

---

# LEARNING RUNGE-KUTTA INTEGRATION SCHEMES FOR ODE SIMULATION AND IDENTIFICATION

---

A PREPRINT

Said Ouala<sup>1</sup>, Laurent Debreu<sup>2</sup>, Bertrand Chapron<sup>3</sup>  
Ananda Pascual<sup>4</sup>, Fabrice Collard<sup>5</sup>, Lucile Gaultier<sup>5</sup>, Ronan Fablet<sup>1</sup>

(1) IMT Atlantique; Lab-STICC, 29200 Brest, France

(2) AIRSEA INRIA GRA, 38000 Grenoble, France

(3) Ifremer, LOPS, 29200 Brest, France

(4) IMEDEA, UIB-CSIC, 07190 Esporles, Spain

(5) ODL, 29200 Brest, France

{said.ouala, ronan.fablet}@imt-atlantique.fr

Laurent.Debreu@inria.fr, Bertrand.Chapron@ifremer.fr,

ananda.pascual@imedea.uib-csic.es

{dr.fab, lucile.gaultier}@oceandatalab.com

May 12, 2021

## ABSTRACT

Deriving analytical solutions of ordinary differential equations is usually restricted to a small subset of problems and numerical techniques are considered. Inevitably, a numerical simulation of a differential equation will then always be distinct from a true analytical solution. An efficient integration scheme shall further not only provide a trajectory throughout a given state, but also be derived to ensure the generated simulation to be close to the analytical one. Consequently, several integration schemes were developed for different classes of differential equations. Unfortunately, when considering the integration of complex non-linear systems, as well as the identification of non-linear equations from data, this choice of the integration scheme is often far from being trivial. In this paper, we propose a novel framework to learn integration schemes that minimize an integration-related cost function. We demonstrate the relevance of the proposed learning-based approach for non-linear equations and include a quantitative analysis w.r.t. classical state-of-the-art integration techniques, especially where the latter may not apply.

**Keywords** Dynamical systems · Numerical integration · Automatic differentiation · Runge-Kutta · Data-driven models · Neural ODE

## 1 Introduction

Only linear differential equations and a restricted class of non-linear ones have closed-form analytical solutions. Therefore, for most real-world applications, the resolution of differential equations relies on numerical integration algorithms to derive an approximation of the true solution of a given equation.

From a mathematical point of view, applying a numerical integration scheme to a differential equation corresponds to map a continuous time-varying equation to a discrete one Parker and Chua [1989]. Consequently, the properties of the simulated solution will depend on the properties of both the continuous equation and the integration scheme. It is however extremely important for the properties of the numerical approximation (typically its asymptotic behaviour) to match the true solution according to some given error and stability criteria.

A large literature discusses such error and stability criteria Thomas [1995] to assess the applicability of an integration scheme for the numerical resolution of a differential equation Friedmann et al. [1977], Berland et al. [2006], Ruuth [2006] with respect to a given discretization and an error tolerance. However, these criteria are most of the time built

on linear equations leading to little to no guarantees when considering their behaviour on non-linear systems Parker and Chua [1989]. In this situation, the integration error depends non-linearly on the state of the system in the phase space which makes the prior characterization of the dynamics not as straightforward as in the linear case (since it becomes equation dependent). Some works developed stability criteria for specific classes of non-linear equations (e.g., B-stability for instance in the case of monotonic equations Hairer and Wanner [1996a]), however, these criteria are usually restricted to a limited class of equations. As a consequence, integration methods are, in practice, often used as black-box routines that simulate a given equation and achieve a given performance criterion.

From this point of view, new black-boxes, written in the language of data-driven models have quickly joined the arsenal of numerical integration techniques both for the resolution of Ordinary Differential Equations (ODEs) and Partial Differential Equations (PDEs). The success of such techniques is mainly due to their ability to match sufficiently well a performance criterion given a sufficient amount of training data. The main investigations of such approaches relate to the exploration of neural solvers for the resolution of ODEs and PDEs Stiller et al. [2020], KHOO et al. [2020], Li et al. [2020] and the derivation of some kind of an optimal spatial discretization for the resolution of PDEs Suzuki [2019], Bar-Sinai et al. [2019]. Unfortunately, little to no work have been investigated when considering time integration techniques and their role and potential improvements when considering the implementation, resolution and identification of differential equations in a data-driven fashion.

Here, we address the design of time integration schemes as a machine learning problem. We focus on Ordinary Differential Equations (ODEs) and show that one can train numerically data-driven or application-specific integration schemes, possibly jointly to data-driven representations in identification scenarios Wang and Lin [1998], Brunton et al. [2016], Fablet et al. [2017]. Focusing on an explicit Runge-Kutta formulation, the proposed approach relies on the exploitation of the automatic differentiation Bückner et al. [2006] of the computational graphs associated with integration schemes. We demonstrate its relevance for the integration and data-driven identification of non-linear dynamics and assess the properties of the trained integration schemes. One of the main findings of this work is the ability of such methodology to output application-adapted integration schemes without resorting to a complex analysis of the underlying equations. Overall, our key contributions comprise:

- the definition of a generic workflow for integration-related problems where the integration scheme is defined as a solution of an optimization problem;
- the data-driven identification of computationally-efficient integration schemes for the time integration of non-linear differential equations;
- the data-driven identification of governing laws stated as the joint learning of an approximate ODE and of the associated integration scheme;
- the evaluation of the key features of the trained integration schemes w.r.t. classical state-of-the-art integration techniques, especially through a study of precision, stability and generalization properties for configurations that were never seen during the training phase.

The paper is organized as follows. In Section 2, we briefly review state-of-the-art integration techniques and the associated performance criteria. Section 3 presents the proposed framework, followed by the experiments and results in Section 4 and 5. We close this paper with conclusions and perspectives for future works in Section 6.

## 2 Integration methods and performance criteria

Let us assume a continuous  $s$ -dimensional dynamical system  $\mathbf{z}_t$  governed by the following non-autonomous time varying ODE

$$\dot{\mathbf{z}}_t = f(t, \mathbf{z}_t) \quad (1)$$

Assuming that, given an initial condition  $\mathbf{z}_{t_0}$ , we aim to solve this equation for an interval  $t \in [t_0, t_f]$ , the corresponding solution (or flow) can be written as

$$\Phi_t(\mathbf{z}_{t_0}) = \mathbf{z}_{t_0} + \int_{t_0}^t f(w, \mathbf{z}_w) dw \quad (2)$$

As stated in the introduction, solving the flow integral  $\int_{t_0}^t f(w, \mathbf{z}_w) dw$  is only possible for a small subset of non-linear ODEs. Simulating the differential equation is then done using numerical integration methods. Formally, the interval  $t \in [t_0, t_f]$  is discretized using a time-step  $h > 0$  as  $h = \frac{t_f - t_0}{N}$  and  $t_n = t_0 + nh$ , where  $0 \leq n \leq N$  an integer and  $N$  is the number of grid points, the problem is then formulated as the approximation of the values of variable  $\mathbf{z}_t$  at each grid point given the initial condition  $\mathbf{z}_0$  as follows:

$$\begin{cases} \hat{\mathbf{z}}_{t_0} = \mathbf{z}_{t_0} = \mathbf{z}_0 \\ \hat{\mathbf{z}}_{t_{n+1}} = \Phi_{\mathcal{D}, t_{n+1}}(\hat{\mathbf{z}}_{t_n}) \approx \mathbf{z}_{t_{n+1}} = \Phi_{t_{n+1}}(\mathbf{z}_{t_n}) \end{cases} \quad (3)$$

with  $\hat{\mathbf{z}}$  a numerical solution computed using the approximation  $\Phi_{\mathcal{D},t_n}$  of the analytical solution (2) and  $\mathcal{D}$  is a given integration scheme.

In this section we will briefly introduce state-of-the-art integration techniques and the associated performance and stability criteria. We focus on single-step explicit schemes as they are one of the most used techniques when considering non-stiff problems.

## 2.1 Single-step numerical integration

Single-step integration schemes can be roughly divided into two main categories: explicit and implicit integration schemes Parker and Chua [1989]. Implicit schemes allow larger stability intervals than their explicit counterparts Zhang et al. [2016] which make them extremely valuable when considering stiff problems Wanner [1977], Pareschi and Russo [2000]. By contrast, explicit integration schemes are straightforward to implement and are generally applied whenever stiffness is not detected within a system.

Single-step explicit integration schemes use a single evaluation of the state of the system at a given grid point to compute the approximation of the solution at the next grid point. Their general form can be written as follows:

$$\hat{\mathbf{z}}_{t_{n+1}} = \Phi_{\mathcal{D},t_{n+1}}(\hat{\mathbf{z}}_{t_n}) = \hat{\mathbf{z}}_{t_n} + h\Psi_{\mathcal{D}}(t_n, \hat{\mathbf{z}}_{t_n}, h) \quad (4)$$

with  $\Psi_{\mathcal{D}}$  an approximation operator related to a truncation of the Taylor expansion of the true solution. We may remind the exact Taylor expansion of the solution of (2) given the true state  $\mathbf{z}_{t_n}$  as:

$$\mathbf{z}_{t_{n+1}} = \mathbf{z}_{t_n} + \sum_{k=1}^{p=+\infty} h^k \frac{1}{k!} f^{k-1}(t_n, \mathbf{z}_{t_n}) \quad (5)$$

with  $f^k$  the  $k^{th}$ -order derivative of ODE operator  $f$ .

The Forward Euler scheme and Runge-Kutta methods are among the most widely used single-step explicit schemes:

- **Forward Euler scheme:** The Forward Euler algorithm may be viewed as a truncation of the Taylor series up to  $p = 1$

$$\hat{\mathbf{z}}_{t_{n+1}} = \Phi_{\mathcal{E},t_n}(\hat{\mathbf{z}}_{t_n}) = \hat{\mathbf{z}}_{t_n} + hf(t_n, \hat{\mathbf{z}}_{t_n}^T) \quad (6)$$

- **Explicit Runge-Kutta schemes:** Approximating the integral of (2) using a low-order truncation of its Taylor series as for instance shown in the Euler algorithm is a quite simple and intuitive approach for approximate solutions of ODEs. However, a simple forward Euler integration can lead to large integration errors and the acceptable integration time-step of such technique is usually small. Considering a high-order truncation usually leads to more accurate and/or stable results. Unfortunately, evaluating high order derivatives of (1) is computationally intensive.

Runge-Kutta methods Ixaru and Vanden Berghe [2004] rely on an iterative evaluation of the ODE function  $f$  in order to mimic high order error truncations of the Talyor expansion of a linear equation. A  $q$ -stage Runge-Kutta ( $\mathcal{RK}_q$ ) technique can be defined by its set of parameters as follows:

$$\begin{aligned} \mathcal{RK}_q &= \{A = [a_{i,j}] \in \mathbb{R}^{q \times q}, b = [b_i] \in \mathbb{R}^q, c = [c_i] \in \mathbb{R}^q \\ &\text{with } \sum_{j=1}^q a_{i,j} = c_i, 0 < c_i < 1 \text{ and } \sum_{i=1}^q b_i = 1\} \end{aligned} \quad (7)$$

Given a Runge-Kutta scheme  $\mathcal{RK}_q$ , the solution of problem (1) can be written as follows:

$$\hat{\mathbf{z}}_{t_{n+1}} = \Phi_{\mathcal{RK}_q,t_{n+1}}(\hat{\mathbf{z}}_{t_n}) = \hat{\mathbf{z}}_{t_n} + \sum_{i=1}^q b_i k_i \quad (8)$$

where  $\Phi_{\mathcal{RK}_q,t_{n+1}}(\hat{\mathbf{z}}_{t_n})$  is the approximate version of (2) written through the Runge-kutta scheme  $\mathcal{RK}_q$ . The coefficients  $k_i$  are given by:

$$k_i = f(t_n + c_i h, \hat{\mathbf{z}}_{t_n} + h(\sum_{j=1}^{i-1} a_{i,j} k_j)) \quad (9)$$

When  $q = 1$ , it simply corresponds to the explicit Euler method. For a given number of stages  $q$ , the coefficients of the Runge-Kutta method shall satisfy some extra conditions (by matching it to the corresponding Taylor

series) to reach a given order  $p$  Butcher [1963], Wanner and Hairer [1996]. Formally, the order  $p$  of the Runge-Kutta method is always inferior or equal to the number of stages  $q$ . For  $q = 4$ , we can retrieve the well-known Runge-Kutta-4 method, when  $p > 4$ , we need more integration stages  $q$  to truly reach a given error order  $p$  Wanner and Hairer [1996].

Explicit integration schemes approximate the solution at  $t_{n+1}$  using a forward Taylor expansion, resulting in an expression of  $\hat{\mathbf{z}}_{t_{n+1}}$  as a function of the previous steps (as illustrated for instance by (4)). Implicit schemes Cooper and Vignesvaran [1993a] in the other hand exploit a backward formulation, resulting in the following general formulation

$$\hat{\mathbf{z}}_{t_{n+1}} = \hat{\mathbf{z}}_{t_n} + h\Psi_{\mathcal{D}}(t_{n+1}, \hat{\mathbf{z}}_{t_n}, \hat{\mathbf{z}}_{t_{n+1}}, h) \quad (10)$$

Note that now, the right hand side of (10) also depends on  $\hat{\mathbf{z}}_{t_{n+1}}$ . The solution of this equation is not as straightforward as in the explicit case since, when considering a non-linear ODE, (10) becomes non-linear and solving for  $\hat{\mathbf{z}}_{t_{n+1}}$  requires using a non-linear (or linearized) solver Cooper and Butcher [1983]. For a Partial Differential Equation (PDE), it also requires the inversion of a potentially large system. Here, we focus on explicit schemes. We further discuss in Section 6 possible applications to explicit schemes.

## 2.2 Integration Errors

Integration errors can be seen as the difference between the true analytical solution and the output of the discrete equation at the same time-step (or steps, depending on whether we are considering local or global errors). We may distinguish round-off errors and truncation errors.

The Round-off error is simply the error resulting from using finite-precision arithmetics. This error is inversely proportional to the integration step as a larger integration step would mean a smaller number of time-steps to be computed and stored Parker and Chua [1989].

