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

Maximal aggregation of polynomial dynamical systems

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

[a]Microsoft Research, Cambridge CB1 2FB, United Kingdom; [b]Department of Computing, University of Oxford, Oxford OX1 3QD, United Kingdom; and [c]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 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.

polynomial dynamical systems | aggregation | partition refinement

Several models in natural and engineering sciences can be described as a system of ordinary differential equations (ODEs) with polynomial derivatives. A frequent concern is the treatment of highly dimensional ODEs because of their unintelligibility as well as the numerical difficulties caused by the large computational cost of the analysis. Reduction techniques based on singular value decomposition and Krylov subspace methods have proved effective in producing reduced models with small approximation errors (1, 2). However, albeit advantageous for numerical simulations, these transformations often lead to loss of structure and physical interpretability. This is a major limitation when the model is used for predictive purposes or for validating mechanistic assumptions (3, 4). An alternative is to aggregate groups of variables into macrovariables, for which an ODE system can be explicitly derived. This has been successfully pursued, for instance, using domain-specific techniques in computational systems biology, where detailed mechanistic mass action ODE models of protein interaction networks may incur combinatorial explosion of the state space (5–7).

Here we propose a generic, domain-agnostic aggregation method for ODEs with polynomial derivatives of any degree based on equivalence relations (i.e., partitions) over the ODE variables. It can be seen as an instance of a family of techniques that consider arbitrary linear transformations of the state space studied for a long time across many disciplines, such as chemistry (8), ecology (9), and control theory (1). In this context, checking whether a generic linear transformation induces an exact aggregation is well-understood. Instead, it has been frequently pointed out that a major limitation concerns the automatic generation of a candidate transformation (10–13). This drawback does not allow one to unravel simpler dynamics from systems of realistic size in practice.

By contrast, we develop an efficient method for computing the largest equivalence, leading to the maximal aggregation of an ODE system. This is achieved by means of a partition refinement algorithm that iteratively splits an initial partition of variables until a fixed point. The maximal aggregation can be obtained by starting the algorithm with the trivial partition having all ODE variables in a single block. Furthermore, the freedom in choosing an arbitrary initial partition is instrumental to producing reductions that preserve the dynamics of desired original variables, which are then not aggregated.

Mathematically, our approach is a generalization of well-known equivalence relations for Markov chains named lumpability (14). Ordinary lumpability relates states that have the same aggregate transition rate toward every equivalence class (thus, it is a forward criterion); in exact lumpability, two equivalent states have the same aggregate rate from every equivalence class (thus, it is a backward criterion). In a conceptually similar spirit, we define forward equivalence as a relation whereby each equivalence class describes the evolution of the sum of ODE variables in the original model. Backward equivalence identifies variables that have the same solutions at all time points (hence, they must start from the same initial conditions). Indeed, forward and backward equivalence collapse to ordinary and exact lumpability, respectively, when the (linear) ODE system is the equation of motion for the transient probability distribution of a continuous time Markov chain (15).

Our technique describes the equivalences in finitary terms, despite that they involve continuous ODE variables. We encode a polynomial ODE system into a reaction network (RN), a structure akin to a formal chemical reaction network (CRN), with one species per ODE variable and one reaction per monomial in the derivatives. The equivalences are then relations over species based on quantities computed by inspecting the reactions. This structural interpretation allows the development of an algorithm for computing maximal equivalences, building on analogous partition refinement techniques developed for Markov chain lumping (16, 17). These enjoy polynomial time and space complexity, owing to the seminal work on foundational problems of computer science by Paige and Tarjan (18).

Our contribution extends recent works that presented an alternative aggregation method based on a logical encoding into a satisfiability problem (19) (however applicable only to ODE systems of moderate size) and an RN encoding for ODE systems with polynomial derivatives of degree at most two (15), with the further limitation that the criterion for forward equivalence was only a sufficient condition for aggregation.

