Piecewise smooth hybrid systems as models for networks in molecular biology

Vincent NOEL[1], Sergei VAKULENKO[4] and Ovidiu RADULESCU[2,3]

[1] Université de Rennes 1 - CNRS 6025 (IRMAR), Campus de Beaulieu, 35042 Rennes, France
vincent.noel@univ-rennes1.fr
[2] Université de Montpellier 2, DIMNP - UMR 5235 CNRS/UM1/UM2, Pl. E. Bataillon, Bat 24, CP 107, 34095 Montpellier Cedex 5, France
[3] INRIA Rennes Bretagne Atlantique, Campus de Beaulieu, 35042 Rennes, France
ovidiu.radulescu@univ-montp2.fr
[4] Saint Petersburg State University of Technology and Design, Bolshoya Morskaya 18, Saint Petersburg, Russia

Abstract *We discuss piecewise smooth hybrid systems as models for regulatory networks in molecular biology. These systems involve both continuous and discrete variables. In the context of gene networks, the discrete variables allow to switch on and off some of the molecular interactions in the model of the biological system. Piecewise smooth hybrid models are well adapted to approximate the dynamics of multiscale dissipative systems that occur in molecular biology. We show how to produce such models by a top down approach that use biological knowledge for a guided choice of important variables and interactions. Then we propose an algorithm for fitting parameters of the piecewise smooth models from data. We illustrate some of the possibilities of this approach by proposing a minimal piecewise smooth model for the cell cycle.*

Keywords systems biology, hybrid models, cell cycle

1 Introduction

Hybrid systems are widely used in automatic control theory to cope with situations arising when a finite-state machine is coupled to mechanisms that can be modeled by differential equations [11]. It is the case of robots, plant controllers, computer disk drives, automated highway systems, flight control, etc. The general behavior of such systems is to pass from one type of smooth dynamics (mode) described by one set of differential equations to another smooth dynamics (mode) described by another set of differential equations. The command of the modes can be performed by changing one or several discrete variables. The mode change can be accompanied or not by jumps (discontinuities) of the trajectories.

Depending on how the discrete variables are changed there may be several types of hybrid systems: switched systems [14], multivalued differential automata [15], piecewise smooth systems [2]. Notice that in the last case, the mode changes when the trajectory attains some smooth manifolds.

Piecewise affine hybrid systems have been used to model dynamics of gene networks [1,3]. In these networks, most of the time, the gene variables are close to discrete values (attractors) and the transitions between discrete attractors are dictated by the relative position of the transient values of these variables with respect to some thresholds. The transient dynamics leading to attractors is considered to be piecewise affine where the linear part of the dynamical equations is defined by a diagonal matrix with negative entries. This approximation allows to reduce the dynamics of simple genetic circuits to a discrete automaton, and can be used for various application such as model checking. However, the study of large networks with this approach suffers from combinatorial explosion.

We must emphasize that piecewise affine models are not always good approximations for the dynamics of the modes. The machinery of the cell cycle is an example. Proteolytic degradation of the cyclins is switched on rapidly by the cyclin dependent kinase complexes but between two successive switchings the complexes have non-linear dynamics implying several positive (autocatalytic processes) and negative feedback loops. These non-linear processes contribute to the robustness of the mechanism. Another example is the dynamics of the genetically regulated metabolism. Genetic changes could be considered as boolean variables that are turned on and off by their mutual interaction and by the interaction with the metabolites, but between two successive switchings of the gene expression the dynamics of metabolism is not linear. More generally, the dynamics of multi-scale network belongs to a patchy landscape formed by smooth, low

dimensional, but curved manifolds, connected by discontinuous transitions. The patches represent low dimensional local invariant manifolds, typical for multiscale dissipative systems, and the transitions correspond to bifurcations of these manifolds [7,6]. Piecewise smooth systems can provide more realistic and more robust models describing these situations.