The local truncation error is the error associated with the numerical integration scheme on a single integration time-step<sup>1</sup> Parker and Chua [1989]. This error streams from the Taylor expansion as follows. Given the true state at time  $t_n$ , a  $p$ -order integration algorithm<sup>2</sup> leads, from the Taylor expansion of the solution at  $t_{n+1}$  given by (5), to the following integration error, referred to as the truncation error  $\epsilon_n$  Parker and Chua [1989]

$$\epsilon_n = E(p, f, \mathbf{z}_{t_n}, t_n)h^{p+1} \quad (11)$$

where  $E \in \mathbb{R}$  depends on  $p, f, \mathbf{z}$  and  $t_n$ . The global truncation error is the summation of the local truncation errors for the  $N$  grid points. Supposing that the initial condition of each integration step is perfect, it can be computed as  $\epsilon^g = \sum_{n=0}^N \epsilon_n$ .

We may discuss the truncation error equation. First of all, the truncation error is, for a given numerical integration scheme, proportional to the integration time-step  $h$  at a given order. This means that any integration algorithm will produce a larger error as  $h$  grows. Furthermore, if the integration time-step is arbitrarily small, high-order algorithms (larger  $p$ ) will achieve a smaller truncation error than low-order algorithms. It is important to note that this statement is only true for small-enough integration time-steps as when  $h$  is big enough high-order methods will have higher integration errors due to the term  $h^{p+1}$ . These considerations can be illustrated on a simple linear ODE. Let us assume that  $\mathbf{z} \in \mathbb{R}$  and  $f(t_n, \mathbf{z}_{t_n}) = \lambda \mathbf{z}_{t_n}$  with  $\lambda < 0$ . The solution of this dissipative equation is a stable equilibrium point at zero and the corresponding truncation error of a given  $p$ -order numerical integration scheme becomes

$$\epsilon_t = \frac{(\lambda h)^{p+1}}{(p+1)!} \mathbf{z}_{t_n} + O(h^{p+2}) \quad (12)$$

Figure 1 depicts a graphical representation of (12). As stated above, for an arbitrarily small integration time-step (and similarly for an arbitrarily slow time-scale  $\lambda$ ), the higher the order of the integration scheme, the smaller its truncation error. However, and as counter-intuitive as it may sound, above a certain threshold for the values of  $(\lambda h)$ , lower-order integration schemes may perform better than higher-order ones. This is typically due to the term  $(\lambda h)^{p+1}$ . These considerations suggest that choosing an adapted integration scheme for a given application is not an easy task especially when constrained by the integration time-step. In practice, complex systems involve different values of  $\lambda$ , which relate to different characteristic time-scales. The choice of  $h$  and of the order of the time integration scheme may then be difficult to optimize. One may also notice that the truncation error equation, in the linear case, depends on the state  $\mathbf{z}_{t_n}$ . In the non-linear case, the error rather depends on the  $p^{th}$  derivative of the vector field  $f(t_n, \mathbf{z}_{t_n})$ , which generally requires further equation-dependent analyses.

<sup>1</sup>Please note that the truncation error is derived assuming perfect knowledge of the initial conditions along the temporal grid.

<sup>2</sup>The order of an integration algorithm is determined by matching its Taylor expansion to the Taylor expansion of the true solution of a linear equation. The value of  $p$  up to which these two series match and they differ afterwards defines the order of the integration scheme. We may note that this definition implicitly involves a matching criterion which might involve some tolerance or decision threshold.

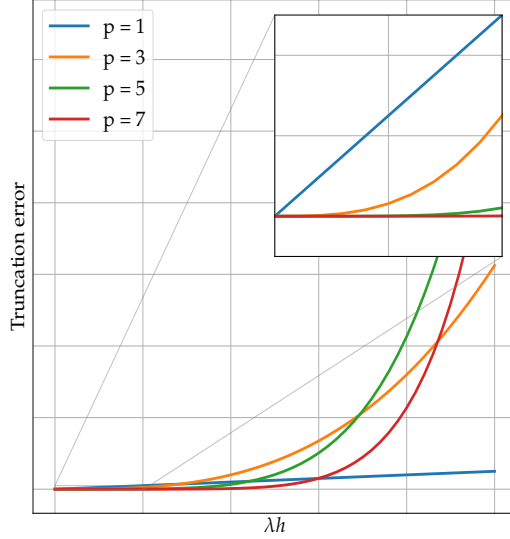


Figure 1: *Truncation error of single-step integration schemes.* The truncation error of an integration scheme when integrating the linear equation  $\dot{\mathbf{z}}_t = \lambda \mathbf{z}_t$ . This error was computed using (12) for several orders of single-step integration schemes.

### 2.3 Integration Stability

The characterization of local and global errors provides a relevant framework to assess about the performance of a numerical integration scheme. However, and since they assume perfect knowledge of the initial condition at each time-step, they cannot be used to assess about the convergence properties of an integration scheme when used to simulate a trajectory of an ODE. In practice, local errors accumulate from a time-step to an other and may become unbounded making the integration scheme unstable.

The analysis of the stability of an integration technique usually relies on the study of the numerical solution issued from its application to a first-order linear equation

$$\dot{\mathbf{z}}_t = \lambda \mathbf{z}_t, \quad \mathbf{z}_{t_0} = \mathbf{z}_0 \quad (13)$$

with  $\mathbf{z}_t, \lambda \in \mathbb{C}$  and  $\text{Real}(\lambda) \leq 0$ . Since the exact solution of this linear ODE is not growing in time, the corresponding numerical solution must have the same property. This requires stating the discrete system in the form

$$\begin{cases} \hat{\mathbf{z}}_{t_0} = \mathbf{z}_{t_0} = \mathbf{z}_0 \\ \hat{\mathbf{z}}_{t_{n+1}} = \mathcal{R}_D(\lambda h) \hat{\mathbf{z}}_{t_n} \end{cases} \quad (14)$$

where  $\mathcal{R}_D$  represents the stability function of the integration scheme. For a  $q$ -stage explicit Runge-Kutta scheme, the stability function is a polynomial of order less or equal to the number of stages  $q$ . It can be written as a function of the Runge-Kutta coefficients  $\mathcal{RK}_q$  given in (7) as follows Hairer and Wanner [1996b]:

$$\mathcal{R}_{\mathcal{RK}}(\lambda h) = \sum_{i=1}^q \alpha_i (\lambda h)^i = \mathbf{b}^T (I - (\lambda h)A)^{-1} \mathbf{1} \quad (15)$$

with  $\alpha_i$  the coefficients of the polynomial  $\mathcal{R}_{\mathcal{RK}}(\lambda h)$  and  $\mathbf{1}$  a vector of ones with an appropriate dimension.

The region of stability of  $\Phi_D$  is given by the values  $\lambda h$  for which  $|\mathcal{R}_D| \leq 1$ . We may point out that this widely used stability analysis is only valid for linear equations and does not guarantee the applicability of a given integration scheme for a non-linear ODE.

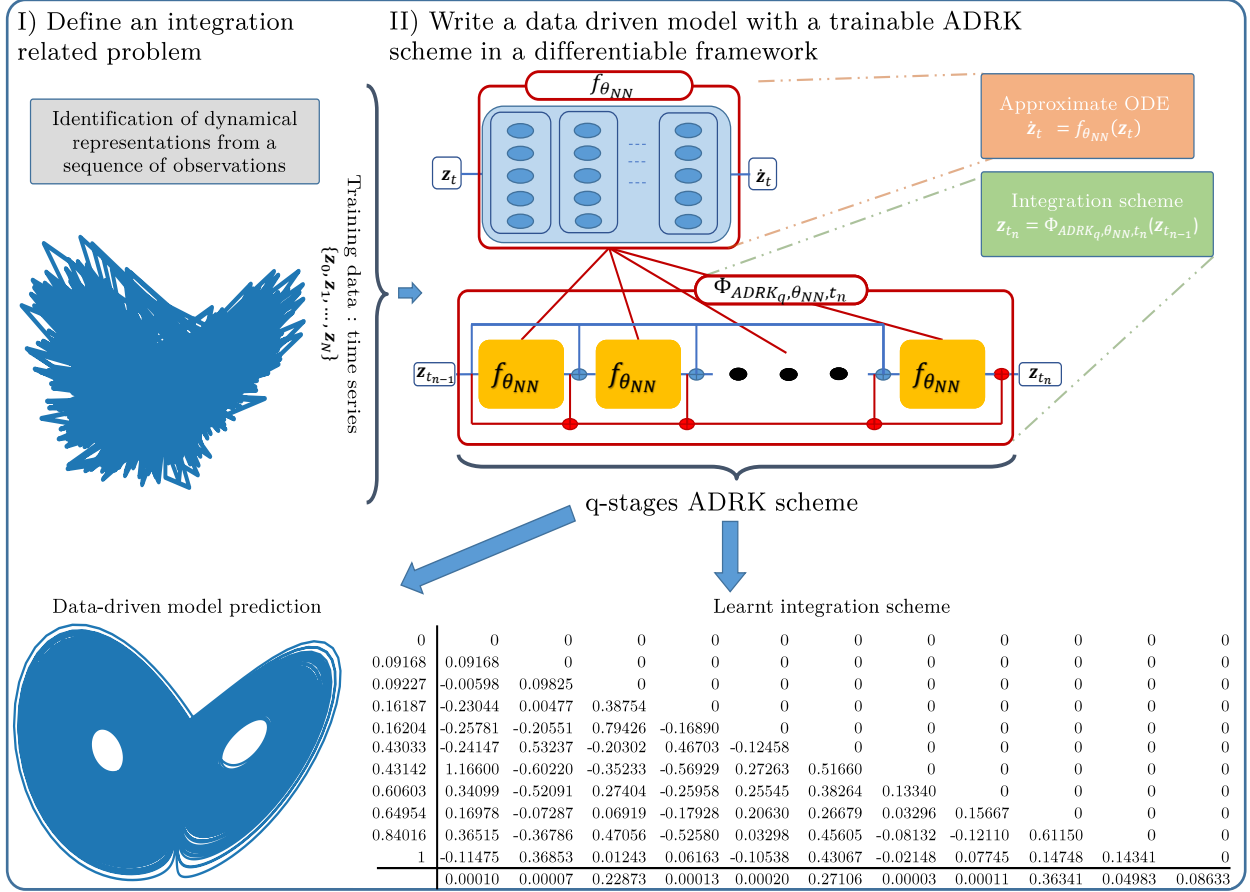


Figure 2: *Sketch of the proposed ADRK framework.* We highlight the main steps of the proposed framework. We first start by defining an integration related problem, the solution of this problem is written as a Runge-Kutta scheme in a differentiable framework. The parameters of the optimization problem are optimized with respect to a predefined cost. After optimization, the learnt integration scheme can be evaluated using classical stability/precision criterions.

### 3 Automatic Differentiation based Runge-Kutta (ADRK)

In this section, we introduce a trainable Runge-Kutta integration framework, referred to as Automatic-Differentiation-based Runge-Kutta (ADRK) scheme, to solve integration-related problems using learning strategies. We first introduce the proposed learning-based framework (Section 3.1). We then address three different applications: the identification of stability-constrained integration schemes (Section 3.2), the integration of ODEs (Section 3.3) and the data-driven identification of ODEs (Section 3.4).

#### 3.1 Problem statement

In this section, we introduce trainable Runge-Kutta integration schemes framework to solve initial value problems (IVPs).

Let us recall the general form of an IVP governed by a non-autonomous ODE:

$$\begin{cases} \dot{\mathbf{z}}_t = f(t, \mathbf{z}_t) \\ \mathbf{z}_{t_0} = \mathbf{z}_0 \end{cases} \quad (16)$$

with  $\mathbf{z}_0$  is a given initial condition. We may also remind the analytical form of the solution, for an interval  $t \in [t_0, t_f]$ , as:

$$\Phi_t(\mathbf{z}_{t_0}) = \mathbf{z}_{t_0} + \int_{t_0}^t f(w, \mathbf{z}_w) dw \quad (17)$$

As stated in Section 2, Runge-Kutta schemes are among the most popular techniques for the numerical resolution of (17).

Given a number of stages  $q$ , the identification of the parameters of a Runge-Kutta scheme  $\mathcal{RK}_q$  given in (7) may derive from different approaches Butcher and Wanner [1996], Owren and Zennaro [1992]. Methods that explicitly enforce stability and/or precision (*i.e.* order) constraints on the integration of a linear equation are the most popular ones Ketcheson and Ahmadi [2013]. They rely on the polynomial form of the integration scheme (given for instance by (15)). Such a formulation provides little to no knowledge regarding the stability and/or the performance of the integration scheme on non-linear ODEs. Furthermore, it makes non trivial the choice of the integration scheme when considering identification applications, for which the ODE is not known.

Here, we adopt a machine learning perspective so that we identify the parameters of a  $q$ -stage Runge-Kutta scheme which lead to the minimization of a predefined performance criterion. We depict in Figure 2 a general sketch of the proposed framework, which involves the following key steps:

- We first state an integration-related problem using a trainable Runge-Kutta scheme with a predefined number of stages. This integration scheme is implemented in a computational framework that supports automatic differentiation Bückner et al. [2006] such as pytorch and TensorFlow. Formally, a trainable Runge-Kutta (ADRK) method is defined similarly to (7) by its Runge-Kutta coefficients as follows

$$\begin{aligned} \text{ADRK}_q &= \{A_{\text{ADRK}_q} = [a_{i,j}] \in \mathbb{R}^{q \times q}, b_{\text{ADRK}_q} = [b_i] \in \mathbb{R}^q, c_{\text{ADRK}_q} = [c_i] \in \mathbb{R}^q \\ &\text{with } \sum_{j=1}^q a_{i,j} = c_i, 0 < c_i < 1 \text{ and } \sum_{i=1}^q b_i = 1\} \end{aligned} \quad (18)$$

- In addition to the parameters of the trainable Runge-Kutta scheme, the considered problem may involve other trainable parameter sets such as unknown parameters of the differential equation. Let us denote by  $\nu$  these additional unknown and trainable parameters and  $\Theta = \{\text{ADRK}_q, \nu\}$  the set of all trainable parameters. The derivation of optimal parameters  $\hat{\Theta}$  comes to the following constrained optimization problem with respect to a problem-dependent criterion  $\mathcal{L}$ , referred to as the training loss in a deep learning formulation

$$\begin{aligned} \hat{\Theta} &= \arg \min_{\Theta} \mathcal{L}(\Theta) \\ &\text{subject to (18)} \end{aligned} \quad (19)$$

We will illustrate different problem-dependent definitions for the training loss in the following sections. We may point out that these definitions may involve some priors, such as stability constraints for Runge-Kutta schemes, as well as training data when considering a data-driven identification issues.

