# Formal lumping of polynomial differential equations through approximate equivalences

Cardelli, Luca; Squillace, Giuseppe; Tribastone, Mirco; Tschaikowski, Max; Vandin, Andrea

*Published in:*
Journal of Logical and Algebraic Methods in Programming

*DOI (link to publication from Publisher):*
10.1016/j.jlamp.2023.100876



*Publication date:*
2023

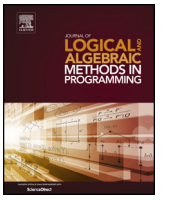
*Document Version*
Publisher's PDF, also known as Version of record

Journal of Logical and Algebraic Methods in Programming 134 (2023) 100876

Contents lists available at ScienceDirect

Journal of Logical and Algebraic Methods in Programming

journal homepage: www.elsevier.com/locate/jlamp



# Formal lumping of polynomial differential equations through approximate equivalences

Luca Cardelli [a], Giuseppe Squillace [b,*], Mirco Tribastone [b], Max Tschaikowski [c], Andrea Vandin [d,e]

[a] University of Oxford, UK
[b] IMT School for Advanced Studies Lucca, Italy
[c] Aalborg University, Denmark
[d] Sant'Anna School for Advanced Studies, Pisa, Italy
[e] DTU Technical University of Denmark, Denmark

ARTICLE INFO

*Article history:*
Received 19 June 2022
Received in revised form 24 February 2023
Accepted 7 May 2023
Available online 12 May 2023

*Keywords:*
Lumping
Model reduction
Polynomial differential equations
Equivalence relations

ABSTRACT

It is well known that exact notions of model abstraction and reduction for dynamical systems may not be robust enough in practice because they are highly sensitive to the specific choice of parameters. In this paper we consider this problem for nonlinear ordinary differential equations (ODEs) with polynomial derivatives. We introduce a model reduction technique based on *approximate differential equivalence*, i.e., a partition of the set of ODE variables that performs an aggregation when the variables are governed by nearby derivatives. We develop algorithms to (i) compute the largest approximate differential equivalence; (ii) construct an approximately reduced model from the original one via an appropriate perturbation of the coefficients of the polynomials; and (iii) provide a formal certificate on the quality of the approximation as an error bound, computed as an over-approximation of the reachable set of the reduced model. Finally, we apply approximate differential equivalences to case studies on electric circuits, biological models, and polymerization reaction networks.


## 1. Introduction

Ordinary differential equations (ODEs) are a fundamental dynamical model across many branches of science and engineering. In particular, ODEs with polynomial right-hand sides are prominent for the study of nonlinear interactions occurring under the assumption of opportunistic contacts between classes of interacting populations, such as in biology [1], chemistry [2], and ecology [3] but can also be used to encode electric circuits [4] and control systems [5]. Obviously, polynomial ODE systems cover linear ODEs as special cases.

Since closed-form solutions of initial value problems with polynomial ODEs (abbreviated as PIVPs throughout the remainder of this paper) are available only in special cases, a major problem when dealing with complex models regards the computational cost of the analysis, which generally is conducted by means of numerical integration. This problem has spurred a considerable amount of cross-disciplinary research on model reduction (e.g., [6,7]). Under this term one generally

* Corresponding author.
*E-mail addresses:* luca.cardelli@cs.ox.ac.uk (L. Cardelli), giuseppe.squillace@imtlucca.it (G. Squillace), mirco.tribastone@imtlucca.it (M. Tribastone), tschaikowski@cs.aau.dk (M. Tschaikowski), andrea.vandin@santannapisa.it (A. Vandin).

https://doi.org/10.1016/j.jlamp.2023.100876

refers to (automatic) techniques that provide a lower-dimensional representation of a given ODE system which retains an appropriate projection of its dynamics through a reduced cost of analysis.

Roughly speaking, model reduction methods can be categorized into two main classes: exact methods provide a reduced model which introduces no error in projected dynamics that it preserves (e.g., the early work by Aoki [8]). Approximate methods, on the other hand, may introduce errors; depending on the specific technique, these errors may be equipped with formal bounds (e.g., [9] and references therein). Among the types of model-reduction methods including time-scale separation, balanced truncation, and singular value decomposition (see [6,2,7] for general reviews on this subject), *lumping* is an established concept (e.g., [10]) whereby each variable in the reduced model (the *macro-variable*) is associated with an ODE giving the dynamics of a (linear or nonlinear) combination of the ODE variables in the original system.

Here we are concerned with a specific class of lumping induced by a so-called *differential equivalence*—i.e., a partition of the set of variables in a given PIVP—such that each macro-variable represents the sum of the variables in a partition block [11]. Specifically, we consider both *backward* and *forward* differential equivalence (abbreviated BDE and FDE, respectively). In the former case, variables in the same block have exactly the same solution if starting from the same initial condition; in the latter case, each macro-variable exactly represents the sum of the original variables, but from the reduced model one cannot recover the solution of an individual original variable in general.

It has been shown that, for both equivalences, one can prove the existence of a maximal exact reduction (or equivalently, of a coarsest partition) [11], which can be computed by building on fundamental algorithmic results in computer science related to partition-refinement algorithms [12]. In particular, restricting to polynomial derivatives enables the computation of differential equivalences on a suitable encoding of the ODEs into a hyper-graph akin to a formal chemical reaction network [13]. This finitary encoding ultimately allows for algorithms that efficiently run in polynomial time and space with respect to the size of the original ODE system, related to the number of variables and of monomials appearing in the right-hand sides [14]. (This is in contrast to significantly more expensive symbolic checks for a more general class of nonlinear ODE systems studied in [11,15].)

Although exact reduction methods prove effective in a variety of models (e.g., performance engineering [16–19], biochemistry [20–22] or physical engineering [23,24]), especially when they have "structural" symmetries that are not dependent on specific values of the parameters, they may not be effective in applications domains where parameter uncertainty, error tolerances, and calibration from finite-precision measurements are common, such as in biology or engineering [7,25]. To cope with this challenge, in this paper we introduce approximate variants of differential equivalence for polynomial ODE systems. Our main idea is to consider a threshold parameter $\varepsilon \geq 0$, which intuitively captures perturbations in polynomials coefficients. This allows relating ODE variables in the same partition block that would be distinct otherwise. The extension is conservative in the sense that the case $\varepsilon = 0$ corresponds to an exact differential equivalence of [14]. In addition to defining criteria for approximate differential equivalences, we provide an algorithm for obtaining the maximal one, still in a partition-refinement framework.

The reduction is represented as a *reference model*, obtained through a perturbation of the coefficients of the original model which makes the given approximate differential equivalence an exact one. By considering a metric (the Euclidean norm) to measure the degree of perturbation, the reference model is the one which minimizes such perturbation. This can be done efficiently by solving an optimization problem which runs polynomially with the size of the ODE system [26]. This approach is analogous to optimal approximate lumping for Markov chains (e.g., [27]), although our theory can be applied to other choices of reference models.

The bound of the error produced by the reference model with respect to the original system can be computed by studying the reachable set of the reference model from an uncertain set of initial conditions that covers the applied perturbation. Therefore, this reachable set becomes a formal bound that relates the reference model to the original model. Section 4 presents a bound which relies on a linearization of the reference model (which can be efficiently computed in the case of polynomial ODE systems). First, we bound the reachable set of the linearized model using closed form solutions, similarly to [28]. Then, we provide a conservative condition (i.e., an over-approximation) that ensures that the linearized model describes the original nonlinear behavior dynamics well. Our bound is given in terms of an $\varepsilon$-$\delta$ argument (similar in spirit to the ones routinely used in calculus). Informally, it states the following: for any choice of the tolerance $\varepsilon$, there exist a degree of perturbation $\delta$ and an *amplifier* $\lambda$ such that, for any ODE system obtained by applying a perturbation to the reference model of at most $\delta$, at all time points the difference between the solution of the reference model and the perturbed one is at most $\lambda$ times the perturbation.

Being based on a linearization, it is perhaps not surprising that the as-computed $\delta$ will account only for *small* perturbations of the parameters. Yet numerical experiments in Section 5 show that these can be enough to explain quasi-symmetric behavior due to parametric tolerances in components of real electric network designs [29]. We show that our bounding technique can complement a state-of-the-art overapproximation technique, CORA [30,31], in that it can scale to larger systems while being more conservative in the size of the initial uncertain set that it supports.

*Further related work* Differential equivalence is related but not comparable to the notion of bisimulation for differential systems [32,20] since it partitions ODE variables rather than the state space. Likewise, it complements [23] in that captures nonlinear relations between ODE variables but does not enjoy a polynomial time algorithm like [14]. Instead, classic approximation approaches relying on Lyapunov-like functions [33,34] may provide tight bounds but their automatic computation remains a challenging task. A restriction to special classes of Lyapunov-like functions (e.g., sum-of-squares polynomials [32]),

instead, leads to efficient construction algorithms which may provide tight bounds, but existence is not guaranteed. On the other hand, approaches based on monotonic systems and differential inequalities [35] are more efficient, the current approach can be expected to provide tighter bounds when applicable. *Abstraction*, supported e.g. by CORA, locally approximates the nonlinear model by a multivariate polynomial or an affine system, see [36,37] and references therein. In that spirit, we linearize across a reference trajectory, a concept known also from gain scheduling [38]. A closer approach to ours is discussed in [39]. It combines local Lyapunov-like functions and techniques based on sensitivity analysis [28]. Our bound is however different because the nonlinear part is bounded analytically by restricting to polynomial derivatives.

More in general, research on approximate reduction of ODE systems spans many disciplines. In chemistry, it can be traced back to Kuo and Wei [40]. They studied monomolecular reaction networks, which give rise to affine ODE systems. The approximation consists in *nearly exact lumping*, i.e., a linear transformation of the state space that would be exact up to a perturbation of the parameters (hence we are similar in spirit). The approximation, however, only applies when the transition matrix underlying the linear system is diagonalizable. Li and Rabitz extend approximate lumping to general CRNs [41], but an explicit error bound is not given. In a similar vein, approximate quotients in ecology have been studied from the point of view of finding a reduced ODE system whose derivatives are as close as possible (in norm) to the derivatives of the original ODE system, where the 0-distance induce the exact quotient [42]. The justification that variables underlying similar ODEs have nearby solutions is grounded on Gronwall's inequality which is also at the basis of more recent quotient constructions [43,44], which however are not algorithmic.

*Relationship with [45]* This work extends the conference paper [45] by presenting full detail of all the proofs and a more extensive evaluation section covering biological, cyber-physical and chemical case studies in addition to the electric circuits already analyzed in [45].

## 2. Background

In this section we introduce the basic background useful for the rest of the paper. Let us consider a polynomial initial value problem (PIVP) over the set of ODE variables $\mathcal{S} = \{x_1, \ldots, x_n\}$. We let $x_i(t)$ be the unique solution for variable $x_i$ at time point $t$. The initial conditions are given by $\sigma : \mathcal{S} \to \mathbb{R}$ such that $x_i(0) = \sigma(x_i)$. We assume to work with PIVPs that do not exhibit explosion in finite time, i.e., they admit solutions on arbitrary long time intervals. This assumption is not restrictive because the majority of practical models satisfy this property. In any case this could be checked via numerical ODE solvers.

The PIVP is composed of ODEs in the form $\dot{x}_i = q_i$, $1 \le i \le n$, where $q_i$ is a multivariate polynomial over $\mathcal{S}$. We say that a PIVP satisfies the *normal form* when each monomial $x^\alpha \equiv \prod_{x_i \in \mathcal{S}} x_i^{\alpha_{x_i}}$, where $\alpha \in \mathbb{N}_0^{\mathcal{S}}$ is a multi-index, appears in $q_i$ at most once. Without loss of generality, we assume that the polynomials $q_i$ of in the PIVPs are in normal form and we indicate this with the notation $\mathcal{N}(q_i)$.

Let $c(q_i, x^\alpha)$ denote the coefficient of the monomial $x^\alpha$ in a normal form polynomial $q_i$ with variables in $\mathcal{S}$, where $\alpha \in \mathbb{N}_0^{\mathcal{S}}$.

**Example 1.** We use the following ODE system, with variables $\mathcal{S} = \{x_1, x_2, x_3\}$, as a running example.

$$\dot{x}_1 = -4.00x_1 + x_2 + x_3$$
$$\dot{x}_2 = 1.99x_1 - x_2 \tag{1}$$
$$\dot{x}_3 = 2.01x_1 - x_3 \tag{2}$$

BDE and FDE were originally provided in [11] for a class of nonlinear ODE systems covering derivatives more general than polynomials. Here we restate these notions for an PIVP. (The proofs for this correspondence are straightforward hence we omit them.)

BDE assures that variables in the same equivalence class have the same solutions at all time points. These relations are built by making pairwise comparisons between the coefficients of the polynomials related to any two variables in the same equivalence class.

**Definition 1** *(Backward differential equivalence (BDE)).* Fix a PIVP, a partition $\mathcal{H}$ of $\mathcal{S}$ and write $x_i \sim^B_{\mathcal{H}} x_j$ if all coefficients of the following polynomial are zero,

$$\wp^{\mathcal{H}}_{i,j} := (q_i - q_j)\left[x_{\langle H',1\rangle}/x_{H'}, \ldots, x_{\langle H',|H'|\rangle}/x_{H'} : H' \in \mathcal{H}\right]$$

i.e., when

$$\sum_{\alpha \in \mathbb{N}_0^{\mathcal{S}}} |c(\wp^{\mathcal{H}}_{i,j}, x^\alpha)| = 0. \tag{3}$$

A partition $\mathcal{H}$ is a BDE if $\mathcal{H} = \mathcal{S}/(\sim^{B^*}_{\mathcal{H}} \cap \sim_{\mathcal{H}})$.

