Approximate Differential Equivalences

Luca Cardelli[1], Mirco Tribastone[2], Max Tschaikowski[2], and Andrea Vandin[2]

[1] Microsoft Research & University of Oxford, UK
[2] IMT School for Advanced Studies Lucca, Italy

Abstract. The problem of comparison and minimization for ordinary differential equations (ODEs) through behavioral equivalences has been tackled so far using exact notions. However these may not be robust enough in practice because they are highly sensitive to the specific choice of parameters. We study ODEs with polynomial derivatives of any degree, covering linear/affine systems and chemical reaction networks with mass-action kinetics. For this class, we introduce *approximate differential equivalence* as a more permissive notion that allows variables of an ODE system to be related even if they vary in their parameters up to a given tolerance. We develop algorithms to *(i)* compute the largest approximate differential equivalence, *(ii)* construct a quotient approximate model from the original one via an appropriate perturbation of the parameters, and *(iii)* provide a formal certificate on the quality of the approximation. Finally, we apply approximate differential equivalences for symmetry reductions in protein interaction networks and models of polymerization in organic chemistry.

1 Introduction

Ordinary differential equations (ODEs) are a prominent model of dynamical systems across many branches of science and engineering, and have enjoyed increasing popularity in computer science, for instance, in computational systems biology [12,4,6,9,32], as an approximation to large-scale Markov models [21,34], and as the governing laws of continuous motion in hybrid systems [15]. This has motivated techniques for the comparison and minimization of ODEs based on behavioral relations, along the lines of analogous efforts in other foundational quantitative models of computation, e.g. [27]. Here we consider *differential equivalence* [8], developed as an equivalence over ODE variables yielding a quotient system that preserves the dynamics of the original one. However differential equivalences (reviewed in Section 2) are exact, hence highly sensitive to parameter values and initial conditions. This may hinder their practical usability in some applications domains, for instance due to parameter uncertainty arising from the well-known difficulty in observing and measuring events in biological processes.

Our objective is to develop approximate variants of differential equivalence (in Section 3). A key assumption is to study systems with derivatives that are multivariate polynomials (over the ODE variables) of any degree, thus restricting the scope of [8]. However we remark that this is still quite a generous class since it includes linear/affine systems (hence, for instance continuous-time Markov chains represented through their Kolmogorov equations) and chemical reaction networks (CRNs) with mass-action kinetics. Further, Bernstein polynomials can interpolate derivatives that are

arbitrary continuous functions, with higher accuracy of the interpolation the higher the polynomial degree.

Considering polynomial derivatives allows us to introduce a notion of equivalence that just concerns the ODE "syntax" (while the nonlinear class of ODEs of [8] required symbolic SMT-based checks). Our main idea is to consider a threshold parameter $\varepsilon \geq 0$ which, roughly speaking, measures the tolerance with which we accept the coefficients of the polynomials to differ for still relating the respective variables. Like in other established approaches such as behavioral pseudometrics (e.g., [13,14,3]), $\varepsilon = 0$ corresponds to an exact differential equivalence of [8]. In addition to defining criteria for approximate differential equivalences, we provide an algorithm for obtaining the largest one (in Section 3). This is done via partition refinement, computing the coarsest refinement of a given initial partition of ODE variables for a given tolerance ε.

In Section 3 we define a quotient ODE system by considering a "perturbation" of the coefficients of the original model which makes the given approximate differential equivalence an exact one. We consider a metric (the Euclidean norm) to measure the degree of perturbation, and take the model that minimizes it. This can be done efficiently by solving an optimization problem, which runs polynomially with the size of the ODE system [25]. This approach is analogous to optimal approximate lumpings for Markov chains (e.g., [17,29]), although our theory can be applied to other choices of quotients.

The quotient ODE represents an approximation to the original model. This approximation is obtained through a transformation, induced by the choice of ε, that involves the structure/syntax of the model. In Section 4 we study the repercussion on the *semantics* (i.e., on the ODE solution), developing a technique to derive a *formal certificate* for the quality of such approximation. This is expressed in terms of an ε-δ argument (similar in spirit to the ones routinely used in calculus). Informally, it states the following: for any choice of the structural tolerance ε, there exist a degree of perturbation δ and an *amplifier* λ such that, for any ODE obtained by applying a perturbation to the reference of at most δ, at all time points the difference between the solution of the reference and such perturbation is at most λ times the perturbation. We take the quotient ODE and compute its solution, which we call the *reference trajectory*. Then, the original ODE system is linearized around the reference trajectory (which is straightforward in the case of polynomial ODE systems). On the linearized system we set up a control problem that *maximizes* the deviation from the reference trajectory, which represents our desired bound (the λ) stated with respect to the linearized system. Owing to linearity, this problem admits a closed form solution which can be solved efficiently. In a second step we take also into account the nonlinear part of the dynamics and provide a conservative condition (the δ) that ensures that the linear system does describe the behavior of the original nonlinear one sufficiently well.

In Section 5 we will present numerical results on case studies of approximate symmetry reductions in protein interaction networks and models of polymerization in organic chemistry.

Further related work. Research on approximate aggregations of ODE systems spans many disciplines. Here we consider the most closed related approaches. In chemistry, the analysis of approximate aggregation can be traced back to Kuo and Wei [26]. They studied monomolecular reaction networks, which give rise to affine ODE systems. The

approximation consists in a *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 error analysis of [26] allows the estimation of the maximum difference between the original traces and the approximately aggregated ones, but considers the special case where the transition matrix is diagonalizable; in addition, it is assumed that a lumping scheme is given, whereas in this paper we compute a candidate quotient ODE automatically. Li and Rabitz extend approximate lumping to general CRNs (e.g., bimolecular ones) [28], but an explicit error bound is not given.

In a similar vein, approximate aggregation in ecology has been studied from the point of view of finding an aggregate ODE system whose derivatives are as close as possible (in norm) to the derivatives of the original ODE system [24], where the 0-distance corresponds to exact (also called *perfect* [23]) aggregation. The justification that variables with nearby derivatives will have nearby solutions is grounded on Gronwall's inequality [38], which is also at the basis of more recent approximation techniques developed for process algebra with ODE semantics [35] and for a class of ODEs modeling so-called *heterogeneous systems* [22]. Unfortunately, it is well-known that Gronwall's inequality offers only a result of asymptotical soundness, in the sense that the solutions of related variables may be shown to correspond only in the limit of infinitesimally small perturbations. The actual bound estimates, instead, cannot in general be used in practice as they tend to grow exponentially with respect to time.