The idea of piecewise smooth patchy landscape arises naturally from the model reduction theory. The dynamics of a multiscale, but nonlinear large model, can be reduced to the one of a dominant subsystem [12,9,8]. In dynamical systems with separation of timescales the dominant subsystem depends on the relative contributions of different variables to the timescales and on the comparison between timescales. Both the contributions of different variables to the timescales of the dynamics and the comparison among timescales (which timescale is slower which one is quicker) can change along a trajectory of the system. Considering that the set of dominant subsystems is finite, the changes are necessarily discrete. Thus, although one may try and sometimes succeed to find a global reduced model, the general picture in the case of multiscale non-linear dissipative systems is a sequence of several approximations (modes) valid locally. The modes integrate the degrees of freedom of the system that are active for a certain time interval [12,9,8].

The problem of how the modes can be rigorously approximated for a given multiscale nonlinear model will be approached elsewhere. In this paper we propose a heuristic to construct appropriate modes and adequate piecewise smooth models by using a top-down approach. Then, we show how the parameters of the hybrid model can be fitted from data or from trajectories produced by existing smooth, but more complex models.

2 Hybrid models

We consider the so-called hybrid dynamical systems (HDS) consisting of two components: a continuous part, u, defined by

$$\frac{du_i}{dt} = f_i(u(t), s(t)), \quad t > 0, \tag{2.1}$$

where $u = (u_1, u_2, ..., u_n) \in \mathbf{R}^n$, and a discrete part $s(t) \in S$, where S is a finite set of states. For molecular networks, the continuous variables are protein concentrations and the discrete states may be gene activities described by boolean variables $s = (s_1(t), s_2(t), ..., s_m(t))$, where $s_j \in \{0, 1\}$ (such boolean gene models are popular, see [4,10] among many others).

There are several possible ways to define the evolution of the s variables. Rather generally, this can be done by a time continuous Markov chain with transition probabilities $p(s, s', u)$ from the state s to the state s' (per unit time) depending on current state $u(t)$. However, in gene networks, transition probabilities dependence on u is not smooth. For instance, the probability for s to jump is close to one if u goes above some threshold value, and close to zero if u is smaller than the threshold. We can, in certain cases, neglect the transition time with respect to the time needed for u variables to change. Assuming that some of the discrete variables contribute to production of u and that other contribute to the degradation of u we obtain a general model of hybrid piece-wise smooth dynamical system :

$$\begin{aligned}
\frac{du_i}{dt} &= \sum_{k=1}^{N} s_k P_{ik}(u) + P_i^0(u) \\
&\quad - \sum_{l=1}^{M} \tilde{s}_l Q_{il}(u) - Q_i^0(u), \\
s_j &= H(\sum_{k=1}^{n} w_{jk} u_k - h_j), \\
\tilde{s}_l &= H(\sum_{k=1}^{n} \tilde{w}_{lk} u_k - \tilde{h}_l)
\end{aligned} \tag{2.2}$$

where H is the unit step function $H(y) = 1$, $y \geq 0$, and $H(y) = 0$, $y < 0$, $P_{ik}, P_i^0, Q_{il}, Q_i^0$ are positive, smooth functions of u_i representing production, basal production, consumption, and basal consumption, respectively. Here $w, \tilde{w}$ are matrices describing the interactions between the u variables, $i = 1, 2, ..., n$, $j = 1, 2, ..., N$, $l = 1, ..., M$ and $h, \tilde{h}$ are thresholds.

The class of models (2.2) is still too general. We shall restrict ourselves to a subclass of piecewise smooth systems where smooth production and degradation terms will be assumed multivariate monomials in u, plus some basal terms:

$$\begin{aligned}
P_{ik}(\mathbf{u}) &= a_{ik} u_1^{\alpha_1^{ik}} \dots u_n^{\alpha_n^{ik}}, \\
P_i^0(\mathbf{u}) &= a_i^0 \\
Q_{il}(\mathbf{u}) &= \tilde{a}_{il} u_1^{\tilde{\alpha}_1^{il}} \dots u_n^{\tilde{\alpha}_n^{il}} \\
Q_i^0(\mathbf{u}) &= \tilde{a}_i^0 u_i
\end{aligned} \tag{2.3}$$

which will be chosen according to a heuristic presented in the next section.

