Symbolic Reachability Analysis of Genetic Regulatory Networks using Qualitative Abstractions

Grégory Batt — Delphine Ropers — Hidde de Jong — Michel Page — Johannes Geiselmann

R apport de recherche

INRIA
RHÔNE-ALPES

Symbolic Reachability Analysis of Genetic Regulatory Networks using Qualitative Abstractions[*]

Grégory Batt[†‡] , Delphine Ropers[†] , Hidde de Jong[†] , Michel Page[†§] , Johannes Geiselmann[¶]

Abstract: The switch-like character of gene regulation has motivated the use of hybrid, discrete-continuous models of genetic regulatory networks. While powerful techniques for the analysis, verification, and control of hybrid systems have been developed, the specificities of the biological application domain pose a number of challenges, notably the absence of quantitative information on parameter values and the size and complexity of networks of biological interest. We introduce a method for the analysis of reachability properties of genetic regulatory networks that is based on a class of discontinuous piecewise-affine (PA) differential equations well-adapted to the above constraints. More specifically, we introduce a hyperrectangular partition of the state space that forms the basis for a discrete abstraction preserving the sign of the derivatives of the state variables. The resulting discrete transition system provides a conservative approximation of the qualitative dynamics of the network and can be efficiently computed in a symbolic manner from inequality constraints on the parameters. The method has been implemented in the computer tool Genetic Network Analyzer (GNA), which has been applied to the analysis of a regulatory system whose functioning is not well-understood by biologists, the nutritional stress response in the bacterium *Escherichia coli*.

Key-words: Piecewise-affine differential equations, qualitative analysis, discrete abstraction, hybrid systems, genetic regulatory networks, systems biology, nutritional stress response, *Escherichia coli*

Analyse symbolique d'atteignabilité de réseaux de régulation génique par abstractions qualitatives

Résumé : Le fonctionnement binaire *allum/teint* des gnes a motiv l'utilisation de modles hybrides pour reprsenter les rseaux de rgulations gniques. Bien que des techniques efficaces aient t dveloppes pour l'analyse, la vrification et le contrle des systmes hybrides, les spcificits des systmes biologiques considrs, en particulier l'absence d'informations quantitatives sur les valeurs des paramtres et la taille et la complexit des rseaux d'intrt biologique, posent un certain nombre de problmes nouveaux. Nous proposons une mthode pour l'analyse des proprits d'atteignabilit des rseaux de rgulations gniques base sur une classe de modles bien adapts aux contraintes mentionnes ci-dessus, utilisant des quations diffrentielles discontinues et affines par morceaux. Plus prcisment, nous introduisons une partition hyperrectangulaire de l'espace d'tat, servant dfinir une abstraction discrte qui prserve les signes des drives des variables d'tat. Le systme de transitions discret ainsi dfini est une approximation conservative de la dynamique du rseau et peut tre calcul symboliquement de faon efficace partir de contraintes d'ingalit sur les paramtres. Cette mthode a t implmente dans l'outil Genetic Network Analyzer (GNA), et utilise pour l'analyse d'un systme biologique encore mal compris, celui de la rponse au stress nutritionnel chez la bactrie *Escherichia coli*.

Mots-clés : Equations diffrentielles affines par morceaux, analyse qualitative, abstraction discrte, systmes hybrides, rseaux de rgulations gniques, biologie des systmes, rponse au stress nutritionnel, *Escherichia coli*

1   Introduction

The functioning and development of living organisms is controlled on the molecular level by networks of genes, proteins, small molecules, and their mutual interactions, so-called *genetic regulatory networks*. The dynamics of these networks is hybrid in nature, in the sense that the continuous evolution of the concentration of proteins and other molecules is punctuated by discrete changes in the activity of genes coding for the proteins. The switch-like character of the dynamics of genetic regulatory networks has attracted much attention from researchers on hybrid systems (*e.g.*, [1, 11, 21, 28, 36]).

While powerful techniques for the analysis, verification, and control of hybrid systems have been developed (see [5, 41] for reviews), the specificities of the biological application domain pose a number of challenges [19, 46]. First, most genetic regulatory networks of interest consist of a large number of genes that are involved in complex, interlocking feedback loops. Second, the data available on both the structure and the dynamics of the networks is currently essentially qualitative in nature, meaning that numerical values for concentration variables and kinetic parameters describing the interactions are generally absent. The above characteristics require hybrid-system methods and tools to be upscalable and capable of dealing with qualitative information.

In this paper, we will show that a class of piecewise-affine differential equation (PADE) models, originally introduced by Glass and Kauffman in the seventies [30], is particularly suitable for dealing with the above challenges. The properties of these PADE models have been well-studied in mathematical biology [15, 18, 22, 24, 25, 26, 29, 31, 32, 42, 43, 45]. The variables in the PADE models are the concentrations of the proteins encoded by the genes, while step functions account for the discrete changes in gene activity occasioned by the regulatory interactions. On a formal level, the PADE models are related to a class of asynchronous logical models proposed by Thomas and colleagues [47, 48]. PADE models and their logical relatives have been used for the study of a number of prokaryotic and eukaryotic regulatory networks (see [23] for a review and references).

The particular form of the PADE models allows the continuous state space to be partitioned into hyperrectangular regions with equivalent qualitative dynamics, preserving the sign pattern of the time derivatives of the solutions. We exploit this partition by associating to each PADE model a continuous transition system having equivalent reachability properties. By means of a *discrete* or *qualitative abstraction* [2, 3, 4, 16, 34, 49], the continuous transition system is turned into a discrete transition system providing a compact and qualitative description of the dynamics of the continuous system. Formally, there exists a simulation relation between the continuous and discrete transition systems, that is, the latter is a conservative approximation of the former. We show that the discrete transition system is invariant over large ranges of parameter values and can be computed in a symbolic manner from inequality constraints. Moreover, it is possible to design tailored algorithms for computing the discrete transition system, which scale up to large and complex genetic regulatory networks.

The paper continues our previous work on the qualitative analysis of the dynamics of genetic regulatory networks by means of PADE models [21, 22]. The main novelty of this

contribution, a preliminary version of which was presented in [10], is the use of a more fine-grained discrete abstraction, preserving the derivative sign pattern. This is an important feature for the experimental validation of models of genetic regulatory networks, since measurements of gene expression often result in observations of changes in the sign of derivatives. Notwithstanding this increase in precision, the refinement does not threaten the applicability of the approach to large and complex networks. This is illustrated by the analysis of a network that is not yet well understood by biologists, the network controlling the carbon starvation response of the bacterium *Escherichia coli*. The application of the method has led to novel insights into how the adaptation of cell growth to nutritional stress emerges from the network of molecular interactions.

The paper is organized as follows. In Sections 2 and 3 we specify the PADE models of genetic regulatory networks in detail and discuss their mathematical properties, paying special attention to complications arising from the discontinuities in the righthand-side of the differential equations. In Section 4 we define a qualitative abstraction of the dynamics of PA systems, based on a hyperrectangular partition of the state space. Section 5 introduces rules to actually compute the discrete transition system induced by the qualitative abstraction and the implementation of the rules in a computer tool called GNA. In Section 6 we illustrate the application of the method to the analysis of the *E. coli* network. In the final section we present our conclusions and discuss the results in the context of related work.

2   PADE models of genetic regulatory networks

The dynamics of genetic regulatory networks can be described by a class of piecewise-affine differential equations (PADE) models of the following general form [30, 42]:

$$\dot{x} = h(x) = f(x) - g(x)\, x, \quad x \in \Omega \setminus \Theta, \tag{1}$$

where $x = (x_1, \ldots, x_n)'$ is a vector of cellular protein concentrations, $f = (f_1, \ldots, f_n)'$, $g = \mathrm{diag}(g_1, \ldots, g_n)$, $\Omega \subset \mathbb{R}^n_{\geq 0}$ is a bounded n-dimensional state space region, and Θ a zero-measure subset of Ω (see below). The rate of change of each protein concentration x_i, $i \in \{1, \ldots, n\}$, is thus defined as the difference of the rate of synthesis $f_i(x)$ and the rate of degradation $g_i(x)\, x_i$ of the protein.

The function $f_i : \Omega \setminus \Theta \to \mathbb{R}_{\geq 0}$ expresses how the rate of synthesis of the protein encoded by gene i depends on the concentrations x of the proteins in the cell. More specifically, the function f_i is defined as

$$f_i(x) = \sum_{l \in L_i} \kappa_i^l\, b_i^l(x), \tag{2}$$

where $\kappa_i^l > 0$ is a rate parameter, $b_i^l : \Omega \setminus \Theta \to \{0, 1\}$ a piecewise-constant *regulation function*, and L_i a possibly empty set of indices of regulation functions. The function g_i expresses the regulation of protein degradation. It is defined analogously to f_i, except that we demand that g_i is strictly positive. In addition, in order to formally distinguish degradation rate

parameters from synthesis rate parameters, we will denote the former by γ instead of κ. Notice that with the above definitions, h is a *piecewise-affine (PA)* vector-valued function.

A regulation function b_i^l describes the conditions under which the protein encoded by gene i is synthesized (degraded) at a rate κ_i^l ($\gamma_i^l x_i$). It is defined in terms of step functions and is the arithmetic equivalent of a Boolean function expressing the logic of gene regulation [30, 47]. More precisely, the conditions for synthesis and degradation are expressed using the step functions s^+, s^-:

$$s^+(x_j, \theta_j) = \begin{cases} 1, & \text{if } x_j > \theta_j, \\ 0, & \text{if } x_j < \theta_j, \end{cases} \quad s^-(x_j, \theta_j) = 1 - s^+(x_j, \theta_j), \tag{3}$$

where x_j is an element of the state vector x and θ_j a constant denoting a threshold concentration. Notice that the step functions are not defined at the thresholds.

Figure 1(a) gives an example of a simple genetic regulatory network consisting of two genes, a and b. When a gene (a or b) is expressed, the corresponding protein (A or B) is synthesized at a specified rate (κ_a or κ_b). Proteins A and B regulate the expression of genes a and b. More specifically, protein B inhibits the expression of gene a, above a certain threshold concentration θ_b, while protein A inhibits the expression of gene b above a threshold concentration θ_a^1, and the expression of its own gene above a second, higher threshold concentration θ_a^2. The degradation of the proteins is not regulated and proportional to the concentration of the proteins (with degradation parameters γ_a or γ_b). The PADE model of this network is shown in Figure 1(b).

$$\dot{x}_a = \kappa_a \, s^-(x_a, \theta_a^2) \, s^-(x_b, \theta_b) - \gamma_a \, x_a,$$
$$\dot{x}_b = \kappa_b \, s^-(x_a, \theta_a^1) - \gamma_b \, x_b.$$

(a)   (b)
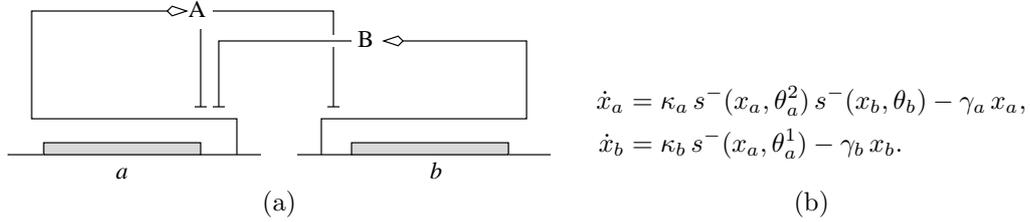
Figure 1: (a) Example of a genetic regulatory network of two genes (a and b), each coding for a regulatory protein (A and B). See Figure 6 for the legend. (b) PADE model corresponding to the network in (a).