In control theory, Pappas *et al.* introduce the notion of approximate bisimulation for general nonlinear dynamical systems [20]. The distance between the trajectories of the original ODE model and those of the reduced model can be bounded by a Lyapunov-like function. A similar approach is taken in [39] which makes use of so-called incremental Lyapunov functions. In general these functions are known to be hard to find automatically; if the search is constrained to a specific class of candidate functions for which computational methods are available (e.g., semidefinite programming for sum-of-squares functions), the existence of a solution may not be guaranteed. In addition, these techniques require a candidate reduced model to be known a priori.

Another approach is based on the idea of hybridization [1] where the state space is partitioned into small cells and the original nonlinear system is approximated by a continuous function that is affine on each cell. By decreasing the maximal diameter of all cells, the nonlinear system converges to the continuous piecewise affine function. Since tight bounds on the reachable set of uncertain ODE systems with affine derivatives can be efficiently computed, see references in [1], the approach provides tight bounds on the reachable set of nonlinear systems if the maximal diameter is sufficiently small. Unfortunately, the number of cells grows exponentially as the maximal diameter is decreased, meaning that the approach suffers from the curse of dimensionality.

Technically, our bound is closely connected to gain scheduling [18], a well established concept in control theory. While this approach has been used to estimate the reachable set of time-varying linear systems in the past [16,15], to the best of our knowledge, for the case of nonlinear systems no formal bound has been established. By taking advantage of the fact that we are working with polynomials, we accomplish this task.

Notation and basic definitions. We denote the infinity norm $\|\cdot\|_\infty$ by $\|\cdot\|$, while $\|\cdot\|_2$ is the Euclidian norm. Whenever convenient, for a given partition of variables $\mathcal{H}$, we write

$H = \{x_{H,1}, \ldots, x_{H,|H|}\}$ for any $H \in \mathcal{H}$. We denote by $\psi[t/s]$ the term that arises by replacing each occurrence of t in ψ by s. Let S be a set, and $\mathcal{H}_1$, $\mathcal{H}_2$ two partitions of S. Then, $\mathcal{H}_1$ is a *refinement* of $\mathcal{H}_2$ if for any block $H_1 \in \mathcal{H}_1$ there exists a block $H_2 \in \mathcal{H}_2$ such that $H_1 \subseteq H_2$. For any partition $\mathcal{H}$ of S, let $\sim_{\mathcal{H}}$ denote the unique equivalence relation with $\mathcal{H} = S/\!\sim_{\mathcal{H}}$. The transitive closure of a relation $\sim$ is denoted by $\sim^*$.

2 Background

Throughout the paper we consider a polynomial initial value problem (PIVP) over the set of ODE variables $\mathcal{S} = \{x_1, \ldots, x_n\}$. It is defined by the ODEs $\dot{x}_i = q_i$, $1 \leq i \leq n$, where q_i is a multivariate polynomial over $\mathcal{S}$. A polynomial q_i is given in the normal form if 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. The normal form of a polynomial q_i is denoted by $\mathcal{N}(q_i)$. Without loss of generality, we assume that the polynomials q_i of a PIVP $\dot{x} = q(x)$, where $q = (q_i)_{x_i \in \mathcal{S}}$, are given in normal form.

For a polynomial q_i in normal form with variables in $\mathcal{S}$, let $c(q_i, x^\alpha)$ denote the coefficient of the monomial x^α, where $\alpha \in \mathbb{N}_0^{\mathcal{S}}$. The initial condition of the PIVP is given by $\sigma : \mathcal{S} \to \mathbb{R}$; $x_i(t)$ denotes the unique solution for variable x_i at time point t starting from $x_i(0) = \sigma(x_i)$. We use the following ODE system as a running example.

Example 1. Consider the ODE system with variables $\mathcal{S} = \{x_1, x_2, x_3\}$, given by

$$\dot{x}_1 = -4.00x_1 + x_2 + x_3 \qquad \dot{x}_2 = 1.99x_1 - x_2 \qquad \dot{x}_3 = 2.01x_1 - x_3 \quad (1)$$

In [8] two variants of differential equivalence were introduced for IDOL, a class of nonlinear ODE systems covering derivatives more general than polynomials. Here we find it convenient to restate them for a PIVP. (The proofs for this correspondence are straightforward hence we omit them.)

We begin with *backward differential equivalence* (BDE), which relates variables that have the same solutions at all time points. The definition of BDE for PIVP makes pairwise comparisons between the coefficients of any two variables in the same equivalence class.

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

$$\wp_{i,j}^{\mathcal{H}} := (q_i - q_j)\big[x_{H',1}/x_{H'}, \ldots, x_{H',|H'|}/x_{H'} : H' \in \mathcal{H}\big]$$

i.e., when

$$\sum_{\alpha \in \mathbb{N}_0^{\mathcal{S}}} |c(\wp_{i,j}^{\mathcal{H}}, x^\alpha)| = 0. \quad (2)$$

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

Essentially, establishing that a candidate partition is a BDE consists in comparing the coefficients of the monomials of the ODEs of related variables, up to the natural equivalence class induced on monomials by the equivalence relation through $\wp_{i,j}^{\mathcal{H}}$. For instance, the partition $\{\{x_1\}, \{x_2, x_3\}\}$ will equate the monomials x_1x_2 and x_1x_3. Then, for any two variables in the same block it must hold that the differences between the coefficients of the same monomials (modulo the induced equivalence class) 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

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

Forward differential equivalence (FDE) identifies a partition that induces a quotient ODE that tracks sums of variables in each equivalence class. For instance, for any initial condition we have that $\{\{x_1\}, \{x_2, x_3\}\}$ is an FDE for (1) because we can find an ODE system for $x_2 + x_3$:

$$\dot{x}_1 = -4.00x_1 + (x_2 + x_3) \qquad (\dot{x_2 + x_3}) = 4.00x_1 - (x_2 + x_3)$$

The change of variable $x_{23} = x_2 + x_3$ gives us the quotient ODE $\dot{x}_1 = -4.00x_1 + x_{23}$, $\dot{x}_{23} = 4.00x_1 - x_{23}$. From this we conclude that the solution satisfies $x_{23}(t) = x_2(t) + x_3(t)$ for all time t if this holds for the initial condition, i.e., $x_{23}(0) = x_2(0) + x_3(0)$.

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

Definition 2 (FDE). *Fix a PIVP, a partition $\mathcal{H}$ of $\mathcal{S}$ and write $x_i \sim^F_{\mathcal{H}} x_j$ if all coefficients of the polynomial $\sum_{H\in\mathcal{H}} \wp^H_{i,j}$ are zero, where*

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

That is, when

$$\sum_{k=1}^{m} \sum_{\alpha \in \mathbb{N}_0^{\mathcal{S}\cup\{s\}}} |c(\wp^{H_k}_{i,j}, x^\alpha)| = 0. \tag{3}$$

$\mathcal{H}$ is an FDE when $\mathcal{H} = \mathcal{S}/(\sim^{F^*}_{\mathcal{H}} \cap \sim_{\mathcal{H}})$.