- Within the considered differentiable framework, we solve this constrained minimization problem using a Stochastic Gradient Descent (SGD) as follows:

$$\theta_{k+1}^i = \theta_k^i - \mu_i \frac{\partial \mathcal{L}(\Theta)}{\partial \theta^i} \quad (20)$$

with  $\mu_i$  a predefined learning rate and  $\theta^i \in \Theta$ . The constrained optimization for the parameters of the integration scheme is solved in a hard fashion by clipping the Runge-Kutta weights to exactly satisfy the Runge-Kutta constraints (18) after each training step  $k$ . At each iteration  $k$ , the SGD step typically applies to a subset of the training dataset, referred to as mini-batch in deep learning frameworks LeCun et al. [2015].

Deep learning frameworks also make highly flexible the definition of the considered training loss, which in turn provide new tools to derive integration schemes according to a variety of criteria as illustrated hereafter.

### 3.2 Stability-constrained ADRK scheme

As first case-study of the proposed ADRK framework, we address the derivation of stable Runge-Kutta schemes Ketcheson and Ahmadi [2013]. Let us consider the following linear ODE:

$$\begin{cases} \dot{\mathbf{z}}_t = \lambda \mathbf{z}_t \\ \mathbf{z}_{t_0} = \mathbf{z}_0 \end{cases} \quad (21)$$

with  $\lambda, \mathbf{z}_0 \in \mathbb{C}$  and  $\text{Real}(\lambda) \leq 0$ .

The application of a given  $q$ -stage ADRK scheme (defined by its Runge-Kutta coefficients  $\mathcal{ADRK}_q$ ) in the numerical integration of the linear ODE (21) leads to a polynomial form similarly to (14) as follow:

$$\begin{cases} \hat{\mathbf{z}}_{t_0} = \mathbf{z}_{t_0} = \mathbf{z}_0 \\ \hat{\mathbf{z}}_{t_{n+1}} = \mathcal{R}_{\mathcal{ADRK}_q}(\lambda h) \hat{\mathbf{z}}_{t_n} \end{cases} \quad (22)$$

with  $\mathcal{R}_{\mathcal{ADRK}_q}$  expressed similarly to (15) as a function of the coefficients  $\mathcal{ADRK}_q$  as follows:

$$\mathcal{R}_{\mathcal{ADRK}}(\lambda h) = \sum_{i=1}^q \alpha_i (\lambda h)^i = \mathbf{b}_{\mathcal{ADRK}}^T (I - (\lambda h) A_{\mathcal{ADRK}})^{-1} \mathbf{1} \quad (23)$$

The derivation of the integration coefficients is classically done by i) following section 2.2, matching the coefficients  $\alpha_i$  to  $\frac{1}{i!}$  up to a given order  $\hat{p}$  or ii) by optimizing the coefficients  $\alpha_i$  to match a stability region  $|\mathcal{R}_{\mathcal{ADRK}_q}(\lambda h)| \leq 1$  for a range of values of  $\lambda h$ . In the latter case, most case studies in previous works focus on specific cases such as dissipative processes *i.e.*  $\lambda \in \mathbb{R}$  with  $\lambda < 0$  (also stated when considering the derivation of stability polynomials as real axis inclusion Abdulle [2000, 2002], Bogatyrev [2005], Ketcheson and Ahmadi [2013], for which the maximal stable region of a  $q$ -stage Runge-Kutta method is reached, for first order schemes, for  $\lambda h = 2q^2$  Ketcheson and Ahmadi [2013]) and advection processes characterized by  $\lambda \in \mathbb{C}$  with  $\text{Real}(\lambda) = 0$  (imaginary axis inclusion Kinnmark and Gray [1984a,b], Mead and Renaut [1999], Ketcheson and Ahmadi [2013] for which the maximal stable region of a  $q$  stages Runge-Kutta method is reached, for first order schemes, for  $\lambda h = q - 1$  Ketcheson and Ahmadi [2013]) and other geometries (disk regions, thin rectangles ...etc.). Overall, such a methodology requires the definition of a potentially complex optimization procedure and usually the choice of a polynomial basis for  $\mathcal{R}_{\mathcal{ADRK}_q}$ .

We apply the proposed ADRK framework to derive stability polynomials without relying on a polynomial basis or on a specific optimization technique. Following the general workflow described in the previous section, the proposed framework relies on the following steps:

- We first define a linear equation (21) for a specific value of  $\lambda$  with  $\text{Real}(\lambda) \leq 0$ . We consider  $\lambda \in \mathbb{R}$  when deriving optimal stability polynomial in the real axis and similarly  $\lambda \in \mathbb{C}$  with  $\text{Real}(\lambda) = 0$  when considering imaginary axis stability inclusion. Other geometries can be considered.
- the set of trainable parameters comprises the ADRK coefficients

$$\Theta = \{\mathcal{ADRK}_q\} \quad (24)$$

- As training loss  $\mathcal{L}(\Theta)$ , we consider the following stability-related criterion, such that the integration scheme will be stable or values of  $\lambda h$  up to a maximum step-size  $h_f$ :

$$\mathcal{L}(\Theta) = \max \left( 0, \sum_{h=0}^{h_f} (|\mathcal{R}_{\mathcal{ADRK}}(h\lambda)| - 1) \right) \quad (25)$$

If the minimization of this training loss reaches 0, it guarantees that the trained ADRK scheme is stable over a finite set of time steps from 0 to  $h_f$ . To enforce the stability over a continuous range of values, we noted experimentally that we could gradually decrease the discretization spacing of  $h$ , typically up to  $h_f/100$ , until the training loss no longer changes at zero.

The maximum step size  $h_f$  is defined depending on the integration problem. For example, and following Ketcheson and Ahmadi [2013], when considering  $\lambda \in \mathbb{R}$  with  $\text{Real}(\lambda) < 0$  *i.e.* real axis inclusion, a maximum value can be set as  $h_f = 2q^2/|\lambda|$ . Similarly, if  $\lambda$  is purely imaginary, the limit values of stable step-size can be set as  $h_f = (q-1)/|\lambda|$ . The value of the stability polynomial  $\mathcal{R}_{\mathcal{ADRK}}$  can be either computed using the analytical formulation (23) or based on a ratio between  $\mathbf{z}_{t_{n+1}}$  and  $\mathbf{z}_{t_n}$  given in (22). Both these two formulations lead to the same value of  $\mathcal{R}_{\mathcal{ADRK}}$ , however, the analytical formulation (23), of the stability polynomial, have a faster converges rate when considering large values of  $\lambda h$ . In this work, real axis inclusion case studies were implemented using the analytical formulation (23) and imaginary axis inclusion experiments were implemented based on a ratio between  $\mathbf{z}_{t_{n+1}}$  and  $\mathbf{z}_{t_n}$ .

The above framework leads to the derivation of stable Runge-Kutta methods up to a given stability limit. We may apply the same workflow for the derivation of high-order integration schemes. Criterion (25) can be replaced by matching the ADRK coefficients  $\alpha_i$  to  $\frac{1}{i!}$  up to a given order  $\hat{p}$ . An example of application of this framework is given in section 4.1

### 3.3 ODE-adapted ADRK scheme

Given a non-linear ODE in (16), the stability and precision properties discussed above provide little to no knowledge regarding the performance of an integration scheme (specifically a Runge-Kutta one) in the time integration of non-linear



equations. From this point of view, adaptive step-size solvers are typically used as an efficient framework that deals with non-linear differential equations. This however comes at the expense of an increased computational complexity. Here, we apply the proposed ADRK scheme to derive computationally-efficient integration schemes for a given non-linear differential equation.

The proposed ADRK technique can exploit simulations from complex integration tools (adaptive solvers, integration schemes with a high number of stages) to optimize adapted integration routines that run at a smaller cost. Formally, let us assume that we are provided with representative time series of a full state vector  $\{\mathbf{z}\}$  of the ODE (16) with a given time sampling rate  $h$ . This training dataset may result from the application of a predefined high-complexity solver. For the sake of simplicity, we consider below a single time series of length  $N + 1$ ,  $\{\mathbf{z}_{t_0}, \mathbf{z}_{t_1}, \dots, \mathbf{z}_{t_N}\}$  with  $t_n = t_0 + nh$ . We could also consider similarly a dataset formed by different time series possibly of varying lengths.

We apply the proposed ADRK framework to learn a  $q$ -stage ADRK scheme which shall minimize the forecasting error. This leads to the following workflow:

- We first implement the non-linear equation (16) in a differentiable framework.
- We define the parameters that require optimization as the ADRK coefficients

$$\Theta = \{\text{ADRK}_q\} \quad (26)$$

- We specify the training loss  $\mathcal{L}(\Theta)$  as a trajectory matching cost such that the ADRK scheme will minimize the simulation error of a given trajectory  $\{\mathbf{z}_0, \mathbf{z}_1, \dots, \mathbf{z}_N\}$  as follows:

$$\mathcal{L} = \sum_{n=1}^N \|\mathbf{z}_{t_n} - \Phi_{\text{ADRK}_q, t_n}(\mathbf{z}_{t_{n-1}})\| \quad (27)$$

with  $\Phi_{\text{ADRK}_q, t_n}(\mathbf{z}_{t_{n-1}})$  is the numerical ADRK integration of (16).

Once the training procedure (20) has converged, we may apply the trained ADRK scheme to compute relevant numerical simulations for the ODE (16). The application of this framework is reported in sections 4.2 and 5.1 when considering respectively the integration of linear and non-linear ODEs.

### 3.4 Joint identification of ADRK schemes and ODE models

Deriving an ODE representation that reproduces the variability of a given dataset is a substantially different, and obviously harder, problem than the integration of a known equation. Particularly, choosing an integration algorithm for a known non-linear ODE may be complex and selecting an integration technique to learn an ODE appears even more complex. Using adaptive step size solvers is a practical solution of the integration issue, however, adaptive step size solvers as proposed in Chen et al. [2018] may be subject to memory and stability issues Gholami et al. [2019], Zhang et al. [2019]. The proposed framework appears as an interesting alternative to classical integration routines since an ADRK scheme can be learnt jointly with an approximate data-driven model from data.

Similarly to the previous case-study, let us assume that we are provided with a time series of  $N + 1$  samples of the full state vector  $\{\mathbf{z}_{t_0}, \mathbf{z}_{t_1}, \dots, \mathbf{z}_{t_N}\}$  with  $t_n = t_0 + nh$  and  $h$  a given time sampling rate. Contrary to the previous case, we assume that this state vector is governed by an unknown ODE. We apply the proposed framework to address the joint identification of the governing ODE and of the associated integration scheme. We proceed as follows:

- We first define a parameterization of the data-driven ODE

$$\dot{\mathbf{z}}_t = f_{\theta_{NN}}(t, \mathbf{z}_t) \quad (28)$$

in a differentiable framework, which usually corresponds to the definition of a neural network architecture with parameters  $\theta_{NN}$ . This architecture may rely on physics-informed parameterizations Brunton et al. [2016], Raissi et al. [2018], Fablet et al. [2017], de Bezenac et al. [2017] or on fully data-driven formulations. Regarding the numerical integration of the ODE (28), we aim to apply a trainable  $q$ -stage ADRK scheme  $\text{ADRK}_q$  given by (18).

- Here, we aim to jointly learn the parameters of the ADRK integration scheme as well as of the ODE, such that the set of trainable parameters  $\Theta$  are given by

$$\Theta = \{\text{ADRK}_q, \theta_{NN}\} \quad (29)$$

- Similarly to the previous case-study, the training procedure comes to optimize all trainable parameters with a view to minimizing the forecasting error from a predefined initial condition, which leads to the following training loss:

$$\mathcal{L}(\Theta) = \sum_{n=1}^N \|\mathbf{z}_{t_n} - \Phi_{\text{ADRK}_q, \theta_{NN}, t_n}(\mathbf{z}_{t_{n-1}})\| \quad (30)$$

with  $\Phi_{\text{ADRK}_q, \theta_{NN}, t_n}(\mathbf{z}_{t_{n-1}})$  the numerical ADRK integration of (28).

We report an example of application of this framework to the data-driven identification of ODE representations of chaotic dynamics when the available observation data involve a scarce time sampling 5.2

## 4 linear ODE case-study

In order to illustrate the key principles of the proposed framework, let us consider the following linear ODE:

$$\begin{cases} \dot{\mathbf{z}}_t = \lambda \mathbf{z}_t \\ \mathbf{z}_{t_0} = \mathbf{z}_0 \end{cases} \quad (31)$$

with  $\lambda$  either a real or complex constant.

We first perform numerical experiments of the proposed framework to learn stability-constrained ADRK scheme as introduced in Section 3.2. In a second experiment, we address the optimization of an ADRK scheme for the integration of this linear ODE,  $\lambda$  being known.

### 4.1 Learning stability-constrained integration schemes

We first report numerical experiments for the learning of stability-constrained ADRK schemes for a given range of integration time-steps. Specifically, we investigate in this section the ability of an  $\text{ADRK}_q$  scheme in the derivation of a stable polynomial  $\mathcal{R}_{\text{ADRK}_q}(\lambda h)$  that can integrate (31) up to the the limit of  $\lambda h = 2q^2$  for real axis inclusion (*i.e*  $\lambda \in \mathbb{R}$ ) and  $\lambda h = q - 1$  for imaginary axis inclusion (*i.e*  $\lambda \in \mathbb{C}$  with  $\text{Real}(\lambda) = 0$ ). For this purpose, the formulation of section 3.2 is studied on a 4-stage ADRK scheme. We include additional experiments, with a higher number of stages, in Supplementaray Material B.

#### 4.1.1 Real-axis inclusion

In this experiment, we state a  $\text{ADRK}_4$  scheme in the framework described in section 3.2 with  $h_f = 2q^2/|\lambda| = 32/|\lambda|$  and  $\lambda = -4$ . This case-study relates to the design of stable integration schemes for dissipation processes. Figure 3 illustrates the stability region of the trained ADRK scheme. A visual comparison of the stability region of this scheme with respect to the classical Euler and Runge Kutta 4 is also given in figure 4.