The definition states that a candidate partition is a BDE if for any two variables in same block, the differences between the coefficients on the same monomials are zero.

**Example 2.** In our running example let us consider the partition of variables $\mathcal{H} = \{H_1, H_2\}$, with $H_1 = \{x_1\}$ and $H_2 = \{x_2, x_3\}$. Then $\mathcal{H}$ is not a BDE because

$$\wp_{2,3}^{\mathcal{H}} = -0.02x_1 \qquad \text{and} \qquad c(\wp_{2,3}^{\mathcal{H}}, x_1) = -0.02 \neq 0.$$

FDE relates sums of variables. More in detail, it identifies a partition that induces a quotient ODE that tracks sums of variables in each equivalence class. This can be done by replacing any two variables in the same equivalence class with their sum.

**Definition 2** *(Forward differential equivalence (FDE)).* Fix a PIVP, a partition $\mathcal{H}$ of $\mathcal{S}$ and write $x_i \sim_{\mathcal{H}}^{F} x_j$ if all coefficients of the polynomial $\sum_{H\in\mathcal{H}} \wp_{i,j}^{H}$ are zero, where

$$\wp_{i,j}^{H} := \sum_{x_k\in H} q_k - \sum_{x_k\in H} q_k[x_i/s(x_i + x_j), x_j/(1-s)(x_i + x_j)]$$

and $s$ is an auxiliary variable which does not denote any state. That is, when

$$\sum_{k=1}^{m} \sum_{\alpha\in\mathbb{N}_0^{\mathcal{S}\dot\cup\{s\}}} |c(\wp_{i,j}^{H_k}, x^\alpha)| = 0. \tag{4}$$

$\mathcal{H} = \{H_1, \ldots, H_m\}$ is an FDE when $\mathcal{H} = \mathcal{S}/(\sim_{\mathcal{H}}^{F^*} \cap \sim_{\mathcal{H}})$.

For a PIVP, FDE can be checked by requiring that the evaluation of the polynomial that represents the quotient derivative for an equivalence class is invariant with respect to a redistribution of the values of any two variables within that equivalence class.

**Example 3.** Let us consider the partition $\{\{x_1\}, \{x_2, x_3\}\}$ in our example, it is an FDE because we can find the following system for $x_2 + x_3$:

$$\begin{aligned} \dot{x}_1 &= -4.00x_1 + (x_2 + x_3) \\ (x_2 \dot{+} x_3) &= 4.00x_1 - (x_2 + x_3) \end{aligned}$$

Replacing the variable $x_{23}$ with $x_2 + x_3$ we obtain the quotient ODE

$$\begin{aligned} \dot{x}_1 &= -4.00x_1 + x_{23} \\ \dot{x}_{23} &= 4.00x_1 - x_{23} \end{aligned}$$

Thus we can conclude that the solution satisfies $x_{23}(t) = x_2(t) + x_3(t)$ for all times $t$ if this holds for the initial condition, i.e., $x_{23}(0) = x_2(0) + x_3(0)$.

## 3. Approximate differential equivalences

In this section we present two different approximate versions of BDE and FDE. In an approximate differential equivalence, we allow the conditions (3)-(4) to be satisfied with a certain amount of tolerance.

**Definition 3** *(Approximate BDE).* Fix a PIVP, a partition $\mathcal{H} = \{H_1, \ldots, H_m\}$ of $\mathcal{S}$, and $\varepsilon \geq 0$. We write $x_i \sim_{\mathcal{H},\varepsilon}^{B} x_j$ if $\sum_{\alpha\in\mathbb{N}_0^{\mathcal{S}}} |c(\wp_{i,j}^{\mathcal{H}}, x^\alpha)| \leq \varepsilon$, where $\wp_{i,j}^{\mathcal{H}}$ is as in Definition 1. A partition $\mathcal{H}$ is an $\varepsilon$-BDE if $\mathcal{H} = \mathcal{S}/(\sim_{\mathcal{H},\varepsilon}^{B^*} \cap \sim_{\mathcal{H}})$.

**Definition 4** *(Approximate FDE).* Fix a PIVP, a partition $\mathcal{H} = \{H_1, \ldots, H_m\}$ of $\mathcal{S}$, and $\varepsilon \geq 0$. We write $x_i \sim_{\mathcal{H},\varepsilon}^{F} x_j$ if $\sum_{k=1}^{m} \sum_{\alpha\in\mathbb{N}_0^{\mathcal{S}\dot\cup\{s\}}} |c(\wp_{i,j}^{H_k}, x^\alpha)| \leq \varepsilon$, where $\wp_{i,j}^{H}$ is as in Definition 2. A partition $\mathcal{H}$ is an $\varepsilon$-FDE when $\mathcal{H} = \mathcal{S}/(\sim_{\mathcal{H},\varepsilon}^{F^*} \cap \sim_{\mathcal{H}})$.

In the definitions above $*$ represents the transitive closure of this relation while $B$ and $F$ stand, respectively, for BDE and FDE. Setting $\varepsilon = 0$ recovers the exact counterparts in both cases. That is, $\mathcal{H}$ is an BDE (resp., FDE) partition if and only if $\mathcal{H}$ is a 0-BDE (resp., 0-FDE) partition. The two approximate differential equivalences are not comparable since their exact counterparts are not [11]. Since these two notions have similar structure, in the rest of this paper we will illustrate only approximate BDE using simple examples. Instead, both notions will be discussed in more detail for the numerical evaluation of Section 5.

**Example 4.** Let us consider our running example (1). Then, the partition $\{\{x_1\}, \{x_2, x_3\}\}$ is a 0.02-BDE partition, as can be easily seen from Example 2.

The next two theorems are concerned with providing an algorithm to compute an approximate differential equivalence. Specifically, Theorem 1 shows the existence of the largest approximate differential equivalence; Theorem 2 proves the correctness of the partition-refinement algorithm to compute it as the coarsest refinement of a given initial partition of variables.

**Theorem 1.** *Fix a PIVP, a partition $\mathcal{G}$ of $\mathcal{S}$, and $\varepsilon \geq 0$. Then, there exists a unique coarsest $\varepsilon$-FDE ($\varepsilon$-BDE) partition refining $\mathcal{G}$.*

**Proof.** Assume that $\mathcal{H}_1, \ldots, \mathcal{H}_n$ are $\varepsilon$-FDE partitions of $\mathcal{S}$ and define $\sim_l := \sim^F_{\mathcal{H}_l,\varepsilon} \cap \sim_{\mathcal{H}_l}$ and $\sim := \sim_{\mathcal{H}}$, where $\mathcal{H} := \mathcal{S}/\big(\bigcup_{l=1}^n \sim_l\big)^*$. Note that the definition of $\varepsilon$-FDE implies that $\mathcal{H}_l = \mathcal{S}/\sim_l^*$ for all $1 \leq l \leq n$.

Let us fix arbitrary $1 \leq l \leq n$ and $x_i \sim_l x_j$. It can be easily seen that for any $H \in \mathcal{H}$ there exist unique blocks $G_1^H, \ldots, G_{m_H}^H \in \mathcal{H}_l$ such that $\biguplus_{k=1}^{m_H} G_k^H = H$. With this, it holds that

$$\begin{aligned}
\sum_{H\in\mathcal{H}} \sum_{\alpha\in\mathbb{N}_0^{\mathcal{S}\dot\cup\{s\}}} |c(\wp_{i,j}^H, x^\alpha)| &= \sum_{H\in\mathcal{H}} \sum_{\alpha\in\mathbb{N}_0^{\mathcal{S}\dot\cup\{s\}}} |c(\sum_{k=1}^{m_H} \wp_{i,j}^{G_k^H}, x^\alpha)| \\
&\leq \sum_{H\in\mathcal{H}} \sum_{k=1}^{m_H} \sum_{\alpha\in\mathbb{N}_0^{\mathcal{S}\dot\cup\{s\}}} |c(\wp_{i,j}^{G_k^H}, x^\alpha)| \\
&= \sum_{G\in\mathcal{H}_l} \sum_{\alpha\in\mathbb{N}_0^{\mathcal{S}\dot\cup\{s\}}} |c(\wp_{i,j}^{G}, x^\alpha)| \\
&\leq \varepsilon,
\end{aligned}$$

where the first estimation follows from the triangle inequality, while the second estimation is thanks to the definition of $\sim_l$. The above readily implies that $\mathcal{H}$ is an $\varepsilon$-FDE partition.

We now turn to the case of $\varepsilon$-BDE. Similarly to the $\varepsilon$-FDE case, we assume that $\mathcal{H}_1, \ldots, \mathcal{H}_n$ are $\varepsilon$-BDE partitions of $\mathcal{S}$ and $\sim_l$, $\sim := \sim_{\mathcal{H}}$ and $\mathcal{H}$ are as above. Note that the definition of $\varepsilon$-BDE implies that $\mathcal{H}_l = \mathcal{S}/\sim_l^*$ for all $1 \leq l \leq n$. For arbitrary $1 \leq l \leq n$ and $x_i \sim_l x_j$, it holds that

$$\sum_{\alpha\in\mathbb{N}_0^{\mathcal{S}}} |c(\wp_{i,j}^{\mathcal{H}}, x^\alpha)| \leq \sum_{\alpha\in\mathbb{N}_0^{\mathcal{S}}} |c(\wp_{i,j}^{\mathcal{H}_l}, x^\alpha)|$$

since $\mathcal{H}_l$ refines $\mathcal{H}$. The above implies that $\mathcal{H}$ is an $\varepsilon$-BDE partition.

So far, we have shown that the coarsening $\mathcal{S}/\big(\bigcup_{l=1}^n \sim_l\big)^*$ of $\varepsilon$-FDE/BDE partitions $\mathcal{H}_1, \ldots, \mathcal{H}_n$ is again an $\varepsilon$-FDE/BDE partition. The claim follows by noting that Lemma 26 in [46] ensures that $\mathcal{S}/\big(\bigcup_{l=1}^n \sim_l\big)^*$ is a refinement of $\mathcal{G}$ if each $\mathcal{S}/\sim_l$ is a refinement of $\mathcal{G}$. □

The following lemma will be needed in the proof of Theorem 2.

**Lemma 1.** *Let $\mathcal{G}, \mathcal{H}$ be two partitions of $\mathcal{S}$. Then, for any $\varepsilon > 0$, the following can be shown.*

*i)* $x_i \sim^{F^*}_{\mathcal{H},\varepsilon} x_j$ *implies* $x_i \sim^{F^*}_{\mathcal{G},\varepsilon} x_j$ *if $\mathcal{H}$ is a refinement of $\mathcal{G}$.*
*ii)* $x_i \sim^{B^*}_{\mathcal{H},\varepsilon} x_j$ *implies* $x_i \sim^{B^*}_{\mathcal{G},\varepsilon} x_j$ *if $\mathcal{H}$ is a refinement of $\mathcal{G}$.*

**Proof.** Let us assume that $x_i \sim^F_{\mathcal{H},\varepsilon} x_j$, which is equivalent to

$$\sum_{H\in\mathcal{H}} \sum_{\alpha\in\mathbb{N}_0^{\mathcal{S}\dot\cup\{s\}}} |c(\wp_{i,j}^H, x^\alpha)| \leq \varepsilon$$

**Algorithm 1** Template partition refinement algorithm for the computation of the coarsest $\varepsilon$-FDE/$\varepsilon$-BDE partition that refines a given initial partition $\mathcal{G}$.

**Require:** A PIVP over variables $\mathcal{S}$, a partition $\mathcal{G}$ of $\mathcal{S}$, a threshold $\varepsilon \geq 0$, and a mode $\chi \in \{F, B\}$.
  $\mathcal{H} \longleftarrow \mathcal{G}$
  **while true do**
    $\mathcal{H}' \longleftarrow \mathcal{S}/(\sim^{\chi^*}_{\mathcal{H},\varepsilon} \cap \sim_{\mathcal{H}})$
    **if** $\mathcal{H}' = \mathcal{H}$ **then**
      **return** $\mathcal{H}$
    **else**
      $\mathcal{H} \longleftarrow \mathcal{H}'$
    **end if**
  **end while**

Since $\mathcal{H}$ is a refinement of $\mathcal{G}$, for any $G \in \mathcal{G}$ there exist unique blocks $H^G_1, \ldots, H^G_{m_G} \in \mathcal{H}$ such that $\biguplus_{k=1}^{m_G} H^G_k = G$. With this, it holds that

$$\begin{aligned}
\sum_{G\in\mathcal{G}} \sum_{\alpha\in\mathbb{N}_0^{\mathcal{S}\dot\cup\{s\}}} |c(\wp^G_{i,j}, x^\alpha)| &= \sum_{G\in\mathcal{G}} \sum_{\alpha\in\mathbb{N}_0^{\mathcal{S}\dot\cup\{s\}}} |c(\sum_{k=1}^{m_G} \wp^{H^G_k}_{i,j}, x^\alpha)| \\
&\leq \sum_{G\in\mathcal{G}} \sum_{k=1}^{m_G} \sum_{\alpha\in\mathbb{N}_0^{\mathcal{S}\dot\cup\{s\}}} |c(\wp^{H^G_k}_{i,j}, x^\alpha)| \\
&= \sum_{H\in\mathcal{H}} \sum_{\alpha\in\mathbb{N}_0^{\mathcal{S}\dot\cup\{s\}}} |c(\wp^{H}_{i,j}, x^\alpha)| \\
&\leq \varepsilon,
\end{aligned}$$

thus showing $x_i \sim^F_{\mathcal{G},\varepsilon} x_j$. This yields the first statement. Let us now assume that $x_i \sim^B_{\mathcal{H},\varepsilon} x_j$ which corresponds by definition to $\sum_{\alpha\in\mathbb{N}_0^{\mathcal{S}}} |c(\wp^{\mathcal{H}}_{i,j}, x^\alpha)| \leq \varepsilon$. Moreover,

$$\sum_{\alpha\in\mathbb{N}_0^{\mathcal{S}}} |c(\wp^{\mathcal{G}}_{i,j}, x^\alpha)| \leq \sum_{\alpha\in\mathbb{N}_0^{\mathcal{S}}} |c(\wp^{\mathcal{H}}_{i,j}, x^\alpha)|$$

because $\mathcal{H}$ is a refinement of $\mathcal{G}$. Hence, we infer that $x_i \sim^B_{\mathcal{G},\varepsilon} x_j$. This readily implies the second statement. □

**Theorem 2.** *Fix a PIVP, a partition $\mathcal{G}$ of $\mathcal{S}$, and $\varepsilon \geq 0$. Then, Algorithm 1 computes the coarsest $\varepsilon$-FDE ($\varepsilon$-BDE) that refines $\mathcal{G}$ if $\chi = F$ ($\chi = B$).*

**Proof.** Let $\mathcal{G}'$ denote the coarsest $\varepsilon$-FDE ($\varepsilon$-BDE) partition that refines $\mathcal{H}_0 := \mathcal{G}$ and set $\mathcal{H}_{k+1} := \mathcal{S}/(\sim^{\chi^*}_{\mathcal{H}_k,\varepsilon} \cap \sim_{\mathcal{H}_k})$ for all $k \geq 0$. Then, the sequence $(\mathcal{H}_k)_{k\geq 0}$ is such that $\mathcal{G}'$ is a refinement of $\mathcal{H}_k$ for all $k \geq 1$. We prove this by induction on $k$.

