Maximal aggregation of polynomial dynamical systems

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

Published in:
Proceedings of the National Academy of Sciences of the United States of America

Link to article, DOI:
10.1073/pnas.1702697114

Publication date:
2017

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

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Maximal aggregation of polynomial dynamical systems

Luca Cardelli\textsuperscript{1,2,3}, Mirco Tribastone\textsuperscript{4,1,2}, Max Tschaikowski\textsuperscript{5,1,}, and Andrea Vandin\textsuperscript{6,1}

\textsuperscript{1}Microsoft Research, Cambridge CB1 2FB, United Kingdom; \textsuperscript{2}Department of Computing, University of Oxford, Oxford OX1 3QD, United Kingdom; and \textsuperscript{3}Scuola IMT Alti Studi Lucca, 55100 Lucca, Italy

Edited by Moshe Y. Vardi, Rice University, Houston, TX, and approved July 28, 2017 (received for review February 16, 2017)

Ordinary differential equations (ODEs) with polynomial derivatives are a fundamental tool for understanding the dynamics of systems across many branches of science, but our ability to gain mechanistic insight and effectively conduct numerical evaluations is critically hindered when dealing with large models. Here we propose an aggregation technique that rests on two notions of equivalence relating ODE variables whenever they have the same solution (backward criterion) or if a self-consistent system can be written for describing the evolution of sums of variables in the same equivalence class (forward criterion). A key feature of our proposal is to encode a polynomial ODE system into a finitary structure akin to a formal chemical reaction network. This enables the development of discrete constraints that efficiently compute the largest equivalence, building on approaches rooted in computer science to minimize basic models of computation through iterative partition refinements. The physical interpretability of the aggregation is shown on polynomial ODE systems for biochemical reaction networks, gene regulatory networks, and evolutionary game theory.

Significance

Large-scale dynamical models hinder our capability of effectively analyzing them and interpreting their behavior. We present an algorithm for the simplification of polynomial ordinary differential equations by aggregating their variables. The reduction can preserve observables of interest and yields a physically intelligible reduced model, since each aggregate corresponds to the exact sum of original variables.

Author contributions: L.C., M. Tribastone, M. Tschaikowski, and A.V. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

1L.C., M. Tribastone, M. Tschaikowski, and A.V. contributed equally to this work.

2To whom correspondence should be addressed. Email: mirco.tribastone@imtlucca.it.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1702697114/-/DCSupplemental.
Model Definitions

We consider first-order ODEs in the form
\[ \frac{dx(t)}{dt} = P(x(t)), \quad x = (x_1, \ldots, x_n), \]
where \( P \) is a vector of multivariate polynomials over variables \( x_1, \ldots, x_n \). Let \( x(0) \) denote the initial condition.

Equivalences. Given a partition \( H = \{ H_1, \ldots, H_m \} \) over the variables of Eq. 1, we construct an aggregation matrix \( A \in \mathbb{R}^{m \times n} \), \( A = (a_{ij}) \), by setting \( a_{ij} = 1 \) if \( x_i \in H_j \) and \( a_{ij} = 0 \) otherwise. We say that \( A \) induces a forward equivalence if it is possible to explicitly describe the dynamics for each partition block. Following ref. 20, this amounts to requiring that
\[ AP(x) = AP(AAx), \quad \text{for all } x, \]
where \( A \) is any generalized right inverse of \( A \).

Instead, backward equivalence captures that the solution \( x(t) \) is “uniform” on a partition of variables \( H \). That is, for every \( H \in \mathcal{H} \) and \( x_i, x_j \in H \), it holds that \( x_i(t) = x_j(t) \) for all \( t \). This can be characterized by requiring that
\[ \mathcal{P}(A^T H_1) \subseteq A^T H_1, \quad \text{with} \quad A^T H_1 = \{ x \in \mathbb{R}^n | x \text{ uniform on } H \}. \]

To obtain a reduced model, we set \( \mathcal{P} = AP \), with the generalized inverse \( A = (\tilde{a}_{ij}) \), such that \( \tilde{a}_{ij} = \mu_i / \sum_k a_{jk} \). Let us consider the reduced ODE system:
\[ \frac{dy(t)}{dt} = \mathcal{P}(y(t)), \quad y = (y_1, \ldots, y_m). \]

If \( A \) is an aggregation matrix representing a forward equivalence, then the solution for the initial condition \( y(0) = Ax(0) \) satisfies \( y(t) = Ax(t) \) for all \( t \). Thus, \( y(t) \) preserves sums of variables, but the individual trajectories cannot be recovered.