The trained  $\text{ADRK}_4$  is able to reach the upper bound of  $\lambda h = 2q^2$ . Furthermore, we can write the polynomial form of the trained ADRK scheme as follows:

$$\mathcal{R}_{\text{ADRK}_4}(\lambda h) = 1 + \lambda h + 0.1567(\lambda h)^2 + 0.0786(\lambda h)^3 + 0.0001(\lambda h)^4 \quad (32)$$

We can remind the Taylor series of the true solution  $\exp(\lambda h)$

$$\exp(\lambda h) = \sum_{p=0}^{p=\infty} \frac{(\lambda h)^p}{p!} = 1 + \lambda h + \frac{1}{2}(\lambda h)^2 + \frac{1}{6}(\lambda h)^3 + \dots \quad (33)$$

As discussed in Section 2.2, from the comparison of (32) to the expansion of the true solution (33), this stability-constrained scheme is only first order, which can be interpreted as trading the precision performance of the scheme for a larger stability region. Related works Ketcheson and Ahmadi [2013] drew similar observations, as the stability upper bound of  $\lambda h = 2q^2$  with  $\lambda < 0$  can only be reached using first-order integration schemes.

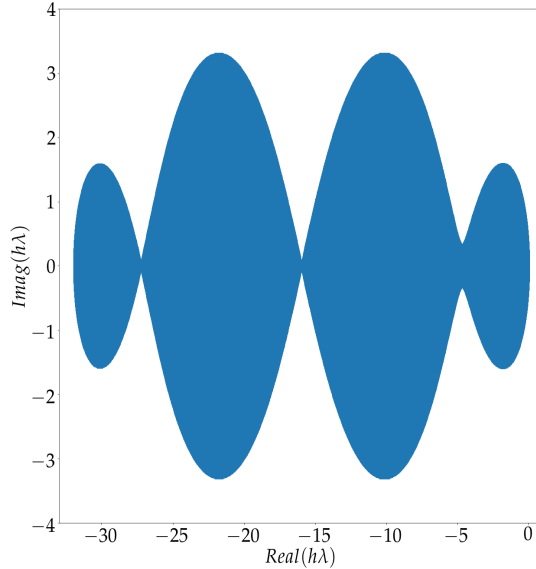


Figure 3: **Two dimensional stability region of the proposed  $ADRK_4$  integration scheme for dissipation processes:** The  $ADRK_4$  scheme was trained based on a stability criterion, up to  $\lambda h = 2q^2$  for  $\lambda \in \mathbb{R}$ ,  $\lambda < 0$ .

#### 4.1.2 Imaginary-axis inclusion

Here, we apply the stability-constrained  $ADRK_4$  framework described in section 3.2 with  $h_f = 3/|\lambda| = (q-1)/|\lambda|$  and  $\lambda = -j$  ( $j^2 = -1$ ), which refers to advection process Kinnmark and Gray [1984b]. Figure 5 illustrates the stability region of the trained ADRK scheme. The trained  $ADRK_4$  is again able to reach the theoretical upper bound of  $\lambda h = q - 1$  for imaginary-axis inclusion. We can write the polynomial form of the trained ADRK scheme as follows:

$$\mathcal{R}_{ADRK_4}(\lambda h) = 1 + \lambda h + 0.5478(\lambda h)^2 + 0.1490(\lambda h)^3 + 0.0480(\lambda h)^4 \quad (34)$$

The learned integration scheme, represented here for instance by the stability polynomial given in (34), is close to a second order scheme, and shows good stability properties along the real axis (the scheme is still stable slightly above  $\text{Real}(\lambda h) = -2$ ) even if it was only trained on purely imaginary values of  $\lambda h$ . This is principally due to the consistency condition  $\sum_{i=1}^q b_{ADRK_i} = 1$  enforced during the training phase. Finally, this stability-constrained integration scheme can also be used to solve other ODEs. Fig. 6 shows the simulated trajectories of Lorenz-63 Lorenz [1963] and Lorenz-96 Lorenz [1996] systems using the stability-constrained  $ADRK_4$  scheme (given by the stability polynomial (34)). Using this integration scheme on such non-linear equations leads to less accurate simulations when compared to the classical fourth order Runge-Kutta since the learnt scheme is only approximately second order and thus less accurate. The learnt scheme is however more stable when considering larger integration time-steps.

#### 4.2 ADRK scheme optimized for a known value of $\lambda$

Assuming the value of parameter  $\lambda$  is known in (31), here  $\lambda = -2$ , we address the derivation of an optimal 4-stage ADRK to reproduce a simulation of (31). This simulation is computed using the analytical solution of (31) from  $t = 0$  up to  $t = 10$  with a sampling rate  $h = 0.01$ . The initial condition  $\mathbf{z}_0$  was set to 0.5.

We apply the training framework introduced in Section 3.3, which relies on trajectory matching training loss. We report the forecasting performance of the trained model, including a comparison to the one of a Euler and RK4 scheme in figure 7.

Besides, the forecasting performance, we analyse the stability region of the trained  $ADRK_4$  scheme as depicted in Fig. 8 with respect to state-of-the-art techniques. The algorithm is stable for the integration time-step of the training data ( $\lambda h = 0.2$ ). Our integration scheme also generalizes to a range of values of  $\lambda h$  (up to  $\lambda h = 2$ ) even if they were never used in the training phase. This is due to the integration scheme constraints (18) that were forced during the training

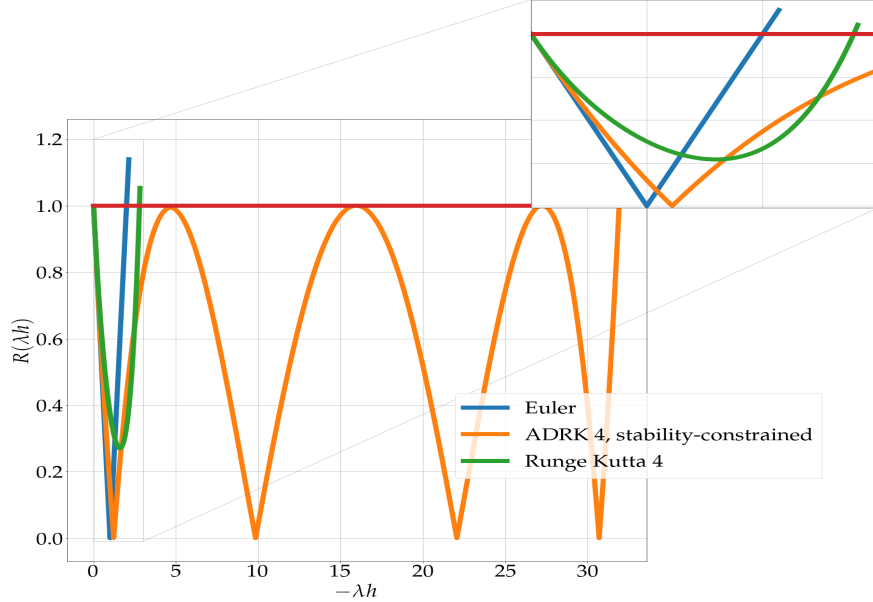


Figure 4: **One-dimensional stability region of the proposed stability constrained ADRK integration scheme with respect to Euler and Runge-Kutta 4 schemes.** We highlight the stability region of the stability constrained ADRK scheme (given in the complex plan in figure 3) in the real plan with respect to the Euler and Runge-Kutta 4 integration techniques.

phase. Furthermore, we may also deduce the order of the trained  $\mathcal{ADRK}_4$  from the analytical form of the stability polynomial  $\mathcal{R}_{\mathcal{ADRK}_4}(\lambda h)$ . We can derive the stability polynomial of the trained ADRK as follows:

$$\mathcal{R}_{\mathcal{ADRK}_4}(\lambda h) = 1 + \lambda h + 0.4990(\lambda h)^2 + 0.0037(\lambda h)^3 + 1.39 \times 10^{-6}(\lambda h)^4 \quad (35)$$

From the comparison of the polynomial (35) to the expansion of the true solution (33), we determine that the trained ADRK scheme is a second-order integration scheme.

## 5 Non-linear ODE case-study

In this section we report numerical experiments for the data-driven identification of Lorenz-63 dynamics. The Lorenz-63 dynamical system is a 3-dimensional model governed by the following ODE:

$$\begin{cases} \frac{dz_{t,1}}{dt} = \sigma(\mathbf{z}_{t,2} - \mathbf{z}_{t,1}) \\ \frac{dz_{t,2}}{dt} = \rho \mathbf{z}_{t,1} - \mathbf{z}_{t,2} - \mathbf{z}_{t,1} \mathbf{z}_{t,3} \\ \frac{dz_{t,3}}{dt} = \mathbf{z}_{t,1} \mathbf{z}_{t,2} - \beta \mathbf{z}_{t,3} \end{cases} \quad (36)$$

Under parameterization  $\sigma = 10$ ,  $\rho = 28$  and  $\beta = 8/3$ , this system involves chaotic dynamics with a strange attractor Lorenz [1963].

We simulate Lorenz-63 state sequences using the LOSDA ODE solver Hindmarsh [1983]. The integrated time-series was then sub-sampled based on a regular sampling rate. This state sequence is used as training data. In this section, we consider two experiments: i) the learning of an ADRK scheme to best reproduce Lorenz-63 trajectories knowing (36) and its parameterization ii) the joint data-driven identification of the underlying ODE and of an ADRK scheme Lorenz model to best reproduce the trajectories of the training dataset.

### 5.1 ADRK scheme for Lorenz-63 system Integration

In this experiment, we compare a 4-stage ADRK scheme to the classical Runge-Kutta 4 method ( $\mathcal{RK}_4$ ) for the integration of the Lorenz 63 system. The  $\mathcal{ADRK}_4$  is optimized to match a single Lorenz 63 trajectory sampled at  $h = 0.1, 0.15, 0.16, 0.17, 0.18, 0.19$ . In table 1, we report the results of this experiment by specifying the ability of the  $\mathcal{RK}_4$  and the  $\mathcal{ADRK}_4$  schemes to correctly integrate the Lorenz 63 equation at these integration time-steps.

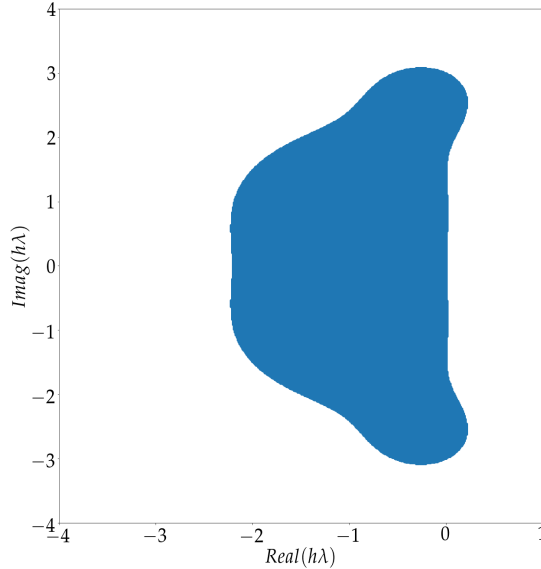


Figure 5: **Two dimensional stability region of the proposed  $ADRK_4$  integration scheme** for advection processes: The  $ADRK_4$  scheme was trained based on a stability criterion, up to  $\lambda h = q - 1$  for  $\lambda \in \mathbb{C}$  with  $\text{Real}(\lambda) = 0$ .

h	0.1	0.15	0.16	0.17	0.18	0.19
$ADRK_4$	✓	✓	✓	✓	✗	✗
$RK_4$	✓	✗	✗	✗	✗	✗

Table 1: **Integration ability of both the  $RK_4$  and the trained  $ADRK_4$  on the Lorenz 63 system with different integration time steps.**

Specifically, an integration scheme succeeds if it generates chaotic Lorenz 63 trajectories and fails if it makes the integration blow up, generates stable equilibrium points or other non-chaotic limit-sets.

The  $ADRK_4$  is able to correctly generate chaotic trajectories from the Lorenz 63 ODE up to a sampling rate  $h = 0.17$ . By contrast, the classical Runge-Kutta 4 technique either blows up or get stuck at an equilibrium point for  $h \geq 0.15$ . Whereas the latter is derived independently of any non-linear ODE to match a fourth order truncation of the Taylor expansion (33) leading to a smaller acceptable integration time-step for the Lorenz equation, the good properties of the trained  $ADRK_4$  scheme comes from the data-driven training procedure. Figure 9 illustrates the stability region of the trained ADRK schemes on each integration time-step. When compared to the classical RK(4) scheme, the ADRK schemes involve larger stability regions as the integration time-step  $h$  increases from 0.1 to 0.17, which makes them relevant for the integration of Lorenz 63 ODE.

Interestingly, the stability region mostly increase in the imaginary axis for the Lorenz equation which emphasises that this equation is probably dominated by an advection process. This observation can be explained through a linearization of the Lorenz 63 model based on the Koopman operator Koopman [1931]. Although such a linearization would theoretically lead to an infinite dimensional system, an approximate finite dimensional model can be written as proposed in Brunton et al. [2017] with an additional forcing term. In such representation, the linear dynamical system, shadowing the true non-linear Lorenz-63 equation, will have only imaginary eigenvalues. This property can explain the stability growth towards the imaginary axis rather than the real one. The  $ADRK_4$  scheme is able to match these requirements leading to a larger acceptable step-size for Lorenz-63 system.