- $k = 1$: Since $\mathcal{G}'$ is a refinement of $\mathcal{H}_0$, Lemma 1 ensures the first claim.
- $k \to k+1$: Thanks to the fact that $\mathcal{G}'$ is a refinement of $\mathcal{H}_k$ by induction, Lemma 1 ensures the first claim.

From the fact that $\mathcal{G}'$ is a refinement of any $\mathcal{H}_k$, we conclude that $\mathcal{G}' = \mathcal{H}_k$ whenever $\mathcal{H}_k$ is an $\varepsilon$-FDE ($\varepsilon$-BDE) partition. Since $\mathcal{H}_k$ is a refinement of $\mathcal{H}_{k-1}$ for all $k \geq 1$ and $\mathcal{S}$ is finite, we can fix the smallest $k \geq 1$ such that $\mathcal{H}_k = \mathcal{H}_{k-1}$. This, in turn, implies that $\mathcal{H}_{k-1} = \mathcal{H}_k = \mathcal{S}/(\sim^{\chi^*}_{\mathcal{H}_{k-1},\varepsilon} \cap \sim_{\mathcal{H}_{k-1}})$. □

**Example 5.** We decided to apply the Algorithm 1 on the running example. For this purpose we set an $\varepsilon = 0.02$ and the initial partition $\mathcal{G} = \{x_1, x_2, x_3\}$. In the first iteration the partition was split in two blocks $\mathcal{H} = \{\{x_1\}, \{x_2, x_3\}\}$. In the next iteration $\mathcal{H}$ can not be split anymore and the algorithm return it as the 0.02-BDE partition.

We now study how efficiently the conditions for approximate differential equivalence can be computed. Since the reductions techniques are concerned with the coefficients of the polynomials, we define the complexity in terms of $\wp^H_{i,j}$ and $\wp^{\mathcal{H}}_{i,j}$. In the case of $\varepsilon$-FDE, we estimate an exponential complexity due to term replacement. Suppose to have the PIVP $\dot{x}_1 = x_2^k, \dot{x}_2 = x_1^k$, for some $k > 0$, then, for $\mathcal{H} = \{\{x_1, x_2\}\}$, the term $q_1[x_1/s(x_1+x_2), x_2/(1-s)(x_1+x_2)]$ will be of size $\mathcal{O}(2^k)$.

The $\varepsilon$-BDE case is different because the conditions involve a difference between polynomials terms with no term rewritings. This discussion can be formalized as follows.

**Theorem 3.** *There exists a polynomial $\Pi$ such that, under the assumptions of Theorem 2, the number of steps done by Algorithm 1 is $\mathcal{O}(\Pi(2^d \cdot p))$ if $\chi = F$ and $\mathcal{O}(\Pi(p))$ if $\chi = B$, respectively, where $d$ is the maximum degree of the polynomial and $p$ is the number of monomials present in the PIVP.*

**Proof.** Follows from the proof of Theorem 2. □

As already mentioned, the approximate differential equivalences are a relaxation of the exact counterparts; for this reason the above results provide a complexity bound for a subclass of ODE considered in [11].

In practice, $d$ is not large. Indeed, in the numerical evaluations in Section 5, $d$ was no larger than two. For instance, in a PIVP of a chemical reaction network with mass-action kinetics, one typically has $d = 2$ because, in nature, at most two species interact in a given reaction [47].

For the sake of completeness, we take into account another complexity measure. Another natural definition could be the maximal distance between derivatives "semantically", i.e., under all possible evaluations within a given domain of interest. For example, consider the PIVP $\dot{x}_1 = x_1^3 - x_2$, $\dot{x}_2 = x_1 - x_2^3$. Establishing that $\{\{x_1, x_2\}\}$ is an $\varepsilon$-BDE would require checking that the difference between the derivatives satisfies

$$|\dot{x}_1 - \dot{x}_2| = |x_1^3 - x_1 + x_2^3 - x_2| \leq \varepsilon, \tag{5}$$

for all $0 \leq x_1, x_2 \leq C$ and for some finite bound $C > 0$. Since this question is in general equivalent to solving a non-convex optimization problem, we infer that the problem is NP-hard [48].

Despite this, if a partition $\mathcal{H}$ satisfies constraints like (5) with respect to some $\varepsilon > 0$, then we can prove that $\mathcal{H}$ is an $\mathcal{O}(\epsilon)$-FDE/BDE, and vice versa. The basic idea is to observe that a polynomial is the zero function if and only if its coefficients are all zero. In this sense, our techniques defined through the coefficients of the polynomials correspond to check if (5) holds.

Given a partition of variables that represents an approximate differential equivalence, we construct a *reference PIVP* by finding a "perturbation" of the original PIVP — i.e., a modification of the initial condition $\sigma$ and the coefficients present in $q_1, \ldots, q_n$ — which ensures that partition becomes an exact differential equivalence. On this reference PIVP one can use the quotienting algorithms for FDE/BDE developed in [11] (and not restated here formally for brevity). Therefore, the as-obtained quotient represents an approximate reduction of the original PIVP. We obtain the desired perturbation by treating the original initial conditions and polynomial coefficients uniformly as initial conditions on an *extended* PIVP where every coefficient is parameterized and turned into a new ODE variable.

The perturbation to use in order to achieve this result can be found by considering an *extended* PIVP where every coefficient and initial condition is parameterized and turned in a new ODE variable.

**Definition 5.** The parameterization of a polynomial $q_i$ in normal form with variables $\mathcal{S}$ is denoted by $\hat{q}_i$ and arises from $q_i$ by replacing, for each $\alpha \in \mathbb{N}_0^{\mathcal{S}}$, the constant $c(q_i, x^\alpha)$ with the parameter $\mathfrak{c}(\hat{q}_i, x^\alpha)$.

**Example 6.** The polynomials $q_2 = 1.99x_1 - x_2$ and $q_3 = 2.01x_1 - x_3$ from Example 1 give rise to the parameterized polynomials $\hat{q}_2 = \mathfrak{c}(\hat{q}_2, x_1)x_1 + \mathfrak{c}(\hat{q}_2, x_2)x_2$ and $\hat{q}_3 = \mathfrak{c}(\hat{q}_3, x_1)x_1 + \mathfrak{c}(\hat{q}_2, x_3)x_3$, respectively.

**Definition 6** *(Extended PIVP).* For a PIVP $\mathcal{P}$ with variables $\mathcal{S}$, set $\Theta = \{\mathfrak{c}(\hat{q}_i, x^\alpha) \mid 1 \leq i \leq n, \alpha \in \mathbb{N}_0^{\mathcal{S}}\}$. Its extended version $\hat{\mathcal{P}}$ has variables $\mathcal{S} \cup \Theta$ and is given by $\dot{x}_i = \hat{q}_i$ and $\dot{\mathfrak{c}}(\hat{q}_i, x^\alpha) = 0$, where $x_i \in \mathcal{S}$ and $\alpha \in \mathbb{N}_0^{\mathcal{S}}$. For a given $\hat{\sigma} \in \mathbb{R}^{\mathcal{S}\cup\Theta}$, let $\hat{\mathcal{P}}(\hat{\sigma})$ denote the PIVP which arises from $\hat{\mathcal{P}}$ by replacing each $v \in \mathcal{S} \cup \Theta$ by the corresponding real value $\sigma(v) \in \mathbb{R}$ in $\hat{\mathcal{P}}$. In particular, let $\hat{\sigma}_0 \in \mathbb{R}^{\mathcal{S}\cup\Theta}$ be such that $\mathcal{P}(\sigma) = \hat{\mathcal{P}}(\hat{\sigma}_0)$.

**Example 7.** If $\mathcal{P}$ is the PIVP from Example 1, its extended version $\hat{\mathcal{P}}$ is

$$\begin{aligned}
\dot{x}_1 &= \mathfrak{c}(\hat{q}_1, x_1)x_1 + \mathfrak{c}(\hat{q}_1, x_2)x_2 + \mathfrak{c}(\hat{q}_1, x_3)x_3,\\
\dot{\mathfrak{c}}(\hat{q}_1, x_i) &= 0, \quad i = 1, 2, 3,\\
\dot{x}_2 &= \mathfrak{c}(\hat{q}_2, x_1)x_1 + \mathfrak{c}(\hat{q}_2, x_2)x_2,\\
\dot{\mathfrak{c}}(\hat{q}_2, x_i) &= 0, \quad i = 1, 2, 3,\\
\dot{x}_3 &= \mathfrak{c}(\hat{q}_3, x_1)x_1 + \mathfrak{c}(\hat{q}_3, x_2)x_2,\\
\dot{\mathfrak{c}}(\hat{q}_3, x_i) &= 0, \quad i = 1, 2, 3.
\end{aligned}$$

The corresponding $\hat{\sigma}_0$ satisfies $\hat{\sigma}_0(x_i) = \sigma(x_i)$ for $1 \leq i \leq 3$ and

$$\hat{\sigma}_0\big(\mathfrak{c}(\hat{q}_1, x_1)\big) = -4.00, \qquad \hat{\sigma}_0\big(\mathfrak{c}(\hat{q}_1, x_2)\big) = 1.00,$$
$$\hat{\sigma}_0\big(\mathfrak{c}(\hat{q}_1, x_3)\big) = 1.00, \qquad \hat{\sigma}_0\big(\mathfrak{c}(\hat{q}_2, x_1)\big) = 1.99,$$
$$\hat{\sigma}_0\big(\mathfrak{c}(\hat{q}_2, x_2)\big) = -1.00, \qquad \hat{\sigma}_0\big(\mathfrak{c}(\hat{q}_3, x_1)\big) = 2.01,$$
$$\hat{\sigma}_0\big(\mathfrak{c}(\hat{q}_3, x_2)\big) = -1.00.$$

The following is needed for the definition of the reference PIVP.

**Definition 7.** Given constant free polynomial $\hat{\wp}$ (i.e., such that $\hat{\wp}(0) = 0$) and $\Xi \subseteq \mathcal{S} \cup \Theta \cup \{s\}$, let $\mathfrak{t}(\hat{\wp}, x^\alpha, \Xi)$ denote the coefficient term of $x^\alpha$ in $\mathcal{N}(\hat{\wp}, \Xi)$, where $\alpha \in \mathbb{N}_0^\Xi$ and $\mathcal{N}(\hat{\wp}, \Xi)$ is the normal form of $\hat{\wp}$ where variables outside $\Xi$ are treated as parameters.

**Example 8.** With $\hat{q}_2$ and $\hat{q}_3$ as in Example 6 and $\Xi = \{x_1, x_2, x_3\}$, the normal form $\mathcal{N}(\hat{q}_2 - \hat{q}_3, \Xi)$ is given by $(\mathfrak{c}(\hat{q}_2, x_1) - \mathfrak{c}(\hat{q}_3, x_1))x_1 + (\mathfrak{c}(\hat{q}_2, x_2) - \mathfrak{c}(\hat{q}_3, x_2))x_2$, while $\mathfrak{t}(\hat{q}_2 - \hat{q}_3, x_1, \Xi) = \mathfrak{c}(\hat{q}_2, x_1) - \mathfrak{c}(\hat{q}_3, x_1)$.

For an $\mathcal{H}$ to be an FDE/BDE, the coefficients of certain polynomials need to coincide or, alternatively, the corresponding differences to be zero. These constraints can be described by linear equations introduced next.