The use of step functions $s^\pm(x_j, \theta_j)$ in (1) gives rise to mathematical complications, because the step functions are undefined and discontinuous at $x_j = \theta_j$. Therefore, h is undefined and may be discontinuous on the threshold hyperplanes $\Theta = \bigcup_{i \in \{1,\dots,n\}, l_i \in \{1,\dots,p_i\}} \{x \in \Omega \mid x_i = \theta_i^{l_i}\}$, where p_i denotes the number of thresholds for the protein encoded by gene i. In order to deal with this problem, we can follow an approach originally proposed by Filippov [27] and widely used in control theory. It consists in extending the differential *equation* $\dot{x} = h(x)$, $x \in \Omega \setminus \Theta$, to the differential *inclusion*

$$\dot{x} \in K(x), \text{ with } K(x) = \overline{co}(\{\lim_{y \to x, \, y \notin \Theta} h(y)\}), \ x \in \Omega, \tag{4}$$

where $\overline{co}(P)$ denotes the smallest closed convex set containing the set P, and $\{\lim_{y \to x, \, y \notin \Theta} h(y)\}$ the set of all limit values of $h(y)$, for $y \notin \Theta$ and $y \to x$. This approach has been applied in the context of genetic regulatory network modeling by Gouzé and Sari [32].

In practice, $K(x)$ may be difficult to compute because the smallest closed convex set can be a complex polyhedron in Ω. We therefore employ an alternative extension of the differential equation:

$$\dot{x} \in H(x), \text{ with } H(x) = \overline{rect}(\{\lim_{y \to x, \, y \notin \Theta} h(y)\}), \, x \in \Omega, \tag{5}$$

where $\overline{rect}(P)$ denotes the smallest closed *hyperrectangular* set containing the set P. The advantage of using $\overline{rect}$ is that we can rewrite $H(x)$ as a system of differential inclusions $\dot{x}_i \in H_i(x)$, $i \in \{1, \ldots, n\}$. Notice that $H(x)$ is an overapproximation of $K(x)$, for all $x \in \Omega$.

Formally, we define the *PA system* Σ as the triple (Ω, Θ, H), that is, the set-valued function H given by (5), and defined on the n-dimensional state space Ω, with Θ the union of the threshold hyperplanes. A *solution* of the PA system Σ on a time interval I is a solution of the differential inclusion (5) on I, that is, an absolutely-continuous vector-valued function $\xi(t)$ such that $\dot{\xi}(t) \in H(\xi(t))$ almost everywhere on I. In particular, the derivative of $\xi(t)$ may not exist, and therefore $\dot{\xi}(t) \in H(\xi(t))$ may not hold, if ξ reaches or leaves Θ at t.

For all $x_0 \in \Omega$ and $\tau \in \mathbb{R}_{>0} \cup \{\infty\}$, $\Xi_\Sigma(x_0, \tau)$ denotes the set of solutions $\xi(t)$ of the PA system Σ, for the initial condition $\xi(0) = x_0$, and $t \in [0, \tau]$. The existence of at least one solution ξ on some time interval $[0, \tau]$, $\tau > 0$, with initial condition $\xi(0) = x_0$ is guaranteed for all x_0 in Ω [27]. However, there is, in general, not a unique solution. The set $\Xi_\Sigma = \bigcup_{x_0 \in \Omega, \tau > 0} \Xi_\Sigma(x_0, \tau)$ is the set of all solutions, on a finite or infinite time interval, of the PA system Σ. We restrict our analysis to the solutions in Ξ_Σ that reach and leave a threshold hyperplane finitely-many times. The dynamics of Σ is thus defined by the set of solutions Ξ_Σ.

3   Mathematical analysis of PA systems

3.1   Mode domains

The dynamical properties of the solutions of Σ can be analyzed in the n-dimensional state space hyperrectangle $\Omega = \Omega_1 \times \ldots \times \Omega_n$, where $\Omega_i = \{x_i \in \mathbb{R} \mid 0 \le x_i \le max_i\}$ and max_i denotes a maximum concentration for each protein, $i \in \{1, \ldots, n\}$. In particular, we set $max_i > \max_{x \in \Omega \setminus \Theta} f_i(x)/g_i(x)$. It is not difficult to show that under this condition Ω is a positively invariant set.

For subsequent use, we introduce the notion of *hyperrectangular partition* of a set $R = R_1 \times \ldots \times R_n \subseteq \Omega$. The partition is induced by sets of hyperplanes orthogonal to one of the axes x_i, $i \in \{1, \ldots, n\}$. More precisely, the hyperplanes orthogonal to the x_i-axis are given by the finite sets $\Lambda_i \subset \Omega_i$, where every $\lambda \in \Lambda_i$ corresponds to a hyperplane

$\{x \in \Omega \mid x_i = \lambda\}$. The hyperrectangular partition of R induced by $\Lambda = \{\Lambda_1, \ldots, \Lambda_n\}$ is defined as $\mathcal{P} = \mathcal{P}_1 \times \ldots \times \mathcal{P}_n$, where $\mathcal{P}_i$, $i \in \{1, \ldots, n\}$, is the *interval partition* of R_i induced by Λ_i. That is, $\mathcal{P}_i$ is the partition of minimal cardinality of R_i, such that for every $P \in \mathcal{P}_i$, either $P = \{\lambda\}$ for some $\lambda \in \Lambda_i$, or P is an interval containing no $\lambda \in \Lambda_i$. As an example, consider the interval $R_i = (r_1, r_2]$ and $\Lambda_i = \{\lambda_1, \lambda_2\}$, with $r_1 < \lambda_1 < r_2 < \lambda_2$. The interval partition of R_i induced by Λ_i equals $\mathcal{P}_i = \{(r_1, \lambda_1), \{\lambda_1\}, (\lambda_1, r_2]\}$.

The $(n-1)$-dimensional threshold hyperplanes introduced in Section 2 define a hyperrectangular partition of Ω.

Definition 1 (Mode domain partition). $\mathcal{M}$ is the hyperrectangular partition of Ω induced by $\{\theta_i^1, \ldots, \theta_i^{p_i}\}$, $i \in \{1, \ldots, n\}$. The sets $M \in \mathcal{M}$ are called *mode domains*.

The reason for speaking of mode domains will become clearer below, when we show that the regulation of gene expression is identical in every $M \in \mathcal{M}$, and thus corresponds to a regulatory mode of the system.
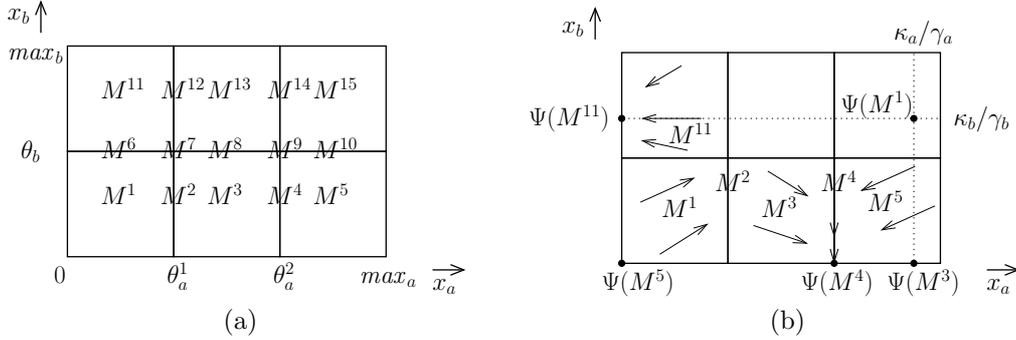

Figure 2:   (a) Mode domain partition of the state space for the model of Figure 1(b). (b) Focal sets and vector fields associated with the mode domains M^1 to M^5, and M^{11}.

Figure 2(a) shows the mode domain partition of the two-dimensional state space of the example network. We distinguish between mode domains like M^2 and M^7, which are located on (intersections of) threshold hyperplanes, and mode domains like M^1, which are not. The former domains are called *singular* mode domains and the latter *regular* mode domains. We denote by $\mathcal{M}_r$ and $\mathcal{M}_s$ the sets of regular and singular mode domains, respectively. Note that $\mathcal{M} = \mathcal{M}_r \cup \mathcal{M}_s$.

We next introduce some basic topological concepts. For every hyperrectangular set, $R \subseteq \Omega$, of dimension k, $k \in \{0, \ldots, n\}$, we define the supporting hyperplane of R, $supp(R) \subseteq \Omega$, as the k-dimensional hyperplane containing R. The *boundary* of R in $supp(R)$, denoted by ∂R, is defined as the set $\overline{R} \setminus \mathring{R}$, where $\overline{R}$ is the closure of R and $\mathring{R}$ its interior, both relative to $supp(R)$. Suppose that M is a singular mode domain, *i.e.* $M \in \mathcal{M}_s$. Then $R(M)$ is defined as the set of regular mode domains M' having M in their boundary, *i.e.* $R(M) = \{M' \in \mathcal{M}_r \mid M \subseteq \partial M'\}$.

Using the definition of the differential inclusion (5), it can be easily shown that in a regular mode domain M,

$$H(x) = \{\mu^M - \nu^M x\}, \ x \in M, \tag{6}$$

where μ^M is a vector of (sums of) positive synthesis rate constants and $\nu^M = \mathrm{diag}(\nu_1^M, \ldots, \nu_n^M)$ a diagonal matrix of (sums of) positive degradation rate constants [22]. We define the *focal set*

$$\Psi(M) = \{\psi(M)\}, \tag{7}$$

where $\psi(M) = (\nu^M)^{-1} \mu^M$, and repeat the following standard result [30].

Lemma 1. Let $M \in \mathcal{M}_r$. Every solution $\xi(t) \in \Xi_\Sigma$ in M monotonically converges towards $\Psi(M)$.

Proof. From (6) it follows that the solutions in a regular mode domain M are given by $\xi(t) = \psi(M) - e^{-\nu^M t}(\psi(M) - x_0)$, for initial conditions $x_0 \in M$. As a consequence, $\xi(t)$ monotonically converges towards $\Psi(M) = \{\psi(M)\}$ as long as $\xi(t) \in M$. □

We will make the generic assumption that the focal sets $\Psi(M)$, for all $M \in \mathcal{M}_r$, are not located in the threshold hyperplanes Θ. Figure 2(b) shows the focal sets of four regular mode domains (M^1, M^3, M^5 and M^{11}). In the case of M^{11}, we see that $\Psi(M^{11}) \subseteq M^{11}$, so that $\psi(M^{11})$ is an asymptotically stable equilibrium point of Σ.

In a singular mode domain, the right-hand side of the differential inclusion (5) reduces to the following set:

$$H(x) = \overline{rect}(\{\mu^{M'} - \nu^{M'} x \mid M' \in R(M)\}), \ x \in M. \tag{8}$$

The *focal set* associated with the domain is now defined as

$$\Psi(M) = supp(M) \cap \overline{rect}(\{\psi(M') \mid M' \in R(M)\}), \tag{9}$$

which is generally not a single point in higher-dimensional systems. As will be shown below, two different cases can be distinguished. If $\Psi(M)$ is empty, then every solution passes through M instantaneously (Lemma 2). If not, then some (but not necessarily all) solutions arriving at M will remain in M for some time, sliding along the threshold hyperplanes containing M (Lemma 3). In the latter case, weaker monotonicity properties hold (Lemmas 3 and 4).