Reaction Networks. According to conditions in either Eq. 2 or 3, checking whether a candidate partition is an equivalence involves reasoning over uncountable state spaces. Here, we develop appropriate finitary characterizations of forward and backward equivalence by encoding an ODE into an RN. Formally, this is a pair \((S, R)\) consisting of a set of species \( S \) and a set of reactions \( R \). We denote by \( MS(S) \) the set of all multisets with elements in \( S \). Each reaction is in the form \( \mu \xrightarrow{\alpha} \mu' \), where \( \mu \) and \( \mu' \) are multisets of species (called reagents and products, respectively), and the coefficient \( \alpha \) is a real number. A formal mass action CRN is, therefore, a special case where \( \alpha > 0 \) is the kinetic constant.

We encode each variable \( x \) with species \( S \) and each monomial \( x^\rho \) appearing in the ODE of \( x \) with the reaction
\[ \sum_{i=1}^n p_i S_i \Rightarrow S_k + \sum_{i=1}^n p_i S_i, \]
where the operator + denotes multisets union and \( p_i S_i \) is a multiset with \( p_i \) occurrences of \( S_i \).

For the RN equivalence conditions, we define the notion of net instantaneous stoichiometry of a species \( S \in S \) and of a set of species \( G \subseteq S \) because of reagents \( \rho \):
\[ \phi(\rho, S) := \sum_{\pi \in \rho} (\pi - \rho_i) \cdot \phi(\rho), \quad \phi(\rho, G) := \sum_{S_i \in G} \phi(\rho, S_i), \]
where \( \rho_i \) and \( \pi_i \) denote the multiplicity of species \( S_i \) in the reagents and products, respectively.

To characterize forward equivalence, we further define
\[ \text{fr}(S, \rho, G) := \frac{\phi(S + \rho, G)}{|S| + |\rho|}, \]
where the operator \(|.|\) denotes the multinomial coefficient induced by a multiset of species:
\[ |\rho| := \left( \sum_{i=1}^n \rho_i \right). \]

Our main result (SI Appendix, SI Text) is that a partition of species \( H \) is a forward equivalence on the respective ODE variables if and only if, for any two blocks \( H, H' \in \mathcal{H} \) and any two \( S_i, S_j \in H \), it holds that
\[ \text{fr}(S_i, \rho, H') = \text{fr}(S_j, \rho, H') \]
for all \( \rho \), such that \( S_i + \rho \) or \( S_j + \rho \) is a reagent in at least one reaction of \( R \) (thus including \( \rho = \emptyset \) for reactions with one species

---

Fig. 1. Example of reduction with forward equivalence. (A) A CRN of a basic mechanism of reversible binding between molecular species, \( A \) and \( B \), through \( A \)'s two identical binding sites. The state of the site is denoted by the subscripts \( u \) (unphosphorylated) and \( p \) (phosphorylated). For simplicity, we assume that binding occurs when a site is phosphorylated, at most one molecule of \( B \) can bind to \( A \), and at most one binding site can be phosphorylated. We set \( k_i = i \) for \( i = 1, \ldots, 4 \). (B) We compute the largest forward equivalence refining the singleton initial partition of species; \( sp \) refers to a block in the set of splitters initialized with the initial partition. At each iteration, the algorithm computes all values \( \text{fr}(., sp) \) (of which those equal to zero are not shown). Each block is refined, such that any two species \( S \) and \( S' \) in the same subblock have the same values of \( \text{fr}(S, \rho, sp) \) and \( \text{fr}(S', \rho, sp) \) for every partner \( \rho \). The first iteration produces the subblocks \( \{ A_{up} \}, \{ A_{wp}, A_{up} \}, \{ B \}, \{ A_{up}, A_{wp}, B \} \), which form the new partition at the next iteration. For each refined block, one among the subblocks of maximal size (here, \{A_{up}, A_{wp}, B\}) is not added to the set of splitters. (C) Each splitter is removed and considered in turn; however, no blocks can be refined further. At the end of the fourth iteration, the set of splitters is empty. The resulting largest forward equivalence aggregates \( A \) molecules that have the same phosphorylation level, abstracting from the identity of the sites and automatically revealing the assumption on their identical dynamics that was made. Computations for backward equivalence proceed similarly using Eq. 7 and the parameter \( H ' \) of \( br \) as splitter.
only as reagent. Intuitively, two species are related whenever they provide the same net instantaneous stoichiometry, for any reaction partner \( \rho \), to all equivalence classes of species.