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.

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

[2]To 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

$$\frac{dx(t)}{dt} = \mathcal{P}(x(t)), \qquad x = (x_1, \ldots, x_n), \qquad \textbf{[1]}$$

where $\mathcal{P}$ 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 $x_j \in H_i$ 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

$$A\mathcal{P}(x) = A\mathcal{P}(\bar{A}Ax), \qquad \text{for all } x, \qquad \textbf{[2]}$$

where $\bar{A}$ 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

$$\mathcal{P}(\mathcal{U}_{\mathcal{H}}) \subseteq \mathcal{U}_{\mathcal{H}}, \text{ with } \mathcal{U}_{\mathcal{H}} = \{x \in \mathbb{R}^n \mid x \text{ uniform on } \mathcal{H}\}. \qquad \textbf{[3]}$$

To obtain a reduced model, we set $\tilde{\mathcal{P}} = A\mathcal{P}\bar{A}$, with the generalized inverse $\bar{A} = (\bar{a}_{ij})$, such that $\bar{a}_{ij} = a_{ji} / \sum_k a_{jk}$. Let us consider the reduced ODE system:

$$\frac{dy(t)}{dt} = \tilde{\mathcal{P}}(y(t)), \qquad y = (y_1, \ldots, y_m). \qquad \textbf{[4]}$$

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 in general, the individual trajectories cannot be recovered. Instead, if 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 $\mathcal{MS}(S)$ the set of all multisets with elements in S. Each reaction is in the form $\mu \xrightarrow{\alpha} \mu'$, where μ and μ' are multisets of species (called reagents and products, respectively), and the coefficient α 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_i with species S_i and each monomial $\alpha \prod_i x_i^{p_i}$ appearing in the ODE of x_k with the reaction

$$\sum_{i=1}^{n} p_i S_i \xrightarrow{\alpha} S_k + \sum_{i=1}^{n} p_i S_i, \qquad \textbf{[5]}$$

where the operator $+$ denotes multiset 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 $S_i \in S$ and of a set of species $G \subseteq S$ because of reagents ρ:

$$\phi(\rho, S_i) := \sum_{(\rho \xrightarrow{\alpha} \pi) \in R} (\pi_i - \rho_i) \cdot \alpha \text{ and } \phi(\rho, G) := \sum_{S_i \in G} \phi(\rho, S_i),$$

where ρ_i and π_i denote the multiplicity of species S_i in the reagents and products, respectively.

To characterize forward equivalence, we further define

$$\mathbf{fr}(S_i, \rho, G) := \frac{\phi(S_i + \rho, G)}{[S_i + \rho]!},$$

where the operator $[\cdot]!$ denotes the multinomial coefficient induced by a multiset of species:

