ such that μ is the only eigenvalue of L in $B(\mu, r) \subset \mathbb{C}$. Then there exists a positive integer N such that, for every $n \geq N$, L_n has exactly ν eigenvalues $\mu_{n,j}$, $j \in \{1, \dots, \nu\}$ (counted with their multiplicities) in $B(\mu, r)$ and, moreover, if M_μ is the generalized eigenspace of μ and $\epsilon_n := \|(L_n - L) \restriction_{M_\mu} \|_{\mathbb{B}\leftarrow M_\mu}$, then

$$\max_{j\in\{1,\dots,\nu\}} |\mu_{n,j} - \mu| = O(\epsilon_n^{1/l}), \qquad \max_{j\in\{1,\dots,\nu\}} \|\varphi_{n,j} - \varphi\|_{\mathbb{B}} = O(\epsilon_n^{1/l}),$$

where $\varphi_{n,j}$ is the eigenfunction associated to $\mu_{n,j}$ normalized as $\|\varphi_{n,j}\|_{\mathbb{B}} = 1$.

Proof. By [34, Example 3.8 and Theorem 5.22], for all μ in the resolvent set of L the strongly stable convergence $L_n - \mu \xrightarrow{ss} L - \mu I$ holds. By [34, Theorem 5.6], this implies the existence of $N \in \mathbb{N}$ satisfying the hypotheses in the statement. Finally, the results on the order of convergence follow from [34, Theorem 6.7]. □

3.3 COLLOCATION METHODS

Collocation methods are numerical methods to find an approximation of the solution of nonlinear functional equations. In particular, they are commonly used to solve several kinds of integro-differential equations (see, e.g., [32, 33, 60, 98]). They require to choose a finite-dimensional subspace of the solutions space, which plays the role of the space of candidate (approximate) solutions, and a finite set of *collocation points* (or *nodes*) in the domain of the relevant equation. Such choices must be made so that there exists a *unique* element of the finite-dimensional subspace which satisfies the original equation at the collocation points, taking into account for possible initial or boundary conditions. The goal is, indeed, to find such unique element, which will be the numerical solution.

Typical examples for the choice of the finite-dimensional subspace are the polynomial space Π_n (for some fixed n) or the piecewise polynomial space

$$\Pi_n^{\mathtt{t}} := \left\{ p \in C(\mathbb{R}, \mathbb{R}^d) \,\middle|\, \forall i < m \; p|_{[t_i, t_{i+1}]} \in \Pi_n \right\},$$

where the *mesh* $\mathtt{t} := (t_0, \dots, t_m)$ satisfies $t_0 < \cdots < t_m$. The mesh can be *adaptive*, i.e., not fixed but rather varying according to some information on

the problem which was acquired, e.g., during previous computations (see, e.g., [89]). Adaptivity can be costly to implement but sometimes necessary: higher resolution can be required to deal with difficult regions (e.g., containing discontinuities or steep gradients).

In order to show how collocation works in practice when differentiation is involved, the following IVP for ODEs can be considered as a didactic example. Given

$$\begin{cases} u'(t) = g(u(t)), & t \in [a,b], \\ u(a) = u_0 \end{cases}$$

and $n > 0$, the polynomial space of degree up to n can be chosen as the space of candidate solutions. Moreover, let $a = \theta_0 < \cdots < \theta_n = b$ be the collocation nodes. Then, the collocation problem consists in finding a polynomial p of degree n satisfying

$$\begin{cases} p'(\theta_i) = g(p(\theta_i)), & i = 1, \dots, n, \\ p(\theta_0) = u_0. \end{cases}$$

Such p can be obtained, as described below, using the *differentiation matrix* $D_n := D_n(\theta_0, \dots, \theta_n)$ of the collocation nodes (see, e.g., [98, 104]), which transforms a vector of data at the collocation points into approximate derivatives (of the interpolating function) at those points. It is straightforward to check that such a matrix must be given by

$$D_n = \begin{pmatrix} d_{0,0} & d_{0,1} & \dots & d_{0,n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n,0} & d_{n,1} & \dots & d_{n,n} \end{pmatrix} := \begin{pmatrix} \ell_0'(\theta_0) & \ell_1'(\theta_0) & \dots & \ell_n'(\theta_0) \\ \vdots & \vdots & \ddots & \vdots \\ \ell_0'(\theta_n) & \ell_1'(\theta_n) & \dots & \ell_n'(\theta_n) \end{pmatrix}$$

with $\{\ell_i\}_{0 \le i \le n}$ the Lagrange basis associated to the collocation nodes, defined as in Subsection 3.1.2.

In order to conclude that p is unique, and therefore the collocation method above is well defined, it can be observed that p is the n-degree polynomial interpolating the values u_i, $i = 0, \dots, n$, defined from the system

$$A \begin{pmatrix} u_0 \\ u_1 \\ \vdots \\ u_n \end{pmatrix} = \begin{pmatrix} u_0 \\ g(u_1) \\ \vdots \\ g(u_n) \end{pmatrix},$$

where A is obtained after editing the first row in order to force $p(\theta_0) = u_0$, that is

$$A = \begin{pmatrix} 1 & 0 & \dots & 0 \\ d_{1,0} & d_{1,1} & \dots & d_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n,0} & d_{n,1} & \dots & d_{n,n} \end{pmatrix}.$$

3.3.1 Collocation error

Collocation methods can be used to obtain better and better approximations of the solutions of interest, by increasing the discretization level.

In the case of polynomial collocation, the discretization level is given by the maximum degree n which defines the space Π_n of candidate solutions. As a consequence of Corollary 3.12, the resulting method exhibits *spectral accuracy* whenever the approximand is smooth, i.e., the error decays faster than $O(n^{-k})$ for any integer k.

In the case of piecewise polynomial collocation, the discretization level is determined by both the degree n and the number m of mesh intervals. As a result, there are two possibilities:

- the *Spectral Element Method* (SEM), where the mesh $\mathtt{t}$ is fixed and the degree n goes to infinity;
- the *Finite Element Method* (FEM), where n is fixed and m goes to infinity.

SEM is a straightforward generalization of the spectral collocation using a single polynomial. Thus, Corollary 3.12 holds, and the regularity of the approximand (or better, its lack of regularity) gives an upper bound to the order of convergence.

In the FEM case, the Lebesgue constant Λ_n does not vary as m goes to infinity, and a bound for the order of convergence is given by the Cauchy remainder formula (Theorem 3.7) in $[t_i, t_{i+1}]$. Thus, if $f \in C^{n+1}([a,b], \mathbb{R}^d)$, then

$$\|f - \mathcal{L}_n f\|_\infty = O(h^{n+1}),$$

where $h := \max_{0 \le i < m} |t_{i+1} - t_i|$.

3.4 NUMERICAL CONTINUATION

Numerical continuation is a widely used method in dynamical system theory, in particular to compute equilibria or periodic solutions and analyze their stability and bifurcations under parameter variation. This section, which is based on [46], contains some basic results that establish under what conditions continuation can actually be applied, and then provides the basics of some examples of continuation algorithms, which will be mentioned in Chapter 4.

The content of this section is somehow less standard than the content of the other sections of this chapter. For this reason, the proofs of most of the relevant results will be provided as well.

Let $G : \mathbb{R}^{n+1} \to \mathbb{R}^n$. We consider the problem

$$G(v) = 0. \tag{3.3}$$

Since continuation methods are applied in a parameter-dependent setting, the problem (3.3) is usually described with an explicit *continuation parameter* $\lambda \in \mathbb{R}$, and can therefore be formulated as

$$G(u, \lambda) = 0, \tag{3.4}$$

where $u \in \mathbb{R}^n$. In principle, however, any component of v can serve as a parameter.

In particular, we are interested in identifying *solution branches*, i.e., one-dimensional continua of solutions parametrized by λ, defined by

$$u = u(\lambda) \Leftrightarrow G(u, \lambda) = 0.$$

Numerical continuation aims, indeed, at approximating a branch through a sequence of points $\{(u_i, \lambda_i)\}_{i \in \mathbb{N}}$. The algorithms that implement it are largely based on the implicit function theorem.

Theorem 3.17 (Implicit function theorem, e.g., [88, Theorem 9.28]). *Let $G : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ satisfy*

- *$G(u_0, \lambda_0) = 0$ for some $u_0 \in \mathbb{R}^n$ and $\lambda_0 \in \mathbb{R}^m$;*
- *$G_u(u_0, \lambda_0)$ is nonsingular;*
- *there is $\rho > 0$ such that G and G_u are Lipschitz continuous in $\mathcal{B}((u_0, \lambda_0), \rho)$, where $\mathbb{R}^n \times \mathbb{R}^m$ is equipped with the maximum norm.*

Then, there exists $0 < \delta \leq \rho$ such that there is a unique function $u(\lambda)$ defined in $\mathcal{B}(\lambda_0, \delta)$ such that $u(\lambda_0) = u_0$ and

$$G(u(\lambda), \lambda) = 0$$

for all $\lambda \in \mathcal{B}(\lambda_0, \delta)$. Moreover, $u(\lambda)$ is uniformly continuous in $\mathcal{B}(\lambda_0, \delta)$.

$(\overline{u}, \overline{\lambda})$ is a *regular* solution if $G_{(u,\lambda)}(\overline{u}, \overline{\lambda})$ has maximal rank.

Theorem 3.18 (Existence of a solution branch, [46, Theorem 5]). *Let $G : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$. Let $v_0 := (u_0, \lambda_0)$ be a regular solution of G. Assume that there is $\rho > 0$ such that G and G_u are Lipschitz continuous in $\mathcal{B}(v_0, \rho)$. Then there exists $0 < \delta \leq \rho$ such that $\mathcal{B}(v_0, \delta)$ contains a unique solution branch v such that $v(0) = v_0$.*

Proof. Since $G_{(u,\lambda)}(v_0)$ has maximal rank, then either $G_u(v_0)$ has in turn maximal rank, or there exists $i \leq n$ such that the matrix obtained from $G_{(u,\lambda)}(v_0)$ by removing the i-th column does. In the latter case, we can rearrange the columns of $G_{(u,\lambda)}(v_0)$ and consistently change parametrization, in order to fall in the former case.
Therefore, $G_u(v_0)$ has a bounded inverse, the hypotheses of the implicit function theorem are satisfied, and the thesis follows. □

The following results concern the regularity of the solution branch.

Lemma 3.19 (Banach lemma, [46, Lemma 1]). *Let $\mathbb{B}$ be a Banach space and $L : \mathbb{B} \to \mathbb{B}$ a linear operator such that $\|L\|_{\mathbb{B} \leftarrow \mathbb{B}} < 1$. Then the operator $I + L$ is invertible and*

$$\|(I + L)^{-1}\|_{\mathbb{B} \leftarrow \mathbb{B}} \leq \frac{1}{1 - \|L\|_{\mathbb{B} \leftarrow \mathbb{B}}}.$$

Proof. If there existed $b \in \mathbb{B} \setminus \{0\}$ such that $(I + L)b = 0$, then we would have

$$\|b\|_{\mathbb{B}} = \|Lb\|_{\mathbb{B}} \leq \|L\|_{\mathbb{B} \leftarrow \mathbb{B}} \|b\|_{\mathbb{B}} < \|b\|_{\mathbb{B}},$$

contradiction. Therefore, $I + L$ is invertible. From $(I + L)(I + L)^{-1} = I$ we get

$$(I + L)^{-1} = I - L(I + L)^{-1},$$

and therefore

$$\|(I+L)^{-1}\|_{\mathbb{B}\leftarrow\mathbb{B}} \leq 1 + \|L\|_{\mathbb{B}\leftarrow\mathbb{B}}\|(I+L)^{-1}\|_{\mathbb{B}\leftarrow\mathbb{B}},$$

from which the bound follows. □

Lemma 3.20 ([46, Lemma 2]). *Let $G : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ satisfy the conditions of the implicit function theorem (Theorem 3.17). Then there exists $\delta > 0$ such that, for all $(u,\lambda) \in \mathcal{B}((u_0,\lambda_0),\delta)$, $G_u(u,\lambda)$ is nonsingular and $\|G_u^{-1}\|_{\mathbb{R}^n\leftarrow\mathcal{B}((u_0,\lambda_0),\delta)}$ is bounded.*

Proof. Let $L : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ be given by

$$L(u,\lambda) := (G_u(u_0,\lambda_0))^{-1}(G_u(u,\lambda)u - G(u_0,\lambda_0)).$$

Let $M := \|G_u(u_0,\lambda_0)^{-1}\|_{\mathbb{R}^m\times\mathbb{B}}$ and let K be the Lipschitz constant of both G and G_u in $\mathcal{B}((u_0,\lambda_0),\rho)$. Let $\delta := \frac{1}{4MK}$. Then, for all $(u,\lambda) \in \mathcal{B}((u_0,\lambda_0),\delta)$,

$$\begin{aligned}\|L(u,\lambda)\|_{\mathbb{B}} \leq M\|G_u(u,\lambda) - G_u(u_0,\lambda_0)\|_{\mathbb{B}} &\leq MK(\|u-u_0\|_{\mathbb{B}} + \|\lambda-\lambda_0\|)\\ &\leq 2MK\delta < 1.\end{aligned}$$

Thus, $\|L\|_{\mathbb{R}^n\leftarrow\mathcal{B}((u_0,\lambda_0),\delta)} < 1$. Therefore, by the Banach lemma (Lemma 3.19), we have

$$\|(I+L)^{-1}\|_{\mathbb{R}^n\leftarrow\mathcal{B}((u_0,\lambda_0),\delta)} \leq \frac{1}{1-\|L\|_{\mathbb{R}^n\leftarrow\mathcal{B}((u_0,\lambda_0),\delta)}}.$$

For all $u \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$,

$$G_u(u,\lambda) = G_u(u_0,\lambda_0) + G_u(u,\lambda) - G_u(u_0,\lambda_0) = G_u(u_0,\lambda_0)(I + L(u,\lambda)).$$

Therefore,

$$\|G_u^{-1}\|_{\mathbb{R}^n\leftarrow\mathcal{B}((u_0,\lambda_0),\delta)} \leq M\|1+L\|_{\mathbb{R}^n\leftarrow\mathcal{B}((u_0,\lambda_0),\delta)} \leq \frac{M}{1-\|L\|_{\mathbb{R}^n\leftarrow\mathcal{B}((u_0,\lambda_0),\delta)}},$$

which means in particular that $\|G_u^{-1}\|_{\mathbb{R}^n\leftarrow\mathcal{B}((u_0,\lambda_0),\delta)}$ is bounded. □

Theorem 3.21 (Differentiability of a solution branch, [46, Theorem 3]). *Let $G : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ satisfy the conditions of the implicit function theorem (Theorem 3.17). Moreover, assume that $G_\lambda(u,\lambda)$ is continuous in $\mathcal{B}((u_0,\lambda_0),\delta)$, where δ is defined in Lemma 3.20. Then, $u(\lambda)$ has a continuous derivative in $\mathcal{B}((u_0,\lambda_0),\delta)$.*

Proof. Let $h_1 : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}$ and $h_2 : \mathbb{R}^n \times \mathbb{R} \times \mathbb{R}$ be defined by

$$h_1(u_1,u_2,\lambda) := G(u_1,\lambda_1) - G(u_2,\lambda_1) - G_u(u_1,\lambda_1)(u_1-u_2),$$

and

$$h_2(u_1,\lambda_1,\lambda_2) := G(u_1,\lambda_1) - G(u_1,\lambda_2) - G_\lambda(u_1,\lambda_1)(\lambda_1-\lambda_2).$$

From the hypotheses, it follows that, for all $(u_1,\lambda_1) \in \mathcal{B}((u_0,\lambda_0),\delta)$,

$$\frac{|h_1(u_1,u,\lambda_1)|}{|u_1-u|} \to 0 \text{ as } |u_1-u| \to 0$$

and

$$\frac{|h_2(u_1, \lambda_1, \lambda)|}{|\lambda_1 - \lambda|} \to 0 \text{ as } |\lambda_1 - \lambda| \to 0. \tag{3.5}$$

Let $\lambda_2 \in \mathcal{B}(\lambda_0, \delta)$. Then

$$\begin{aligned} 0 &= G(u(\lambda_1), \lambda_1)) - G(u(\lambda_2), \lambda_2) \\ &= G(u(\lambda_1), \lambda_1)) - G(u(\lambda_2), \lambda_1)) + G(u(\lambda_2), \lambda_1) - G(u(\lambda_2), \lambda_2) \\ &= G_u(u(\lambda_1), \lambda_1)(u(\lambda_1) - u(\lambda_2)) + h_1(u(\lambda_1), u(\lambda_2), \lambda_1)) \\ &+ G_\lambda(u(\lambda_2), \lambda_1)(\lambda_1 - \lambda_2) + h_2(u(\lambda_2), \lambda_1, \lambda_2). \end{aligned} \tag{3.6}$$

Let $h : \mathcal{B}(\lambda_0, \delta) \to \mathbb{R}^n$ be defined by

$$\begin{aligned} h(\lambda_1, \lambda_2) := &(G_u(u(\lambda_1), \lambda_1) - G_u(u(\lambda_2), \lambda_2))(\lambda_1 - \lambda_2) \\ &+ h_1(u(\lambda_1), u(\lambda_2), \lambda_1) + h_2(u(\lambda_2), \lambda_1, \lambda_2). \end{aligned}$$

By the continuity of u and G_λ, it follows that

$$\frac{|(G_u(u(\lambda_1), \lambda_1) - G_u(u(\lambda_2), \lambda_2))(\lambda_1 - \lambda_2)|}{|\lambda_1 - \lambda_2|} \to 0 \text{ as } |\lambda_1 - \lambda_2| \to 0.$$

By the uniform continuity of u and (3.5), it follows that

$$\frac{|h_1(u(\lambda_1), u(\lambda_2), \lambda_1)|}{|\lambda_1 - \lambda_2|} = \frac{|h_1(u(\lambda_1), u(\lambda_2), \lambda_1)|}{|u(\lambda_1) - u(\lambda_2)|} \cdot \frac{|u(\lambda_1) - u(\lambda_2)|}{|\lambda_1 - \lambda_2|} \to 0$$

as $|\lambda_1 - \lambda_2| \to 0$. Finally, by (3.5),

$$\frac{|h_2(u(\lambda_1), \lambda_2, \lambda_1)|}{|\lambda_1 - \lambda_2|} \to 0 \text{ as } |\lambda_1 - \lambda_2| \to 0,$$

and thus

$$\frac{|h(\lambda_1, \lambda_2)|}{|\lambda_1 - \lambda_2|} \to 0 \text{ as } |\lambda_1 - \lambda_2| \to 0. \tag{3.7}$$

From the existence of $G_u(u(\lambda_1), \lambda_1)^{-1}$ (Lemma 3.20), (3.6) can be rewritten as

$$u(\lambda_1) - u(\lambda_2) = -G_u(u(\lambda_1), \lambda_1)^{-1}(G_\lambda(u(\lambda_1), \lambda_1)(\lambda_1 - \lambda_2) - h(\lambda_1, \lambda_2)). \tag{3.8}$$

$\|G_u^{-1}\|_{\mathbb{R}^n \leftarrow \mathcal{B}((u_0, \lambda_0), \delta)}$ is bounded by Lemma 3.20, thus, by (3.7),

$$\frac{|G_u(u(\lambda_1), \lambda_1)^{-1} h(\lambda_1, \lambda_2)|}{|\lambda_1 - \lambda_2|} \to 0 \text{ as } |\lambda_1 - \lambda_2| \to 0,$$

which implies, by (3.8), that $-G_u(u(\lambda), \lambda)^{-1} G_\lambda(u(\lambda), \lambda)$ is the derivative of $u(\lambda)$ in $\mathcal{B}((u_0, \lambda_0), \delta)$. Therefore, its continuity follows from the continuity of $-G_u(u(\lambda), \lambda)^{-1}$, which, in turn, follows from the inequality

$$\begin{aligned} &|G_u(u(\lambda_1), \lambda_1)^{-1} - G_u(u(\lambda_2), \lambda_2)^{-1}| \\ &= |G_u(u(\lambda_1), \lambda_1)^{-1}(G_u(u(\lambda_2), \lambda_2) - G_u(u(\lambda_1), \lambda_1)) G_u(u(\lambda_2), \lambda_2)^{-1}| \\ &\leq K \|G_u^{-1}\|^2_{\mathbb{R}^n \leftarrow \mathcal{B}((u_0, \lambda_0), \delta)} (|u(\lambda_1) - u(\lambda_2)| + |\lambda_1 - \lambda_2|), \end{aligned}$$

where K is the Lipschitz constant of G_u in $\mathcal{B}((u_0, \lambda_0), \delta)$. □

The algorithms for numerical continuation compute a sequence of points in the relevant branch using a *predictor-corrector* procedure (see, e.g., [7, 55, 58]). The rest of the section will be dedicated to the description of some examples of such algorithms.

3.4.1 Pseudo-arclength continuation

The algorithm described in this subsection constitutes the foundation for many widely used continuation-based software packages (e.g., AUTO [1], MatCont [4] and XPPAUT [6]) and is described in [46].
Given a known point of the solution branch $v_0 := (u_0, \lambda_0)$ and a length Δs, the prediction for the following point consists in computing the unit vector $\dot{v}_0$ which is tangent to the branch (*Euler tangent prediction*, Figure 3.1, left), and taking a step of length Δs along that vector. The correction consists in computing the hyperplane which is perpendicular to $\dot{v}_0$ and contains v_0, and then looking for the intersection with the relevant branch. This means solving the nonlinear determined system

$$\begin{cases} G(v_1) = 0 \\ (v_1 - v_0)^T \dot{v}_0 - \Delta s = 0. \end{cases} \tag{3.9}$$

The last equation is called *pseudo-arclength condition* and imposes the length of the projection of $v_1 - v_0$ and the tangent vector.

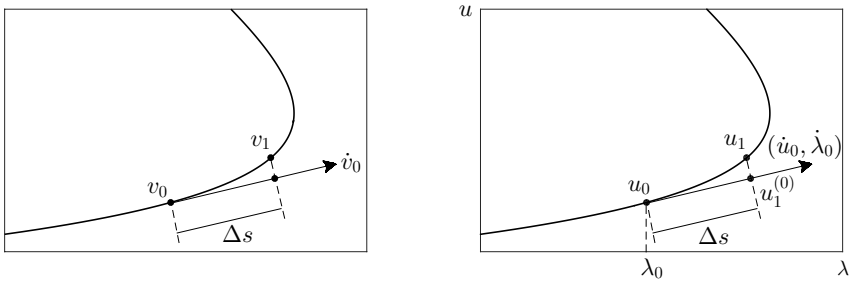

Figure 3.1: Pseudo-arclength continuation (left) with natural parameterization (right). Original figure from [11], courtesy of AIMS.

The method can vary according to the chosen way to solve (3.9), usually a Newton-like method. In particular, using the classical Newton's method, the k-th iteration would read

$$\begin{pmatrix} G_v(v_1^{(k)}) \\ \dot{v}_0^T \end{pmatrix} \Delta v_1^{(k)} = - \begin{pmatrix} G(v_1^{(k)}) \\ (v_1^{(k)} - v_0)^T \dot{v}_0 - \Delta s \end{pmatrix}, \quad k \geq 0, \tag{3.10}$$

with initial guess given by the prediction step

$$v_1^{(0)} = v_0 + \Delta s \, \dot{v}_0$$

and updates

$$v_1^{(k+1)} = v_1^{(k)} + \Delta v_1^{(k)}, \quad k \geq 0.$$

Note that the computation of the new tangent vector requires very little computational effort, once v_1, as well as the Jacobian at v_1, have been computed. Indeed, it is (a unit multiple of) the solution of

$$\begin{pmatrix} G_v(v_1) \\ \dot{v}_0^T \end{pmatrix} \dot{v}_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

where the equation $\dot{v}_0^T \dot{v}_1 = 1$ guarantees the preservation of the orientation of the branch (if Δs is small enough).
The applicability of Newton's method to solve (3.9) is supported by the following theorem

Theorem 3.22. *The Jacobian in system* (3.10) *is nonsingular at a regular solution point.*

Proof. Let v_0 be a regular solution point. The relevant Jacobian matrix is

$$\begin{pmatrix} G_v(v_0) \\ \dot{v}_0^T \end{pmatrix}.$$

From the regularity of v_0, we know that $G_v(v_0)$ has maximal rank and, therefore, nullspace of dimension 1. This means that its nullspace is $\text{span}\{\dot{v}_0\}$. Assume for a contradiction that there exists $z \neq 0$ in the null space of the Jacobian. This would imply $G_v(v_0)z = 0$, from which we get $z = c\dot{v}_0$ for some $c \in \mathbb{R}$. But it also implies

$$c = c\|\dot{v}_0\|^2 = cz^T z = \dot{v}_0^T z = 0,$$

contradiction. □

Other Newton-like methods can be used, for instance the Broyden's update [31], which might reduce the computational effort by avoiding the calculation of the Jacobian matrix in (3.10) at each iteration. Indeed, at the k-th iteration for $k > 1$, $G_v(v_1^{(k)})$ is approximated using the secant equation in the finite-difference approximation, i.e.,

$$G_v(v_1^{(k)}) \approx G_v(v_1^{(k-1)}) + \frac{G(v_1^{(k)}) - G(v_1^{(k-1)}) - G_v(v_1^{(k-1)})\Delta v_1^{(k-1)}}{\|\Delta v_1^{(k-1)}\|^2} \cdot \left(\Delta v_1^{(k-1)}\right)^T.$$

As mentioned at the beginning of section 3.4, the pseudo-arclength continuation can be formulated according to different parametrizations. In the case of (3.4) we talk about *natural parameterization* and we look for the solution branch $(u(\lambda), \lambda)$ or, basically, for $u(\lambda)$. Then (3.9) becomes

$$\begin{cases} G(u_1, \lambda_1) = 0 \\ (u_1 - u_0)^T \dot{u}_0 + (\lambda_1 - \lambda_0)\dot{\lambda}_0 - \Delta s = 0, \end{cases}$$

Figure 3.1 (right). The k-th iteration of Newton's method reads

$$\begin{pmatrix} G_u(u_1^{(k)}, \lambda_1^{(k)}) & G_\lambda(u_1^{(k)}, \lambda_1^{(k)}) \\ \dot{u}_0^T & \dot{\lambda}_0 \end{pmatrix} \begin{pmatrix} \Delta u_1^{(k)} \\ \Delta \lambda_1^{(k)} \end{pmatrix} = - \begin{pmatrix} G(u_1^{(k)}, \lambda_1^{(k)}) \\ (u_1^{(k)} - u_0)^T \dot{u}_0 + (\lambda_1^{(k)} - \lambda_0)\dot{\lambda}_0 - \Delta s \end{pmatrix},$$

with initial prediction

$$\begin{pmatrix} u_1^{(0)} \\ \lambda_1^{(0)} \end{pmatrix} = \begin{pmatrix} u_0 \\ \lambda_0 \end{pmatrix} + \Delta s \begin{pmatrix} \dot{u}_0 \\ \dot{\lambda}_0 \end{pmatrix}$$

and updates

$$\begin{pmatrix} u_1^{(k+1)} \\ \lambda_1^{(k+1)} \end{pmatrix} = \begin{pmatrix} u_1^{(k)} \\ \lambda_1^{(k)} \end{pmatrix} + \begin{pmatrix} \Delta u_1^{(k+1)} \\ \Delta\lambda_1^{(k+1)} \end{pmatrix}, \quad k \geq 0.$$

The new tangent vector is obtained by normalizing the solution of

$$\begin{pmatrix} G_u(u_1, \lambda_1) & G_\lambda(u_1, \lambda_1) \\ \dot{u}_0^T & \dot{\lambda}_0 \end{pmatrix} \begin{pmatrix} \dot{u}_1 \\ \dot{\lambda}_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

3.4.2 Natural continuation

Natural continuation is based on a simpler idea. Given a known point of the solution branch and a fixed real number $\Delta\lambda$, the predictor computes the direction of the vector which is tangent to the branch and takes a step along that vector so that the continuation parameter increases by $\Delta\lambda$. The correction consists in computing the hyperplane which is perpendicular to the direction orthogonal to the one of the continuation parameter and contains the predicted point, and then looking for the intersection with the relevant branch, Figure 3.2 (left). The method can be further simplified by substituting the prediction along the tangent vector with that along the secant through the two preceding steps, Figure 3.2 (right).

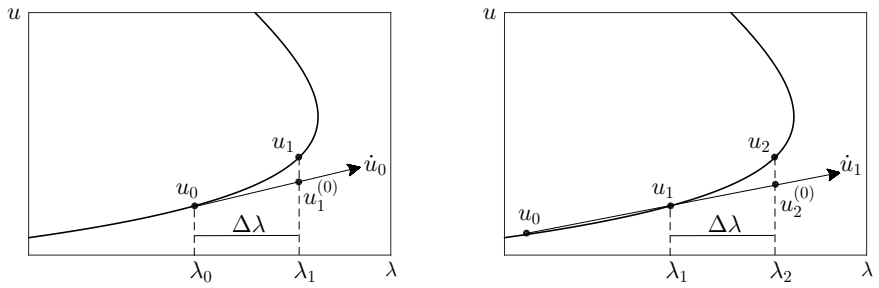

Figure 3.2: Natural continuation with tangent (left) and secant (right) prediction. Original figure from [11], courtesy of AIMS.

Nevertheless, natural continuation is less commonly used than pseudo-arclength continuation since the former may fail close to *fold bifurcations* in the solution branch (see, e.g., [72, Sections 3.2 and 3.3]), i.e., in the latter case mentioned in the proof of Theorem 3.18. Indeed, if that holds, then the unique solution branch cannot be parametrized by λ.

3.5 DISCRETIZATION OF FIXED POINT PROBLEMS

This section describes some general notions and principles concerning the discretization of an infinite-dimensional problem, which will play a role in Chapter 6. Indeed, many of the proofs contained therein are rather technical. Thus, it might be worth to outline here the relevant discretization theory principles separately, in order to better understand where in the general framework each technical result fits.

In particular, the focus will be on fixed point problems on Banach spaces, i.e., problems of the form

$$x = \Phi(x) \tag{3.11}$$

for some operator $\Phi : X \to X$, where X is a Banach space. It is assumed that a solution x^* to (3.11) exists and is locally unique.

A discretization scheme for (3.11) is defined by a finite-dimensional subset $\hat{X} \subset X$, equipped with a projection operator $P : X \to \hat{X}$. The resulting *discrete problem* is

$$x = \hat{\Phi}(x), \tag{3.12}$$

where $\hat{\Phi} : \hat{X} \to \hat{X}$ is defined as $P \circ \Phi$. The dimension of $\hat{X}$ corresponds to the level of the discretization.

For a discretization scheme to be meaningful, the problem (3.12) must be (locally) uniquely solvable. Subtracting $x^* = \Phi(x^*)$ to equation (3.12), one obtains

$$\begin{aligned} x - x^* &= \hat{\Phi}(x) - \Phi(x^*) \\ &= \hat{\Phi}(x) - \hat{\Phi}(x^*) + \hat{\Phi}(x^*) - \Phi(x^*) \\ &= P\Phi(x) - P\Phi(x^*) + P\Phi(x^*) - \Phi(x^*) \\ &= P(\Phi(x) - \Phi(x^*)) + P\Phi(x^*) - x^*. \end{aligned} \tag{3.13}$$

The term $\varepsilon_C := P\Phi(x^*) - x^*$ in the right-hand side measures the extent to which the solution of (3.11) satisfies (3.12), and is called the *consistency error*. The scheme is *consistent* if the norm of the consistency error goes to 0 as the discretization level goes to infinity.

The following are classical assumptions in discretization theory (see, e.g., [69, Lemma 19.1]).

Assumption 3.23. The operator $\Phi : X \to X$ has a bounded Fréchet derivative in a neighborhood of the fixed point x^*.

Assumption 3.24. The operator $I - D\Phi(x^*) : X \to X$ has a bounded inverse.

If Assumption 3.23 holds, then (3.13) can be rewritten as

$$x - x^* = P(\Phi(x) - \Phi(x^*) - D\Phi(x^*)(x - x^*)) + PD\Phi(x^*)(x - x^*) + \varepsilon_C,$$

leading to

$$(I - PD\Phi(x^*))(x - x^*) = P(\Phi(x) - \Phi(x^*) - D\Phi(x^*)(x - x^*)) + \varepsilon_C. \tag{3.14}$$

Note that the well-posedeness of the discrete problem (3.12) is equivalent to the existence and (local) uniqueness of $\hat{x} \in \hat{X}$ satisfying (3.14). This is, in turn, equivalent to the existence and uniqueness of the *discretization error* $\hat{x} - x^* \in X$ satisfying (3.14). Given the expression of the left-hand side of the latter, a necessary condition for the sought well-posedeness is the invertibility of the operator $I - PD\Phi(x^*)$. In this regard, the scheme is *stable* if the *stability constant* $\|(I - PD\Phi(x^*))^{-1}\|_{X \leftarrow X}$ is uniformly bounded as the discretization level goes to infinity.

Stability is closely related to Assumption 3.24. Indeed, it can be derived from the latter if, e.g., the hypotheses of the Banach perturbation lemma

(Theorem 3.15) are satisfied. In some cases, however, the derivation is not so immediate (see paragraph after the proof of Proposition 6.18 in Chapter 6).

Note that, if Φ is a linear operator, the right-hand side of (3.14) reduces to ε_C and, in particular, does not contain x. This means that, in this case, the stability of the method is also a sufficient condition for the well-posedeness.

If Φ is not linear, thanks to Assumption 3.23 the right-hand side of (3.14) reduces to $o(\|x - x^*\|_X) + \varepsilon_C$ (recall Definition 2.8), leading to

$$x - x^* = (I - PD\Phi(x^*))^{-1}(h(x) + \varepsilon_C) \tag{3.15}$$

for some $h(x) \in o(\|x - x^*\|_X)$. If, in addition, the bound

$$\|(I - PD\Phi(x^*))^{-1}\|_{X \leftarrow X} \cdot |h(x)| \leq r\|x - x^*\|_X \tag{3.16}$$

holds uniformly for some $r < 1$, then (3.15) gives

$$\|x - x^*\|_X = \frac{1}{1-r}\|(I - PD\Phi(x^*))^{-1}\|_{X \leftarrow X} \cdot \varepsilon_C. \tag{3.17}$$

(3.16) holds near x^* if, e.g., $D\Phi$ (or, indeed, $PD\Phi$) is Lipschitz continuous with a small Lipschitz constant. To sum up, if a discretization scheme is stable, consistent, and satisfies (3.16), then by (3.17) the discretization error goes to 0 as the discretization level goes to infinity. In other words, the scheme in *convergent*. In particular, the discrete method is eventually well-posed.

The general notions described so far in this Section will be applied to analyze the convergence of the method described in Chapter 6. In particular, Proposition 6.2 proves the validity of Assumption 3.23, while Proposition 6.11 deals with Assumption 3.24.

On the other hand, obtaining the bound (3.16) can possibly be considered the subtlest part of the overall proof. The reason is that the condition on the Lipschitz continuity of $D\Phi$ (which, as anticipated, would be sufficient) cannot hold due to the nature of the problem, which concerns periodic BVPs (see Section 2.2). In particular, the map $D_\omega\Phi$ cannot be continuous since differentiating with respect to the period ω involves the composition with the map s_ω. However, the problem is overcome through the (weaker but sufficient) condition that $D\Phi$ is locally Lipschitz continuous in the solution x^* (see Propositions 6.7, 6.18) with a sufficiently small Lipschitz constant (see Proposition 6.23). The latter is possible thanks to the fact that x^* lies in a more regular subspace than the space X (see Lemma 7.5 in Chapter 7).

Finally, stability is proved through Lemma 6.22, while consistency is proved within Proposition 6.23. Section 6.5 is dedicated to a conclusive analysis of the consistency and the convergence error.

4 CONTINUATION OF EQUILIBRIA

As shown in section 1.4, the *Daphnia* model [43] is described by a coupled RE/DDE and has several complications. This chapter, the content of which is mostly included in the paper [11], deals indeed with the difficulties that arise from the presence of these external complications when standard continuation techniques are applied to, e.g., compute equilibria which depend on some parameter.

4.1 CONTINUATION FOR DAPHNIA

In the *Daphnia* model, defined by equations (1.9)-(1.12), the growth rate and the survival probability are only defined by means of solutions of external ODEs, and the delay $a_A(S_t)$ in the RE depends on the state through a nonlinear equation. In addition, the dependence on the state concerns also the breaking point between the juvenile and the mature life stages.

The *Daphnia* model can have *trivial* equilibria, i.e., $(b, S) \equiv (0, \overline{S})$ for $\overline{S}$ any zero of f in (1.12), as well as *nontrivial* equilibria $(b, S) \equiv (\overline{b}, \overline{S})$, where $\overline{S}$ satisfies

$$\int_{\overline{a}_A}^{a_{\max}} \beta(\xi(a, \overline{S}), \overline{S}) \mathcal{F}(a, \overline{S}) \, \mathrm{d}a = 1 \tag{4.1}$$

for $\overline{a}_A := a_A(\overline{S})$ and $\overline{b}$ is consequently given by

$$\overline{b} = \frac{f(\overline{S})}{\int_0^{a_{\max}} \gamma(\xi(a, \overline{S}), \overline{S}) \mathcal{F}(a, \overline{S}) \, \mathrm{d}a}.$$

The focus of this chapter is only on the nontrivial equilibria, in particular the component $\overline{S}$ given by (4.1), since the computation of the trivial ones is not affected by the difficulties mentioned above. The objective is to continue a solution branch of $\overline{S}$ with respect to a selected model parameter, while the values of all the other model parameters remain fixed. The choices for those values are the same as in [25] and are reported in Table 4.1. Most of them are hidden in (4.1), appearing only in the definitions of the various rates defining the model.

4.2 EXTERNAL CONTINUATION

This section describes and compares two approaches to solve the problem of the continuation for *Daphnia*, which are examples of *external* continuation, in that the solutions of the external IVPs (1.9) and (1.10) and of the maturation condition (1.11) are computed externally with respect to the continuation process.

resource intrinsic rate of change	$f(S) = a_1 S(1 - S/C)$
consumer growth rate	$g(\xi, S) = \gamma_g (\xi_m f_r(S) - \xi)$
consumer mortality rate	$\mu(\xi, S) = \mu$
consumer adults reproduction rate	$\beta(\xi, S) = r_m f_r(S) \xi^2$
consumer ingestion rate	$\gamma(\xi, S) = \nu_S f_r(S) \xi^2$
Holling type II functional response	$f_r(S) := \sigma S/(1 + \sigma S)$
size at birth	$\xi_b = 0.8$
size at maturation	$\xi_A = 2.5$
maximum size	$\xi_m = 6.0$
growth time constant	$\gamma_g = 0.15$
functional response shape parameter	$\sigma = 7.0$
maximum feeding rate	$\nu_S = 1.8$
maximum reproduction rate	$r_m = 0.1$
mortality rate parameter	$\mu =$ varying
environment carrying capacity	$C = 0.5$
flow-through rate	$a_1 = 0.5$
maximum age	$a_{\max} = 70$

Table 4.1: Rates (top) and parameters (bottom) of the considered *Daphnia* model.