These models have several advantages with respect to standard models in molecular biology and neuroscience based on differential equations. They allow

us to simulate, in a fairly simple manner, discontinuous transitions occurring in such systems (see a typical graph describing time evolution of protein concentration within cellular cell cycle, Fig. 4.1). The discontinuous transitions result either from fast processes or from strongly non-linear (thresholding) phenomena. This class of models is also scalable in the sense that more and more details can be introduced at relatively low cost, by increasing the number of discrete variables and the size of the interaction matrices.

The definition of the modes slightly extends the one of S-systems, introduced by Savageau [13]. Our choice was motivated by the fact that S-systems proved their utility as models for metabolic networks whose dynamics we want to encompass by considering the modes. The introduction of basal terms avoids spurious long living states when some products have zero concentrations.

The monomial rates can be fully justified for linear networks of biochemical reactions with totally separated constants. The same is true for nonlinear mechanisms resulting from mass action law for instance. In general simplified rates of complex mechanisms can be rational functions of the concentrations. However, when concentrations are very large or very small the monomial power laws are recovered. For a multiscale system changing regime (for instance a Michaelis Menten reaction switching from a saturated enzyme regime to a small concentration substrate regime) one can use the discrete variables to illustrate the change.

In the next section we illustrate the possibilities of this model and show that (2.2) can simulate the mitotic oscillations of the cell cycle.

3 Heuristic for choosing the discrete variables and the multivariate monomial terms

The interactions between the molecular variables of the model can occur at several levels:

i) The discrete interactions.

Discrete interactions manifest themselves punctually as a consequence of thresholding of rapid phenomena. They contribute to changing the discrete variables s_j, $\tilde{s}_j$.

One protein can contribute to switching on or off the discrete variables commanding the production or the degradation of another protein. The action of u_i on u_j is positive (an activation) if $w_{ji} > 0$ (contribute to turn on production) or if $\tilde{w}_{ji} < 0$ (contribute to turn off degradation). Conversely the action of u_i on u_j is negative if $w_{ji} < 0$ or if $\tilde{w}_{ji} > 0$.

ii The continuous interactions.

The continuous interactions guide the dynamics of the modes. During the mode dynamics the variables s_j, $\tilde{s}_j$ are fixed. The continuous variable u_i activates u_j if either $\alpha_j^{ik} > 0$ or $\tilde{\alpha}_j^{il} < 0$, for some k, l. Conversely, u_i inactivates u_j if either $\alpha_j^{ik} < 0$ or $\tilde{\alpha}_j^{il} > 0$, for some k, l.

In the following we provide a heuristic allowing to produce hybrid models.

In order to define a hybrid model we need a hybrid interaction scheme. This consists in saying, for each given species, whether its production/degradation can be switched on and off and by which species, also which species modulate the production/degradation of a given species in a smooth way. The representation of the hybrid interaction scheme can be given as a regulated reaction graph.

A regulated reaction graph is a quadruple (V, R, E, E_r). The triplet (V, R, E), where $E \subset V \times R \cup R \times V$, defines a reaction bipartite graph, ie $(x, y) \in E$ iff $x \in V, y \in R$ and x is a substrate of R, or $x \in R, y \in V$ and y is a product of x. $E_r \subset V \times R$ is the set that defines regulations, for instance $(x, z) \in E_r$ if $x \in V$ regulates $z \in R$.

Consistently with the choice (2.2),(2.3) for piecewise-smooth systems the stoichiometry of the reaction graph (V, R, E) is mono-molecular, any reaction has at most one substrate and at most one product (generalizations are possible, but will not be discussed here).