$$[\rho]! := \binom{\sum_{i=1}^{n} \rho_i}{\rho_1, \ldots, \rho_n}.$$

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

$$\mathbf{fr}(S_i, \rho, H') = \mathbf{fr}(S_j, \rho, H') \qquad \textbf{[6]}$$

for all ρ, 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

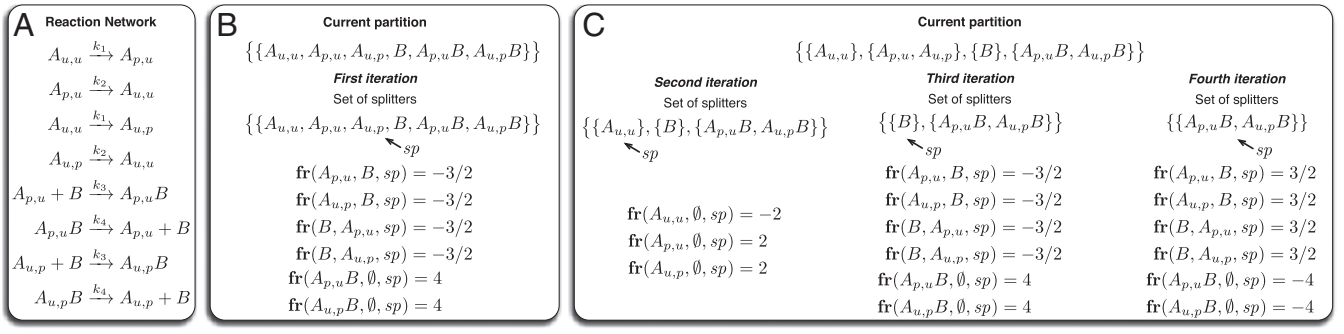

Fig. 1. Example of reduction with forward equivalence. (*A*) A CRN of a basic mechanism of reversible binding between molecular species, *A* and *B*, through *A*'s two identical binding sites. The state of the site is denoted by the subscripts *u* (unphosphorylated) and *p* (phosphorylated). For simplicity, 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 $\mathbf{fr}(\cdot, \cdot, sp)$ (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 $\mathbf{fr}(S_i, \rho, sp)$ and $\mathbf{fr}(S_j, \rho, sp)$ for every partner ρ. The first iteration produces the subblocks $\{A_{u,u}\}$, $\{A_{p,u}, A_{u,p}\}$, $\{B\}$, and $\{A_{p,u}B, A_{u,p}B\}$, which will form the new partition at the next iteration. For each refined block, one among the subblocks of maximal size (here, $\{A_{p,u}, A_{u,p}\}$) 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.

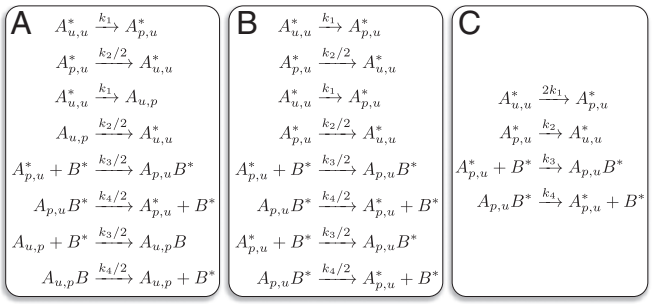

Fig. 2. Illustrative example of the three-step reduction up to forward/backward equivalence using the RN of Fig. 1*A* and the largest forward equivalence of Fig. 1*C*. In the first step, (*A*) the rate of each reaction is divided by the product of the cardinalities of the equivalence classes to which the reagents belong; we use an asterisk to indicate the representative of each equivalence class. (*B*) Each reaction is rewritten by replacing every species with its representative. (*C*) Reactions with same reagents and products are merged by summing the rates.

only as reagent). Intuitively, two species are related whenever they provide same net instantaneous stoichiometry, for any reaction partner ρ, 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:

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

Also, we define

$$\mathbf{br}(S_i, \mathcal{M}, H') := \sum_{S_k \in H'} \sum_{\rho \in \mathcal{M}} \frac{\phi(S_k + \rho, S_i)}{|S_k + \rho|_{H'}},$$

where $|\pi|_{H'} = |\{S_l \in H' | \pi_l > 0\}|$ counts the number of different species in H' occurring in π. Then, $\mathcal{H}$ 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

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

for all $H' \in \mathcal{H}$ and $\mathcal{M} \in \{\rho | (S_k + \rho \xrightarrow{\alpha} \pi) \in R, S_k \in S\}/\approx_{\mathcal{H}}$ (*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