Lemma 2. Let $M \in \mathcal{M}_s$. Every solution $\xi(t) \in \Xi_\Sigma$ arriving at M instantaneously crosses the domain if and only if $\Psi(M) = \{\}$.

Proof. We first prove sufficiency. From the definition of $\Psi(M)$ in a singular mode domain and $\Psi(M) = \{\}$, it follows that for some $i \in \{1, \ldots, n\}$ such that M_i is included in a threshold hyperplane, the intersection of $supp_i(M) = \{\theta_i^{l_i}\}$, $l_i \in \{1, \ldots, p_i\}$, and $\overline{rect}(\{\psi_i(M') \mid M' \in R(M)\})$ is empty. As a consequence, either $\psi_i(M') > \theta_i^{l_i}$ for all

$M' \in R(M)$, or $\psi_i(M') < \theta_i^{l_i}$ for all $M' \in R(M)$. From (8) and the hyperrectangular shape of $H(x)$, we obtain $H_i(x) = \overline{rect}(\{\nu_i^{M'} \ (\psi_i(M') - \theta_i^{l_i}) \mid M' \in R(M)\})$, for all $x \in M$. As a consequence, either $\min H_i(x) > 0$ or $\max H_i(x) < 0$, so that $0 \notin H_i(x)$ and no solution with $\dot{\xi}_i(t) = 0$ exists in M. Therefore, every solution arriving at M instantaneously crosses the domain.

Necessity is proven by contraposition, by an argument similar to that given for sufficiency. From $\Psi(M) \neq \{\}$ we show that $0 \in H_i(x)$, for every $x \in M$ and $i \in \{1, \ldots, n\}$ such that M_i is included in a threshold hyperplane. We can then define solutions $\xi(t)$ that satisfy $\dot{\xi}_i(t) = 0$ on some time interval, that is, solutions that do not instantaneously cross M. $\quad\square$

Lemma 3. Let $M \in \mathcal{M}_s$ and $\Psi(M) \neq \{\}$. For every solution $\xi(t) \in \Xi_\Sigma$ in M, and every $i \in \{1, \ldots, n\}$ such that M_i is included in a threshold hyperplane, it holds that $\dot{\xi}_i(t) = 0$. For all other i, $\xi_i(t)$ monotonically converges towards $\Psi_i(M)$.

Proof. In order to remain in M, $\xi(t)$ must slide along the threshold hyperplanes containing M. Therefore, $\dot{\xi}_i(t) = 0$ for every $x \in M$ and $i \in \{1, \ldots, n\}$ such that M_i is included in a threshold hyperplane. For all other i, we must have $\dot{\xi}_i(t) \in H_i(\xi(t))$ almost everywhere, by the definition of solutions of PA systems. Moreover, $H_i(\xi(t)) = \overline{rect}(\{\nu_i^{M'} \ (\psi_i(M') - \xi_i(t)) \mid M' \in R(M)\})$. Let $\xi_i(t) \notin \Psi_i(M)$, and $\xi_i(t) < \psi_i(M')$ for all $M' \in R(M)$ (the case $\xi_i(t) > \psi_i(M')$ goes analogously). It follows that $\min H_i(\xi(t)) > 0$, and hence that $\dot{\xi}_i(t) > 0$, provided that $\dot{\xi}(t) \in H(\xi(t))$. We thus infer that $\xi_i(t)$ monotonically converges towards $\Psi_i(M)$. $\quad\square$

Lemma 4. Let $M \in \mathcal{M}_s$ and $\Psi(M) \neq \{\}$. For every $\psi \in \Psi(M)$ and $x \in M$, there exists a solution $\xi(t) \in \Xi_\Sigma$ in M, with $\xi(0) = x$, such that $\xi(t)$ monotonically converges towards ψ.

Proof. Let $M^+(i) = \arg\max_{M' \in R(M)} \ \psi_i(M')$ and $M^-(i) = \arg\min_{M' \in R(M)} \ \psi_i(M')$, for all $i \in \{1, \ldots, n\}$. Notice that due to the hyperrectangular shape of $\Psi(M)$, we can write $\psi_i = \alpha_i \ \psi_i(M^-(i)) + (1 - \alpha_i) \ \psi_i(M^+(i))$, for some $\alpha_i \in [0, 1]$. Moreover, after defining $\beta_i \in [0, 1]$ such that $\beta_i = \alpha_i \ \nu_i^{M^+(i)} / (\alpha_i \ \nu_i^{M^+(i)} + (1 - \alpha_i) \ \nu_i^{M^-(i)})$, we introduce μ_i and ν_i such that

$$\psi_i = \frac{\mu_i}{\nu_i} = \frac{\beta_i \ \mu_i^{M^-(i)} + (1 - \beta_i) \ \mu_i^{M^+(i)}}{\beta_i \ \nu_i^{M^-(i)} + (1 - \beta_i) \ \nu_i^{M^+(i)}}. \tag{10}$$

That is, ψ_i is written as the quotient of μ_i and ν_i, which in turn can be defined as a convex combination of $\mu_i^{M^-(i)}$, $\mu_i^{M^+(i)}$ and $\nu_i^{M^-(i)}$, $\nu_i^{M^+(i)}$, respectively.

Now, construct a function $\xi(t) = \psi - e^{-\nu t} \ (\psi - x)$. This yields $\dot{\xi}_i(t) = \nu_i \ e^{-\nu_i t} \ (\psi_i - x_i) = \nu_i \ (\psi_i - \xi_i(t)) = \mu_i - \nu_i \ \xi_i(t)$, for every $i \in \{1, \ldots, n\}$. Following the definition of μ_i and ν_i in (10), this gives

$$\dot{\xi}_i(t) \ = \ \beta_i \ \ (\mu_i^{M^-(i)} - \nu_i^{M^-(i)} \xi_i(t)) \ + \ (1 - \beta_i) \ \ (\mu_i^{M^+(i)} - \nu_i^{M^+(i)} \xi_i(t)). \tag{11}$$

For $\xi(t)$ to be a solution, it must satisfy $\dot{\xi}_i(t) \in H_i(\xi(t))$ almost everywhere, for every $i \in \{1, \ldots, n\}$. It can be directly verified that the expression for $\dot{\xi}_i(t)$ in (11), with $\beta_i \in [0, 1]$,

is included in $H_i(\xi(t))$ as defined by (8). Hence, the condition is satisfied and $\xi(t)$ is a solution of Σ in M. Moreover, $\xi(0) = x$ and $\xi(t)$ monotonically converges towards ψ. □

In the sequel, domains which every solution crosses instantaneously will be called *instantaneous* domains, whereas domains in which at least some solutions remain for some time will be referred to as *persistent*. Figure 2(b) shows two singular mode domains, M^2 and M^4. M^2 is an instantaneous mode domain ($\Psi(M^2) = \emptyset$), whereas M^4 is a persistent mode domain ($\Psi(M^4) = \{(\theta_a^2, 0)'\}$) in which solutions slide along the threshold plane. In this simple example, it is intuitively clear how to define the flow in M^4, given the dynamics in M^3 and M^5. The use of differential inclusions as described above makes it possible to define the flow in singular domains in a systematic and mathematically proper way.

The fact that every mode domain is associated with a unique focal set has provided the basis for a discrete abstraction criterion employed in our previous work [15, 21, 22]. However, this criterion disregards the fact that the system does not always exhibit the same qualitative dynamics in different parts of a mode domain, in the sense that the sign pattern of the derivatives of the solutions may not be unique. Consider a solution $\xi(t) \in \Xi_\Sigma$ in M^{11} in Figure 2(b): depending on whether $\xi_b(t)$ is larger than, equal to, or smaller than the focal concentration κ_b/γ_b in M^{11}, ξ_b will be decreasing, steady, or increasing. As a consequence, if we abstract the domain M^{11} away into a single discrete state, we will not be able to unambiguously infer that solutions entering this domain from M^6 are increasing in the x_b-dimension. This may lead to problems when comparing predictions from the model with gene expression data, for instance the observed variation of the sign of $\dot{x}_b$. Today's experimental techniques, such as quantitative RT-PCR, reporter genes, and DNA microarrays, usually produce information on changes in the sign of the derivatives of the concentration variables.

3.2   Flow domains

The mismatch between the description levels of the mathematical analysis and the experimental data calls for a finer partitioning of the state space, which can then provide the basis for a more adequate abstraction criterion. Along these lines, the regular and singular mode domains distinguished above are repartitioned by means of the $(n-1)$-dimensional hyperplanes corresponding to the focal concentrations.

Definition 2 (Flow domain partition). $\mathcal{D}^M$ is the hyperrectangular partition of a mode domain $M \in \mathcal{M}$ induced by $\{\psi_i(M)\}$, if M is regular, and by $\{\psi_i(M') \mid M' \in R(M)\}$, if M is singular, $i \in \{1, \ldots, n\}$. $\mathcal{D} = \cup_{M \in \mathcal{M}} \mathcal{D}^M$ is a partition of Ω and the sets $D \in \mathcal{D}$ are called *flow domains*.

The reason for speaking about flow domains is that, as will be seen below, in every $D \in \mathcal{D}$ the flow of the system is qualitatively identical. The partitioning of the state space into 27 flow domains is illustrated for the example system in Figure 3(a). Every flow domain is included in a single mode domain, a relation captured by the surjective function *mode*: $\mathcal{D} \to \mathcal{M}$, defined as $mode(D) = M$, if and only if $D \subseteq M$. Similarly, the function *flow*:

$\Omega \to \mathcal{D}$ denotes the surjective mapping that associates a point in the state space to its flow domain: $flow(x) = D$, if and only if $x \in D$.
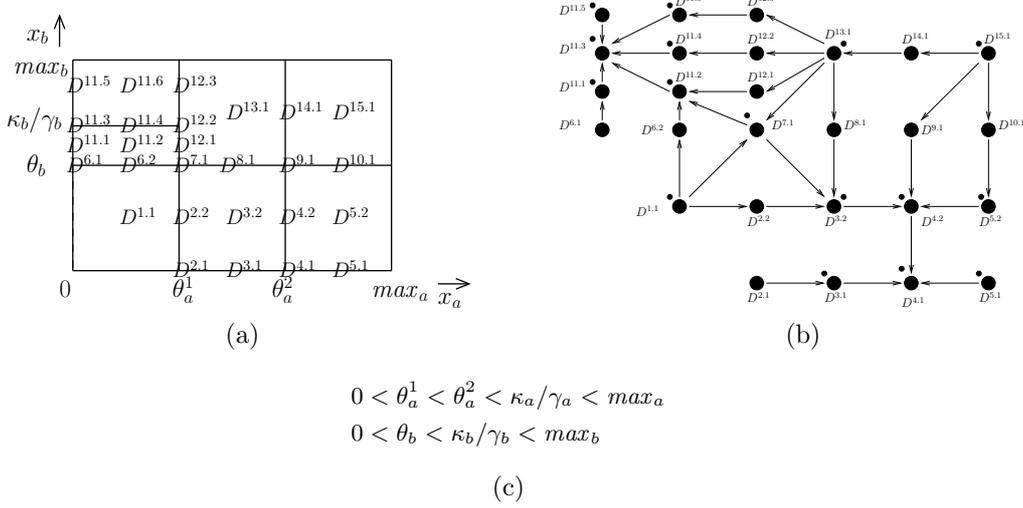

Figure 3: (a) Flow domain partition of the state space of the model in Figure 1(b). (b) State transition graph of the corresponding qualitative transition system. For the sake of clarity, self-transitions are represented by dots and transition labels are omitted. (c) Parameter inequality constraints for which the graph in (b) is invariant.