The technique proposed in [22] can be used to reduce a nonlinear delay system to a system of ODEs, having in mind the idea of applying standard continuation tools for ODEs to such system via a pseudospectral discretization [57]. It extends the concept of infinitesimal generator for linear delay systems (e.g., [51, 3.29, Chapter II]), and consequently the IG-approach for linear delay systems (e.g., [28]) to nonlinear ones. The first step consists in transforming the delay equation into an equivalent nonlinear abstract differential equation (ADE). For example, consider an IVP of the form

$$\begin{cases} y'(t) = G(y_t), t \geq 0 \\ y_0 = \psi, \end{cases} \tag{4.2}$$

with $\psi \in Y := C([-\tau, 0], \mathbb{R}^d)$ and $G : Y \to \mathbb{R}^d$. If $\mathcal{A}_G$ is the corresponding generator and ψ belongs to its domain, then (4.2) is equivalent to the ADE

$$\begin{cases} v'(t) = \mathcal{A}_G(v(t)), t \geq 0 \\ v(0) = \psi, \end{cases} \tag{4.3}$$

in the sense that if y is a solution of (4.2), then the map defined by $v(t) := y_t$ is a solution of (4.3) and viceversa.

The pseudospectral discretization of degree M is defined as follows. The discretization of the state space Y is given by $Y_M := \mathbb{R}^{d(M+1)}$. A state $\psi \in Y$ is discretized by the vector ψ_M containing the values of ψ at the $M + 1$ Chebyshev extrema $\{\theta_{M,i}\}_{i \leq M}$ in $[-\tau, 0]$. By means of the reconstruction operator $R_M : Y_M \to Y$, which interpolates ψ_M at the Chebyshev extrema, the discrete version of G can be defined by

$$G_M(\psi_M) = G(R_M(\psi_M))$$

and, consequently, the discrete version of $\mathcal{A}_G$ is given by

$$\mathcal{A}_{G,M}(\psi_M) = \left(G_M(\psi_M), \frac{\mathrm{d}}{\mathrm{d}\theta} R_M(\psi_M)(\theta)|_{\theta=\theta_{M,1}}, \ldots, \frac{\mathrm{d}}{\mathrm{d}\theta} R_M(\psi_M)(\theta)|_{\theta=\theta_{M,M}} \right).$$

Eventually, (4.3) is discretized as

$$\begin{cases} v'_M(t) = \mathcal{A}_{G,M}(v_M(t)), t \geq 0 \\ v_M(0) = \psi_M. \end{cases} \tag{4.4}$$

The following theorems are then proved.

Theorem 4.1 (One-to-one correspondence, [22, Theorem 2.4]). *If $\overline{y} \in Y$ is an equilibrium for (4.2), then $(\overline{y}, \overline{v}_M)$ defined by*

$$\overline{v}_{M,i} = \overline{y},\ i = 1, \ldots, M, \tag{4.5}$$

is an equilibrium for (4.4). Conversely, if $(\overline{y}, \overline{v}_M)$ is an equilibrium for (4.4), then (4.5) holds and $\overline{y}$ is an equilibrium for (4.2).

Theorem 4.2 (Commutativity, [22, Theorem 2.5]). *Linearization around an equilibrium and pseudospectral discretization commute.*

Therefore, the problem of approximating the relevant eigenvalues of $\mathcal{A}_G$ through the ones of $\mathcal{A}_{G,M}$ can be reduced to its linear counterpart and the *spectral accuracy* of the approximation can be derived from the following theorem.

Theorem 4.3 ([28]). *Let $\overline{y}$ be an equilibrium of the DDE in (4.2) and let $L := DG(\overline{y})$. Let λ be an eigenvalue of $\mathcal{A}_L$ of multiplicity m. Then, for sufficiently large M, there exist eigenvalues $\lambda_{M,1}, \ldots, \lambda_{M,m}$ of $\mathcal{A}_{L,M}$, counted with their multiplicities, such that*

$$\max_{i=1,\ldots,m} |\lambda_{M,i} - \lambda| \leq C_2 \left(\frac{1}{\sqrt{M}} \left(\epsilon + \frac{C_1 |\lambda| \tau}{M} \right)^M \right)^{\frac{1}{m}},$$

where ϵ takes into account the possible error in the approximation of L, and C_1, C_2 are constants independent of M.

The method is then extended naturally to REs and coupled systems, which had been treated, respectively, in [24] and [25] in the linear case. The discretization obtained is then meant to be given as input to some software tool which is able to perform the continuation of the desired equilibrium automatically (e.g., MatCont).

Another numerical approach to continue equilibria of PSPM was developed in [90]. It is based on a natural continuation with secant prediction, recall Figure 4.1 (right), where the correction step is made by using the Broyden's update. Integrals and external IVPs are solved simultaneously via the embedded Runge-Kutta pair `DOPRI54` [49].

4.2.1 Preliminary numerical tests

Figure 4.1 shows the results of the continuation of the $\overline{S}$ component of the equilibria where the mortality rate μ is chosen as the continuation parameter, and was already presented in [11]. The curves in the left panel have been obtained as described next. The solid curve with bullets is the result

of the approach in [22] using MatCont. The resolution of the external IVPs (1.9) and (1.10) (by Matlab's `ode45`) are performed automatically inside the routine. The dashed curve with diamonds is, instead, the result of the approach in [90]. Both `DOPRI54` and `ode45` are capable of *event detection*, i.e., can automatically detect when the maturation condition (1.11) is satisfied during the integration of (1.9). However, all these calculations are repeated from scratch at every continuation step, done by MatCont or the natural continuation used in [90], with no possibility of exploiting the information available from the previous step.

The marked points represent the continuation steps. Some work was required so that each continuation step corresponded to (approximately) the same value of μ in the two cases, which was necessary in order to make a fair comparison. The difficulties were mostly due to the fact that MatCont contains a routine to control the step size of the continuation automatically; the user can only give a minimum and maximum step size as input, but cannot otherwise influence the way it varies during the computation. It was nevertheless possible to control the step size in the natural continuation based on [90], so that eventually the same range for μ was covered using the same amount of continuation steps.

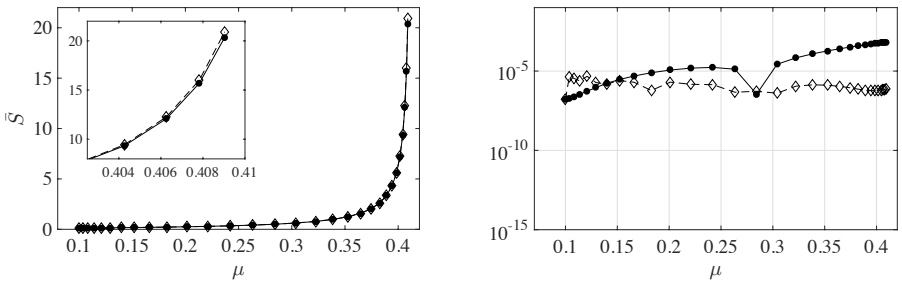

Figure 4.1: Equilibrium branch $\overline{S}(\mu)$ with zoom (left) and relevant residual (right) of the *Daphnia* model, computed with [22] (solid line with bullets) and [90] (dashed line with diamonds). Original figure from [11], courtesy of AIMS. See text for more details.

As for the left panel of the figure, the inner zoom highlightens the part of the outer plot where the two curves can be distinguished, which correspond to the greater values of μ.

The right panel shows the *residual* obtained along the branch with the two methods, i.e., the absolute difference between the right-hand and left-hand side of (4.1), where the continuation solution is plugged in at the left-hand side. Thanks to the choices in Table 4.1 the integral in (4.1) could be evaluated analytically as follows at the various input values for $\overline{S}$.

The solution of the IVP (1.9) for g given by Table 4.1 is

$$\overline{\xi}(\alpha) = \xi_b e^{-\gamma_g \alpha} + \xi_m f(S)(1 - e^{-\gamma_g \alpha}),$$

which gives $\xi(a, S) = \xi_b e^{-\gamma_g a} + \xi_m f(S)(1 - e^{-\gamma_g a})$. From (1.11), it follows that

$$a_A = -\frac{\log\left(\dfrac{\xi_a - \xi_m f(S)}{\xi_b - \xi_m f(S)}\right)}{\gamma_g}.$$

The solution of the IVP (1.10) gives simply $\mathcal{F}(a, S) = e^{-\mu a}$, being μ a constant function. Thus, the right-hand side of (4.1) reads

$$\begin{aligned} r_m f_r(S) \cdot \Bigg(\int_{a_A}^{a_{\max}} ((\xi_b - \xi_m f(S))^2 e^{-2\gamma_g a} + 2(\xi_b - \xi_m f(S))\xi_m f(S) e^{-\gamma_g a} \\ + \xi_m^2 f(S)^2) e^{-\mu a} \, \mathrm{d}a \Bigg) \\ = -r_m f_r(S) \cdot \Bigg(\frac{(\xi_b - \xi_m f(S))^2}{2\gamma_g + \mu} e^{-(2\gamma_g + \mu)a} + \frac{2(\xi_b - \xi_m f(S))\xi_m f(S)}{\gamma_g + \mu} e^{-(\gamma_g + \mu)a} \\ + \frac{\xi_m^2 f(S)^2}{\mu} e^{-\mu a} \Bigg) \Bigg|_{a_A}^{a_{\max}}. \end{aligned}$$

The outcomes of the computations with either [22] of [90] are comparable in terms of order of magnitude, even though a slightly increasing trend emerges for [22]. It is not immediate to tell which of the many possible sources of error (quadrature, IVPs, maturation condition, correction procedures, MatCont inner tolerances) determine the difference.

Similarly, there is no clear explanation for the relatively high difference in terms of the computational time needed to trace the equilibrium branch, but the fact that the approach in [90] is specific to the class of models of interest, and not as general as MatCont. The total elapsed time amounts indeed to 257.59 s with [22] and 59.32 s with [90], both implemented in Matlab and run on a MacBook Pro 2.3GHz Intel Core i7 16GB, the same hardware used to perform also the tests in Section 4.4. Regardless of this difference, these data on the computational time are what motivate the attempt described in the next section to improve the continuation strategy for complex models like the *Daphnia* one.

4.3 INTERNAL CONTINUATION

As anticipated in subsection 4.2.1, none of the standard techniques to continue the equilibria of the models of interest for this work involves the possibility of taking advantage of the information available from the previous step. The new strategy proposed in this section is indeed based on the idea that exploiting such information can be crucial in determining the overall computational cost. Having in mind the class of *Daphnia*-like models, the *internal* continuation method consists in including into the continuation framework the solution of the external IVPs (1.9) and (1.10) as well as the solution of the maturation condition (1.11). Aiming at proving its validity, the following subsections show the results of some tests on prototype problems, obtained by simplifying the *Daphnia* model. At first, all the techincalities of the model are dropped, but for the presence of an external ODE. Then, each of the other prototype models tackles one single challenge of the original model, namely a state-dependent maturation age, possible discontinuities among juveniles and adults and, finally, systems of external ODEs. It is also described how each of these challenges is addressed in the framework of the newly proposed internal continuation.

The choice of the specific instances of prototype problems allows to obtain a known analytic expression of the solution branch, and therefore measure

the *true error* due to the applied continuation. As for the notation used within the prototype models, the letter $\lambda \in \mathbb{R}$ indicates the varying parameter, and the objective is that of determining a quantity $x \in \mathbb{R}$ as a function of λ, defined implicitly through an integral condition. With reference to the case of *Daphnia*, x corresponds to $\overline{S}$, λ to the the mortality parameter μ (recall Table 4.1) and the integral condition to (4.1). In what follows, unless otherwise specified, the true error of a continuation curve $\{(x_i, \lambda_i)\}_{i<k}$ obtained after k continuation steps with respect to exact curve $x(\lambda)$, will be defined as

$$\epsilon_x = \max_{i<k} |x_i - x(\lambda_i)|. \tag{4.6}$$

Note that in the case of the search of the nontrivial equilibrium the family of ODEs (1.9) parametrized by the age $a \in [0, a_{\max}]$ reduces to a single ODE since the resource $\psi = \overline{S}$ is constant in time, being at equilibrium. However, this is not the case for other solutions (e.g., periodic ones), and the study of equilibria is just the first step of the dynamical analysis which shall be continued and extended beyond equilibria, in the future, as mentioned in Section 1.2. This is the reason why the internal continuation, as well as the relevant prototype problems, have been formulated in this more general way.

4.3.1 Alternative PDE formulation

The *Daphnia* model can be derived from the classical PDE formulation used to describe a size-structured population (feeding on some resource), as done in [37, Chapter 6].

The unknown function of the relevant PDE is given by $n(t, \xi)$, the density of individuals with size ξ at time t, while $g(\xi, S)$ is the growth rate of an individual with size ξ under resource S, and individuals are born with minimal size ξ_b.

The usual PDE formulation of a size-structured model is (assume μ constant for simplicity, as in Table 4.1)

$$\frac{\partial}{\partial t} n(t, \xi) + \frac{\partial}{\partial \xi} [g(\xi, S(t)) n(t, \xi)] = -\mu n(t, \xi) \tag{PDE}$$

$$g(\xi_b, S(t)) n(t, \xi_b) = \int_{\xi_b}^{\infty} \beta(\xi, S(t)) n(t, \xi) \, \mathrm{d}\xi. \tag{BC}$$

plus the equation for S, which reads

$$S'(t) = f(S(t)) - \int_{\xi_b}^{\infty} \gamma(\xi, S(t)) n(t, \xi) \, \mathrm{d}\xi.$$

The renewal equation is obtained by defining

$$b(t) := g(\xi_b, S(t)) n(t, \xi_b) \tag{4.7}$$

and using (BC) with $\xi = \xi(t,a)$, $d\xi/da = g(\xi(t,a), S(t))$ to write

$$\begin{aligned} b(t) &= \int_{\xi_b}^{\infty} \beta(\xi, S(t)) n(t,\xi)\, \mathrm{d}\xi \\ &= \int_0^{\infty} \beta(\xi(t,a), S(t)) n(t,\xi(t,a)) \frac{d\xi(t,a)}{da}\, \mathrm{d}a \\ &= \int_0^{\infty} \beta(\xi(a), S(t)) n(t,\xi(a)) g(\xi(a), S(t))\, \mathrm{d}a. \end{aligned}$$

So the classical renewal equation can be obtained from

$$n(t,\xi(t,a)) g(\xi(t,a), S(t)) = b(t-a) e^{-\mu a}, \tag{4.8}$$

which in turn can be proved by *integration along characteristics* of (PDE). The latter method consists in reducing the PDE to a family of ODEs which in turn define a family of curves in the (ξ, t) plane (namely, the *characteristics)*. This is achieved by introducing an extra variable $\alpha \geq 0$ (which, in this case, plays the role of age) and using a change of variables $t = t(\alpha)$ and $\xi = \xi(\alpha)$ such that

$$\xi'(\alpha) = g(\xi(\alpha), S(t(\alpha))), \quad \xi(0) = \xi_b$$

and

$$t'(\alpha) = 1, \quad t(0) = t - a.$$

a is an extra parameter that corresponds to the age of an individual that has size ξ at time t (so, in practice, $t(\alpha) = t - a + \alpha$) Thus, from (PDE) it follows that

$$\begin{aligned} \frac{d}{d\alpha} n(t(\alpha), \xi(\alpha)) &= \frac{\partial}{\partial t} n(t(\alpha), \xi(\alpha)) + \frac{d\xi(\alpha)}{d\alpha} \frac{\partial}{\partial \xi} n(t(\alpha), \xi(\alpha)) \\ &= \frac{\partial}{\partial t(\alpha)} n(t(\alpha), \xi(\alpha)) + \frac{\partial}{\partial \xi} [g(\xi(\alpha), S(t)) n(t(\alpha), \xi(\alpha))] \\ &\quad - n(t(\alpha), \xi(\alpha)) \frac{\partial}{\partial \xi} g(\xi(\alpha), S(t)) \\ &= -[\mu + \frac{\partial}{\partial \xi} g(\xi(\alpha), S(t))] n(t(\alpha), \xi(\alpha)). \end{aligned} \tag{4.9}$$

The (linear) ODE for $n(t(\alpha), \xi(\alpha))$ can be integrated, giving

$$\begin{aligned} n(t(\alpha), \xi(\alpha)) &= n(t(0), \xi(0)) e^{-\mu\alpha - \int_0^{\alpha} \frac{\partial}{\partial \xi} g(\xi(\theta), S(t(\theta)))\, \mathrm{d}\theta} \\ &= n(t(0), \xi(0)) e^{-\mu\alpha - \int_{\xi_b}^{\xi} \frac{\frac{\partial}{\partial \xi} g(\eta, S(t-a+\theta(\eta)))}{g(\eta, S(t-a+\theta(\eta)))}\, \mathrm{d}\eta}, \end{aligned}$$

where η represents size and $\eta := \xi(\theta)$ represents age. From

$$\int_{\xi_b}^{\xi} \frac{\frac{\partial}{\partial \xi} g(\eta, S(t-a+\theta(\eta)))}{g(\eta, S(t-a+\theta(\eta)))}\, \mathrm{d}\eta = \log \frac{g(\xi(\alpha), S(t-a+\alpha))}{g(\xi_b, S(t-a))}$$

and (4.7), for $\alpha \in [0, a]$ (4.9) becomes

$$\begin{aligned} n(t(\alpha), \xi(\alpha)) &= n(t(0), \xi(0)) \frac{g(\xi_b, S(t(0)))}{g(\xi(\alpha), S(t-a+\alpha))} e^{-\mu\alpha} \\ &= n(t-a, \xi_b) \frac{g(\xi_b, S(t-a))}{g(\xi(\alpha), S(t-a+\alpha))} e^{-\mu\alpha} \\ &= \frac{1}{g(\xi(\alpha), S(t-a+\alpha))} b(t-a) e^{-\mu\alpha}. \end{aligned}$$

Finally, (4.8) follows by taking $\alpha = a$, so that $t(a) = t$ and $\xi(a) = \xi$.

Note that the collocation of the external ODEs in $[0, a]$ for every $a \in [0, a_{\max}]$ is only naturally linked to the renewal formulation. Indeed, the age does not even appear explicitly in the original PDE since the latter refers to a size-structured population rather than to an age-structured one. This constitutes the main drawback of this PDE formulation in the context of parameter continuation of solutions, for the reasons mentioned prior to the beginning of this Subsection, i.e., the need of formulating the internal approach in order to be able to deal with non-steady solutions. Nevertheless, the PDE could also be equivalently formulated using age as the structuring variable, so that the unknown functions $n(t, a)$ and $\xi(t, a)$ depend on the age a and are defined through the PDEs

$$\begin{aligned} \frac{\partial}{\partial t} n(t,a) + \frac{\partial}{\partial a} n(t,a) &= -\mu n(t,a) \\ \frac{\partial}{\partial t} \xi(t,a) + \frac{\partial}{\partial a} \xi(t,a) &= g(\xi, S(t)) \end{aligned}$$

with boundary conditions

$$n(t,0) = \int_{\bar{a}_A}^{a_{\max}} \beta(\xi(t,a), S(t)) n(t,a)\, \mathrm{d}a, \quad \xi(t,0) = \xi_b \tag{4.10}$$

plus the equation for S, which reads

$$S'(t) = f(S(t)) - \int_0^{a_{\max}} \gamma(\xi(t,a), S(t)) n(t,a)\, \mathrm{d}a.$$

In particular, this reformulation has a state-dependent boundary condition (the first of (4.10)), while avoiding the state-dependence of the characteristic speed.

4.3.2 The basic prototype problem

The basic prototype problem is represented by the continuation of the curve $x(\lambda)$ defined by

$$\int_0^1 f(a, x, \lambda)\, \mathrm{d}a = 0, \tag{4.11}$$

where $f(a, x, \lambda) := \varphi(a; a, x, \lambda)$ and $\varphi(\alpha; a, x, \lambda) \in \mathbb{R}$ is the solution of

$$\begin{cases} \varphi'(\alpha; a, x, \lambda) = g(\varphi(\alpha; a, x, \lambda), a, x, \lambda), & \alpha \in [0, a], \\ \varphi(0; a, x, \lambda) = \varphi_0 \end{cases} \tag{4.12}$$

for some $g : \mathbb{R} \times [0,1] \times \mathbb{R}^2 \to \mathbb{R}$ and $\varphi_0 \in \mathbb{R}$.

As anticipated, the only challenge in this basic case is the fact that the integral corresponding to the one in (4.1) is defined through the solution of one external IVP (systems are dealt with in Subsection 4.3.4). The integration extrema are kept fixed in order to avoid state-dependency (which is treated in Subsection 4.3.3).

The definition of the problem depends on the choices on how to approximate both the integral in (4.11) and the solution of the IVP (4.12) needed to compute the relevant values of the integrand function.

The former is approximated through the Clenshaw-Curtis quadrature described in Subsection 3.1.5. The same quadrature is used also for the forthcoming prototypes.

The IVP (4.12) is solved numerically by means of polynomial collocation, described in Section 3.3. This is done in order to obtain the value $f(a_j, x, \lambda)$ for each $j = 1, \dots, N$ given x and λ, where $a_0, \dots, a_N$ are the Chebyshev nodes in $[0, 1]$ (a_0 is excluded with these quadrature nodes because $f(a_0, x, \lambda) = \varphi_0$ is given). Thus, collocation is used to compute an n-degree polynomial $p^{(j)}(\alpha) := p(\alpha; a_j, x, \lambda)$ such that

$$\begin{cases} p^{(j)\prime}(\alpha_i^{(j)}) = g(p^{(j)}(\alpha_i^{(j)}), a_j, x, \lambda), & i \in \{1, \dots, n\}, \\ p^{(j)}(\alpha_0^{(j)}) = \varphi_0, \end{cases}$$

where the collocation points $0 = \alpha_0^{(j)} < \cdots < \alpha_n^{(j)} = a_j$, $j \in \{1, \dots, N\}$, are the Chebyshev nodes in $[0, a_j]$.

The variables included in the continuation framework are those consituting the vector u given by

$$u := (p^{(1)}(\alpha_1^{(1)}), \dots, p^{(1)}(\alpha_n^{(1)}), \dots, p^{(N)}(\alpha_1^{(N)}), \dots, p^{(N)}(\alpha_n^{(N)}), x)^T \in R^{nN+1}, \tag{4.13}$$

and include, in particular, all the collocation variables $p^{(j)}(\alpha_i^{(j)})$, $i = 1, \dots, n$, $j = 1, \dots, N$.

Then G in (3.4) is given componentwise by

$$\begin{cases} p^{(j)\prime}(\alpha_i^{(j)}) - g(p^{(j)}(\alpha_i^{(j)})), & i = 1, \dots, n,\ j = 1, \dots, N \\ \sum_{j=0}^{N} w_j p^{(j)}(\alpha_n^{(j)}). \end{cases}$$

which leads to a bordered block diagonal structure of the resulting Jacobian matrix, as shown in figure 4.2.

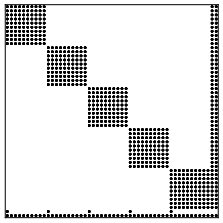

Figure 4.2: An example of bordered block diagonal structure of the Jacobian matrix for $n = 10$ (determining the size of the diagonal blocks) and $N = 5$ (determining the number of the diagonal blocks). Original figure from [11], courtesy of AIMS.

Note that the presence of a continuation framework eliminates the classical problem with solving IVPs (or any kind of nonlinear equations) through collocation, that is, the choice of a suitable initial guess to start the chosen iterative solver. Since the collocation variables are included into the continuation framework, such an initial guess is given by the same variables as computed at the previous continuation step.

The above cannot happen within the external continuation, since the standard initial value solvers (e.g., `ode45` in Matlab) can only use the information on the initial value φ_0 in order to start the computation, and none of the quantities computed at the preceding continuation step. This should be the main source of computational advantage of the internal strategy proposed.

Note that the increased number of equations of the continuation problem, namely $O(nN)$ as opposed to $O(N)$ for the external approaches (quadrature is anyway necessary in all cases), does not correspond to an increase in computational complexity. Indeed, using a classical external integration method of order $O(n)$ also has, in general, computational complexity $O(nN)$.

4.3.3 The state dependent prototype problem

In the state-dependent problem, one extremum of integration in (4.11) depends on the unknown x.

Such problem is represented by the continuation of the curve $x(\lambda)$ defined by

$$\int_{\overline{a}}^{1} f(a, x, \lambda)\, \mathrm{d}a = 0, \tag{4.14}$$

where $f(a, x, \lambda)$ is defined as in Subsection 4.3.2 ans $\overline{a} = \overline{a}(x, \lambda)$ is implicitly defined by

$$f(\overline{a}, x, \lambda) = \overline{\varphi} \tag{4.15}$$

for given $\overline{\varphi}$. In this case, polynomial collocation is used to compute the value of $f(a_j, x, \lambda)$ for each $j = 1, \dots, N$ given x and λ, where $a_0, \dots, a_N$ are the Chebyshev nodes in $[\overline{a}, 1]$ ($a_0 = \overline{a}$ is excluded with these quadrature nodes because $f(a_0, x, \lambda) = \overline{\varphi}$ is given).

The extra challenge in this case is the addition of the nonlinear condition (4.15) to the problem. This corresponds to the addition of an extra continuation variable, which is indeed $\overline{a}$. Hence (4.13) becomes

$$u := (p^{(1)}(a_1^{(1)}), \dots, p^{(1)}(a_n^{(1)}), \dots, p^{(N)}(a_1^{(N)}), \dots, p^{(N)}(a_n^{(N)}), x, \overline{a})^T \in \mathbb{R}^{nN+2},$$

implying that also the computation of the value of $\overline{a}$ takes advantage of the correspondent value computed at the previous continuation step, which is used to start the iterative solver at the current step.

Discontinuous right-hand side

As briefly anticipated in section 1.4, vital rates may be discontinuous at a finite number of points, corresponding to the beginning of the various life stages. This holds in particular in the case of *Daphnia*, where the growth rate of the consumer population is in general different between juveniles and adults, and therefore the right-hand side of the ODE in (1.9) may change across $\overline{a}$.

The state-dependent prototype problem with discontinuous right-hand side is represented by the continuation of the curve $x(\lambda)$ defined by equa-

tion (4.14), where $f(a,x,\lambda) := \varphi(a;a,x,\lambda)$, $\overline{a}$ is defined by (4.15), while $\varphi(\alpha;a,x,\lambda) \in \mathbb{R}$ is the solution of

$$\begin{cases} \varphi'(\alpha;a,x,\lambda) = g_1(\varphi(\alpha;a,x,\lambda),a,x,\lambda), & \alpha \in [0,a], \\ \varphi(0;a,x,\lambda) = \varphi_0 \end{cases}$$

if $a \leq \overline{a}$, while for $a > \overline{a}$ it is the solution of

$$\begin{cases} \varphi'(\alpha;a,x,\lambda) = g_2(\varphi(\alpha;a,x,\lambda),a,x,\lambda), & \alpha \in [\overline{a},a], \\ \varphi(\overline{a};a,x,\lambda) = \overline{\varphi}, & \\ \varphi'(\alpha;a,x,\lambda) = g_1(\varphi(\alpha;a,x,\lambda),a,x,\lambda), & \alpha \in [0,\overline{a}], \\ \varphi(0;a,x,\lambda) = \varphi_0. & \end{cases} \tag{4.16}$$

for some $g_1, g_2 : \mathbb{R} \times [0,1] \times \mathbb{R}^2 \to \mathbb{R}$ and $\varphi_0 \in \mathbb{R}$.

The extra challenge with respect to the simpler state-dependent problem is due to the different choices concerning the collocation in the case $a > \overline{a}$. Now, the collocation solution is a continuous piecewise polynomial in $[0,a]$, with $\overline{a}$ the only breaking point. Although, in principle, a different number of collocation nodes can be used in the two intervals, the chosen nodes are the Chebyshev nodes $0 = \alpha_0 < \cdots < \alpha_n = \overline{a}$ in $[0,\overline{a}]$ and the Chebyshev nodes $\overline{a} = \alpha_0^{(j)} < \cdots < \alpha_n^{(j)} = a_j$, $j = 1,\ldots,N$ in $[\overline{a},a_j]$. The values of the solution at all those nodes are all included into the internal continuation, and the dimension of the problem becomes $O(2nN)$.

4.3.4 The double size prototype problem

As shown in section 1.4, the *Daphnia* model is, in fact, defined by two external ODEs.

The double size prototype problem accomodates this feature, and is represented by the continuation of the curve $x(\lambda)$ defined by

$$\int_0^1 f(a,x,\lambda)c(a,x,\lambda)\,\mathrm{d}a = 0, \tag{4.17}$$

where $f(a,x,\lambda) := \varphi(a;a,x,\lambda)$, $c(a,x,\lambda) := \gamma(a;a,x,\lambda)$, while $\varphi(\alpha;a,x,\lambda)$ and $\gamma(\alpha;a,x,\lambda) \in \mathbb{R}$ constitute the solution of the system

$$\begin{cases} \varphi'(\alpha;a,x,\lambda) = g(\varphi(\alpha;a,x,\lambda),\gamma(\alpha;a,x,\lambda),a,x,\lambda), & \alpha \in [0,a], \\ \gamma'(\alpha;a,x,\lambda) = h(\varphi(\alpha;a,x,\lambda),\gamma(\alpha;a,x,\lambda),a,x,\lambda), & \alpha \in [0,a], \\ \varphi(0;a,x,\lambda) = \varphi_0, & \\ \gamma(0;a,x,\lambda) = \gamma_0 & \end{cases} \tag{4.18}$$

for given $g,h : \mathbb{R}^2 \times [0,1] \times \mathbb{R}^2 \to \mathbb{R}$ and $\varphi_0, \gamma_0 \in \mathbb{R}$. The choice of the integrand in (4.17), which may in principle be any function of $f(a,x,\lambda)$ and $c(a,x,\lambda)$, is motivated by the need to obtain an exact solution. Other simple choices, such as linear combinations, are possible. The dimension of the continuation problem (which is not state-dependent) is again $O(2nN)$. With respect to (4.13), the collocation unknowns are grouped as block-vectors with blocks of size 2, so that (4.13) remains unchanged but for the range of p which is now in $\mathbb{R}^2$.

4.4 NUMERICAL TESTS

Subsections 4.4.1 to 4.4.3 show the results of numerical simulations using internal continuation on the prototype models described in Section 4.3. The relevant algorithms are implemented in Python - which is an arbitrary choice, in fact, everything could be implemented just as well using other software. In order to make a fair comparison with the external continuation approach, the behavior of the relevant classical continuation-based software tools is emulated using the correspondent Python routines. The codes for the simulations are available at `http://cdlab.uniud.it/software#int-cont`.

In particular, the IVPs are solved by `scipy.integrate.odeint`, a Python IVP solver from the `scipy` package [66], based on the `LSODA` solver from the FORTRAN library `odepack`, which is able to switch automatically between non-stiff problems (solved using the implicit Adams formula) and stiff ones (solved using backward differentiation formulas), according to the method proposed in [84].

The nonlinear equation representing the maturation condition is solved by `scipy.optimize.fsolve`, which is based on the `HYBRD` and the `HYBRJ` solvers from the FORTRAN library `minpack`, which implement a variant of Powell's hybrid method [85].

Both the internal and the external approaches are implemented using pseudo-arclength continuation with tangent prediction and Newton's correction (see Subsection 3.4.1), the tolerance of which is set to 10^{-13}. All the tests are run on a MacBook Pro 2.3GHz Intel Core i7 16GB.

Subsection 4.4.4 shows the results of similar numerical simulations on the complete *Daphnia* model, and compares them with the results relevant to Figure 4.1 as described at the end of Section 4.1.

All the results of the simulations which are shown in this Section were presented in [11].

4.4.1 The basic prototype problem

Consider $g : \mathbb{R} \times [0,1] \times \mathbb{R}^2 \to \mathbb{R}$ defined by

$$g(\varphi(\alpha;a,x,\lambda),a,x,\lambda) := \lambda\varphi(\alpha;a,x,\lambda) + 2xe^{-\lambda a}. \tag{4.19}$$

The solution of (4.12) for g defined as in (4.19) can be obtained, e.g., using the Variation of Constants Formula, and reads

$$\varphi(\alpha;a,x,\lambda) = e^{\lambda a}\varphi_0 + \frac{x}{\lambda}(e^{\lambda(\alpha-a)} - e^{\lambda a}).$$

With reference to Subsection 4.3.2, it follows that

$$f(a,x,\lambda) = \varphi(a;a,x,\lambda) = e^{\lambda a}\varphi_0 + \frac{x}{\lambda}(1 - e^{\lambda a}).$$

Thus,

$$\int_0^1 f(a,x,\lambda)\,\mathrm{d}a = \int_0^1 e^{\lambda a}\varphi_0 + \frac{x}{\lambda}(1 - e^{\lambda a}) = \frac{e^{\lambda}}{\lambda}\varphi_0 + \frac{x}{\lambda} + \frac{x}{\lambda^2}(e^{-\lambda} - 1)$$

and the solution branch defined by (4.11) reads

$$x(\lambda) = \frac{\varphi_0}{2} \cdot \frac{\lambda(1 - e^{\lambda})}{\lambda + e^{-\lambda} - 1}. \tag{4.20}$$

The analytical expression of $x(\lambda)$ in (4.20) allows to evaluate the true error (4.6) on the continuation curve by varying the number of collocation and quadrature nodes, while running a fixed number of continuation steps (precisely 10) so that the same range for the continuation parameter is covered with both the internal and external approaches. As anticipated in Section 4.3, this is what motivates the choice of the specific instances of the prototype problems used for the simulations, and in fact it holds also for all the forthcoming tests.

Figure 4.3 (top) shows the error (4.6) obtained with respect to the curve (4.20) when the right-hand side of (4.12) is given by (4.19) with $\varphi_0 = 1$, using $N = 10$ quadrature nodes. The error decays *spectrally* as the number n of collocation nodes increases in the internal case (line with bullets), which is the expected behavior of collocation when the problem is smooth [100]. Horizontal lines are the result of the external continuation where the tolerances of `odeint` and `fsolve` are both set to $10^{-8}, 5 \times 10^{-10}, 10^{-13}$ respectively. For each of those values there is (at least) a value of n for which the internal continuation performs better in terms of both time and error. The diamond markers in Figure 4.3 (bottom) show that this holds for $n = 7, 8, 11$ respectively.

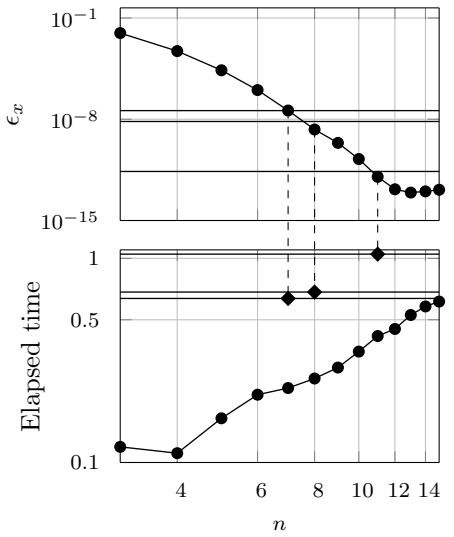

Figure 4.3: Internal (lines with bullets) versus external continuation (horizontal lines) for (4.19): error (4.6) on the true curve (4.20) (top) and elapsed time (bottom, s) using n collocation points and $N = 10$ quadrature nodes. Original figure from [11], courtesy of AIMS. See text for more details.

Different choices for N are possible. In fact, the simulations above were replicated using different values for N, up to 100, and always gave, qualitatively speaking, the same results. Figure 4.4 is a partial evidence of this. It is obtained fixing $n = 12$ (for which, as shown in Figure 4.3, the internal continuation approach turns out to be superior to the external one when $N = 10$

and the tolerance is set to 10^{-13}) and increasing N. Figure 4.4 (top) shows the error (4.6) obtained on the true curve (4.20), while Figure 4.4 (bottom) compares the elapsed time. Lines marked with bullets refer to the internal continuation, while the ones with squares refer to the external continuation with external tolerances fixed to 10^{-13}. For all values N in the Figure (between 8 and 15), the internal continuation performs better in terms of both time and error.

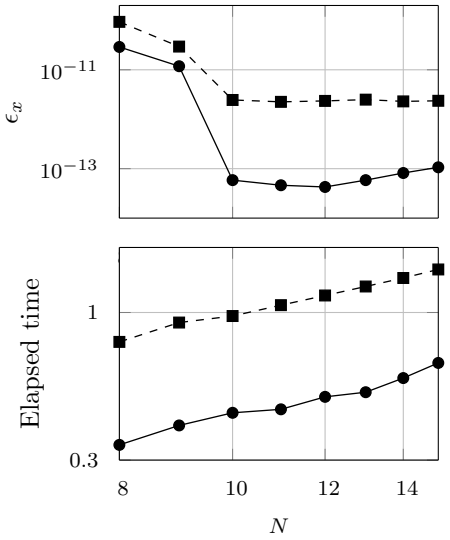

Figure 4.4: Internal (lines with bullets) versus external continuation (lines with squares) for (4.19): error (4.6) on the true curve (4.20) (top) and elapsed time (bottom, s) using $n = 12$ collocation points and N quadrature nodes. Original figure from [11], courtesy of AIMS. See text for more details.

4.4.2 The state dependent prototype problem

Consider $g : \mathbb{R} \times [0,1] \times \mathbb{R}^2 \to \mathbb{R}$ defined by

$$g(\varphi(\alpha;a,x,\lambda),a,x,\lambda) := -(a+1)x(\lambda^2+2)\left(\frac{\varphi(\alpha;a,x,\lambda)-\varphi_0}{x(\lambda^2+2)(a+1)}+1\right)^2. \tag{4.21}$$

The solution of (4.12) for g defined as in (4.21) reads

$$\varphi(\alpha;a,x,\lambda) = (a+1)x(\lambda^2+2)\left(\frac{1}{\alpha+1}-1\right)+\varphi_0,$$

which is well defined for $\alpha \in [0,1]$. With reference to Subsection 4.3.2, consider the choice

$$\overline{\varphi} = \varphi_0 - 1 \tag{4.22}$$

in (4.15). It follows that

$$f(a,x,\lambda) = \varphi(a;a,x,\lambda) = -ax(\lambda^2+2)+\varphi_0.$$

In particular, from (4.22),

$$\overline{a} = \frac{1}{x(\lambda^2+2)},$$

thus

$$\int_{\overline{a}}^{1} f(a,x,\lambda)\,\mathrm{d}a = -x(\lambda^2+2)\frac{a^2}{2} + \varphi_0 a\Big|_{\overline{a}}^{1} = -\frac{x(\lambda^2+2)}{2} + \varphi_0 + \frac{-2\varphi_0+1}{2x(\lambda^2+2)}.$$

The solution of (4.14) is given by either $x(\lambda^2+2) = 1$ or $x(\lambda^2+2) = 2\varphi_0 - 1$. Since the former gives $\overline{a} = 1$, the only nontrivial solution branch reads

$$x(\lambda) = \frac{2\varphi_0 - 1}{\lambda^2 + 2}. \tag{4.23}$$

Note that this instance of a prototype problem does not have any immediate biological meaning since, when g represents a growth rate as in *Daphnia*, it must be $\varphi(\alpha; a, x, \lambda) > 0$ and in particular $\overline{\varphi} > \varphi_0 > 0$. However, as far as only prototype problems are concerned, these and similar constraints are ignored, the focus being instead on obtaining exact expressions of the solution branches.