Some of the regulations in E_r are discrete and some are continuous and we can define the partition $E_r = E_r^d \cup E_r^c$. Similarly, there is a partition of the reactions $R = R^c \cup R^s$. A reaction y belongs to the switched reactions $y \in R^s$ if $(x, y) \in E_r^d$, for some $x \in V$.

The role of the regulators (continuous if they modulate the reaction rate, discrete if they contribute to switching it on and off) should be indicated on the graph together with the signs of the regulations.

Given a reaction, we identify its substrate and the regulators. The non-basal term in the reaction rate is a product of the concentrations of the substrates, concentrations of activators, divided by the concentrations of inhibitors. The basal term is constant if there is no substrate, or proportional to the concentration of the substrate (for instance in consumption reactions).

Assuming that there are n species $u \in \mathbb{R}^n$ and that the reactions have stoichiometric vectors ν_j, $1 \leq j \leq m$, one obtains the following piecewise-smooth model:

$$\frac{d\mathbf{u}}{dt} = \sum_{j\in R^c} \nu_j R_j(\mathbf{u}) + \sum_{k\in R^d} \nu_k (R_k(\mathbf{u})\sigma_k(\mathbf{u}) + R_k^0(\mathbf{u})) \tag{3.1}$$

where $\sigma_k(\mathbf{u}) = H(\sum_{(i,j)\in E^r} w_{kj}u_j - h_k)$. The relation between σ_k and $s_i, \tilde{s}_j$ from Eq.2.2 is straightforward.

The reaction rates have the forms given by (2.3). The monomial exponents α_{ij}, $\tilde{\alpha}_j^i$ and the final rates can be obtain from the following heuristic rules:

i) If a reaction j is activated then $\alpha_j^i = 1$ for all activators and $\alpha_j^i = -1$ for all inhibitors i in the absence of cooperativity. Cooperativity may be taken into account by considering $|\alpha_j^i| > 1$.

ii) Basal rates are constant for reactions without substrates and proportional to the concentration of the substrate otherwise.

iii) If activated reactions are present with intermittence, their non-basal rates are multiplied by discrete variables s_i.

As an example let us consider the minimal model proposed by Goldbeter for mitotic oscillations of the cell cycle [5]. Basically, this consists of three variables C (cyclin), M (cyclin dependent kinase complex) and X (proteolytic enzyme, most probably a polo-like kinase). The production of M is activated by C (also by M which is auto-catalytic), the production of X is activated by M and the degradation of C is activated by X. The hybrid interaction scheme contains six reactions. We decided that the degradation of the cyclin C acts discretely (on/off mechanism) and that all the other reactions are always present in the model (their rates are smoothly regulated). Then the hybrid model is the following:

$$\begin{aligned} \frac{dC}{dt} &= k_1 - \tilde{k}_1 C H(X - \tilde{h}_1) - \tilde{k}_1^0 C \\ \frac{dM}{dt} &= (k_2 MC + k_1^0) - \tilde{k}_2^0 M \\ \frac{dX}{dt} &= (k_3 M + k_2^0) - \tilde{k}_3^0 X \end{aligned} \tag{3.2}$$

where H is the Heaviside unit step function.

4 Reverse engineering of hybrid models

We would like to develop a method allowing to find the parameters of a model from the class introduced above that best describes the observed trajectories of a biological system. These trajectories can come from experiments or can be produced by non-hybrid models. In both situations we obtain a model whose parameters can be easily interpreted in biological terms. The hybrid model can be further analyzed or used to model more complex situations.