The repartitioning of mode domain M^{11} leads to six flow domains (Figure 3(a)). The finer partition guarantees that in every flow domain of M^{11}, the derivatives have a unique sign pattern. In $D^{11.2}$, for instance, the x_a-derivative is negative and the x_b-derivative is positive, whereas in $D^{11.3}$ both derivatives equal zero (in fact, $D^{11.3}$ coincides with $\psi(M^{11})$ and is an equilibrium point of the system). This property is true more generally. Consider a point x in a flow domain $D \in \mathcal{D}$. We denote by $S(x) \in 2^{\{-1,0,1\}^n}$ the derivative sign pattern at x, that is, the set of derivative sign vectors of the solutions in D passing through x. More formally,

$$S(x) = \{sign(\dot{\xi}(\tau)) \mid \xi(t) \in \Xi_\Sigma \text{ in } D, \text{ and } \exists \tau \geq 0 : \xi(\tau) = x \text{ and } \dot{\xi}(\tau) \in H(\xi(\tau))\}. \quad (12)$$

This gives the following central result.

Theorem 1 (Qualitatively-identical dynamics). For all $x, y \in D$ and $D \in \mathcal{D}$, $S(x) = S(y)$.

Proof. Note that due to the hyperrectangular shape of $H(x)$, $S(x)$ can be decomposed into $S_1(x) \times \ldots \times S_n(x)$, where $S_i(x)$ denotes the set of the ith components of the derivative sign vectors of the solutions in D passing through x, $i \in \{1, \ldots, n\}$ (idem $S(y)$). Define $M = mode(D)$ and distinguish the cases (a)-(c).

(a) $M \in \mathcal{M}_r$. Suppose $1 \in S_i(x)$, but $1 \notin S_i(y)$, for any $i \in \{1, \ldots, n\}$ (for 0 and -1 the argument is similar). By definition of $S_i(x)$ this means that $\dot{\xi}_i(\tau) > 0$ for some $\xi(t) \in \Xi_\Sigma$ in D and $\tau \geq 0$, such that $\xi(\tau) = x$. On the other hand, $\dot{\varphi}_i(\sigma) \leq 0$ for all $\varphi(t) \in \Xi_\Sigma$ in D and $\sigma \geq 0$, such that $\varphi(\sigma) = y$. Given that $\dot{\xi}_i(\tau) \in H_i(\xi(\tau)) = \{\nu_i^M \ (\psi_i(M) - \xi_i(\tau))\}$, due to the definition of $S(x)$, it follows that $\xi_i(\tau) = x_i < \psi_i(M)$. Similarly, $\varphi_i(\sigma) = y_i \geq \psi_i(M)$. But then, by Definition 2, x and y are not located in the same flow domain D, contrary to the assertion of the theorem. Therefore, $1 \in S_i(x)$ implies $1 \in S_i(y)$. The converse is shown in the same way.

(b) $M \in \mathcal{M}_s$ and $\Psi(M) = \{\}$. No solution remains in D, so by definition $S(x) = S(y) = \{\}$.

(c) $M \in \mathcal{M}_s$ and $\Psi(M) \neq \{\}$. For every $i \in \{1, \ldots, n\}$ such that D_i is located in a threshold or focal hyperplane, we must have $\dot{\xi}_i(t) = 0$ for solutions remaining in D. Consequently, $S_i(x) = \{0\}$ (idem $S_i(y) = \{0\}$). For all other i, suppose $1 \in S_i(x)$, but $1 \notin S_i(y)$ (the argument for 0 and -1 is similar). It follows that $\dot{\xi}_i(\tau) > 0$ for some $\xi(t) \in \Xi_\Sigma$ in D and $\tau \geq 0$, such that $\xi(\tau) = x$. However, $\dot{\varphi}_i(\sigma) \leq 0$ for all $\varphi(t) \in \Xi_\Sigma$ in D and $\sigma \geq 0$, such that $\varphi(\sigma) = y$. From $\dot{\xi}_i(\tau) \in H_i(\xi(\tau)) = \overline{rect}(\{\nu_i^{M'} \ (\psi_i(M') - \xi_i(\tau)) \mid M' \in R(M)\})$, we conclude that $\xi_i(\tau) = x_i < \psi_i(M')$, for some $M' \in R(M)$. Similarly, we must have $\dot{\varphi}_i(\sigma) \in H_i(\varphi(\sigma)) = \overline{rect}(\{\nu_i^{M'} \ (\psi_i(M') - \varphi_i(\sigma)) \mid M' \in R(M)\})$. Now, if there were some $M' \in R(M)$ such that $\varphi_i(\sigma) = y_i < \psi_i(M')$, then by Lemma 4 there would exist a solution such that $\dot{\varphi}_i(\sigma) > 0$. Since this contradicts the assumption, we conclude that $\varphi_i(\sigma) = y_i \geq \psi_i(M')$, for all $M' \in R(M)$. This implies, again by Definition 2, that x and y are not located in the same flow domain D. Therefore, $1 \in S_i(x)$ implies $1 \in S_i(y)$ (and conversely). $\qquad\qquad\square$

Notice that the definition of S as a set is a direct consequence of the use of differential inclusions. Since the solutions of differential inclusions are not unique, several solutions may pass through x and their derivatives may have a different sign in some dimension $i \in \{1, \ldots, n\}$. This situation does not occur in our two-gene example.

Theorem 1 suggests that the partition of the state space introduced in this section can be used as an abstraction criterion better-adapted to the available experimental data on gene expression. This idea will be further developed in the next section.

4   Qualitative abstraction of the dynamics of PA systems

4.1   Qualitative PA transition systems

As a preparatory step, we define a *continuous transition system* having the same reachability properties as the original PA system Σ. Consider $x \in D$ and $x' \in D'$, where $D, D' \in \mathcal{D}$ are flow domains. If there exists a solution $\xi(t)$ of Σ passing through x at time $\tau \in \mathbb{R}_{\geq 0}$ and reaching x' at time $\tau' \in \mathbb{R}_{>0} \cup \{\infty\}$, without leaving $D \cup D'$ in the time interval $[\tau, \tau']$, then the absolute continuity of $\xi(t)$ implies that D and D' are either equal or contiguous.

More precisely, one of the following three cases holds: $D = D'$, $D \subseteq \partial D'$, or $D' \subseteq \partial D$. We consequently distinguish three corresponding types of continuous transitions: *internal*, denoted by $x \xrightarrow{int} x'$, *dimension increasing*, denoted by $x \xrightarrow{dim^+} x'$, and *dimension decreasing*, denoted by $x \xrightarrow{dim^-} x'$. The latter two terms refer to the increase or decrease in dimension of the flow domain when going from D to D'. This leads to the following definition:

Definition 3 (PA transition system). $\Sigma\text{-TS} = (\Omega, L, \Pi, \rightarrow, \models)$ is the transition system corresponding to the PA system $\Sigma = (\Omega, \Theta, H)$, where:

1. Ω is the state space;

2. $L = \{int, dim^+, dim^-\}$ is a set of labels denoting the three different types of transitions;

3. $\Pi = \{Dsign = S \mid S \in 2^{\{-1,0,1\}^n}\}$ is a set of propositions describing the derivative sign pattern of the concentration variables;

4. $\rightarrow$ is the transition relation describing the continuous evolution of the system, defined by $\rightarrow \subseteq \Omega \times L \times \Omega$, such that $x \xrightarrow{l} x'$ if and only if there exists $\xi(t) \in \Xi_\Sigma$ and τ, τ', such that $0 \leq \tau < \tau'$, $\xi(\tau) = x$, $\xi(\tau') = x'$, and

   - if $l = int$, then for all $t \in [\tau, \tau']$: $\xi(t) \in \mathit{flow}(x) = \mathit{flow}(x')$,
   - if $l = dim^+$, then for all $t \in (\tau, \tau']$: $\xi(t) \in \mathit{flow}(x') \neq \mathit{flow}(x)$,
   - if $l = dim^-$, then for all $t \in [\tau, \tau')$: $\xi(t) \in \mathit{flow}(x) \neq \mathit{flow}(x')$;

5. $\models$ is the satisfaction relation of the propositions in Π, defined by $\models \subseteq \Omega \times \Pi$, such that $x \models Dsign = S$ if and only if $S = S(x)$.

The satisfaction relation $\models$ thus associates to each point x in the state space a qualitative description of the dynamics of the system at x. We define any sequence of points in Ω, $(x^0, \ldots, x^m)$, $m \geq 0$, as a *run* of Σ-TS if for all $j \in \{0, \ldots, m-1\}$, there exists some $l \in L$ such that $x^j \xrightarrow{l} x^{j+1}$. It is not difficult to show that a PA system Σ and its corresponding PA transition system Σ-TS have equivalent reachability properties (see the proof in Appendix A).

Proposition 1 (Reachability equivalence). For all $x, x' \in \Omega$, there exists a solution $\xi(t)$ of Σ and τ, τ', such that $0 \leq \tau \leq \tau'$, $\xi(\tau) = x$, and $\xi(\tau') = x'$ if and only if there exists a run $(x^0, \ldots, x^m)$ of Σ-TS such that $x^0 = x$ and $x^m = x'$.

The continuous PA transition system has an infinite number of states and transitions, which makes it difficult to verify dynamical properties of interest, for instance by means of conventional tools for model checking [17]. However, we can define a discrete transition system, with a finite number of states and transitions, which preserves important properties of the qualitative dynamics of the system. In order to achieve this, we introduce the equivalence relation $\sim_\Omega \subseteq \Omega \times \Omega$ induced by the partition $\mathcal{D}$ of the state space: $x \sim_\Omega x'$ if and only

if $flow(x) = flow(x')$. From Theorem 1 it follows that $\sim_\Omega$ is *proposition-preserving* [4, 16], in the sense that for all $x, x' \in D$ and for all $\pi \in \Pi$, $x \models \pi$ if and only if $x' \models \pi$.

The discrete or *qualitative abstraction* of a PA transition system Σ-TS, called *qualitative PA transition system*, is now defined as the quotient transition system of Σ-TS, given the equivalence relation $\sim_\Omega$ [4, 16].

Definition 4 (Qualitative PA transition system). The qualitative PA transition system corresponding to the PA transition system Σ-TS $= (\Omega, L, \Pi, \rightarrow, \models)$ is Σ-QTS $= (\Omega/_{\sim_\Omega}, L, \Pi, \rightarrow_{\sim_\Omega}, \models_{\sim_\Omega})$.