Figure 4.5 (top) shows the error (4.6) obtained with respect to the curve (4.23) when the right-hand side of (4.12) is given by (4.21) and the maturation condition by (4.22), with $\varphi_0 = \frac{3}{2}$, using $N = 10$ quadrature nodes. Again, horizontal lines are the result of external continuation where the tolerances of `odeint` and `fsolve` are both set to $10^{-8}, 5\times10^{-10}, 10^{-13}$ respectively. For each of those values there is (at least) a number n of collocation points for which the internal continuation performs better in terms of both time and error. The diamond markers in Figure 4.5 (bottom) show that this holds for $n = 13, 12, 17$ respectively.

As a side remark, note that the external continuation takes slightly less time for 5×10^{-10} than for 10^{-8}: this can be somehow related to the automatic error control of either `odeint` or `fsolve`.

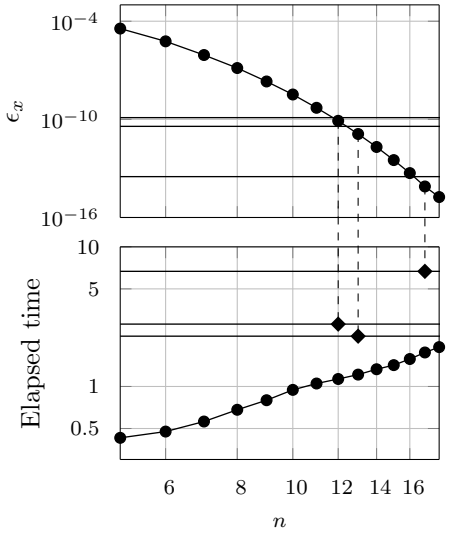

Figure 4.5: Internal (lines with bullets) versus external continuation (horizontal lines) for (4.21) and (4.22): error (4.6) on the true curve (4.23) (top) and elapsed time (bottom, s) using n collocation points and $N = 10$ quadrature nodes. Original figure from [11], courtesy of AIMS. See text for more details.

Discontinuous right hand side

Consider $g_1 : \mathbb{R} \times [0,1] \times \mathbb{R}^2 \to \mathbb{R}$ defined by

$$g_1(\varphi(\alpha;a,x,\lambda),a,x,\lambda) := -x(\lambda^2+1)\left(\frac{\varphi(\alpha;a,x,\lambda)-\varphi_0}{x(\lambda^2+1)} - \frac{1}{2}\right)^2. \tag{4.24}$$

The solution of (4.12) for g_1 defined as in (4.24) reads

$$\varphi_1(\alpha;a,x,\lambda) = x(\lambda^2+1)\left(\frac{1}{\alpha-2}+\frac{1}{2}\right)+\varphi_0,$$

which is well defined for $\alpha \in [0,1]$. With reference to Subsection 4.3.3, consider the choice

$$\overline{\varphi} = \frac{1}{2}\varphi_0 \tag{4.25}$$

in (4.15). It follows that

$$\frac{1}{2}\varphi_0 = f(\overline{a},x,\lambda) = \varphi_1(\overline{a};\overline{a},x,\lambda) = x(\lambda^2+1)\left(\frac{1}{\overline{a}-2}+\frac{1}{2}\right)+\varphi_0$$

and thus

$$\overline{a} = \frac{2\varphi_0}{x(\lambda^2+2)}.$$

Consider $g_2 : \mathbb{R} \times [0,1] \times \mathbb{R}^2 \to \mathbb{R}$ defined by

$$g_2(\varphi(\alpha;a,x,\lambda),a,x,\lambda) := -(a+1)x(\lambda^2+1)\left(\frac{\varphi(\alpha;a,x,\lambda)-\overline{\varphi}}{x(\lambda^2+1)(a+1)} + \frac{1}{\overline{a}+1}\right)^2. \tag{4.26}$$

The solution of (4.16) for g_2 defined as in (4.26) reads

$$\varphi_2(\alpha;a,x,\lambda) = (a+1)x(\lambda^2+1)\left(\frac{1}{\alpha+1}-\frac{1}{\overline{a}+1}\right)+\overline{\varphi},$$

which is well defined for $\alpha \in [0,1]$. It follows that, for $a > \overline{a}$,

$$f(a,x,\lambda) = \varphi_2(a;a,x,\lambda) = x(\lambda^2+1)\frac{\overline{a}-a}{\overline{a}+1}+\overline{\varphi},$$

thus

$$\begin{aligned}\int_{\overline{a}}^1 f(a,x,\lambda)\,\mathrm{d}a &= x(\lambda^2+1)\frac{\overline{a}a-a^2/2}{\overline{a}+1}+\overline{\varphi}a\Big|_{\overline{a}}^1 \\ &= (1-\overline{a})\left(x(\lambda^2+1)\frac{\overline{a}-1}{2(\overline{a}+1)}+\overline{\varphi}\right).\end{aligned}$$

Apart from the trivial solution $\overline{a} = 1$ of (4.14), the other solutions are given by $x(\lambda^2+1) = -2\varphi_0$ and $x(\lambda^2+1) = 3\varphi_0$. Since the former gives $\overline{a} = -2$, the only nontrivial solution branch reads

$$x(\lambda) = \frac{3\varphi_0}{\lambda^2+1}. \tag{4.27}$$

Figure 4.6 (top) shows the error (4.6) obtained with respect of the true curve (4.27), when the right hand side of (4.16) is given by (4.24) and (4.26), and the maturation condition by (4.25), with $\varphi_0 = 1$, using $N = 10$ quadrature nodes. Horizontal lines are the result of external continuation where the tolerances of `odeint` and `fsolve` are both set to $10^{-8}, 5\times10^{-10}, 10^{-13}$ respectively. For each of those values there is (at least) a number n of collocation points for which the internal continuation performs better in terms of both time and error. The diamond markers in Figure 4.6 (bottom) show that this holds for $n = 8, 10, 11$ respectively.

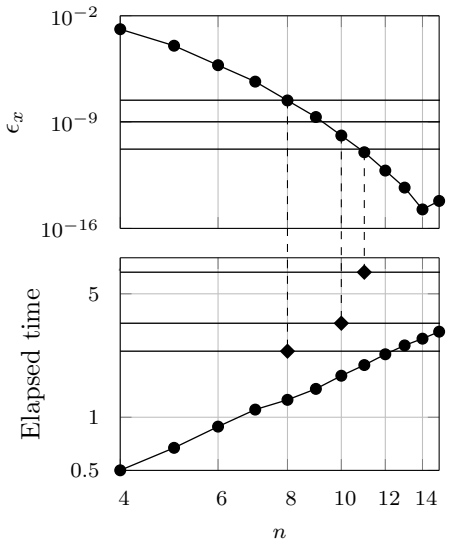

Figure 4.6: Internal (lines with bullets) versus external continuation (horizontal lines) for (4.24), (4.26) and (4.25): error (4.6) on the true curve (4.27) (top) and elapsed time (bottom, s) using n collocation points and $N = 10$ quadrature nodes. Original figure from [11], courtesy of AIMS. See text for more details.

4.4.3 The double size prototype problem

Consider $g, h : \mathbb{R} \times [0,1] \times \mathbb{R}^2 \to \mathbb{R}$ defined by

$$\begin{cases} g(\varphi(\alpha;a,x,\lambda),\gamma(\alpha;a,x,\lambda),a,x,\lambda) := \\ \qquad -(a+1)x(\lambda^2+1)\left(\dfrac{\varphi(\alpha;a,x,\lambda)-\varphi_0}{x(\lambda^2+1)(a+1)}+1\right)\left(\dfrac{\gamma(\alpha;a,x,\lambda)-\gamma_0}{(\lambda^2+1)(a+1)}+1\right) \\ h(\varphi(\alpha;a,x,\lambda),\gamma(\alpha;a,x,\lambda),a,x,\lambda) := \\ \qquad -(a+1)(\lambda^2+1)\left(\dfrac{\varphi(\alpha;a,x,\lambda)-\varphi_0}{x(\lambda^2+1)(a+1)}+1\right)\left(\dfrac{\gamma(\alpha;a,x,\lambda)-\gamma_0}{(\lambda^2+1)(a+1)}+1\right). \end{cases} \tag{4.28}$$

The solution of (4.18) for g, h defined as in (4.28), reads

$$\begin{cases} \varphi(\alpha;a,x,\lambda) = (a+1)x(\lambda^2+1)\left(\dfrac{1}{\alpha+1}-1\right)+\varphi_0 \\ \gamma(\alpha;a,x,\lambda) = (a+1)(\lambda^2+1)\left(\dfrac{1}{\alpha+1}-1\right)+\gamma_0, \end{cases}$$

which are well defined for $\alpha \in [0,1]$. With reference to Subsection 4.3.4, it follows that

$$\begin{cases} f(a,x,\lambda) = \varphi(a;a,x,\lambda) = -ax(\lambda^2+1)+\varphi_0 \\ c(a,x,\lambda) = \gamma(a;a,x,\lambda) = -a(\lambda^2+1)+\gamma_0, \end{cases}$$

thus

$$\int_0^1 f(a,x,\lambda)c(a,x,\lambda)\,\mathrm{d}a = \frac{1}{3}x(\lambda^2+1)^2 - \frac{1}{2}(\lambda^2+1)(\varphi_0+x\gamma_0)+\varphi_0\gamma_0$$

and the solution of (4.17) is given by

$$x(\lambda) = \frac{3\varphi_0}{\lambda^2+1}\cdot\frac{\lambda^2+1-2\varphi_0}{2(\lambda^2+1)-3\gamma_0}. \tag{4.29}$$

Figure 4.7 (top) shows the error (4.6) obtained with respect of the curve (4.29) when the right-hand sides of (4.18) are given by (4.28), with $\varphi_0 = \gamma_0 = 1$, using $N = 10$ quadrature nodes. Horizontal lines are the result of external continuation where the tolerances of `odeint` and `fsolve` are both set to $10^{-8}, 5 \times 10^{-10}, 10^{-13}$ respectively. Unlike the previous scalar cases, the external continuation seems to perform slightly better in terms of both time and error. The diamond markers in Figure 4.7 (bottom) show that this holds for $n = 12, 13, 18$ respectively (mind anyway the time scale in the bottom panel). The explanation might reside in the increased dimension of the continuation problem for the internal approach, as mentioned in Subsection 4.3.4, and the denser block structure of the resulting Jacobian matrix with respect to the state-dependent prototype problem with discontinuous right-hand side, which shares the same increase in dimension (see Subsection 4.3.3).

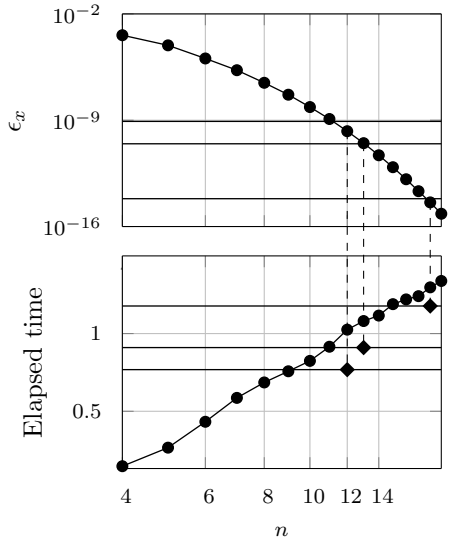

Figure 4.7: Internal (lines with bullets) versus external continuation (horizontal lines) for (4.28): error (4.6) on the true curve (4.29) (top) and elapsed time (bottom, s) using n collocation points and $N = 10$ quadrature nodes. Original figure from [11], courtesy of AIMS. See text for more details.

4.4.4 The complete Daphnia problem

The last numerical results presented in this Section concern the continuation of the branch $\overline{S}(\mu)$, as the ones in Subsection 4.2.1 with the approaches described in [22] and in [90]. In particular, both these approaches are compared with the internal one.

As it was done for the [90] approach, within the internal one it is also possible to control the step size, in order to cover approximately the same range for μ as the [22] approach (Figure 4.1) with 30 steps. The tolerance of the Newton's corrections is set to 10^{-6} in all cases.

Figure 4.8 shows the results of three runs, which use the same number of quadrature and collocation nodes ($n = N$) and differ only by the choice of such number, respectively $n = N = 10, 15, 20$. The results obtained are

superposed to Figure 4.1. In particular, the left panel only differs from the one in Figure 4.1 in that the curve obtained with $n = N = 20$ (dash-dot line with stars) is added. In the right panel, the residual for all the three choices of $n = N$, defined as in Section 4.1 (again dash-dot lines with stars). is added: lines with smaller residual correspond to larger values of $n = N$. Table 4.2 shows the results concerning the computational time in relation to

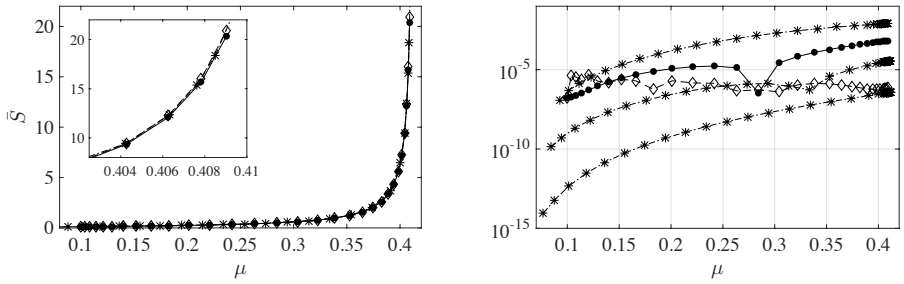

Figure 4.8: Equilibrium branch $\overline{S}(\mu)$ with zoom (left) and relevant residual (right) of the *Daphnia* model, computed with the internal continuation (dash-dot line with stars), superposed to Figure 4.1 for comparison. Original figure from [11], courtesy of AIMS. See text for more details.

the maximal residual, in all cases. As expected from the results on the prototype problems, the outcome demonstrates the superiority of the internal continuation with respect to either [22] or [90].

Note that the internal continuation is implemented in Python, whereas [22] and [90] are implemented in Matlab (for the former there is no alternative due to MatCont, for the latter the relevant codes are available only in Matlab). However, in neither case the different language is responsible for such an evident speed-up. Moreover, recall that [90] is implemented with secant prediction and Broyden's update, both choices favoring the latter (in terms of computational time) with respect to the internal approach for avoiding the computation of the Jacobian.

method	computational time	maximal residual
[22]	257.59s	6.6357×10^{-4}
[90]	59.32s	4.6768×10^{-6}
internal continuation with $n = N = 10$	1.73s	8.1723×10^{-3}
internal continuation with $n = N = 15$	4.40s	3.6517×10^{-5}
internal continuation with $n = N = 20$	9.18s	3.6854×10^{-7}

Table 4.2: Computational time and maximal residual for the continuation of the *Daphnia* model. Original table from [11], courtesy of AIMS.

5 APPROXIMATION OF PERIODIC SOLUTIONS

This chapter, the content of which is mostly included in the paper [9][1], deals with the computation of periodic solutions of (1.1). This is, somehow, another instance of *internal* approach, in that the computation of the solution takes advantage of the results obtained in previous steps (i.e., to compute solutions for other values of the model parameters). The technique which will be described in the following was developed and experimentally tested in [53] for DDEs, but was never extended to REs or coupled systems. The possibility of such extension, is supported by the numerical tests which will be shown, performed first separately on DDEs and REs, then on coupled systems.

5.1 COLLOCATION OF THE PERIODIC BOUNDARY VALUE PROBLEM

(2.13) can be solved numerically through (e.g., polynomial) collocation. As explained in Section 3.3, this would mean looking for m-degree polynomials u and v in $[0,1]$ such that

$$\begin{cases} u(\theta_j) = F(\overline{u}_{\theta_j} \circ s_\omega, \overline{v}_{\theta_j} \circ s_\omega), & j = 1, \dots, m, \\ v'(\theta_j) = \omega G(\overline{u}_{\theta_j} \circ s_\omega, \overline{v}_{\theta_j} \circ s_\omega), & j = 1, \dots, m, \\ (u(0), v(0)) = (u(1), v(1)) \\ p(u, v) = 0 \end{cases}$$

for given collocation points $0 \leq \theta_1 < \cdots < \theta_m \leq 1$, where $s_\omega, \overline{u}$ and $\overline{v}$ are defined as in Section 2.2.

Moreover, following [53], the method can be improved using piecewise polynomial collocation. As explained in Section 3.3, in this case the numerical solution in $[0,1]$ is obtained by solving the following system having dimension $(1 + Lm) \times (d_X + d_Y) + 1$:

$$\begin{cases} u(\theta_j) = F(\overline{u}_{\theta_{i,j}} \circ s_\omega, \overline{v}_{\theta_{i,j}} \circ s_\omega), & j \in \{1, \dots, m\},\ i \in \{0, \dots, L-1\}, \\ v'(\theta_j) = \omega G(\overline{u}_{\theta_{i,j}} \circ s_\omega, \overline{v}_{\theta_{i,j}} \circ s_\omega), & j \in \{1, \dots, m\},\ i \in \{0, \dots, L-1\}, \\ (u(0), v(0)) = (u(1), v(1)) \\ p(u, v) = 0 \end{cases} \tag{5.1}$$

for a given mesh $0 = t_0 < \cdots < t_L = 1$ and collocation points

$$t_i \leq \theta_{i,1} < \cdots < \theta_{i,m} \leq t_{i+1}$$

for all $i \in \{0, \dots, L-1\}$.

[1] In particular, almost all the figures in the chapter have already been accepted for publication in [9], but not yet officially published at the time of the submission of this thesis.

The variables which are included in the continuation framework are, other than ω, those of the form $u_{i,j} := u(t_{i,j})$ and $v_{i,j} := v(t_{i,j})$ for $i \in \{0, \dots, L-1\}$, $j \in \{0, \dots, m\}$, and fixed *representation nodes*

$$t_i = t_{i,0} < \cdots < t_{i,m} = t_{i+1},$$

which can be chosen independently from the collocation nodes. If $\{\ell_{i,j}\}_{0 \le j \le m}$ is the Lagrange basis associated to the representation nodes in $[t_i, t_{i+1}]$ (see Subsection 3.1.2), then such variables consitute the solution of the system

$$\begin{cases} \sum_{k=0}^{m} \ell_{i,k}(\overline{u}_{\theta_{i,k}} \circ s_\omega, \overline{v}_{\theta_{i,k}} \circ s_\omega) = F(\overline{u}_{\theta_{i,j}} \circ s_\omega, \overline{v}_{\theta_{i,j}} \circ s_\omega), & j \in \{1, \dots, m\}, \\ & i \in \{0, \dots, L-1\}, \\ \sum_{k=0}^{m} \ell'_{k,j}(\overline{u}_{\theta_{i,j}} \circ s_\omega, \overline{v}_{\theta_{i,j}} \circ s_\omega) = \omega G(\overline{u}_{\theta_{i,j}} \circ s_\omega, \overline{v}_{\theta_{i,j}} \circ s_\omega), & j \in \{1, \dots, m\}, \\ & i \in \{0, \dots, L-1\}, \\ u_{0,0} - u_{L,0} = 0 \\ v_{0,0} - v_{L,0} = 0 \\ p(u, v) = 0, \end{cases}$$

where all the values $\ell'_{i,k}(\theta_{i,k})$ are computed through the differentiation matrix, defined in Section 3.3.

As far as the phase condition p is concerned, note that in the case of numerical approximations through iterative methods either some $\hat{x}$ (or $\hat{y}$) or some $\tilde{x}$ (or $\tilde{y}$), defined as in Section 2.2, is available: indeed, if the continuation starts close to a Hopf bifurcation (see, e.g., [72, Sections 3.4 and 3.5]), i.e., at a "newborn" limit cycle, a coordinate of the equilibrium giving rise to it is a natural choice for $\hat{x}$ (or $\hat{y}$). Alternatively, a possible reference solution $\tilde{x}$ (or $\tilde{y}$) can be given by a cycle with period $2\pi/\beta$, where β is (the absolute value of) the imaginary part of the conjugate pair determining the Hopf bifurcation. This cycle is intended to represent an approximation of a periodic solution corresponding to a value of the parameter obtained by slightly perturbing the Hopf one, and a reasonable guess for the amplitude is given by $\sqrt{\alpha}$, where α is the real part of the aforementioned conjugate pair at the perturbed value of the parameter. On the other hand, at the subsequent continuation steps, $\tilde{x}$ (or $\tilde{y}$) can be defined as a component of the periodic solution computed at the previous continuation step.

5.2 PERIODIC BVP FOR DDES

The following theorem provides a bound on the collocation error, in the case of IVPs for DDEs, assuming that the relevant *breaking points* (see, e.g., [18] for a general reference) are included in the discretization mesh.

Theorem 5.1 ([53, Theorem 4.1]). *Consider an IVP of the form*

$$\begin{cases} y'(t) = G(y_t), & t \in [0, \omega], \\ y_0 = \phi \end{cases}$$

such that G and ϕ are arbitrarily smooth. Let z be its exact solution and let $u \in \Pi_n^t$ be a solution obtained through a piecewise polynomial collocation scheme, using any mesh $\mathtt{t}$ and any set of collocation nodes. Let $h := \max_{0 \le i < n} t_{i+1} - t_i$. Then

$$\max_{t \in [0,\omega]} \|z(t) - u(t)\| = O(h^m).$$

Moreover, if the collocation nodes are chosen as the Gauss-Legendre nodes in each $[t_i, t_{i+1}]$, then

$$\max_{t \in [0,\omega]} \|z(t) - u(t)\| = O(h^{m+1}).$$

Since, in the case of ODEs, the same convergence orders hold for IVPs and BVPs (using similar meshes), in [53], the same order of convergence is conjectured also for the BVP

$$\begin{cases} y'(t) = G(y_t), \quad t \in [0, \omega], \\ y_0 = y_\omega \\ p(y_0, y_\omega) = 0 \end{cases} \tag{5.2}$$

and supported by several numerical tests. It is worth remarking that, anyway, a theoretical proof of convergence is still missing, and the one in [13] does not consider the presence of any unknown parameters (in particular, the period is assumed to be known).

The first of these test was replicated here, and the results are shown below. The relevant DDE is the delayed logistic equation

$$y'(t) = (\lambda - y(t-1))y(t), \tag{5.3}$$

for which $\lambda = \pi/2$ is a Hopf bifurcation point. Starting from a sinusoidal perturbation of the corresponding equilibrium, the branch of periodic orbits was continued up to $\lambda = 1.7$, using a trivial phase condition at the first step with $\hat{y} = \pi/2$, and an integral phase condition at the other steps, with $\tilde{y}$ the solution found at the previous continuation step.

Note that this does not mean that the continuation is dependent on the parameter λ in the sense of Section 3.4. In fact, formally speaking, λ is not even included in the continuation, since otherwise the continuation variables would outnumber the equations. Rather, this *manual* continuation requires, at each step, to increase λ by some $\Delta\lambda > 0$, and simply substitute $\lambda + \Delta\lambda$ to λ in all the relevant positions of the Jacobian matrix. The solution obtained at the previous continuation step will be the prediction at the current step, without further modifications.

At $\lambda = 1.7$, a periodic solution of period $T \approx 4.0964$ was found, Figure 5.1. Gauss-Legendre collocation points were used. Since there is no way to obtain an exact expression of the relevant periodic solution, the error was computed with respect to a reference solution obtained with $L = 1000$ and $m = 4$. Indeed, Figure 5.2 confirms the $O(h^{m+1})$ behavior for both the continuous (uniform) and discrete (maximum at the mesh points) errors, where the former was computed by taking the maximum of the errors at the 4001 representation points relevant to the reference solution.

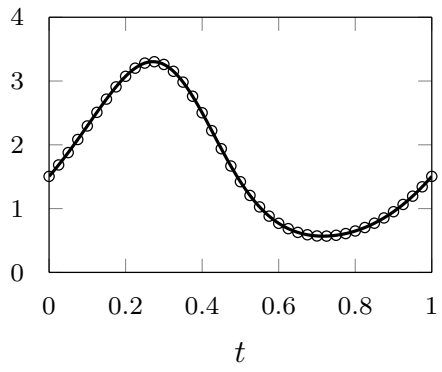

Figure 5.1: Periodic solution of (5.3) at $\lambda = 1.7$: reference solution obtained using $L = 1000$ and $m = 4$ (solid line) compared with the solution approximated using $L = 10$ and $m = 4$ (circles). Original figure from [9][1].

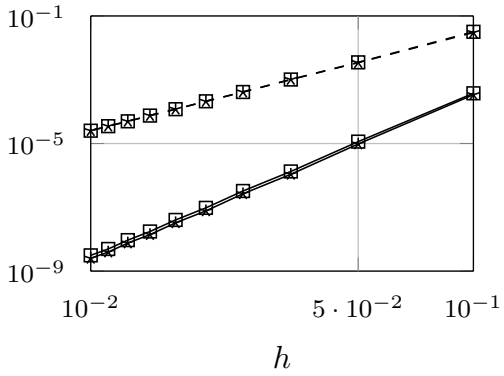

Figure 5.2: Periodic solution of (5.3) at $\lambda = 1.7$: continuous (squares) and discrete (stars) errors for $m = 3$ (dashed line) and $m = 4$ (solid line). Original figure from [9][1].

5.3 PERIODIC BVP FOR RES

Given the numerical proof, found in [53], that the same order of convergence as Theorem 5.1 holds also for (5.2), the same behavior can be expected in the case of REs with a smoothing right-hand side, such as those of the form (1.3) and (1.4). Our expectation is indeed satisfied as shown by the test described below. The relevant code is available at `http://cdlab.uniud.it/software#per-sol`.

The relevant RE is

$$x(t) = \frac{\gamma}{2} \int_{-3}^{-1} x(t+\theta)(1 - x(t+\theta))d\theta \tag{5.4}$$

(with trivial phase condition), for which, as shown in [23], the exact expression of the periodic solution between a Hopf bifurcation point and the first period doubling is

$$x(t) = \sigma + A \sin\left(\frac{\pi}{2}t\right), \tag{5.5}$$

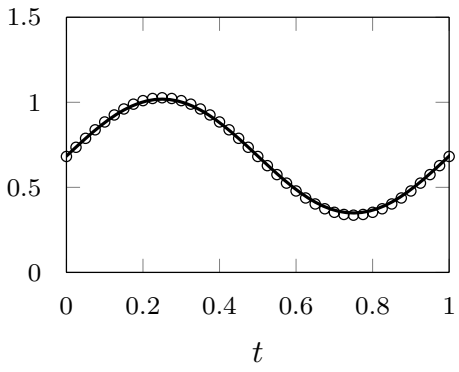

Figure 5.3: Periodic solution of (5.4) at $\gamma = 4.327$: exact solution (solid line) compared with the solution approximated using $L = 10$, $m = 4$ and Chebyshev points (circles). Original figure from [9][1].

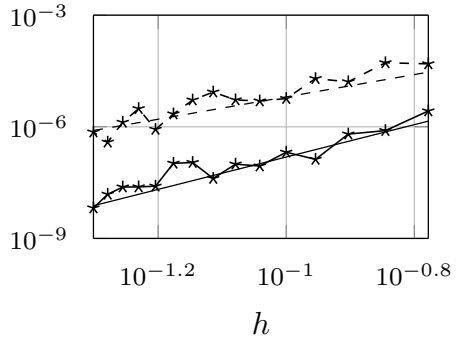

Figure 5.4: Periodic solution of (5.4) at $\gamma = 4.327$: continuous error for $m = 3$ (dashed line) and $m = 4$ (solid line) using Chebyshev points, compared to straight lines having angular coefficient 3 (dashed) and 4 (solid).

where

$$\begin{cases} \sigma = \dfrac{1}{2} + \dfrac{\pi}{4\gamma}, \\ A^2 = 2\sigma\left(1 - \dfrac{1}{\gamma} - \sigma\right), \end{cases}$$

and $\gamma = 2 + \pi/2$ is a Hopf bifurcation point. Here, γ plays the same role as λ from (5.3) in the continuation, meaning that it is updated manually.

The integral representing the distributed delay was approximated through a Clenshaw-Curtis quadrature [98] rescaled to the interval $[-3, -1]$ (see Subsection 3.1.5).

Starting from the exact solution at $\gamma = 4$, the branch of periodic orbits was continued up to $\gamma = 4.327$, corresponding to the first period doubling after the Hopf bifurcation, Figure 5.3.

The continuation was performed using trivial phase condition with $\hat{x} = \sigma$. Both Chebyshev and Gauss-Legendre collocation points were used to test convergence. Figure 5.5 confirms the $O(h^m)$ behavior for both the continuous and discrete errors for the former, while Figure 5.4 confirms the $O(h^{m+1})$ behavior for the latter.

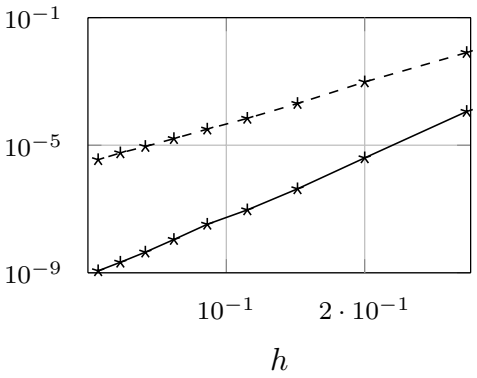

Figure 5.5: Periodic solution of (5.4) at $\gamma = 4.327$: continuous error for $m = 3$ (dashed line) and $m = 5$ (solid line) using Gauss-Legendre points. Original figure from [9][1].

5.4 PERIODIC BVP FOR COUPLED SYSTEMS

Thanks to the scalar phase condition introduced in Section 2.2, the dimension of system (5.1) is as high as the number of unknowns, regardless of the quantities d_X and d_Y. This is a necessary condition for the corresponding numerical problem to be well-posed.

It is, in principle, not sufficient, in the sense that poor choices of the phase conditions might lead to a numerical problem having multiple solutions or no solutions at all. This can happen, e.g., if $\hat{x}$ (or $\hat{y}$) and k defined as in Section 2.2 are chosen so that any periodic solution of the original equation satisfies $x_k(t) = \hat{x}$ (or $y_k(t) = \hat{y}$) for multiple values of $t \in [0, 1)$, or for no values at all. However, in the generic case of a single DDE or a single RE, finding a suitable phase condition is not an issue.

A general assumption when dealing with problems that are solved through Newton's method is that the Fréchet derivative of the original problem has a bounded inverse in the solution (recall Assumption 3.24). In many realistic cases, the assumption seems to guarantee the convergence of Newton's method. However, the question is not so trivial in the general case $d_X + d_Y \geq 2$, which includes, in particular, all coupled systems. For instance, in the extreme case of a system given by two completely independent equations (with the exception of the period ω, which needs to be the same for the two corresponding solutions), local uniqueness cannot hold when using one of the typical phase conditions, which only involve one of the components. In other words, the sought solution is not isolated. Keeping this example in mind, one can conclude that the various components of the system must be interdependent to some extent, in order to apply (the natural extension of) the method in [53].

This complicates the search for a possible coupled system to test the method on, especially if requiring the corresponding periodic solutions to

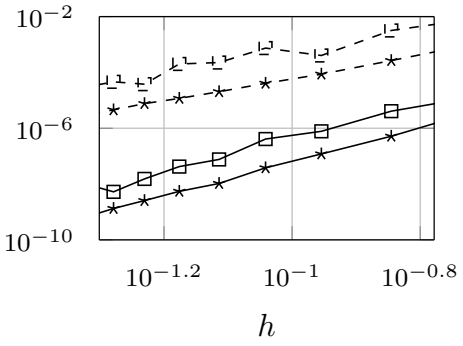

Figure 5.6: Periodic solution of (5.6) at $\gamma = 4.327$: continuous error for $m = 3$ (dashed line) and $m = 5$ (solid line) of the x-component (stars) and the y-component (squares), using Gauss-Legendre points.

have an exact expression (in order to monitor the true error). An example of coupled system leading to an ill-posed periodic BVP is

$$\begin{cases} x(t) = \frac{\gamma}{2}\int_{-3}^{-1} x(t+\theta)(1-x(t+\theta))\,\mathrm{d}\theta, & t \in [0,\omega], \\ y'(t) = \gamma x(t)(x(t-1)(1-x(t-1)) - x(t-3)(1-x(t-3)), & t \in [0,\omega], \\ (x_0, y_0) = (x_\omega, y_\omega), \\ p(x|_{[0,\omega]}) = 0, \end{cases} \tag{5.6}$$

defined starting from (5.4). Indeed, for all $y_0 \in \mathbb{R}$, the functions given by (5.5) and

$$y(t) = x^2(t) + y_0$$

constitute a continuum of solutions of (5.6), and the corresponding Newton's method was not able to reach convergence within the tests run by the author. However, one could think of using a 2-dimensional phase condition, at the expense of one collocation condition (in order to keep the right number of equations) although, normally, this would not be a good idea (see Remark 5.2). Figure 5.6 shows the error obtained, after this modification, by continuing the branch of periodic orbits from $\gamma = 4$ up to $\gamma = 4.327$, as done in Section 5.3. The continuation was performed using a trivial phase condition with $\hat{x} = \sigma$ and $\hat{y} = \sigma^2$, and removing the last collocation condition relevant to the DDE. Both Chebyshev and Gauss-Legendre collocation points were used to test convergence. In particular, Figure 5.6 confirms the $O(h^{m+1})$ behavior in the latter case.

On the other hand, the test described below (code available at `http://cdlab.uniud.it/software#per-sol`) shows that the method in [53] can be

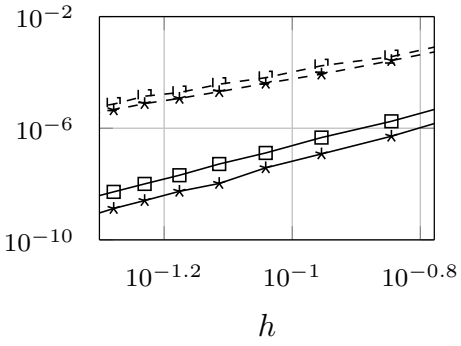

Figure 5.7: Periodic solution of (5.7) at $\gamma = 4.327$: continuous error for $m = 3$ (dashed line) and $m = 5$ (solid line) of the x-component (stars) and the y-component (squares), using Gauss-Legendre points.

extended naturally to well-posed periodic BVPs defined from coupled systems. This is the case, for instance, of the BVP

$$\begin{cases} x(t) = \frac{\gamma}{2}\int_{-3}^{-1} x(t+\theta)(1-x(t+\theta))\,\mathrm{d}\theta, \\ y'(t) = \gamma x(t)(x(t-1)(1-x(t-1)) - x(t-3)(1-x(t-3)) + y(t), \\ (x_0, y_0) = (x_\omega, y_\omega), \\ p(x|_{[0,\omega]}) = 0, \end{cases} \tag{5.7}$$

defined for $t \in [0, \omega]$, starting from (5.6). Indeed, thanks to the variation of constant formula (2.3), the unique ω-periodic solution is given by (5.5) and

$$y(t) = \frac{2e^4}{1-e^4} \cdot \int_0^4 e^{-\theta} x'(t+\theta)x(t+\theta)\,\mathrm{d}\theta.$$

Figure 5.7 shows the error obtained when continuing the branch of periodic orbits from $\gamma = 4$ up to $\gamma = 4.327$, as done in Section 5.3. The continuation was performed using a trivial phase condition with $\hat{x} = \sigma$. Both Chebyshev and Gauss-Legendre collocation points were used to test convergence. In particular, Figure 5.7 confirms the $O(h^{m+1})$ behavior in the latter case.

In conclusion, the method in [53] can be extended to any well-posed periodic BVP, leading to the same order of convergence obtained in the case of periodic BVPs defined from DDEs only. However, the role of the phase condition in determining the well-posedeness of the BVP is still being investigated, as well as some correspondence between the well-posedeness of the theoretical problem and that of the numerical one. Such a correspondence is, indeed, not necessarily trivial even for ODEs (see, e.g., [20]).

Remark 5.2. A more natural way to treat ill-posed problems such as (5.6) consists in using a 2-dimensional phase condition while adding a new variable, in order to obtain again the same number of equations and variables. For instance, one could use a trivial phase condition with $\hat{x} = \sigma$ and $\hat{y} = \sigma^2$ while adding a variable d (which will be 0) to the relevant DDE as

$$y'(t) = \gamma x(t)(x(t-1)(1-x(t-1)) - x(t-3)(1-x(t-3)) + d, \quad t \in [0, \omega].$$

Figure 5.8 shows the error obtained, after this modification, by continuing the branch of periodic orbits from $\gamma = 4$ up to $\gamma = 4.327$, as done in Section

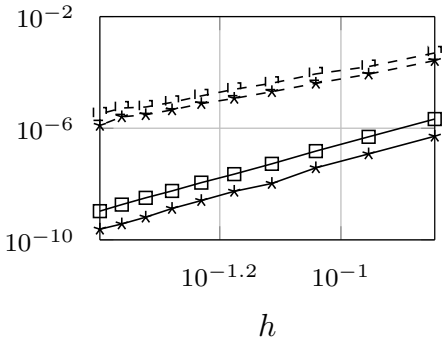

Figure 5.8: Periodic solution of (5.6) at $\gamma = 4.327$: continuous error for $m = 3$ (dashed line) and $m = 5$ (solid line) of the x-component (stars) and the y-component (squares), using Gauss-Legendre points.

5.3. Both Chebyshev and Gauss-Legendre collocation points were used to test convergence. In particular, Figure 5.8 confirms the $O(h^{m+1})$ behavior in the latter case. A special thanks goes to one of the referees, who indicated this alternative. ◁

6 THEORETICAL CONVERGENCE OF THE COLLOCATION METHOD

This chapter represents one of the main original contributions of this thesis from a theoretical point of view, and most of it is included in a work which was recently submitted [10]. Such work is highly reliant on the paper [79], where a general framework for solving a certain class of BVPs is presented, and accompanied by a rigorous proof of convergence of the corresponding iterative method. In particular, it concerns BVPs for *neutral* delay differential equations, i.e., functional equations which involve also evaluation in the past of the derivative.