3 Approximate Differential Equivalences

Definitions. Approximate differential equivalence relaxes the equality conditions (2)-(3) of Definition 1 and 2 to inequalities with respect to a tolerance level ε.

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 $x_i \sim^B_{\mathcal{H},\varepsilon} x_j$ if $\sum_{\alpha\in\mathbb{N}_0^{\mathcal{S}}} |c(\wp^{\mathcal{H}}_{i,j}, x^\alpha)| \leq \varepsilon$, where $\wp^{\mathcal{H}}_{i,j}$ is as in Definition 1. A partition $\mathcal{H}$ is an ε-BDE if $\mathcal{H} = \mathcal{S}/(\sim^{B^*}_{\mathcal{H},\varepsilon} \cap \sim_{\mathcal{H}})$.*

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 $x_i \sim^F_{\mathcal{H},\varepsilon} x_j$ if $\sum_{k=1}^{m} \sum_{\alpha\in\mathbb{N}_0^{\mathcal{S}\cup\{s\}}} |c(\wp^{H_k}_{i,j}, x^\alpha)| \leq \varepsilon$, where $\wp^H_{i,j}$ is as in Definition 2. A partition $\mathcal{H}$ is an ε-FDE when $\mathcal{H} = \mathcal{S}/(\sim^{F^*}_{\mathcal{H},\varepsilon} \cap \sim_{\mathcal{H}})$.*

Algorithm 1 Template partition refinement algorithm for the computation of the coarsest ε-FDE/ε-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
    $\mathcal{H}' \longleftarrow S/(\sim^{\chi^*}_{\mathcal{H},\varepsilon} \cap \sim_h)$
    if $\mathcal{H}' = \mathcal{H}$ **then**
      return $\mathcal{H}$
    else
      $\mathcal{H} \longleftarrow \mathcal{H}'$
    end if
  end while

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 [8]. Since these two notions have similar structure in the rest of this paper we will illustrate only approximate BDE using small examples. Instead, both notions will be discussed in more detail for the numerical evaluation of Section 5.

Example 3. Let us consider our running example (1). Then, the partition $\big\{\{x_1\}, \{x_2, x_3\}\big\}$ is a 0.02-BDE partition, as can be easily seen from Example 2.

The next two results show the existence of largest approximate differential equivalences and of a partition-refinement algorithm to compute it. [1]

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

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

We now study how efficiently the conditions for approximate differential equivalence can be computed. The notions are defined with respect to the coefficients of the polynomials $\wp^H_{i,j}$ and $\wp^{\mathcal{H}}_{i,j}$ and thus require the computation of their normalization. In the case of ε-FDE this yields exponential complexity, due to term replacement. To see this, consider for instance the program $\dot{x}_1 = x_2^k, \dot{x}_2 = x_1^k$, for some $k > 0$. Then, for $\mathcal{H} = \big\{\{x_1, x_2\}\big\}$, the term $q_1[x_1/s(x_1 + x_2), x_2/(1 - s)(x_1 + x_2)]$ will be of size $\mathcal{O}(2^k)$. This stands in stark contrast to ε-BDE, where 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 Π such that, under the assumptions of Theorem 2, the number of steps done by Algorithm 1 is $\mathcal{O}\big(\Pi(2^d \cdot p)\big)$ if $\chi = F$ and $\mathcal{O}\big(\Pi(p)\big)$ if $\chi = B$, respectively, where d is the maximum degree of the polynomial and p is the number of monomials present in the PIVP.*

[1] Note to reviewers: proofs are provided as supplementary material.

In practice, d is usually not large. For example, mass-action CRNs feature ODEs with degree-two polynomials because in nature at most two species interact in every reaction [19]. An experimental comparison between the reduction runtimes of ε-FDE and ε-BDE will be presented in Section 5. Since 0-FDE/BDE coincides with FDE/BDE, it is interesting to note that the above result provides, in its own right, a novel complexity bound for a subclass of ODE systems considered in [8].

Computational considerations also motivate the choice of the definitions of approximate differential equivalence given in this paper. A different natural definition could have been to compute 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 ε-BDE would require checking that the difference between the derivatives satisfies

$$|\dot{x}_1 - \dot{x}_2| = |x_1^3 - x_1 + x_2^3 - x_2| \leq \varepsilon, \text{for all } 0 \leq x_1, x_2 \leq C \tag{4}$$

for some finite $C > 0$ that represents some bounded domain where the trajectories are assumed to live. Since this example shows that this question is in general equivalent to solving a nonconvex optimization problem, we infer that the problem is NP-hard [31].

However it can be easily shown that our approximate differential equivalence, defined through the coefficients of the polynomials, is equivalent to checks such as (4) in the following sense: If a partition $\mathcal{H}$ satisfies constraints similar to (4) with respect to some $\varepsilon > 0$, then there exists an $\varepsilon' > 0$ such that $\mathcal{H}$ is an ε'-FDE/BDE, and vice versa.

Finally, we remark that our structural/syntactic criteria can be used for PIVPs only. It is the lack of analogous conditions in the case of more general functions like minima or roots which prevents our approximate differential equivalences to be extended in a straightforward way to the full class of nonlinear ODEs of [8].

Reference PIVP. 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 σ and the coefficients present in $q_1, \ldots, q_n$ — which ensures that that very partition becomes an exact differential equivalence. On this reference PIVP one can use the quotienting algorithms for FDE/BDE developed in [8] (and not restated here formally for brevity). 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 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 $c(q_i, x^\alpha)$ by the parameter $\mathfrak{c}(\hat{q}_i, x^\alpha)$.*

Example 4. 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* $\mathcal{P} = \hat{\mathcal{P}}(\hat{\sigma}_0)$*.*

Let us provide the extended version of our running example.

Example 5. 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 $1 \le i \le 3$ and

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

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

Definition 7. *Given constant free polynomial* $\hat{\wp}$ *and* $\Xi \subseteq \mathcal{S} \cup \Theta \cup \{s\}$*, let* $\mathfrak{t}(\hat{\wp}, x^\alpha, \Xi)$ *denote the coefficient term of* x^α *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* Ξ *are treated as parameters.*

Example 6. With $\hat{q}_2$ and $\hat{q}_3$ as in Example 4 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 $\mathfrak{t}(\hat{q}_2 - \hat{q}_3, x_1, \Xi) = \mathfrak{c}(\hat{q}_2, x_1) - \mathfrak{c}(\hat{q}_3, x_1)$.