Proposition 2 (Qualitative PA transition system). Let Σ-QTS $= (\Omega/_{\sim_\Omega}, L, \Pi, \rightarrow_{\sim_\Omega}, \models_{\sim_\Omega})$ be the qualitative PA transition system corresponding to the PA transition system Σ-TS $= (\Omega, L, \Pi, \rightarrow, \models)$. Then

1. $\Omega/_{\sim_\Omega} = \mathcal{D}$;

2. $\rightarrow_{\sim_\Omega} \subseteq \mathcal{D} \times L \times \mathcal{D}$, such that $D \xrightarrow{l}_{\sim_\Omega} D'$ if and only if there exists $\xi(t) \in \Xi_\Sigma$ and τ, τ', $0 \le \tau < \tau'$, such that $\xi(\tau) \in D$, $\xi(\tau') \in D'$, and

   - if $l = int$, then for all $t \in [\tau, \tau']$: $\xi(t) \in D = D'$,
   - if $l = dim^+$, then for all $t \in (\tau, \tau']$: $\xi(t) \in D' \neq D$,
   - if $l = dim^-$, then for all $t \in [\tau, \tau')$: $\xi(t) \in D \neq D'$;

3. $\models_{\sim_\Omega} \subseteq \mathcal{D} \times \Pi$, such that $D \models Dsign = S$ if and only if for all $x \in D$, $S(x) = S$.

Proof. First, the quotient space $\Omega/_{\sim_\Omega}$ equals $\mathcal{D}$ by the definition of the equivalence relation $\sim_\Omega$. The second part of the proposition follows from the definition of $\rightarrow_{\sim_\Omega}$ as a relation on $\Omega/_{\sim_\Omega} \times L \times \Omega/_{\sim_\Omega}$ such that $D \xrightarrow{l}_{\sim_\Omega} D'$ if and only if there exist $x \in D$ and $x' \in D'$ such that $x \xrightarrow{l} x'$ [4]. The expression for $\rightarrow_{\sim_\Omega}$ is a direct consequence of $\Omega/_{\sim_\Omega} = \mathcal{D}$ and Definition 3. Third, $\models_{\sim_\Omega}$ is a relation defined on $\Omega/_{\sim_\Omega} \times \Pi$ such that $D \models_{\sim_\Omega} \pi$ if and only if there exists $x \in D$ such that $x \models \pi$ [4]. The properties considered here are of the type $Dsign = S$ (Definition 3). Theorem 1 implies the invariance of S for all $x \in D$. $\square$

Notice that the transitions labeled by dim^+ or dim^- connect two different flow domains, since in Proposition 2 we require that $D \neq D'$. This corresponds to a continuous evolution of the system along which it switches from one flow domain to another. On the contrary, the transitions labeled by int correspond to the continuous evolution of the system in a single flow domain.

As for Σ-TS, we define any sequence of flow domains $(D^0, \ldots, D^m)$, $m \ge 0$, as a *run* of Σ-QTS if and only if for all $j \in \{0, \ldots, m-1\}$, there exists $l \in L$ such that $D^j \xrightarrow{l}_{\sim_\Omega} D^{j+1}$. The satisfaction relation $\models_{\sim_\Omega}$ associates to every run a qualitative description of the evolution of the derivatives over time. Σ-QTS can be represented by a directed graph $G = (\mathcal{D}, \rightarrow_{\sim_\Omega})$, called the *state transition graph*. The paths in G represent

the runs of the system. The state transition graph corresponding to the two-gene example is represented in Figure 3(b), and $(D^{1.1}, D^{2.2}, D^{3.2}, D^{4.2}, D^{4.1})$ is an example of a run.

It directly follows from the definitions of quotient transition system and simulation of transition systems [4, 16] that Σ-QTS is a *simulation* of Σ-TS. The converse is not true in general, because Σ-QTS and Σ-TS are not bisimilar.

Proposition 3. Σ-QTS is a simulation of Σ-TS.

As a consequence of Proposition 3, if there exists a run $(x^0, \ldots, x^m)$ of Σ-TS, then there also exists a run $(D^0, \ldots, D^m)$ of Σ-QTS such that $x^i \in D^i$, for all $i \in \{0, \ldots, m\}$. In other words, Σ-QTS is a *conservative approximation* of Σ-TS.

4.2 Invariance of qualitative PA transition systems in parameter space

Σ-QTS provides a qualitative picture of the dynamics of a genetic regulatory network. This picture generally depends on the values of the parameters in the PADE model. Since exact numerical values for the thresholds θ and the production and degradation constants κ and γ are usually not available, it is important to verify to which extent Σ-QTS is robust to variations in the parameter values.

We therefore introduce a second equivalence relation $\sim_\Gamma \subseteq \Gamma \times \Gamma$, defined on the *parameter space* $\Gamma \subseteq \mathbb{R}_{>0}^q$ of the PA system, with q the number of parameters in Σ. Two parameter vectors $p, p' \in \Gamma$ are equivalent, if their corresponding qualitative PA transition systems, and hence their state transition graphs, are isomorphic. Given the equivalence relation $\sim_\Gamma$, we denote by $\Gamma/_{\sim_\Gamma}$ the quotient parameter space. That is, $\Gamma/_{\sim_\Gamma}$ is a partition of the parameter space consisting of sets over which the qualitative PA transition system is invariant.

For our purpose, subsets of Γ defined by the following *parameter inequality constraints* are particularly interesting.

Definition 5 (Parameter inequality constraints). The parameter inequality constraints of Σ are a set of total strict orders on $\{\theta_i^1, \ldots, \theta_i^{p_i}\} \cup \{\psi_i(M) \mid M \in \mathcal{M}_r\}$, for every $i \in \{1, \ldots, n\}$.

The following theorem states that Σ-QTS is invariant over the subsets of Γ defined by the inequality constraints of Definition 5. That is, the qualitative dynamics of Σ is robust to changes in parameter values that do not change the total order of the thresholds and focal concentrations.

Theorem 2 (Invariance). Let $P \subseteq \Gamma$ be a set defined by the parameter inequality constraints of Σ. Then, there exists some $Q \in \Gamma/_{\sim_\Gamma}$ such that $P \subseteq Q$.

Proof. Let $p, p' \in P$ be two parameter vectors. Given that p and p' satisfy the same parameter inequality constraints, they lead to the same mode and flow domain partitions (Definitions 1 and 2), and thus the same quotient space $\Omega/_{\sim_\Omega} = \mathcal{D}$. As will be seen in the next section, the relations $\rightarrow_{\sim_\Omega}$ and $\models_{\sim_\Omega}$ can be characterized by means of rules involving

inequality constraints (Propositions 4 to 8). It can be directly verified that the necessary and sufficient conditions in the rules apply in exactly the same way to p and p'. $\qquad\square$

Suppose that protein A inhibits the expression of gene b at a concentration lower than that required for the inhibition of the expression of its own gene. This gives the inequality constraint $\theta_a^1 < \theta_a^2$ in Figure 3(c). Moreover, if we further assume that when gene a is active, the concentration of protein A tends towards a level above which autoinhibition occurs, we obtain the inequality constraint $\kappa_a/\gamma_a > \theta_a^2$. For these inequality constraints, and similar constraints for protein B, the state transition graph in Figure 3(b) is invariant. Whereas exact numerical values for the parameters are usually not available, the weaker information required for the formulation of the inequality constraints can often be obtained from the experimental literature. This will be illustrated in Section 6 for the *E. coli* example.

5   Symbolic computation of qualitative PA transition system

The computation of the qualitative PA transition system Σ-QTS is greatly simplified by the fact that the domains D and the focal sets $\Psi(M)$ are hyperrectangular, which allows them to be expressed as product sets, *i.e.* $D = D_1 \times \ldots \times D_n$ and $\Psi(M) = \Psi_1(M) \times \ldots \times \Psi_n(M)$. As a consequence, the computation can be carried out for each dimension separately. In this section, we will describe rules to determine the set of states $\mathcal{D}$, the satisfaction relation $\models_{\sim_\Omega}$, and the transition relation $\rightarrow_{\sim_\Omega}$, as well as their implementation in the computer tool GNA. In practice the computations reduce to simple checks of ordering relations, which can be carried out symbolically by means of the parameter inequality constraints.

5.1   Computation of states

In order to compute the states of Σ-QTS, we need to determine the flow domain partition $\mathcal{D}$ of Ω (Proposition 2). This requires a total ordering of the threshold concentrations $\{\theta_i^1, \ldots, \theta_i^{p_i}\}$ and the focal concentrations $\{\psi_i(M) \mid M \in \mathcal{M}_r\}$, $i \in \{1, \ldots, n\}$ (Definition 2). The parameter inequality constraints of Σ provide this information.

Each flow domain has associated to it a number of properties, defined by the satisfaction relation $\models_{\sim_\Omega}$. From Proposition 2 it follows that computing the relation $\models_{\sim_\Omega}$ amounts to computing $S(x)$. The following two propositions, which are direct consequences of the basic mathematical properties of solutions of PA systems discussed in Section 3, show how to compute the derivative sign patterns.

Proposition 4 (Computation of *Dsign*). Let $D \in \mathcal{D}$ be an instantaneous flow domain. $D \models_{\sim_\Omega} Dsign = \{\}$.

Proof. There exists no solution remaining in D for some time, so by the definition of $S(x)$, we have $S(x) = \{\}$, for every $x \in D$. $\qquad\square$

Proposition 5 (Computation of *Dsign*). Let $D \in \mathcal{D}$ be a persistent flow domain. $D \models_{\sim_\Omega} Dsign = S \neq \{\}$, $S = S_1 \times \ldots \times S_n$, and for every $i \in \{1, \ldots, n\}$,

if D_i is located in a threshold or focal concentration hyperplane, then $S_i = \{0\}$;

otherwise,

- $-1 \in S_i$, if and only if for all $x \in D$ there exists $\psi \in \Psi(mode(D))$ such that $\psi_i - x_i < 0$;
- $0 \in S_i$, if and only if for all $x \in D$ there exists $\psi \in \Psi(mode(D))$ such that $\psi_i - x_i = 0$;
- $1 \in S_i$, if and only if for all $x \in D$ there exists $\psi \in \Psi(mode(D))$ such that $\psi_i - x_i > 0$.

Proof. Let $M = mode(D)$. Because D is persistent, there exists a solution $\xi(t) \in \Xi_\Sigma$ remaining in D for some time. Let $\xi(\tau) = x \in D$ for some $\tau \geq 0$. From (12) we infer $S(x) \neq \{\}$. By Theorem 1 this must hold for all $x \in D$, so $S \neq \{\}$.

Let D_i be located in a threshold or focal concentration hyperplane. For every solution $\xi(t)$ in D, such that $\xi(\tau) = x \in D$ for some $\tau \geq 0$, the derivative of $\xi(\tau)$ exists, and $\dot{\xi}(\tau) \in H(\xi(\tau))$, it holds that $\dot{\xi}_i(\tau) = 0$ and hence $S_i(x) = \{0\}$. By Theorem 1 this must hold for all $x \in D$, so $S_i = \{0\}$.