The goal of this work is to adapt the approach in [79] to compute periodic solutions of (non-neutral) DDEs, which, after scaling time through (2.10), can be written in the form

$$y'(t) = \omega G(y_t \circ s_\omega), \quad t \in [0,1], \tag{6.1}$$

where G is defined on some state space $\mathtt{Y}$ as described at the beginning of Chapter 1, and s_ω is defined as in (2.10). Recall that $\mathtt{Y}$ does not need to be a space of continuous functions. Indeed, from now on, $\mathtt{Y}$ will be a generic subspace of $\mathbb{F}([-\tau,0],\mathbb{R}^{d_Y})$. Although the problem of computing periodic solutions of DDEs has already received some consideration in literature, little has been done on the theoretical analysis of the error and the convergence of the relevant iterative methods (see Section 6.1). Indeed, as far as the author's knowledge goes, this would be the first work addressing the problem of convergence of piecewise collocation methods for the computation of periodic solutions of general DDEs of the form (1.5), with no limitations on the number or type of delays or constraints on the relation between delays and period.

Recalling the formulations (2.11) and (2.13) of a BVP for a delay system, the correspondent formulations for the DDE (6.1) would read

$$\begin{cases} y'(t) = \omega G(y_t \circ s_\omega), \quad t \in [0,1], \\ y_0 = y_1 \\ p(y|_{[0,1]}) = 0, \end{cases} \tag{6.2}$$

where the periodicity condition is on the state and the solution is intended in $[-1,1]$, and

$$\begin{cases} y'(t) = \omega G(\overline{y}_t \circ s_\omega), \quad t \in [0,1], \\ y(0) = y(1) \\ p(y) = 0, \end{cases} \tag{6.3}$$

where the use of periodic states allows to formulate a finite dimensional periodicity condition and the solution is intended in $[0,1]$. Recall that derivatives with respect to time are defined from the right in the context of DDEs.

The contributions of this research consist in the proofs of the validity of the assumptions required to apply the abstract approach of [79] in the case of periodic BVPs. The assumptions concerning exclusively the formulation of the original problem will be addressed in Section 6.3, while the ones concerning the discretization method will be treated in Section 6.4.

Although the general BVP in [79] considers unknown parameters explicitly, in the periodic case the period plays the role of the (main) unknown parameter of the problem. The troubles in the effort of validating the assumptions are mostly due to the special role that the period plays in the BVP, since it is directly linked to the course of time through (2.10). This also affects the regularity that must be required from the functionals involved, and therefore the choice of the relevant spaces where the solution, its derivative or the states must lie.

The more technical parts of the relevant proofs will be presented separately in Chapter 7, which plays the role of an appendix for this Chapter.

6.1 STATE OF ART

[79, Section 1.1] contains an exhaustive description of the literature on the BVPs for (neutral) functional differential equations. As far as non-neutral differential equations are concerned, a brief introduction of the two equivalent alternatives can be found in [53, Section 2], where it is also recalled that (6.2) is an instance of *Halanay's BVP* (so named in [68]), i.e., a BVP with boundary condition of the form

$$\mathcal{N}(y_0, y_\omega) = 0$$

for some $\mathcal{N} : \mathtt{Y} \times \mathtt{Y} \to \mathtt{Y}$. However, the majority of works address formulation (6.3), e.g., [12, 13, 14, 15, 16, 17, 52, 53, 73, 77, 78, 79, 86], while only few treat the BVP using formulation (6.2) [54, 74, 101].

Among the works cited above, very few include a theoretical proof of the convergence of the method, e.g., [13, 52]. In particular, [13] does not consider the presence of unknown parameters, and [52], despite dealing explicitly with periodicity, assumes the period to be known (and equal to 1) and restricts its analysis to linear problems. Moreover, neither of the works considers a general right hand side, but rather one containing only (a finite number of) discrete delays.

The approach proposed in [79] is, on the other hand, very general and abstract, while two more concrete instances of the method are illustrated in [77, 78]. In particular, in the former the problem is discretized through collocation, in the latter through the *Fourier series method*. However, in all cases the treatment is devoted to general BVPs, not necessarily restricted to the periodic case, and not addressing explicitly the presence of the period as a parameter.

For these reasons the work described in this chapter aims at applying this general approach to both (6.2) and (6.3). Note that the latter, through the definition of periodic state, is the periodic instance of the *side condition* considered in [79] (eq. (7), page 526), while the former is not even mentioned.

In fact, the entire analysis is carried out while assuming that the boundary condition is finite-dimensional, although it is mentioned that it can also be applied to cases where it is infinite-dimensional, by adding further discretization. In spite of this, it will be shown in Section 6.3 that only (6.2) is amenable of the treatment in [79]. Therefore, in the subsequent Sections the proofs will only be given for (6.2), reserving to comment about (6.3) up to the point where it fails to fit into [79].

6.2 THE PROBLEM IN ABSTRACT FORM

This section describes the general form of the BVP addressed in [79], and shows how both formulations (6.2) and (6.3) of the periodic boundary value problem can fit into it.

The BVP considered in [79] has the form

$$\begin{cases} u = \mathcal{F}(\mathcal{G}(u,\alpha), u, \beta) \\ \mathcal{B}(\mathcal{G}(u,\alpha), u, \beta) = 0, \end{cases}$$

and its relevant solution is $v := \mathcal{G}(u,\alpha)$, which lies in a normed space $\mathbb{V} \subseteq \mathbb{F}([a,b],\mathbb{R}^d)$, while u is its derivative and lies in a Banach space $\mathbb{U} \subseteq \mathbb{F}([a,b],\mathbb{R}^d)$. The operator $\mathcal{G} : \mathbb{U} \times \mathbb{A} \to \mathbb{V}$ represents a (linear) Green operator which reconstructs the solution $v = \mathcal{G}(u,\alpha)$ given its derivative u and some α in a Banach space $\mathbb{A}$ which plays the role of an initial state or initial value. A classic example of the second instance is

$$\mathcal{G}(u,\alpha)(t) := \alpha + \int_c^t u(s)\,\mathrm{d}s, \quad t \in [a,b],$$

for some $c \in [a,b]$.

Finally, β is a vector of parameters which vary together with the solution and live in a Banach space $\mathbb{B}$.

The first line defines the functional equation of neutral type from the function $\mathcal{F} : \mathbb{V} \times \mathbb{U} \times \mathbb{B} \to \mathbb{U}$, which represents its right-hand side. The second line represents the boundary condition through a function $\mathcal{B} : \mathbb{V} \times \mathbb{U} \times \mathbb{B} \to \mathbb{A} \times \mathbb{B}$, which usually includes a classical boundary condition (the component in $\mathbb{A}$) and an extra condition which poses the necessary constraints on the parameters (the component in $\mathbb{B}$) .

The method in [79] is based on the translation of the BVP into a fixed point problem. The *Problem in Abstract Form* (PAF) consists in finding $(v^*,\beta^*) \in \mathbb{V} \times \mathbb{B}$ with $v^* := \mathcal{G}(u^*,\alpha^*)$ and $(u^*,\alpha^*,\beta^*) \in \mathbb{U} \times \mathbb{A} \times \mathbb{B}$ such that

$$(u^*,\alpha^*,\beta^*) = \Phi(u^*,\alpha^*,\beta^*) \tag{6.4}$$

for $\Phi : \mathbb{U} \times \mathbb{A} \times \mathbb{B} \to \mathbb{U} \times \mathbb{A} \times \mathbb{B}$ given by

$$\Phi(u,\alpha,\beta) := \begin{pmatrix} \mathcal{F}(\mathcal{G}(u,\alpha), u, \beta) \\ (\alpha,\beta) - \mathcal{B}(\mathcal{G}(u,\alpha), u, \beta) \end{pmatrix}. \tag{6.5}$$

In the sequel, the apex * will denote quantities relevant to fixed points.

6.2.1 Equivalent formulations

This subsection shows that (6.2) and (6.3) are both instances of (6.4).

In the case of (6.3) the domain of the BVP is $[a,b] = [0,1]$. $\mathbb{U} = \mathbb{U}_1$ and $\mathbb{V} = \mathbb{V}_1$ are chosen so that $\mathbb{U}_1, \mathbb{V}_1 \subseteq \mathbb{F}([0,1], \mathbb{R}^d)$, while $\mathbb{A} = \mathbb{A}_1 = \mathbb{R}^d$. The only unknown parameter is the original period, therefore $\mathbb{B} = \mathbb{B}_1 = \mathbb{R}$. In the sequel, ω will be used in place of β. The Green operator $\mathcal{G} = \mathcal{G}_1$ is chosen as the operator $\mathcal{G}_1 : \mathbb{U}_1 \times \mathbb{A}_1 \to \mathbb{F}([0,1], \mathbb{R}^d)$ defined as

$$\mathcal{G}_1(u, \alpha)(t) := \alpha + \int_0^t u(s)\, \mathrm{d}s, \quad t \in [0,1]. \tag{6.6}$$

The solutions of (2.13) are exactly the pairs $(v^*, \omega^*) \in \mathbb{V}_1 \times \mathbb{B}_1$ with $v^* := \mathcal{G}_1(u^*, \alpha^*)$ and $(u^*, \alpha^*, \omega^*) \in \mathbb{U}_1 \times \mathbb{A}_1 \times \mathbb{B}_1$ the fixed points of the map $\Phi_1 : \mathbb{U}_1 \times \mathbb{A}_1 \times \mathbb{B}_1 \to \mathbb{U}_1 \times \mathbb{A}_1 \times \mathbb{B}_1$ defined by

$$\Phi_1(u, \alpha, \omega) := \begin{pmatrix} \omega G(\overline{\mathcal{G}_1(u,\alpha)_{(\cdot)}} \circ s_\omega) \\ \mathcal{G}_1(u,\alpha)(1) \\ \omega - p(\mathcal{G}_1(u,\alpha)) \end{pmatrix}.$$

Above α plays the role of the initial value $v(0)$, $v_{(\cdot)}$ denotes the map $t \mapsto v_t$. Thus, with the above choices, (6.3) leads to an instance of (6.5) with $\mathcal{F} = \mathcal{F}_1 : \mathbb{V}_1 \times \mathbb{U}_1 \times \mathbb{B}_1 \to \mathbb{U}_1$ and $\mathcal{B} = \mathcal{B}_1 : \mathbb{V}_1 \times \mathbb{U}_1 \times \mathbb{B}_1 \to \mathbb{A}_1 \times \mathbb{B}_1$ given respectively by

$$\mathcal{F}_1(v, u, \omega) := \omega G(\overline{v_{(\cdot)}} \circ s_\omega)$$

and

$$\mathcal{B}_1(v, u, \omega) := \begin{pmatrix} v(0) - v(1) \\ p(v) \end{pmatrix}.$$

Note that the boundary operator is linear and only includes the periodicity and the phase conditions, none of which depend on ω. Moreover, note that, using this formulation, G needs to be defined on discontinuous functions as well. Indeed, $\mathcal{G}_1(u, \alpha)$ is generally not periodic, which means that $\overline{\mathcal{G}_1(u,\alpha)_{(\cdot)}}$ is not continuous, according to the definition of periodic state. This will be addressed in Section 6.3.

In the case of (6.2) the domain of the BVP is again $[a,b] = [0,1]$. $\mathbb{U} = \mathbb{U}_2$ and $\mathbb{V} = \mathbb{V}_2$ are chosen so that $\mathbb{U}_2 \subseteq \mathbb{F}([0,1], \mathbb{R}^d)$, and $\mathbb{V}_2 \subseteq \mathbb{F}([-1,1], \mathbb{R}^d)$ while $\mathbb{A} = \mathbb{A}_2 \subseteq \mathbb{F}([-1,0], \mathbb{R}^d)$. $\mathbb{A}$ is meant to represent a subset of the state space, which consists of functions in $\mathbb{F}([-\frac{\tau}{\omega}, 0], \mathbb{R}^d)$. However, $\mathbb{A}$ is not allowed to vary together with ω. Therefore, it must be defined as a subset of the *enlarged* state space $Y \subseteq \mathbb{F}([-1,0], \mathbb{R}^d)$, defining $y_t \in Y$ as

$$y_t(\theta) := y(t + \theta), \quad \theta \in [-1, 0]. \tag{6.7}$$

In fact, Y is enlarged thanks to the assumption $\tau \leq \omega$ (recall Section 2.2). Again, the only parameter is ω, therefore $\mathbb{B} = \mathbb{B}_2 = \mathbb{R}$. The Green operator $\mathcal{G} = \mathcal{G}_2$ is chosen as the operator $\mathcal{G}_2 : \mathbb{U}_2 \times \mathbb{A}_2 \to \mathbb{F}([-1,1], \mathbb{R}^d)$ defined as

$$\mathcal{G}_2(u, \psi)(t) := \begin{cases} \psi(0) + \int_0^t u(s)\, \mathrm{d}s, & t \in [0,1], \\ \psi(t), & t \in [-1,0], \end{cases} \tag{6.8}$$

which corresponds to the operator V first introduced in [29]. The solutions of (6.2) are exactly the pairs $(v^*, \omega^*) \in \mathbb{V}_2 \times \mathbb{B}_2$ with $v^* := \mathcal{G}_2(u^*, \psi^*)$ and $(u^*, \psi^*, \omega^*) \in \mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2$ the fixed points of the map $\Phi_2 : \mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2 \to \mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2$ defined by

$$\Phi_2(u, \psi, \omega) := \begin{pmatrix} \omega G(\mathcal{G}_2(u, \psi)_{(\cdot)} \circ s_\omega) \\ \mathcal{G}_2(u, \psi)_1 \\ \omega - p(\mathcal{G}_2(u, \psi)|_{[0,1]}) \end{pmatrix}. \tag{6.9}$$

Above ψ plays the role of the initial state v_0, and this motivates the choice of changing the notation concerning to $\mathbb{A}$, i.e., in this case it contain states $\psi \in \mathbb{A}_2 \subseteq Y$ rather than solution values $\alpha \in \mathbb{A}_1 = \mathbb{R}^d$. With these choices it follows that (6.2) leads to an instance of (6.5) with $\mathcal{F} = \mathcal{F}_2 : \mathbb{V}_2 \times \mathbb{U}_2 \times \mathbb{B}_2 \to \mathbb{U}_2$ and $\mathcal{B} = \mathcal{B}_2 : \mathbb{V}_2 \times \mathbb{U}_2 \times \mathbb{B}_2 \to \mathbb{A}_2 \times \mathbb{B}_2$ given respectively by

$$\mathcal{F}_2(v, u, \omega) := \omega G(v_{(\cdot)} \circ s_\omega) \tag{6.10}$$

and

$$\mathcal{B}_2(v, u, \omega) := \begin{pmatrix} v_0 - v_1 \\ p(v|_{[0,1]}) \end{pmatrix}. \tag{6.11}$$

Again, the boundary operator is linear and independent of either u or ω.

Note that the need for the relevant Banach spaces to remain fixed and independent from the values of the parameters is also the main reason why the change of variable (2.10) is needed. Indeed, without it, the domain of the BVP would be $[0, \omega]$ and the spaces would be consequently defined according to the current (unknown) value of ω.

6.3 VALIDATION OF THE THEORETICAL ASSUMPTIONS

As shown in Subsection 6.2.1, both formulations (6.2) and (6.3) can be translated into instances of the PAF in multiple ways: indeed, in principle, different choices for the Banach spaces $\mathbb{U}_1, \mathbb{U}_2, \mathbb{V}_1, \mathbb{V}_2, \mathbb{A}_2$ (and relevant norms) are possible, according to the regularity of the sought solution. However, this does not imply that the the convergence framework in [79] can be applied either way. In fact, several theoretical assumptions are required, and their validity depends on the choices of the spaces, as well as on the regularity of the right-hand side G. This section includes the definitions of such assumptions and their statements as propositions, instanced according to the problems of interest in this chapter.

In the case of formulation (6.2), their validity will be proved under specific choices of the relevant Banach spaces (the norms of which are indicated in Section 1.6) and regularity properties of G. For ease of reference throughout the text, the corresponding hypotheses are collected below.

(T1) $\mathtt{Y} = B^\infty([-\tau, 0], \mathbb{R}^d)$, $Y = B^\infty([-1, 0], \mathbb{R}^d)$.

(T2) $\mathbb{U}_2 = B^\infty([0, 1], \mathbb{R}^d)$, $\mathbb{V}_2 = B^{1,\infty}([-1, 1], \mathbb{R}^d)$, $\mathbb{A}_2 = B^{1,\infty}([-1, 0], \mathbb{R}^d)$.

(T3) $G : \mathtt{Y} \to \mathbb{R}^d$ is Fréchet-differentiable at every $\mathtt{y} \in \mathtt{Y}$.

(T4) $G \in \mathcal{C}^1(\mathrm{Y}, \mathbb{R}^d)$.

(T5) There exist $r > 0$ and $\kappa \geq 0$ such that

$$\|DG(\mathrm{y}) - DG(v_t^* \circ s_{\omega^*})\|_{\mathbb{R}^d \leftarrow \mathrm{Y}} \leq \kappa \|\mathrm{y} - v_t^* \circ s_{\omega^*}\|_{\mathrm{Y}}$$

for every $\mathrm{y} \in \overline{B}(v_t^* \circ s_{\omega^*}, r)$, uniformly with respect to $t \in [0,1]$.

The reason why $\mathbb{V}_2$ must be contained in $B^{1,\infty}([-1,1],\mathbb{R}^d)$ will be clear in the proof of Proposition 6.2, concerning the first assumption that needs to be verified. Indeed, due to the role played by the parameter ω in the course of time, the states v_t will need to have measurable and bounded derivative for all $v \in \mathbb{V}_2$. On the other hand $\mathbb{V}_2$ cannot be restricted to $C^1([-1,1],\mathbb{R}^d)$. Even if one chose to work with $\mathrm{Y} = C^1([-\tau,0],\mathbb{R}^d)$, the derivatives of the elements of $\mathbb{V}_2$ cannot all be continuous at 0, since that would imply $\psi'(0^-) = G(\psi)$ for all $\psi \in \mathrm{Y}$.

The argument above is what motivates the choice of $\mathbb{V}_2$ (and, consequently, of $\mathbb{U}_2$ and $\mathbb{A}_2$) in (T2). Note that spaces of bounded and measurable functions are also used in the theory of DDEs, instead of the standard continuous ones, if one weakens the notion of solution (see, e.g., [45, Exercise 2.1, Chapter 0]).

It is worth mentioning that (T5) could never be satisfied if state-dependent delays were taken into account. Indeed, in that case, the expression of the partial derivative of $G(v_t \circ s_\omega)$ with respect to v would contain v' and could not be (Lipschitz) continuous with respect to v. Thus, the framework described is limited to constant delays.

Meanwhile, as far as formulation (6.3) is concerned, it will be shown that there are no possible choices for the Banach spaces which allow to satisfy all the assumptions, as well as some comments on the validity of some of them under specific choices. The first theoretical assumption concerns the Fréchet-differentiability (see Section 2.3) of the operators $\mathcal{F}$ and $\mathcal{B}$ appearing in (6.5).

Assumption 6.1 (A$\mathfrak{FB}$, [79, page 534])**.** The operators $\mathfrak{F}$ and $\mathfrak{B}$ are Fréchet-differentiable at any point $(v_0, u_0, \beta_0) \in \mathbb{V} \times \mathbb{U} \times \mathbb{B}$.

The trickiest part while proving this assumption in our case is the differentiation with respect to ω, since it involves the composition with s_ω.

Since p is linear in (6.11), so it is $\mathcal{B}$, hence it is Fréchet-differentiable. Thus, with respect to the validity of Assumption A$\mathfrak{FB}$ using the formulation (6.2), it is sufficient to prove the following.

Proposition 6.2. *Under (T1), (T2) and (T3), there exists $r \in (0, \omega^*)$ such that $\mathcal{F}_2$ in (6.10) is Fréchet-differentiable, from the right with respect to ω, at every $(\hat{v}, \hat{u}, \hat{\omega}) \in \overline{B}((v^*, u^*, \omega^*), r)$, and*

$$D\mathcal{F}_2(\hat{v}, \hat{u}, \hat{\omega})(v, u, \omega) = \mathfrak{L}_2(\cdot; \hat{v}, \hat{\omega})[v_{(\cdot)} \circ s_{\hat{\omega}}] + \omega \mathfrak{M}_2(\cdot; \hat{v}, \hat{\omega}) \tag{6.12}$$

for $(v, u, \omega) \in \mathbb{V}_2 \times \mathbb{U}_2 \times (0, +\infty)$, where, for $t \in [0,1]$,

$$\mathfrak{L}_2(t; v, \omega) := \omega DG(v_t \circ s_\omega) \tag{6.13}$$

and

$$\mathfrak{M}_2(t; v, \omega) := G(v_t \circ s_\omega) - \mathfrak{L}_2(t; v, \omega)[v_t' \circ s_\omega] \cdot \frac{s_\omega}{\omega}. \tag{6.14}$$

Recall that derivatives with respect to time are defined from the right. Thus, the derivative with respect to the period is intended from the right since the period affects the course of time in the domain of the state space through (2.10), which is increasing with respect to ω as far its argument is negative.

Note that the Fréchet derivative cannot, in any case, be continuous, as explained at the end of Section 3.5. However, the problem will be overcome thanks to its Lipschitz continuity in the solution (v^*, u^*, ω^*) (Proposition 6.7) and the fact that (v^*, u^*, ω^*) lies in a more regular subspace than the space $\mathbb{V}_2 \times \mathbb{U}_2 \times \mathbb{R}$ (Lemma 7.5 in Chapter 7).

Proof. The thesis holds once that (6.12) is proved according to Definition 2.8, i.e., $u \in \mathbb{U}, v \in \mathbb{V}$ and $\omega > 0$,

$$\begin{aligned}\|\mathcal{F}_2(\hat{v}+v, \hat{u}+u, \hat{\omega}+\omega) - \mathcal{F}_2(\hat{v}, \hat{u}, \hat{\omega}) - D\mathcal{F}_2(\hat{v}, \hat{u}, \hat{\omega})(v, u, \omega)\|_{\mathbb{U}_2} \\ = o\left(\|(v, u, \omega)\|_{\mathbb{V}_2 \times \mathbb{U}_2 \times \mathbb{B}_2}\right).\end{aligned}$$

As for the left-hand side, by using (6.10), the choice of $\mathbb{V}_2$ in (T2) leads to evaluate

$$\begin{aligned}(\hat{\omega}+\omega)G((\hat{v}+v)_t \circ s_{\hat{\omega}+\omega}) - \hat{\omega}G(\hat{v}_t \circ s_{\hat{\omega}}) - \hat{\omega}DG(\hat{v}_t \circ s_{\hat{\omega}})[v_t \circ s_{\hat{\omega}}] \\ - \omega G(\hat{v}_t \circ s_{\hat{\omega}}) + \omega DG(\hat{v}_t \circ s_{\hat{\omega}})\hat{[}v'_t \circ s_{\hat{\omega}}] \cdot s_{\hat{\omega}} \\ = (\hat{\omega}+\omega)[G((\hat{v}+v)_t \circ s_{\hat{\omega}+\omega}) - G(\hat{v}_t \circ s_{\hat{\omega}})] \\ - \hat{\omega}DG(\hat{v}_t \circ s_{\hat{\omega}})[v_t \circ s_{\hat{\omega}}] + \omega DG(\hat{v}_t \circ s_{\hat{\omega}})[\hat{v}'_t \circ s_{\hat{\omega}}] \cdot s_{\hat{\omega}}\end{aligned} \tag{6.15}$$

for $t \in [0, 1]$. From Definition 2.8, (T3) allows to write

$$G((\hat{v}+v)_t \circ s_{\hat{\omega}+\omega}) - G(\hat{v}_t \circ s_{\hat{\omega}}) = DG(\hat{v}_t \circ s_{\hat{\omega}})\xi^t + o(\|\xi^t\|_{\mathtt{Y}}) \tag{6.16}$$

for

$$\xi^t := (\hat{v}+v)_t \circ s_{\hat{\omega}+\omega} - \hat{v}_t \circ s_{\hat{\omega}}.$$

This, in turn, leads to evaluate $\xi^t(\sigma)$ for every $\sigma \in [-\tau, 0]$ given the choice of $\mathtt{Y}$ in (T1). Then (1.2) gives

$$\begin{aligned}\xi^t(\sigma) &= \hat{v}(t + s_{\hat{\omega}+\omega}(\sigma)) - \hat{v}(t + s_{\hat{\omega}}(\sigma)) + v(t + s_{\hat{\omega}+\omega}(\sigma)) \\ &= \hat{v}'(t + s_{\hat{\omega}}(\sigma))\eta(\sigma) + o(|\eta(\sigma)|) + v(t + s_{\hat{\omega}+\omega}(\sigma))\end{aligned} \tag{6.17}$$

for

$$\eta(\sigma) := s_{\hat{\omega}+\omega}(\sigma) - s_{\hat{\omega}}(\sigma),$$

which follows from Taylor's theorem (Theorem 3.2) to $\hat{v}$ thanks to the choice of $\mathbb{V}_2$ in (T2). Since

$$\eta(\sigma) = \frac{\sigma}{\hat{\omega}+\omega} - \frac{\sigma}{\hat{\omega}} = -s_{\hat{\omega}}(\sigma) \cdot \frac{\omega}{\hat{\omega}+\omega} > 0 \tag{6.18}$$

follows from (2.10), substitution into (6.17) leads to

$$\xi^t = -\hat{v}'_t \circ s_{\hat{\omega}} \cdot s_{\hat{\omega}} \cdot \frac{\omega}{\hat{\omega}+\omega} + v_t \circ s_{\hat{\omega}+\omega} + o(\omega)$$

with

$$\|\xi^t\|_{\mathtt{Y}} = O(\omega + \|v\|_{\mathbb{V}_2}).$$

Substitution first into (6.16) and then into (6.15) leads to

$$\begin{aligned}
&(\hat{\omega}+\omega)G\left((\hat{v}+v)_t \circ s_{\hat{\omega}+\omega}\right) - \hat{\omega}G(\hat{v}_t \circ s_{\hat{\omega}}) - \hat{\omega}DG(\hat{v}_t \circ s_{\hat{\omega}})[v_t \circ s_{\hat{\omega}}] \\
&\quad - \omega G(\hat{v}_t \circ s_{\hat{\omega}}) + \omega DG(\hat{v}_t \circ s_{\hat{\omega}})[\hat{v}'_t \circ s_{\hat{\omega}}] \cdot s_{\hat{\omega}} \\
&\quad = (\hat{\omega}+\omega)DG(\hat{v}_t \circ s_{\hat{\omega}})\left([-\hat{v}'_t \circ s_{\hat{\omega}}] \cdot s_{\hat{\omega}} \cdot \frac{\omega}{\hat{\omega}+\omega} + v_t \circ s_{\hat{\omega}}\right) \\
&\quad + o(\omega + \|v\|_{\mathbb{V}_2}) - \hat{\omega}DG(\hat{v}_t \circ s_{\hat{\omega}})[v_t \circ s_{\hat{\omega}}] + \omega DG(\hat{v}_t \circ s_{\hat{\omega}})[\hat{v}'_t \circ s_{\hat{\omega}}] \cdot s_{\hat{\omega}} \\
&\quad = o(\omega + \|v\|_{\mathbb{V}_2}) + O(\omega \cdot \|v\|_{\mathbb{V}_2}).
\end{aligned}$$

The thesis is now straightforward since

$$\|(v,u,\omega)\|_{\mathbb{V}_2 \times \mathbb{U}_2 \times \mathbb{B}_2} = \max\{\|v\|_{\mathbb{V}_2}, \|u\|_{\mathbb{U}_2}, \omega\}$$

by (1.14). □

As far as formulation (6.3) is concerned, it can checked whether there are any choices for $\mathbb{V}_1$ and $\mathbb{U}_1$ in (T2) such that, assuming again (T1) and (T3) as well, the theorem above still holds. To this end, one can go through the steps of the proof and look for the ones that depend on the choices in (T2). Differentiabilty of $\overline{v}_t$ is still required, even to formulate the proposition, and in general this is not satisfied at $\sigma = -\omega t$ by a solution of (6.3), due to the definition of the periodic state in (2.12). However, recall that derivatives with respect to time are only considered from the right. The key step in the proof is (6.17), where the application of Taylor's theorem is indeed subject to the differentiability of $\overline{v}_t$, but $\eta(\sigma)$ is positive in (6.18) and thus differentiability from the right is sufficient. Thus, the theorem holds under (T1), (T3) and

$$\mathbb{U}_1 = B^\infty([0,1],\mathbb{R}^d), \quad \mathbb{V}_1 = B^{1,\infty}([0,1],\mathbb{R}^d).$$

Note that, just like in the case of formulation (6.2), the spaces above cannot be restricted to $C([0,1],\mathbb{R}^d)$ and $C^1([0,1],\mathbb{R}^d)$ respectively, since the map $t \mapsto \overline{y}_t$ introduced in (2.12) is not even continuous, but rather includes a jump discontinuity of $y(0) - y(1)$ at $-t$. Right continuity is thus preserved.

The second theoretical assumption concerns the Fréchet-differentiability (see Section 2.3) of the Green operator $\mathcal{G}$ appearing in (6.5).

Assumption 6.3 (A$\mathfrak{G}$, [79, page 534])**.** The linear operator $\mathfrak{G}$ is bounded.

The following proposition concerns the validity of Assumption A$\mathfrak{G}$ using the formulation (6.2).

Proposition 6.4. *Under (T2), $\mathcal{G}_2$ is bounded.*

Proof. Thanks to the inequality

$$\begin{aligned}
\frac{\|\mathcal{G}_2(u,\psi)\|_{\mathbb{V}_2}}{\|(u,\psi)\|_{\mathbb{U}_2 \times \mathbb{A}_2}} &= \frac{\max\{\|\psi(0) + \int_0^\cdot u(s)\,\mathrm{d}s\|_\infty + \|u\|_\infty, \|\psi\|_{\mathbb{A}_2}\}}{\max\{\|u\|_{\mathbb{U}_2}, \|\psi\|_{\mathbb{A}_2}\}} \\
&\leq \frac{\max\{\|\psi\|_\infty + \|u\|_\infty + \|u\|_\infty, \|\psi\|_{\mathbb{A}_2}\}}{\max\{\|u\|_{\mathbb{U}_2}, \|\psi\|_{\mathbb{A}_2}\}},
\end{aligned}$$

which holds for all nonzero pairs $(u,\psi) \in \mathbb{U}_2 \times \mathbb{A}_2$, it can be concluded that

$$\|\mathcal{G}_2\|_{\mathbb{V}_2 \leftarrow \mathbb{U}_2} \leq 3.$$

□

Note that the choice of the relevant Banach spaces also play a role in the well-posedeness of the PAF. In fact, the latter requires the range of $\mathcal{G}$ to lie in $\mathbb{V}$. Indeed, under (T2), the derivative of $\psi \in \mathbb{A}_2$ is measurable and bounded, and so is $u \in \mathbb{U}_2$. Due to the fact that time derivatives are intended from the right, it is not required that $\psi'(0) = u(0)$: $\mathcal{G}_2$ verifies anyway the sought requirement under (T2).

Since the operator $\mathcal{G}_2$ is linear, it is Fréchet-differentiable. Consequently, Proposition 6.2 guarantees the Fréchet-differentiability of the fixed point operator (6.9) as stated next.

Corollary 6.5. *Under (T1), (T2) and (T3), there exists $r \in (0, \omega^*)$ such that Φ_2 in (6.9) is Fréchet-differentiable at every $(\hat{u}, \hat{\psi}, \hat{\omega}) \in \overline{B}((u^*, \psi^*, \omega^*), r)$, from the right with respect to ω, and*

$$D\Phi_2(\hat{u}, \hat{\psi}, \hat{\omega})\,(u, \psi, \omega) = \begin{pmatrix} \mathfrak{L}_2(\cdot; \mathcal{G}_2(\hat{u}, \hat{\psi}), \hat{\omega})[\mathcal{G}_2(u, \psi)_{(\cdot)} \circ s_{\hat{\omega}}] + \omega \mathfrak{M}_2(\cdot; \mathcal{G}_2(\hat{u}, \hat{\psi}), \hat{\omega}) \\ \mathcal{G}_2(u, \psi)_1 \\ \omega - p(\mathcal{G}_2(u, \psi)|_{[0,1]}) \end{pmatrix}$$

for $(u, \psi, \omega) \in \mathbb{U}_2 \times \mathbb{A}_2 \times (0, +\infty)$, $\mathfrak{L}_2$ in (6.13) and $\mathfrak{M}_2$ in (6.14).

Proof. The only nonlinear component of Φ_2 in (6.9) is the first one, i.e., the one in $\mathbb{U}_2$ given by $\mathcal{F}_2$ in (6.10). The result is thus provided directly by Proposition 6.2. □

As far as formulation (6.3) is concerned, the same result holds, given that $\mathcal{G}_1$ is still linear, hence Fréchet-differentiable, and the range of $\mathcal{G}_1$ is in $\mathbb{V}_1$ under

$$\mathbb{U}_1 = B^\infty([0,1], \mathbb{R}^d), \quad \mathbb{V}_1 = B^{1,\infty}([0,1], \mathbb{R}^d).$$

The third theoretical assumption concerns the local Lipschitz continuity of the Fréchet derivative of the fixed point operator at the relevant fixed points.

Assumption 6.6 (Ax^*1, [79, page 536])**.** There exist $r_0 > 0$ and $L \geq 0$ such that

$$\begin{aligned} \|D\Phi(u, \alpha, \beta) - D\Phi(u^*, \alpha^*, \beta^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B} \leftarrow \mathbb{U}\times\mathbb{A}\times\mathbb{B}} \\ \leq L\|(u, \alpha, \beta) - (u^*, \alpha^*, \beta^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \end{aligned}$$

for all $(u, \alpha, \beta) \in \mathcal{B}((u^*, \alpha^*, \beta^*), r_0))$.

In the sequel $(u^*, \psi^*, \omega^*) \in \mathbb{U}_2 \times \mathbb{A}_2 \times (0, +\infty)$ is a fixed point of Φ_2 in (6.9) and y^* is the corresponding 1-periodic solution of (1.5). With respect to the validity of Assumption Ax^*1 using the formulation (6.2), the following holds.

Proposition 6.7. *Under (T1), (T2), (T3) and (T5), there exist $r_2 \in (0, \omega^*)$ and $\kappa_2 \geq 0$ such that*

$$\begin{aligned} \|D\Phi_2(u, \psi, \omega) - D\Phi_2(u^*, \psi^*, \omega^*)\|_{\mathbb{U}_2\times\mathbb{A}_2\times\mathbb{B}_2 \leftarrow \mathbb{U}_2\times\mathbb{A}_2\times(0,+\infty)} \\ \leq \kappa_2\|(u, \psi, \omega) - (u^*, \psi^*, \omega^*)\|_{\mathbb{U}_2\times\mathbb{A}_2\times\mathbb{B}_2} \end{aligned}$$

for all $(u, \psi, \omega) \in \mathcal{B}((u^*, \psi^*, \omega^*), r_2)$.

Proof. In this proof, the notations

$$v := \mathcal{G}_2(u,\psi), \quad v^* := \mathcal{G}_2(u^*,\psi^*), \quad \overline{v} := \mathcal{G}_2(\overline{u},\overline{\psi})$$

will be used for brevity. According to (1.15), the thesis is equivalent to the existence of $r_2 > 0$ and $\kappa_2 \geq 0$ such that

$$\begin{aligned}\|D\Phi_2(u,\psi,\omega)\,(\overline{u},\overline{\psi},\overline{\omega}) - D\Phi_2(u^*,\psi^*,\omega^*)(\overline{u},\overline{\psi},\overline{\omega})\|_{\mathbb{U}_2\times\mathbb{A}_2\times\mathbb{B}_2} \\ \leq \kappa_2\|(\overline{u},\overline{\psi},\overline{\omega})\|_{\mathbb{U}_2\times\mathbb{A}_2\times\mathbb{B}_2}\cdot\|(u,\psi,\omega)-(u^*,\psi^*,\omega^*)\|_{\mathbb{U}_2\times\mathbb{A}_2\times\mathbb{B}_2}\end{aligned}$$

for all $(u,\psi,\omega) \in \mathcal{B}((u^*,\psi^*,\omega^*),r_2)$ and all $(\overline{u},\overline{\psi},\overline{\omega}) \in \mathbb{U}_2\times\mathbb{A}_2\times(0,+\infty)$. Since both $\mathcal{G}_2$ and p are linear, following the proof of Corollary 6.5, it is sufficient to prove the statement with respect to the first component of $D\Phi_2$, i.e., the one in $\mathbb{U}_2$. Then, by defining

$$P(t) := \omega DG(v_t\circ s_\omega)\overline{v}_t\circ s_\omega - \omega^* DG(v_t^*\circ s_{\omega^*})\overline{v}_t\circ s_{\omega^*}, \tag{6.19}$$

$$Q(t) := \overline{\omega}[G(v_t\circ s_\omega) - G(v_t^*\circ s_{\omega^*})] \tag{6.20}$$

and

$$R(t) := -\overline{\omega}[DG(v_t\circ s_\omega)v_t'\circ s_\omega\cdot s_\omega - DG(v_t^*\circ s_{\omega^*})v_t^{*\prime}\circ s_{\omega^*}\cdot s_{\omega^*}] \tag{6.21}$$

through (6.13) and (6.14), the goal is to bound

$$|P(t)+Q(t)+R(t)|$$

for all $t\in[0,1]$ given the choice of $\mathbb{U}_2$ in (T2).

