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):
Maximal aggregation of polynomial dynamical systems

Luca Cardelli1,a,b,1, Mirco Tribastone1,c,2, Max Tschaikowski1,c,1, and Andrea Vandin1,c,1

1Microsoft Research, Cambridge CB1 2FB, United Kingdom; 2Department of Computing, University of Oxford, Oxford OX1 3QD, United Kingdom; and 3Scuola IMT Alti Studi Lučka, 55100 Lučka, 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 a discrete algorithm to 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@mimicluca.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
\[
dx(t) \over dt = P(x(t)), \quad x = (x_1, \ldots, x_n),
\]
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 \( \mathcal{H} = \{ H_1, \ldots, H_m \} \) 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
\[
A P(x) = A P(A A x), \quad \text{for all } x,
\]
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 \( \mathcal{H} \). 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
\[
\mathcal{P} (\mathcal{H}_j) \subseteq \mathcal{U}_H, \quad \text{with } \mathcal{H}_j = \{ x \in \mathbb{R}^n | x \text{ uniform on } H_j \}.
\]

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

If \( \tilde{A} \) is an aggregation matrix representing a forward equivalence, then the solution for the initial condition \( y(0) = A x(0) \) satisfies \( y(t) = A x(t) \) for all \( t \). Thus, \( y(t) \) preserves sums of variables, but in general, the individual trajectories cannot be recovered. Instead, if \( \tilde{A} \) represents a backward equivalence, then the original solution is obtained by dividing each trace \( y_i(t) \) by the size of the equivalence class \( |H_i| \), provided that the initial conditions \( x(0) \) are uniform on \( \mathcal{H} \).

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 \( \mathcal{M}(\mathcal{S}) \) 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 \( S_1 \) and each monomial \( \alpha ! \frac{\partial^n x}{\partial t^n} \) appearing in the ODE of \( x \), with the reaction
\[
\sum_{i=1}^n p_i S_i \xrightarrow{\alpha} S_k + \sum_{i=1}^n p_i S_i,
\]
where the operator \( + \) denotes multiset union and \( p_i \) 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 \( S_i \in S \) and of a set of species \( G \subseteq S \) because of reagents \( \rho \),
\[
\phi(\rho, S_i) := \sum_{(\rho \xrightarrow{\alpha} x) \in R} (\tau_i - \rho_i) \alpha \quad \text{and} \quad \phi(\rho, G) := \sum_{S_i \in G} \phi(\rho, S_i),
\]
where \( \rho_i \) and \( \tau_i \) denote the multiplicity of species \( S_i \) in the reagents and products, respectively.

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

Our main result (SI Appendix, SI Text) is that a partition of species \( \mathcal{H} \) 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 \( \rho = 0 \) 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) The current partition. 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 \( fr(S_i, \rho, sp) \) (of which those equal to zero are not shown). Each block is refined, such that any two species \( S_i \) and \( S_j \) in the same subblock have the same values of \( fr(S_i, \rho, sp) \) and \( fr(S_j, \rho, sp) \) for every partner \( \rho \). The first iteration produces the subblocks \( \{A_{pu}, A_{up}, p_u, B, A_{pu}B, A_{up}B\} \) and \( \{A_{up}, A_{pu}, B, A_{up}B, A_{pu}B\} \), which will form the new partition at the next iteration. For each refined block, one among the subblocks of maximal size (here, \( \{A_{pu}, A_{up}, B\} \)) 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.
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
equivalence 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 equivalence 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).

Fig. 4. Forward equivalence for the FcR-I model of early events of ref. 32. (A) Graphical representation of the components involved in the pathway. Lyn kinase is recruited by the β subunit of FcR-I, while Syk kinase binds to the γ site. Syk is modeled with two phosphorylation units for the linker region and the activation loop. (B) Example of a class of the maximal forward equivalence: the complex conformation is equal up to the states of Syk’s phosphorylation units. The gray boxes represent a refinement which aggregates complexes equal up to the state of the linker region only, corresponding to the exactly reduced model discussed in ref. 33. We use solid and hollow circles to represent phosphorylated and unphosphorylated sites, respectively.

Fig. 5. Graphical description of representative equivalence classes (rounded boxes) obtained by computing the largest forward equivalence on the model of JAK activation by Barua et al. (35). (A) Singleton block where the complex consists of two JAK molecules with unphosphorylated sites Y1/Y2 (hollow circles) bound to a GH dimer. Under these conditions, the sites can be phosphorylated (solid circles). (B) Phosphorylation (solid circles) at site Y1 allows the binding of SH2-Bβ, which can dimer and bind to a phosphoinositide lipid. (C) Basic forward equivalence class with three species when a GH ligand/receptor complex undergoes constitutive turnover or endocytosis. (D) Since JAK2 molecules cannot bind to degraded/internalized complexes, the three complexes have effectively equivalent dynamics because they may only give rise to unbinding of the JAK2 molecule. A similar symmetry can be observed among the complexes in E, where additionally, SH2-Bβ and phosphoinositide can unbind.
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 reveals 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 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

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

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 $\sum_{i \in P} a_{ik} = \sum_{j \in \mathcal{P}} a_{jk}$ for any two equivalent players $i,j$ and for any equivalence 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.