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

$$\big\{\mathfrak{t}(\tilde{\wp}^H_{i,j}, x^\alpha, \mathcal{S} \cup \{s\}) = 0 \mid \alpha \in \mathbb{N}_0^{\mathcal{S}\cup\{s\}}, H \in \mathcal{H} \text{ and } x_i \sim_{\mathcal{H}} x_j\big\} \tag{5}$$

with $\tilde{\wp}^H_{i,j} = \sum_{x_k \in H} \hat{q}_k - \sum_{x_k \in H} \hat{q}_k[x_i/s(x_i + x_j), x_j/(1-s)(x_i + x_j)]$.

If $\mathcal{H}$ *is an* ε*-BDE partition of* $\mathcal{S}$*, the corresponding set of linear constraints is*

$$\begin{aligned}
&\big\{\mathfrak{t}(\tilde{\wp}^{\mathcal{H}}_{i,j}, x^\alpha, \mathcal{S}) = 0 \mid \alpha \in \mathbb{N}_0^{\mathcal{S}}, x_i \sim_{\mathcal{H}} x_j\big\}\\
&\qquad \cup \big\{x_{i_j} - x_{i_{j+1}} = 0 \mid 1 \le j \le k-1 \text{ and } \{x_{i_1}, \ldots, x_{i_k}\} \in \mathcal{S}/{\sim_{\mathcal{H}}}\big\},
\end{aligned} \tag{6}$$

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

Example 7. 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 approximate BDE is "useful" under the further constraint that related variables have the same initial conditions. This translates into adding the constraints in (6) 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 ε-FDE, instead, only constraints on the parameters Θ are made.

Theorem 4. *Given a PIVP $\mathcal{P}$ with variables $\mathcal{S}$, an ε-FDE/BDE partition $\mathcal{H}$ and a configuration $\hat{\sigma} \in \mathbb{R}^{\mathcal{S}\cup\Theta}$ that satisfies (5)/(6), it holds that $\mathcal{H}$ is an FDE/BDE of $\hat{\mathcal{P}}(\hat{\sigma})$.*

The system from Theorem 4 can be shown to be underdetermined, hence there are infinitely many perturbations that lead to an exact differential equivalence. This observation is an instance of the well-known fact that, in general, an approximate reduction is not unique. 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.

We are interested in finding a configuration $\hat{\sigma}$ which satisfies the constraints of Theorem 4 and minimizes the distance $\|\hat{\sigma} - \hat{\sigma}_0\|_2$. Mathematically, this corresponds to the optimization problem

$$\hat{\sigma}_* = \underset{\hat{\sigma}:\text{Eq. (5)/(6) holds}}{\operatorname{argmin}} \|\hat{\sigma} - \hat{\sigma}_0\|_2 \tag{7}$$

Since the solution space of a linear system is convex, the Euclidian norm yields a convex quadratic program that can be solved in polynomial time [25].

Example 8. Let us continue Example 7 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:

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

The above discussions are summarized in the following.

Theorem 5. *Given a PIVP, $\varepsilon \geq 0$, and an ε-FDE/BDE partition $\mathcal{H}$, the solution of (7) exists and can be computed in polynomial time.*

The solution of the optimization problem (7) 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 tight bound on the difference between the solution of the original PIVP and the reference. More specifically, we will show how to compute two values $\delta > 0$ and $\lambda > 0$ such that for all initial conditions $\hat{\sigma}_1 \in \mathbb{R}^{S\cup\Theta}$ with $\|x^{\hat{\sigma}_1}(0) - x^{\hat{\sigma}_*}(0)\| \leq \delta$, it holds that $\max_{0\leq t\leq\hat{\tau}} \|x^{\hat{\sigma}_1}(t) - x^{\hat{\sigma}_*}(t)\| \leq \lambda\|x^{\hat{\sigma}_1}(0) - x^{\hat{\sigma}_*}(0)\|$, where $x^{\hat{\sigma}}$ denotes the solution underlying $\hat{\mathcal{P}}(\hat{\sigma})$ and $\hat{\tau} > 0$ is a previously fixed finite time horizon.

The quantity δ gives the size of the ball around the initial condition $\hat{\sigma}_*$ of the reference PIVP, whereas λ is the *amplifier* that relates the maximum distance between trajectories

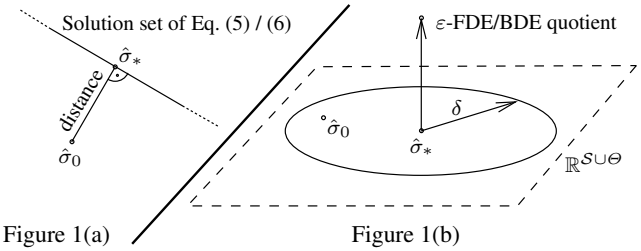


Figure 1(a) Figure 1(b)

Fig. 1: Given a PIVP $\mathcal{P}$, a partition $\mathcal{G}$ of $\mathcal{S}$, and an $\varepsilon > 0$, the coarsest ε-FDE/BDE partition $\mathcal{H}$ that refines $\mathcal{G}$ is constructed. Afterwards, the solution $\hat{\sigma}_*$ of the optimization problem (7) is computed in Fig. 1(a). This allows to compute the ε-FDE/BDE quotient $\hat{\mathcal{P}}(\hat{\sigma}_*)$ of $\mathcal{H}$. With this, λ and δ from Theorem 7 are calculated. In the case the distance between $\hat{\sigma}_0$ and $\hat{\sigma}_*$ does not exceed δ, the tight bounds of Theorem 7 can be applied and relate, in particular, the trajectories of $\hat{\mathcal{P}}(\hat{\sigma}_*)$ and $\hat{\mathcal{P}}(\hat{\sigma}_0) = \mathcal{P}$, as depicted in Fig. 1(b).

to the distance between the initial conditions. Therefore, if the initial condition of the original PIVP $\hat{\mathcal{P}}(\hat{\sigma}_0)$ falls within the prescribed δ ball, then the above statement will provide a formal bound of the error made in approximating the original PIVP $\hat{\mathcal{P}}(\hat{\sigma}_0)$ with the reference PIVP. This idea is visualized in Fig. 1(b).