As for (6.19), it can be rewritten as

$$\begin{aligned}P(t) &= (A_1+A_2)(B_1+B_2)(C_1+C_2) - A_2B_2C_2 \\ &= A_1B_1C_1 + A_1B_1C_2 + A_1B_2C_1 + A_1B_2C_2 + \\ &\quad + A_2B_1C_1 + A_2B_1C_2 + A_2B_2C_1\end{aligned} \tag{6.22}$$

for

$$A_1 := \omega - \omega^*, \qquad A_2 := \omega^*,$$

$$B_1 := DG(v_t\circ s_\omega) - DG(v_t^*\circ s_{\omega^*}), \qquad B_2 := DG(v_t^*\circ s_{\omega^*})$$

and

$$C_1 := \overline{v}_t\circ s_\omega - \overline{v}_t\circ s_{\omega^*}, \qquad C_2 := \overline{v}_t\circ s_{\omega^*}.$$

The plan is to bound every single term A_i, B_i and C_i, $i=1,2$, to eventually get the desired bound for P. Then the same will be done for R in (6.21), while a bound for Q in (6.20) can be obtained more straightforwardly. Note that all quantities with the apex * are constant since related to the fixed point. Clearly

$$|A_1| \leq \|(u,\psi,\omega)-(u^*,\psi^*,\omega^*)\|_{\mathbb{U}_2\times\mathbb{A}_2\times\mathbb{B}_2}$$

follows from (1.14), while

$$|A_2| = \omega^*.$$

As for B_1, (T5) gives

$$\begin{aligned}\|B_1\|_{\mathbb{R}^d \leftarrow \mathtt{Y}} &\leq \kappa \|v_t \circ s_\omega - v_t^* \circ s_{\omega^*}\|_{\mathtt{Y}} \\ &\leq \kappa \|v_t \circ s_\omega - v_t^* \circ s_\omega\|_{\mathtt{Y}} + \kappa \|v_t^* \circ s_\omega - v_t^* \circ s_{\omega^*}\|_{\mathtt{Y}}.\end{aligned}$$

As for the first addend in the right-hand side above,

$$\|v_t \circ s_\omega - v_t^* \circ s_\omega\|_{\mathtt{Y}} \leq 2\|(u, \psi, \omega) - (u^*, \psi^*, \omega^*)\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2} \tag{6.23}$$

follows directly from (6.8) and (1.14) again. As for the second addend,

$$\|v_t^* \circ s_\omega - v_t^* \circ s_{\omega^*}\|_{\mathtt{Y}} \leq \|(u^*, \psi^*, \omega^*)\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2} \cdot \|s_\omega - s_{\omega^*}\|_\infty$$

follows by Lemma 7.4 in Chapter 7. Since

$$|s_\omega(\sigma) - s_{\omega^*}(\sigma)| = \left|\frac{\sigma}{\omega} - \frac{\sigma}{\omega^*}\right| \leq \frac{\tau|\omega - \omega^*|}{\omega^* \omega} \tag{6.24}$$

holds for every $\sigma \in [-\tau, 0]$, then $\|s_\omega - s_{\omega^*}\|_\infty \leq \frac{\tau|\omega - \omega^*|}{\omega^* \omega}$. Moreover, from $|\omega - \omega^*| \leq \|(u, \psi, \omega) - (u^*, \psi^*, \omega^*)\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2}$, it follows that

$$\begin{aligned}&\|v_t^* \circ s_\omega - v_t^* \circ s_{\omega^*}\|_{\mathtt{Y}} \\ &\qquad \leq \frac{\tau \|(u^*, \psi^*, \omega^*)\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2}}{\omega^*(\omega^* - r)} \cdot \|(u, \psi, \omega) - (u^*, \psi^*, \omega^*)\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2}\end{aligned} \tag{6.25}$$

for every $\omega \in \overline{B}(\omega^*, r)$, and r in (T5). Eventually,

$$\|B_1\|_{\mathbb{R}^d \leftarrow \mathtt{Y}} \leq \kappa \left(2 + \frac{\tau \|(u^*, \psi^*, \omega^*)\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2}}{\omega^*(\omega^* - r)}\right) \cdot \|(u, \psi, \omega) - (u^*, \psi^*, \omega^*)\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2}.$$

As for B_2, it is possible to define

$$\kappa_{2,1} := \max_{t \in [0,1]} \|DG(v_t^* \circ s_{\omega^*})\|_{\mathbb{R}^d \leftarrow \mathtt{Y}},$$

since the map $t \mapsto v_t^*$ is uniformly continuous and so is DG at the state corresponding to the fixed point, under (T5). This leads to the bound

$$\|B_2\|_{\mathbb{R}^d \leftarrow \mathtt{Y}} \leq \kappa_{2,1}. \tag{6.26}$$

The same arguments used above for B_1 and B_2 lead also, respectively, to

$$\|C_1\|_{\mathtt{Y}} \leq \frac{\tau \|(\overline{u}, \overline{\psi}, \overline{\omega})\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2}}{\omega^*(\omega^* - r)} \cdot \|(u, \psi, \omega) - (u^*, \psi^*, \omega^*)\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2}$$

and

$$\|C_2\|_{\mathtt{Y}} \leq 2\|(\overline{u}, \overline{\psi}, \overline{\omega})\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2}.$$

Eventually, note that every triple $A_i B_j C_k$ for $i, j, k \in \{1, 2\}$ in the last member of (6.22) contains exactly a C-term, which is either bounded by a constant times $\|(\overline{u}, \overline{\psi}, \overline{\omega})\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2}$ or by a constant times

$$\|(\overline{u}, \overline{\psi}, \overline{\omega})\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2} \|(u, \psi, \omega) - (u^*, \psi^*, \omega^*)\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2}.$$

Moreover, every triple contains at least a factor of index 1, which, in the case of A and B, are bounded by some constant times

$$\|(u, \psi, \omega) - (u^*, \psi^*, \omega^*)\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2}.$$

Therefore, for each triple there exist $\kappa_{i,j,k}$ such that

$$\|A_iB_jC_k\|_{\mathbb{U}_2} \leq \kappa_{i,j,k}\|(\overline{u},\overline{\psi},\overline{\omega})\|_{\mathbb{U}_2\times\mathbb{A}_2\times\mathbb{B}_2}\cdot\|(u,\psi,\omega)-(u^*,\psi^*,\omega^*)\|^m_{\mathbb{U}_2\times\mathbb{A}_2\times\mathbb{B}_2}$$

for some $m \in \{1,2,3\}$. Note that, for $\|(u,\psi,\omega)-(u^*,\psi^*,\omega^*)\| \leq 1$, the inequality

$$\|(u,\psi,\omega)-(u^*,\psi^*,\omega^*)\|^m \leq \|(u,\psi,\omega)-(u^*,\psi^*,\omega^*)\|$$

holds for $m \in \{1,2,3\}$. Thus, for $r_{2,P} := \min\{r/2,1\}$, where r is as in (T5), by virtue of (6.23) there exist $\kappa_{2,P} := \max_{i,j,k}\kappa_{i,j,k} \geq 0$ such that

$$\|P\|_{\mathbb{U}_2} \leq \kappa_{2,P}\|(\overline{u},\overline{\psi},\overline{\omega})\|_{\mathbb{U}_2\times\mathbb{A}_2\times\mathbb{B}_2}\cdot\|(u,\psi,\omega)-(u^*,\psi^*,\omega^*)\|_{\mathbb{U}_2\times\mathbb{A}_2\times\mathbb{B}_2}$$

for all $(u,\psi,\omega) \in \mathcal{B}((u^*,\psi^*,\omega^*),r_{2,P})$. Since every triple contains exactly one A-term, one B-term and one C-term, the constant $\kappa_{2,P}$ can be defined as

$$\begin{aligned}\kappa_{2,P} := \max\{1,\omega^*\}\cdot\max\left\{\kappa\left(2+\tfrac{\tau\|(u^*,\psi^*,\omega^*)\|_{\mathbb{U}_2\times\mathbb{A}_2\times\mathbb{B}_2}}{\omega^*(\omega^*-r)}\right),\kappa_{2,1}\right\}\\ \cdot\max\left\{\tfrac{\tau}{\omega^*(\omega^*-r)},2\right\}.\end{aligned}$$

A bound for the term Q in (6.20) can be retrieved by noting that, under (T3), G satisfies the hypotheses of the mean value theorem, i.e., Theorem 3.2 with $n=1$, in a neighborhood of $v_t^*\circ s_{\omega^*}$, which leads, under (T5), to

$$\begin{aligned}\|Q\|_{\mathbb{U}_2} &\leq |\overline{\omega}|\max_{\mathrm{y}\in\mathcal{B}(v_t^*\circ s_{\omega^*},r)}\|DG(\mathrm{y})\|_{\mathbb{R}^d\leftarrow\mathrm{Y}}\|v_t\circ s_\omega - v_t^*\circ s_{\omega^*}\|_{\mathrm{Y}}\\ &\leq \|(\overline{u},\overline{\psi},\overline{\omega})\|_{\mathbb{U}_2\times\mathbb{A}_2\times\mathbb{B}_2}(\kappa r+\kappa_{2,1})\|v_t\circ s_\omega - v_t^*\circ s_{\omega^*}\|_{\mathrm{Y}}.\end{aligned}$$

Then, by (6.23) and (6.25), for $r_{2,Q} := r/2$ and

$$\kappa_{2,Q} := (\kappa r+\kappa_{2,1})\left(2+\frac{\tau\|(u^*,\psi^*,\omega^*)\|_{\mathbb{U}_2\times\mathbb{A}_2\times\mathbb{B}_2}}{\omega^*(\omega^*-r)}\right)$$

the bound

$$\|Q\|_{\mathbb{U}_2} \leq \kappa_{2,Q}\|(\overline{u},\overline{\psi},\overline{\omega})\|_{\mathbb{U}_2\times\mathbb{A}_2\times\mathbb{B}_2}\cdot\|(u,\psi,\omega)-(u^*,\psi^*,\omega^*)\|_{\mathbb{U}_2\times\mathbb{A}_2\times\mathbb{B}_2}$$

holds for all $(u,\psi,\omega) \in \mathcal{B}((u^*,\psi^*,\omega^*),r_{2,Q})$.

The term R in (6.21), as well as the term P above, can be written in the form

$$\begin{aligned}R(t) &= -\overline{\omega}[(B_1+B_2)(D_1+D_2)(E_1+E_2)-B_2D_2E_2]\\ &= -\overline{\omega}[B_1D_1E_1+B_1D_1E_2+B_1D_2E_1+B_1D_2E_2+\\ &\quad +B_2D_1E_1+B_2D_1E_2+B_2D_2E_1]\end{aligned} \tag{6.27}$$

for the same B_1 and B_2 above plus

$$D_1 := v_t'\circ s_\omega - v_t^{*\prime}\circ s_{\omega^*}, \qquad D_2 := v_t^{*\prime}\circ s_{\omega^*}$$

and

$$E_1 := s_\omega - s_{\omega^*}, \qquad E_2 := s_{\omega^*}.$$

$R(t)$ is the most subtle term in proving the Proposition. The sought bounds can be obtained thanks to the fact that the fixed point lies indeed

in a more regular subspace than $\mathbb{U}_2 \times \mathbb{A}_2 \times [0, +\infty)$ (see Lemma 7.5). In particular, D_1 can be bounded as

$$\begin{aligned}\|D_1\|_{\mathtt{Y}} &\leq \|v_t' \circ s_\omega - v_t^{*\prime} \circ s_\omega\|_{\mathtt{Y}} + \|v_t^{*\prime} \circ s_\omega - v_t^{*\prime} \circ s_{\omega^*}\|_{\mathtt{Y}} \\ &\leq 2\|(u, \psi, \omega) - (u^*, \psi^*, \omega^*)\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2} \\ &+ \|(u^{*\prime}, \psi^{*\prime})\|_{\mathbb{U}_2 \times \mathbb{A}_2} \cdot \|s_\omega - s_{\omega^*}\|_\infty,\end{aligned}$$

where the inequality for the former addend follows from (6.8) and (1.14), and the one for the latter follows from Lemma 7.5. By (6.24) and the inequality $|\omega - \omega^*| \leq \|(u, \psi, \omega) - (u^*, \psi^*, \omega^*)\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2}$, this implies

$$\|D_1\|_{\mathtt{Y}} \leq \left(2 + \frac{\tau\|(u^{*\prime}, \psi^{*\prime})\|_{\mathbb{U}_2 \times \mathbb{A}_2}}{\omega^*(\omega^* - r)}\right) \cdot \|(u, \psi, \omega) - (u^*, \psi^*, \omega^*)\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2}.$$

D_2, instead, is bounded by a constant, namely

$$\|D_2\|_{\mathtt{Y}} \leq 2\|(u^*, \psi^*, \omega^*)\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2}.$$

As for E_1, the bound

$$\|E_1\|_\infty \leq \frac{\tau}{\omega^*(\omega^* - r)} \cdot \|(u, \psi, \omega) - (u^*, \psi^*, \omega^*)\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2}$$

follows directly from (6.24), while E_2 is bounded by a constant, that is,

$$\|E_2\|_\infty \leq \frac{\tau}{\omega^*}.$$

Eventually, note that every triple $B_i D_j E_k$ for $i, j, k \in \{1, 2\}$ in the last member of (6.27) contains a factor of index 1, which are always bounded by some constant times $\|(u, \psi, \omega) - (u^*, \psi^*, \omega^*)\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2}$, while the factors of index 2 are simply bounded by constants. Arguing as done above for P, for $r_{2,R} := \min\{1, r/2\}$ there exists $\kappa_{2,R} \geq 0$ such that

$$\|R\|_{\mathbb{U}_2} \leq \kappa_{2,R}\|(\overline{u}, \overline{\psi}, \overline{\omega})\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2} \cdot \|(u, \psi, \omega) - (u^*, \psi^*, \omega^*)\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2}$$

for all $(u, \psi, \omega) \in \mathcal{B}((u^*, \psi^*, \omega^*), r_{2,R})$. In particular, $\kappa_{2,R}$ can be defined as

$$\begin{aligned}\kappa_{2,R} := \max&\left\{\kappa\left(2 + \tfrac{\tau\|(u^*, \psi^*, \omega^*)\|_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2}}{\omega^*(\omega^* - r)}\right), \kappa_{2,1}\right\} \\ \cdot \max&\left\{\left(2 + \tfrac{\tau\|(u^{*\prime}, \psi^{*\prime})\|_{\mathbb{U}_2 \times \mathbb{A}_2}}{\omega^*(\omega^* - r)}\right), 2\right\} \cdot \tfrac{\tau}{\omega^*(\omega^* - r)}.\end{aligned}$$

The thesis eventually follows by choosing

$$r_2 = \min\{r_{2,P}, r_{2,Q}, r_{2,R}\} = \min\{1, r/2\}, \quad \kappa_2 = \kappa_{2,P} + \kappa_{2,Q} + \kappa_{2,R}.$$

□

Remark 6.8. Observe that the Lipschitz constant κ_2 grows unbounded as $\omega^* \to 0$ due to the presence of the latter at the denominator of several of its terms. ◁

As far as formulation (6.3) is concerned, for the previous choice

$$\mathbb{U}_1 = B^\infty([0,1],\mathbb{R}^d), \quad \mathbb{V}_1 = B^{1,\infty}([0,1],\mathbb{R}^d)$$

the above proof would fail because of the relevant term C_1, i.e.,

$$\overline{\mathcal{G}_1(\overline{u},\overline{\alpha})_t} \circ s_\omega - \overline{\mathcal{G}_1(\overline{u},\overline{\alpha})_t} \circ s_{\omega^*}.$$

Indeed, as already observed, the function $\overline{\mathcal{G}_1(\overline{u},\overline{\alpha})_t}$ is always discontinuous at time $-t$, preventing the achievement of the necessary Lipschitz condition. Alternatively, a possible remedy is that of restricting to the spaces

$$\mathbb{U}_1 = B^\infty_\pi([0,1],\mathbb{R}^d) := \left\{ u \in B^\infty([0,1],\mathbb{R}^d) \ : \ \int_0^1 u(s)\,\mathrm{d}s = 0 \right\} \tag{6.28}$$

and

$$\mathbb{V}_1 = B^{1,\infty}_\pi([0,1],\mathbb{R}^d := \{ v \in B^{1,\infty}([0,1],\mathbb{R}^d) \ : \ v(0) = v(1) \}. \tag{6.29}$$

These choices guarantee not only that $\overline{\mathcal{G}_1(\overline{u},\overline{\alpha})_t}$ is continuous, but also Lipschitz continuous thanks to the constraint of zero mean imposed to the derivative u. Indeed,

$$\overline{\mathcal{G}_1(\overline{u},\overline{\alpha})}_t(\theta_1) - \overline{\mathcal{G}_1(\overline{u},\overline{\alpha})}_t(\theta_2) = \begin{cases} \displaystyle\int_{t+\theta_2}^{t+\theta_1} \overline{u}(s)\,\mathrm{d}s, & \theta_1,\theta_2 \geq -t, \\ \displaystyle\int_0^{t+\theta_1} \overline{u}(s)\,\mathrm{d}s + \int_{1+t+\theta_2}^1 \overline{u}(s)\,\mathrm{d}s, & \theta_1 \geq -t > \theta_2, \\ \displaystyle\int_{1+t+\theta_2}^{1+t+\theta_1} \overline{u}(s)\,\mathrm{d}s, & \theta_1,\theta_2 < -t. \end{cases}$$

Thus,

$$|\overline{\mathcal{G}_1(\overline{u},\overline{\alpha})}_t(\theta_1) - \overline{\mathcal{G}_1(\overline{u},\overline{\alpha})}_t(\theta_2)| \leq |\theta_1 - \theta_2| \|\overline{u}\|_{\mathbb{U}_1}.$$

Note that the choices (6.28) and (6.29) allow to satisfy the previous assumptions as well, since $\mathbb{U}_1$ and $\mathbb{V}_1$ are subsets, respectively, of

$$\mathbb{U}_1 = B^\infty([0,1],\mathbb{R}^d), \quad \mathbb{V}_1 = B^{1,\infty}([0,1],\mathbb{R}^d),$$

defined in the previous choice, but share the same norm.

Finally, note that in principle there can be other ways to deal with the need of restricting to spaces of functions having a null average. In [92], e.g., the problem is dealt with by considering

$$\mathcal{G}(u,\alpha) := \alpha + \int_0^t [Q_0 u](s)\,\mathrm{d}s, \qquad [Q_0 u](t) = u(t) - \int_0^1 u(s)\,\mathrm{d}s. \tag{6.30}$$

The fourth (and last) theoretical assumption concerns the well-posedness of a linear(ized) inhomogeneous version of the PAF (6.4).

Assumption 6.9 (Ax^*2, [79, page 536])**.** The linear bounded operator $I - D\Phi(u^*,\alpha^*,\beta^*)$ is invertible, i.e., for any $(u_0,\alpha_0,\beta_0) \in \mathbb{U} \times \mathbb{A} \times \mathbb{B}$ the linear problem

$$\begin{cases} u = D\mathfrak{F}^*(\mathcal{G}(u,\alpha),u,\alpha) + u_0 \\ D\mathfrak{B}^*(\mathcal{G}(u,\alpha),u,\alpha) = (\alpha_0,\beta_0), \end{cases}$$

where $D\mathfrak{F}^* = D\mathfrak{F}(v^*,u^*,\beta^*)$ and $D\mathfrak{B}^* = D\mathfrak{B}(v^*,u^*,\beta^*)$, has a unique solution $(u,\alpha,\beta) \in \mathbb{U} \times \mathbb{A} \times \mathbb{B}$.

One of the requirements for its validity is a consequence of *hyperbolicity* of the concerned periodic solution, which is a standard assumption in the context of application of the principle of linearized stability (see, e.g., [45, Chapter XIV]) or [61, Chapter 10]), in which one derives information on the stability of such solution by investigating the stability of the zero solution of (6.1) linearized around the periodic solution itself. The linearization of (1.5) around the ω^*-periodic solution y^* leads to the linear homogeneous DDE

$$y'(t) = \mathfrak{L}_2(t; v^*, \omega^*)[y_t \circ s_{\omega^*}] \tag{6.31}$$

for $\mathfrak{L}_2$ in (6.13). Under (T4) the associated initial value problem is well-posed, and so is $T_2^*(t,s) : Y \to Y$, the relevant evolution operator (see Section 1.1) for $s \in \mathbb{R}$ and $t \geq s$. $T_2^*(1,0)$ represents thus the corresponding monodromy operator. Then hyperbolicity implies the required additional hypothesis of 1 being a simple Floquet multiplier, i.e., a simple eigenvalue of $T_2^*(1,0)$, besides having no other Floquet multipliers on the unit circle.

Remark 6.10. 1 is always a Floquet multiplier due to linearization. Indeed, as a general fact the derivative of a solution of a nonlinear problem is always a solution of the problem obtained by linearizing around this solution. For instance, if y^* is a solution of (1.5), then $y^{*\prime}$ satisfies equation (6.31). But if the original solution is periodic then so is its derivative, i.e., it is an eigenvector of the monodromy operator with respect to the eigenvalue 1. ◁

In the following, the notation will refer to the one of Proposition 6.7. It is also convenient to introduce the abbreviations

$$\mathfrak{L}^* := \mathfrak{L}(\cdot; v^*, \omega^*), \qquad \mathfrak{M}^* := \mathfrak{M}(\cdot; v^*, \omega^*). \tag{6.32}$$

Concerning the validity of Ax^*2, the following will be proved.

Proposition 6.11. *Under (T1), (T2) and (T4), if $1 \in \sigma(T_2^*(1,0))$ is simple, then the linear bounded operator $I_{\mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2} - D\Phi_2(u^*, \psi^*, \omega^*)$ is invertible, i.e., for all $(u_0, \psi_0, \omega_0) \in \mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2$ there exists a unique $(u, \psi, \omega) \in \mathbb{U}_2 \times \mathbb{A}_2 \times \mathbb{B}_2$ such that*

$$\begin{cases} u = \mathfrak{L}_2^*[\mathcal{G}_2(u,\psi)_{(\cdot)} \circ s_{\omega^*}] + \omega \mathfrak{M}_2^* + u_0 \\ \psi = \mathcal{G}_2(u,\psi)_1 + \psi_0 \\ p(\mathcal{G}_2(u,\psi)|_{[0,1]}) = \omega_0. \end{cases} \tag{6.33}$$

The proof will be divided in two parts. The first part follows next, and is rather technical. The second one, on the other hand, is not as immediate, despite its appearence: in fact, it only concerns the outrule of a nongeneric case ($k_1 = 0$ in the proof below). Since this "secondary" proof represents a main contribution of this thesis, it will be presented afterwards in a dedicated Subsection, and the fact will be taken as granted until then, so that its analysis can be expanded in detail without interrupting the main reading flow. Below $\sigma(\mathcal{A})$ denotes the spectrum of an operator $\mathcal{A}$, i.e., the set

$$\{\lambda \in \mathbb{C} : \lambda I - \mathcal{A} \text{ is not invertible}\}.$$

Proof. The proof is based on treating (6.33) as an initial value problem for $v = \mathcal{G}_2(u, \psi)$, i.e.,

$$\begin{cases} v'(t) = \mathfrak{L}_2^*(t)[v_t \circ s_{\omega^*}] + \omega \mathfrak{M}_2^*(t) + u_0(t) \\ v_0 = \psi \end{cases} \tag{6.34}$$

for $t \in [0,1]$, imposing then the boundary conditions in (6.33). Being the DDE in (6.34) linear inhomogeneous with continuous linear part under (T4), for every $\psi \in \mathbb{A}_2$ there exists a unique solution v whose state can be expressed through the variation of constants formula (2.3)

$$v_t = T_2^*(t,0)\psi + \int_0^t [T_2^*(t,s)X_0][\omega\mathfrak{M}_2^*(s) + u_0(s)]\,\mathrm{d}s, \quad t \in [0,1],$$

see Theorem 2.1. The first boundary condition in (6.33) gives then

$$\psi = T_2^*(1,0)\psi + \int_0^1 [T_2^*(1,s)X_0][\omega\mathfrak{M}_2^*(s) + u_0(s)]\,\mathrm{d}s + \psi_0. \tag{6.35}$$

Let now R and K be, respectively, the range and the kernel of $I_Y - T_2^*(1,0)$. Then, from a well-known result in spectral theory (see, e.g., [45, Theorem 2.5, Chapter IV]),

$$Y = R \oplus K \tag{6.36}$$

and, by the hypothesis on the multiplier 1, $K = \mathrm{span}\{\varphi\}$ for φ an eigenfunction of the multiplier 1 itself. Moreover, it is reasonable to assume $p(v(\cdot;\varphi)|_{[0,1]}) \neq 0$ (see Remark 6.12 below), where $v(\cdot;\varphi)$ denotes the solution of (6.34) exiting from φ.

From (6.35) let us define the elements of Y

$$\xi_1^* := \int_0^1 [T_2^*(1,s)X_0]\mathfrak{M}_2^*(s)\,\mathrm{d}s, \qquad \xi_2^* := \int_0^1 [T_2^*(1,s)X_0]u_0(s)\,\mathrm{d}s + \psi_0,$$

so that (6.35) becomes

$$[I_Y - T_2^*(1,0)]\psi = \omega\xi_1^* + \xi_2^*. \tag{6.37}$$

Note that $\psi_0 \in Y$ since $\mathbb{A}_2 \subseteq Y$. From (6.36) it follows that ξ_1^* can be written uniquely as

$$\xi_1^* = r_1 + k_1\varphi, \tag{6.38}$$

where $r_1 \in R$ and $k_1 \in \mathbb{R}$. Similarly, $\xi_2^* = r_2 + k_2\varphi$. Then from (6.37) it must be $\omega\xi_1^* + \xi_2^* \in R$, which implies $\omega k_1 + k_2 = 0$. Therefore, by assuming $k_1 \neq 0$, it follows that

$$\omega = -k_2/k_1 \tag{6.39}$$

is the only possible solution. For the time being, as anticipated, $k_1 \neq 0$ will just be assumed, while the case $k_1 = 0$ will be ruled out in the forthcoming Subsection.

Eventually, let η be such that $\omega\xi_1^* + \xi_2^* = \eta - T_2^*(1,0)\eta$. Then every ψ satisfying (6.37) can be written uniquely as $\eta + \lambda\varphi$ for some $\lambda \in \mathbb{R}$. The value of λ is fixed by imposing the second boundary condition in (6.33), i.e., $p(v(\cdot;\eta)|_{[0,1]}) + \lambda p(v(\cdot;\varphi)|_{[0,1]}) = \omega_0$. Uniqueness follows from $p(v(\cdot;\varphi)|_{[0,1]}) \neq 0$. □

Remark 6.12. The condition $p(v(\cdot;\varphi)|_{[0,1]}) \neq 0$ is generic and not restrictive at all. In any case, it is always possible to change p in order to meet the above requirement. ◁

As for (6.3), it is immediate to verify that the similar result cannot be obtained under the choices (6.28) and (6.29). Indeed, these choices would require $v(1) = v(0) = \alpha$ for $v = \mathcal{G}_1(u, \alpha)$ according to (6.6), that is, for every $\alpha_0 \in \mathbb{A}_1$ one should find a unique $\alpha \in \mathbb{A}_1$ satisfying $\alpha = \mathcal{G}_1(u, \alpha)(1) + \alpha_0$, but since $u \in \mathbb{U}_1$ implies that u has zero mean, it must necessarily be $\mathcal{G}_1(u, \alpha)(1) = \alpha$, i.e., $\alpha_0 = 0$. In this sense, the approach proposed in [79] is not applicable to formulation (6.3), although, in principle, one could try to reformulate the latter using, e.g., (6.30) and replacing the periodicity condition with

$$\int_0^1 G(\mathcal{G}(u, \alpha)_t)\, \mathrm{d}t = 0.$$

The rest of the Chapter only focuses on formulation (6.2). In particular, in the remaining of the work the index 2 referring to (6.2) will be dropped to lighten the notation.

6.3.1 The nongeneric case

The content of this Subsection completes the proof of Proposition 6.11 by showing that $k_1 \neq 0$ must necessarily hold in (6.38). A key step is represented by equation (6.44) which, indeed, is an instance of the more general fact that left and right eigenvectors of simple eigenvalues are not orthogonal in some sense. The proof will be done by contradiction, assuming that $k_1 = 0$, which is equivalent to $\xi_1^* \in R$. Therefore, there exists $\gamma \in Y$ such that

$$[I_Y - T_2^*(1, 0)]\gamma = \omega \xi_1^*,$$

which amounts to say that

$$y'(t) = \mathfrak{L}_2^*(t)[y_t \circ s_{\omega^*}] + \omega \mathfrak{M}_2^*(t) \tag{6.40}$$

has a 1-periodic solution (with $y_0 = \gamma$). The proof given next requires several tools (anticipated in Section 2.1).

By (T4) and Theorem 2.2, $\mathfrak{L}_2^*$ can be expressed through the Riemann-Stieltjes integral (2.2) as

$$\mathfrak{L}_2^*(t)\psi \circ s_{\omega^*} = \int_{-\tau}^0 \mathrm{d}_\sigma \mathtt{n}^*(t, \sigma)\psi(s_{\omega^*}(\sigma)), \quad \psi \in Y,$$

where, for every $\sigma \in [-\tau, 0]$, $\mathtt{n}^*(\cdot, \sigma) : \mathbb{R} \to \mathbb{R}^{d \times d}$ is 1-periodic. Then, by (6.14), $\mathfrak{M}_2^*$ becomes

$$\begin{aligned} \mathfrak{M}_2^*(t) &= G(v_t^* \circ s_{\omega^*}) - \int_{-\tau}^0 \mathrm{d}_\sigma \mathtt{n}^*(t, \sigma) v^{*\prime}(t + s_{\omega^*}(\sigma)) \cdot \frac{s_{\omega^*}(\sigma)}{\omega^*} \\ &= \frac{1}{\omega^*} \left(v^{*\prime}(t) - \int_{-r}^0 \mathrm{d}_\theta n^*(t, \theta) v^{*\prime}(t + \theta)\theta \right) \end{aligned} \tag{6.41}$$

for $n(t, \theta) = n(t, s_{\omega^*}(\sigma)) := \mathtt{n}(t, \sigma)$ thanks to the change of variable (2.10) for $\omega = \omega^*$, and by using the fact that v^* is the 1-periodic solution of (6.1) for $\omega = \omega^*$ again.

The rest of the proof requires tools from the theory of adjoint equations for DDEs. Note that periodic solutions are defined on the whole line. Thus, through (2.4), (6.31) reads

$$y'(t) = \int_{-r}^{0} \mathrm{d}_\theta n^*(t,\theta) y(t+\theta), \tag{6.42}$$

while (2.5) reads

$$z(t) + \int_t^\infty z(\theta) n^*(\theta, t-\theta)\,\mathrm{d}\theta = \text{constant}. \tag{6.43}$$

From Theorem 2.3, it follows that

$$(z^t, y_t)_t = c$$

for some constant $c \in \mathbb{R}$ independently of t.

It was observed in Remark 6.10 that (6.42) has a 1-periodic solution (viz. $v^{*\prime}$). Under the hypothesis of Proposition 6.11, this is the only 1-periodic solution. Thus, by Lemma 2.4, the adjoint equation (6.43) has also a (unique) 1-periodic solution, say z^*. By Proposition 2.6, it follows that

$$(z^{*t}, v_t^{*\prime})_t = c^* \neq 0. \tag{6.44}$$

Integrating over one period gives

$$\begin{aligned} c^* = \int_0^1 c^*\,\mathrm{d}t = & \int_0^1 z^*(t) v^{*\prime}(t)\,\mathrm{d}t \\ & + \int_0^1 \int_{-r}^0 \mathrm{d}_\beta \left[\int_0^r z^*(t+\xi) n^*(t+\xi, \beta-\xi)\,\mathrm{d}\xi \right] v^{*\prime}(t+\beta)\,\mathrm{d}t. \end{aligned}$$

As for the last integral, thanks to the periodicity of $v^{*\prime}$, z^* and n^*, by exchanging the order of integration and by observing that

$$\int_a^b \mathrm{d}_\theta n^*(t,\theta) v^{*\prime}(t+\theta) = 0$$

whenever $a > 0$ or $b < -r$, it follows that

$$\begin{aligned}
&\int_0^1 \int_{-r}^0 \mathrm{d}_\beta \left[\int_0^r z^*(t+\xi) n^*(t+\xi, \beta-\xi)\, \mathrm{d}\xi \right] v^{*\prime}(t+\beta)\, \mathrm{d}t \\
&\qquad = \int_{-r}^0 \mathrm{d}_\beta \int_0^r \left[\int_0^1 z^*(t+\xi) n^*(t+\xi, \beta-\xi) v^{*\prime}(t+\beta)\, \mathrm{d}t \right] \mathrm{d}\xi \\
&\qquad = \int_{-r}^0 \mathrm{d}_\beta \int_0^r \left[\int_0^1 z^*(t) n^*(t, \beta-\xi) v^{*\prime}(t+\beta-\xi)\, \mathrm{d}t \right] \mathrm{d}\xi \\
&\qquad = \int_0^1 z^*(t) \int_{-r}^0 \mathrm{d}_\beta \left[\int_0^r n^*(t, \beta-\xi) v^{*\prime}(t+\beta-\xi)\, \mathrm{d}\xi \right] \mathrm{d}t \\
&\qquad = \int_0^1 z^*(t) \int_0^r \left[\int_{-r}^0 \mathrm{d}_\beta n^*(t, \beta-\xi) v^{*\prime}(t+\beta-\xi) \right] \mathrm{d}\xi\, \mathrm{d}t \\
&\qquad = \int_0^1 z^*(t) \int_0^r \left[\int_{-\xi-r}^{-\xi} \mathrm{d}_\theta n^*(t, \theta) v^{*\prime}(t+\theta) \right] \mathrm{d}\xi\, \mathrm{d}t \\
&\qquad = \int_0^1 z^*(t) \int_0^r \left[\int_{-r}^{-\xi} \mathrm{d}_\theta n^*(t, \theta) v^{*\prime}(t+\theta) \right] \mathrm{d}\xi\, \mathrm{d}t \\
&\qquad = \int_0^1 z^*(t) \int_{-r}^0 \mathrm{d}_\theta \left[\int_0^{-\theta} n^*(t, \theta) v^{*\prime}(t+\theta)\, \mathrm{d}\xi \right] \mathrm{d}t \\
&\qquad = \int_0^1 z^*(t) \int_{-r}^0 \mathrm{d}_\theta n^*(t, \theta) v^{*\prime}(t+\theta) \left[\int_0^{-\theta} \mathrm{d}\xi \right] \mathrm{d}t \\
&\qquad = -\int_0^1 z^*(t) \int_{-r}^0 \mathrm{d}_\theta n^*(t, \theta) v^{*\prime}(t+\theta) \theta\, \mathrm{d}t.
\end{aligned}$$

Finally, by using (6.41),

$$\begin{aligned}
c^* &= \int_0^1 z^*(t) v^{*\prime}(t) dt - \int_0^1 z^*(t) \int_{-r}^0 \mathrm{d}_\theta n^*(t, \theta) v^{*\prime}(t+\theta) \theta\, \mathrm{d}t \\
&= \int_0^1 z^*(t) \left(v^{*\prime}(t) - \int_{-r}^0 \mathrm{d}_\theta n^*(t, \theta) v^{*\prime}(t+\theta) \theta \right) \mathrm{d}t \qquad (6.45) \\
&= \omega^* \int_0^1 z^*(t) \mathfrak{M}_2^*(t)\, \mathrm{d}t.
\end{aligned}$$

Going back to (6.40), note that it has a 1-periodic solution if $k_1 = 0$. But then Theorem 2.7 gives $c^* = 0$ by (6.45), which is a contradiction thanks to (6.44).

6.4 VALIDATION OF THE NUMERICAL ASSUMPTIONS

As shown in Section 6.3, formulation (6.2) satisfies the theoretical assumptions required in [79] under the choices (T1) and (T2) of the relevant spaces and the hypotheses (T3), (T4) and (T5) on the regularity of the right-hand side. On the other hand, this does not hold for formulation (6.3) under any choices of the relevant spaces. Thus, in the sequel only formulation (6.2) will be considered, and the index 2 will be dropped in order to lighten the notation.

This section includes the definitions of the other assumptions required in [79], namely the *numerical* ones, and their statements as propositions in-

stanced according to formulation (6.2). Such assumptions concern the chosen discretization scheme for the numerical method, which is defined by the *primary* and *secondary* discretizations, described below.

The primary discretization consists in reducing the spaces $\mathbb{U}$ and $\mathbb{A}$ to finite-dimensional spaces $\mathbb{U}_L$ and $\mathbb{A}_L$, given a level of discretization L. This happens by means of *restriction* operators

$$\rho_L^+ : \mathbb{U} \to \mathbb{U}_L, \qquad \rho_L^- : \mathbb{A} \to \mathbb{A}_L$$

and *prolongation* operators

$$\pi_L^+ : \mathbb{U}_L \to \mathbb{U}, \qquad \pi_L^- : \mathbb{A}_L \to \mathbb{A},$$

which extend respectively to

$$R_L : \mathbb{U} \times \mathbb{A} \times \mathbb{B} \to \mathbb{U}_L \times \mathbb{A}_L \times \mathbb{B}, \quad R_L(u, \psi, \omega) := (\rho_L^+ u, \rho_L^- \psi, \omega) \tag{6.46}$$

and

$$P_L : \mathbb{U}_L \times \mathbb{A}_L \times \mathbb{B} \to \mathbb{U} \times \mathbb{A} \times \mathbb{B}, \quad P_L(u_L, \psi_L, \omega) := (\pi_L^+ u_L, \pi_L^- \psi_L, \omega). \tag{6.47}$$

All the operators above can in principle be defined freely, as long as they are linear and bounded. The specific choices made in this context, based on piecewise polynomial interpolation, will be described below.

The discretization of the space $\mathbb{U}$ concerns the interval $[0,1]$. Consider then an *outer* mesh

$$\Omega_L^+ := \{t_i^+ = ih \ : \ i = 0, 1, \dots, L, \ h = 1/L\} \subset [0,1], \tag{6.48}$$

and *inner* meshes

$$\Omega_{L,i}^+ := \{t_{i,j}^+ := t_{i-1}^+ + c_j h \ : \ j = 1, \dots, m\} \subset [t_{i-1}^+, t_i^+], \quad i = 1, \dots, L, \tag{6.49}$$

where $0 < c_1 < \cdots < c_m < 1$ are given abscissae for m a positive integer. Correspondingly, the discretized space is defined as

$$\mathbb{U}_L := \mathbb{R}^{(1+Lm)\times d}, \tag{6.50}$$

whose elements u_L are indexed as

$$u_L := (u_{1,0}, u_{1,1}, \dots, u_{1,m}, \dots, u_{L,1} \dots, u_{L,m})^T$$

with components in $\mathbb{R}^d$. Moreover, the restriction operator reads

$$\rho_L^+ u := (u(0), u(t_{1,1}^+), \dots, u(t_{1,m}^+), \dots, u(t_{L,1}^+) \dots, u(t_{L,m}^+))^T \in \mathbb{U}_L \tag{6.51}$$

for any $u \in \mathbb{U}$ and, for $u_L \in \mathbb{U}_L$, its prolongation $\pi_L^+ u_L \in \mathbb{U}$ is the unique element of the space

$$\Pi_{L,m}^+ := \{p \in C([0,1], \mathbb{R}^d) \ : \ p|_{[t_{i-1}^+, t_i^+]} \in \Pi_m, \ i = 1, \dots, L\} \tag{6.52}$$

such that

$$\pi_L^+ u_L(0) = u_{1,0}, \quad \pi_L^+ u_L(t_{i,j}^+) = u_{i,j}, \quad j = 1, \dots, m, \ i = 1, \dots, L. \tag{6.53}$$

The piecewise polynomial $p \in \Pi^+_{L,m}$ can be represented through its pieces as

$$p|_{[t^+_{i-1},t^+_i]}(t) = \sum_{j=0}^{m} \ell_{m,i,j}(t)p(t^+_{i,j}), \quad t \in [0,1], \tag{6.54}$$

where

$$t^+_{i,0} := t^+_{i-1}, \quad i = 1,\dots,L, \tag{6.55}$$

and $\{\ell_{m,i,0}, \ell_{m,i,1}, \dots, \ell_{m,i,m}\}$ is the Lagrange basis relevant to the nodes $\{t^+_{i,0}\} \cup \Omega^+_{L,i}$. Observe that the latter is invariant with respect to i as long as the abscissae c_j, $j = 1,\dots,m$, defining the inner meshes (6.49), are fixed. Indeed, for $i = 1,\dots,L$ and $j = 1,\dots,m$,

$$\ell_{m,i,j} = \ell_{m,j}\left(\frac{t - t^+_{i-1}}{h}\right), \quad t \in [t^+_{i-1}, t^+_i].$$

where $\{\ell_{m,0}, \ell_{m,1}, \dots, \ell_{m,m}\}$ is the Lagrange basis relevant to the nodes $c_0, \dots, c_m$ in $[0,1]$ with $c_0 = 0$. Thus, the corresponding Lebesgue constant (see Subection 3.1.2)

$$\Lambda_{m,i} = \max_{t\in[t^+_{i-1},t^+_i]} \sum_{j=0}^{m} |\ell_{m,i,j}(t)| \tag{6.56}$$

is independent of the index i, and in the sequel will be simply denoted by Λ_m. The constant

$$\Lambda'_{m,i} := \max_{t\in[t^+_{i-1},t^+_i]} \sum_{j=0}^{m} |\ell'_{m,i,j}(t)| \tag{6.57}$$

is independent of i as well, and in the sequel will be denoted by Λ'_m.