Overall, this experiment highlights one of the main potentials of the proposed framework when considering the time integration of non-linear differential equations. In these applications, and especially when considering complex problems for which a detailed analytical study of the underlying dynamical properties can hardly be achieved, the

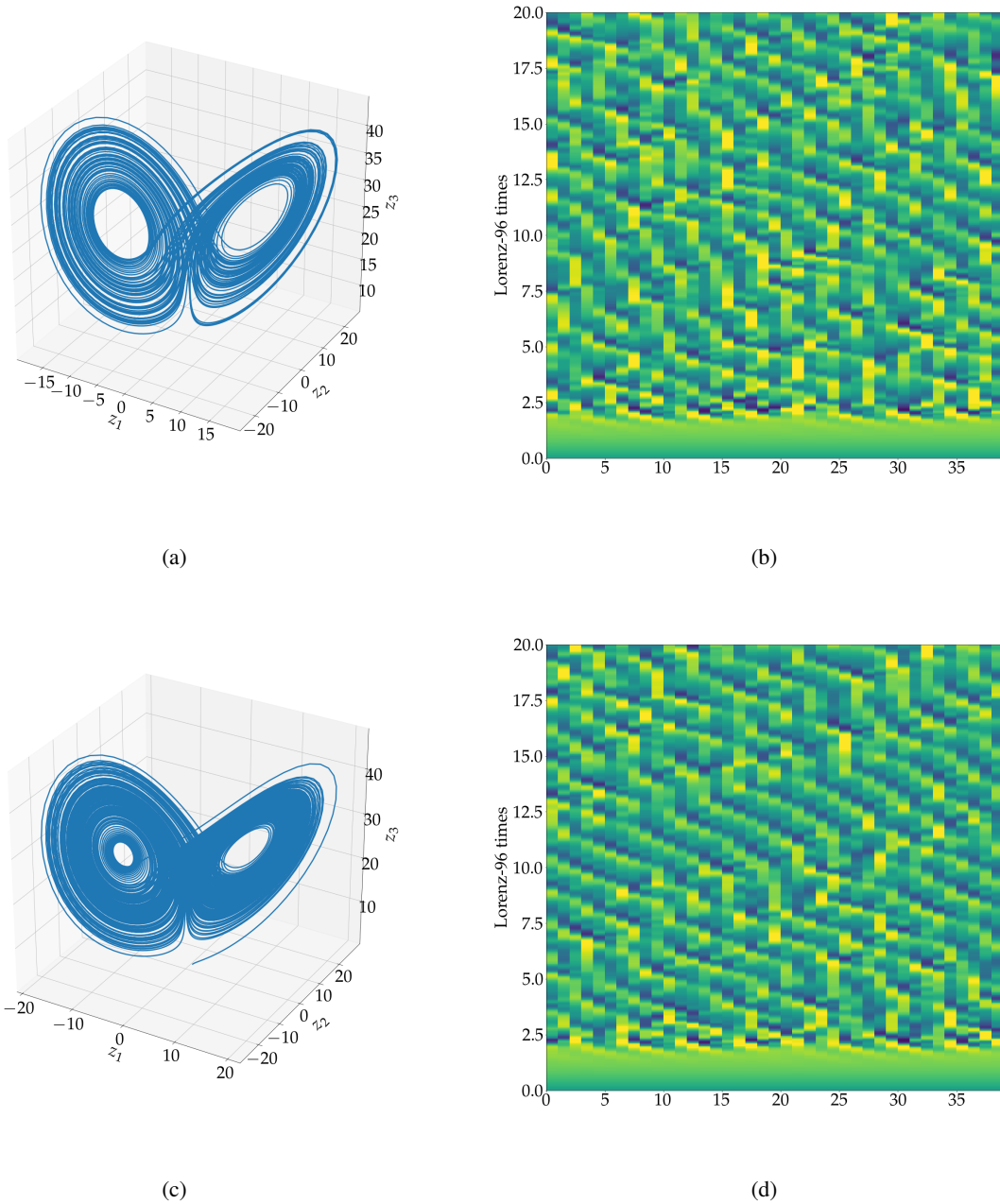


Figure 6: **Application of the stability-constrained  $ADRK_4$  integration scheme on non-linear dynamics.** Integration of Lorenz 63 (a) and 96 (b) equations using our stability-constrained  $ADRK_4$  integration scheme given by the stability polynomial (34). Figures (c) and (d) correspond to the integration of the Lorenz 63 and 96 models respectively using the classical adaptive step-size Adams technique.

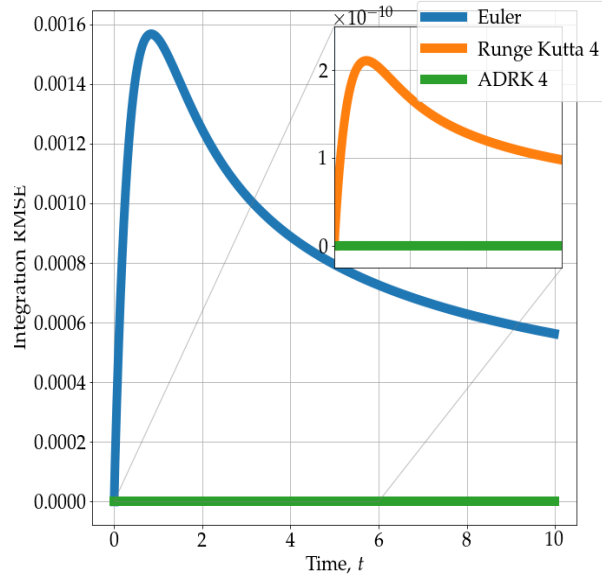


Figure 7: **Root Mean Squared Error (RMSE) of several numerical resolutions of (31) with respect to the analytical solution.** The numerical resolution is carried given the same initial condition and integration time step for all the integration schemes.

proposed ADRK framework adapts to the observed non-linear dynamics leading to an equation-adapted integration scheme.

## 5.2 Lorenz-63 system Identification

In this experiment we subsampled the Lorenz 63 trajectories to a low regular sampling rate  $h_1 = 0.2$ ,  $h_2 = 0.3$  and  $h_3 = 0.4$ . The goal of this experiment is to try to discover a model for the Lorenz system given temporally sparse data that can not be integrated using classical integration schemes. For this purpose, an approximate model  $f_{\theta_{NN}}$  given in (28) is optimized to generate a Lorenz state sequence sampled at a sparse sampling rate. The approximate model is a Linear Quadratic Model (LQM) as proposed in Fablet et al. [2017]. This architecture ensures that the true model lies within the space of possible model parameterizations. For benchmarking purpose, the following models were tested:

- **Sparse regression model** Brunton et al. [2016] (SR): This model computes a sparse regression over an augmented states vector based on second order polynomial representations of the Lorenz states. The learnt dynamical model is then integrated to compute forecasts using the LOSDA ODE solver Hindmarsh [1983].
- **Euler based model** (Euler): The approximate model  $f_{\theta_{NN}}$  is integrated using an explicit Euler scheme.
- **Runge-Kutta 4 based model** ( $\mathcal{RK}_4$ ): The approximate model  $f_{\theta_{NN}}$  is integrated using the classical Runge-Kutta 4 scheme.
- **Neural ODE** (DOPRI8): The approximate model  $f_{\theta_{NN}}$  is integrated using the adaptive step size dopri8 solver. The backward pass is computed here using the adjoint sensitivity method L. S. Pontryagin [1962], as proposed in Chen et al. [2018].
- **11-stage ADRK scheme**: Proposed automatic differentiation Runge Kutta scheme with a number of stages equal to 11. In this architecture, the weights of the integration scheme are learnt jointly with the parameters of the approximate model as explained in section 3.4.

Besides the proposed ADRK scheme, all the benchmark models use a specified integration routines and the optimization is carried only with respect to the parameters of the model  $f_{\theta_{NN}}$ . Finally the data-driven integration schemes are noted  $ADRK_{h_1}$ ,  $ADRK_{h_2}$  and  $ADRK_{h_3}$  with each index ( $h_1$ ,  $h_2$  and  $h_3$ ) corresponding to the time sampling of the training time series.

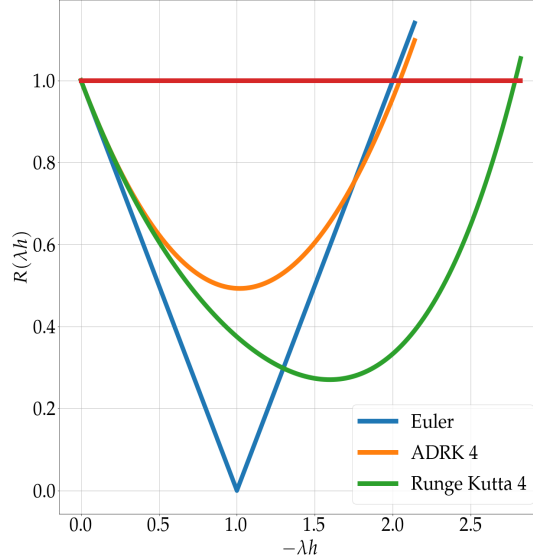


Figure 8: **One-dimensional stability region of the proposed ADRK integration scheme with respect to classical state-of-the-art algorithms.** We compare the stability region of the proposed  $\mathcal{ADRK}_4$  optimized on a short term forecasting cost with respect to classical state-of-the-art Euler and Runge-Kutta-4 integration schemes.

Model		$h = 0.2$	$h = 0.3$	$h = 0.4$
SR	$t_0 + h$	9.06	> 10	9.34
	$t_0 + 4h$	6.02	5.81	6.97
Euler	$t_0 + h$	4.27	2.57	1.99
	$t_0 + 4h$	> 10	> 10	7.89
$\mathcal{RK}_4$	$t_0 + h$	2.05	3.10	2.48
	$t_0 + 4h$	3.82	7.33	> 10
Dopri8	$t_0 + h$	<b>0.005</b>	<b>0.0001</b>	3.1305
	$t_0 + 4h$	0.021	<b>0.0003</b>	> 10
$\mathcal{ADRK}$	$t_0 + h$	0.017	0.077	<b>0.189</b>
	$t_0 + 4h$	<b>0.020</b>	0.23	<b>1.93</b>

Table 2: **Forecasting performance of data-driven models for Lorenz-63 dynamical model.** Mean Root Mean Squared Error (RMSE) for different forecasting time-steps of the tested models.

**ADRK Performance analysis:** We report the forecasting performance in Table 2. Figures 10 and 11 illustrate the trajectories generated using the trained  $\mathcal{ADRK}$  and Dopri8 based models respectively. When compared to classical fixed step-size models, the proposed  $\mathcal{ADRK}$  is the only model able to discover the hidden dynamics of the system. The reason is that classical fixed step-size solvers such as Euler and Runge-Kutta 4 lead to high truncation errors making the identifiability impossible, the proposed model in the other hand, when deployed with 11 stages, can mimic adapted, stable and precise integration schemes, that are able to unfold the true Lorenz 63 dynamics. These results can be explained by analysing the truncation error of a  $q$ -stage ADRK scheme. The truncation error of numerical integration scheme provides a straightforward tool to state about the one step ahead integration performance of a given integration technique as a function of its order  $p$ . Assuming that the  $q$ -stage ADRK corresponds to a  $\hat{p}$ -order numerical integration scheme, the loss function of the ADRK relates, following (11), to the truncation error of the learnt integration scheme as follows:

$$\mathcal{L}(\Theta) \propto (\epsilon_n)^2 = (E(\hat{p}, f_{\theta_{NN}}, \mathbf{z}_{t_n}, t_n)h^{p+1})^2 \quad (37)$$



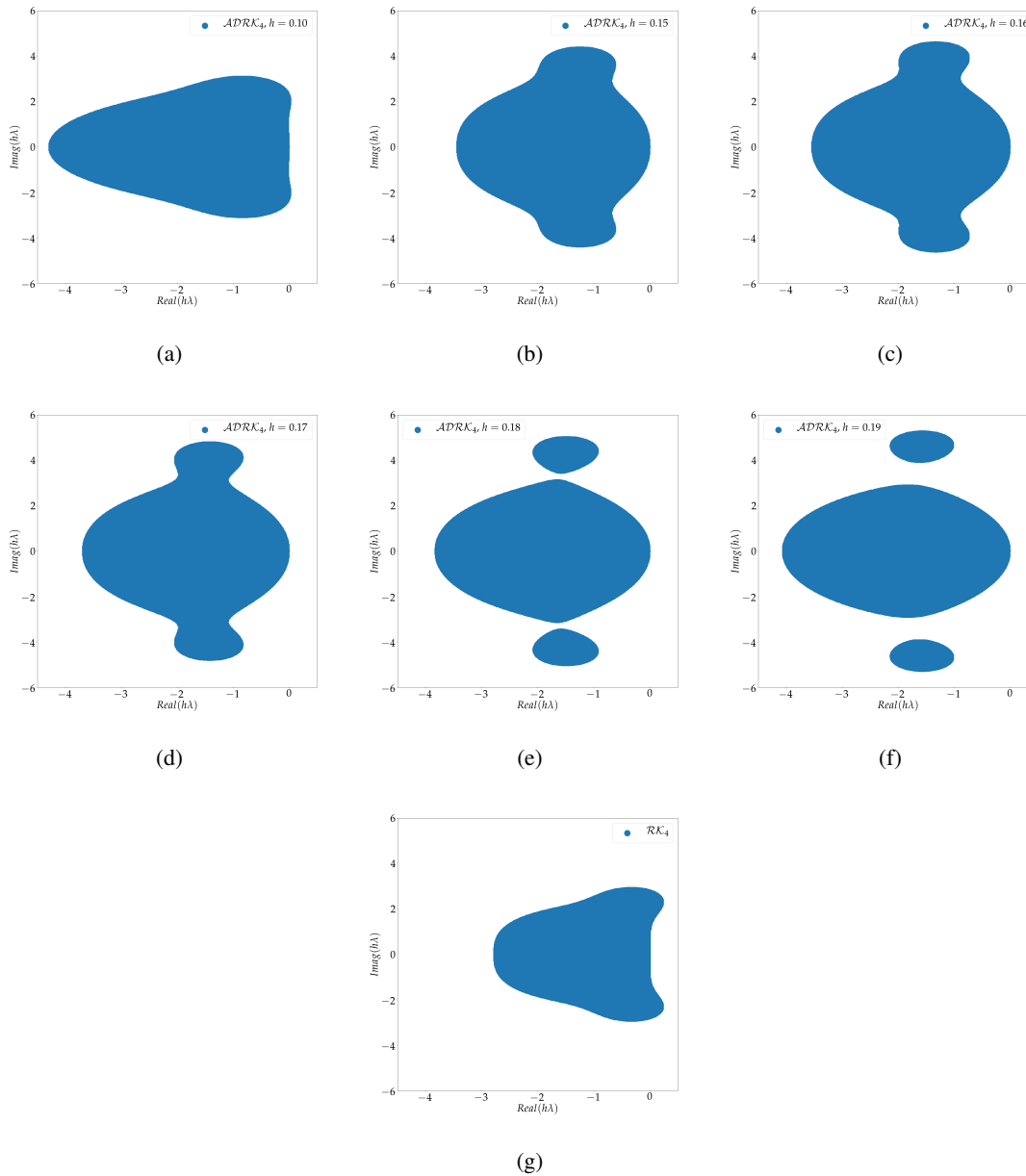


Figure 9: *Stability region of the  $ADRK_4$  schemes trained to match the Lorenz 63 states at different integration time-steps.* Stability region of the  $ADRK_4$  scheme trained based on an integration time-step ranging from  $h = 0.1$  to  $h = 0.19$  are given in figures (a) to (f) respectively. The stability region of the classical  $RK_4$  is given as a reference in fig. (g).

where  $E \in \mathbb{R}$  a function of  $\hat{p}$ ,  $f_{\theta_{NN}}$ ,  $\mathbf{z}$  and  $t_n$ .