Let us recall 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 ∂_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 $\dot{y}(t) = A(t)y(t)$, where $y(t_0) = y_0$ and $t_0 \leq t \leq \hat{\tau}$, is given by $y(t) = \Lambda(t_0, t)y_0$ for all $t_0 \leq t \leq \hat{\tau}$.*

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 $\Lambda(t_0, t_1)e_{x_i}$ is the x_i-column of $\Lambda(t_0, t_1)$, this shows that $\Lambda(t_0, t_1)$ can be efficiently computed by solving $|\mathcal{S} \cup \Theta|$ instances of the linear ODE system from Theorem 6. Those quantities are needed in the following major result.

Theorem 7. *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* $\|y\| \leq 1$ *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* $\sum_{k=2}^{\deg(\hat{\mathcal{P}})} d_k (2\lambda_0\delta)^{k-1} \leq (2\lambda_1\hat{\tau})^{-1}$.

Theorem 7 provides a bound on the difference $x^{\hat{\sigma}_1}(t) - x^{\hat{\sigma}_*}(t)$ in terms of the initial perturbation $x^{\hat{\sigma}_1}(0) - x^{\hat{\sigma}_*}(0)$ if the latter is sufficiently small, i.e., does not exceed δ. 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 are bounded by d_k along the reference trajectory $x^{\hat{\sigma}_*}$.

The next result shows that our bound is tight.

Theorem 8. *If an extended PIVP* $\hat{\mathcal{P}}$ *satisfies* $\deg(\hat{\mathcal{P}}) = 1$ *and* $\lambda = 2\lambda_0$*, it holds that*

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

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

Note that the amplifier in Theorem 7 is twice as large as the amplifier in Theorem 8. This is because the proof of Theorem 7 has to account for nonlinear terms present in the remainder function r. More importantly, Theorem 8 shows that the amplifier of Theorem 7 cannot be substantially improved.

The following results simplifies the constraints on δ from Theorem 7 in the case of $\deg(\hat{\mathcal{P}}) \leq 3$.

Lemma 1. *In the case where* $\deg(\hat{\mathcal{P}}) \leq 3$*, the constraint on* δ *of Theorem 7 simplifies to* $\delta \leq \left[2\hat{\tau}\lambda_0\lambda_1\left(d_2 + \sqrt{d_2^2 + \frac{2d_3}{\lambda_1\hat{\tau}}}\right)\right]^{-1}$.

The above lemma applies, for instance, to most CRNs, 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 2. *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 7 to* $C^{\deg(\hat{\mathcal{P}})-k} MD$*.*

Example 9. Since $\deg(\hat{\mathcal{P}}) = 2$ in Example 5, coefficients $d_3, d_4, \ldots$ are zero and we only need to bound d_2. Moreover, the constraint in Theorem 7 simplifies to $\delta \leq (4\hat{\tau}\lambda_0\lambda_1 d_2)^{-1}$ thanks to Lemma 1. By applying Lemma 2, 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 $\delta \leq 0.02$.

Table 1: Binding model results. The time horizon $\hat{\tau}$ is given in 10^{-3} time units.

Model			*Rates of bindings*							0.18-FDE	0.04-BDE		*Reference model*			
BS	$\lvert R\rvert$	$\lvert S\rvert$	k_{b_1}	k_{b_2}	k_{b_3}	k_{b_4}	k_{b_5}	k_{b_6}	k_{b_7}	*Runtime (s)*		$\lvert\mathcal{H}\rvert$	$\hat{\tau}$	λ	δ	$\lVert\cdot\rVert$
2	8	5	9.99000	10.01000	—	—	—	—	—	0.2	0.001	4	6.0	1.60	2.6E-2	1.0E-2
3	24	9	9.99000	10.00000	10.01000	—	—	—	—	0.1	0.002	5	2.0	1.98	1.9E-2	1.0E-2
4	64	17	9.99800	9.99900	10.00000	10.00100	—	—	—	1.1	0.005	6	1.4	2.07	4.2E-3	2.0E-3
5	160	33	9.99920	9.99960	10.00000	10.00040	10.00080	—	—	4.6	0.020	7	1.1	2.10	1.3E-3	8.0E-4
6	384	65	9.99985	9.99990	9.99995	10.00000	10.00005	10.00010	—	23	0.070	8	0.9	2.12	3.0E-4	1.5E-4
7	896	129	9.99997	9.99998	9.99999	10.00000	10.00001	10.00002	10.00003	190	0.400	9	0.7	2.21	6.9E-5	3.0E-5

5 Case Studies

Protein-interaction networks. A recurring case in computational systems biology is the dynamics of complexes such as receptors and scaffold proteins, which have multiple binding domains (e.g., [11,10,5]). 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:

$$A + B \xrightarrow{k_{b_1}} A_{10} \quad A_{10} \xrightarrow{k_{u_1}} A + B \qquad A + B \xrightarrow{k_{b_2}} A_{01} \quad A_{01} \xrightarrow{k_{u_2}} A + B$$
$$A_{01} + B \xrightarrow{k_{b_1}} A_{11} \quad A_{11} \xrightarrow{k_{u_1}} B + A_{01} \quad A_{10} + B \xrightarrow{k_{b_2}} A_{11} \quad A_{11} \xrightarrow{k_{u_2}} B + A_{10}$$

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., [33,30]). 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 [7,8], as well as by domain-specific techniques [11,10,5]. 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.

To answer this question, we consider variants of this binding model by increasing the number of sites of molecule A from 2 to 7, 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 values of k_{b_i} were spread uniformly around an arbitrarily fixed value of 10, as shown in Table 1. In this way, all binding rates are equal to 10 in the reference PIVP.

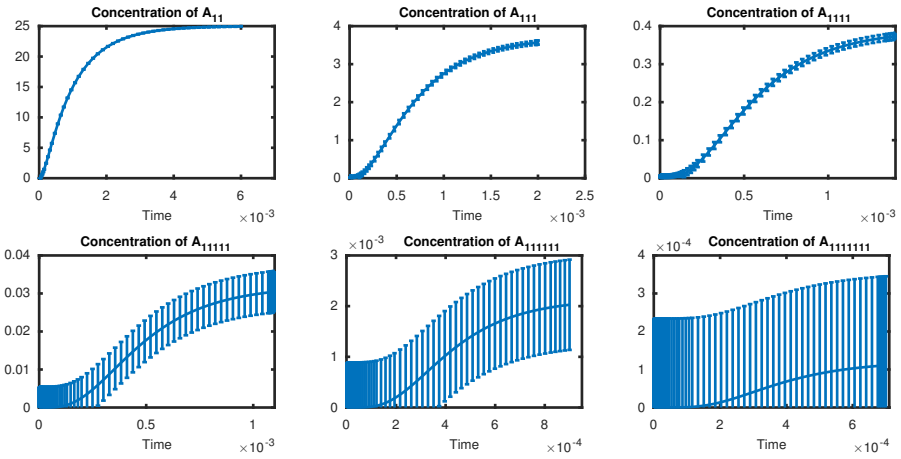


Fig. 2: Representative traces (solid line) in the reference models for the concentration of the complex A where every site is bound to a B molecule, together with the formal bound (error bar) on the respective traces in the original model.