Similarly, the discretization of the space $\mathbb{A}$ concerns the interval $[-1,0]$. The corresponding outer mesh is given by

$$\Omega^-_L := \{t^-_i = ih - 1 \ : \ i = 0,1,\dots,L,\ h = 1/L\} \subset [-1,0], \tag{6.58}$$

while the inner meshes are

$$\Omega^-_{L,i} := \{t^-_{i,j} := t^-_{i-1} + c_j h \ : \ j = 1,\dots,m\} \subset [t^-_{i-1}, t^-_i], \quad i = 1,\dots,L.$$

Correspondingly, the discretized space is defined as

$$\mathbb{A}_L := \mathbb{R}^{(1+Lm)\times d} \tag{6.59}$$

with indexing

$$\psi_L := (\psi_{1,0}, \psi_{1,1}, \dots, \psi_{1,m}, \dots, \psi_{L,1} \dots, \psi_{L,m})^T.$$

The restriction operator reads

$$\rho^-_L\psi := (\psi(-1), \psi(t^-_{1,1}), \dots, \psi(t^-_{1,m}), \dots, \psi(t^-_{L,1}) \dots, \psi(t^-_{L,m}))^T \in \mathbb{A}_L \tag{6.60}$$

for any $\psi \in \mathbb{A}$ and, for $\psi_L \in \mathbb{A}_L$, its prolongation $\pi^-_L\psi_L \in \mathbb{A}$ is the unique element of the space

$$\Pi^-_{L,m} := \{p \in C([-1,0],\mathbb{R}^d) \ : \ p|_{[t^-_{i-1},t^-_i]} \in \Pi_m,\ i = 1,\dots,L\} \tag{6.61}$$

such that

$$\pi_L^- \psi_L(-1) = \psi_{1,0}, \quad \pi_L^- \psi_L|_{[t_{i-1}^-, t_i^-]}(t_{i,j}^-) = \psi_{i,j}, \quad j = 1, \dots, m, \ i = 1, \dots, L. \tag{6.62}$$

Elements in $\Pi_{L,m}^-$ are represented in the same way as those of $\Pi_{L,m}^+$ by suitably adapting both (6.54) and (6.55), so that also Λ_m and Λ'_m in (6.57) are unchanged.

Note that, under the choices above, the actual discretization level is dependent on both L and m. Fixing the polynomial degree m and choosing the number L of mesh intervals as discretization index corresponds to the FEM (see Subsection 3.3.1), in that the interest is towards the behavior as $L \to \infty$. This is also the traditional approach followed in practical implementations, as, e.g., in MatCont for ODEs [4] or in older versions DDE-Biftool for DDEs [2]. Being the SEM of less interest in the context of practical applications, its convergence under this framework will only be briefly commented at the end of this Chapter (Subsection 6.5.3). However, it is worth noting that the newer versions of DDE-Biftool [93] allow to work with polynomials of arbitrarily large degree.

If the operator $\mathcal{F}$ in (6.10) cannot be computed exactly, then a secondary discretization is needed as well. It consists in defining, for a given level of discretization M, an operator $\mathcal{F}_M$ that is meant to be used in place of $\mathcal{F}$. In particular, $\mathcal{F}_M$ is defined through an approximated version G_M of the right-hand side G as

$$\mathcal{F}_M(u, \psi, \omega) := \omega G_M(\mathcal{G}(u, \psi)_{(\cdot)} \circ s_\omega). \tag{6.63}$$

Correspondingly, Φ_M is the operator obtained by replacing $\mathcal{F}$ in Φ in (6.9) with its approximated version, i.e., $\Phi_M : \mathbb{U} \times \mathbb{A} \times \mathbb{B} \to \mathbb{U} \times \mathbb{A} \times \mathbb{B}$ defined by

$$\Phi_M(u, \psi, \omega) := \begin{pmatrix} \omega G_M(\mathcal{G}(u, \psi)_{(\cdot)} \circ s_\omega) \\ \mathcal{G}(u, \psi)_1 \\ \omega - p(\mathcal{G}(u, \psi)|_{[0,1]}) \end{pmatrix}. \tag{6.64}$$

A typical reason why the introduction of G_M is needed is the presence in G of integrals defining distributed delays, which might need indeed the application of suitable quadrature rules. The operator $\mathcal{G}$, on the other hand, does not need to be discretized since it can be evaluated exactly in $\pi_L^+ \mathbb{U}_L \times \pi_L^- \mathbb{A}_L$ according to (6.50) and (6.59), if the quadrature formula has a sufficient degree of precision (recall Subsection 3.1.4). Moreover, it is reasonable to assume that the operator p defining the phase condition in (6.2) can be evaluated exactly in $\mathcal{G}(\pi_L^+ \mathbb{U}_L, \pi_L^- \mathbb{A}_L)|_{[0,1]})$, since this is the case, for instance, of the trivial (2.8) and integral (2.9) phase conditions. In the latter case, this follows from the possibility of applying the piecewise quadrature based on the mesh of the primary discretization, which is indeed the standard approach used in practical applications.

Using the two discretizations together, one can define the discrete version

$$\Phi_{L,M} := R_L \Phi_M P_L : \mathbb{U}_L \times \mathbb{A}_L \times \mathbb{B} \to \mathbb{U}_L \times \mathbb{A}_L \times \mathbb{B} \tag{6.65}$$

of the fixed point operator Φ in (6.9) as

$$\Phi_{L,M}(u_L, \psi_L, \omega) := \begin{pmatrix} \omega\rho_L^+ G_M(\mathcal{G}(\pi_L^+ u_L, \pi_L^- \psi_L)_{(\cdot)} \circ s_\omega) \\ \rho_L^- \mathcal{G}(\pi_L^+ u_L, \pi_L^- \psi_L)_1 \\ \omega - p(\mathcal{G}(\pi_L^+ u_L, \pi_L^- \psi_L)|_{[0,1]}) \end{pmatrix}.$$

The results in Section 6.5 will involve fixed points $(u^*_{L,M}, \psi^*_{L,M}, \omega^*_{L,M})$ of $\Phi_{L,M}$, which can be found by standard solvers for nonlinear systems of algebraic equations. $P_L(u^*_{L,M}, \psi^*_{L,M}, \omega^*_{L,M})$ will be the sought approximation of a fixed point (u^*, ψ^*, ω^*) of Φ in (6.9) and, correspondingly, the solution $v^* = \mathcal{G}(u^*, \psi^*)$ of (6.2) will be approximated by $v^*_{L,M} := \mathcal{G}(\pi_L^+ u^*_{L,M}, \pi_L^- \psi^*_{L,M})$.

In the rest of the Section, the validity of the numerical assumptions in [79] will be proved under specific choices of the discretization scheme and regularity properties of the discretized right-hand side. As done in Section 6.3, for ease of reference throughout the text, all the hypotheses that will be used are collected below.

(N1) The primary discretization of the space $\mathbb{U}$ is based on (6.48)–(6.52).

(N2) The primary discretization of the space $\mathbb{A}$ is based on (6.58)–(6.61).

(N3) For all positive integers M, G_M is Fréchet-differentiable at every $\mathtt{y} \in \mathtt{Y}$.

(N4) For all positive integers M, $G_M \in \mathcal{C}^1(\mathtt{Y}, \mathbb{R}^d)$.

(N5) There exist $r > 0$ and $\kappa \geq 0$ such that

$$\|DG_M(\mathtt{y}) - DG_M(v_t^* \circ s_{\omega^*})\|_{\mathbb{R}^d \leftarrow \mathtt{Y}} \leq \kappa \|\mathtt{y} - v_t^* \circ s_{\omega^*}\|_{\mathtt{Y}}$$

for every $\mathtt{y} \in \overline{B}(v_t^* \circ s_{\omega^*}, r)$, uniformly with respect to $t \in [0,1]$ and for every positive integer M.

(N6)

$$\lim_{M \to \infty} |G_M(v_t^* \circ s_{\omega^*}) - G(v_t^* \circ s_{\omega^*})| = 0$$

holds uniformly with respect to $t \in [0,1]$.

(N7)

$$\lim_{M \to \infty} \|DG_M(v_t^* \circ s_{\omega^*}) - DG(v_t^* \circ s_{\omega^*})\|_{\mathbb{R}^d \leftarrow \mathtt{Y}} = 0$$

holds uniformly with respect to $t \in [0,1]$.

Note that the uniformity with respect to M of r and κ in (N5) is not as restrictive as it might seem. As anticipated, among the main reasons to introduce G_M is the quadrature of distributed delays. Indeed, if one considers right-hand sides G of the form (1.3) for some integration kernel H with locally Lipschitz continuous derivative with respect to the second argument, uniformly with respect to the first argument, then (T5) is satisfied and also (N5) follows from the application of any convergent interpolatory formula. The same argument holds also if G is of the form (1.4), for some g with locally Lipschitz continuous derivative and any integration kernel H.

The first numerical assumption concerns the Fréchet-differentiability of the operator $\mathcal{F}_M$ defined in (6.63), and is the discrete version of Assumption 6.1.

Assumption 6.13 ($A\mathfrak{F}_K\mathfrak{B}_K$, [79, page 535])**.** For every positive integer M, the operators $\mathfrak{F}_M$ and $\mathfrak{B}_M$ are Fréchet-differentiable at any point $(v_0, u_0, \beta_0) \in \mathbb{V} \times \mathbb{U} \times \mathbb{B}$.

As anticipated, the boundary operator for formulation (6.2) does not need a secondary discretization, thus $\mathcal{B}_M = \mathcal{B}$ for all M, and thus the validity of the assumption is a consequence of the following theorem.

Proposition 6.14. *Under (T1), (T2) and (N3), there exists $r \in (0, \omega^*)$ such that $\mathcal{F}_M$ is Fréchet-differentiable at every $(\hat{v}, \hat{u}, \hat{\omega}) \in \overline{B}((v^*, u^*, \omega^*), r)$, from the right with respect to ω, and*

$$D\mathcal{F}_M(\hat{v}, \hat{u}, \hat{\omega})(v, u, \omega) = \mathfrak{L}_M(\cdot; \hat{v}, \hat{\omega})[v_{(\cdot)} \circ s_{\hat{\omega}}] + \omega \mathfrak{M}_M(\cdot; \hat{v}, \hat{\omega})$$

for $(v, u, \omega) \in \mathbb{V} \times \mathbb{U} \times (0, +\infty)$, where, for $t \in [0, 1]$,

$$\mathfrak{L}_M(t; v, \omega) := \omega DG_M(v_t \circ s_\omega) \tag{6.66}$$

and

$$\mathfrak{M}_M(t; v, \omega) := G_M(v_t \circ s_\omega) - \mathfrak{L}_M(t; \hat{v}, \hat{\omega})[v'_t \circ s_\omega] \cdot \frac{s_\omega}{\omega}. \tag{6.67}$$

Proof. The result follows from applying Proposition 6.2 to $\mathcal{F}_M$. □

Proposition 6.14 guarantees the Fréchet-differentiability of the fixed point operator Φ_M in (6.64), as stated next.

Corollary 6.15. *Under (T1), (T2) and (N3), there exists $r \in (0, \omega^*)$ such that Φ_M in (6.64) is Fréchet-differentiable at every $(\hat{u}, \hat{\psi}, \hat{\omega}) \in \overline{B}((u^*, \psi^*, \omega^*), r)$, from the right with respect to ω, and*

$$\begin{aligned} &D\Phi_M(\hat{u}, \hat{\psi}, \hat{\omega})(u, \psi, \omega) \\ &\quad = \begin{pmatrix} \mathfrak{L}_M(\cdot; \mathcal{G}(\hat{u}, \hat{\psi}), \hat{\omega})[\mathcal{G}(u, \psi)_{(\cdot)} \circ s_{\hat{\omega}}] + \omega \mathfrak{M}_M(\cdot; \mathcal{G}(\hat{u}, \hat{\psi}), \hat{\omega}) \\ \mathcal{G}(u, \psi)_1 \\ \omega - p(\mathcal{G}(u, \psi)|_{[0,1]}) \end{pmatrix} \end{aligned}$$

for $(u, \psi, \omega) \in \mathbb{U} \times \mathbb{A} \times (0, +\infty)$, $\mathfrak{L}_M$ in (6.66) and $\mathfrak{M}_M$ in (6.67).

Proof. The result follows from applying Corollary 6.5 to the map Φ_M. □

Note that the operator $\Phi_{L,M}$ in (6.65), used to compute the discrete approximations, is not defined on the same space as Φ, and therefore cannot be used directly to analyze the convergence of the method. Rather, the operator $P_L R_L \Phi_M$ will play that role. This and the relation between all the relevant fixed points are arguments of Section 6.5. Meanwhile, in order to ease the notation, it is useful to define $\Psi, \Psi_{L,M} : \mathbb{U} \times \mathbb{A} \times \mathbb{B} \to \mathbb{U} \times \mathbb{A} \times \mathbb{B}$ as

$$\Psi := I_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}} - \Phi, \qquad \Psi_{L,M} := I_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}} - P_L R_L \Phi_M. \tag{6.68}$$

Note that both Ψ and $\Psi_{L,M}$ are Fréchet-differentiable, the first thanks to Corollary 6.5 and the second thanks to Corollary 6.15 and the linearity of both P_L and R_L.

The other numerical assumptions concern the stability of the chosen discretization. The first one is somehow the discrete version of Assumption 6.6.

Assumption 6.16 (CS1, [79, page 537])**.** There exists $r_1 > 0$ and, for any positive integers M and L, $L_{L,M} \geq 0$ such that

$$\begin{aligned} &\|D\Psi_{L,M}(u,\alpha,\beta) - D\Psi_{L,M}(u^*,\alpha^*,\beta^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}\leftarrow\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \\ &= \|P_L R_L(D\Phi_M(u,\alpha,\beta) - D\Phi_M(u^*,\alpha^*,\beta^*))\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}\leftarrow\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \\ &\leq L_{L,M}\|(u,\alpha,\beta) - (u^*,\alpha^*,\beta^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \end{aligned}$$

for all $(u,\alpha,\beta) \in \mathcal{B}((u^*,\alpha^*,\beta^*),r_1))$.

Correspondingly, the second one is somehow the discrete version of Assumption 6.9.

Assumption 6.17 (CS2, [79, page 537])**.** There exists a positive integer $\overline{N}$ such that, for all $L, M \geq \overline{N}$, $D\Psi_{L,M}(u^*,\alpha^*,\beta^*)$ is invertible and

$$\begin{aligned} \lim_{L,M\to\infty} \frac{1}{r_2(L,M)}&\|[D\Psi_{L,M}(u^*,\psi^*,\omega^*)]^{-1}\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}\leftarrow\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \\ &\cdot\|\Psi_{L,M}(u^*,\psi^*,\omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}} = 0, \end{aligned}$$

where

$$r_2(L,M) := \min\left\{r_1, \frac{1}{2L_{L,M}\|[D\Psi_{L,M}(u^*,\psi^*,\omega^*)]^{-1}\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}\leftarrow\mathbb{U}\times\mathbb{A}\times\mathbb{B}}}\right\}$$

with r_1 and $L_{L,M}$ given as in Assumption 6.16.

For the sequel, it is convenient to introduce the abbreviations

$$\mathfrak{L}_M^* := \mathfrak{L}_M(\cdot;v^*,\omega^*), \qquad \mathfrak{M}_M^* := \mathfrak{M}_M(\cdot;v^*,\omega^*) \tag{6.69}$$

in accordance with (6.32).

The validity of Assumption 6.16 is address by the following proposition.

Proposition 6.18. *Under (T1), (T2), (N1), (N2), (N3) and (N5), there exist $r_1 \in (0,\omega^*)$ and $\kappa \geq 0$ such that*

$$\begin{aligned} \|D\Psi_{L,M}(u,\psi,\omega) - D\Psi_{L,M}(u^*,\psi^*,\omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}\leftarrow\mathbb{U}_2\times\mathbb{A}\times(0,+\infty)} \\ \leq \kappa\|(u,\psi,\omega) - (u^*,\psi^*,\omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \end{aligned}$$

for all $(u,\psi,\omega) \in \overline{B}((u^*,\psi^*,\omega^*),r_1)$ and for all positive integers L and M.

Proof. Applying Proposition 6.7 to the map Φ_M, it follows that there exist $r_1 \in (0,\omega^*)$ and $\kappa_1 \geq 0$ such that

$$\begin{aligned} \|D\Phi_M(u,\psi,\omega) - D\Phi_M(u^*,\psi^*,\omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}\leftarrow\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \\ \leq \kappa_1\|(u,\psi,\omega) - (u^*,\psi^*,\omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \end{aligned}$$

for all $(u,\psi,\omega) \in \overline{B}((u^*,\psi^*,\omega^*),r_1)$. In particular, correspondingly to what was obtained in Proposition 6.7, one can choose $r_1 = \min\{1, r/2\}$ for r in (N5). By Corollary 7.3, the thesis follows directly from the second of (6.68) by choosing $\kappa = 1 + \kappa_1 \cdot \max\{\Lambda_m + \Lambda'_m, 1\}$. □

Note that κ is independent of L thanks to (7.6) and independent of M thanks to (N5). However, it depends on m.

The proof of the validity of Assumption 6.17 consists of two parts. The main one concerns the invertibility of $D\Psi_{L,M}(u^*, \psi^*, \omega^*)$ and will be, in turn, divided into several lemmas. Indeed, it is not as straightforward as it might appear at a first glance. Although it is known, by Proposition 6.11, that $D\Psi(u^*, \psi^*, \omega^*)$ is invertible, a trivial application of the Banach perturbation lemma (Theorem 3.15) is not feasible. That is, it is not true that

$$\lim_{L,M\to\infty} \|D\Psi_{L,M}(u^*, \psi^*, \omega^*) - D\Psi(u^*, \psi^*, \omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}\leftarrow\mathbb{U}\times\mathbb{A}\times(0,+\infty)} = 0.$$

Indeed, this in turn would require

$$\lim_{L\to\infty} \|(I_{\mathbb{A}} - \pi_L^- \rho_L^-)\mathcal{G}(u, \psi)_1\|_{\mathbb{A}} = 0$$

through (6.46), (6.47), (6.64) and (6.68). The latter cannot hold for all $(u, \psi) \in \mathbb{U} \times \mathbb{A}$ given the choice $\mathbb{A}$ in (T2) (recall the definition of its norm in the second of (1.13)).

However, the Banach perturbation lemma represents a sufficient but not necessary criterion. In the following, the invertibility of $D\Psi_{L,M}(u^*, \psi^*, \omega^*)$ will be proved by following the lines of the proof of Proposition 6.11 and, indeed, other instances of the Banach perturbation lemma will be applied at several points.

The first step of the proof consists in showing that the initial value problem for

$$y'(t) = [\pi_L^+ \rho_L^+ \mathfrak{L}_M^*[y_{(\cdot)} \circ s_{\omega^*}]](t) \tag{6.70}$$

is well-posed, and thus an associated evolution operator $T_{L,M}^*(t, s) : Y \to Y$ can be defined for $t, s \in [0, 1]$ and $t \geq s$. In the sequel it is also convenient to use the abbreviations

$$\begin{aligned}
&\mathcal{G}^+ u := \mathcal{G}(u, 0), && \mathcal{G}^- \psi := \mathcal{G}(0, \psi),\\
&\mathcal{K}^{*,+} u := \mathfrak{L}^*[(\mathcal{G}^+ u)_{(\cdot)} \circ s_{\omega^*}], && \mathcal{K}^{*,-} \psi := \mathfrak{L}^*[(\mathcal{G}^- \psi)_{(\cdot)} \circ s_{\omega^*}],\\
&\mathcal{K}_M^{*,+} u := \mathfrak{L}_M^*[(\mathcal{G}^+ u)_{(\cdot)} \circ s_{\omega^*}], && \mathcal{K}_M^{*,-} \psi := \mathfrak{L}_M^*[(\mathcal{G}^- \psi)_{(\cdot)} \circ s_{\omega^*}].
\end{aligned} \tag{6.71}$$

Lemma 6.19. *Under (T1), (T2), (T4), (N1), (N2), (N4) and (N7), there exist positive integers $\overline{L}$ and $\overline{M}$ such that, for every $L \geq \overline{L}$ and $M \geq \overline{M}$, the initial value problem*

$$\begin{cases} y'(t) = [\pi_L^+ \rho_L^+ \mathfrak{L}_M^*[y_{(\cdot)} \circ s_{\omega^*}]](t), \quad t \in [0, 1], \\ y_0 = \psi \end{cases} \tag{6.72}$$

for $\psi \in Y$ has a unique solution $y_{L,M}$.

Proof. Set $u(t) := y'(t)$ for $t \in [0, 1]$ and use $y = \mathcal{G}(u, \psi)$ according to (6.8). By virtue of (6.71), (6.72) becomes

$$u = \pi_L^+ \rho_L^+ \mathcal{K}_M^{*,+} u + \pi_L^+ \rho_L^+ \mathcal{K}_M^{*,-} \psi.$$

Well-posedness is thus equivalent to the invertibility of $I_{\mathbb{U}} - \pi_L^+ \rho_L^+ \mathcal{K}_M^{*,+} : \mathbb{U} \to \mathbb{U}$. Since the invertibility of $I_{\mathbb{U}} - \mathcal{K}^{*,+} : \mathbb{U} \to \mathbb{U}$ is guaranteed, under (T4), by the well-posedness of the initial value problem for (6.31), thanks to (7.7) in Lemma 7.6 the thesis follows by applying the Banach perturbation lemma. □

Lemma 6.20. *Under (T1), (T2), (T4), (N1), (N2), (N4) and (N7),*

$$\lim_{L,M\to\infty} \|T^*_{L,M}(t,s) - T^*(t,s)\|_{Y\leftarrow Y} = 0 \tag{6.73}$$

uniformly with respect to $t,s \in [0,1]$, $t \geq s$. If, in addition, $1 \in \sigma(T^*(1,0))$ is simple with eigenfunction φ normalized as $\|\varphi\|_Y = 1$ and $r > 0$ is such that 1 is the only eigenvalue of $T^*(1,0)$ in $B(1,r) \subset \mathbb{C}$, then there exist positive integers $\overline{L}$ and $\overline{M}$ such that, for every $L \geq \overline{L}$ and $M \geq \overline{M}$, $T^*_{L,M}(1,0)$ has only a simple eigenvalue $\mu_{L,M}$ in $B(1,r)$ and, moreover,

$$\lim_{L,M\to\infty} |\mu_{L,M} - 1| = 0, \qquad \lim_{L,M\to\infty} \|\varphi_{L,M} - \varphi\|_Y = 0, \tag{6.74}$$

where $\varphi_{L,M}$ is the eigenfunction associated to $\mu_{L,M}$ normalized as $\|\varphi_{L,M}\|_Y = 1$.

Proof. The proof will be given for $s = 0$, the extension to $s \in (0,1)$ being straightforward.

Let $\mathcal{G}(u,\psi)$ be the solution of (6.31) exiting from a given $\psi \in Y$, where u satisfies $u = \mathfrak{L}^*[\mathcal{G}(u,\psi)_{(\cdot)} \circ s_{\omega^*}]$. Correspondingly, thanks to Lemma 6.19, let $\mathcal{G}(u_{L,M},\psi)$ be *the* solution of (6.70) exiting from the same ψ, where $u_{L,M}$ satisfies $u_{L,M} = \pi^+_L \rho^+_L \mathfrak{L}^*_M[\mathcal{G}(u_{L,M},\psi)_{(\cdot)} \circ s_{\omega^*}]$. The relevant evolution operators are defined, for $t \in [0,1]$, respectively by

$$T^*(t,0)\psi = \mathcal{G}(u,\psi)_t \quad \text{and} \quad T^*_{L,M}(t,0)\psi = \mathcal{G}(u_{L,M},\psi)_t.$$

The linearity of $\mathcal{G}$ in (6.8) leads to

$$T^*_{L,M}(t,0)\psi - T^*(t,0)\psi = \mathcal{G}(u_{L,M} - u, 0)_t = \mathcal{G}^+(u_{L,M} - u)_t.$$

Therefore, (6.73) is equivalent to showing that

$$\lim_{L,M\to\infty} \|e_{L,M}\|_{\mathbb{U}} = 0 \tag{6.75}$$

for $e_{L,M} := u_{L,M} - u$. Using the abbreviations (6.71), one can write

$$u_{L,M} = \pi^+_L \rho^+_L \mathcal{K}^{*,+}_M u_{L,M} + \pi^+_L \rho^+_L \mathcal{K}^{*,-}_M \psi, \quad u = \mathcal{K}^{*,+} u + \mathcal{K}^{*,-}\psi,$$

therefore

$$e_{L,M} = \pi^+_L \rho^+_L \mathcal{K}^{*,+}_M e_{L,M} + r^+_{L,M} + r^-_{L,M},$$

where

$$r^+_{L,M} := (\pi^+_L \rho^+_L \mathcal{K}^{*,+}_M - \mathcal{K}^{*,+})u, \quad r^-_{L,M} := (\pi^+_L \rho^+_L \mathcal{K}^{*,-}_M - \mathcal{K}^{*,-})\psi.$$

Through the Banach perturbation lemma, upon showing in the proof of Lemma 6.19 that $I_{\mathbb{U}} - \pi^+_L \rho^+_L \mathcal{K}^{*,+}_M$ is invertible, one can also show that

$$\|(I_{\mathbb{U}} - \pi^+_L \rho^+_L \mathcal{K}^{*,+}_M)^{-1}\|_{\mathbb{U}\leftarrow\mathbb{U}} \leq 2\|(I_{\mathbb{U}} - \mathcal{K}^{*,+})^{-1}\|_{\mathbb{U}\leftarrow\mathbb{U}}$$

holds for L and M sufficiently large. Now (6.75) follows since both

$$\|r^+_{L,M}\|_{\mathbb{U}} \leq \|\pi^+_L \rho^+_L \mathcal{K}^{*,+}_M - \mathcal{K}^{*,+}\|_{\mathbb{U}\leftarrow\mathbb{U}} \|u\|_{\mathbb{U}}$$

and

$$\|r^-_{L,M}\|_{\mathbb{U}} \leq \|\pi^+_L \rho^+_L \mathcal{K}^{*,-}_M - \mathcal{K}^{*,-}\|_{\mathbb{U}\leftarrow\mathbb{A}} \|\psi\|_{\mathbb{A}}$$

vanish by Lemma 7.6.

The second part follows by Lemma 3.16 for $\nu = l = 1$. □

The last step needed to prove the invertibility of $D\Psi_{L,M}(u^*, \psi^*, \omega^*)$ is given by the following proposition.

Proposition 6.21. *Under (T1), (T2), (T4), (N1), (N2), (N4), (N6) and (N7), there exist positive integers $\overline{L}$ and $\overline{M}$ such that, for every $L \geq \overline{L}$ and $M \geq \overline{M}$, $D\Psi_{L,M}(u^*, \psi^*, \omega^*)$ is invertible, i.e., for all $(u_0, \psi_0, \omega_0) \in \mathbb{U} \times \mathbb{A} \times \mathbb{B}$ there exists a unique $(u_{L,M}, \psi_{L,M}, \omega_{L,M}) \in \mathbb{U} \times \mathbb{A} \times \mathbb{B}$ such that*

$$\begin{cases} u_{L,M} = \pi_L^+ \rho_L^+ \mathfrak{L}_M^* [\mathcal{G}(u_{L,M}, \psi_{L,M})_{(\cdot)} \circ s_{\omega^*}] + \omega_{L,M} \pi_L^+ \rho_L^+ \mathfrak{M}_M^* + u_0 \\ \psi_{L,M} = \pi_L^- \rho_L^- \mathcal{G}(u_{L,M}, \psi_{L,M})_1 + \psi_0 \\ p(\mathcal{G}(u_{L,M}, \psi_{L,M})|_{[0,1]}) = \omega_0. \end{cases} \tag{6.76}$$

Proof. As anticipated, this proofs follows the lines of the one of Proposition 6.11. Consider (6.76) as an initial value problem for $v_{L,M} := \mathcal{G}(u_{L,M}, \psi_{L,M})$, i.e.,

$$\begin{cases} v'_{L,M}(t) = (\pi_L^+ \rho_L^+ \mathfrak{L}_M^* [v_{L,M,\cdot} \circ s_{\omega^*}])(t) + \omega_{L,M} \pi_L^+ \rho_L^+ \mathfrak{M}_M^*(t) + u_0(t) \\ v_{L,M,0} = \psi_{L,M} \end{cases} \tag{6.77}$$

for $t \in [0,1]$, where $v_{L,M,t}$ is a shortcut for $(v_{L,M})_t$, defined in (6.7). The variation of constants formula (2.3) gives, for $t \in [0,1]$,

$$v_{L,M,t} = T_{L,M}^*(t,0)\psi_{L,M} + \int_0^t [T_{L,M}^*(t,s)X_0][\omega_{L,M}\pi_L^+\rho_L^+\mathfrak{M}_M^*(s) + u_0(s)]\,\mathrm{d}s,$$

which, together with the first boundary condition in (6.76), leads to

$$\begin{aligned} \psi_{L,M} = {} & \pi_L^- \rho_L^- T_{L,M}^*(1,0)\psi_{L,M} \\ & + \pi_L^- \rho_L^- \int_0^1 [T_{L,M}^*(1,s)X_0][\omega_{L,M}\pi_L^+\rho_L^+\mathfrak{M}_M^*(s) + u_0(s)]\,\mathrm{d}s + \psi_0. \end{aligned}$$

For $L \geq \overline{L}$ and $M \geq \overline{M}$, with $\overline{L}$ and $\overline{M}$ given by Lemma 6.20, let $\mu_{L,M}$ be the relevant simple multiplier of $T_{L,M}^*(1,0)$. Then the last equation reads

$$\begin{aligned} \mu_{L,M}\psi_{L,M} = {} & T_{L,M}^*(1,0)\psi_{L,M} \\ & + \pi_L^- \rho_L^- \int_0^1 [T_{L,M}^*(1,s)X_0][\omega_{L,M}\pi_L^+\rho_L^+\mathfrak{M}_M^*(s) + u_0(s)]\,\mathrm{d}s \\ & + \psi_0 + \nu_{L,M} \end{aligned} \tag{6.78}$$

for

$$\nu_{L,M} := (\pi_L^- \rho_L^- - I_{\mathbb{A}}) T_{L,M}^*(1,0)\psi_{L,M} + (\mu_{L,M} - 1)\psi_{L,M}. \tag{6.79}$$

Note that, under (T2), $T_{L,M}^*(1,0)\psi_{L,M} = \mathcal{G}(u_{L,M}, \psi_{L,M})_1 \in \mathbb{A}$.

The state space can be decomposed as

$$Y = R_{L,M} \oplus K_{L,M} \tag{6.80}$$

for $R_{L,M}$ and $K_{L,M}$ the range and the kernel of $\mu_{L,M} I_Y - T_{L,M}^*(1,0)$, respectively. Since $\mu_{L,M}$ is simple, $K_{L,M} = \mathrm{span}\{\varphi_{L,M}\}$ for some $\varphi_{L,M}$ an eigenfunction of the multiplier $\mu_{L,M}$. Recalling Remark 6.12, it is reasonable to assume $p(v(\cdot;\varphi_{L,M})|_{[0,1]}) \neq 0$ for $v(\cdot;\varphi_{L,M})$ the solution of (6.77) exiting from $\varphi_{L,M}$, thanks to the linearity of p and to the second of (6.74) in Lemma 6.20 .

Consider the elements of Y

$$\begin{aligned}\xi^*_{L,M,1} &:= \pi^-_L \rho^-_L \int_0^1 [T^*_{L,M}(1,s)X_0]\pi^+_L \rho^+_L \mathfrak{M}^*_M(s)\,\mathrm{d}s,\\ \xi^*_{L,M,2} &:= \pi^-_L \rho^-_L \int_0^1 [T^*_{L,M}(1,s)X_0]u_0(s)\,\mathrm{d}s + \psi_0.\end{aligned} \tag{6.81}$$

Then, (6.78) becomes

$$[\mu_{L,M} I_Y - T^*_{L,M}(1,0)]\psi_{L,M} = \omega_{L,M}\xi^*_{L,M,1} + \xi^*_{L,M,2} + \nu_{L,M}. \tag{6.82}$$

From (6.80), there exist unique $r_{L,M,1}, r_{L,M,2}, s_{L,M} \in R_{L,M}$ and $k_{L,M,1}, k_{L,M,2}$, $h_{L,M} \in \mathbb{R}$ such that

$$\begin{aligned}\xi^*_{L,M,1} &= r_{L,M,1} + k_{L,M,1}\varphi_{L,M},\\ \xi^*_{L,M,2} &= r_{L,M,2} + k_{L,M,2}\varphi_{L,M},\\ \nu_{L,M} &= s_{L,M} + h_{L,M}\varphi_{L,M}.\end{aligned} \tag{6.83}$$

Thus, from (6.82) it must be $\omega_{L,M}\xi^*_{L,M,1} + \xi^*_{L,M,2} + \nu_{L,M} \in R_{L,M}$, which implies $\omega_{L,M}k_{L,M,1} + k_{L,M,2} + h_{L,M} = 0$. By (7.14) in Proposition 7.8, $k_{L,M,1} \to k_1$ for k_1 in the proof of Proposition 6.11. As the latter is proved to be different from 0 in Subsection 6.3.1, the same holds for $k_{L,M,1}$ for L and M sufficiently large. Therefore, $k_{L,M,1} \neq 0$ can be assumed, leading to

$$\omega_{L,M} = -\frac{k_{L,M,2} + h_{L,M}}{k_{L,M,1}} \tag{6.84}$$

being the only possible solution. Eventually, let $\eta_{L,M}$ be such that

$$[\mu_{L,M} I_Y - T^*_{L,M}(1,0)]\eta_{L,M} = \omega_{L,M}\xi^*_{L,M,1} + \xi^*_{L,M,2} + \nu_{L,M}.$$

Then, every $\psi_{L,M}$ satisfying (6.82) can be written as $\eta_{L,M} + \lambda_{L,M}\varphi_{L,M}$ for some $\lambda_{L,M} \in \mathbb{R}$. The value of the latter can be fixed uniquely by imposing the phase condition, i.e., $p(v(\cdot;\eta_{L,M})|_{[0,1]}) + \lambda_{L,M}p(v(\cdot;\varphi_{L,M})|_{[0,1]}) = \omega_0$. □

The last step to complete the proof of the validity of Assumption 6.17 is represented by Proposition 6.23 below, and consists in showing that the inverse of $D\Psi_{L,M}(u^*, \psi^*, \omega^*)$ is bounded uniformly with respect to L and M.