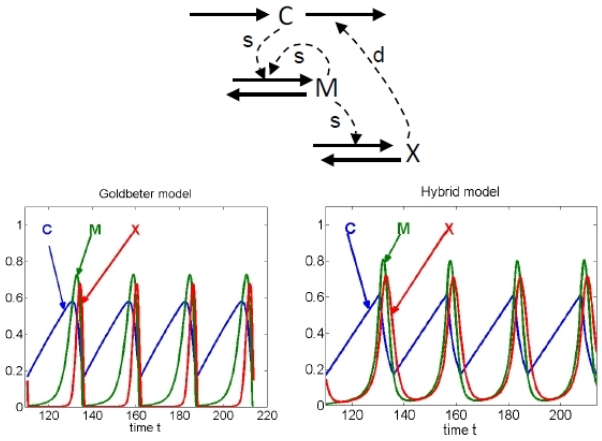

Fig. 4.1. (Middle) Regulated reaction graph for the minimal cell cycle model. Continuous arrows represent reactions, dotted arrows represent regulations. (s) regulations smoothly modulate the rates. (d) regulations discretely turn on and off the reaction rates. (Left) Trajectories of the non-hybrid model by Golbeter [5]. (Right) Trajectories of the hybrid model.

In the following we present a reverse engineering algorithm that works well for systems with sharp transitions.

Data. n trajectories (time series) $u_1(t), ..., u_n(t)$ given at time moments $t_0, t_1, ..., t_N$. A regulated reaction graph (the smooth/discrete partition of the regulations can be unspecified).

Output. A model of the type (2.2),(2.3) with values of the parameters that fit well the data.

The algorithm has several steps.

I. Splitting of the trajectory into smooth parts.

We look for K time intervals $I_1, I_2, ..., I_K$. The dynamics on each of the intervals is smooth, it is given by (2.2) with the s variables fixed. Mode transitions (change of the variables) occur at the borders of these intervals. We denote the switching times as $\tau_1, ...\tau_K$.

Finding τ_k is a problem of singularity detection. This could be done by various methods, for example by wavelet analysis. We have chosen as criterium the value of the second derivative of u_i. For piecewise smooth systems, the derivatives of the trajectories are discontinuous at the switching times τ_k. The second derivative has delta-Dirac components located at τ_k, which will show up as peaks in the numerically estimated second derivatives.

II. Identification of the mode transitions.

Given a switching time τ_k the mode transition is defined by the set of values values σ_j indicating reactions to be turned on or off at τ_k. The presence of a discontinuity is indicated by a peak in the second derivative of one or several species u_i. Without knowing which reaction in the regulated reaction graph has discrete behavior, there are several possible choices for such reactions. Each one of this choices could lead to a different hybrid model corresponding to a different characterization of the interactions as discrete and continuous. This step is supervised and could take into account biologist's intuition.

The discontinuities of the trajectories give the transitions but not the first mode. This choice is also supervised and takes into account periodicity constraints. From the first mode and from the transitions, all the modes (values of σ_j on the intervals I_k) are straightforwardly obtained.

III. Determining the mode internal parameters.

The mode internal parameters are obtained by simulating annealing. Let $u_i^{modes}(t)$ be the continuous hybrid trajectories obtained by integrating the modes between the calculated transition times. The simulated annealing algorithm minimizes the following objective function:

$$F = \sum_{i,k} C_k (u_i^{modes}(t_k) - u_i(t_k))^2$$

C_k are positive weights that increase with time. We thus penalize large time deviations that can arise from period misfit.

IV. Determining the mode control parameters.

Let $\sigma_m = H(\sum_{(m,j)\in E^r} w_{mj} u_j - h_j)$ be the discrete variables determined above. Let σ_k^m be the constant values of σ_m on T_k. Consider now the optimal trajectories $u_i^{modes*}(t_l)$ obtained before.

Then, one should have

$$(\sum_{(m,j)\in E^r} w_{mj} u_j^{modes*}(t_l) - h_j)\sigma_k^m > 0, \text{for all } t_l \in T_k \tag{4.1}$$

which is a linear programming problem for w_{mj} that can be resolved (if it has a solution) in polynomial time.