**Definition 8.** Given a PIVP with variables $\mathcal{S}$ and an $\varepsilon$-FDE partition $\mathcal{H}$ of $\mathcal{S}$, the set of linear constraints of $\mathcal{H}$ is given by

$$\big\{\mathfrak{t}(\tilde{\wp}^H_{i,j}, x^\alpha, \mathcal{S} \cup \{s\}) = 0 \mid \alpha \in \mathbb{N}_0^{\mathcal{S}\cup\{s\}}, H \in \mathcal{H} \text{ and } x_i \sim_{\mathcal{H}} x_j\big\} \tag{6}$$

with $\tilde{\wp}^H_{i,j} = \sum_{x_k \in H} \hat{q}_k - \sum_{x_k \in H} \hat{q}_k[x_i/s(x_i + x_j), x_j/(1-s)(x_i + x_j)]$. Instead, if $\mathcal{H}$ is an $\varepsilon$-BDE partition of $\mathcal{S}$, the corresponding set of linear constraints is

$$\big\{\mathfrak{t}(\tilde{\wp}^{\mathcal{H}}_{i,j}, x^\alpha, \mathcal{S}) = 0 \mid \alpha \in \mathbb{N}_0^{\mathcal{S}}, x_i \sim_{\mathcal{H}} x_j\big\} \cup \big\{x_{i_j} - x_{i_{j+1}} = 0 \mid 1 \le j \le k-1 \text{ and } \{x_{i_1}, \ldots, x_{i_k}\} \in \mathcal{S}/\!\sim_{\mathcal{H}}\big\}, \tag{7}$$

where $\tilde{\wp}^{\mathcal{H}}_{i,j} = (\hat{q}_i - \hat{q}_j)\big[x_{H',1}/x_{H'}, \ldots, x_{H',|H'|}/x_{H'} : H' \in \mathcal{H}\big]$.

We next showcase the linear equations on our example.

**Example 9.** From Example 2, we know that $\mathcal{H} = \{\{x_1\}, \{x_2, x_3\}\}$ is a 0.02-BDE partition of the PIVP (1). The set of linear constraints underlying $\mathcal{H}$ is given by $\mathfrak{c}(\hat{q}_2, x_1) - \mathfrak{c}(\hat{q}_3, x_1) = 0$ and $x_2 - x_3 = 0$.

**Remark 1.** In line with its exact counterpart, an $\varepsilon$-BDE is "useful" under the further constraint that related variables have the same initial conditions in the reference model, as a necessary condition for having equal solutions at all time points. This translates into adding the constraints in (7) that perturbed initial conditions of related variables are equal. This leads, for instance, to the constraint $x_2 - x_3 = 0$ in the running example. For $\varepsilon$-FDE, instead, only constraints on the parameters $\Theta$ are made.

**Theorem 4.** *Given a PIVP $\mathcal{P}$ with variables $\mathcal{S}$, an $\varepsilon$-FDE/BDE partition $\mathcal{H}$ and a configuration $\hat{\sigma} \in \mathbb{R}^{\mathcal{S}\cup\Theta}$ that satisfies (6)/(7), it holds that $\mathcal{H}$ is an FDE/BDE of $\hat{\mathcal{P}}(\hat{\sigma})$.*

**Proof.** Fix a partition $\mathcal{H}$ and consider the system of equations (6)

$$\big\{\mathfrak{t}(\tilde{\wp}^H_{i,j}, x^\alpha, \mathcal{S} \cup \{s\}) = 0 \mid \alpha \in \mathbb{N}_0^{\mathcal{S}\cup\{s\}}, H \in \mathcal{H} \text{ and } x_i \sim_{\mathcal{H}} x_j\big\}$$

By definition, $\mathfrak{t}(\tilde{\wp}^H_{i,j}, x^\alpha, \mathcal{S} \cup \{s\}) = 0$ states that the coefficient of monomial $x^\alpha$ in polynomial $\tilde{\wp}^H_{i,j}$ is zero. Likewise, by definition, $\mathcal{H}$ is an FDE if and only if all polynomials $\tilde{\wp}^H_{i,j}$, where $x_i \sim_{\mathcal{H}} x_j$, are zero. Since a polynomial is zero if and only if all its coefficients are zero, we thus obtain that any choice of parameters satisfying the above set of equations yields an FDE. The proof for BDE proceeds along the same lines. □

In general, an approximate quotient is not unique; indeed, the linear system of constraints from Theorem 4 is underdetermined. Here, we fix one candidate perturbation by assuming that nearby initial conditions yield nearby trajectories. This fact is asymptotically true due to Gronwall's inequality, as mentioned in Section 1. For this purpose, we set the Euclidian norm as objective function in a linear system with the constraints defined by 4.

$$\hat{\sigma}_* = \operatorname*{argmin}_{\hat{\sigma}:\text{Eq. (6)/(7) holds}} \|\hat{\sigma} - \hat{\sigma}_0\|_2 \tag{8}$$

This yields a convex quadratic program that can be solved in polynomial time [26].

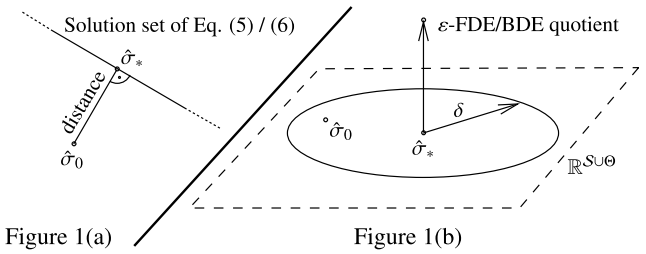



**Fig. 1.** Given a PIVP $\mathcal{P}$, a partition $\mathcal{G}$ of $\mathcal{S}$, and an $\varepsilon > 0$, the coarsest $\varepsilon$-FDE/BDE partition $\mathcal{H}$ that refines $\mathcal{G}$ is constructed. Afterwards, the solution $\hat{\sigma}_*$ of the optimization problem (8) is computed in Fig. 1(a). This allows to compute the $\varepsilon$-FDE/BDE quotient $\hat{\mathcal{P}}(\hat{\sigma}_*)$ of $\mathcal{H}$. With this, $\lambda$ and $\delta$ from Theorem 8 are calculated. In the case the distance between $\hat{\sigma}_0$ and $\hat{\sigma}_*$ does not exceed $\delta$, the tight bounds of Theorem 8 can be applied and relate the trajectories of $\hat{\mathcal{P}}(\hat{\sigma}_*)$ and $\hat{\mathcal{P}}(\hat{\sigma}_0) = \mathcal{P}(\sigma)$, as depicted in Fig. 1(b).

**Example 10.** Let us continue Example 9 and assume that $\sigma(x_2) = \sigma(x_3)$. In such a case, it can be easily seen that $\hat{\sigma}_*$ and $\hat{\sigma}_0$ satisfy $\hat{\sigma}_*(\mathfrak{c}(\hat{q}_2, x_1)) = \hat{\sigma}_*(\mathfrak{c}(\hat{q}_3, x_1)) = \big(\hat{\sigma}_0(\mathfrak{c}(\hat{q}_2, x_1)) + \hat{\sigma}_0(\mathfrak{c}(\hat{q}_3, x_1))\big)/2 = 2.00$ and coincide on all other parameters. In other words, the closest PIVP that enjoys an exact BDE relating $x_2$ and $x_3$ is given, as expected, by perturbing the coefficients 1.99 and 2.01 of (1) to their average value, yielding:

$$\dot{x}_1 = -4.00x_1 + x_2 + x_3$$
$$\dot{x}_2 = 2.00x_1 - x_2$$
$$\dot{x}_3 = 2.00x_1 - x_3$$

The above discussions are summarized in the following.

**Theorem 5.** *Given a PIVP, $\varepsilon \geq 0$, and an $\varepsilon$-FDE/BDE partition $\mathcal{H}$, the solution of (8) exists and can be computed in polynomial time.*

**Proof.** Follows from Theorem 3, Theorem 4 and [26]. □

The solution of the optimization problem (8) stated in Theorem 5 is informally depicted in Fig. 1a.

The reference PIVP is the extended, exactly reducible PIVP with the optimum initial condition $\hat{\sigma}_*$, i.e., $\hat{\mathcal{P}}(\hat{\sigma}_*)$. Its ODE solution is called the *reference trajectory*.

## 4. Error bounds

The objective of this section is to provide a formal bound between the solution of the original PIVP and the reference one. For this, we consider two parameters $\delta > 0$ and $\lambda > 0$. The former corresponds to the size of a ball around the initial condition $\hat{\sigma}_*$ of the reference PIVP; the latter is an *amplifier* that relates the maximum distance between trajectories to the distance between the initial conditions. Every time the initial condition of the original PIVP $\hat{\mathcal{P}}(\hat{\sigma}_0)$ stays inside the $\delta$ ball it is possible to show that the maximum error is limited by a formal bound affected by $\lambda$. This idea is visualized in Fig. 1(b).

The next definitions and theorems build the theory necessary to formalize this idea. We start recalling the notion of Jacobian matrix.

**Definition 9.** Given an extended PIVP with variables $\mathcal{S} \cup \Theta$, the entries of the Jacobian matrix $A = (A_{i,j})_{x_i, x_j \in \mathcal{S}\cup\Theta}$ are given by $A_{i,j} = \partial_{x_j} \hat{q}_i$, where $\partial_x$ denotes the partial derivative with respect to $x$.

Let $A(t) \in \mathbb{R}^{\mathcal{S}\cup\Theta \times \mathcal{S}\cup\Theta}$ denote the Jacobian obtained by plugging in the reference trajectory $x^{\hat{\sigma}_*}(t)$. We will need the following result from the theory of ODEs.

**Theorem 6.** *There exists a family of matrices $\Lambda(t_0, t_1)$, with $0 \leq t_0 \leq t_1 \leq \hat{\tau}$, such that the solution of $\dot{y}(t) = A(t)y(t) + u(t)$, where $y(t_0) = y_0$ and $u$ is continuous, is given by $y(t) = \Lambda(t_0, t)y_0 + \int_{t_0}^{t} \Lambda(s, t)u(s)ds$ for all $0 \leq t_0 \leq t \leq \hat{\tau}$.*

**Proof.** See [49, Theorem 1, Section 1.10]. □

Fix some arbitrary $\hat{\sigma}_1 \in \mathbb{R}^{\mathcal{S}\cup\Theta}$ and let $\hat{\mathcal{P}}$ be given by $\dot{x} = \hat{q}(x)$. Using Taylor's expansion of $\hat{q}$ at point $x^{\hat{\sigma}_*}(s)$, it holds that $\hat{q}(x^{\hat{\sigma}_1}(s)) = \hat{q}(x^{\hat{\sigma}_*}(s)) + A(s)(x^{\hat{\sigma}_1}(s) - x^{\hat{\sigma}_*}(s)) + r(s, x^{\hat{\sigma}_1}(s) - x^{\hat{\sigma}_*}(s))$, where $r$ is the *remainder function* which accounts for the higher order terms of $\hat{q}$ at point $x^{\hat{\sigma}_*}(s)$. This implies that

$$(\dot{x}^{\hat{\sigma}_1}(t) - \dot{x}^{\hat{\sigma}_*}(t)) = A(t)(x^{\hat{\sigma}_1}(t) - x^{\hat{\sigma}_*}(t)) + r(s, x^{\hat{\sigma}_1}(t) - x^{\hat{\sigma}_*}(t)),$$

meaning that $x^{\hat{\sigma}_1} - x^{\hat{\sigma}_*}$ can be interpreted as a solution of the linear ODE system from Theorem 6 with input function $u(s) = r(s, x^{\hat{\sigma}_1}(s) - x^{\hat{\sigma}_*}(s))$. Let $\Delta x$ denote the solution of $\Delta\dot{x}(t) = A(t)\Delta x(t)$ with $\Delta x(0) = x^{\hat{\sigma}_1}(0) - x^{\hat{\sigma}_*}(0)$. The above discussion and Theorem 6 then ensure that the auxiliary function $z = x^{\hat{\sigma}_1} - x^{\hat{\sigma}_*} - \Delta x$ satisfies $z(t) = \Lambda(0,t)z(0) + \int_0^t \Lambda(s,t)r(s, x^{\hat{\sigma}_1}(s) - x^{\hat{\sigma}_*}(s))ds$. With $z(0) = 0$, we thus get the following.

**Theorem 7.** *With the notation from above, it holds that* $\|x^{\hat{\sigma}_1}(t) - x^{\hat{\sigma}_*}(t)\| \leq \|\Delta x(t)\| + \|\int_0^t \Lambda(s,t) \cdot \big(r(s, x^{\hat{\sigma}_1}(s) - x^{\hat{\sigma}_*}(s))\big)ds\|$.

Since $\Delta x \equiv x^{\hat{\sigma}_1} - x^{\hat{\sigma}_*}$ only if the remainder function $r$ is zero (that is, only if the original ODE system is linear), we call $\Delta x$ the linearization of the true difference function $x^{\hat{\sigma}_1} - x^{\hat{\sigma}_*}$. With this, we are in a position to show Theorem 8 and 9.

**Theorem 8.** *Consider an extended PIVP* $\hat{\mathcal{P}}$ *with variables* $\mathcal{S} \cup \Theta$ *and define* $\lambda_0 = \max_{0\leq t\leq \hat{\tau}} \|\Lambda(0,t)\|$ *and* $\lambda_1 = \max_{0\leq t_0\leq t_1\leq \hat{\tau}} \|\Lambda(t_0,t_1)\|$. *Further, define the remainder function* $r : [0;\hat{\tau}] \times \mathbb{R}^{\mathcal{S}\cup\Theta} \to \mathbb{R}^{\mathcal{S}\cup\Theta}$ *via*

$$r(t, x - x^{\hat{\sigma}_*}(t)) = \hat{q}(x) - \hat{q}(x^{\hat{\sigma}_*}(t)) - A(t)(x - x^{\hat{\sigma}_*}(t))$$

*and let* $0 \leq d_2, d_3, \ldots$ *be such that* $\|r(t,y)\| \leq \sum_{k=2}^{\deg(\hat{\mathcal{P}})} d_k \|y\|^k$ *for all* $y \in \mathbb{R}^{\mathcal{S}\cup\Theta}$ *and* $0 \leq t \leq \hat{\tau}$. *Then, with* $\lambda = 2\lambda_0$, *for any* $x^{\hat{\sigma}_1}(0) \in \mathbb{R}^{\mathcal{S}\cup\Theta}$, *it holds that*

$$\|x^{\hat{\sigma}_1}(0) - x^{\hat{\sigma}_*}(0)\| \leq \delta \Rightarrow \max_{0\leq t\leq\hat{\tau}} \|x^{\hat{\sigma}_1}(t) - x^{\hat{\sigma}_*}(t)\| \leq$$
$$\lambda\|x^{\hat{\sigma}_1}(0) - x^{\hat{\sigma}_*}(0)\|$$

*whenever* $\delta > 0$ *satisfies* $\sum_{k=2}^{\deg(\hat{\mathcal{P}})} d_k(2\lambda_0\delta)^{k-1} \leq (2\lambda_1\hat{\tau})^{-1}$.