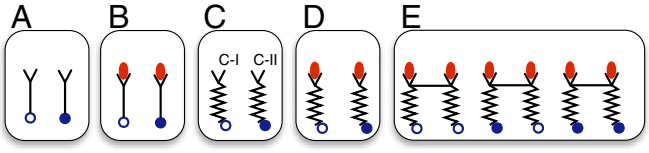

Fig. 3. Representative forward equivalence classes for the model of ref. 28. Forward equivalence aggregates molecular complexes that are equal up to the states of the phosphorylation site (hollow/solid blue circles) of EGFR. *A* shows the two-species equivalence class for EGFR (Y-shaped) without conformational change of the cytosolic tail. (*B*) The equivalence class aggregates EGFR when it is bound to EGF (solid red ellipses) as well as when (*C* and *D*) the cytosolic tail has undergone conformational change (wiggled lines). These basic patterns of equivalence carry over to all oligomers that are formed through ectodomain cross-linking, such as (*E*) the three possible phosphorylation states of dimers with changed cytosolic tail and bound to EGF.

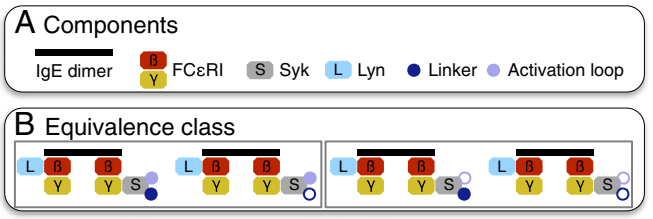

Fig. 4. Forward equivalence for the FcεRI model of early events of ref. 32. (*A*) Graphical representation of the components involved in the pathway. Lyn kinase is recruited by the β subunit of FcεRI, while Syk kinase binds to the γ site. Syk is modeled with two phosphorylation units for the linker region 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.

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. 3*C*, 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. 3*C*, 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 nontrivial 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).

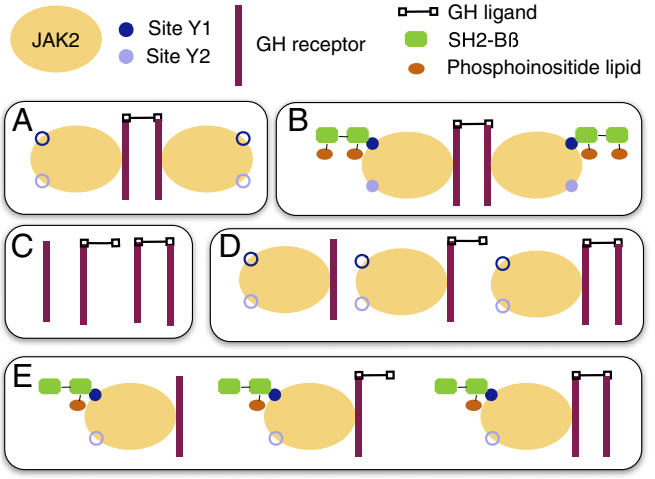

Fig. 5. Graphical description of representative equivalence classes (rounded boxes) obtained by computing the largest forward equivalence on the model of JAK activation by Barua et al. (35). (*A*) Singleton block where the complex consists of two JAK molecules with unphosphorylated sites Y1/Y2 (hollow 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.