In this case study, we looked for differences between the rate values which ensured that the original model was contained in the δ neighborhood around the reference. The results of the analysis, in Table 1, show that the original model size grows combinatorially with respect to the number of binding sites (columns $|R|$ and $|S|$ provide the number of reactions and species in each model, respectively). Since distinct binding rates are used, none of the models can be reduced by FDE or BDE. Instead, this combinatorial explosion is tackled applying ε-FDE or ε-BDE for an appropriate ε, here chosen large enough to reduce all models (we have set $\varepsilon = 0.18$ and $\varepsilon = 0.04$ for ε-FDE and ε-BDE, respectively). Similarly to the exact counterpart where all the binding rates are equal, all models feature the same ε-FDE and ε-BDE coarsest partition (column $|\mathcal{H}|$ gives the number of blocks). The fact that ε-BDE has much better runtime performance than ε-FDE backs the complexity result in Theorem 3.

For each model, the time horizon $\hat{\tau}$ was set large enough to exercise most of its transient dynamics, starting from an initial condition where the concentrations were set to 100 for the free molecules A and B, and to zero for the other species. With this set-up, the values of δ were computed using Lemma 1. We find that δ decreases as the number of asymmetric binding sites increases; correspondingly the values of λ increase, but remaining in the acceptable range 1.60-2.21. The last column reports the infinity norm of the difference between the rates of the original model and that of the reference, confirming that the δ neighborhood does contain the original model in all cases. To do so, however, the relative differences between the binding rates were decreased by about one order of magnitude at each added binding site.

Figure 2 shows representative plots of the reference trajectory with error bars that indicate the formal bound on approximation error of any trajectory starting within the computed δ neighborhood of the reference. For each model, we show the time course

of the concentration of complex A with every site occupied by a B molecule, i.e., A_{11}, A_{111}, and so on. With these parameterizations, this turns out to be the species with the lowest concentration levels, decreasing of roughly one order of magnitude with increasing number of binding sites. Thus relatively larger regions of uncertainty are provided for species that have low outputs in absolute terms. We remark that such situations may occur especially in large-dimensional systems where low outputs indicate low probability of finding a certain complex configuration in the system.

With the next example we study another pattern of quasi-symmetries in CRNs. We also use it to some numerical evidence of the robustness with which the reference model can still be taken as an empirical approximate even when the δ neighborhood of the formal bound does not contain the original model.

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. For exact aggregation purposes, 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 [37].

Here, we consider the polymerization scenario taken from [37, Chapter 7] illustrating the formation of polycyclic aromatic hybrocarbons 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

$$\begin{aligned}
\mathrm{A}_i + \mathrm{H} &\xrightarrow{\alpha_i} \mathrm{A}_i\tilde{} + \mathrm{H}_2 \quad (i,1) & \mathrm{A}_i\tilde{} + \mathrm{H}_2 &\xrightarrow{\overline{\alpha}_i} \mathrm{A}_i + \mathrm{H} \quad (i,\overline{1})\\
\mathrm{A}_i\tilde{} + \mathrm{C_2H_2} &\xrightarrow{\beta_i} \mathrm{A}_i\mathrm{CHCH}\tilde{} \quad (i,2) & \mathrm{A}_i\mathrm{CHCH}\tilde{} &\xrightarrow{\overline{\beta}_i} \mathrm{A}_i\tilde{} + \mathrm{C_2H_2} \quad (i,\overline{2})\\
\mathrm{A}_i\mathrm{CHCH}\tilde{} + \mathrm{C_2H_2} &\xrightarrow{\gamma_i} \mathrm{A}_{i+1} + \mathrm{H} \quad (i,3) & \dots & \quad (i+1,1)
\end{aligned}$$

Here $\mathrm{A}_i\tilde{}$ is an aromatic radical formed by H abstraction from A_i, and $\mathrm{A}_i\mathrm{CHCH}\tilde{}$ is a radical formed by adding $\mathrm{C_2H_2}$ 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 the following finite truncation of such a CRN:

$$\begin{aligned}
\mathrm{A_1} + \mathrm{H} &\xrightarrow{\alpha_1} \mathrm{A_1}\tilde{} + \mathrm{H_2} \quad (1,1) & \mathrm{A_2} + \mathrm{H} &\xrightarrow{\alpha_2} \mathrm{A_2}\tilde{} + \mathrm{H_2} \quad (2,1)\\
\mathrm{A_1}\tilde{} + \mathrm{H_2} &\xrightarrow{\overline{\alpha}_1} \mathrm{A_1} + \mathrm{H} \quad (1,\overline{1}) & \mathrm{A_2}\tilde{} + \mathrm{H_2} &\xrightarrow{\overline{\alpha}_2} \mathrm{A_2} + \mathrm{H} \quad (2,\overline{1})\\
\mathrm{A_1}\tilde{} + \mathrm{C_2H_2} &\xrightarrow{\beta_1} \mathrm{A_1CHCH}\tilde{} \quad (1,2) & \mathrm{A_2}\tilde{} + \mathrm{C_2H_2} &\xrightarrow{\beta_2} \mathrm{A_2CHCH}\tilde{} \quad (2,2)\\
\mathrm{A_1CHCH}\tilde{} &\xrightarrow{\overline{\beta}_1} \mathrm{A_1}\tilde{} + \mathrm{C_2H_2} \quad (1,\overline{2}) & \mathrm{A_2CHCH}\tilde{} &\xrightarrow{\overline{\beta}_2} \mathrm{A_2}\tilde{} + \mathrm{C_2H_2} \quad (2,\overline{2})\\
\mathrm{A_1CHCH}\tilde{} + \mathrm{C_2H_2} &\xrightarrow{\gamma_1} \mathrm{A_2} + \mathrm{H} \quad (1,3) & \mathrm{A_2CHCH}\tilde{} + \mathrm{C_2H_2} &\xrightarrow{\gamma_2} \mathrm{A_1} + \mathrm{H} \quad (2,3)
\end{aligned}$$

where 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 $\mathrm{A_3}$ to $\mathrm{A_1}$, see reaction $(2,3)$. Intuitively, this mimics the fact that polymers of length 3 become unstable due to their length.