**Fig. 6.** Graphical representation, using ref. 36, of the Boolean model for T-cell receptor signaling studied in refs. 37 and 38. Each node is a Boolean variable, whereas a directed arc describes an influence represented by the source variable appearing in the update function of the target variable. The network has three inputs (nodes with no incoming arcs): CD8, CD45, and the T-cell receptor TCRlig. Similar to ref. 38, we consider two variants with and without the feedback loops Fyn $\rightarrow$ PAGCsk and ZAP70 $\rightarrow$ cCbl (dashed arrows). Using the technique of ref. 38, we obtained a multivariate polynomial ODE system of degree five. On this, we fixed an initial partition where the input variables are singletons, ensuring that the largest backward equivalence that refines this partition reveals nodes with equivalent dynamics for any input. Nontrivial backward equivalence classes are represented with colored nodes with the same background. The class $\{cCbl, LAT\}$ is found only when the feedback loops are active. In this case, they are simultaneously subjected to the same influence by ZAP70. Indeed, backward equivalence turns out to aggregate the ODEs of nodes with update functions that are equal up to a renaming of nodes in the same equivalence class.
growth factor (EGF)-induced intermolecular autophosphorylation of the EGF recep-

Kozer N, et al. (2013) Exploring higher-order egfr oligomerisation and
covalent ligand-receptor interactions with steric constraints on configurations of cell-

the order of phosphate processing and protein-protein interactions. *FEBS Lett* 514:106-
1061.

Kozer N, et al. (2013) Exploring higher-order egfr oligomerisation and
phosphorylation-a combined experimental and theoretical approach. *Mol Biosyst* 9:
1849-1863.

growth factor (EGF)-induced intermolecular autophosphorylation of the EGF recep-


bisimulations for chemical reaction networks. *Proceedings of the 26th International
Conference on Concurrency Theory (CONCUR)* (Schloss Dagstuhl-Leibniz-Zentrum
für Informatik, Wadern, Germany), Vol 9207, pp 226-239.


Sneddon MW, Faeder JR, Emonet T (2011) Efficient modeling, simulation and coarse-

Hutchcroft JE, Geahlen RL, Deenin GG, Oliver JM (1992) FcR-mediated tyrosine phos-
phorylation and activation of the 72-kda protein-tyrosine kinase, ptk72, in rii-h3 rat

Barus D, Faeder JR, Haugh JM (2009) A bipolar clamp mechanism for activation of

Chaoiuia C, Naldi A, Thieffy D (2012) Bacterial Molecular Networks: Methods and
Protocols, eds van Helden J, Tousaint A, Thieffy D (Springer, New York), pp 463-
495.

for the structural and functional analysis of signaling and regulatory networks. *BMC

Wittmann DM, et al. (2009) Transforming boolean models to continuous models:
Methodology and application to t-cell receptor signaling. *BMC Syst Biol* 3:1–21.

Le Novère N (2015) Quantitative and logical modeling of molecular and gene net-
works. *Nat Rev Genet* 16:146-158.

Glass L, Kauffman SA (1973) The logical analysis of continuous, non-linear biochemical


Acad Sci USA* 111(Suppl 3):10918–10917.


Anderson J, Chang YC, Papachristodoulou A (2011) Model decomposition and reduc-


Probab* 31:95–75.

lumping and reductions for large biological models. *IET Syst Biol* 2:342-351.


Görnerup O, Jacob I MN (2010) A method for finding aggregated representations of

Simon PL, Taylor M, Kios IZ (2010) Exact epidemic models on graphs using graph-

Anderson J, Chang YC, Papachristodoulou A (2011) Model decomposition and reduc-


Probab* 31:95–75.

lumping and reductions for large biological models. *IET Syst Biol* 2:342-351.


Görnerup O, Jacob I MN (2010) A method for finding aggregated representations of

Simon PL, Taylor M, Kios IZ (2010) Exact epidemic models on graphs using graph-

Anderson J, Chang YC, Papachristodoulou A (2011) Model decomposition and reduc-


Probab* 31:95–75.

lumping and reductions for large biological models. *IET Syst Biol* 2:342-351.


Görnerup O, Jacob I MN (2010) A method for finding aggregated representations of

Simon PL, Taylor M, Kios IZ (2010) Exact epidemic models on graphs using graph-

Anderson J, Chang YC, Papachristodoulou A (2011) Model decomposition and reduc-


Probab* 31:95–75.

lumping and reductions for large biological models. *IET Syst Biol* 2:342-351.


Görnerup O, Jacob I MN (2010) A method for finding aggregated representations of

Simon PL, Taylor M, Kios IZ (2010) Exact epidemic models on graphs using graph-

Anderson J, Chang YC, Papachristodoulou A (2011) Model decomposition and reduc-


Probab* 31:95–75.