**Proof.** It can be efficiently checked whether $\mathcal{P}(\hat{\sigma}_*)$ exhibits singularities on $[0;T]$ by using a numerical ODE solver. Hence, if $\mathcal{P}(\hat{\sigma}_*)$ is singularity free, the bound which is to be proven next ensures that $\mathcal{P}(\hat{\sigma}_1)$ is singularity free as long as $\|x^{\hat{\sigma}_1}(0) - x^{\hat{\sigma}_*}(0)\| \leq \delta$.

Let $\delta = \|x^{\hat{\sigma}_1}(0) - x^{\hat{\sigma}_*}(0)\|$ satisfy $\sum_{k=2}^{\deg(\hat{\mathcal{P}})} d_k\delta_+^{k-1} < (2\lambda_1\hat{\tau})^{-1}$ for $\delta_+ = 2\lambda_0\delta$. Since $\delta < \delta_+$, it holds that

$$0 < \tau(\delta) = \inf\{0 \leq t \leq \hat{\tau} \mid \|x^{\hat{\sigma}_1}(t) - x^{\hat{\sigma}_*}(t)\| \geq \delta_+\},$$

where $\inf\emptyset := \infty$ as usual. With this, it holds that

$$\left\| \int_0^t \Lambda(s,t) r(s, x^{\hat{\sigma}_1}(s) - x^{\hat{\sigma}_*}(s))ds \right\| \leq \lambda_1 t \sum_{k=2}^{\deg(\hat{\mathcal{P}})} d_k\delta_+^k$$

for all $t \leq \min\{\tau(\delta), \hat{\tau}\}$. Hence, Theorem 9 and 7 yield

$$\|x^{\hat{\sigma}_1}(t) - x^{\hat{\sigma}_*}(t)\| \leq \lambda_0\delta + \lambda_1\hat{\tau} \sum_{k=2}^{\deg(\hat{\mathcal{P}})} d_k\delta_+^k =$$
$$\frac{\delta_+}{2} + \lambda_1\hat{\tau} \sum_{k=2}^{\deg(\hat{\mathcal{P}})} d_k\delta_+^k < \delta_+$$

for all $t \leq \min\{\tau(\delta), \hat{\tau}\}$, where the last inequality follows from straightforward algebraic manipulation. This implies that $\tau(\delta) = \infty$, thus showing that $\|x^{\hat{\sigma}_1}(t) - x^{\hat{\sigma}_*}(t)\| < 2\lambda_0\|x^{\hat{\sigma}_1}(0) - x^{\hat{\sigma}_*}(0)\|$ for all $0 \leq t \leq \hat{\tau}$ because $\delta_+ = 2\lambda_0\delta$. □

Through the Theorem 8 we show that there exists a formal bound proportional to $\lambda$ in terms of the initial perturbation if this is smaller than $\delta$.

We wish to point out that the maximal $\delta$ satisfying $\sum_{k=2}^{\deg(\hat{\mathcal{P}})} d_k(2\lambda_0\delta)^{k-1} \leq (2\lambda_1\hat{\tau})^{-1}$ is a root of a polynomial in one variable and thus can be efficiently approximated from below via Newton's method. Instead, the assumption $\|r(s,y)\| \leq \sum_{k=2}^{\deg(\hat{\mathcal{P}})} d_k\|y\|^k$ on the remainder function $r$ states essentially that, for any $k \geq 2$, the sum of all $k$-th order derivatives of $r$ is bounded by $d_k$ along the reference trajectory $x^{\hat{\sigma}_*}$.

The previous Theorem could be more precise in the special case of linear systems (i.e., $\deg(\hat{\mathcal{P}}) = 1$) as discussed in [50]. In the next result we consider this context and we show that the amplifier could be halved, this is because Theorem 9 need

not estimate nonlinear terms present in remainder function $r$. More importantly, Theorem 9 shows that the amplifier of Theorem 8 cannot be substantially improved.

**Theorem 9.** *If an extended PIVP $\hat{\mathcal{P}}$ satisfies $\deg(\hat{\mathcal{P}}) = 1$ and $\lambda = 2\lambda_0$, it holds that*

$$\max_{0\le t\le\hat{\tau}} \|x^{\hat{\sigma}_1}(t) - x^{\hat{\sigma}_*}(t)\| \le \frac{\lambda}{2}\|x^{\hat{\sigma}_1}(0) - x^{\hat{\sigma}_*}(0)\|$$

*for any $x^{\hat{\sigma}_1}(0) \in \mathbb{R}^{\mathcal{S}\cup\Theta}$. The bound is tight in the sense that there exist $0 \le t \le \hat{\tau}$ and $x^{\hat{\sigma}_1}(0) \in \mathbb{R}^{\mathcal{S}\cup\Theta}$ such that $\|x^{\hat{\sigma}_1}(t) - x^{\hat{\sigma}_*}(t)\| = \frac{\lambda}{2}\|x^{\hat{\sigma}_1}(0) - x^{\hat{\sigma}_*}(0)\|$.*

**Proof.** As pointed out above, in the case of $\deg(\hat{\mathcal{P}}) = 1$, it holds that $r \equiv 0$. This and Theorem 6 imply $x^{\hat{\sigma}_1}(t) - x^{\hat{\sigma}_*}(t) = \Delta x(t) = \Lambda(0,t)\Delta x(0) = \Lambda(0,t)(x^{\hat{\sigma}_1}(0) - x^{\hat{\sigma}_*}(0))$ for any $x^{\hat{\sigma}_1}(0) \in \mathbb{R}^{\mathcal{S}\cup\Theta}$ and $0 \le t \le \hat{\tau}$, thus yielding the first claim. The second claim, instead, follows by noting that

$$\max_{0\le t\le\hat{\tau}} \max_{\|\Delta x(0)\|=1} \|\Delta x(t)\| = \max_{0\le t\le\hat{\tau}} \max_{\|\Delta x(0)\|=1} \|\Lambda(0,t)\Delta x(0)\| = \max_{0\le t\le\hat{\tau}} \|\Lambda(0,t)\| = \lambda_0 \quad \square$$

**Remark 2.** We note that $\lambda_0, \lambda_1$ can be estimated efficiently. Indeed, let $e_{x_i} \in \mathbb{R}^{\mathcal{S}\cup\Theta}$ be the $x_i$-th unit vector in $\mathbb{R}^{\mathcal{S}\cup\Theta}$, i.e., $e_{x_i}(x_j) = \delta_{i,j}$ where $\delta_{i,j}$ is the Kronecker delta. Then, if $y(t_0) = e_{x_i}$, Theorem 6 implies $y(t_1) = \Lambda(t_0,t_1)e_{x_i}$. Since $\Lambda(0,t_1)e_{x_i}$ is the $x_i$-column of $\Lambda(0,t_1)$ and $\Lambda(t_0,t_1) = \Lambda(0,t_1)\Lambda(0,t_0)^{-1}$, this shows that the matrices $\Lambda(t_0,t_1)$ can be computed by solving $|\mathcal{S}\cup\Theta|$ instances of the linear ODE system from Theorem 6 up to time $\hat{\tau}$.

By calculating a bound $L > 0$ on $\max_{0\le t\le\hat{\tau}}\|A(t)\|$ and by computing the matrices $\Lambda(t_i,t_j)$ for all time points $t_k$ underlying a fixed discretization step $\Delta t > 0$ of $[0;\hat{\tau}]$, the following can be shown.

**Lemma 2.** *Together with $\lambda_0^+ = \max_i\|\Lambda(0,t_i)\|$ and $\lambda_1^+ = \max_{i\le j}\|\Lambda(t_i,t_j)\|$, it holds that $\lambda_0 \le \lambda_0^+ e^{L\Delta t}$ and $\lambda_1 \le \lambda_1^+[1 + L\Delta t(e^{L\Delta t} + 1)]$.*

**Proof.** Fix $0 \le s \le t \le \hat{\tau}$ and assume that $|s-t| \le \Delta$. Since $(\partial_t\Lambda)(s,t) = A(t)\Lambda(s,t)$, it can be shown (see Lemma 1 and 2 in [44]) that $\|\Lambda(s,t)\| \le e^{L(t-s)} \le e^{L\Delta t}$ and the claim holds true. Using this, it can be also seen that

$$\|\Lambda(0,t)\| = \|\Lambda(t_i,t)\Lambda(0,t_i)\| \le \lambda_0^+ e^{L\Delta t}$$

for some $t_i \le t$ such that $t - t_i \le \Delta t$. Let us now assume that $|s-t| > \Delta t$. Then there exist $t_i \le t_j$ such that $s \le t_i \le t_j \le t$ with $t_i - s, t - t_j \le \Delta$. With this it holds that

$$\|\Lambda(t_i,t_j) - \Lambda(s,t)\| \le \|\Lambda(t_i,t_j) - \Lambda(t_i,t)\| + \|\Lambda(t_i,t) - \Lambda(s,t)\|$$

The bound on $\lambda_1$ then follows by noting that

$$\begin{aligned}\|\Lambda(t_i,t_j) - \Lambda(t_i,t)\| &= \|\Lambda(t_i,t_j) - \Lambda(t_j,t)\Lambda(t_i,t_j)\| \\ &= \|(\Lambda(t_j,t_j) - \Lambda(t_j,t))\Lambda(t_i,t_j)\| \\ &\le \lambda_1^+ L\Delta t\end{aligned}$$

and

$$\begin{aligned}\|\Lambda(t_i,t) - \Lambda(s,t)\| &= \|\Lambda(t_i,t) - \Lambda(t_i,t)\Lambda(s,t_i)\| \\ &= \|\Lambda(t_i,t)(\Lambda(s,s) - \Lambda(s,t_i))\| \\ &= \|\Lambda(t_j,t)\Lambda(t_i,t_j)(\Lambda(s,s) - \Lambda(s,t_i))\| \\ &\le e^{L\Delta t}\lambda_1^+ L\Delta t. \quad \square\end{aligned}$$

The next result simplifies the constraints on $\delta$ from Theorem 8 if $\deg(\hat{\mathcal{P}}) \le 3$.

**Lemma 3.** *In the case where $\deg(\hat{\mathcal{P}}) \le 3$, the constraint on $\delta$ of Theorem 8 simplifies to $\delta \le \left[2\hat{\tau}\lambda_0\lambda_1\left(d_2 + \sqrt{d_2^2 + \frac{2d_3}{\lambda_1\hat{\tau}}}\right)\right]^{-1}$.*

**Proof.** For $\deg(\hat{\mathcal{P}}) = 3$, the constraint writes as $d_2(2\lambda_0\delta)^2 + d_3(2\lambda_0\delta)^1 \leq (2\lambda_1\hat{\tau})^{-1}$. Since the left-hand side is monotonic increasing in $\delta$, it thus suffices to solve the quadratic equation $d_2(2\lambda_0\delta)^2 + d_3(2\lambda_0\delta)^1 = (2\lambda_1\hat{\tau})^{-1}$. The following formula, known as Muller's variant, can be readily checked to solve a quadratic equation $ax^2 + bx + c$:

$$x = \frac{-2c}{b \pm \sqrt{b^2 - 4ac}}$$

Matching the coefficients gives the claim. The claim for $\deg(\hat{\mathcal{P}}) = 2$, instead, is trivial. □

The above lemma applies, for instance, to most biochemical systems, as discussed in Section 3. The next result, instead, allows for an efficient estimation of $d_k$, with $2 \leq k \leq \deg(\hat{\mathcal{P}})$.

**Lemma 4.** *Given an extended PIVP $\hat{\mathcal{P}}$ with variables $\mathcal{S} \cup \Theta$, let $\#_k(\hat{q}_i)$ be the number of degree $k$ monomials in $\mathcal{N}(\hat{q}_i)$ and $D(\hat{q}_i, \hat{\sigma})$ the largest coefficient of $\mathcal{N}(\hat{q}_i)$ for configuration $\hat{\sigma} \in \mathbb{R}^{\mathcal{S}\cup\Theta}$. With $C = \max_{0\leq t\leq\hat{\tau}} \|x^{\hat{\sigma}_*}(t)\|$, $M = \max_{x_i\in\mathcal{S}} \max_{k\geq 2} \#_k(\hat{q}_i)$ and $D = \max_{x_i\in\mathcal{S}} D(\hat{q}_i, \hat{\sigma}_*)$, it suffices to set $d_k$ in Theorem 8 to $C^{\deg(\hat{\mathcal{P}})-k}MD$.*

**Proof.** *Fix some $x_i \in \mathcal{S} \cup \Theta$ and $0 \leq s \leq \hat{\tau}$. Then, the multidimensional Taylor expansion of $\hat{q}_i$ at point $x_s := x^{\hat{\sigma}_*}(s)$ is given by*

$$\hat{q}_i(x) = \sum_{|\alpha|\leq\deg(\hat{\mathcal{P}})} \frac{(D^\alpha \hat{q}_i)(x_s)}{\alpha!}(x - x_s)^\alpha,$$

*where $D^\alpha = \frac{\partial^{|\alpha|}}{\prod_{x_i\in\mathcal{S}} \partial x_i^{\alpha_{x_i}}}$ is the standard compact notation of the partial derivative underlying the multi-index $\alpha \in \mathbb{N}_0^{\mathcal{S}\cup\Theta}$ with $\alpha! = \prod_{x_i\in\mathcal{S}\cup\Theta} \alpha_{x_i}!$ and $|\alpha| = \sum_{x_i\in\mathcal{S}} \alpha_{x_i}$. Using the concept of Jacobi matrix, the above formula rewrites to*

$$\hat{q}_i(x) = \hat{q}_i(x_s) + e_{x_i}^T A(x_s)(x - x_s) + \sum_{2\leq|\alpha|\leq\deg(\hat{\mathcal{P}})} \frac{(D^\alpha \hat{q}_i)(x_s)}{\alpha!}(x - x_s)^\alpha \tag{9}$$

*Since this shows that*

$$r_i(s, y) = \sum_{2\leq|\alpha|\leq\deg(\hat{\mathcal{P}})} \frac{(D^\alpha \hat{q}_i)(x_s)}{\alpha!} y^\alpha,$$

*a straightforward estimation of the terms yields the claim.* □

For the case of linear systems whose parameters are subject to perturbation, instead, the following lemma can be applied. It provides a sharper estimate on $d_2$ but comes at the price of more involved computation.