A similar result holds for backward equivalence. Here, two species are related whenever they have the same instantaneous stoichiometry aggregated across all reagents that are equal up to the considered equivalence. To formally express this, let \( \mathcal{H} \) be a partition of species and \( \approx_{\mathcal{H}} \) be the equivalence relation naturally induced by \( \mathcal{H} \) over multisets of species as follows:

\[
\approx_{\mathcal{H}} = \left\{ (\rho, \pi) \in \mathcal{M} \mathcal{S}(S) \times \mathcal{M} \mathcal{S}(S) \mid \sum_{S_i \in \mathcal{H}} \rho_i = \sum_{S_i \in \mathcal{H}} \pi_i, \forall H \in \mathcal{H} \right\}.
\]

Also, we define

\[
\mathsf{br}(S_i, M, H') := \sum_{S_i \in H'} \sum_{\rho \in \mathcal{M} \mathcal{S}(S_i)} \phi(S_k + \rho, S_i) / \phi(S_k, S_i + \rho),
\]

where \( |\pi|_{H'} = |\{ S_i \in H' | \pi_i > 0 \} | \) counts the number of different species in \( H' \) occurring in \( \pi \). Then, \( \mathsf{br} \) is a backward equivalence for the corresponding ODE variables if and only if, for any block \( H \in \mathcal{H} \) and any two \( S_i, S_j \in H \), it holds that

\[
\phi(\emptyset, S_i) = \phi(\emptyset, S_j) \quad \text{and} \quad \mathsf{br}(S_i, M, H') = \mathsf{br}(S_j, M, H') \quad [7]
\]

for all \( H' \in \mathcal{H} \) and \( M \in \{ \rho(S_k + \rho \xrightarrow{\pi} \pi) \in R, S_k \in S \} / \approx_{\mathcal{H}} \) (SI Appendix, SI Text).

The first condition of Eq. 7 regards reactions that encode constants (i.e., degree zero monomials). These are ignored in Eq. 6 because they do not affect forward equivalence.

**Reduction Algorithm.** The largest equivalence that refines an initial partition of species is computed via iterative refinements. Briefly, a set of “splitters” is initialized with the blocks of the initial partition. Each splitter is considered as a candidate block that prevents the current partition from being an equivalence. In both Eqs. 6 and 7, the parameter \( H' \) represents the splitter. If the equivalence criteria, checked with respect to the splitter, do not hold, the partition is refined, such that any two species in the same subblock will now satisfy them. The resulting subblocks are added to the set of splitters, except for the largest one, following an argument similar to ref. 18. There exists a unique fixed point corresponding to the case where no more splitters have to be considered, yielding the desired largest equivalence (Fig. 1).

The algorithm runs in polynomial time and space with respect to the number of variables and monomials in the derivatives (SI Appendix, SI Text).

For a given equivalence, a reduced RN can be obtained by transforming the original one in three steps that preserve the structure of the reactions (Fig. 2). For both forward and backward equivalence, a species in the reduced RN represents the sum of species belonging to that equivalence class; in the case of a backward equivalence, the individual trajectory of an original species can then be recovered by simply dividing the ODE solution for each representative by the cardinality of its equivalence class. From the reduced RN, we can compute the reduced ODE system of Eq. 4 by reversing the encoding of Eq. 5 (SI Appendix, SI Text). This corresponds to interpreting the reduced RN with mass action kinetics, straightforwardly generalized to nonpositive reaction systems.

**Applications**

**Molecular Biology.** Multisite protein phosphorylation is a widely studied signal transduction mechanism responsible for many regulatory roles in eukaryotic cells, such as threshold setting and switch-like behavior (21–23). The RN in Fig. 1 is a simple model of unordered phosphorylation, where the sites are assumed to be equivalent. In this case, it is common to consider the same kinetic rates when describing their interactions (21, 24, and 25) as a mathematical simplification backed by experimental evidence (26). In general, the full dynamics of a protein with \( n \) phosphorylation sites would require \( 2^n \) variables to keep track of the state of each individual site. Both forward and backward equivalence explain the assumption of identical sites, yielding \( n + 1 \) equivalence classes that group variables related to proteins with the same number of phosphorylated sites. This confirms an earlier lumping scheme developed specifically for this scenario (27). A similar aggregation can be observed in the modeling of mechanisms of complex formation in the case where a receptor protein has multiple binding sites (SI Appendix, SI Text and Table S1).