Alternatively, D_i is not located in a threshold or focal concentration hyperplane. We only consider the case $-1 \in S_i$ (the argument for 0 and 1 is similar). We first prove necessity of the inequality condition. $-1 \in S_i$ means that for all $x \in D$, there exists a solution $\xi(t) \in \Xi_\Sigma$ with $\xi(\tau) = x$ and $\tau \geq 0$, such that $\dot{\xi}_i(\tau) < 0$ (Proposition 2 and (12)). If M is regular, then $H_i(\xi(\tau)) = \{\nu_i^M (\psi_i(M) - \xi_i(\tau))\}$. Since $\dot{\xi}_i(\tau) \in H_i(\xi(\tau))$, it follows that $\psi_i(M) - x_i < 0$. If M is singular, then $H_i(\xi(\tau)) = \overline{rect}(\{\nu_i^{M'} (\psi_i(M') - \xi_i(\tau)) \mid M' \in R(M)\})$. $\dot{\xi}_i(\tau) < 0$ implies that there exists some $\psi(M')$, $M' \in R(M)$, such that $\psi_i(M') - x_i < 0$. Conversely, by Lemmas 1 and 4 there exist a solution $\xi(t) \in \Xi_\Sigma$ in M monotonically converging from x to ψ, with $\xi(0) = x$. Because $\psi_i - x_i < 0$, we have $\dot{\xi}_i(t) < 0$, $t \geq 0$, and thus $-1 \in S_i(x)$. By Theorem 1 it follows that $-1 \in S_i$, which proves sufficiency. $\qquad\square$

Notice that the total ordering on the threshold and focal concentrations expressed by the parameter inequality constraints (Definition 5) allows one to decide which of the conditions in the second part of Proposition 5 are satisfied. As a prerequisite, $\Psi(mode(D))$ needs to be determined, which is also straightforward, even for singular mode domains, given the definition of the focal set and the parameter inequality constraints. We illustrate the application of Propositions 4 and 5 to our two-gene example.

First, consider the flow domain $D^{1.1}$ in Figure 4(a). $\Psi(mode(D^{1.1}))$ equals $\{(\kappa_a/\gamma_a, \kappa_b/\gamma_b)'\}$, so that $D^{1.1}$ is persistent. We therefore apply Proposition 5 to determine $S = S_a \times S_b$, such that $D^{1.1} \models_{\sim_\Omega} Dsign = S$. Given that $D_a^{1.1} = [0, \theta_a^1)$ and $\psi_a(mode(D^{1.1})) = \kappa_a/\gamma_a$, and $\theta_a^1 < \kappa_a/\gamma_a$ according to the parameter inequality constraints in Figure 3(c), it follows that $\psi_a - x_a > 0$, for all $x \in D^{1.1}$ and $\psi \in \Psi(mode(D^{1.1}))$. Consequently, we infer $S_a = \{1\}$. Repeating this procedure in the x_b-dimension, we similarly find
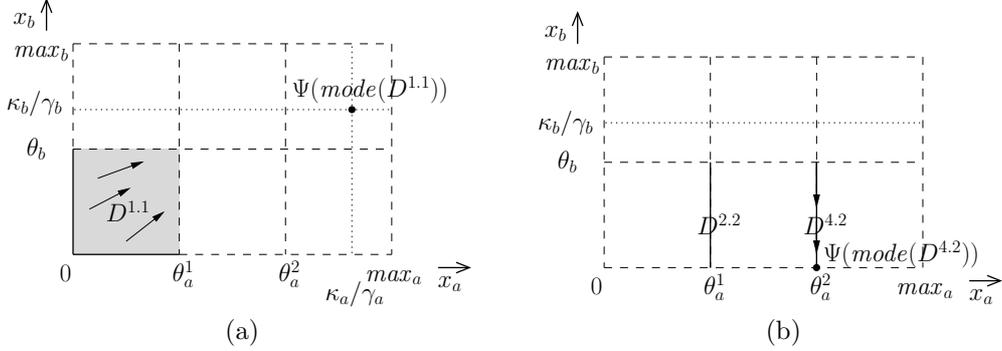
Figure 4: Derivative signs of flow domains: (a) $Dsign(D^{1.1}) = \{(1,1)'\}$ and (b) $Dsign(D^{4.2}) = \{(0,-1)'\}$ and $Dsign(D^{2.2}) = \{\}$.

$S_b = \{1\}$, so that $D^{1.1} \models_{\sim_\Omega} Dsign = \{(1,1)'\}$. This means that the solutions in $D^{1.1}$ are strictly increasing in both dimensions.

As a second example, consider the flow domain $D^{4.2}$, represented in Figure 4(b). $M^4 = mode(D^{4.2})$ is a singular mode domain, so we first have to compute $\Psi(M^4)$. It can be immediately verified that $R(M^4) = \{M^3, M^5\}$, where $\psi(M^3) = (\kappa_a/\gamma_a, 0)'$ and $\psi(M^5) = (0,0)'$ (Figure 2(b)). As a consequence, $\overline{rect}(\{\psi(M^3), \psi(M^5)\}) = [0, \kappa_a/\gamma_a] \times \{0\}$. From the parameter inequality constraints in Figure 3(c) it follows that $0 < \theta_a^2 < \kappa_a/\gamma_a$, so that $supp(M^4)$ and $\overline{rect}(\{\psi(M^3), \psi(M^5)\})$ intersect at $\Psi(M^4) = \{(\theta_a^2, 0)'\}$. $D^{4.2}$ is persistent, so that Proposition 5 applies. We determine $S = S_a \times S_b$, such that $D^{4.2} \models_{\sim_\Omega} Dsign = S$. Since $D_a^{4.2}$ coincides with a threshold hyperplane, $S_a = \{0\}$. Bearing in mind that $D_b^{4.2} = (0, \theta_b)$ and $\psi_b(M^4) = 0$, we find $S_b = \{-1\}$. This results in $D^{4.2} \models_{\sim_\Omega} Dsign = \{(0,-1)'\}$, which is of course consistent with the fact that the solutions sliding along $D^{4.2}$ monotonicaly converge towards $\Psi(M^4)$, and are therefore strictly decreasing in the x_b-dimension.

5.2   Computation of transitions between states

In Section 4.1 we have distinguished three types of transitions: int, dim^-, and dim^+. We will formulate, for each of these three cases, transition rules that can be applied by means of the parameter inequality constraints of Σ.

The transitions of type int are easy to determine, since by Proposition 2 they are necessarily self-transitions, from a flow domain D to itself, which occur if and only if D is persistent.

Proposition 6 (Computation of int transition). Let $D \in \mathcal{D}$. $D \xrightarrow{int}_{\sim_\Omega} D$ if and only if D is persistent.

Proof. By Proposition 2 an int transition occurs if and only if there exist solutions $\xi(t) \in \Xi_\Sigma$ remaining in D. That is, if and only if D is persistent.   □

Recall that the persistence of a domain can be determined by checking whether $\Psi(mode(D)) \neq \{\}$ (Lemmas 1 and 2). In our two-gene example, the focal set $\Psi(M^1)$ is not empty. Therefore, $D^{1.1}$ is persistent and there exists an *int* transition on $D^{1.1}$. On the other hand, $\Psi(M^2)$ is empty, so $D^{2.2}$ is instantaneous and does not have an *int* transition (Figure 3(b)).

A dim^+ transition $D \xrightarrow{dim^+} {}_{\sim_\Omega} D'$ is dimension increasing, and therefore requires that $D \subseteq \partial D'$ (Section 4.1). In order to make D' reachable from D, the solutions in D' must point away from D in the dimensions $i \in \{1, \dots, n\}$ for which $D_i \subseteq \partial D'_i$. This is expressed by the following proposition.

Proposition 7 (Computation of dim^+ transition). Let $D, D' \in \mathcal{D}$ and $D \subseteq \partial D'$. $D \xrightarrow{dim^+} {}_{\sim_\Omega} D'$ if and only if $\Psi(mode(D')) \neq \{\}$ and there exist $x \in D$, $x' \in D'$, and $\psi' \in \Psi(mode(D'))$, such that for all $i \in \{1, \dots, n\}$ for which $D_i \subseteq \partial D'_i$, it holds that

$$(\psi'_i - x_i)(x'_i - x_i) > 0. \tag{13}$$

Proof. Let $M' = mode(D')$. We first prove necessity by contraposition. If $\Psi(M') = \{\}$, then no solutions remain in D' and a transition from D to D' is not possible. Otherwise, suppose that for all $x \in D$, $x' \in D'$, and $\psi' \in \Psi(M')$, there exists some $i \in \{1, \dots, n\}$ for which $D_i \subseteq \partial D'_i$ and $(\psi'_i - x_i)(x'_i - x_i) \leq 0$. Furthermore, assume $x'_i - x_i > 0$ for all $x \in D$, $x' \in D'$ (the case $x'_i - x_i < 0$ goes analogously). As a consequence, $\psi'_i \leq x_i < x'_i$, for all $x \in D$, $x' \in D'$, and $\psi' \in \Psi(M')$. By Lemmas 1 and 3, for all solutions $\xi(t) \in \Xi_\Sigma$ in D', $\xi_i(t)$ monotonically converges towards $\Psi_i(M')$. As a consequence, no solution enters D' from D and there does not exist a transition $D \xrightarrow{dim^+} {}_{\sim_\Omega} D'$.

Next, we prove sufficiency. Suppose $\Psi(M') \neq \{\}$ and there exist $x \in D$, $x' \in D'$, and $\psi' \in \Psi(M')$, such that for all $i \in \{1, \dots, n\}$ for which $D_i \subseteq \partial D'_i$, it holds that $(\psi'_i - x_i)(x'_i - x_i) > 0$. By the definition of the flow domain partition (Definition 2) this holds for all $x \in D$ and $x' \in D'$. We further assume that $x'_i - x_i > 0$ for all $x \in D$, $x' \in D'$ (the case $x'_i - x_i < 0$ goes analogously). As a consequence, $x_i < x'_i < \psi'_i$ for all $x \in D$ and $x' \in D'$, and some $\psi' \in \Psi(M')$. By Lemma 4, for all $x' \in D'$ there exists a solution $\xi(t) \in \Xi_\Sigma$ in D', with $\xi(0) = x'$, which monotonically converges towards ψ'. More precisely, by the construction of $\xi(t)$ in the proof of Lemma 4, we have $\dot{\xi}_i(t) > 0$ as long as $\xi_i(t) < \phi'_i$. As a consequence, it is possible for solutions to enter D' from D and there exists a transition $D \xrightarrow{dim^+} {}_{\sim_\Omega} D'$. $\square$

The total strict ordering defined by the parameter inequality constraints allows the necessary and sufficient conditions for the existence of a dim^+ transition to be straightforwardly tested.

As an illustration of the proposition, consider the dim^+ transition from $D^{2.2}$ to $D^{3.2}$ (Figure 5). Notice that $D^{2.2} \in \partial D^{3.2}$. Moreover, $D_a^{2.2} \in \partial D_a^{3.2}$, so that $x'_a - x_a > 0$, for all $x \in D^{2.2}$ and $x' \in D^{3.2}$. Since $\psi_a(mode(D^{3.2})) = \kappa_a/\gamma_a$, and $\theta_a^2 < \kappa_a/\gamma_a$ due to the parameter inequality constraints (Figure 3(c)), we have $\psi'_a - x'_a > 0$, for all $x' \in D^{3.2}$ and $\psi' \in \Psi(mode(D^{3.2}))$. As a consequence, Proposition 7 allows us to infer that $D^{2.2} \xrightarrow{dim^+} {}_{\sim_\Omega}$
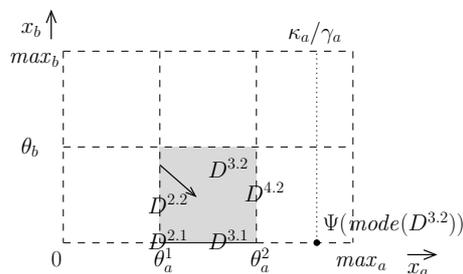
Figure 5: dim^+ transitions between flow domains. Representation of the state space, with a trajectory entering $D^{2.2}$ from $D^{3.2}$.