Equation (37) illustrates a main characteristic about learning data-driven representations of dynamical models, using fixed step-size integration schemes: (ii) we may jointly tune  $f_{\theta_{NN}}$  and  $\hat{p}$  in the ADRK architecture to lower the training loss function.

Assuming that the integration time-step  $h$  is set by the temporal sampling of our training data, one may decrease the training error through tuning the approximation  $f_{NN}$  of the true dynamical model  $f$  Brunton et al. [2016], Fablet et al. [2017]. One may also decrease the training loss function through an adapted order  $\hat{p}$  of the integration scheme, parameterized in the proposed ADRK framework by the number of stages  $q$ . Overall, the training loss lower bound motivates the use of the proposed framework in identification applications since if we start an identification problem with a fixed integration scheme or family that can not integrate properly the dynamics (which is a fair enough claim since we are interested in non-linear system identification where stability and precision criteria are not known a priori), we might not find a good fit for our data simply due to the choice of the integration method.

The poor results of the sparse regression method is in the other hand simply due to a wrong estimation of the derivatives. This step is inevitable using such technique and when provided with temporally sparse data, a decent estimation becomes impossible. Adaptive step-size based models in the other hand lead to an overall better short term forecast and, similarly to the proposed framework, are able to correctly identify the Lorenz 63 model when provided with data sampled at  $h = 0.2$  and  $h = 0.3$ . This method however is unable to derive a decent approximation when provided with data sampled at  $h = 0.4$ . We believe that this is principally due to the backward pass formulation proposed in Chen et al. [2018] which integrates an augmented ODE backward in time in the training phase (to avoid heavy memory usage due to classical backpropagation through residual network). These assumptions were then verified on the same case study by computing the backward pass through storing every evaluation of the adaptive solver. Within this configuration, the Dopri8 technique is able to unfold the true structure of the attractor and achieves a decent short term forecast, similar to the proposed  $ADRK$  (0.18 at  $t_0 + h$  and 0.94 at  $t_0 + 4h$ ). Figure 13 illustrates a simulation example carried using both the  $ADRK$  and the Dopri8 based models on the dataset sampled at  $h = 0.4$  (the Dopri8 based model is trained without the use of the adjoint method). The number of function evaluations of the adaptive solver varies between 19 and 14313 (please refer to figure 13 for a visualisation of a lower bound of the number of function evaluations per initial condition) against only 11 for the proposed framework making our method highly computationally efficient, especially when considering more complex systems. Furthermore, storing all the integration steps of an adaptive solver in order to backpropagate the training error is not guaranteed to be feasible in first place on more complex, high dimensional systems due to memory blowups.

**Generalizability of the learnt ADRK schemes to new integration configurations:** Another interesting experiment is illustrated in Fig. 12. The trained integration scheme is dissociated from the data-driven model  $f_{NN}$  and used to integrate the Lorenz 96 differential equation. Since the integration scheme is always consistent, there is a range of integration time-steps for which the integration scheme is stable given a new ODE and thus, converges to give a correct simulation of the equation.

**Stability and precision analysis of the learnt ADRK schemes:** Figure 15 illustrates the stability region of the trained schemes. The Euler and Runge-Kutta-4 schemes are shown as references. Interestingly, the higher the integration time-step of the trained model, the larger the stability region of the optimized scheme. Furthermore, analysing the stability polynomials of the ADRK schemes given in (38) reveals an interesting aspect of the proposed framework. The order of the integration scheme increases when moving from the data sampled at  $h = 0.2$  to  $h = 0.3$  but decreases when trained on the trajectory sampled at  $h = 0.4$ . These observations can be highlighted through the computation of the normalized coefficients error<sup>3</sup> given in figure 14. We believe that the order of the  $ADRK_{h_3}$  drops to lower the truncation error of the integration scheme as when considering high values of  $\lambda h$ , and as illustrated in figure 1, low order schemes tend to have a smaller truncation error, which guarantees a larger stability region, necessary for both the identification and the integration of the data driven model at  $h = 0.4$ . From this point of view, the learnt integration

<sup>3</sup>This score is computed as the sum of the normalized root squared error of each coefficient of the stability polynomials, given in 38, with respect to the true Taylor expansion of the analytical solution up to a given order  $p$ .

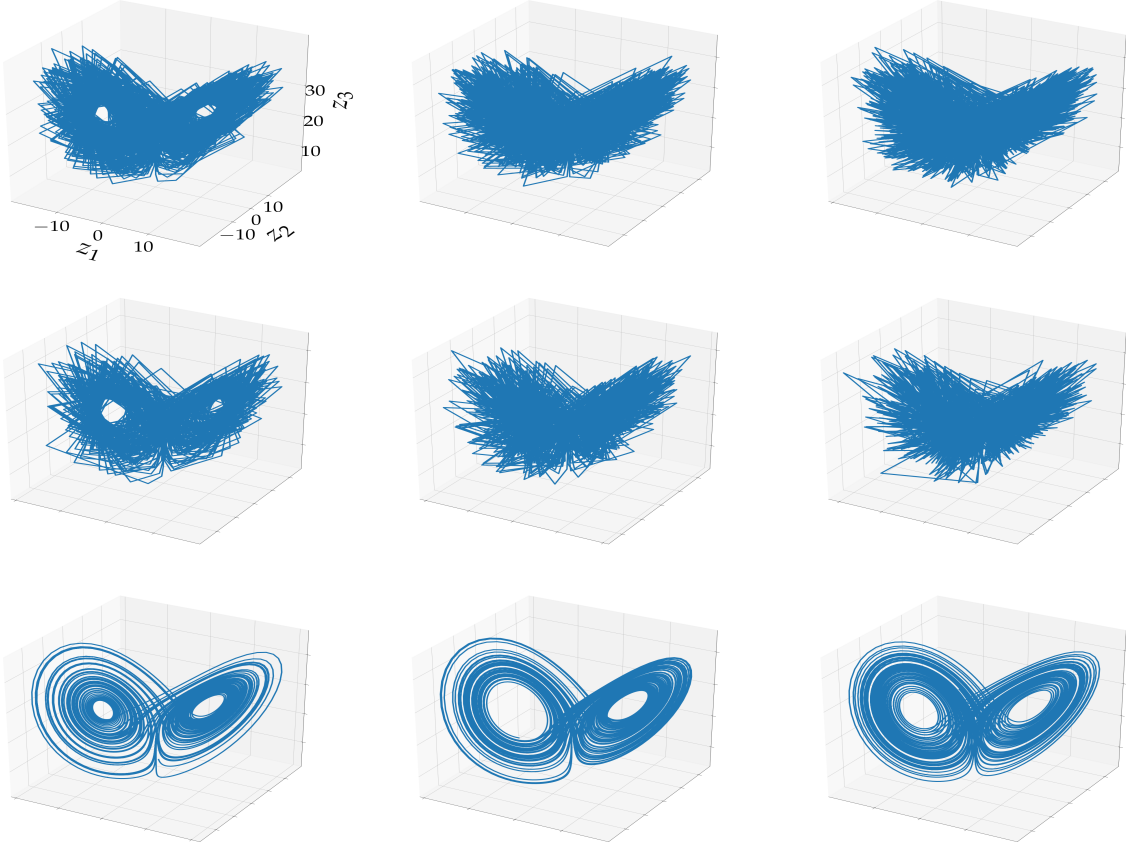


Figure 10: **Training data and simulated attractors from the corresponding data-driven ADRK based models.** First row, training sequence generated with different regular time sampling; second row, simulated attractors of the data-driven models, third row; simulated attractors using a smaller integration time-step. The columns correspond to the sampling rate of the training series (ranging from 0.2 to 0.4.) which corresponds to the integration time-step  $h$  when considering fixed step-size algorithms.

scheme was able to adapt, similarly to the previous case study to the dynamics of the observations, but also to their sampling leading to a correct identification of the dynamics.

$$\begin{aligned}
 \mathcal{R}_{ADRK_{h_1}}(h\lambda) &= 1 + h\lambda + 0.5009(h\lambda)^2 + 0.1646(h\lambda)^3 + 0.04010(h\lambda)^4 + 0.007592(h\lambda)^5 + \\
 &\quad 0.001136(h\lambda)^6 + \dots \\
 \mathcal{R}_{ADRK_{h_2}}(h\lambda) &= 1 + h\lambda + 0.5021(h\lambda)^2 + 0.1691(h\lambda)^3 + 0.04273(h\lambda)^4 + 0.008826(h\lambda)^5 + \\
 &\quad 0.001510(h\lambda)^6 + 2.147^{-4}(h\lambda)^7 + 2.550E^{-5}(h\lambda)^8 + \dots \\
 \mathcal{R}_{ADRK_{h_3}}(h\lambda) &= 1 + h\lambda + 0.4971(h\lambda)^2 + 0.1630(h\lambda)^3 + 0.03935(h\lambda)^4 + 0.007438(h\lambda)^5 + \\
 &\quad 0.001138(h\lambda)^6 + 1.411^{-4}(h\lambda)^7 + 1.423E^{-5}(h\lambda)^8 + \dots \\
 \mathcal{R}_{exp}(h\lambda) &= 1 + h\lambda + 0.5(h\lambda)^2 + 0.16666(h\lambda)^3 + 0.04166(h\lambda)^4 + 0.008333(h\lambda)^5 + \\
 &\quad 0.001388(h\lambda)^6 + 1.984^{-4}(h\lambda)^7 + 2.480E^{-5}(h\lambda)^8 + \dots
 \end{aligned} \tag{38}$$

## 6 Conclusion

In this work, we address the data-driven identification of numerical integration schemes for the simulation and the identification of ODE representations of dynamical systems. Formally, we focus on explicit Runge-Kutta integration

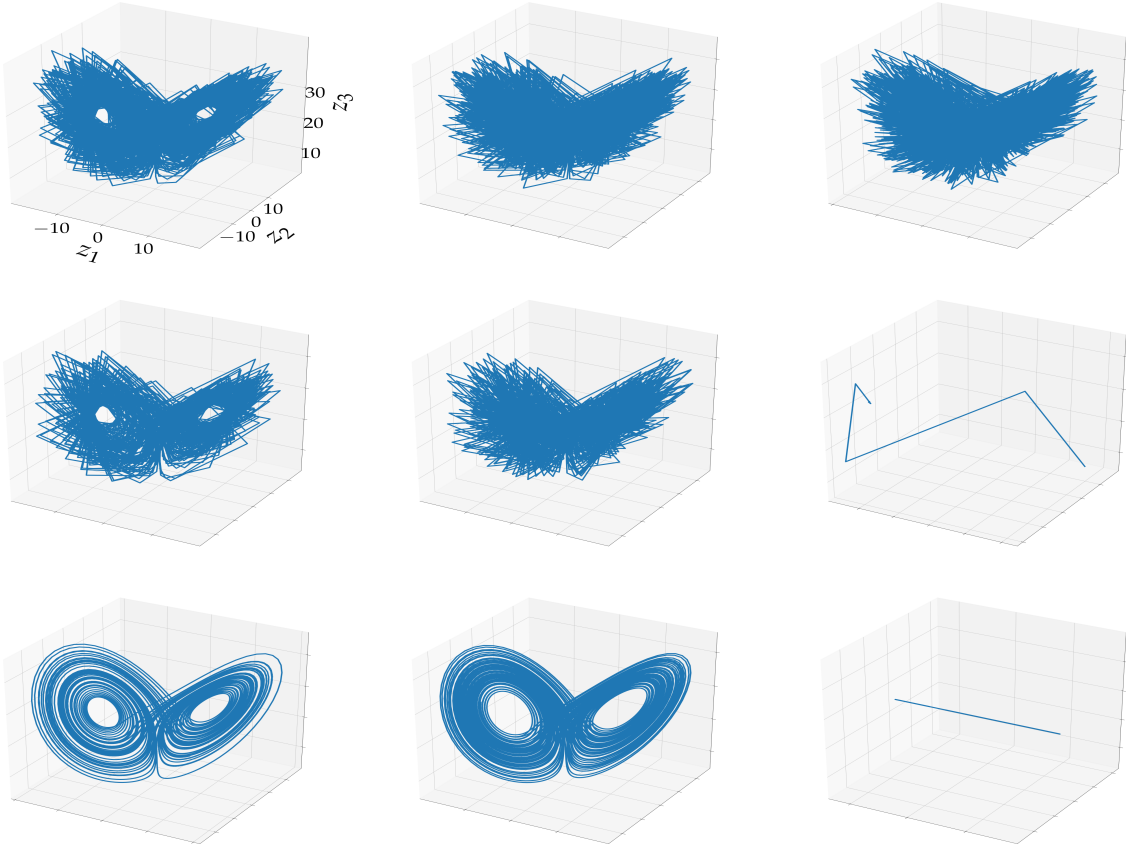


Figure 11: *Training data and simulated attractors from the corresponding data-driven Dopri8 based models.* First row, training sequence generated with different regular time sampling; second row, simulated attractors of the data-driven models, third row; simulated attractors using a smaller integration time-step. The columns correspond to the sampling rate of the training series (ranging from 0.2 to 0.4.).

schemes and we state the identification of these schemes as an optimization problem according to some predefined criteria. We implement numerically the proposed approach using deep learning framework to benefit from embedded automatic differentiation tools. Interestingly, we can consider both theory-guided and data-driven criterion. We demonstrate the relevance of the proposed framework on three case studies, namely the derivation of stable integration schemes for the time integration of linear equations, the identification of an integration scheme optimized for the resolution of a predefined non-linear ODE and the joint identification of numerical integration schemes and ODE representations from a sequence of observations.

The reported experiments demonstrate the relevance of the proposed framework. Specifically, we demonstrate experimentally that stability-constrained AD RK schemes can reach stability upper bounds found in related works Van der Houwen [1972], Kinnmark and Gray [1984c], Ketcheson and Ahmadi [2013] without resorting to any extra-parameterization or requiring to choose basis functions that generate stability functions as in Ketcheson and Ahmadi [2013]. We also show that trainable AD RK schemes can adapt to given dynamics and to the sampling of the observations to properly integrate non-linear dynamics. Besides, they make tractable the joint identification of an integration scheme and of an ODE representation of some observed dynamics. We may point out that the reported results do not relate to the parametric form of the ODE but truly to the fact that using classical integration schemes might not be appropriate for the considered identification task.