Forward equivalence may also aggregate complexes exhibiting different phosphorylation levels. Kozer et al. (28) propose a model of oligomerization of the EGF receptor (EGFR) kinase. It accounts for ligand binding, conformational changes of the EGFR cytosolic tail induced by the presence of the ligand and formation of dimers, trimers, and tetramers as well as EGFR phosphorylation/dephosphorylation occurring at a single site. The original network consists of 923 species and 11,918 reactions. The maximal forward equivalence aggregates oligomers that are equal up to the phosphorylation state of their sites (Fig. 3). This leads to a reduced network with only 87 species and 705 reactions, still useful to answer biologically relevant questions, such as those in ref. 28 concerning the distribution of the cluster sizes.

An inspection of the members of the equivalence classes suggests that the dynamics of phosphorylation/dephosphorylation and oligomer formation are independent. Effectively, the
equivocation classes internalize the phosphorylation dynamics in the following sense. On any phosphorylation event, the complex undergoes a change of state, turning into another complex within the same equivocation class. Different members of the same class, however, may have functionally distinct behavior. This also prevents an aggregation by backward equivalence. For instance, phosphorylation may not occur in a single EGFR (Fig. 3C, C-I) because it depends on the context, requiring two receptors to be bound with a conformationally changed tail (29). Instead, a phosphorylated EGFR (Fig. 3C, C-II) may always undergo dephosphorylation, since this is modeled as a spontaneous reaction. Situations such as these, which feature site interactions that are controlled or dependent on other sites, may block the use of domain-specific reduction techniques (5–7, 30), since they exploit assumptions of independence within interaction domains (SI Appendix, SI Text). We refer to ref. 31 for a recent discussion on the complementarity between equivalence-based CRN aggregations and rule-based reduction techniques (7).

A similar aggregation pattern arises in a model of early events of the signaling pathway of the high-affinity receptor for IgE (FcεRI) in mast cells and basophils. The pathway includes phosphorylation of the tyrosine residues on both the β and γ subunits of FcεRI by the Lyn kinase, which then recruits the protein tyrosine kinase Syk (34). An experimentally validated model has been proposed to provide mechanistic insights into these processes (32). Here, a bivalent IgE ligand aggregates FcεRI; Syk presents two phosphorylation units, the linker region and the activation loop, transphosphorylated by Lyn and Syk, respectively. The overall pathway is described by 354 molecular species and 3,680 reactions. The maximal forward equivalence shows that complexes that have the same formation up to the phosphorylation state of both units can be aggregated (Fig. 4), yielding a reduced network with 105 species and 775 reactions. This finding extends and provides a formal proof for the observation made in ref. 33, supplementary note 8, where an exactly reduced network (with 172 species and 1,433 reactions) was obtained by abstracting from the phosphorylation site of the linker region only. Indeed, that network corresponds to the refinement of the maximal forward equivalence which separates complexes according to the phosphorylation status of Syk (SI Appendix, SI Text).

We now discuss an example where members of the same equivalence class do not have the same structure, using a detailed model of activation of JAK, a family of enzymes that mediate gene transcription (35). The mechanism is explained by the formation of a macrocomplex by JAK binding to growth hormone (GH) receptor dimers. The maximal forward equivalence aggregates the dynamics of ligand/receptor complexes undergoing constitutive turnover or endocytosis (Fig. 5). This gives non-trivial equivalence classes containing complexes that differ in the structure of the GH ligand/receptor. The full network (35) (SI Appendix, SI Text), consisting of 471 species and 5,033 reactions, is reduced to 345 species and 4,068 reactions. We find that every complex in the same equivalence class features the same number of phosphorylated sites of JAK (Y1 and Y2). The maximal aggregation still allows all of the analyses of ref. 35, which concern the concentrations of certain complexes and the phosphorylation level of Y2.

We note that all previous models obey the law of mass action and underlie ODE systems with polynomial derivatives of degree two at most. Our technique is also applicable to ODEs with other nonlinearities, such as biochemical networks with Michaelis–Menten kinetics. Following, for instance, ref. 2, this can be done by constructing an equivalent polynomial ODE system with auxiliary variables for rational expressions, sigmoids, and trigonometric functions (SI Appendix, SI Text).