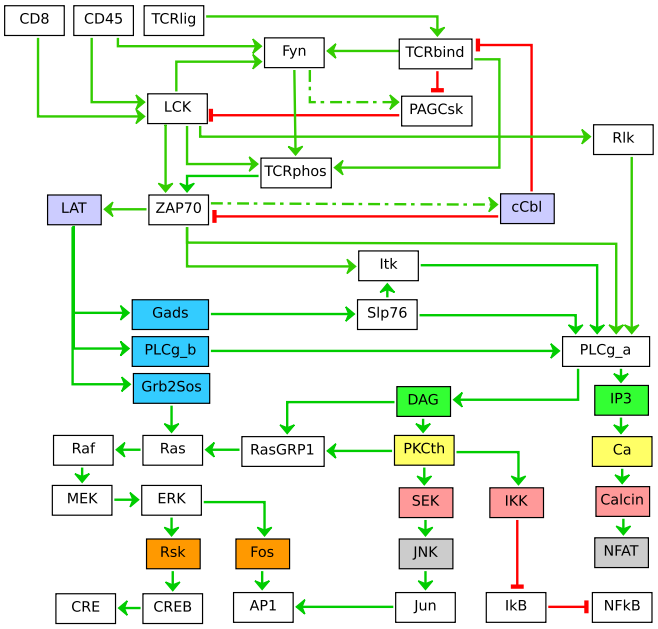

Fig. 6. Graphical representation, using ref. 36, of the Boolean model for T-cell receptor signaling studied in refs. 37 and 38. Each node is a Boolean variable, whereas a directed arc describes an influence represented by the source variable appearing in the update function of the target variable. The 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.

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 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 ith strategy, with $1 \leq i \leq n$. Its evolution is governed by the polynomial ODE system

$$\dot{x}_i = x_i\big((Bx)_i - x^T Bx\big), \qquad 1 \leq i \leq n, \qquad \textbf{[8]}$$

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

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

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 $\sum_{k \in P} a_{ik} = \sum_{k \in P} a_{jk}$ 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.