**Lemma 5.** *Given an extended PIVP $\hat{\mathcal{P}}$ with variables $\mathcal{S} \cup \Theta$, the Hessian matrix $H^k = (H^k_{i,j})_{x_i,x_j}$ of $\hat{q}_k$ is given by $H^k_{i,j} = \partial_{x_i}\partial_{x_j}\hat{q}_k$. With this, $d_2$ can be chosen as $d_2 = \frac{1}{2} \cdot \max_{x_i\in\mathcal{S}\cup\Theta} \max_{0\leq t\leq\hat{\tau}} \|H^i(x^{\hat{\sigma}_*}(t))\|$.*

**Proof.** Using the concept of Hessian matrix, (9) from the proof of Lemma 4 can be rewritten into

$$\hat{q}_i(x) = \hat{q}_i(x_s) + e_{x_i}^T A(x_s)(x - x_s) + \tfrac{1}{2}(x - x_s)^T H^i(x_s)(x - x_s) + \sum_{|\alpha|=3} \frac{(D^\alpha \hat{q}_i)(x_s)}{\alpha!}(x - x_s)^\alpha,$$

thus readily implying the claim. □

**Example 11.** Since $\deg(\hat{\mathcal{P}}) = 2$ in Example 7, coefficients $d_3, d_4, \ldots$ are zero and we only need to bound $d_2$. Moreover, the constraint in Theorem 8 simplifies to $\delta \leq (4\hat{\tau}\lambda_0\lambda_1 d_2)^{-1}$ thanks to Lemma 3. By applying Lemma 4, instead, we see that it suffices to choose $d_2 = 2.00$ because $M = 2.00$ and $D = 1.00$. In the case of $\hat{\tau} = 3.00$, we thus get $\lambda_0 = \lambda_1 = 1.40$ which yields $\delta \leq 0.02$.

## 5. Evaluation

In this section let show the application of $\varepsilon$-BDE (FDE) on real case studies. More in detail, we take into account three different examples: an electrical network, a protein interaction network and a polymerization model. For every model we provide the results achieved applying the reductions implemented in the ERODE tool and we compare them against CORA. The results reported were computed on a laptop with Windows 10 64 bits over an actual Intel Core i7 machine with 16 GB of RAM.

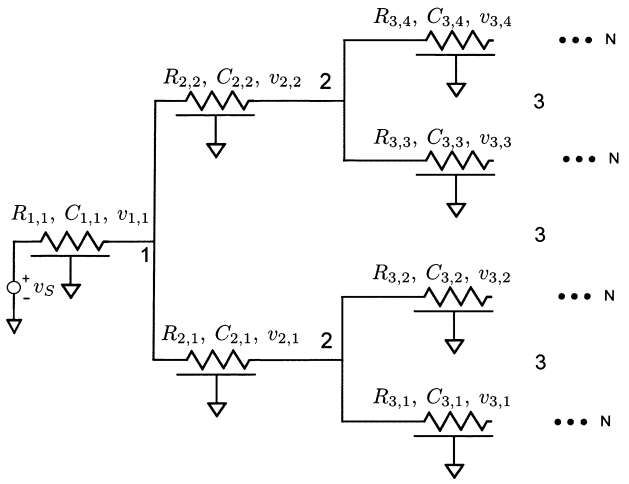



**Fig. 2.** H-tree network adapted from [29].

**Table 1**
Nominal parameters of electronic components at different depths $i$.

| $i$ | $R_i^*$ (mΩ) | $C_i^*$ (fF) |
|---|---|---|
| 1 | 3.19 | 0.280 |
| 2 | 6.37 | 0.300 |
| 3 | 12.75 | 0.130 |
| 4 | 25.50 | 0.140 |
| 5 | 50.00 | 0.070 |
| 6 | 100.00 | 0.070 |

### 5.1. Electrical network

We consider a simplified (inductance free) version of a power distribution electrical network from [29], arranged as a tree called *H-tree* (Fig. 2). We let $N$ be the depth of the tree and denote the resistances and the capacitances at depth $i$ by $R_{i,k}$ and $C_{i,k}$, respectively. The source voltage is $v_s$, here assumed to be constant, $v_s = 2.0\text{V}$. Then, the voltage across $C_{i,k}$, denoted by $v_{i,k}$, obeys the affine ODE

$$\begin{aligned}\dot{v}_{1,1} &= \frac{v_S - v_{1,1}}{R_{1,1}C_{1,1}} - \frac{v_{1,1} - v_{2,1}}{R_{2,1}C_{1,1}} - \frac{v_{1,1} - v_{2,2}}{R_{2,2}C_{1,1}},\\ \dot{v}_{i,k} &= \frac{v_{i-1,l} - v_{i,k}}{R_{i,k}C_{i,k}},\end{aligned} \tag{10}$$

where $1 \le i \le N$, $k = 1, \ldots, 2^{i-1}$, and $l = \lceil k/2 \rceil$, where $\lceil \cdot \rceil$ denotes the ceil function. Here we considered networks with depth up to $N = 6$. For depths $i \le 4$, the nominal parameter values $R_i^*$ and $C_i^*$ were taken from [29]; for $i \ge 5$, instead, we extrapolated them. The parameters are summarized in Table 1.

In [29] the authors show that when the values of resistors and capacitors of any depth are equal, i.e., $R_{i,\cdot} \equiv R_i^*$ and $C_{i,\cdot} \equiv C_i^*$ then the network is *symmetric*. That is, the voltages at the capacitors in any level are equal at all time points. Indeed, $\left\{\{v_{i,k} \mid 1 \le k \le 2^{i-1}\} \mid 1 \le i \le N\right\}$ is an exact BDE partition (with $N$ equivalence classes).

We now study the robustness of the symmetry under the realistic assumption that resistances and capacitances are only approximately equal. In particular, we test whether it is possible to explain quasi-symmetries when the parameters have tolerance $\eta = 0.01\%$. This corresponds to a practical situation when components or measurements parameters enjoy high accuracy.

We considered networks of different size by varying the maximum depth $N$ from 2 to 6. For each size, we built an ODE models by sampling values for $R_{i,k}$ and $C_{i,k}$ uniformly at random within $\eta$ percent from their nominal values.

To each model we applied the $\varepsilon$-BDE reduction algorithm; choosing $\varepsilon = 6.00\text{E-}4$, it returned a quotient corresponding to a perfectly symmetrical case. We computed the values of $\delta$ and $\lambda$ over a time horizon of 7 time units. This was chosen as a representative time point, for any $N$, of the transient state of the network (to account for the fact that, typically, circuits are analyzed in the time domain for transient analysis only).

**Table 2**
H-tree model results. The *Bound* column in the $\varepsilon$-BDE side refers to the quantity $\lambda \|\cdot\|$.

| | $\epsilon$-BDE | | | | CORA | |
|---|---|---|---|---|---|---|
| N | Time (s) | $\delta$ | $\|\cdot\|$ | Bound | Time (s) | Bound |
| 2 | 7.06E-1 | 7.95E−4 | 6.64E−5 | 3.59E−4 | 2.05E+1 | 3.90E−3 |
| 3 | 1.33E+0 | 6.34E−4 | 6.70E−5 | 4.20E−4 | 5.60E+1 | 3.00E−3 |
| 4 | 2.62E+0 | 4.71E−4 | 8.79E−5 | 6.84E−4 | 3.04E+2 | 2.30E−3 |
| 5 | 9.20E+0 | 4.71E−4 | 1.27E−4 | 9.88E−4 | 4.02E+3 | 1.60E−3 |
| 6 | 2.57E+1 | 4.71E−4 | 1.58E−4 | 1.23E−3 | – | |

The presence of uncertain parameters required us to transform the originally affine system (10) into a polynomial system of degree two (by substituting each $1/(R_{i,k}C_{j,l})$ with a corresponding new state variable) with $2^{N+1}$ states. Since this required nonlinear over-approximation techniques, it ruled out standard over-approximation approaches for linear systems.

In Table 2 we present the results for the random models generated. The runtimes (second column) refer to the computation cost of the $\lambda$-$\delta$ pair. In all cases, $\delta$ turned out to be larger than the distance between the original model and its quotient $\|\hat{\sigma}_0 - \hat{\sigma}_*\| = \|x^{\hat{\sigma}_0}(0) - x^{\hat{\sigma}_*}(0)\|$ (shown in column $\|\cdot\|$). This demonstrates that the 0.01% tolerance can be formally explained by approximate differential equivalence, indeed the precondition of Theorem 8 hold and we can provide a formal bound.

For CORA we set time step equal to 0.01 as this led to tight enough bounds. For our approach, instead, we used time step 0.023 because this ensured tight approximations of $\lambda_0$ and $\lambda_1$ via Lemma 2. In all cases the time-out was set to 3 hours.

The comparison results are also reported in Table 2. For a network of depth $N$, the over-approximations for CORA is reported as the maximal diameter of the flowpipe underlying $v_{N,1}$ across all time points. As such, it can be compared to the product $\lambda \cdot \|\hat{\sigma}_0 - \hat{\sigma}_*\|$ given by our bound, which is also explicitly reported in the table for the sake of easy comparability (column *Bound* in Table 2).

CORA reported tight bounds, but ours are better, approximately 1 order of magnitude less. CORA failed to compute the symbolic Jacobian matrices for $N > 5$. Our approach, instead, reaches good performance also for $N = 6$, not only, we wish to point out that our algorithm naturally applies to parallelization. Indeed, its bottleneck is in the computation of the set of linear ODE systems discussed in Remark 2, which can be trivially solved independently from each other.

### 5.2. Polymerization model

In chemistry, polymerization is the process by which basic compounds, called *monomers*, react and bind to form chains of several units. A prototypical CRN for modeling such situation in the case of *homo-polymerization* (when monomers are of the same species), may be as follows: $A + A \to AA$, $AA + A \to AAA$, and so on until the polymer grows so large that other phenomena will cause instability, preventing further binding of monomers. To allow for an exact quotient, the kinetic rates are typically set equally in all reactions. However this assumption is challenged by measurements, explained by the fact that the geometrical conformation of polymers of different length affects reactivity [51].

Here, we consider the polymerization scenario taken from [51, Chapter 7] illustrating the formation of polycyclic aromatic hydrocarbons in flame combustion. The CRN describes the growth of a molecule with $i$ aromatic rings, denoted by the formal chemical species $A_i$, according to the infinite reaction scheme

$$\mathrm{A}_i + \mathrm{H} \xrightarrow{\alpha_i} \mathrm{A}_i\tilde{} + \mathrm{H}_2 \qquad (i,1) \tag{11}$$

$$\mathrm{A}_i\tilde{} + \mathrm{H}_2 \xrightarrow{\overline{\alpha}_i} \mathrm{A}_i + \mathrm{H} \qquad (i,\overline{1})$$

$$\mathrm{A}_i\tilde{} + \mathrm{C}_2\mathrm{H}_2 \xrightarrow{\beta_i} \mathrm{A}_i\mathrm{CHCH}\tilde{} \qquad (i,2) \tag{12}$$

$$\mathrm{A}_i\mathrm{CHCH}\tilde{} \xrightarrow{\overline{\beta}_i} \mathrm{A}_i\tilde{} + \mathrm{C}_2\mathrm{H}_2 \qquad (i,\overline{2})$$

$$\mathrm{A}_i\mathrm{CHCH}\tilde{} + \mathrm{C}_2\mathrm{H}_2 \xrightarrow{\gamma_i} \mathrm{A}_{i+1} + \mathrm{H} \qquad (i,3) \qquad \dots \qquad (i+1,1)$$

Here $\mathrm{A}_i\tilde{}$ is an aromatic radical formed by H abstraction from $\mathrm{A}_i$, and $\mathrm{A}_i\mathrm{CHCH}\tilde{}$ is a radical formed by adding $\mathrm{C}_2\mathrm{H}_2$ to $\mathrm{A}_i\tilde{}$. The reactions $(i,1)$ and $(i,\overline{1})$, and similarly $(i,2)$ and $(i,\overline{2})$, model reversible mechanisms. In this section we consider finite truncations of the previous model, like the following one.