Logic Models of Regulatory Networks. Logic models are another established method to describe regulatory networks as a means of expressing qualitative interactions between biomolecular processes (39). Each process is associated with a Boolean variable that describes two discrete states (e.g., on–off). A Boolean update function defines how each variable may change state depending on the values of the other variables (e.g., to represent promotion and inhibition). Boolean models may be too coarse when a more detailed evolution is required: for instance, to compare predictions against experimental data or when they are to be coupled with a quantitative models. For this, Boolean models can be translated into ODEs with derivatives that agree with the Boolean update function whenever inputs are only either zero (false) or one (true) (38, 40, 41).
Here we consider the multivariate polynomial interpolation of ref. 38. On a model of T-cell receptor signaling studied in refs. 37 and 38, backward equivalence reduces processes that exhibit the same behavior because they are updated by functions that are equal up to variables in the same equivalence class (Fig. 6, and SI Appendix, Figs. S1 and S2 for further examples). From the reduced model, we can exactly recover the original solution in terms of continuous signals in the [0, 1] interval. Instead, forward equivalence leads to variables living in larger domains. On this example, the maximal forward equivalence reduces the ODE model, such that it is still possible to analyze full activation/deactivation of the downstream transcription factors CRE, AP1, NFAT, and NFKB, which belong to the same equivalence class (SI Appendix, Figs. S3 and S4).

**Evolutionary Game Theory.** The replicator equation is a well-studied model for several natural, social, and economic systems (42). It describes the dynamics of populations of individuals that choose strategies with a rate of growth that depends on the comparison between an individual’s own payoff and the population’s average. In its first formulation (43), the replicator equation considers a state represented by the vector $x = (x_1, \ldots, x_n)$, where $x_i$ denotes the probability of an individual choosing the $i$th strategy, with $1 \leq i \leq n$. Its evolution is governed by the polynomial ODE system

$$\dot{x}_i = x_i ((Bx)_i - x^T B x), \quad 1 \leq i \leq n,$$

where $B$ is the $n \times n$ payoff matrix, together with an initial condition, such that it represents an initial proportion of strategies [i.e., $\sum_{i=1}^{n} x_i(0) = 1$]. Here, backward equivalence may detect strategies chosen with the same frequency within the population. For instance, given the following payoff matrix

$$B = \begin{bmatrix} 1 & 3 & 2 \\ 3 & 1 & 2 \\ 4 & 4 & 1 \end{bmatrix}.$$

backward equivalence relates $x_1$ and $x_2$.

Similar investigations can be made on variants of the replicator equations that model evolutionary dynamics over networks (44). Here, a vertex represents a player that can interact with its neighbors only. In this context, the problem of network aggregation has also been studied using graph lumpability, a criterion that involves conditions on the (weighted) adjacency matrix of the network as well as on the players’ payoff matrices (45). It is related to backward equivalence in that it captures an equivalence relation between players/vertices, such that any two equivalent players choose any strategy with the same frequency at all time points. Graph lumpability turns out to be a sufficient condition for backward aggregation. For instance, let us consider a network with four players playing two strategies characterized by adjacency matrix $A = (a_{ij})_{1 \leq i, j \leq 4}$ and payoff matrices $B_i$, $1 \leq i \leq 4$ given by

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix}, \quad B_{1,2} = \begin{bmatrix} 1 & 3 \\ 4 & 1 \\ 7 & 1 \\ 2 & 2 \end{bmatrix}, \quad B_{3,4} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{bmatrix}.$$

Then, players 1 and 2 as well as players 3 and 4 have the same ODE solutions, but this is not captured by an equivalence relation in the sense of graph lumpability, since it requires $\sum_{i \in P} a_{ik} = \sum_{i \in P} a_{il}$ for any two equivalent players $i, j$ and for any equivalent class $P$ of players. Clearly, this condition is not satisfied by taking $i = 1, j = 2$, and $P = \{3, 4\}$.

**Conclusion**

We presented a technique to reduce polynomial ODE systems up to an equivalence relation over its variables. Our method exactly preserves observables of interest across the whole time course. Hence, the reduced model can be used as an input to complementary techniques that sacrifice exactness, such as timescale decomposition (46).

In the notable case where the model is a formal chemical RN, the reduction preserves structure, in that the original reactions are only subjected to renaming and merging. For other domain-specific applications, such as rule-based systems, Boolean networks, payoff matrices, and so on, one would seek to directly obtain reduced models of the corresponding nature induced by a backward/forward equivalence. Technically, this does not seem to be straightforward. For example, in the case of Boolean networks, forward equivalence yields a reduced ODE system where each aggregated variable will take values in the continuous interval $[0, n]$, where $n$ is the cardinality of the corresponding equivalence class. Thus, in general, there is no Boolean network, such that its polynomial ODE interpolation corresponds to an aggregated ODE system up to forward equivalence because by construction, each interpolated ODE variable takes values in the interval $[0, 1]$. In this paper, we have privileged a domain-agnostic view. We aim to address domain-specific challenges in future work.

**ACKNOWLEDGMENTS.** L.C. is partially funded by a Royal Society Research Professorship.