1. Antoulas A (2005) *Approximation of Large-Scale Dynamical Systems*, Advances in Design and Control (SIAM, Philadelphia).
2. Gu C (2011) QLMOR: A projection-based nonlinear model order reduction approach using quadratic-linear representation of nonlinear systems. *IEEE Trans Comput Aided Des Integrated Circ Syst* 30:1307–1320.
3. Sunnaker M, Cedersund G, Jirstrand M (2011) A method for zooming of nonlinear models of biochemical systems. *BMC Syst Biol* 5:140.
4. Apri M, de Gee M, Molenaar J (2012) Complexity reduction preserving dynamical behavior of biochemical networks. *J Theor Biol* 304:16–26.
5. Conzelmann H, Saez-Rodriguez J, Sauter T, Kholodenko B, Gilles E (2006) A domain-oriented approach to the reduction of combinatorial complexity in signal transduction networks. *BMC Bioinformatics* 7:34.
6. Borisov NM, Chistopolsky AS, Faeder JR, Kholodenko BN (2008) Domain-oriented reduction of rule-based network models. *IET Syst Biol* 2:342–351.
7. Feret J, Danos V, Krivine J, Harmer R, Fontana W (2009) Internal coarse-graining of molecular systems. *Proc Natl Acad Sci USA* 106:6453–6458.
8. Okino MS, Mavrovouniotis ML (1998) Simplification of mathematical models of chemical reaction systems. *Chem Rev* 2:391–408.
9. Iwasa Y, Andreasen V, Levin S (1987) Aggregation in model ecosystems. I. Perfect aggregation. *Ecol Modell* 37:287–302.
10. Görnerup O, Jacobi MN (2010) A method for finding aggregated representations of linear dynamical systems. *Adv Complex Syst* 13:199–215.
11. Simon PL, Taylor M, Kiss IZ (2010) Exact epidemic models on graphs using graph-automorphism driven lumping. *J Math Biol* 62:479–508.
12. Anderson J, Chang YC, Papachristodoulou A (2011) Model decomposition and reduction tools for large-scale networks in systems biology. *Automatica* 47:1165–1174.
13. Turanyi T, Tomlin AS (2014) *Analysis of Kinetic Reaction Mechanisms* (Springer, Berlin).
14. Buchholz P (1994) Exact and ordinary lumpability in finite Markov chains. *J Appl Probab* 31:59–75.
15. Cardelli L, Tribastone M, Tschaikowski M, Vandin A (2016) Efficient syntax-driven lumping of differential equations. *Proceedings of the 21st International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)*, Lecture Notes in Computer Science (Springer, Berlin), Vol 9636, pp 93–111.
16. Derisavi S, Hermanns H, Sanders WH (2003) Optimal state-space lumping in Markov chains. *Inform Process Lett* 87:309–315.
17. Valmari A, Franceschinis G (2010) Simple $O(m \log n)$ time Markov chain lumping. *Proceedings of the 16th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)*, Lecture Notes in Computer Science (Springer, Berlin), Vol 6015, pp 38–52.
18. Paige R, Tarjan R (1987) Three partition refinement algorithms. *SIAM J Comput* 16:973–989.
19. Cardelli L, Tribastone M, Tschaikowski M, Vandin A (2016) Symbolic computation of differential equivalences. *Proceedings of the 43rd Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL)* (ACM, New York), pp 137–150.
20. Tóth J, Li G, Rabitz H, Tomlin AS (1997) The effect of lumping and expanding on kinetic differential equations. *SIAM J Appl Math* 57:1531–1556.
21. Salazar C, Höfer T (2009) Multisite protein phosphorylation – from molecular mechanisms to kinetic models. *FEBS J* 276):3177–3198.
22. Thomson M, Gunawardena J (2009) Unlimited multistability in multisite phosphorylation systems. *Nature* 460:274–277.
23. Gunawardena J (2005) Multisite protein phosphorylation makes a good threshold but can be a poor switch. *Proc Natl Acad Sci USA* 102:14617–14622.
24. Perelson AS, DeLisi C (1980) Receptor clustering on a cell surface. I. Theory of receptor cross-linking by ligands bearing two chemically identical functional groups. *Math Biosci* 48:71–110.
25. Kocieniewski P, Faeder JR, Lipniacki T (2012) The interplay of double phosphorylation and scaffolding in MAPK pathways. *J Theor Biol* 295:116–124.
26. Monine MI, Posner RG, Savage PB, Faeder JR, Hlavacek WS (2010) Modeling multivalent ligand-receptor interactions with steric constraints on configurations of cell-surface receptor aggregates. *Biophys J* 98:48–56.
27. Salazar C, Höfer T (2007) Versatile regulation of multisite protein phosphorylation by the order of phosphate processing and protein–protein interactions. *FEBS J* 274:1046–1061.
28. Kozer N, et al. (2013) Exploring higher-order egfr oligomerisation and phosphorylation-a combined experimental and theoretical approach. *Mol Biosyst* 9: 1849–1863.
29. Honegger AM, Schmidt A, Ullrich A, Schlessinger J (1990) Evidence for epidermal growth factor (EGF)-induced intermolecular autophosphorylation of the EGF receptors in living cells. *Mol Cell Biol* 10:4035–4044.