Future work will first investigate the proposed framework for the identification of data-driven representations of geophysical dynamics. Applications to satellite-derived sea surface observations appear very appealing. The typical sampling of these data range from a few hours to 10 days. The proposed AD RK then seem more appropriate compared with fixed step-size techniques to explore data-driven representations.

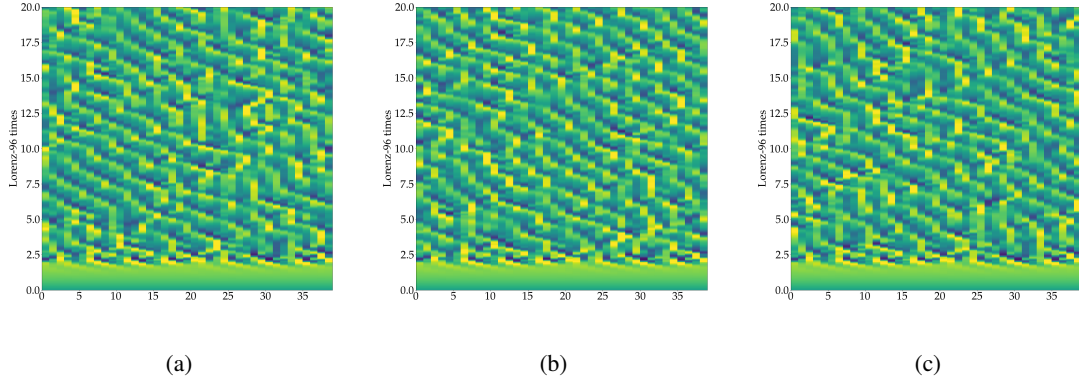


Figure 12: *Application of the learnt integration scheme on the Lorenz 96 equation.* Integration of the Lorenz 96 model using the  $ADRK_{h_1}$  in (a), the  $ADRK_{h_2}$  in (b) and the  $ADRK_{h_3}$  in (c).

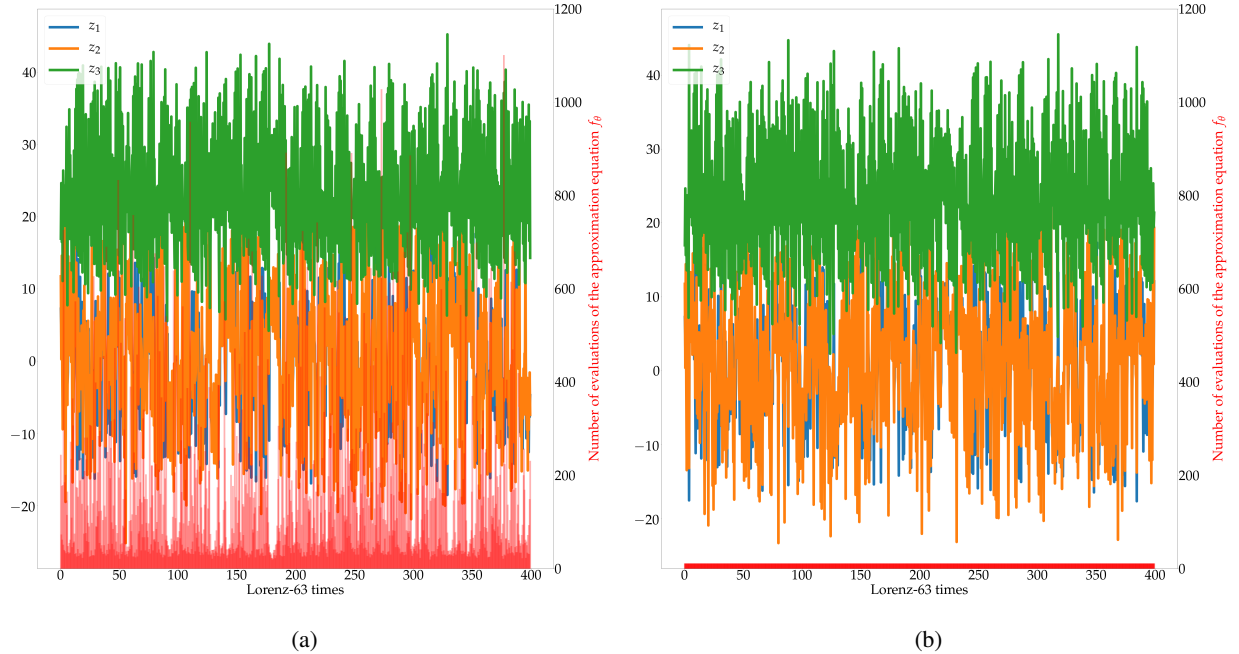


Figure 13: *Simulated time series of the Lorenz 63 dynamics from two data-driven models.* (a) Dopri8 based model trained without the adjoint technique proposed in Chen et al. [2018]. (b) Proposed  $ADRK$ . The training of both these models was carried on the dataset sampled with  $h = 0.4$ . The red bars correspond, in the  $ADRK$  figure (b) to the exact number of evaluations of the approximate function  $f_\theta$  for every point within the simulated trajectory. In the DOPRI8 figure, the bars correspond to the minimum number of evaluations, computed when considering each integration carried within a single stage integration scheme.

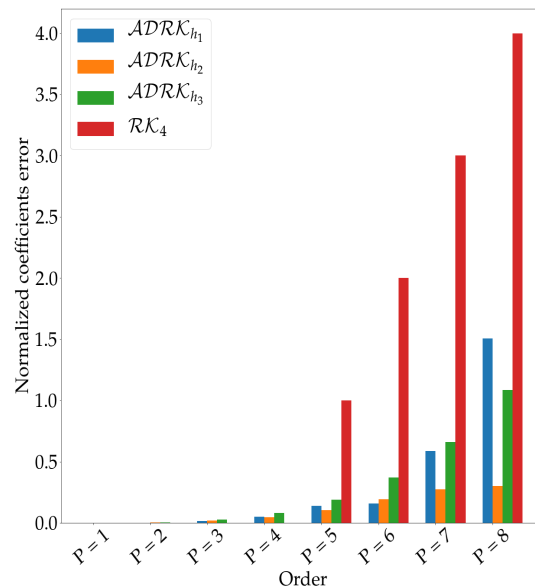


Figure 14: **Normalized coefficients error of the stability polynomials of the ADRK schemes (Lorenz 63 identification case-study) with respect to the Taylor expansion of the analytical solution.** Normalized cumulative error of the ADRK stability polynomials coefficients up to the order  $p$ , with respect to the Taylor expansion of the true solution. The Runge-Kutta-4 algorithm is given as a reference.

From a methodological point of view, future work may first investigate the identification of ADRK schemes using other data-driven or stability-based criteria, including including ones defined as weighted versions of different criteria. Whereas this work focuses on explicit integration schemes, the proposed framework could be extended to implicit ones including a non-linear differentiable solver. This could apply for instance to classical implicit solvers such as modified newton iteration Cooper and Vignesvaran [1993b].

## References

- Thomas S. Parker and Leon O. Chua. *Integration of Trajectories*, pages 83–114. Springer New York, New York, NY, 1989. ISBN 978-1-4612-3486-9. doi:10.1007/978-1-4612-3486-9\_4. URL [https://doi.org/10.1007/978-1-4612-3486-9\\_4](https://doi.org/10.1007/978-1-4612-3486-9_4).
- J. W. Thomas. *Stability*, pages 97–145. Springer New York, New York, NY, 1995. ISBN 978-1-4899-7278-1. doi:10.1007/978-1-4899-7278-1\_4. URL [https://doi.org/10.1007/978-1-4899-7278-1\\_4](https://doi.org/10.1007/978-1-4899-7278-1_4).
- P Friedmann, CE Hammond, and Tze-Hsin Woo. Efficient numerical treatment of periodic systems with application to stability problems. *International Journal for Numerical Methods in Engineering*, 11(7):1117–1136, 1977.
- Julien Berland, Christophe Bogey, and Christophe Bailly. Low-dissipation and low-dispersion fourth-order runge-kutta algorithm. *Computers and Fluids*, 35(10):1459–1463, 2006. ISSN 0045-7930. doi:<https://doi.org/10.1016/j.compfluid.2005.04.003>. URL <https://www.sciencedirect.com/science/article/pii/S0045793005000575>.
- Steven Ruuth. Global optimization of explicit strong-stability-preserving runge-kutta methods. *Mathematics of Computation*, 75(253):183–207, 2006.
- Ernst Hairer and Gerhard Wanner. *B-Stability and Contractivity*, pages 180–200. Springer Berlin Heidelberg, Berlin, Heidelberg, 1996a. ISBN 978-3-642-05221-7. doi:10.1007/978-3-642-05221-7\_12. URL [https://doi.org/10.1007/978-3-642-05221-7\\_12](https://doi.org/10.1007/978-3-642-05221-7_12).
- Patrick Stiller, Friedrich Bethke, Maximilian Böhme, Richard Pausch, Sunna Torge, Alexander Debus, Jan Vorberger, Michael Bussmann, and Nico Hoffmann. Large-scale neural solvers for partial differential equations. *CoRR*, abs/2009.03730, 2020. URL <https://arxiv.org/abs/2009.03730>.

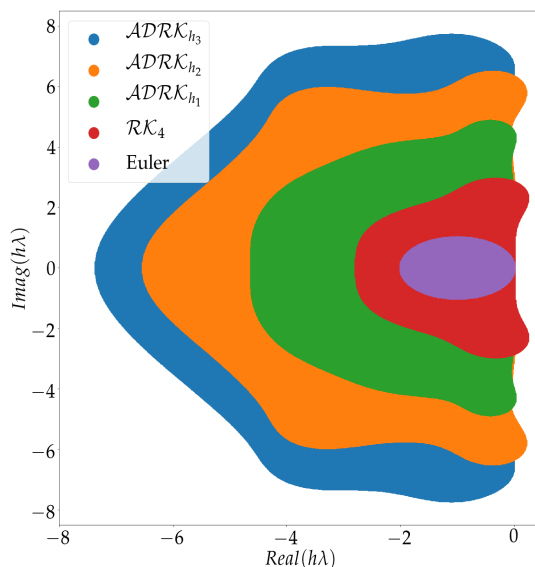


Figure 15: *Stability region of the ADRK schemes (Lorenz 63 identification case-study).*

- YUEHAW KHOO, JIANFENG LU, and LEXING YING. Solving parametric pde problems with artificial neural networks. *European Journal of Applied Mathematics*, page 1–15, 2020. doi:10.1017/S0956792520000182.
- Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. *arXiv preprint arXiv:2010.08895*, 2020.
- Yoshiro Suzuki. Neural network-based discretization of nonlinear differential equations. *Neural Computing and Applications*, 31(7):3023–3038, 2019.
- Yohai Bar-Sinai, Stephan Hoyer, Jason Hickey, and Michael P Brenner. Learning data-driven discretizations for partial differential equations. *Proceedings of the National Academy of Sciences*, 116(31):15344–15349, 2019.
- Yi-Jen Wang and Chin-Teng Lin. Runge-kutta neural network for identification of dynamical systems in high accuracy. *IEEE Transactions on Neural Networks*, 9(2):294–307, 1998.
- Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, April 2016. ISSN 0027-8424, 1091-6490. doi:10.1073/pnas.1517384113. URL <http://www.pnas.org/lookup/doi/10.1073/pnas.1517384113>.
- Ronan Fablet, Said Ouala, and Cedric Herzet. Bilinear residual Neural Network for the identification and forecasting of dynamical systems. *SciRate*, dec 2017. URL <https://scirate.com/arxiv/1712.07003>.
- H Martin Bücker, George Corliss, Paul Hovland, Uwe Naumann, and Boyana Norris. *Automatic differentiation: applications, theory, and implementations*, volume 50. Springer Science & Business Media, 2006.
- Hong Zhang, Adrian Sandu, and Sebastien Blaise. High order implicit-explicit general linear methods with optimized stability regions. *SIAM Journal on Scientific Computing*, 38(3):A1430–A1453, 2016.
- Gerhard Wanner. *On the integration of stiff differential equations*, pages 209–226. Birkhäuser Basel, Basel, 1977. ISBN 978-3-0348-5575-4. doi:10.1007/978-3-0348-5575-4\_11. URL [https://doi.org/10.1007/978-3-0348-5575-4\\_11](https://doi.org/10.1007/978-3-0348-5575-4_11).
- Lorenzo Pareschi and Giovanni Russo. Implicit-explicit runge-kutta schemes for stiff systems of differential equations. *Recent trends in numerical analysis*, 3:269–289, 2000.
- Liviu Gr. Ixaru and Guido Vanden Berghe. *Runge-Kutta Solvers for Ordinary Differential Equations*, pages 223–304. Springer Netherlands, Dordrecht, 2004. ISBN 978-1-4020-2100-8. doi:10.1007/978-1-4020-2100-8\_6. URL [https://doi.org/10.1007/978-1-4020-2100-8\\_6](https://doi.org/10.1007/978-1-4020-2100-8_6).