Lemma 6.22. *Under (T1), (T2), (T4), (N1), (N2), (N4), (N6) and (N7), the inverse of $D\Psi_{L,M}(u^*, \psi^*, \omega^*)$ is uniformly bounded with respect to both L and M.*

Proof. Proposition 6.21 guarantees that, given $(u_0, \psi_0, \omega_0) \in \mathbb{U} \times \mathbb{A} \times \mathbb{B}$, there exists a unique $(u_{L,M}, \psi_{L,M}, \omega_{L,M}) \in \mathbb{U} \times \mathbb{A} \times \mathbb{B}$ satisfying

$$D\Psi_{L,M}(u^*, \psi^*, \omega^*)(u_{L,M}, \psi_{L,M}, \omega_{L,M}) = (u_0, \psi_0, \omega_0).$$

The thesis is thus equivalent to the fact that $\|(u_{L,M}, \psi_{L,M}, \omega_{L,M})\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}}$ is bounded uniformly with respect to both L and M. In order to prove that, it is useful to relate $(u_{L,M}, \psi_{L,M}, \omega_{L,M})$ to the solution of the collocation of an equivalent reformulation of (the secondary discretization of) (6.33), defined as follows, according to the primary discretization under (N1) and (N2). Such a reformulation comes from the need to give a proper sense to the collocation problem since, in general, u is not continuous therein, while the

range of $\pi_L^+\rho_L^+$ contains only continuous functions. The terms of (6.33) can be rearranged as

$$\begin{cases} z = \mathfrak{L}^*[\mathcal{G}(z,\gamma)_{(\cdot)} \circ s_{\omega^*}] + \omega\mathfrak{M}^* + \mathfrak{L}^*[\mathcal{G}(u_0,\psi_0)_{(\cdot)} \circ s_{\omega^*}] \\ \gamma = \mathcal{G}(z,\gamma)_1 + \mathcal{G}(u_0,\psi_0)_1 \\ p(\mathcal{G}(z,\gamma)|_{[0,1]}) = \omega_0 - p(\mathcal{G}(u_0,\psi_0)|_{[0,1]}) \end{cases} \tag{6.85}$$

obtained from (6.33) by setting $z := u - u_0$ and $\gamma := \psi - \psi_0$. Note that z is continuous as it follows from the first equation in (6.85) under (T4). Correspondingly, (6.76) can be rewritten as

$$\begin{cases} z_{L,M} = \pi_L^+\rho_L^+\mathfrak{L}_M^*[\mathcal{G}(z_{L,M},\gamma_{L,M})_{(\cdot)} \circ s_{\omega^*}] + \omega_{L,M}\pi_L^+\rho_L^+\mathfrak{M}_M^* \\ \qquad\qquad + \pi_L^+\rho_L^+\mathfrak{L}_M^*[\mathcal{G}(u_0,\psi_0)_{(\cdot)} \circ s_{\omega^*}] \\ \gamma_{L,M} = \pi_L^-\rho_L^-\mathcal{G}(z_{L,M},\gamma_{L,M})_1 + \pi_L^-\rho_L^-\mathcal{G}(u_0,\psi_0)_1 \\ p(\mathcal{G}(z_{L,M},\gamma_{L,M})|_{[0,1]}) = \omega_0 - p(\mathcal{G}(u_0,\psi_0)|_{[0,1]}) \end{cases} \tag{6.86}$$

for $z_{L,M} := u_{L,M} - u_0$ and $\gamma_{L,M} := \psi_{L,M} - \psi_0$. It follows

$$u_{L,M} = e_{L,M}^+ + u, \qquad \psi_{L,M} = e_{L,M}^- + \psi, \tag{6.87}$$

where $e_{L,M}^+ := z_{L,M} - z$ and $e_{L,M}^- := \gamma_{L,M} - \gamma$ are the collocation errors of the components in $\mathbb{U}$ and $\mathbb{A}$, respectively, given that $(z_{L,M}, \gamma_{L,M}, \omega_{L,M})$ is the collocation solution of (the secondary discretization of) (6.85) according to (N1) and (N2). Subtracting (6.85) from (6.86) leads to

$$\begin{cases} e_{L,M}^+ = \pi_L^+\rho_L^+\mathfrak{L}_M^*[\mathcal{G}(e_{L,M}^+,e_{L,M}^-)_{(\cdot)} \circ s_{\omega^*}] + \varepsilon_{\omega,L,M} + \varepsilon_{L,M}^+ \\ e_{L,M}^- = \pi_L^-\rho_L^-\mathcal{G}(e_{L,M}^+,e_{L,M}^-)_1 + \varepsilon_{L,M}^- \\ p(\mathcal{G}(e_{L,M}^+,e_{L,M}^-)|_{[0,1]}) = 0 \end{cases} \tag{6.88}$$

for

$$\begin{aligned} \varepsilon_{\omega,L,M} &:= \omega_{L,M}\pi_L^+\rho_L^+\mathfrak{M}_M^* - \omega\mathfrak{M}^*, \\ \varepsilon_{L,M}^+ &:= \pi_L^+\rho_L^+\mathfrak{L}_M^*[\mathcal{G}(u_0,\psi_0)_{(\cdot)} \circ s_{\omega^*}] - \mathfrak{L}^*[\mathcal{G}(u_0,\psi_0)_{(\cdot)} \circ s_{\omega^*}], \\ \varepsilon_{L,M}^- &:= (\pi_L^-\rho_L^- - I_{\mathbb{A}})\mathcal{G}(u_0,\psi_0)_1. \end{aligned} \tag{6.89}$$

By (6.71), the first two equations of (6.88) read

$$\begin{cases} e_{L,M}^+ = \pi_L^+\rho_L^+\mathcal{K}_M^{*,+}e_{L,M}^+ + \pi_L^+\rho_L^+\mathcal{K}_M^{*,-}e_{L,M}^- + \varepsilon_{\omega,L,M} + \varepsilon_{L,M}^+ \\ e_{L,M}^- = \pi_L^-\rho_L^-\mathcal{G}_1^+e_{L,M}^+ + \pi_L^-\rho_L^-\mathcal{G}_1^-e_{L,M}^- + \varepsilon_{L,M}^-, \end{cases} \tag{6.90}$$

where $\mathcal{G}(e_{L,M}^+,e_{L,M}^-)_1 = \mathcal{G}_1^+e_{L,M}^+ + \mathcal{G}_1^-e_{L,M}^-$ for

$$(\mathcal{G}_1^+e_{L,M}^+)(t) := \int_0^{1+t} e_{L,M}^+(s)\,\mathrm{d}s, \qquad (\mathcal{G}_1^-e_{L,M}^-)(t) = e_{L,M}^-(0) \tag{6.91}$$

and $t \in [-1,0]$ according to the definition of $\mathcal{G}$ in (6.8). Allowing for a block-wise definition of operators in $\mathbb{U} \times \mathbb{A}$, which should be self-explaining in the following, (6.90) becomes

$$\begin{pmatrix} e_{L,M}^+ \\ e_{L,M}^- \end{pmatrix} = \begin{pmatrix} \pi_L^+\rho_L^+\mathcal{K}_M^{*,+} & \pi_L^+\rho_L^+\mathcal{K}_M^{*,-} \\ \pi_L^-\rho_L^-\mathcal{G}_1^+ & \pi_L^-\rho_L^-\mathcal{G}_1^- \end{pmatrix} \begin{pmatrix} e_{L,M}^+ \\ e_{L,M}^- \end{pmatrix} + \begin{pmatrix} \varepsilon_{\omega,L,M} + \varepsilon_{L,M}^+ \\ \varepsilon_{L,M}^- \end{pmatrix}.$$

In order to get a bound on $\|(e^+_{L,M}, e^-_{L,M})\|_{C([0,1],\mathbb{R}^d)\times\mathbb{A}}$, that is, a bound on the collocation error, it is crucial to observe that $e^+_{L,M}$ is continuous since so is z by construction, while $z_{L,M}$ belongs to $\Pi^+_{L,m}$. Moreover, also $\varepsilon_{\omega,L,M}$ and $\varepsilon^+_{L,M}$ in (6.89) are continuous under (T4) and (N4). In what follows, the notation

$$C^+ := C([0,1],\mathbb{R}^d) \tag{6.92}$$

will be used for brevity.

Note that existence and uniqueness of $(e^+_{L,M}, e^-_{L,M})$ follows already from Propositions 6.21 and 6.11, so that the invertibility of the operator

$$\begin{pmatrix} I_{C^+} & 0 \\ 0 & I_{\mathbb{A}} \end{pmatrix} - \begin{pmatrix} \pi^+_L\rho^+_L\mathcal{K}^{*,+}_M & \pi^+_L\rho^+_L\mathcal{K}^{*,-}_M \\ \pi^-_L\rho^-_L\mathcal{G}^+_1 & \pi^-_L\rho^-_L\mathcal{G}^-_1 \end{pmatrix} : C^+\times\mathbb{A}\to C^+\times\mathbb{A}$$

is already proved. The following step consists in proving that

$$\lim_{L,M\to\infty}\left\| \begin{pmatrix} \pi^+_L\rho^+_L\mathcal{K}^{*,+}_M & \pi^+_L\rho^+_L\mathcal{K}^{*,-}_M \\ \pi^-_L\rho^-_L\mathcal{G}^+_1 & \pi^-_L\rho^-_L\mathcal{G}^-_1 \end{pmatrix} - \begin{pmatrix} \mathcal{K}^{*,+} & \mathcal{K}^{*,-} \\ \mathcal{G}^+_1 & \mathcal{G}^-_1 \end{pmatrix} \right\|_{C^+\times\mathbb{A}\leftarrow C^+\times\mathbb{A}} = 0,$$

meaning to apply then the Banach perturbation lemma to recover the bound

$$\begin{aligned} &\left\| \left[\begin{pmatrix} I_{C^+} & 0 \\ 0 & I_{\mathbb{A}} \end{pmatrix} - \begin{pmatrix} \pi^+_L\rho^+_L\mathcal{K}^{*,+}_M & \pi^+_L\rho^+_L\mathcal{K}^{*,-}_M \\ \pi^-_L\rho^-_L\mathcal{G}^+_1 & \pi^-_L\rho^-_L\mathcal{G}^-_1 \end{pmatrix} \right]^{-1} \right\|_{C^+\times\mathbb{A}\leftarrow C^+\times\mathbb{A}} \\ &\qquad \le 2 \left\| \left[\begin{pmatrix} I_{C^+} & 0 \\ 0 & I_{\mathbb{A}} \end{pmatrix} - \begin{pmatrix} \mathcal{K}^{*,+} & \mathcal{K}^{*,-} \\ \mathcal{G}^+_1 & \mathcal{G}^-_1 \end{pmatrix} \right]^{-1} \right\|_{C^+\times\mathbb{A}\leftarrow C^+\times\mathbb{A}}, \end{aligned} \tag{6.93}$$

for sufficiently large L and M, which is also uniform with respect to both L and M. Indeed, Proposition 6.11 gives the invertibility of the operator

$$\begin{pmatrix} I_{C^+} & 0 \\ 0 & I_{\mathbb{A}} \end{pmatrix} - \begin{pmatrix} \mathcal{K}^{*,+} & \mathcal{K}^{*,-} \\ \mathcal{G}^+_1 & \mathcal{G}^-_1 \end{pmatrix} : C^+\times\mathbb{A}\to C^+\times\mathbb{A}.$$

Note that Lemma 7.6 holds as well if one replaces $\mathbb{U}$ with C^+ since the norm is the same. Therefore, (6.93) holds thanks to Lemma 7.9, and gives

$$\|(e^+_{L,M}, e^-_{L,M})\|_{C^+\times\mathbb{A}} \le \kappa\|(\varepsilon_{\omega,L,M} + \varepsilon^+_{L,M}, \varepsilon^-_{L,M})\|_{C^+\times\mathbb{A}}$$

for some constant κ independent of L and M. By the definition (6.89), one can write

$$\varepsilon^+_{L,M} = \pi^+_L\rho^+_L(\mathfrak{L}^*_M - \mathfrak{L}^*[)\mathcal{G}(u_0,\psi_0)_{(\cdot)}\circ s_{\omega^*}] + (\pi^+_L\rho^+_L - I_{\mathbb{U}})\mathfrak{L}^*[\mathcal{G}(u_0,\psi_0)_{(\cdot)}\circ s_{\omega^*}], \tag{6.94}$$

so that $\varepsilon^+_{L,M}$ vanishes as $L, M\to\infty$ under (T4) and (N7) by (7.1) of Lemma 7.1 and (7.9) of Lemma 7.7. On the other hand, the derivative of $\mathcal{G}(u_0,\psi_0)_1$, namely u_0, is not necessarily continuous, which means that Lemma 7.2 cannot be used to prove that

$$\varepsilon^-_{L,M} := (\pi^-_L\rho^-_L - I_{\mathbb{A}})\mathcal{G}(u_0,\psi_0)_1.$$

vanishes. However, it is anyway bounded uniformly with respect to L and M since u_0 is bounded, and in particular $\|\varepsilon^-_{L,M}\|_{\mathbb{A}} \le \|\psi_0\|_{\mathbb{A}} + 2\|u_0\|_{\mathbb{U}}$, where the

factor 2 comes from taking into account for possible jumps in u_0. Eventually, as far as $\varepsilon_{\omega,L,M}$ is concerned, note that according to its definition in (6.89)

$$\varepsilon_{\omega,L,M} = \omega_{L,M}\pi_L^+\rho_L^+(\mathfrak{M}_M^* - \mathfrak{M}^*) + \omega_{L,M}(\pi_L^+\rho_L^+ - I_{\mathbb{U}})\mathfrak{M}^* + (\omega_{L,M} - \omega)\mathfrak{M}^*,$$

the first and the second addends of the right-hand side vanish thanks to the same arguments adopted for (6.94) under (T4), (N6) and (N7). As for the third addend, $\omega_{L,M} \to \omega$ follows from Proposition 7.8, and, therefore, the first addend vanishes thanks to (7.1) in Lemma 7.1.

Finally, (7.18) shows that $\psi_{L,M}$ is uniformly bounded. Thus, eventually, it can be concluded that $\|(u_{L,M}, \psi_{L,M}, \omega_{L,M})\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}}$ is bounded uniformly with respect to both L and M, thanks to (6.87) and Proposition 6.11. □

All the steps to prove the validity of Assumption 6.17 are collected by the following Proposition.

Proposition 6.23. *Under (T1), (T2), (T4), (N1), (N2), (N4), (N6) and (N7),*

$$\begin{aligned}\lim_{L,M\to\infty} \frac{1}{r_2(L,M)} &\|[D\Psi_{L,M}(u^*,\psi^*,\omega^*)]^{-1}\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}\leftarrow\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \\ &\cdot\|\Psi_{L,M}(u^*,\psi^*,\omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}} = 0,\end{aligned} \tag{6.95}$$

where

$$r_2(L,M) := \min\left\{r_1, \frac{1}{2\kappa\|[D\Psi_{L,M}(u^*,\psi^*,\omega^*)]^{-1}\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}\leftarrow\mathbb{U}\times\mathbb{A}\times\mathbb{B}}}\right\}$$

with r_1 and κ as in Proposition 6.18.

Proof. Thanks to Lemma 6.22 and to the fact that r_1 and κ in Proposition 6.18 are independent of L and M (as observed right after its proof), it remains to prove that $\|\Psi_{L,M}(u^*,\psi^*,\omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}}$ vanishes. The fact that (u^*,ψ^*,ω^*) is a fixed point for Φ gives

$$\begin{aligned}\|\Psi_{L,M}(u^*,\psi^*,\omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \le &\|(I_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}} - P_LR_L)(u^*,\psi^*,\omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \\ &+\|P_LR_L[\Phi_M(u^*,\psi^*,\omega^*) - \Phi(u^*,\psi^*,\omega^*)]\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}}.\end{aligned} \tag{6.96}$$

The second addend in the right-hand side above vanishes under (N6) and (N7) and thanks to (7.6) of Corollary 7.3. The first addend vanishes as well by Lemma 7.5, which shows in particular that u^* and $\psi^{*\prime}$ are continuous. □

6.5 CONVERGENCE

As shown in Sections 6.3 and 6.4, formulation (6.2) satisfies all the assumptions required in [79] under certain choices on the relevant spaces, the discretization, and the regularity of both the original and the discrete right hand side. The proof of the convergence of the relevant FEM (see Subsection 3.3.1) will be concluded in this section, by stating two theorems, namely [79, Theorems 1 and 2], which ensure the convergence of the general method provided that all the required assumptions are satisfied. The rest of the section will be dedicated to comment on the resulting rate of convergence, as well as some comments on the convergence of the SEM.

Theorem 6.24 ([79, Theorem 1, page 538]). *Under (T1), (T2), (T4), (N1), (N2), (N4), (N5), (N6) and (N7), there exists a positive integer $\overline{N}$ such that, for every $L, M \geq \overline{N}$, given $r_2(L,M)$ defined as in Proposition 6.23, $P_L R_L \Phi_M$ has a unique fixed point $(\tilde{u}^*_{L,M}, \tilde{\psi}^*_{L,M}, \tilde{\omega}^*_{L,M})$ in $\mathcal{B}((u^*,\psi^*,\omega^*), r_2(L,M))$ which satisfies*

$$\begin{aligned}\|(\tilde{u}^*_{L,M}, \tilde{\psi}^*_{L,M}, \tilde{\omega}^*_{L,M}) - (u^*,\psi^*,\omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \\ \leq 2\|[D\Psi_{L,M}(u^*,\psi^*,\omega^*)]^{-1}\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}\leftarrow\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \\ \cdot\|\Psi_{L,M}(u^*,\psi^*,\omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}}.\end{aligned}$$

Moreover, we have the expansion

$$\begin{aligned}(\tilde{u}^*_{L,M}, \tilde{\psi}^*_{L,M}, \tilde{\omega}^*_{L,M}) - (u^*,\psi^*,\omega^*) \\ = -[D\Psi_{L,M}(u^*,\psi^*,\omega^*)]^{-1}\Psi_{L,M}(u^*,\psi^*,\omega^*) + \delta_{L,M},\end{aligned}$$

where

$$\begin{aligned}\|\delta_{L,M}\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \leq 4\kappa\cdot\|[D\Psi_{L,M}(u^*,\psi^*,\omega^*)]^{-1}\|^3_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}\leftarrow\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \\ \cdot\|\Psi_{L,M}(u^*,\psi^*,\omega^*)\|^2_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}}\end{aligned}$$

for κ defined as in Proposition 6.18.

Theorem 6.25 ([79, Theorem 2, page 539]). *Under (T1), (T2), (T4), (N1), (N2), (N4), (N5), (N6) and (N7), there exists a positive integer $\hat{N}$ such that, for all $L, M \geq \hat{N}$, the operator $R_L\Phi_M P_L$ has a fixed point $(u^*_{L,M}, \psi^*_{L,M}, \omega^*_{L,M})$ and*

$$\begin{aligned}\|P_L(u^*_{L,M}, \psi^*_{L,M}, \omega^*_{L,M}) \ - \ (u^*,\psi^*,\omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \\ \leq 2\,\|[D\Psi_{L,M}(u^*,\psi^*,\omega^*)]^{-1}\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}\leftarrow\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \\ \cdot\|\Psi_{L,M}(u^*,\psi^*,\omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}}\end{aligned}$$

and

$$\begin{aligned}P_L(u^*_{L,M}, \psi^*_{L,M}, \omega^*_{L,M}) - (u^*,\psi^*,\omega^*) \\ = -[D\Psi_{L,M}(u^*,\psi^*,\omega^*)]^{-1}\Psi_{L,M}(u^*,\psi^*,\omega^*) + \delta_{L,M},\end{aligned}$$

where $\delta_{L,M}$ is bounded as in Theorem 6.24. Moreover, if $(\hat{u}^*_{L,M}, \hat{\psi}^*_{L,M}, \hat{\omega}^*_{L,M})$ is another fixed point of $R_L\Phi_M P_L$, then

$$\|P_L(\hat{u}^*_{L,M}, \hat{\psi}^*_{L,M}, \hat{\omega}^*_{L,M}) - (u^*,\psi^*,\omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}} > r_2(L,M)$$

and

$$\begin{aligned}\|(\hat{u}^*_{L,M}, \hat{\psi}^*_{L,M}, \hat{\omega}^*_{L,M}) - (u^*_{L,M}, \psi^*_{L,M}, \omega^*_{L,M})\|_{\mathbb{U}_L\times\mathbb{A}_L\times\mathbb{B}} \\ > \frac{r_2(L,M)}{2\cdot\max\{\|\pi^+_L\|_{\mathbb{U}\leftarrow\mathbb{U}_L}, \|\pi^-_L\|_{\mathbb{A}\leftarrow\mathbb{A}_L}, 1\}}\end{aligned}$$

for $r_2(L,K)$ defined as in Proposition 6.23. Finally,

$$\begin{aligned}\|(v^*_{L,M}, \omega^*_{L,M}) - (v^*,\omega^*)\|_{\mathbb{V}\times\mathbb{B}} \leq 2\cdot\max\{\|\mathcal{G}\|_{\mathbb{V}\leftarrow\mathbb{U}\times\mathbb{A}}, 1\} \\ \cdot\|[D\Psi_{L,M}(u^*,\psi^*,\omega^*)]^{-1}\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}\leftarrow\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \\ \cdot\|\Psi_{L,M}(u^*,\psi^*,\omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}}.\end{aligned} \tag{6.97}$$

As a side remark, it is worth observing that Assumption 6.6 is not directly needed to prove the two final theorems above. However, Assumption 6.16 is needed, and its validity implies anyway the one of Assumption 6.6 for a given approximation G_M in place of G, the proof of which would be unchanged. Moreover, in many cases, this implies in turn the validity of Assumption 6.6 for G: indeed, when the approximation G_M is needed to discretize a distributed integral, G is at least as regular as G_M. These are the reasons why the proof of Theorem 6.7 has been presented in Section 6.3 in full detail, as to follow the presentation in [79]. Eventually, observe that the comment given after the proof about the failure of formulation (6.3) holds unaltered, since the mentioned critical step is independent of G or G_M.

Note that the second factor in the right-hand side of (6.97) is well-defined thanks to Proposition 6.4. Thanks to Lemma 6.22, the error on (v^*, ω^*) is determined by the last factor, namely the *consistency error*. The bound (6.96) for the latter, in view of Corollary 7.3, leads to the bound

$$\|\Psi_{L,M}(u^*, \psi^*, \omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \leq \varepsilon_L + \max\{\Lambda_m + \Lambda'_m, 1\}\varepsilon_M, \tag{6.98}$$

where

$$\varepsilon_L := \|(I_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}} - P_L R_L)(u^*, \psi^*, \omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}}$$

and

$$\varepsilon_M := \|\Phi_M(u^*, \psi^*, \omega^*) - \Phi(u^*, \psi^*, \omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}}. \tag{6.99}$$

are the key contributions to the consistency error, called respectively *primary* and *secondary* consistency errors, and will be analyzed separately in the following Subsections.

6.5.1 Primary consistency error

According to (6.46), (6.47) and (1.14), the primary consistency error is bounded as

$$\varepsilon_L \leq \max\{\|u^* - \pi_L^+ \rho_L^+ u^*\|_{\mathbb{U}}, \|\psi^* - \pi_L^- \rho_L^- \psi^*\|_{\mathbb{A}}\}.$$

Therefore an upper bound on ε_L is determined by the regularity of both u^* and ψ^*, according to the result below.

Theorem 6.26. *Let $G \in C^p(\mathbb{Y}, \mathbb{R}^d)$ for some integer $p \geq 1$. Then, under (T1), (T2), (N1) and (N2), it holds that $u^* \in C^p([0,1], \mathbb{R}^d)$, $\psi^* \in C^{p+1}([-1,0], \mathbb{R}^d)$, $v^* \in C^{p+1}([-1,1], \mathbb{R}^d)$ and*

$$\varepsilon_L = O\left(h^{\min\{m,p\}}\right). \tag{6.100}$$

Proof. Note that $v^* = \mathcal{G}(u^*, \psi^*)$ satisfies (6.2), hence its periodic extension to $[-1, \infty]$ is a (periodic) solution of (6.1). Thus, being v^* continuous, if G is continuous, then v^* is continuously differentiable in $[0, +\infty)$, which implies that u^* is continuous. Moreover, ψ^* is continuously differentiable by periodicity and, $\psi^{*\prime}(0) = u^*(0)$ follows again by periodicity since $v^{*\prime}$ is continuous at 1. This means that v^* is continuously differentiable in $[-1, 1]$. As a consequence, if $p = 1$, u^* becomes continuously differentiable and the

whole reasoning can be repeated, proving the first part of the thesis. This is a consequence of the well-known *smoothing effect* of DDEs.

To prove (6.100), note that

$$\|u^* - \pi_L^+ \rho_L^+ u^*\|_{\mathbb{U}} \leq \frac{\|u^{*(m+1)}\|_\infty}{(m+1)!} \cdot h^{m+1} \tag{6.101}$$

holds if $p \geq m+1$, while

$$\|u^* - \pi_L^+ \rho_L^+ u^*\|_{\mathbb{U}} \leq (1+\Lambda_m) \left(\frac{h}{2}\right)^p \frac{c_p}{m^p} \cdot \|u^{*(p)}\|_\infty \tag{6.102}$$

holds if $p \leq m+1$, with c_p a positive constant independent of m. (6.101) is a direct consequence of the standard Cauchy interpolation reminder (Theorem 3.7). (6.102) follows from Theorems 3.3 and 3.6.

Secondly, similar results can be obtained for the component in $\mathbb{A}$, by recalling that $\|\cdot\|_{\mathbb{A}}$ is given by the second of (1.13). Indeed, on the one hand, based on the same arguments used above for (6.101) and (6.102),

$$\|\psi^* - \pi_L^- \rho_L^- \psi^*\|_\infty \leq \frac{\|\psi^{*(m+1)}\|_\infty}{(m+1)!} \cdot h^{m+1}$$

holds if $p \geq m$, while

$$\|\psi^* - \pi_L^- \rho_L^- \psi^*\|_\infty \leq (1+\Lambda_m) \left(\frac{h}{2}\right)^{p+1} \frac{c'_p}{m^{p+1}} \cdot \|\psi^{*(p+1)}\|_\infty$$

holds if $p \leq m$, with c'_p a positive constant independent of m. On the other hand,

$$\|(\psi^* - \pi_L^- \rho_L^- \psi^*)'\|_\infty \leq \frac{\|\psi^{*(m+1)}\|_\infty}{m!} \cdot h^m \tag{6.103}$$

holds if $p \geq m$, while

$$\|(\psi^* - \pi_L^- \rho_L^- \psi^*)'\|_\infty \leq \Lambda_m \left(\frac{h}{2}\right)^p \frac{c''_p}{m^{p-1}} \cdot \|\psi^{*(p+1)}\|_\infty \tag{6.104}$$

holds if $p \leq m$, with c''_p a positive constant independent of m. In particular, (6.103) follows by the Cauchy interpolation reminder (Theorem 3.7), after taking the first derivative of the reminder itself. Moreover, (6.104) follows similarly to (7.5) in the proof of Lemma 7.2 thanks to [81, (12) at page 331] and Theorem 3.3. □

According to the theorem above, $O(h^m)$ is a lower bound for the global consistency error in (6.98), even in the case that the periodic solution were smooth enough, i.e., $p > m+1$, and assuming the absence of a secondary discretization. This is in contrast to the estimate $O(h^{m+1})$ obtained in [77] (see in particular the conclusions therein). This difference is due to the fact that in formulation (6.2) the space $\mathbb{A}$ is infinite-dimensional and needs to be discretized, possibility which is only mentioned in [79], rather than being concretely elaborated and, simultaneously, to the fact that functions in $\mathbb{A}$ must be differentiable, due, in turn, to the need of differentiating with respect to the period, as already remarked several times. After all, formulation (6.3), in which $\mathbb{A}$ is finite-dimensional, does not satisfy all the required assumptions to develop this convergence analysis.

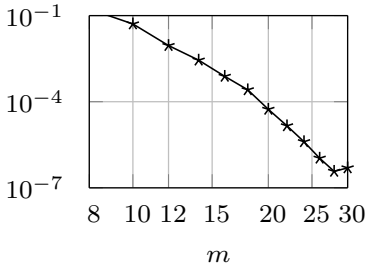

Figure 6.1: Periodic solution of (5.3) at $\lambda = 1.7$: continuous error for $L = 1$ using Chebyshev points.

6.5.2 Secondary consistency error

The error term ε_M in (6.99) concerns only the secondary discretization and, according to (6.9) and (6.64), it reduces to

$$\varepsilon_M := \omega^* \| G_M(v^*_{(\cdot)} \circ s_{\omega^*}) - G(v^*_{(\cdot)} \circ s_{\omega^*}) \|_{\mathbb{U}}. \tag{6.105}$$

Although this error is absent in case a secondary discretization is not needed, it was already remarked that the latter is (at least) necessary when the equation contains distributed delays, in which case it consists in applying suitable quadrature rules to approximate the concerned integrals. In all cases where the presence of integrals is the only thing that prevents the possibility to compute the right hand side exactly, (6.105) can be seen as a quadrature error. Therefore, the secondary discretization plays a minor role in determining the overall consistency error, provided that one chooses a quadrature formula that guarantees at least the same order of the primary consistency error (as far as M varies proportionally to L). Otherwise, assuming that (6.105) falls below a given tolerance, say TOL, the consistency error decays down to TOL as fast as the primary consistency error.

6.5.3 Convergence of the spectral element method

As explained in Subsection 3.3.1 two methods can be considered as far as the convergence of the proposed piecewise collocation strategy is concerned, namely the FEM and the SEM. In particular, with reference to the primary discretization under (N1) and (N2), the FEM consists in letting $L \to \infty$ while keeping m fixed, while the SEM consists in letting $m \to \infty$ while keeping L fixed. The analysis carried out in Sections 6.3 and 6.4 are presented for the FEM, concerning which Theorem 6.26 guarantees an error of magnitude $O(L^{-m})$ under suitable regularity conditions.

On the other hand, it is not yet clear whether the convergence of the SEM is guaranteed under the general framework of reference for the current work. Note that, even if it were not the case, this would not imply that the SEM does not converge for periodic BVPs. Indeed, some numerical experiments run by the author suggest the opposite. Figure 6.1 shows the results of one of such experiments, suggesting a spectral decay of the error.

Thus, it is anyway likely that an error analysis different from the one proposed in [77, 78, 79] would work.

The rest of this subsection will aim at going through the various arguments used in the analysis in Sections 6.4, 7.1, 7.3 and Subsection 6.5.1 which would fail for the SEM based on (N1) and (N2). Some of these arguments can be readapted (possibly by adding further assumptions, e.g., on the requirements of regularity), yet some others seem not amenable of a definitive solution, or at least of a simple one.

Starting from Section 6.4, the first point suggesting that the SEM might fail is in the proof of Lemma 6.19, in particular due to (7.7) in Lemma 7.6. In the case of the SEM indeed,

$$\lim_{m,M\to\infty} \|(\pi_L^+\rho_L^+ - I_{\mathbb{U}})\mathcal{K}_M^{*,+}\|_{\mathbb{U}\leftarrow\mathbb{U}} = 0$$

cannot hold by Faber's theorem (Theorem 3.8), since (N4) only guarantees that functions in the range of $\mathcal{K}_M^{*,+}$ are continuous, as observed in the proof of Lemma 7.6. Nevertheless, the problem can be easily overcome by assuming (N5), which guarantees the functions in the range of $\mathcal{K}_M^{*,+}$ to be Lipschitz continuous (since $\mathcal{G}^+$ already maps to Lipschitz continuous functions), thanks to Theorem 3.11. The same argument is used in the proofs of both Lemma 6.20 and Lemma 6.22, which can be fixed similarly.

There is also another issue which emerges in the proof of Lemma 6.22, concerning the first addend in the right-hand side of (6.94). Indeed, by Theorem 3.5, Λ_m in (7.1) of Lemma 7.1 grows unbounded independently of the choice of the collocation abscissae, at least as $O(\log m)$ (and at most with the same order in case of Chebyshev-type nodes, see Theorems 3.9 and 3.10). Therefore, in order to ensure convergence one should assume to balance this growth with the rate of convergence of the secondary discretization, being the attention focused on the term $\pi_L^+\rho_L^+(\mathfrak{L}_M^* - \mathfrak{L}^*)$. Since primary and secondary discretizations can be chosen independently, this balance can be a reasonable option. Of course, if a secondary discretization is not required, the term is not even present and the issue becomes meaningless. Note that similar issues appear in the proof of the convergence of $\varepsilon_{\omega,L,M}$ in (6.89). In the latter also the convergence of $\omega_{L,M}$ to ω is required, which should follow from Proposition 7.8, whose validity for the SEM will be discussed next.

As for the proof of the first part, namely (7.11), in Proposition 7.8 the convergence of $\xi_{L,M,1}^*$ depends on the four terms at the right-hand side of (7.12). In particular, for the first and the third ones, the same balance between primary and secondary discretization mentioned above has to be considered. The second addend could be made vanishing by ensuring that the interpolation error $\|(\pi_L^+\rho_L^+ - I_{\mathbb{U}})\mathfrak{M}_M^*\|_{\mathbb{U}}$ decays fast enough to override the growth of $\pi_L^-\rho_L^-$. The latter, thanks to Theorems 3.9 and 3.10, according to (7.3) of Lemma 7.2, grows at best as $O(m\log m)$ for Chebyshev-type nodes. Consequently, $\mathfrak{M}_M^*$ should be at least continuously differentiable with Lipschitz continuous first derivative, thus guaranteeing, by Corollary 3.12 with $k = 1$, that the above interpolation error is $O(\log m/m^2)$ and the sought balance is scored. Finally, the fourth term concerns the interpolation error in $\mathbb{A}$ of the function $\int_0^1 [T^*(1,s)X_0]\mathfrak{M}^*(s)\,\mathrm{d}s$. As this is the state at 1 of (7.13), the above map has a Lipschitz continuous first derivative under (T5). Thus, if the abscissae $c_1,\dots,c_{m-1}$ are chosen corresponding to Chebyshev-type zeros, as well as $c_m = 1$, one can apply Theorem 3.13 with $i = q = 1$ to replace (7.5)

in Lemma 7.2 with $\|(\pi_L^- \rho_L^- \psi - \psi)'\|_\infty \leq c\Lambda_m E_{m-1}(\psi')$. Note that, for other choices of the abscissae there are no similar results: indeed, a further factor m may, in principle, appear in the right-hand side (but not more, thanks to [81, (12) at page 331]), thus requiring a degree of regularity that cannot be obtained when the same analysis is carried-out for $\xi^*_{L,M,2}$ (see (7.16) and the relevant comments).

Despite the remedies to the problems above, to complete the proof of Proposition 7.8 boundedness of $\psi_{L,M}$ is required, and the latter becomes mandatory for Lemma 6.22 to hold. However, such boundedness is not guaranteed by (7.18) given (7.3) of Lemma 7.2. Yet it could well be that the norm of $[D\Psi_{L,M}(u^*, \psi^*, \omega^*)]^{-1}$ grows with m, but not as fast as the (square root of the) consistency error $\|\Psi_{L,M}(u^*, \psi^*, \omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}}$ decays, recall in fact (6.95) in Proposition 6.23. The hypothesis is not far from being reasonable, given that the consistency error depends on the regularity of the periodic solution at hands, in view of a possible variant for the SEM of Theorem 6.26. In any case, proving (or disproving) this requires a much deeper analysis than the one carried for FEM in this chapter.

7 PROOFS OF TECHNICAL RESULTS

This chapter contains the proofs of some results which have been used in Chapter 6. Being these proofs rather technical, they were not included therein in order to not interrupt the main reading flow.

7.1 BASIC RESULTS ON THE DISCRETIZATION

Lemma 7.1. *Let ρ_L^+, π_L^+ and Λ_m be defined respectively in (6.51), (6.53) and (6.56) under (N1) and let C^+ be defined in (6.92). Then, under (T2),*

$$\|\pi_L^+\rho_L^+\|_{\mathbb{U}\leftarrow\mathbb{U}} \leq \Lambda_m \tag{7.1}$$

holds for all positive integers L and

$$\lim_{L\to\infty} \|\pi_L^+\rho_L^+ u - u\|_{\mathbb{U}} = 0 \tag{7.2}$$

holds for all $u \in C^+$.

Proof. According to the notation of Section 6.4,

$$\pi_L^+\rho_L^+ u(t) = \sum_{j=0}^{m} \ell_{m,i,j}(t) u(t_{i,j}^+)$$

holds for $u \in \mathbb{U}$ and $t \in [t_{i-1}^+, t_i^+]$, $i = 1, \dots, L$. Then (7.1) follows from

$$\|\pi_L^+\rho_L^+ u\|_{\mathbb{U}} \leq \max_{i=1,\dots,L} \max_{t\in[t_{i-1}^+,t_i^+]} \sum_{j=0}^{m} |\ell_{m,i,j}(t)| \|u\|_{\mathbb{U}} = \Lambda_m \|u\|_{\mathbb{U}},$$

which in turn follows from the fact that the Lebesgue constant is independent of i. As for (7.2),

$$\begin{aligned}
(\pi_L^+\rho_L^+ u - u)(t) &= \sum_{j=0}^{m} \ell_{m,i,j}(t) u(t_{i,j}^+) - \sum_{j=0}^{m} \ell_{m,i,j}(t) u(t) + \sum_{j=0}^{m} \ell_{m,i,j}(t) u(t) - u(t) \\
&= \sum_{j=0}^{m} \ell_{m,i,j}(t) [u(t_{i,j}^+) - u(t)] + \left(\sum_{j=0}^{m} \ell_{m,i,j}(t) - 1 \right) u(t) \\
&= \sum_{j=0}^{m} \ell_{m,i,j}(t) [u(t_{i,j}^+) - u(t)]
\end{aligned}$$

holds always for $t \in [t_{i-1}^+, t_i^+]$, $i = 1, \dots, L$. Therefore

$$\|\pi_L^+\rho_L^+ u - u\|_{\mathbb{U}} \leq \Lambda_m \omega(u; h),$$

where ω denotes the modulus of continuity. The latter vanishes as $h \to 0$ only if u is at least continuous. □

Lemma 7.2. *Let ρ_L^-, π_L^-, Λ_m and Λ'_m be defined respectively in (6.60), (6.62), (6.56) and (6.57) under (N2) and let $C^{1,-} := C^1([-1,0],\mathbb{R}^d)$. Then, under (T2),*

$$\|\pi_L^-\rho_L^-\|_{\mathbb{A}\leftarrow\mathbb{A}} \leq \Lambda_m + \Lambda'_m \tag{7.3}$$

holds for all positive integers L and

$$\lim_{L\to\infty} \|\pi_L^-\rho_L^-\psi - \psi\|_{\mathbb{A}} = 0 \tag{7.4}$$

holds for all $\psi \in C^{1,-}$.

Proof. The proof of (7.3) is analogous to that of (7.1) in Lemma 7.1 once considered that $\|\cdot\|_{\mathbb{A}}$ is given by (1.13). As for (7.4), $\|\pi_L^-\rho_L^-\psi - \psi\|_\infty \leq \Lambda_m\omega(\psi;h)$, follows similarly by the proof of (7.2) in Lemma 7.1, while

$$\|(\pi_L^-\rho_L^-\psi - \psi)'\|_\infty \leq c(m+1)\Lambda_m E_{m-1}(\psi') \tag{7.5}$$

holds for some positive constant c thanks to [81, (12) at page 331], where $E_m(f)$ denotes the best uniform approximation error of f with (piecewise) polynomials of degree m. As for the latter $E_{m-1}(\psi') \leq 6\omega(\psi', h/2m)$ holds thanks to Theorem 3.3, so that it vanishes as $h \to 0$ only if ψ is at least continuously differentiable. □

Corollary 7.3. *Let R_L, P_L and Λ_m be defined respectively in (6.46), (6.47) and (6.56) under (N1) and (N2). Then*

$$\|P_L R_L\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}\leftarrow\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \leq \max\{\Lambda_m + \Lambda'_m, 1\} \tag{7.6}$$

holds for all positive integers L.

Proof.

$$\begin{aligned}\|P_L R_L\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}\leftarrow\mathbb{U}\times\mathbb{A}\times\mathbb{B}} &= \max\{\|\pi_L^+\rho_L^+\|_{\mathbb{U}\leftarrow\mathbb{U}}, \|\pi_L^-\rho_L^-\|_{\mathbb{A}\leftarrow\mathbb{A}}, \|I_{\mathbb{B}}\|_{\mathbb{B}\leftarrow\mathbb{B}}\}\\ &\leq \max\{\Lambda_m, \Lambda_m + \Lambda'_m, 1\}\\ &= \max\{\Lambda_m + \Lambda'_m, 1\},\end{aligned}$$

thanks to (1.14), (7.1) in Lemma 7.1 and (7.3) in Lemma 7.2. □

7.2 RESULTS CONCERNING THE THEORETICAL ASSUMPTIONS

Lemma 7.4. *Let $(u^*, \psi^*, \omega^*) \in \mathbb{U}\times\mathbb{A}\times\mathbb{B}$ be a fixed point of Φ in (6.9). Then, under (T2), $v^* := \mathcal{G}(u^*, \psi^*)$ for $\mathcal{G}$ in (6.8) is Lipschitz continuous, in particular*

$$|v^*(t_1) - v^*(t_2)| \leq \|(u^*, \psi^*, \omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \cdot |t_1 - t_2|$$

holds for all $t_1, t_2 \in [-1, 1]$.