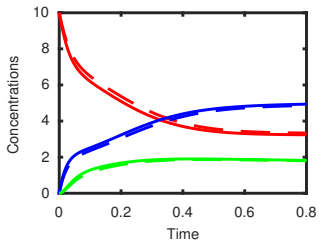
In [37] it is assumed that the dynamics of polymers do not depend on their length. In our truncated model, this corresponds to setting perfectly symmetric values for rates, i.e., $\alpha_1 = \alpha_2$, $\beta_1 = \beta_2$, $\gamma_1 = \gamma_2$, $\overline{\alpha}_1 = \overline{\alpha}_2$, $\overline{\beta}_1 = \overline{\beta}_2$. Then, it can be shown (similarly to [37]) that x_{A_1} and x_{A_2}, $x_{\mathrm{A}_1\text{-}}$ and $x_{\mathrm{A}_2\text{-}}$, and $x_{\mathrm{A}_1\mathrm{CHCH}\text{-}}$ and $x_{\mathrm{A}_2\mathrm{CHCH}\text{-}}$ are related by FDE.

We now consider a variant with approximately equal rates by setting:

$$\begin{aligned} \alpha_1 &= 1.0 - \eta/2, & \beta_1 &= 2.0 - \eta/2, & \gamma_1 &= 3.0 - \eta/2, \\ \alpha_2 &= 1.0 + \eta/2, & \beta_2 &= 2.0 + \eta/2, & \gamma_2 &= 3.0 + \eta/2, \end{aligned}$$

while we kept the reversed rates equal, i.e., $\overline{\alpha}_1 = \overline{\alpha}_2 = 0.1$ and $\overline{\beta}_1 = \overline{\beta}_2 = 0.2$.

With this parameterization, η corresponds to the norm between the original model and the reference model obtained from a 0.01-FDE. As expected, here the parameters α_i, β_i and γ_i (for $i \in \{1, 2\}$) are perturbed to new values 1.0, 2.0 and 3.0, respectively. To study the approximation error, we set the initial concentrations of H, H_2, $\mathrm{C}_2\mathrm{H}_2$ and A_1 to 10, with 0.8 time horizon, which ensures an ODE steady state of the reference model. Then, Lemma 1 yields $\delta = 3.39\text{E}{-}04$, with $\lambda = 1.1$. Thus, values of η up to δ can be explained by the formal bound. Instead, the figure in the right inset compares the ODE traces of the non-singleton 0.01-FDE equivalence classes, i.e., $x_{A_1} + x_{A_2}$ (red lines), $x_{\mathrm{A}_1\text{-}} + x_{\mathrm{A}_2\text{-}}$ (blue lines) and $x_{\mathrm{A}_1\mathrm{CHCH}\text{-}} + x_{\mathrm{A}_2\mathrm{CHCH}\text{-}}$ (green lines), of the reference PIVP (dashed lines) and the original model (solid lines) for $\eta = 0.039$, i.e., two orders of magnitude larger than δ. Although the approximation is not formally justified by the bound, it can effectively recover the dynamics of the original model.



6 Conclusion

Reasoning about quantitative properties approximately can represent an effective way of taming the complexity of real systems. In this paper we have considered ordinary differential equations (ODEs) with polynomial derivatives. We developed notions of equivalence as a relaxation of their exact counterparts, by allowing the derivatives of related ODE variables to vary up to a desired tolerance. We believe that our algorithmic approach can be a useful tool to systematically discover near-symmetries in many situations such as those presented in our case studies.

Within this framework it is possible to pursue further routes for improvement. One would be to consider, at the price of additional computations, several reference trajectories, as proposed in [16,15]. This would allow to extend our formal estimation, in principle, to arbitrarily larger neighborhoods of the reference model, thus helping cope with the case when the neighborhood is smaller than required. It would be also possible to integrate other bounding techniques, such as [36,2] which lacked an automatic synthesis of a reference model but can offer a tradeoff between precision/tightness of the bound and computation cost in its derivation. Finally we plan to extend the numerical investigation by implementing these techniques in a robust software tool.

References

1. Asarin, E., Dang, T., Girard, A.: Reachability Analysis of Nonlinear Systems Using Conservative Approximation. In: HSCC (2003)
2. Bortolussi, L., Gast, N.: Mean field approximation of uncertain stochastic models. In: DSN (2016)
3. van Breugel, F., Worrell, J.: Towards quantitative verification of probabilistic transition systems. In: ICALP (2001)
4. Calzone, L., Fages, F., Soliman, S.: BIOCHAM: an environment for modeling biological systems and formalizing experimental knowledge. Bioinformatics 22(14), 1805–1807 (2006)
5. Camporesi, F., Feret, J.: Formal reduction for rule-based models. Electronic Notes in Theoretical Computer Science 276, 29–59 (2011)
6. Cardelli, L.: On process rate semantics. Theoretical Computer Science 391(3), 190–215 (2008)
7. Cardelli, L., Tribastone, M., Tschaikowski, M., Vandin, A.: Forward and backward bisimulations for chemical reaction networks. In: CONCUR. pp. 226–239 (2015)
8. Cardelli, L., Tribastone, M., Tschaikowski, M., Vandin, A.: Symbolic computation of differential equivalences. In: POPL (2016)
9. Ciocchetta, F., Hillston, J.: Bio-PEPA: A framework for the modelling and analysis of biological systems. Theoretical Computer Science 410(33-34), 3065–3084 (2009)
10. Conzelmann, H., Fey, D., Gilles, E.: Exact model reduction of combinatorial reaction networks. BMC Systems Biology 2(1), 78 (2008)
11. Conzelmann, H., Saez-Rodriguez, J., Sauter, T., Kholodenko, B., Gilles, E.: A domain-oriented approach to the reduction of combinatorial complexity in signal transduction networks. BMC Bioinformatics 7(1), 34 (2006)
12. Danos, V., Laneve, C.: Formal molecular biology. Theoretical Computer Science 325(1), 69–110 (2004)
13. Desharnais, J., Gupta, V., Jagadeesan, R., Panangaden, P.: Metrics for labeled Markov systems. In: CONCUR (1999)
14. Desharnais, J., Jagadeesan, R., Gupta, V., Panangaden, P.: Approximating labeled Markov processes. In: LICS (2000)
15. Donzé, A., Krogh, B.H., Rajhans, A.: Parameter Synthesis for Hybrid Systems with an Application to Simulink Models. In: HSCC. pp. 165–179. Springer (2009)
16. Donzé, A., Maler, O.: Systematic Simulation Using Sensitivity Analysis. In: HSCC. pp. 174–189. Springer (2007)
17. E, W., Li, T., Vanden-Eijnden, E.: Optimal partition and effective dynamics of complex networks. PNAS 105(23), 7907–12 (Jun 2008)
18. Farrell, J.A., Polycarpou, M.M.: Adaptive Approximation Based Control. Wiley-Interscience (2006)
19. Gillespie, D.T.: The chemical Langevin equation. The Journal of Chemical Physics 113(1), 297–306 (2000)
20. Girard, A., Pappas, G.: Approximate bisimulations for nonlinear dynamical systems. In: IEEE Conference on Decision and Control and European Control Conference (2005)
21. Hayden, R.A., Bradley, J.T.: A fluid analysis framework for a Markovian process algebra. Theoretical Computer Science 411(22-24), 2260–2297 (2010)
22. Iacobelli, G., Tribastone, M.: Lumpability of fluid models with heterogeneous agent types. In: DSN (2013)
23. Iwasa, Y., Andreasen, V., Levin, S.: Aggregation in model ecosystems. I. Perfect aggregation. Ecological Modelling 37(3-4), 287–302 (1987)