- J. C. Butcher. Coefficients for the study of runge-kutta integration processes. *Journal of the Australian Mathematical Society*, 3(2):185–201, 1963. doi:10.1017/S1446788700027932.
- Gerhard Wanner and Ernst Hairer. *Solving ordinary differential equations II*. Springer Berlin Heidelberg, 1996.
- G.J. Cooper and R. Vignesvaran. Some schemes for the implementation of implicit runge-kutta methods. *Journal of Computational and Applied Mathematics*, 45(1):213 – 225, 1993a. ISSN 0377-0427. doi:https://doi.org/10.1016/0377-0427(93)90276-H. URL <http://www.sciencedirect.com/science/article/pii/037704279390276H>.
- GJ Cooper and JC Butcher. An iteration scheme for implicit runge—kutta methods. *IMA journal of numerical analysis*, 3(2):127–140, 1983.
- Ernst Hairer and Gerhard Wanner. *Construction of Implicit Runge-Kutta Methods*, pages 71–90. Springer Berlin Heidelberg, Berlin, Heidelberg, 1996b. ISBN 978-3-642-05221-7. doi:10.1007/978-3-642-05221-7\_5. URL [https://doi.org/10.1007/978-3-642-05221-7\\_5](https://doi.org/10.1007/978-3-642-05221-7_5).
- John Charles Butcher and Gerhard Wanner. Runge-kutta methods: some historical notes. *Applied Numerical Mathematics*, 22(1-3):113–151, 1996.
- Brynjulf Owren and Marino Zennaro. Derivation of efficient, continuous, explicit runge–kutta methods. *SIAM journal on scientific and statistical computing*, 13(6):1488–1501, 1992.
- David Ketcheson and Aron Ahmadia. Optimal stability polynomials for numerical integration of initial value problems. *Communications in Applied Mathematics and Computational Science*, 7(2):247–271, 2013.
- Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
- Assyr Abdulle. On roots and error constants of optimal stability polynomials. *BIT Numerical Mathematics*, 40(1): 177–182, 2000.
- Assyr Abdulle. Fourth order chebyshev methods with recurrence relation. *SIAM Journal on Scientific Computing*, 23(6):2041–2054, 2002.
- Andrei Borisovich Bogatyrev. Effective solution of the problem of the optimal stability polynomial. *Sbornik: Mathematics*, 196(7):959, 2005.
- Ingemar P.E. Kinnmark and William G. Gray. One step integration methods with maximum stability regions. *Mathematics and Computers in Simulation*, 26(2):87–92, 1984a. ISSN 0378-4754. doi:https://doi.org/10.1016/0378-4754(84)90039-9. URL <https://www.sciencedirect.com/science/article/pii/0378475484900399>.
- Ingemar P.E. Kinnmark and William G. Gray. One step integration methods of third-fourth order accuracy with large hyperbolic stability limits. *Mathematics and Computers in Simulation*, 26(3):181–188, 1984b. ISSN 0378-4754. doi:https://doi.org/10.1016/0378-4754(84)90056-9. URL <https://www.sciencedirect.com/science/article/pii/0378475484900569>.
- JL Mead and RA Renaut. Optimal runge–kutta methods for first order pseudospectral operators. *Journal of Computational Physics*, 152(1):404–419, 1999.
- Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems*, pages 6571–6583, 2018.
- Amir Gholami, Kurt Keutzer, and George Biros. Anode: Unconditionally accurate memory-efficient gradients for neural odes, 2019.
- Tianjun Zhang, Zhewei Yao, Amir Gholami, Kurt Keutzer, Joseph Gonzalez, George Biros, and Michael Mahoney. Anodev2: A coupled neural ode evolution framework, 2019.
- M. Raissi, P. Perdikaris, and G. E. Karniadakis. Multistep Neural Networks for Data-driven Discovery of Nonlinear Dynamical Systems. *ArXiv e-prints*, January 2018.
- Emmanuel de Bezenac, Arthur Pajot, and Patrick Gallinari. Deep learning for physical processes: Incorporating prior scientific knowledge. *CoRR*, abs/1711.07970, 2017. URL <http://arxiv.org/abs/1711.07970>.
- Edward N. Lorenz. Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, March 1963. ISSN 0022-4928. doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2. URL [http://journals.ametsoc.org/doi/abs/10.1175/1520-0469\(1963\)020%3C0130:DNF%3E2.0.CO;2](http://journals.ametsoc.org/doi/abs/10.1175/1520-0469(1963)020%3C0130:DNF%3E2.0.CO;2).
- Edward N Lorenz. Predictability: A problem partly solved. In *Proc. Seminar on predictability*, volume 1, 1996.
- A. C. Hindmarsh. ODEPACK, a systematized collection of ODE solvers. *IMACS Transactions on Scientific Computation*, 1:55–64, 1983.
- Bernard O Koopman. Hamiltonian systems and transformation in hilbert space. *Proceedings of the national academy of sciences of the united states of america*, 17(5):315, 1931.



- Steven L Brunton, Bingni W Brunton, Joshua L Proctor, Eureka Kaiser, and J Nathan Kutz. Chaos as an intermittently forced linear system. *Nature communications*, 8(1):1–9, 2017.
- R. V. Gamkrelidze L. S. Pontryagin, V. G. Boltyanskii. *The mathematical theory of optimal processes*. Interscience publishers, a division of John Wiley and sons, New York London, 1962.
- PJ Van der Houwen. Explicit runge-kutta formulas with increased stability boundaries. *Numerische Mathematik*, 20(2): 149–164, 1972.
- Ingemar PE Kinnmark and William G Gray. One step integration methods with maximum stability regions. *Mathematics and Computers in Simulation*, 26(2):87–92, 1984c.
- GJ Cooper and R Vignesvaran. Some schemes for the implementation of implicit runge-kutta methods. *Journal of Computational and Applied Mathematics*, 45(1-2):213–225, 1993b.

## Acknowledgement

This work was supported by Labex Cominlabs (grant SEACS), CNES (grant OSTST-MANATEE), Microsoft (AI EU Ocean awards), ANR Projects Melody and OceaniX. It benefited from HPC and GPU resources from Azure (Microsoft EU Ocean awards) and GENCI-IDRIS (Grant 2020-101030).

## A Integration schemes parameters

We present in this section the integration coefficients *i.e.* the matrix  $a$  and the vectors  $b$  and  $c$  of the trained integration schemes of section 5.2. Classically, Runge-Kutta coefficients are arranged in a Butcher table. The general form of this table is given in table 3.

c	a
b	b

Table 3: *Butcher table.*

As an illustrative example, table 4 resumes the well known Runge-Kutta 4 method.

0	0	0	0	0
1/2	1/2	0	0	0
1/2	0	1/2	0	0
1	0	0	1	0
	1/6	1/3	1/3	1/6

Table 4: *Runge-Kutta 4 parameters.*

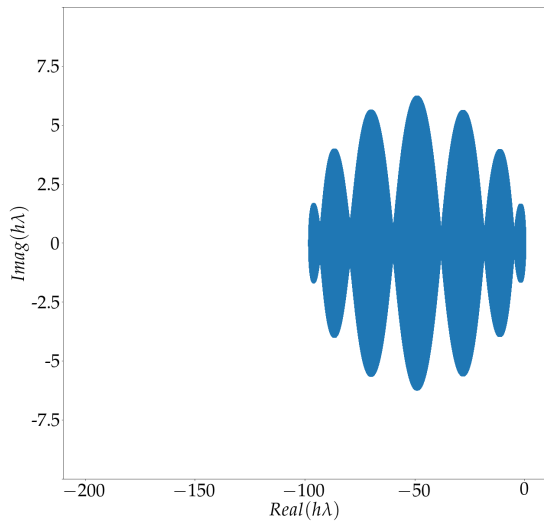
Tables 5, 6 and 7 highlight the parameters of the integration schemes  $ADRK_{h_1}$ ,  $ADRK_{h_2}$  and  $ADRK_{h_3}$  respectively trained on the Lorenz 63 data sampled from  $h = 0.2$  to  $h = 0.4$ .

0	0	0	0	0	0	0	0	0	0	0	0	0
0.00295	0.00295	0	0	0	0	0	0	0	0	0	0	0
0.00299	2.17592	-2.17293	0	0	0	0	0	0	0	0	0	0
0.00300	2.98541	-2.17397	-0.80844	0	0	0	0	0	0	0	0	0
0.24293	-1.48213	3.52131	-0.92958	-0.86667	0	0	0	0	0	0	0	0
0.24293	-1.20813	0.81761	0.44993	0.09160	0.09192	0	0	0	0	0	0	0
0.32598	0.68880	-0.37282	-0.24275	-0.02117	0.06104	0.21287	0	0	0	0	0	0
0.49628	0.28836	0.04642	-0.17632	-0.01809	-0.01089	0.07776	0.28904	0	0	0	0	0
0.71550	0.05976	-0.02908	0.45015	-0.16537	-0.08716	-0.16371	0.10869	0.54223	0	0	0	0
0.83650	0.11298	0.23727	-0.32220	-0.10863	0.21937	0.49812	-0.42969	0.34843	0.28086	0	0	0
1	-0.52891	0.28477	0.30886	-0.10163	0.19143	0.33904	0.06679	0.08134	0.14083	0.21750	0	0
	0.00013	0.00013	0.09499	0.00011	0.00033	0.03125	0.39447	0.06335	0.27661	0.02192	0.11670	0

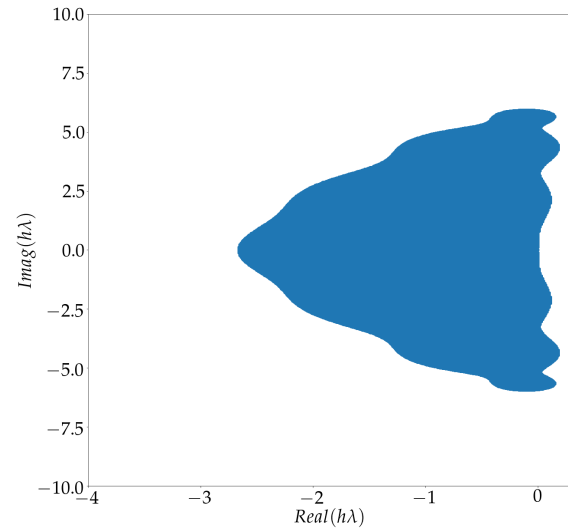
Table 5:  $ADRK_{h_1}$  parameters.

## B Stability-constrained ADRK schemes, additional experiments

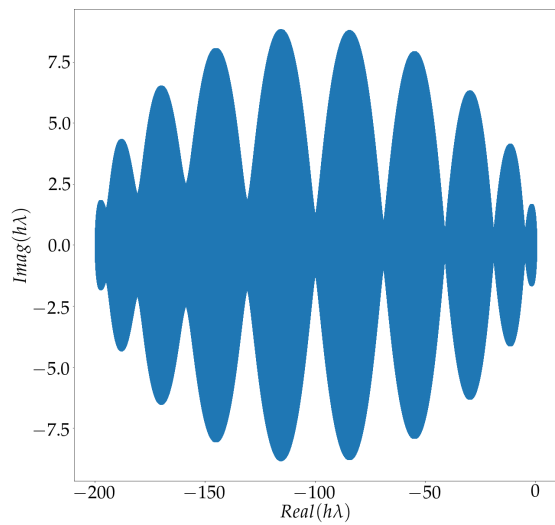
We complement the stability-constrained ADRK schemes experiment with two similar case studies based on an  $ADRK_7$  and an  $ADRK_{10}$ . Specifically, both the Runge-Kutta schemes are optimized to reach the maximum possible value of  $\lambda h = 2q^2$  for the real axis inclusion case study Figure 16.



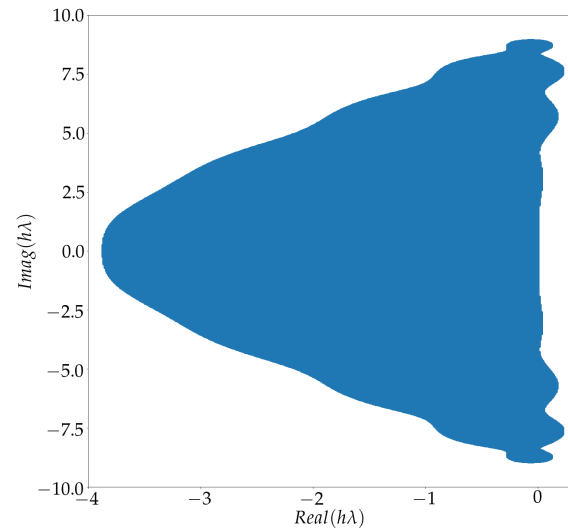
(a) 7-stage ADRK, Real axis inclusion



(b) 7-stage ADRK, Imaginary axis inclusion



(c) 10-stage ADRK, Real axis inclusion



(d) 10-stage ADRK, Imaginary axis inclusion

Figure 16: **Two dimensional stability region of the stability-constrained ADRK integration schemes for several number of stages:** The tested, stability-constrained, ADRK schemes reach the stability limit of both real and imaginary axis inclusion case-studies.

0	0	0	0	0	0	0	0	0	0	0	0
0.09168	0.09168	0	0	0	0	0	0	0	0	0	0
0.09227	-0.00598	0.09825	0	0	0	0	0	0	0	0	0
0.16187	-0.23044	0.00477	0.38754	0	0	0	0	0	0	0	0
0.16204	-0.25781	-0.20551	0.79426	-0.16890	0	0	0	0	0	0	0
0.43033	-0.24147	0.53237	-0.20302	0.46703	-0.12458	0	0	0	0	0	0
0.43142	1.16600	-0.60220	-0.35233	-0.56929	0.27263	0.51660	0	0	0	0	0
0.60603	0.34099	-0.52091	0.27404	-0.25958	0.25545	0.38264	0.13340	0	0	0	0
0.64954	0.16978	-0.07287	0.06919	-0.17928	0.20630	0.26679	0.03296	0.15667	0	0	0
0.84016	0.36515	-0.36786	0.47056	-0.52580	0.03298	0.45605	-0.08132	-0.12110	0.61150	0	0
1	-0.11475	0.36853	0.01243	0.06163	-0.10538	0.43067	-0.02148	0.07745	0.14748	0.14341	0
	0.00010	0.00007	0.22873	0.00013	0.00020	0.27106	0.00003	0.00011	0.36341	0.04983	0.08633

Table 6:  $ADRK_{h_2}$  parameters.

0	0	0	0	0	0	0	0	0	0	0	0
0.13282	0.13282	0	0	0	0	0	0	0	0	0	0
0.18403	-0.01533	0.19936	0	0	0	0	0	0	0	0	0
0.28433	0.06830	0.03882	0.17721	0	0	0	0	0	0	0	0
0.37687	0.05088	0.04762	0.07653	0.20184	0	0	0	0	0	0	0
0.49570	0.05444	0.09223	0.08569	0.06398	0.19936	0	0	0	0	0	0
0.65230	0.10848	0.12104	0.08825	0.09688	0.01685	0.22079	0	0	0	0	0
0.66402	-0.00214	0.11741	0.10766	0.07588	0.12490	0.06258	0.17773	0	0	0	0
0.78754	0.03866	0.07149	0.12547	0.10885	0.07602	0.10123	0.06464	0.20117	0	0	0
0.85205	0.05165	0.10963	0.06966	0.11010	0.09470	0.14948	0.07204	0.10591	0.08888	0	0
1	0.02863	0.11511	0.05827	0.16788	0.08683	0.15047	0.02359	0.10025	-0.09581	0.36479	0
	0.04237	0.10164	0.08155	0.12025	0.09046	0.13617	0.06988	0.11978	0.03882	0.14815	0.05093

Table 7:  $ADRK_{h_3}$  parameters.