$D^{3.2}$. In a similar way, we can infer that there is a dim^+ transition from $D^{2.1}$ to $D^{3.1}$. However, as expected from the direction of the flow in Σ, a dim^+ transition from $D^{2.1}$ to $D^{3.2}$ is excluded, since $\psi'_b - x'_b < 0$ and $x'_b - x_b > 0$, for all $x \in D^{2.1}$, $x' \in D^{3.2}$ and $\psi' \in \Psi(mode(D^{3.2}))$.

The rule for dim^- transitions is almost symmetric and given in Appendix A.

5.3   Computer implementation

In summary, given a PADE system Σ and parameter inequality constraints, we can compute its qualitative abstraction, that is, the qualitative PA transition system Σ-QTS, by means of Propositions 4 to 8. Instead of numerically computing the derivative signs in the domains and the transitions between domains, we have developed symbolic algorithms exploiting the parameter inequality constraints of Definition 5. In particular, we map the total strict order on $B_i = \{\theta_i^1, \ldots, \theta_i^{p_i}\} \cup \{\psi_i(M) \mid M \in \mathcal{M}_r\}$, $i \in \{1, \ldots, n\}$, to the set $C_i = \{0, \ldots, |B_i| + 1\}$. This allows D_i and $\Psi_i(M)$ to be expressed as intervals on C. The conditions in the propositions then naturally translate into simple inequality tests on integer coordinates of domains and focal concentrations (see [7] for details).

The computation of Σ-QTS has been implemented in a new version of the computer tool *Genetic Network Analyzer (GNA)* [9]. In order to facilitate the analysis of Σ-QTS, the state transition graph generated by GNA can be exported to standard model-checking tools like NuSMV and CADP [9].

In practice, since the number of flow domains in the state space grows exponentially with the number of genes in the network, it is not usually possible to compute the complete state transition graph. However, knowledge of the part of the graph reachable from a (set of) initial flow domain(s) is often sufficient to answer most of the questions of biological interest. In GNA it is possible to either compute the complete qualitative PA transition system or carry out a *reachability analysis* from a specified set of initial domains. Moreover, GNA identifies the equilibrium states, that is, the flow domains corresponding to (sets of) equilibrium points, by testing whether $D \models_{\sim_\Omega} 0 \in Dsign$. For each of the equilibrium states,

the *attractor set* is computed, that is, the set of states from which the equilibrium state is reachable.

6 Qualitative analysis of nutritional stress response in *E. coli*

In case of nutritional stress, an *Escherichia coli* population abandons exponential growth and enters a non-growth state called *stationary phase*. This growth-phase transition is accompanied by numerous physiological changes in the bacteria, concerning among other things the morphology and the metabolism of the cell, as well as gene expression [37]. On the molecular level, the transition from exponential phase to stationary phase is controlled by a complex genetic regulatory network integrating various environmental signals. The molecular basis of the adaptation of the growth of *E. coli* to nutritional stress conditions has been the focus of extensive studies for decades [33]. However, notwithstanding the large amount of information accumulated on the genes, proteins, and other molecules known to be involved in the stress adaptation process, there is currently no global understanding of how the response of the cell emerges from the network of molecular interactions. This suggest the use of modeling and simulation tools to study the dynamics of the stress response. However, with some exceptions [14], numerical values for the parameters characterizing the interactions and the molecular concentrations are absent, which makes it difficult to apply traditional methods.
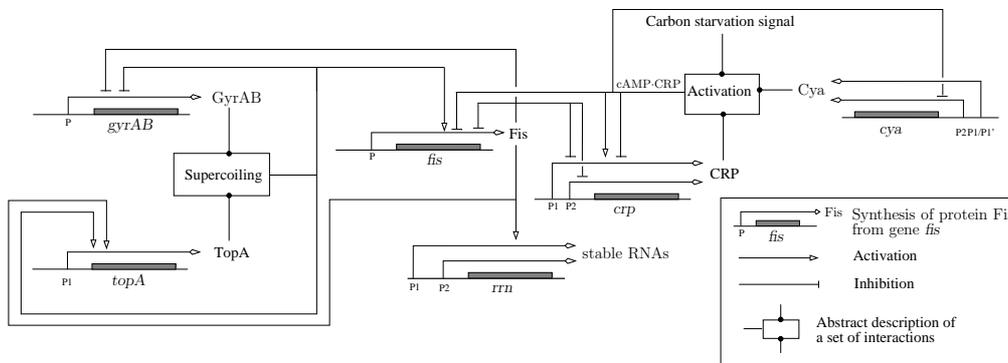

Figure 6: Network of key genes, proteins, and regulatory interactions involved in the nutritional stress network in *Escherichia coli*. The contents of the boxes labelled 'activation' and 'supercoiling' are detailed in [44].

The above circumstances have motivated the qualitative analysis of the nutritional stress response network in *E. coli* by means of the method presented in this paper [44]. On the basis of literature data, we have decided to focus, as a first step, on a network of six genes

that are believed to play a key role in the carbon starvation response (Figure 6). The network includes genes involved in the transduction of the nutritional stress signal (the global regulator *crp* and the adenylate cyclase *cya*), metabolism (the global regulator *fis*), cellular growth (the *rrn* genes coding for stable RNAs), and DNA supercoiling, an important modulator of gene expression (the topoisomerase *topA* and the gyrase *gyrAB*).

Based on this information, a PADE model of seven variables has been constructed, one protein concentration variable for each of the six genes and one input variable (u_{signal}) representing the presence or absence of a carbon starvation signal [44]. As an illustration, the piecewise-affine differential equation and the parameter inequality constraints for the state variable x_{topA} are given below.

$$\dot{x}_{topA} \;=\; \kappa_{topA} \; s^+(x_{gyrAB}, \theta^2_{gyrAB}) \; s^-(x_{topA}, \theta^1_{topA}) \; s^+(x_{fis}, \theta^4_{fis}) - \; \gamma_{topA} \; x_{topA}$$

$$0 < \theta^1_{topA} < \theta^2_{topA} < \kappa_{topA}/\gamma_{topA} < max_{topA}$$

The equation and inequalities state that in the presence of a high concentration of Fis ($s^+(x_{fis}, \theta^4_{fis}) = 1$), and of a high level of DNA supercoiling ($s^+(x_{gyrAB}, \theta^2_{gyrAB})$ $\cdot s^-(x_{topA}, \theta^1_{topA}) = 1$), the concentration of TopA increases, converging towards a high value ($\kappa_{topA}/\gamma_{topA} > \theta^2_{topA}$).

Using the computer tool GNA, we have performed reachability analyses on the qualitative PA transition system associated with the PADE model. The simulation of the entry into stationary phase has given rise to a state transition graph of 66 states, computed in less than 1 s on a PC (800 MHz, 256 Mb). Figure 7 represents the temporal evolution of two of the protein concentrations in a run. The predicted expression profiles are consistent with the observations [44].

The coupling of GNA with model-checking tools [9] has allowed a more systematic verification of observed dynamical properties. In [6], the measured concentration of the global regulator Fis is shown to decrease and become steady in stationary phase, which is characterized by a low concentration x_{rrn} of stable RNAs. This can be expressed by means of the following CTL formula:

$$\mathbf{EF}(\dot{x}_{fis} < 0 \wedge \mathbf{EF}(\dot{x}_{fis} = 0 \wedge x_{rrn} < \theta_{rrn})). \tag{14}$$

The verification of (14) takes a fraction of a second to complete and shows that the formula holds for the state transition graph. The observation that the transcription of *cya* is negatively regulated by cAMP and CRP [38] is also reproduced by the model. In fact, the following CTL formula is satisfied by the graph:

$$\mathbf{AG}(x_{crp} > \theta^3_{crp} \wedge x_{cya} > \theta^3_{cya} \wedge u_{signal} > \theta_{signal} \rightarrow \mathbf{EF} \; \dot{x}_{cya} < 0). \tag{15}$$

The formula expresses that always, for high levels of CRP and Cya, in the presence of the carbon starvation signal, the system can eventually reach a state in which the expression of *cya* decreases.
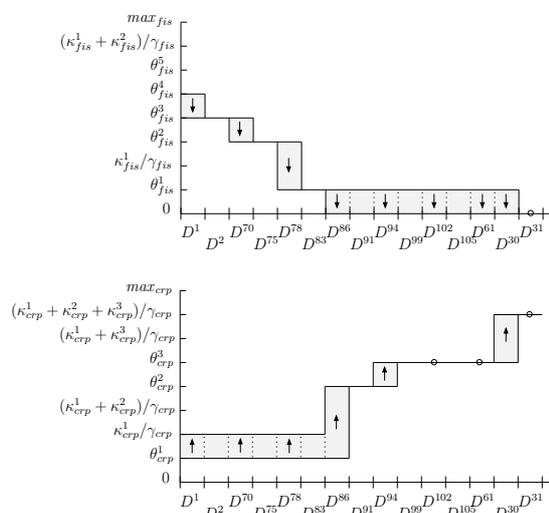
Figure 7: Temporal evolution of Fis and CRP concentrations in the run $(D^1, \ldots, D^{31})$. For every D^j in the run, the profiles display the corresponding concentration intervals D^j_{fis} and D^j_{crp}. The symbols $\uparrow$, $\downarrow$, and $\circ$ indicate the sign of the derivative for persistent states.

The application of the method has led to new insights into how the carbon starvation signal results in the slowing-down of bacterial growth characteristic for the stationary phase [44]. In summary, the analysis has brought to the fore the role of the mutual inhibition of Fis and CRP, which in the presence of a carbon starvation signal results in the inhibition of *fis* and in the activation of *crp*. This causes a decrease of the expression of the *rrn* genes, which code for stable RNAs and are a reliable indicator of cellular growth. In addition to this increased understanding of the transition from exponential to stationary phase, the model has yielded predictions on the occurrence of damped oscillations in some of the protein concentrations after a nutrient upshift, predictions that are being tested in our laboratory.

We are currently working on extended models of the nutritional stress response network. The increase in the number of variables naturally leads to the generation of larger state transition graphs by our method. In order to investigate the upscalability of the method more systematically, we have applied it to a PADE model with nine state and two input variables, describing the initiation of sporulation in the bacterium *Bacillus subtilis* [20]. More specifically, we have analyzed the response of the cell to carbon starvation in the case of the wild-type and a dozen of mutant strains. On average, the state transition graphs generated by our method consist of 585 states, with a maximum of 2234 states (computed in 14 s). The analysis of graphs of this size does not pose any problems for current model-checking tools, which shows that our approach is upscalable to large and complex networks.