24. Iwasa, Y., Levin, S.A., Andreasen, V.: Aggregation in Model Ecosystems II. Approximate Aggregation. Mathematical Medicine and Biology 6(1), 1–23 (1989)
25. Kozlov, M., Tarasov, S., Khachiyan, L.: The polynomial solvability of convex quadratic programming. USSR Computational Mathematics and Mathematical Physics 20(5), 223 – 228 (1980)
26. Kuo, J.C.W., Wei, J.: Lumping analysis in monomolecular reaction systems. Analysis of approximately lumpable system. Industrial & Engineering Chemistry Fundamentals 8(1), 124–133 (1969)
27. Larsen, K.G., Skou, A.: Bisimulation through probabilistic testing. Inf. Comput. 94(1), 1–28 (1991)
28. Li, G., Rabitz, H.: A general analysis of approximate lumping in chemical kinetics. Chemical Engineering Science 45(4), 977–1002 (1990)
29. Milios, D., Gilmore, S.: Component aggregation for PEPA models: An approach based on approximate strong equivalence. Performance Evaluation 94, 43–71 (2015)
30. Monine, M.I., Posner, R.G., Savage, P.B., Faeder, J.R., Hlavacek, W.S.: Modeling multivalent ligand-receptor interactions with steric constraints on configurations of cell-surface receptor aggregates. Biophysical Journal 98(1), 48–56 (2010)
31. Pardalos, P.M., Vavasis, S.A.: Quadratic programming with one negative eigenvalue is NP-hard. Journal of Global Optimization 1(1), 15–22 (1991)
32. Pedersen, M., Plotkin, G.D.: A language for biochemical systems: Design and formal specification. In: Transactions on Computational Systems Biology XII, LNCS, vol. 5945, pp. 77–145. Springer (2010)
33. Sneddon, M.W., Faeder, J.R., Emonet, T.: Efficient modeling, simulation and coarse-graining of biological complexity with NFsim. Nature Methods 8(2), 177–183 (2011)
34. Tribastone, M., Gilmore, S., Hillston, J.: Scalable differential analysis of process algebra models. IEEE Trans. Software Eng. 38(1), 205–219 (2012)
35. Tschaikowski, M., Tribastone, M.: A unified framework for differential aggregations in markovian process algebra. Journal of Logical and Algebraic Methods in Programming 84(2), 238–258 (2015)
36. Tschaikowski, M., Tribastone, M.: Approximate reduction of heterogeneous nonlinear models with differential hulls. IEEE Transactions on Automatic Control (2016)
37. Turányi, T., Tomlin, A.S.: Reduction of Reaction Mechanisms, pp. 183–312. Springer Berlin Heidelberg, Berlin, Heidelberg (2014)
38. Vidyasagar, M.: Nonlinear Systems Analysis. Prentice-Hall, Inc., Upper Saddle River, NJ, USA (1992)
39. Zamani, M., Majumdar, R.: A Lyapunov approach in incremental stability. In: IEEE Conference on Decision and Control and European Control Conference (2011)

A Proofs

Proof (Theorem 1). Assume that $\mathcal{H}_1, \ldots, \mathcal{H}_n$ are ε-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 ε-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

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

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 ε-FDE partition.

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

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

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

So far, we have shown that the coarsening $\mathcal{S}/\big(\bigcup_{l=1}^n\sim_l\big)^*$ of ε-FDE/BDE partitions $\mathcal{H}_1,\dots,\mathcal{H}_n$ is again an ε-FDE/BDE partition. The claim follows by noting that Lemma 26 in [7] 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 3. *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

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

Since $\mathcal{H}$ is a refinement of $\mathcal{G}$, for any $G \in \mathcal{G}$ there exist unique blocks $H_1^G, \dots, H_{m_G}^G \in \mathcal{H}$ such that $\biguplus_{k=1}^{m_G} H_k^G = G$. With this, it holds that

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

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

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

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

Proof (Theorem 2). Let $\mathcal{G}'$ denote the coarsest ε-FDE (ε-BDE) partition that refines $\mathcal{H}_0 := \mathcal{G}$ and set $\mathcal{H}_{k+1} := \mathcal{S}/(\sim_{\mathcal{H}_k,\varepsilon}^{\chi^*} \cap \sim_{\mathcal{H}_k})$ 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 3 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 3 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 ε-FDE (ε-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 $\mathcal{H}_{k-1} = \mathcal{H}_k = \mathcal{S}/(\sim_{\mathcal{H}_{k-1},\varepsilon}^{\chi^*} \cap \sim_{\mathcal{H}_{k-1}})$.

Proof (Theorem 3). Follows from the proof of Theorem 2.

Proof (Lemma 1). Straightforward application of Muller's version of the quadratic formula.

Proof (Lemma 2). 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

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

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.

A.1 Proofs of Theorem 7 and Theorem 8

We begin by stating the following generalization of Theorem 6.

Theorem 9. *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. Any decent book about ODEs will do.

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 9 with input function $u(s) = r(s, x^{\hat{\sigma}_1}(s) - x^{\hat{\sigma}_*}(s))$. Let Δx 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 9 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 10. *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 Δx the linearization of the true difference function $x^{\hat{\sigma}_1} - x^{\hat{\sigma}_*}$. With this, we are in a position to show Theorem 7 and 8.

Proof (Theorem 8). As pointed out above, in the case of $\deg(\hat{\mathcal{P}}) = 1$, it holds that $r \equiv 0$. This and Theorem 9 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 $0 \leq t \leq \hat{\tau}$, thus yielding the first claim. The second claim, instead, follows by noting that

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

Proof (Theorem 7). 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 8 and 10 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 $\delta_+ = 2\lambda_0\delta$.