$$\mathrm{A}_1 + \mathrm{H} \xrightarrow{\alpha_1} \mathrm{A}_1\tilde{} + \mathrm{H}_2 \qquad (1,1) \tag{13}$$

$$\mathrm{A}_2 + \mathrm{H} \xrightarrow{\alpha_2} \mathrm{A}_2\tilde{} + \mathrm{H}_2 \qquad (2,1)$$

$$\mathrm{A}_1\tilde{} + \mathrm{H}_2 \xrightarrow{\overline{\alpha}_1} \mathrm{A}_1 + \mathrm{H} \qquad (1,\overline{1}) \tag{14}$$

**Table 3**
Polymerization models results. The *Bound* column in the $\varepsilon$-BDE side refers to the quantity $\lambda \|\cdot\|$.

| $\varepsilon$-BDE | | | | | CORA | |
|---|---|---|---|---|---|---|
| *N* | *Time (s)* | $\delta$ | $\|\cdot\|$ | *Bound* | *Time (s)* | *Bound* |
| 2 | 4.80 | 4.21E–2 | 1.50E–3 | 3.02E–3 | 21.54 | 1.50E–2 |
| 3 | 6.75 | 2.14E–2 | 4.00E–3 | 8.07E–3 | 48.60 | 1.50E–2 |
| 4 | 9.80 | 1.29E–2 | 4.20E–3 | 8.44E–3 | 81.27 | 1.50E–2 |
| 5 | 11.97 | 8.60E–3 | 4.60E–3 | 9.32E–3 | 231.81 | 1.50E–2 |
| 6 | 14.10 | 6.20E–3 | 5.80E–3 | 1.17E–2 | 432.47 | 1.50E–2 |

$$\begin{aligned}
A_2\tilde{} + H_2 &\xrightarrow{\overline{\alpha}_2} A_2 + H && (2,\overline{1}) \\
A_1\tilde{} + C_2H_2 &\xrightarrow{\beta_1} A_1CHCH\tilde{} && (1,2) \\
A_2\tilde{} + C_2H_2 &\xrightarrow{\beta_2} A_2CHCH\tilde{} && (2,2) \qquad (15) \\
A_1CHCH\tilde{} &\xrightarrow{\overline{\beta}_1} A_1\tilde{} + C_2H_2 && (1,\overline{2}) \\
A_2CHCH\tilde{} &\xrightarrow{\overline{\beta}_2} A_2\tilde{} + C_2H_2 && (2,\overline{2}) \qquad (16) \\
A_1CHCH\tilde{} + C_2H_2 &\xrightarrow{\gamma_1} A_2 + H && (1,3) \qquad (17) \\
A_2CHCH\tilde{} + C_2H_2 &\xrightarrow{\gamma_2} A_1 + H && (2,3)
\end{aligned}$$

In this case we restrict only to the dynamics of polymers up to length 2 (i.e., with $i \in \{1,2\}$), and redirect the flux originally going in $A_3$ to $A_1$, see reaction $(2,3)$. Intuitively, this mimics the fact that polymers of length 3 become unstable due to their length. Let define $N$ the length of the polymers, in this section we take into account three different truncation with N from 2 to 6. For sake of simplicity, the set-up of the experiments is explained below only for the model with $N$ equals to 2, the discussion easily extends for the other models. A model with polymers of length $N$ has $3 \cdot (N+1)$ ODEs.

In [51] it is assumed that the dynamics of polymers do not depend on their length. In our truncated model, this corresponds to setting perfectly symmetric values for rates, i.e., $\alpha_1 = \alpha_2$, $\beta_1 = \beta_2$, $\gamma_1 = \gamma_2$, $\overline{\alpha}_1 = \overline{\alpha}_2$, $\overline{\beta}_1 = \overline{\beta}_2$. Then, it can be shown (similarly to [51]) that $x_{A_1}$ and $x_{A_2}$, $x_{A_1\tilde{}}$ and $x_{A_2\tilde{}}$, and $x_{A_1CHCH\tilde{}}$ and $x_{A_2CHCH\tilde{}}$ are related by FDE.

We now consider a variant with approximately equal rates by perturbing the parameters of 0.1%, while we kept the reversed rates equal, i.e., $\overline{\alpha}_1 = \overline{\alpha}_2 = 0.1$ and $\overline{\beta}_1 = \overline{\beta}_2 = 0.2$. The perturbation was made around the values 1.0, 2.0 and 3.0 respectively for $\alpha$, $\beta$ and $\gamma$.

To study the approximation error, we set the initial concentrations of H, $H_2$, $C_2H_2$ and $A_1$ to 5, with a 0.06 time horizon, which ensures a good part of the models' dynamic.

The CORA bound corresponds to the maximal flowpipe diameter of molecule $A_1$.

The experimental results are reported in Table 3. For each model the norm $\|\cdot\|$ is less than $\delta$ which imply the existence of the formal bound $\lambda \|\cdot\|$ computed in column *Bound*. The running times show that $\epsilon-$FDE reduction is quicker than CORA, not only, also our formal bounds outperform the ones provided by CORA.

### 5.3. Protein interaction networks

We tested approximate differential equivalence on another example of polynomial ODE systems from computational biochemistry. Here a recurring case is the dynamics of complexes such as receptors and scaffold proteins, which have multiple binding domains (e.g., [52,53]). Let us consider a prototypical situation where a molecule $A$ has two independent binding sites to which molecule $B$ can bind reversibly. We denote by $A_{10}$ and $A_{01}$ the species obtained when $A$ and $B$ are bound via the first or second binding site of $A$, respectively, while the other binding site is free. Instead, $A_{11}$ denotes the complex obtained when $A$ is bound to two molecules of $B$. This situation can be described using the following mass-action CRN:

$$\begin{aligned}
A + B &\xrightarrow{k_{b_1}} A_{10} & \qquad A_{10} &\xrightarrow{k_{u_1}} A + B \\
A + B &\xrightarrow{k_{b_2}} A_{01} & \qquad A_{01} &\xrightarrow{k_{u_2}} A + B \\
A_{01} + B &\xrightarrow{k_{b_1}} A_{11} & \qquad A_{11} &\xrightarrow{k_{u_1}} B + A_{01} \\
A_{10} + B &\xrightarrow{k_{b_2}} A_{11} & \qquad A_{11} &\xrightarrow{k_{u_2}} B + A_{10}
\end{aligned}$$

**Table 4**
Binding model: parameters and reduction results.

| | *Rates of bindings* | | | | *Runtime (s)* | | |
|---|---|---|---|---|---|---|---|
| $N$ | $k_{b_1}$ | $k_{b_2}$ | $k_{b_3}$ | $k_{b_4}$ | 0.8-FDE | 0.4-BDE | $\lvert\mathcal{H}\rvert$ |
| 2 | 10.0748 | 9.9864 | – | – | 0.003 | 0.001 | 4 |
| 3 | 9.9174 | 10.0575 | 9.9740 | – | 0.010 | 0.001 | 5 |
| 4 | 10.0886 | 10.0226 | 9.9418 | 9.9505 | 0.078 | 0.002 | 6 |

**Table 5**
Results of $\varepsilon$-BDE and CORA for the model reported in Table 4. The horizon $\tau$ is equals to 8E-4. The *Bound* column in the $\varepsilon$-BDE side refers to the quantity $\lambda \|\cdot\|$.

| | $\varepsilon$-BDE | | | | CORA | |
|---|---|---|---|---|---|---|
| $N$ | *Time (s)* | $\delta$ | $\|\cdot\|$ | *Bound* | *Time (s)* | *Bound* |
| 2 | 4.02 | 2.00E−0 | 6.25E−2 | 1.30E−1 | 5.80 | 1.13E−1 |
| 3 | 6.69 | 5.19E−1 | 9.97E−2 | 2.00E−1 | 14.25 | 1.34E−1 |
| 4 | 11.35 | 1.22E−1 | 1.19E−1 | 2.40E−1 | – | |

In the PIVP each species $S$ will have an associated ODE $\dot{x}_S = q_S$ where $q_S$ is a degree-two polynomial, together with an initial condition $x_S(0)$ that gives its initial concentration. For instance, the first reaction contributes the monomial $-k_{b_1} x_A x_B$ to the ODE of each reagent, i.e. $x_A$ and $x_B$, and $k_{b_1} x_A x_B$ to the ODE of the product $x_{10}$.

Many models in the literature assume perfectly symmetric binding sites, where the kinetic constants do not depend on which particular binding site is involved in the interaction (e.g., [54,55]). Mathematically, this is translated into assuming that $k_{b_1} = k_{b_2}$ and $k_{u_1} = k_{u_2}$. At the ODE level, these perfect symmetries between the species $A_{10}$ and $A_{01}$ are captured by both FDE and BDE [46,11], as well as by domain-specific techniques [56,52,53]. However, it is well understood that in reality distinct binding sites will have different kinetic rates to account for their different conformation, or just for the ineluctable uncertainty introduced by difficulties in measuring such rates. Thus, the assumption of perfect symmetry can be seen as a mathematical convenience to simplify the description of a more heterogeneous real system. With approximate differential equivalences we can study *how strong* the assumption of perfect symmetry actually is.

We considered variants of this binding model by increasing the number of sites $N$ of molecule $A$ from 2 to 4, each site being involved in binding events with its own rate denoted by $k_{b_i}$; for simplicity, instead, we assume that unbinding events have the same kinetic constant $k_{u_i} = 0.1$ for every site $i$. The model with $N$ binding sites has $2^N + 1$ ODEs. Similarly to the H-tree case study, the values of $k_{b_i}$ were sampled uniformly in an interval around an arbitrarily fixed value of 10. The interval considers a 1% perturbation of the fixed value. In this way, the original model was always contained in the $\delta$ neighborhood around the reference PIVP, where all binding rates then become equal to 10 by construction. In Table 4 let show the values for the model taken into account for experimental results.

Since distinct binding rates are used, none of the models can be reduced by FDE or BDE. Instead, we set $\varepsilon = 0.8$ and $\varepsilon = 0.4$ for $\varepsilon$-FDE and $\varepsilon$-BDE, respectively. With these, all models for $2 \leq N \leq 4$ feature the same $\varepsilon$-FDE and $\varepsilon$-BDE coarsest partition (column $\lvert\mathcal{H}\rvert$ gives the number of blocks). This is similar to the exact counterpart where all the binding rates are equal. The fact that $\varepsilon$-BDE has much better runtime performance than $\varepsilon$-FDE backs the complexity result in Theorem 3.

We computed the bounds for the $\varepsilon$-BDE only. We set up the time horizon $\hat{\tau}$ equals to 8.0E–4, starting from an initial condition where the concentrations were always set to 20 for the free molecules $A$ and $B$, and to zero for the other species. With this set-up, the values of $\delta$ were computed using Lemma 3. The results, together with the runtimes for these computations, are reported in Table 5. We find that $\delta$ decreases as the number of asymmetric binding sites increases; The results confirm that the $\delta$ neighborhood does contain the original model in all cases, consequently also in this case we can provide a formal bound for each model.

The comparison with CORA was performed similarly to the H-tree case studies. In particular, in all cases we used the time horizons $\hat{\tau}$ and we set time step equal to $\hat{\tau}/100$ for CORA. For our approach, instead, in order to ensure tight approximations of $\lambda_0$ and $\lambda_1$ via Lemma 2 we used a time steps equal to 1E–6. The CORA bound was computed as the maximal flowpipe diameter for the value of species $A$ across the time interval. In the models with $N$ equals to 2 and 3 the CORA bounds are tighter. Despite this the $\varepsilon$-BDE provide good formal bounds in a smaller amount of time. In the last model CORA did not terminate its computation within the 3 hours timeout, instead our tool returns a good formal bound.

### 5.4. Mobile virus model

We consider a simplified mobile virus propagation model inspired by [57]. Here a region is divided into cells where the mobile virus can spread following an SI model [58]. The SI model describes the spread of an infection in a population composed of two kinds of individuals: susceptible (S) and infected (I). The former are the ones that can contract the virus, while the latter are the ones already infected. The proportion of infected grows according to the differential equation $\dot{I} = \beta SI/N$, where the parameter $\beta$ is called infection rate.

**Table 6**
Results of $\varepsilon$-BDE and CORA for the multiclass SI models. The *Bound* column in the $\varepsilon$-BDE side refers to the quantity $\lambda \|\cdot\|$.

| | | $\varepsilon$-BDE | | | | CORA | |
|---|---|---|---|---|---|---|---|
| $N$ | $\hat{\tau}$ | *Time (s)* | $\delta$ | $\|\cdot\|$ | *Bound* | *Time (s)* | *Bound* |
| 4 | 1.50 | 5.70E–1 | 3.80E–3 | 8.65E–4 | 1.89E–3 | 3.69E+1 | 1.68E–2 |
| 8 | 0.80 | 2.13E+0 | 2.80E–3 | 1.90E–3 | 4.52E–3 | 1.24E+3 | 4.17E–2 |
| 12 | 0.50 | 6.63E+0 | 2.40E–3 | 2.00E–3 | 5.18E–3 | — | |

More in detail, we consider a multiclass SI model where each cell represents a class of the model. In this case, the individuals are mobile phones that infect each other through messages exchange. We can organize the cells in a network where nodes represent a SI model with $S_i$ susceptible and $I_i$ infected phones. An edge from a node $i$ to a node $j$ has a weight that represents the cross-class infection rate $\beta_{ij}$. The self loops represent the spreading of the disease within the cell.