Proof. By the choice of $\mathbb{V}_2$ in (T2), the derivative $v^{*\prime}$ of v^* is bounded, and by the choice of $\mathbb{A}_2$ in (T2) it can be expressed as

$$v^{*\prime}(t) := \begin{cases} u^*(t), & t \in [0,1],\\ \psi^{*\prime}(t), & t \in [-1,0).\end{cases}$$

Therefore,

$$\|v^{*\prime}\|_\infty = \max\{\|u^*\|_\infty, \|\psi^{*\prime}\|_\infty\} \leq \max\{\|u^*\|_{\mathbb{U}_2}, \|\psi^*\|_{\mathbb{A}_2}\} \leq \|(u^*, \psi^*, \omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}},$$

from which it follows that

$$|v^*(t_1) - v^*(t_2)| \leq \|v^{*\prime}\|_\infty \cdot |t_1 - t_2| \leq \|(u^*, \psi^*, \omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \cdot |t_1 - t_2|.$$

□

Lemma 7.5. *Let $(u^*, \psi^*, \omega^*) \in \mathbb{U} \times \mathbb{A} \times \mathbb{B}$ be a fixed point of Φ in (6.9) and $v^* := \mathcal{G}(u^*, \psi^*)$ for $\mathcal{G}$ in (6.8). Then, under (T2) and (T5), u^*, $\psi^{*\prime}$ and $v^{*\prime}$ are Lipschitz continuous, in particular*

$$|v^{*\prime}(t_1) - v^{*\prime}(t_2)| \leq \omega^* \kappa_{2,1} \|(u^*, \psi^*, \omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}} \cdot |t_1 - t_2|$$

hold for all $t_1, t_2 \in [-1, 1]$, where $\kappa_{2,1}$ is defined as in (6.26).

Proof. Thanks to (T5), the constant $\kappa_{2,1}$ in (6.26) is well-defined, and, for $t_1, t_2 \in [0, 1]$,

$$|u^*(t_1) - u^*(t_2)| \leq \omega^* \kappa_{2,1} \|v^*_{t_1} \circ s_{\omega^*} - v^*_{t_2} \circ s_{\omega^*}\|_{\mathbb{Y}}.$$

Thus, u^* is Lipschitz continuous with constant $\omega^* \kappa_{2,1} \|(u^*, \psi^*, \omega^*)\|_{\mathbb{U}\times\mathbb{A}\times\mathbb{B}}$ by Lemma 7.4. The same holds for $\psi^{*\prime}$ by periodicity and, in turn, for $v^{*\prime}$, being it the continuous junction of two Lipschitz continuous functions with the same constant. □

7.3 RESULTS CONCERNING THE LAST NUMERICAL ASSUMPTION

Lemma 7.6. *Let ρ_L^+ and π_L^+ be defined respectively in (6.51) and (6.53) under (N1), $\mathcal{K}_M^{*,+}$ and $\mathcal{K}_M^{*,-}$, $\mathcal{K}^{*,+}$ and $\mathcal{K}^{*,-}$ be defined in (6.71). Then, under (T2), (N4) and (N7),*

$$\lim_{L,M\to\infty} \|\pi_L^+ \rho_L^+ \mathcal{K}_M^{*,+} - \mathcal{K}^{*,+}\|_{\mathbb{U}\leftarrow\mathbb{U}} = 0, \tag{7.7}$$

and

$$\lim_{L,M\to\infty} \|\pi_L^+ \rho_L^+ \mathcal{K}_M^{*,-} - \mathcal{K}^{*,-}\|_{\mathbb{U}\leftarrow\mathbb{A}} = 0. \tag{7.8}$$

Proof. As for (7.7), a bound can be obtained by the inequality

$$\|\pi_L^+ \rho_L^+ \mathcal{K}_M^{*,+} - \mathcal{K}^{*,+}\|_{\mathbb{U}\leftarrow\mathbb{U}} \leq \|(\pi_L^+ \rho_L^+ - I_{\mathbb{U}})\mathcal{K}_M^{*,+}\|_{\mathbb{U}\leftarrow\mathbb{U}} + \|\mathcal{K}_M^{*,+} - \mathcal{K}^{*,+}\|_{\mathbb{U}\leftarrow\mathbb{U}}.$$

The second addend in the right-hand side above vanishes thanks to (N7). It thus also follows that $\mathcal{K}_M^{*,+}$ is uniformly bounded with respect to M. This in turn makes the first addend vanish as well, given that $\mathcal{K}_M^{*,+}\mathbb{U} \subseteq C([0, 1], \mathbb{R}^d)$ as it follows from (6.8) through the definition of $\mathcal{G}^+$ in (6.71) and the continuity of $\mathfrak{L}_M^*$ under (N4). The same arguments hold for (7.8). □

Lemma 7.7. *Let $\mathfrak{L}^*, \mathfrak{M}^*$ and $\mathfrak{L}^*_M, \mathfrak{M}^*_M$ be defined respectively in (6.32) and (6.69). Then, under (N6) and (N7),*

$$\lim_{M\to\infty} \|\mathfrak{L}^*_M - \mathfrak{L}^*\|_{\mathcal{L}(\Upsilon,\mathbb{R}^d)\leftarrow[0,1]} = 0 \tag{7.9}$$

and

$$\lim_{M\to\infty} \|\mathfrak{M}^*_M - \mathfrak{M}^*\|_\infty = 0. \tag{7.10}$$

Proof. (7.9) follows directly by (N7). (7.10) follows by (N6) and (N7). □

Proposition 7.8. *Let ω and $\omega_{L,M}$ be given as in (6.39) and (6.84), respectively. Then, under (T4), (N4), (N6) and (N7),*

$$\lim_{L,M\to\infty} \omega_{L,M} = \omega.$$

Proof. Let ξ^*_1 and ξ^*_2 be as in the proof of Proposition 6.11 and $\xi^*_{L,M,1}$, $\xi^*_{L,M,2}$ and $\nu_{L,M}$ be as in (6.83). The first step consists in proving that

$$\lim_{L,M\to\infty} \|\xi^*_{L,M,1} - \xi^*_1\|_Y = 0. \tag{7.11}$$

From (6.38) and the first of (6.81), it follows that

$$\begin{aligned}
\xi^*_{L,M,1} - \xi^*_1 &= \pi^-_L \rho^-_L \int_0^1 [T^*_{L,M}(1,s)X_0]\pi^+_L \rho^+_L \mathfrak{M}^*_M(s)\,\mathrm{d}s \\
&\quad - \int_0^1 [T^*(1,s)X_0]\mathfrak{M}^*(s)\,\mathrm{d}s \\
&= \pi^-_L \rho^-_L \int_0^1 [(T^*_{L,M}(1,s) - T^*(1,s))X_0]\pi^+_L \rho^+_L \mathfrak{M}^*_M(s)\,\mathrm{d}s \\
&\quad + \pi^-_L \rho^-_L \int_0^1 [T^*(1,s)X_0][\pi^+_L \rho^+_L - I_{\mathbb{U}}]\mathfrak{M}^*_M(s)\,\mathrm{d}s \\
&\quad + \pi^-_L \rho^-_L \int_0^1 [T^*(1,s)X_0](\mathfrak{M}^*_M(s) - \mathfrak{M}^*(s))\,\mathrm{d}s \\
&\quad + (\pi^-_L \rho^-_L - I_{\mathbb{A}}) \int_0^1 [T^*(1,s)X_0]\mathfrak{M}^*(s)\,\mathrm{d}s.
\end{aligned} \tag{7.12}$$

From Lemma 7.7, it follows that $\mathfrak{M}^*_M$ is uniformly bounded. As a consequence, the third addend in the right-hand side of the last equality above vanishes thanks to (7.3) of Lemma 7.2, and the first addend vanishes as well also thanks to Lemma 6.20 and (7.1) of Lemma 7.1. Since $\mathfrak{M}^*_M$ is continuous under (N4) thanks to Lemma 7.5 (recall its definition from (6.14) and the second of (6.32)), the second addend vanishes similarly thanks to (7.2) of Lemma 7.1. Finally, as for the last addend, note that $\int_0^1 [T^*(1,s)X_0]\mathfrak{M}^*(s)\,\mathrm{d}s$ is the state solution at 1 of

$$\begin{cases} v'(t) = \mathfrak{L}^*(t)[v_t \circ s_{\omega^*}] + \mathfrak{M}^*(t), \quad t \in [0,1], \\ v_0 = 0, \end{cases} \tag{7.13}$$

as it can be seen by applying the variation of constants formula as done for (6.77). As such it is continuously differentiable, being the right-hand side of the DDE continuous under (T4) similarly as already observed above for

$\mathfrak{M}_M^*$. Therefore, also this addend vanishes thanks to (7.4) of Lemma 7.2, thus (7.11) holds. Moreover,

$$\lim_{L,M\to\infty} k_{L,M,1} = k_1 \tag{7.14}$$

follows from the second of (6.74) in Lemma 6.20. Note that the definition of ξ_1^* does not change if one considers system (6.85) in place of (6.33). Similarly, the definition of $\xi_{L,M,1}$ does not change if one considers system (6.86) in place of (6.76).

Note now that the same reasoning cannot be used to prove that

$$\lim_{L,M\to\infty} \|\xi_{L,M,2}^* - \xi_2^*\|_Y = 0. \tag{7.15}$$

From the second of (6.81), it follows that

$$\begin{aligned}
\xi_{L,M,2}^* - \xi_2^* &= \pi_L^- \rho_L^- \int_0^1 [T_{L,M}^*(1,s)X_0]u_0(s)\,\mathrm{d}s \\
&\quad - \int_0^1 [T^*(1,s)X_0]u_0(s)\,\mathrm{d}s \\
&= \pi_L^- \rho_L^- \int_0^1 [(T_{L,M}^*(1,s) - T^*(1,s))X_0]u_0(s)\,\mathrm{d}s \\
&\quad + (\pi_L^- \rho_L^- - I_{\mathbb{A}}) \int_0^1 [T^*(1,s)X_0]u_0(s)\,\mathrm{d}s.
\end{aligned} \tag{7.16}$$

However, in the last addend, $\int_0^1 [T^*(1,s)X_0]u_0(s)\,\mathrm{d}s$ is the state solution at 1 of

$$\begin{cases} v'(t) = \mathfrak{L}^*(t)[v_t \circ s_{\omega^*}] + u_0(t), \quad t \in [0,1], \\ v_0 = 0, \end{cases}$$

and the latter is not necessarily continuously differentiable since $u_0 \in \mathbb{U}$ is not necessarily continuous, so that (7.4) of Lemma 7.2 cannot be applied as done above. Nevertheless, one can define ξ_2^* from system (6.85) in place of (6.33), i.e.,

$$\xi_2^* := \int_0^1 [T^*(1,s)X_0](\mathfrak{L}^*[\mathcal{G}(u_0,\psi_0).\circ s_{\omega^*}])(s)\,\mathrm{d}s + \psi_0.$$

Similarly, one can define $\xi_{L,M,2}$ from system (6.86) in place of (6.76), i.e.,

$$\xi_{L,M,2}^* := \pi_L^- \rho_L^- \int_0^1 [T^*(1,s)X_0]\pi_L^+ \rho_L^+ (\mathfrak{L}_M^*[\mathcal{G}(u_0,\psi_0).\circ s_{\omega^*}])(s)\,\mathrm{d}s + \psi_0,$$

giving

$$\begin{aligned}
\xi_{L,M,2}^* - \xi_2^* &= \pi_L^- \rho_L^- \int_0^1 [T_{L,M}^*(1,s)X_0]\pi_L^+ \rho_L^+ (\mathfrak{L}_M^*[\mathcal{G}(u_0,\psi_0).\circ s_{\omega^*}])(s)\,\mathrm{d}s \\
&\quad - \int_0^1 [T^*(1,s)X_0](\mathfrak{L}^*[\mathcal{G}(u_0,\psi_0).\circ s_{\omega^*}])(s)\,\mathrm{d}s \\
&= \pi_L^- \rho_L^- \int_0^1 [(T_{L,M}^*(1,s) - T^*(1,s))X_0]\pi_L^+ \rho_L^+ (\mathfrak{L}_M^*[\mathcal{G}(u_0,\psi_0).\circ s_{\omega^*}])(s)\,\mathrm{d}s \\
&\quad + \pi_L^- \rho_L^- \int_0^1 [T^*(1,s)X_0][\pi_L^+ \rho_L^+ - I_{\mathbb{U}}](\mathfrak{L}_M^*[\mathcal{G}(u_0,\psi_0).\circ s_{\omega^*}])(s)\,\mathrm{d}s \\
&\quad + \pi_L^- \rho_L^- \int_0^1 [T^*(1,s)X_0](\mathfrak{L}_M^*[\mathcal{G}(u_0,\psi_0).\circ s_{\omega^*}] - \mathfrak{L}^*[\mathcal{G}(u_0,\psi_0).\circ s_{\omega^*}])(s)\,\mathrm{d}s \\
&\quad + (\pi_L^- \rho_L^- - I_{\mathbb{A}}) \int_0^1 [T^*(1,s)X_0](\mathfrak{L}^*[\mathcal{G}(u_0,\psi_0).\circ s_{\omega^*}])(s)\,\mathrm{d}s.
\end{aligned}$$

Thus, the proof of (7.15) can be accomplished by using the same arguments adopted in the proof of (7.11), since the term $\mathfrak{L}^*[\mathcal{G}(u_0,\psi_0)_\cdot \circ s_{\omega^*}]$ is continuous under (T4) and therefore the fourth addend vanishes. Moreover,

$$\lim_{L,M\to\infty} k_{L,M,2} = k_2 \tag{7.17}$$

thanks to (6.74) in Lemma 6.20.

Note that $\psi_{L,M}$ is bounded, from

$$\|\psi_{L,M}\|_{\mathbb{A}} \leq (\Lambda_m + \Lambda'_m)\|\mathcal{G}(u_{L,M},\psi_{L,M})_1\|_\infty + \|\psi_0\|_{\mathbb{A}}, \tag{7.18}$$

which, in turn, follows from the second of (6.76), where

$$\|\mathcal{G}(u_{L,M},\psi_{L,M})_1\|_\infty = \|\mathcal{G}(u_{L,M},\psi_{L,M})|_{[0,1]}\|_\infty$$

is bounded as it follows from the third of (6.76) and the continuity of p.

Eventually,

$$\lim_{L,M\to\infty} h_{L,M} = 0 \tag{7.19}$$

follows since in the third of (6.83) $\varphi_{L,M}$ converges to φ thanks to Lemma 6.20 again and $\nu_{L,M}$ in (6.79) vanishes. The latter statement is a consequence of $\psi_{L,M}$ being bounded, $\mu_{L,M} \to 1$ from Lemma 6.20 and that $(\pi_L^- \rho_L^- - I_{\mathbb{A}})T^*_{L,M}(1,0)$ vanishes since

$$\begin{aligned}(\pi_L^- \rho_L^- - I_{\mathbb{A}})T^*_{L,M}(1,0) = {} & (\pi_L^- \rho_L^- - I_{\mathbb{A}})[T^*_{L,M}(1,0) - T^*(1,0)] \\ & + (\pi_L^- \rho_L^- - I_{\mathbb{A}})T^*(1,0).\end{aligned}$$

Indeed, the right-hand side above vanishes under (N2) thanks to (7.4) of Lemma 7.2, Lemma 6.20 again and to the fact that the range of $T^*(1,0)$ contains only continuously differentiable functions.

In conclusion, by (7.14), (7.17) and (7.19),

$$\begin{aligned}\lim_{L,M\to\infty}(\omega_{L,M} - \omega) &= \lim_{L,M\to\infty}\left(-\frac{k_{L,M,2} + h_{L,M}}{k_{L,M,1}} + \frac{k_2}{k_1}\right) \\ &= \lim_{L,M\to\infty}\frac{-k_{L,M,2}k_1 - h_{L,M}k_1 + k_2 k_{L,M,1}}{k_{L,M,1}k_1} \\ &= \frac{-k_2 k_1 - 0\cdot k_1 + k_2 k_1}{k_1 k_1} = 0.\end{aligned}$$

□

Lemma 7.9. *Let ρ_L^- and π_L^- be defined respectively in (6.60) and (6.62) under (N2) and $\mathcal{G}_1^+, \mathcal{G}_1^-$ be defined in (6.91). Then, under (T2),*

$$\lim_{L,M\to\infty} \|\pi_L^- \rho_L^- \mathcal{G}_1^+ - \mathcal{G}_1^+\|_{\mathbb{A}\leftarrow C^+} = 0 \tag{7.20}$$

and

$$\lim_{L,M\to\infty} \|\pi_L^- \rho_L^- \mathcal{G}_1^- - \mathcal{G}_1^-\|_{\mathbb{A}\leftarrow\mathbb{A}} = 0. \tag{7.21}$$

Proof. (7.20) follows from the fact that $\mathcal{G}_1^+ C^+$ contains only continuously differentiable functions by the first of (6.91). (7.21) follows from the fact that $\mathcal{G}_1^- \mathbb{A}$ contains only constant functions by the second of (6.91). □

8 CONCLUDING REMARKS AND FUTURE WORK

As explained in Section 1.2, the infinite dimensionality of delay dynamical systems poses serious challenges in their qualitative and quantitative study. As a result, much work is still required to develop a comprehensive theory of DDEs and REs, as well as relevant numerical methods, particularly in the latter case. This thesis is meant to represent a (small) step towards this direction, proposing means to improve the performance of existing techniques (Chapter 4) and extensions of known numerical methods to new classes of equations (Chapter 5), as well as addressing theoretical problems concerning the convergence of such methods (Chapters 6 and 7).

In particular, it was proved in Chapter 4, through numerical experiments, that the internal approach proposed is more efficient than the standard approaches for the continuation of equilibria. Note that this efficiency comes at the cost of loss of generality, and possibily a higher implementation time, when considering general-purpose software such as MatCont [4] for comparison. However, the internal approach was also shown to be superior to that proposed in [90] (as well as in [35]), which features the same lack of generality.

Given the interest towards periodic solutions in applications of delay equations, extending the approach in order to continue non-steady solutions could be worth the effort. As anticipated at the end of Section 1.2, this is indeed the final goal as far as internal continuation is concerned, and this extension would somehow translate into merging the approach with the one described in Chapter 5 to compute periodic solutions.

Further (heuristic) improvements of both approaches can also be considered. Specifically, one could think of refining the relevant Newton's step by exploiting the band diagonal structure of the corresponding Jacobian matrix.

Finally, the work presented in Chapter 6 (as well as in Chapter 7) is an attempt to carry a convergence analysis of the FEM to compute periodic solutions of DDEs. It constitutes only the first step towards the final goal of proving the convergence of the method for coupled RE/DDE systems. In the case of DDEs, the method does not quite correspond to the (more natural) approach in [53] (described in Section 5.2), in that it involves the collocation of the function appearing on the left-hand side (i.e., the derivative, and not the solution itself). However, the latter approach is indeed the natural one when it comes to REs or neutral DDEs. In both cases, further substantial work would probably be needed, given the different smoothness requirements for the functional spaces involved in the analysis. Indeed, as it was clear at several points of the analysis presented, due to the need for differentiating with respect to parameters, such requirements are highly influenced by the role of the period ω, which is directly linked to the course of

time. Moreover, the problem will probably require fundamentally different Banach spaces.

The possibility of extending the method to DDEs with different type of delays (e.g., state-dependent) is also an open problem. Although, in this case, the relevant Banach spaces do not necessarily have to change, other issues may be encountered (recall the comment concerning (T5) before Assumption 6.1).

It is worth mentioning that verifying the hyperbolicity of the periodic orbit needed in Proposition 6.11 can be reduced to checking the eigenvalues of a certain finite-dimensional characteristic matrix of the periodic orbit (see [94, 97]).

The main result of this Chapter is the $O(L^{-m})$ convergence order for FEM, for a fixed m and an increasing number L of mesh intervals. Note, however, that the whole analysis can also be performed by considering different choices for the discretization, for instance nonuniform (adaptive) outer meshes. In this case, the convergence order would be $O(h^m)$, where h is the (vanishing) size of the largest mesh interval.

It might be worth to make some final observations concerning formulations (6.2) and (6.3). Although the two formulations are formally different, they lead to fundamentally equivalent numerical methods. In fact, when discretizing the problem (6.2) one just introduces redundant variables. Thus, it is reasonable to conjecture that a theoretical convergence analysis can be also carried out using formulation (6.3), using a different choice for the fixed point problem.

BIBLIOGRAPHY

[1] AUTO. http://indy.cs.concordia.ca/auto/.

[2] DDE-BIFTOOL. http://ddebiftool.sourceforge.net/.

[3] Knut. http://rs1909.github.io/knut/.

[4] MatCont. https://sourceforge.net/projects/matcont/.

[5] PyDSTool. https://pypi.org/project/PyDSTool/.

[6] XPPAUT. http://www.math.pitt.edu/~bard/xpp/xpp.html.

[7] E. L. Allgower and K. Georg. *Introduction to Numerical Continuation Methods*, volume 45 of *Classics in Applied Mathematics*. SIAM, Philadelphia, 2003.

[8] A. Ambrosetti and G. Prodi. *A primer of nonlinear analysis*, volume 34 of *Cambridge Studies in Advanced Mathematics*. Cambridge Univeristy Press, New York, 1995.

[9] A. Andò and D. Breda. Collocation techniques for structured populations modeled by delay equations. In M. Aguiar, C. Brauman, B. Kooi, A. Pugliese, N. Stollenwerk, and E. Venturino, editors, *Current Trends in Dynamical Systems in Biology and Natural Sciences*, SEPA SIMAI series. Springer, 2019. To appear.

[10] A. Andò and D. Breda. Convergence analysis of collocation methods for computing periodic solutions of retarded functional differential equations. 2019. Submitted.

[11] A. Andò, D. Breda, and S. Scarabel. Numerical continuation and delay equations: A novel approach for complex models of structured populations. *Discrete Contin. Dyn. S. Ser. S*, 2019. DOI: 110.3934/dcdss.2020165.

[12] N. Azbelev, V. P. Maksimov, and L. F. Rakhmatullina. *Introduction to the theory of functional differential equations: methods and applications*. Number 3 in Contemporary Mathematics and Its Applications. Hindawi Publishing Corporation, 2007.

[13] G. Bader. Solving boundary value problems for functional-differential equations by collocation. In *Numerical boundary value ODEs (Vancouver, B.C., 1984)*, volume 5, pages 227–243. Birkhäuser, 1985.

[14] D. A. W. Barton, B. Krauskopf, and R. E. Wilson. Collocation schemes for periodic solutions of neutral delay differential equations. *J. Differ. Equ. Appl.*, 12(11):1087–1101, 2006.

[15] A. Bellen. Monotone methods for periodic solutions of second order scalar functional differential equations. *Numer. Math.*, 42:15–30, 1983.

[16] A. Bellen. A Runge-Kutta-Nystrom method for delay differential equations. *In: U.M. Ascher, R.D. Russell (eds.) Numerical boundary values ODEs (Vancouver, B.C., 1984),* Progr. Sci. Comput., 5:271–283, 1985.

[17] A. Bellen and M. Zennaro. A collocation method for boundary value problems of differential equations with functional arguments. *Computing*, pages 307–318, 1984.

[18] A. Bellen and M. Zennaro. *Numerical methods for delay differential equations*. Numerical Mathemathics and Scientifing Computing series. Oxford University Press, 2003.

[19] J.-P. Berrut and L. N. Trefethen. Barycentric Lagrange interpolation. *SIAM Rev.*, 46(3):501–517, 2004.

[20] W.-J. Beyn and E. Doedel. Stability and multiplicity of solutions to discretizations of nonlinear ordinary differential equations. *SIAM J. Sci. Stat. Comput.*, 2(1):107–120, 1981.

[21] D. Breda, O. Diekmann, W. de Graaf, A. Pugliese, and R. Vermiglio. On the formulation of epidemic models (an appraisal of Kermack and McKendrick). *J. Biol. Dyn.*, 6(2):103–117, 2012.

[22] D. Breda, O. Diekmann, M. Gyllenberg, F. Scarabel, and R. Vermiglio. Pseudospectral discretization of nonlinear delay equations: new prospects for numerical bifurcation analysis. *SIAM J. Appl. Dyn. Sys.*, 15(1):1–23, 2016.

[23] D. Breda, O. Diekmann, D. Liessi, and F. Scarabel. Numerical bifurcation analysis of a class of nonlinear renewal equations. *Electron. J. Qual. Theory Differ. Equ.*, 65:1–24, 2016.

[24] D. Breda, O. Diekmann, S. Maset, and R. Vermiglio. A numerical approach for investigating the stability of equilibria for structured population models. *J. Biol. Dyn.*, 7(1):4–20, 2013.

[25] D. Breda, P. Getto, J. Sánchez Sanz, and R. Vermiglio. Computing the eigenvalues of realistic Daphnia models by pseudospectral methods. *SIAM J. Sci. Comput.*, 37(6):2607–2629, 2015.

[26] D. Breda and D. Liessi. Approximation of eigenvalues of evolution operators for linear renewal equations. *SIAM J. Numer. Anal.*, 56(3):1456–1481, 2018.

[27] D. Breda and D. Liessi. Floquet theory and stability of periodic solutions of renewal equations. *J Dyn Diff Equat*, 2020. DOI: 10.1007/s10884-020-09826-7.

[28] D. Breda, S. Maset, and R. Vermiglio. Pseudospectral differencing methods for characteristic roots of delay differential equations. *SIAM J. Sci. Comp.*, 27(3):482–495, 2005.

[29] D. Breda, S. Maset, and R. Vermiglio. Approximation of eigenvalues of evolution operators for linear retarded functional differential equations. *SIAM J. Numer. Anal.*, 50(3):1456–1483, 2012.

[30] D. Breda, S. Maset, and R. Vermiglio. *Stability of linear delay differential equations – A numerical approach with MATLAB*. SpringerBriefs in Control, Automation and Robotics. Springer, New York, 2015.

[31] C. G. Broyden. A class of methods for solving nonlinear simultaneous equations. *Math. Comp.*, 19(92):577–593, 1965.

[32] H. Brunner. *Collocation methods for Volterra integral and related functional differential equations*. Number 15 in Cambridge monographs on applied and computational mathematics. Cambridge University Press, Cambridge, 2004.

[33] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. Zang. *Spectral Methods. Evolution to Complex Geometries and Applications to Fluid Dynamics*. Scientific Computation Series. Springer, Berlin, Germany, 2007.

[34] F. Chatelin. *Spectral approximation of linear operators*. Classics in Applied Mathematics. SIAM, New York, 2011.

[35] A. M. de Roos. PSPManalysis. https://cran.r-project.org/package=PSPManalysis.

[36] A. M. de Roos. Numerical methods for structured population models: the escalator boxcar train. *Numer. Methods Partial Differential Equations*, 4(3):173–195, 1988.

[37] A. M. de Roos. A gentle introduction to models of physiologically structured populations. In S. Tuljapurkar and H. Caswell, editors, *Structured-population models in marine, terrestrial and freshwater systems*, pages 119–204, New York, 1997. Chapman and Hall.

[38] A. M. de Roos, O. Diekmann, P. Getto, and M. A. Kirkilionis. Numerical equilibrium analysis for structured consumer resource models. *B. Math. Biol.*, 72:259–297, 2010.

[39] A. M. de Roos, J. A. J. Metz, E. Evers, and A. Leipoldt. A size-dependent predator prey interaction: who pursues whom? *J. Math. Biol.*, 28:609–643, 1990.

[40] O. Diekmann, P. Getto, and M. Gyllenberg. Stability and bifurcation analysis of Volterra functional equations in the light of suns and stars. *SIAM J. Math. Anal.*, 39(4):1023–1069, 2008.

[41] O. Diekmann, P. Getto, and Y. Nakata. On the characteristic equation $\lambda = \alpha_1 + (\alpha_2 + \alpha_3\lambda)e^{-\lambda}$ and its use in the context of a cell population model. *J. Math. Biol.*, 72:877–908, 2016.

[42] O. Diekmann, M. Gyllenberg, H. Huang, M. Kirkilionis, J. A. J. Metz, and H. R. Thieme. On the formulation and analysis of general deterministic structured population models. II. nonlinear theory. *J. Math. Biol.*, 43:157–189, 2001.

[43] O. Diekmann, M. Gyllenberg, J. A. J. Metz, S. Nakaoka, and A. M. de Roos. *Daphnia* revisited: local stability and bifurcation theory for

physiologically structured population models explained by way of an example. *J. Math. Biol.*, 61(2):277–318, 2010.

[44] O. Diekmann, M. Gyllenberg, J. A. J. Metz, and H. R. Thieme. On the formulation and analysis of general deterministic structured population models. I. linear theory. *J. Math. Biol.*, 36:349–388, 1998.

[45] O. Diekmann, S. A. van Gils, S. M. Verduyn Lunel, and H.-O. Walther. *Delay Equations – Functional, Complex and Nonlinear Analysis*. Number 110 in Applied Mathematical Sciences. Springer Verlag, New York, 1995.

[46] E. Doedel. Lecture notes on numerical analysis of nonlinear equations. In H. M. Osinga, B. Krauskopf, and J. Galán-Vioque, editors, *Numerical continuation methods for dynamical systems*, Understanding Complex Systems, pages 1–49. Springer, 2007.

[47] A. d'Onofrio and P. Manfredi. *Modeling the Interplay Between Human Behavior and the Spread of Infectious Diseases*. Science and Business Media. Springer, Berlin, 2013.

[48] A. d'Onofrio, P. Manfredi, and E. Salinelli. Vaccinating behaviour, information, and the dynamics of SIR vaccine preventable diseases. *Theoret. Population Biol.*, 71:301–317, 2007.

[49] J. R. Dormand and P. J. Prince. A family of embedded Runge-Kutta formulae. *J. Comput. Appl. Math.*, 6(1):19–26, 1980.

[50] H. Ehlich and K. Zeller. Auswertung der normen von interpolationsoperatoren. *Math. Ann.*, 164:105–112, 1966.

[51] K. Engel and R. Nagel. *One-Parameter Semigroups for Linear Evolution Equations*. Number 194 in Graduate Texts in Mathematics. Springer-Verlag, New York, 1999.

[52] K. Engelborghs and E. Doedel. Stability of piecewise polynomial collocation for computing periodic solutions of delay differential equations. *Numer. Math.*, 91(4):627–648, 2002.

[53] K. Engelborghs, T. Luzyanina, K. J. in 't Hout, and D. Roose. Collocation methods for the computation of periodic solutions of delay differential equations. *SIAM J. Sci. Comput.*, 22(5):1593–1609, 2001.

[54] K. Engelborghs, T. Luzyanina, and D. Roose. Numerical bifurcation analysis of delay differential equations. *J. Comput. Appl. Math.*, 125(1-2):265–275, 2000.

[55] B. Ermentrout. *Simulating, Analyzing, and Animating Dynamical Systems – A Guide to XPPAUT for Researchers and Students*. Software - Environment - Tools series. SIAM, Philadelphia, 2002.

[56] G. Faber. Über die interpolatorische darstellung stetiger funktionen. *Jahresber. Deut. Math. Verein.*, 23:192–210, 1914.

[57] D. Gottlieb and S. Orszag. *Numerical analysis of spectral methods: theory and applications*, volume 26 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. SIAM, Philadelphia, 1977.

[58] W. J. F. Govaerts. *Numerical methods for bifurcations of dynamical equilibria*. SIAM, Philadelphia, 2000.

[59] G. Gripenberg, S.-O. Londen, and O. Staffans. *Volterra integral and functional equations*. Encyclopedia of Mathematics and its Applications. Cambridge University Press, Cambridge, first edition, 2009.

[60] E. Hairer, S. P. Nörsett, and G. Wanner. *Solving ordinary differential equations I: nonstiff problems*. Number 8 in Computational Mathematics. Springer-Verlag, Berlin, 1987.

[61] J. K. Hale. *Theory of functional differential equations*. Number 99 in Applied Mathematical Sciences. Springer Verlag, New York, first edition, 1977.

[62] E. I. Herbert Keller. *Analysis of Numerical Methods*. Dover, 1966.

[63] G. E. Hutchinson. Circular causal systems in ecology. *Ann. N.Y. Acad. Sci.*, 50:221–246, 1948.

[64] H. Inaba. *Age-Structured Population Dynamics in Demography and Epidemiology*. Springer, New York, 2017.

[65] P. Jagers. The deterministic evolution of general branching populations. In M. de Gunst, C. Klaassen, and A. van der Vaart, editors, *State of the art in probability and statistics*, volume 36 of *Lecture Notes–Monograph Series*. Institute of Mathematical Statistics, 2001.

[66] E. Jones, T. Oliphant, P. Peterson, et al. *SciPy: Open source scientific tools for Python*, 2001.

[67] D. Kincaid and W. Cheney. *Numerical Analysis - Mathematics of Scientific Computing*. Number 2 in Pure and Applied Undergraduate Texts. American Mathematical Society, Providence, 2002.

[68] V. B. Kolmanovskii and A. Myshkis. *Introduction to the Theory and Applications of Functional Differential Equations*. Number 463 in Mathematics and Its Applications. Springer Netherlands, The Netherlands, 1 edition, 1999.

[69] M. A. Krasnosel'skii, G. M. Vainikko, R. P. Zabreyko, R. Y. B, and V. V. Stet'senko. *Approximate Solution of Operator Equations*. Springer Netherlands, 1 edition, 1972.

[70] R. Kress. *Linear integral equations*. Number 82 in Applied Mathematical Sciences. Springer-Verlag, New York, 1989.

[71] Y. Kuang. *Delay Differential Equations: With Applications in Population Dynamics*. Number 191 in Dynamics in Science and Engineering. Academic Press, New York, 1993.

[72] Y. A. Kuznetsov. *Elements of applied bifurcation theory*. Number 112 in Applied Mathematical Sciences. Springer-Verlag, New York, second edition, 1998.

[73] W. M. Liu. Criterion of Hopf bifurcations without using eigenvalues. *J. Math. Anal. Appl.*, 182:250–256, 1994.

[74] T. Luzyanina, K. Engelborghs, K. Lust, and D. Roose. Computation, continuation and bifurcation analysis of periodic solutions of delay differential equations. *Internat. J. Bifur. Chaos Appl. Sci. Engrg.*, 7(11):2547–2560, 1997.

[75] A. M. Lyapunov. *The general problem of the stability of motion*. Number 7 in Collected Works II (in Russian). Kharkov Mathematical Society, 1892. Republished by the University of Tolouse 1908 and Princeton University Press 1949 (in French), in Int. J. Control 55, 531-773 and as a book by Taylor & Francis, London, 1992 (in English).

[76] T. R. Malthus. *An essay on the principle of population*. J. Johnson, London, 1798.

[77] S. Maset. The collocation method in the numerical solution of boundary value problems for neutral functional differential equations. Part I: Convergence results. *SIAM J Numer. Anal.*, 53(6):2771–2793, 2015.

[78] S. Maset. The collocation method in the numerical solution of boundary value problems for neutral functional differential equations. Part II: Differential equations with deviating arguments. *SIAM J Numer. Anal.*, 53(6):2794–2821, 2015.

[79] S. Maset. An abstract framework in the numerical solution of boundary value problems for neutral functional differential equations. *Numer. Math.*, 133(3):525–555, 2016.

[80] G. Mastroianni and G. Milovanovic. *Interpolation Processes - Basic Theory and Applications*. Springer Monographs in Mathematics. Springer-Verlag, Heidelberg, 2008.

[81] G. Mastroianni and D. Occorsio. Optimal systems of nodes for Lagrange interpolation on bounded intervals. a survey. *J. Comput. Appl. Math.*, 134:325–341, 2001.

[82] H. Metz and O. Diekmann. *The dynamics of physiologically structured populations*. Number 68 in Lecture Notes in Biomathematics. Springer-Verlag, New York, 1986.

[83] J. M. Ortega. *Numerical Analysis: a second course*, volume 3 of *Classic Applied Mathematics*. Society for Industrial and Applied Mathematics, 1990.

[84] L. Petzold. Automatic selection of methods for solving stiff and nonstiff systems of ordinary differential equations. *SIAM J. Sci. Stat. Comput.*, 4(1):136–148, 1983.

[85] M. J. D. Powell. A hybrid method for nonlinear equations. In P. Rabinowitz, editor, *Numerical Methods for Nonlinear Algebraic Equations*. Gordon and Breach, 1970.

[86] G. W. Reddien and C. C. Travis. Approximation methods for boundary value problems of differential equations with functional arguments. *J. Math. Anal. Appl.*, 46:62–74, 1974.

[87] T. Rivlin. *An introduction to the approximation of functions*. Blaisdell, Waltham, 1969.

[88] W. Rudin. *Principles of mathematical analysis*. 3rd ed. International Series in Pure and Applied Mathematics. McGraw Hill, 1976.

[89] R. D. Russell and T. Christiansen. Adaptive mesh selection strategies for solving boundary value problems. *SIAM J. on Numer. Anal.*, 15(1):59–80, February 1978.

[90] J. Sánchez Sanz and P. Getto. Numerical bifurcation analysis of physiologically structured populations: Consumer-resource, cannibalistic and trophic models. *B. Math. Biol.*, 78(7):1546–84, 2016.

[91] V. M. Shurenkov. On the theory of Markov renewal. *Theory Probab. Appl.*, 29:247–265, 1984.

[92] J. Sieber. Finding periodic orbits in state-dependent delay differential equations as roots of algebraic equations. *Discrete Contin. Dyn. S. Ser. S*, 32(8):2607–2561, 2012.

[93] J. Sieber, K. Engelborghs, T. Luzyanina, G. Samaey, and D. Roose. Ddebiftool manual - bifurcation analysis of delay differential equations, 2014.

[94] J. Sieber and R. Szalai. Characteristic matrices for linear periodic delay differential equations. *SIAM J. Appl. Dyn. Sys.*, 10(1):129–147, 2011.

[95] H. L. Smith. *An introduction to delay differential equations with applications to the life sciences*. Number 57 in Texts in Applied Mathematics. Springer, New York, 2011.

[96] M. Spivak. *Calculus*. Cambridge University Press, third edition, 1994.

[97] R. Szalai, G. Stépán, and S. J. Hogan. Continuation of bifurcations in periodic delay-differential equations using characteristic matrices. *SIAM J. Sci. Comput.*, 28(4):1301–1317, 2006.

[98] L. N. Trefethen. *Spectral methods in MATLAB*. Software - Environment - Tools series. SIAM, Philadelphia, 2000.

[99] L. N. Trefethen. Is Gauss quadrature better than Clenshaw-Curtis? *SIAM Rev.*, 50(1):67–87, 2008.

[100] L. N. Trefethen. *Approximation theory and approximation practice*. Number 128 in Other Titles in Applied Mathematics. SIAM, Philadelphia, 2013.

[101] K. Verheyden and K. Lust. A Newton-Picard collocation method for periodic solutions of delay differential equations. *BIT*, 45(3):605–625, 2005.

[102] P. F. Verhulst. *Notice sur la loi que la population poursuit dans son accroissement*. Number 10 in Correspondence Mathématique et Physique. A. Quetelet, Bruxelles, 1838.

[103] V. Volterra. Sur la théorie mathématique des phénomènes héréditaires. *J. Math. Pures Appl.*, 7:249–298, 1928.

[104] J. A. Weideman and S. C. Reddy. A MATLAB differentiation matrix suite. *ACM T. Math. Software*, 26(4):465–519, 2000.