30. Koschorreck M, Conzelmann H, Ebert S, Ederer M, Gilles E (2007) Reduced modeling of signal transduction — A modular approach. *BMC Bioinformatics* 8:336.
31. Cardelli L, Tribastone M, Tschaikowski M, Vandin A (2015) Forward and backward bisimulations for chemical reaction networks. *Proceedings of the 26th International Conference on Concurrency Theory (CONCUR)* (Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Wadern, Germany), pp 226–239.
32. Faeder JR, et al. (2003) Investigation of early events in FcεRI-mediated signaling using a detailed mathematical model. *J Immunol* 170:3769–3781.
33. Sneddon MW, Faeder JR, Emonet T (2011) Efficient modeling, simulation and coarse-graining of biological complexity with NFsim. *Nat Methods* 8:177–183.
34. Hutchcroft JE, Geahlen RL, Deanin GG, Oliver JM (1992) FcεRI-mediated tyrosine phosphorylation and activation of the 72-kda protein-tyrosine kinase, ptk72, in rbl-2h3 rat tumor mast cells. *Proc Natl Acad Sci USA* 89:9107–9111.
35. Barua D, Faeder JR, Haugh JM (2009) A bipolar clamp mechanism for activation of Jak-family protein tyrosine kinases. *PLoS Comput Biol* 5:e1000364.
36. Chaouiya C, Naldi A, Thieffry D (2012) *Bacterial Molecular Networks: Methods and Protocols*, eds van Helden J, Toussaint A, Thieffry D (Springer, New York), pp 463–479.
37. Klamt S, Saez-Rodriguez J, Lindquist JA, Simeoni L, Gilles ED (2006) A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics* 7:1–26.
38. 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.
39. Le Novere N (2015) Quantitative and logic modelling of molecular and gene networks. *Nat Rev Genet* 16:146–158.
40. Glass L, Kauffman SA (1973) The logical analysis of continuous, non-linear biochemical control networks. *J Theor Biol* 39:103–129.
41. de Jong H, et al. (2004) Qualitative simulation of genetic regulatory networks using piecewise-linear models. *Bull Math Biol* 66:301–340.
42. Cressman R, Tao Y (2014) The replicator equation and other game dynamics. *Proc Natl Acad Sci USA* 111(Suppl 3):10810–10817.
43. Taylor PD, Jonker LB (1978) Evolutionary stable strategies and game dynamics. *Math Biosci* 40:145–156.
44. Ohtsuki H, Nowak MA (2006) The replicator equation on graphs. *J Theor Biol* 243: 86–97.
45. Iacobelli G, Madeo D, Mocenni C (2016) Lumping evolutionary game dynamics on networks. *J Theor Biol* 407:328–338.
46. Gorban AN, Karlin IV (2005) *Invariant Manifolds for Physical and Chemical Kinetics*, Lecture Notes in Physics (Springer, Berlin), Vol 660.
47. Larsen KG, Skou A (1991) Bisimulation through probabilistic testing. *Inf Comput* 94: 1–28.
48. Baier C, Engelen B, Majster-Cederbaum ME (2000) Deciding bisimilarity and similarity for probabilistic processes. *J Comput Syst Sci* 60:187–231.
49. Blinov ML, Faeder JR, Goldstein B, Hlavacek WS (2004) BioNetGen: Software for rule-based modeling of signal transduction based on the interactions of molecular domains. *Bioinformatics* 20:3289–3291.
50. Cardelli L, Tribastone M, Tschaikowski M, Vandin A (2017) ERODE: A tool for the evaluation and reduction of ordinary differential equations. *Proceedings of the 22nd International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS)*, Lecture Notes in Computer Science (Springer, Berlin), Vol 10206, pp 310–328.
51. Zhang J, Zhang F, Ebert D, Cobb MH, Goldsmith EJ (1995) Activity of the MAP kinase ERK2 is controlled by a flexible surface loop. *Structure* 3:299–307.
52. Wanant S, Quon MJ (2000) Insulin receptor binding kinetics: Modeling and simulation studies. *J Theor Biol* 205:355–364.
53. Kolczyk K, Samaga R, Conzelmann H, Mirschel S, Conradi C (2012) The process-interaction-model: A common representation of rule-based and logical models allows studying signal transduction on different levels of detail. *BMC Bioinformatics* 13:251.
54. Conzelmann H, Fey D, Gilles E (2008) Exact model reduction of combinatorial reaction networks. *BMC Syst Biol* 2:78.
55. Bajpai A, et al. (2013) Dynamics of SIN asymmetry establishment. *PLoS Comput Biol* 9:1–11.
56. Naldi A, Rem E, Thieffry D, Chaouiya C (2009) A reduction of logical regulatory graphs preserving essential dynamical properties. *Proceedings of the 7th Internal Conference on Computational Methods in Systems Biology*, Lecture Notes in Computer Science, eds Degano P, Gorrieri R (Springer, Berlin), Vol 5688, pp 266–280.
57. Sahin Ö, et al. (2009) Modeling ERBB receptor-regulated G1/S transition to find novel targets for de novo trastuzumab resistance. *BMC Syst Biol* 3:1.
58. Niarakis A, et al. (2014) Computational modeling of the main signaling pathways involved in mast cell activation. *Curr Top Microbiol Immunol* 382:69–93.