We generate three fully connected networks with 4, 8, and 12 nodes. We split the nodes in two groups with different parameters. We set up the infection rate $\beta_{ij}$ equal to 1 for node $i$ in the first group and equal to 2 for node $i$ in the second group. In line with other case studies, we perturb these parameters of 0.1% around their fixed value. We expect to get a reduced model where the infected and susceptible belonging to the same group are lumped together. This corresponds to the following partition

$$\left\{\{S_1, ..., S_{\frac{N}{2}}\}, \{S_{\frac{N}{2}+1}, ..., S_N\}, \{I_1, ..., I_{\frac{N}{2}}\}, \{I_{\frac{N}{2}+1}, ..., I_N\}\right\}.$$

As initial conditions, we pick a proportion of the susceptible equal to 0.9 for the first group and 1.0 for the second. We set up different time horizons $\hat{\tau}$ for each network to ensure the system's steady state. We chose a time step equal to 0.01 for both CORA and $\varepsilon$-BDE. We reduce the model with $\varepsilon = 5\text{E}{-3}$. The CORA bound was computed as the maximal flowpipe diameter for the value of species $S_1$ across the time interval.

In Table 6, we report the results for each network considered. Our approach provides *Bounds* one order of magnitude smaller than CORA. Increasing the dimension of the model, the number of parameters grows quickly. CORA goes out of memory for the model with 12 nodes, while we can compute the *Bound* in a small amount of time. The reduced model aggregates variables that are generated from the same group showing that our approach can lump similar equations considering a perturbation of their parameters.

## 6. Conclusion

Reasoning about quantitative properties approximately can represent an effective way of taming the complexity of real systems. Here we have considered ordinary differential equations (ODEs) with polynomial derivatives. We developed notions of equivalence as a relaxation of their exact counterparts, allowing the derivatives of related ODE variables to vary up to a desired tolerance. Our algorithmic approach can be useful to systematically discover quasi-symmetries in situations such as those presented in our case study. The evaluation section shows that our approach is complementary to established over-approximation techniques such as CORA. Our approach handled better smaller uncertainties on larger models, while CORA handled better larger uncertainties on smaller models. In future work, it would be also possible to integrate other bounding techniques, such as [35] which lacks an automatic synthesis of a reference model but can offer a tradeoff between tightness of the bound and computation cost in its derivation.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

No data was used for the research described in the article.

## Acknowledgements

Luca Cardelli is supported by a Royal Society Research Professorship. This work was supported in part by the PRIN project SEDUCE, no. 2017TWRCNB, the Poul Due Jensen Foundation grant no. 883901, and the DFF RP1 project REDUCTO no. 9040-00224B. This work was partially supported by project SERICS (PE00000014) and by project Tuscany Health Ecosystem (THE), CUP: B83C22003920001, under the Italian National Recovery and Resilience Plan funded by the European Union - NextGenerationEU.

## References


[1] J.D. Murray, Mathematical Biology I: An Introduction, 3rd edition, Springer, 2002.
[2] M.S. Okino, M.L. Mavrovouniotis, Simplification of mathematical models of chemical reaction systems, Chem. Rev. 2 (98) (1998) 391–408.
[3] Y. Iwasa, V. Andreasen, S. Levin, Aggregation in model ecosystems. I. Perfect aggregation, Ecol. Model. 37 (3–4) (1987) 287–302.
[4] L. Cardelli, M. Tribastone, M. Tschaikowski, From electric circuits to chemical networks, Nat. Comput. 19 (1) (2020) 237–248.
[5] M. Whitby, L. Cardelli, M. Kwiatkowska, L. Laurenti, M. Tribastone, M. Tschaikowski, PID control of biochemical reaction networks, IEEE Trans. Autom. Control 67 (2) (2022) 1023–1030.
[6] A. Antoulas, Approximation of Large-Scale Dynamical Systems, Advances in Design and Control, SIAM, 2005.
[7] T.J. Snowden, P.H. van der Graaf, M.J. Tindall, Methods of model reduction for large-scale biological systems: a survey of current methods and trends, Bull. Math. Biol. 79 (7) (2017) 1449–1486, https://doi.org/10.1007/s11538-017-0277-2.
[8] M. Aoki, Control of large-scale dynamic systems by aggregation, IEEE Trans. Autom. Control 13 (3) (1968) 246–253, https://doi.org/10.1109/TAC.1968.1098900.
[9] T.P. Prescott, A. Papachristodoulou, Guaranteed error bounds for structured complexity reduction of biochemical networks, J. Theor. Biol. 304 (2012) 172–182, https://doi.org/10.1016/j.jtbi.2012.04.002.
[10] J. Wei, J.C.W. Kuo, Lumping analysis in monomolecular reaction systems. Analysis of the exactly lumpable system, Industrial & Engineering Chemistry Fundamentals 8 (1) (1969) 114–123, https://doi.org/10.1021/I160029a019, http://pubs.acs.org/doi/pdf/10.1021/I160029a019.
[11] L. Cardelli, M. Tribastone, M. Tschaikowski, A. Vandin, Symbolic computation of differential equivalences, in: POPL, 2016.
[12] R. Paige, R. Tarjan, Three partition refinement algorithms, SIAM J. Comput. 16 (6) (1987) 973–989.
[13] L. Cardelli, M. Tribastone, M. Tschaikowski, A. Vandin, Efficient syntax-driven lumping of differential equations, in: Tools and Algorithms for the Construction and Analysis of Systems — 21st International Conference, TACAS, 2016, http://cse.lab.imtlucca.it/~mirco.tribastone/papers/tacas2016.pdf.
[14] L. Cardelli, M. Tribastone, M. Tschaikowski, A. Vandin, Maximal aggregation of polynomial dynamical systems, Proc. Natl. Acad. Sci. 114 (38) (2017) 10029–10034.
[15] L. Cardelli, M. Tribastone, M. Tschaikowski, A. Vandin, Symbolic computation of differential equivalences, Theor. Comput. Sci. 777 (2019) 132–154.
[16] M. Tschaikowski, M. Tribastone, Exact fluid lumpability for Markovian process algebra, in: CONCUR, 2012, pp. 380–394.
[17] M. Tschaikowski, M. Tribastone, Tackling continuous state-space explosion in a Markovian process algebra, Theor. Comput. Sci. 517 (2014) 1–33.
[18] M. Tribastone, Behavioral relations in a process algebra for variants, in: S. Gnesi, A. Fantechi, P. Heymans, J. Rubin, K. Czarnecki, D. Dhungana (Eds.), SPLC, ACM, 2014, pp. 82–91.
[19] M. Tschaikowski, M. Tribastone, Spatial fluid limits for stochastic mobile networks, Perform. Eval. 109 (2017) 52–76.
[20] M.A. Islam, A. Murthy, E. Bartocci, E. Cherry, F.H. Fenton, J. Glimm, S.A. Smolka, R. Grosu, Model-order reduction of ion channel dynamics using approximate bisimulation, Theor. Comput. Sci. 599 (2015) 34–46, https://doi.org/10.1016/j.tcs.2014.03.018.
[21] L. Cardelli, I.C. Pérez-Verona, M. Tribastone, M. Tschaikowski, A. Vandin, T. Waizmann, Exact maximal reduction of stochastic reaction networks by species lumping, Bioinform. 37 (15) (2021) 2175–2182.
[22] L. Cardelli, M. Tribastone, M. Tschaikowski, A. Vandin, Syntactic Markovian bisimulation for chemical reaction networks, in: L. Aceto, G. Bacci, G. Bacci, A. Ingólfsdóttir, A. Legay, R. Mardare (Eds.), Models, Algorithms, Logics and Tools - Essays Dedicated to Kim Guldstrand Larsen on the Occasion of His 60th Birthday, vol. 10460, 2017, pp. 466–483.
[23] M. Boreale, Algebra, coalgebra, and minimization in polynomial differential equations, in: FOSSACS, 2017, pp. 71–87.
[24] M. Althoff, C.L. Guernic, B.H. Krogh, Reachable set computation for uncertain time-varying linear systems, in: M. Caccamo, E. Frazzoli, R. Grosu (Eds.), HSCC, ACM, 2011, pp. 93–102.
[25] A. Das, G.V. Merrett, M. Tribastone, B.M. Al-Hashimi, Workload change point detection for runtime thermal management of embedded systems, IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst. 35 (8) (2016) 1358–1371.
[26] M. Kozlov, S. Tarasov, L. Khachiyan, The polynomial solvability of convex quadratic programming, USSR Comput. Math. Math. Phys. 20 (5) (1980) 223–228.
[27] W. E, T. Li, E. Vanden-Eijnden, Optimal partition and effective dynamics of complex networks, PNAS 105 (23) (2008) 7907–7912.
[28] R. Lal, P. Prabhakar, Bounded error flowpipe computation of parameterized linear systems, in: EMSOFT, 2015, pp. 237–246.
[29] J. Rosenfeld, E.G. Friedman, Design methodology for global resonant H-tree clock distribution networks, IEEE Trans. Very Large Scale Integr. (VLSI) Syst. 15 (2) (2007) 135–148, https://doi.org/10.1109/TVLSI.2007.893576.
[30] M. Althoff, An introduction to CORA 2015, in: Proc. of the Workshop on Applied Verification for Continuous and Hybrid Systems, 2015.
[31] M. Althoff, Reachability analysis of nonlinear systems using conservative polynomialization and non-convex sets, in: HSCC, 2013, pp. 173–182.
[32] A. Girard, G. Pappas, Approximate bisimulations for nonlinear dynamical systems, in: IEEE Conference on Decision and Control and European Control Conference, 2005.
[33] R. Majumdar, M. Zamani, Approximately bisimilar symbolic models for digital control systems, in: CAV, 2012, pp. 362–377.
[34] P.S. Duggirala, S. Mitra, M. Viswanathan, Verification of annotated models from executions, in: EMSOFT, IEEE Press, 2013, pp. 26:1–26:10.
[35] M. Tschaikowski, M. Tribastone, Approximate reduction of heterogeneous nonlinear models with differential hulls, IEEE TAC.
[36] E. Asarin, T. Dang, A. Girard, Reachability analysis of nonlinear systems using conservative approximation, in: HSCC, 2003.
[37] X. Chen, E. Ábrahám, S. Sankaranarayanan, Flow*: an analyzer for non-linear hybrid systems, in: CAV, 2013, pp. 258–263.
[38] J.A. Farrell, M.M. Polycarpou, Adaptive Approximation Based Control, Wiley-Interscience, 2006.
[39] C. Fan, B. Qi, S. Mitra, M. Viswanathan, P.S. Duggirala, Automatic reachability analysis for nonlinear hybrid models with C2E2, in: CAV, 2016, pp. 531–538.
[40] J.C.W. Kuo, J. Wei, Lumping analysis in monomolecular reaction systems. Analysis of approximately lumpable system, Ind. Eng. Chem. Fundam. 8 (1) (1969) 124–133.
[41] G. Li, H. Rabitz, A general analysis of approximate lumping in chemical kinetics, Chem. Eng. Sci. 45 (4) (1990) 977–1002.
[42] Y. Iwasa, S.A. Levin, V. Andreasen, Aggregation in model ecosystems II. Approximate aggregation, Math. Med. Biol. 6 (1) (1989) 1–23.
[43] M. Tschaikowski, M. Tribastone, A unified framework for differential aggregations in Markovian process algebra, JLAMP 84 (2) (2015) 238–258.
[44] G. Iacobelli, M. Tribastone, Lumpability of fluid models with heterogeneous agent types, in: DSN, 2013.
[45] L. Cardelli, M. Tribastone, M. Tschaikowski, A. Vandin, Guaranteed error bounds on approximate model abstractions through reachability analysis, in: A. McIver, A. Horvath (Eds.), Quantitative Evaluation of Systems, Springer International Publishing, Cham, ISBN 978-3-319-99154-2, 2018, pp. 104–121.
[46] L. Cardelli, M. Tribastone, M. Tschaikowski, A. Vandin, Forward and backward bisimulations for chemical reaction networks, in: CONCUR, 2015, pp. 226–239.
[47] D. Gillespie, Exact stochastic simulation of coupled chemical reactions, J. Phys. Chem. 81 (25) (1977) 2340–2361.
[48] P.M. Pardalos, S.A. Vavasis, Quadratic programming with one negative eigenvalue is NP-hard, J. Glob. Optim. 1 (1) (1991) 15–22.
[49] L. Perko, L. Perko, Differential Equations and Dynamical Systems, Texts in Applied Mathematics, Springer, New York, 1996.
[50] A. Donzé, O. Maler, Systematic simulation using sensitivity analysis, in: HSCC, Springer, 2007, pp. 174–189.
[51] T. Turányi, A.S. Tomlin, Reduction of Reaction Mechanisms, Springer Berlin Heidelberg, Berlin, Heidelberg, 2014, pp. 183–312.

[52] H. Conzelmann, D. Fey, E. Gilles, Exact model reduction of combinatorial reaction networks, BMC Syst. Biol. 2 (1) (2008) 78.
[53] F. Camporesi, J. Feret, Formal reduction for rule-based models, Electron. Notes Theor. Comput. Sci. 276 (2011) 29–59.
[54] M.W. Sneddon, J.R. Faeder, T. Emonet, Efficient modeling, simulation and coarse-graining of biological complexity with NFsim, Nat. Methods 8 (2) (2011) 177–183.
[55] M.I. Monine, R.G. Posner, P.B. Savage, J.R. Faeder, W.S. Hlavacek, Modeling multivalent ligand-receptor interactions with steric constraints on configurations of cell-surface receptor aggregates, Biophys. J. 98 (1) (2010) 48–56.
[56] H. Conzelmann, J. Saez-Rodriguez, T. Sauter, B. Kholodenko, E. Gilles, A domain-oriented approach to the reduction of combinatorial complexity in signal transduction networks, BMC Bioinform. 7 (1) (2006) 34.
[57] P. Wang, M.C. González, C.A. Hidalgo, A.-L. Barabási, Understanding the spreading patterns of mobile phone viruses, Science 324 (5930) (2009) 1071–1076.
[58] R.M. Anderson, R.M. May, Infectious Diseases of Humans: Dynamics and Control, Oxford University Press, 1992.