The algorithm has been applied to the minimal cell cycle model by Golbeter and the result is shown in Fig. 4.1. Of course the fit is not perfect and one should by no means expect a perfect fit. One of the reason of the differences is that the model by Goldbeter uses degradation terms that saturate and are practically constant on the descending slope of the variables M, X, while our linear degradation terms lead to exponential decrease.

5 Conclusion

The results that we present are a proof of principle that piecewise smooth hybrid models can be constructed with a simple heuristic from basic information about biochemical interactions. Using this class of hybrid models instead of piecewise-linear approximations provides, in many situations, a better balance between discrete and smooth interactions. For instance, the hybrid cell cycle model presented here has only two discrete transitions per period and it is very robust. A piecewise-linear version of the same model, would need a lot more discrete transitions per period which will reduce robustness and increase the difficulty of the inversion procedure. In the future we will apply the heuristic and the fitting algorithm to obtain a realistic model for the eucaryotic cell cycle.

References

[1] H. De Jong, J.L. Gouzé, C. Hernandez, M. Page, T. Sari, and J. Geiselmann. Qualitative simulation of genetic regulatory networks using piecewise-linear models. *Bulletin of Mathematical Biology*, 66(2):301–340, 2004.

[2] A.F. Filippov and FM Arscott. *Differential equations with discontinuous righthand sides*. Springer, 1988.

[3] J. Gebert, N. Radde, and G.W. Weber. Modeling gene regulatory networks with piecewise linear differential equations. *European Journal of Operational Research*, 181(3):1148–1165, 2007.

[4] L. Glass and S.A. Kauffman. The logical analysis of continuous, non-linear biochemical control networks. *Journal of Theoretical Biology*, 39(1):103–129, 1973.

[5] A. Goldbeter. A minimal cascade model for the mitotic oscillator involving cyclin and cdc2 kinase. *Proceedings of the National Academy of Sciences of the United States of America*, 88(20):9107, 1991.

[6] A.N. Gorban and I.V. Karlin. Method of invariant manifold for chemical kinetics. *Chemical Engineering Science*, 58(21):4751–4768, 2003.

[7] A.N. Gorban and I.V. Karlin. *Invariant manifolds for physical and chemical kinetics, Lect. Notes Phys. 660*. Springer Verlag, Berlin, Heidelberg, 2005.

[8] AN Gorban and O. Radulescu. Dynamic and static limitation in reaction networks, revisited . In David West Guy B. Marin and Gregory S. Yablonsky, editors, *Advances in Chemical Engineering - Mathematics in Chemical Kinetics and Engineering*, volume 34 of *Advances in Chemical Engineering*, pages 103–173. Elsevier, 2008.

[9] AN Gorban, O. Radulescu, and AY Zinovyev. Asymptotology of chemical reaction networks. *Chemical Engineering Science*, 65:2310–2324, 2010.

[10] S.A. Kauffman. *The origins of order: Self organization and selection in evolution*. Oxford University Press, USA, 1993.

[11] A.S. Matveev and A.V. Savkin. *Qualitative theory of hybrid dynamical systems*. Birkhauser, 2000.

[12] O. Radulescu, A.N. Gorban, A. Zinovyev, and A. Lilienbaum. Robust simplifications of multiscale biochemical networks. *BMC systems biology*, 2(1):86, 2008.

[13] M.A. Savageau and E.O. Voit. Recasting nonlinear differential equations as S-systems: a canonical nonlinear form. *Mathematical biosciences*, 87(1):83–115, 1987.

[14] R. Shorten, F. Wirth, O. Mason, K. Wulff, and C. King. Stability Criteria for Switched and Hybrid Systems. *SIAM Review*, 49(4):545–592, 2007.

[15] L. Tavernini. Differential automata and their discrete simulators. *Nonlin. Anal. Theory Methods Applic.*, 11(6):665–683, 1987.