7 Discussion and conclusions

We have presented a method for the qualitative analysis of hybrid models of genetic regulatory networks. The method is based on a class of piecewise-affine differential equation models that has been well-studied in mathematical biology. By defining a qualitative abstraction preserving the sign pattern of the derivatives of concentration variables, the continuous PA transition system associated with a PADE model is transformed into a discrete or qualitative PA transition system whose properties can be analyzed by means of classical model-checking tools. The qualitative PA transition system is a conservative approximation of the underlying continuous PA transition system and can be easily computed in a symbolic manner by exploiting inequality constraints on the parameters. We have applied the implementation of the method to the analysis of a system whose functioning is not well-understood by biologists today, the nutritional stress response in the bacterium *E. coli*.
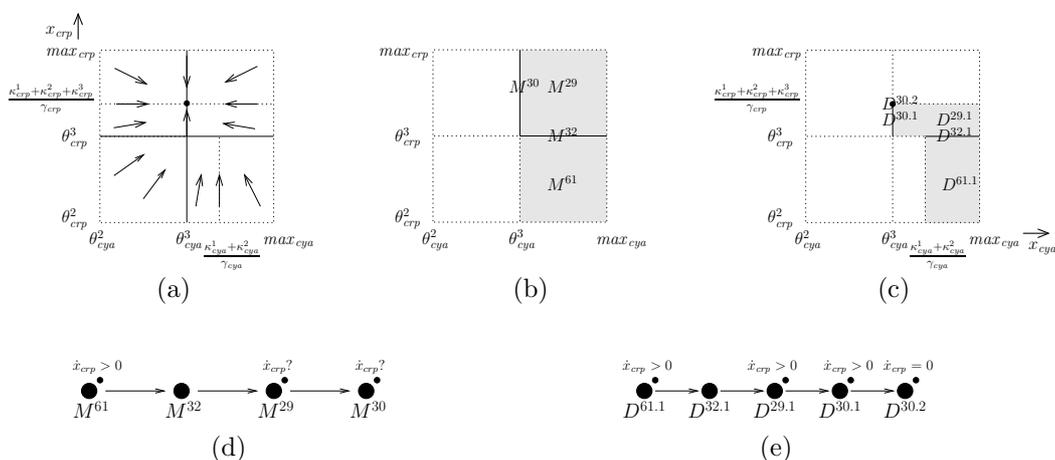
Figure 8: (a) Two-dimensional projection of a slice of the phase space of the *E. coli* stress response model for the variables x_{crp} and x_{cya}. (b)-(c) Partitioning into (b) mode domains and (c) flow domains of the projection. (d)-(e) Excerpts of the state transition graph resulting from the qualitative abstraction based on (d) mode domains and (e) flow domains.

The results of this paper extend our previous work on the qualitative analysis of PADE models of genetic regulatory networks [21, 22]. In particular, we have defined a refined partitioning of the state space which underlies a qualitative abstraction preserving stronger properties of the qualitative dynamics of the system, *i.e.* the derivative sign pattern. The resulting qualitative PA transition system is better adapted to the abstraction level of the experimental data, in the sense that it avoids verification of dynamical properties to be over-conservative. Consider Figure 8, which compares two-dimensional projections of a state-space slice of the stress response model. Depending on whether mode domains or flow domains are used as the abstraction criterion, the state transition graph will be different

(compare (d) and (e) of Figure 8). Whereas the CTL formula $EF(\dot{x}_{crp} > 0 \wedge EF (\dot{x}_{crp} < 0))$ holds for the graph in (d), this is not true in (e), thus revealing that the coarse-grained abstraction may cause models to escape refutation by available experimental data. Judging from our experience with several PADE models of bacterial regulatory networks, the use of a fine-grained abstractions leads to only a modest increase in the size of the state transition graph [10]. That is, the increase in precision does not exclude the application of the refined abstraction to larger systems.

The hybrid character of the dynamics of genetic regulatory networks has stimulated the interest in the application of hybrid-systems methods and tools over the past few years [1, 11, 21, 28, 36]. Our approach differs from this related work on several counts. Whereas we use piecewise-affine deterministic models, other groups have employed multi-affine and related deterministic models [11, 42] or stochastic models [36]. These models are less restrictive and thus provide a more precise description of the network interactions. However, they are more difficult to analyze and in higher dimensions usually require the application of numerical techniques. This is not straightforward to achieve for most biological systems, since numerical information on parameter values is usually imprecise or simply not available.

The PADE models (1) in this paper have been well-studied in mathematical biology [15, 18, 22, 24, 25, 26, 29, 30, 31, 32, 42, 43, 45, 47], and have also formed the basis for other work in the field of hybrid systems [28]. However, contrary to [28], we take into account the dynamics of the system on threshold hyperplanes, where equilibrium points and other phenomena of interest may occur [15, 32] (but see [1] for ideas on how to extend the approach in [28]). In [28] the partition of the state space is dynamically refined, so as to arrive at a qualitative PA transition system that is a better approximation of the original PA system. This requires the use of quantifier elimination methods [35] which allow to decide more general and more powerful properties than the rules proposed in Section 5 of this paper, but that also incur higher computational costs. The approach of this paper allows us to fully exploit the favorable mathematical properties of the PADE models (1), and thus promote the upscalability of the method to large and complex networks (Section 6).

From a more general perspective, our approach can be seen as an application of the notion of discrete abstraction, commonly used to study the dynamics of systems with an infinite number of states [2, 3, 4, 16, 34, 49]. Much work has focused on the identification of classes of continuous and hybrid dynamical systems for which bisimulation relations with finite transition systems are guaranteed to exist. The results of this paper can be seen as showing that the weaker simulation relation may also be of considerable practical interest, especially for classes of systems for which the existence of a finite bisimulation cannot be guaranteed. Discrete abstraction criteria similar to the one used in this paper, based on the sign of the (higher) derivatives of continuous variables, have been proposed by other authors in the fields of hybrids systems [2, 49] and control theory and qualitative reasoning [12, 13, 39, 40]. In comparison with these approaches, our work deals with a less general class of models. However, this allows the development and implementation of efficient, tailored algorithms for the practical computation of the qualitative dynamics of the system, even on (intersections of) threshold hyperplanes, where discontinuities may occur.

The possibility to use efficient algorithms for the computation of the qualitative PA transition system rests, to a large extent, on the approximation of the set $K(x)$ in (4) by the set $H(x)$ in (5). Because the latter set is hyperrectangular, the computation of domains, transitions, and sign patterns can be carried out seperately in every dimension, using the ordering of parameter values fixed by inequality constraints. Because $H(x)$ is an overapproximation of $K(x)$, the state transition graph may contain sequences of states that would not occur in the graph obtained by using $K(x)$. As a consequence, a PADE model may fail to be rejected by an observed time-series of measurements of the concentration variables. However, due to the fact that the approximation of $H(x)$ by $K(x)$ is conservative, a PADE model will never be falsely rejected. An obvious direction for further research would be to see whether finer overapproximations of $K(x)$ can be found that still allow tailored symbolic algorithms to be used that do not compromise the upscalability of the method.

A   Proofs

Proposition 1 (Reachability equivalence). For all $x, x' \in \Omega$, there exists a solution $\xi(t) \in \Sigma$ and τ, τ', such that $0 \leq \tau \leq \tau'$, $\xi(\tau) = x$, and $\xi(\tau') = x'$ if and only if there exists a run $(x^0, \ldots, x^m)$ of Σ-TS such that $x^0 = x$ and $x^m = x'$.

Proof. We first prove necessity. Consider two points $x, x' \in \Omega$, a solution $\xi(t) \in \Xi_\Sigma$, and τ, τ' such that $0 \leq \tau \leq \tau'$, $\xi(\tau) = x$, and $\xi(\tau') = x'$. If $\tau = \tau'$, then $(\xi(\tau))$ is a trivial run satisfying the conditions of the proposition. Otherwise, $\tau < \tau'$ and we denote by $D^0, \ldots, D^m$ the time-ordered sequence of flow domains traversed by $\xi(t)$ on the time interval $[\tau, \tau']$. $D^0, \ldots, D^m$ is a finite sequence, since by definition any solution of Σ reaches and leaves finitely-many times a threshold hyperplane during a time interval. If $m = 0$, then by Definition 3 there exists an *int* transition from $\xi(\tau)$ to $\xi(\tau')$, and consequently, $(\xi(\tau), \xi(\tau'))$ is a run satisfying the conditions of the proposition. Otherwise, $m > 0$, and we denote by $\sigma^0, \ldots, \sigma^{m-1}$, the switching times, formally defined as $\sigma^j = \sup \{t \in [\tau, \tau'] \mid \xi(t) \in D^j\}$, $j \in \{0, \ldots, m-1\}$. Finally, we introduce a sequence of time instants $\tau^0, \ldots, \tau^m$ such that $\xi(\tau^j) \in D^j$, $j \in \{0, \ldots, m\}$. More precisely, we define τ^0 as τ, τ^j as $(\sigma^{j-1} + \sigma^j)/2$, for all $j \in \{1, \ldots, m-1\}$, and τ^m as τ'. It is not difficult to show by induction on j that $(\xi(\tau^0), \ldots, \xi(\tau^j))$ is a run, for every $j \in \{0, \ldots, m\}$ [7].

Next, we prove sufficiency. Consider a run $(x^0, \ldots, x^m)$ of Σ-TS, with $x^0 = x$ and $x^m = x'$. If $m = 0$, then $x = x'$ and any solution $\xi(t) \in \Xi_\Sigma$ with $\xi(0) = x$ satisfies the conditions of the proposition for $\tau = \tau' = 0$. In the sequel, we suppose $m > 0$. Then, by Definition 3, there exists a sequence of solutions $(\xi^0(t), \ldots, \xi^{m-1}(t))$, and of time instants

$(\tau^0, \ldots, \tau^m)$, such that for all $j \in \{0, \ldots m-1\}$, $\xi^j(t)$ is defined on the time interval $[\tau^j, \tau^{j+1}]$, with $\tau^j < \tau^{j+1}$, and satisfies $\xi^j(\tau^j) = x^j$ and $\xi^j(\tau^{j+1}) = x^{j+1}$. It can be straightforwardly shown [7] that the concatenation of the solutions $\xi^j(t)$, $j \in \{0, \ldots, m-1\}$, is a solution that satisfies the conditions of the proposition. □

Proposition 8 (Computation of dim^- transition). Let $D, D' \in \mathcal{D}$ and $D' \subseteq \partial D$. $D \xrightarrow{dim^-}_{\sim_\Omega} D'$ if and only if $\Psi(mode(D)) \neq \{\}$ and there exist $x \in D$, $x' \in D'$, and $\psi \in \Psi(mode(D))$, such that either (a) for all $i \in \{1, \ldots, n\}$ for which $D'_i \subseteq \partial D_i$, it holds that

$$(\psi_i - x'_i)(x'_i - x_i) > 0, \tag{16}$$

or (b) it holds that $\psi = x'$.

Proof. Let $M = mode(D)$. We first prove necessity by contraposition. If $\Psi(M) = \{\}$, then there exists no solution remaining in D for some time, and a transition from D to D' is not possible. Otherwise, suppose that for all $x \in D$, $x' \in D'$, and $\psi \in \Psi(M)$, there exists some $i \in \{1, \ldots, n\}$ for which $D'_i \subseteq \partial D_i$, $(\psi_i - x'_i)(x'_i - x_i) \leq 0$, and $\psi \neq x'$. We assume $x'_i - x_i > 0$ for all $x \in D$, $x' \in D'$ (the case $x'_i - x_i < 0$ goes analogously).

If the inequality is strict, then $\psi_i < x'_i$, for all $x' \in D'$ and $\psi \in \Psi(M)$. By Lemma 3, for all solutions $\xi(t) \in \Xi_\Sigma$ in D, $\xi_i(t)$ monotonically converges towards $\Psi_i(M)$. As a consequence, no solution enters D' from D, and there does not exist a transition $D \xrightarrow{dim^-}_{\sim_\Omega} D'$. If the inequality is not strict, then by definition of the flow domain partition (Definition 2), we have $\psi_i = x'_i$ for all $x' \in D'$ and $\psi \in \Psi(M)$. From the same definition it also follows that either $\psi_i = \psi_i(M)$ (if M is regular) or $\psi_i = \max_{M' \in R(M)} \psi_i(M')$ (if M is singular). It can be shown by construction in the way of the proof of Lemma 4 that solutions reach D' only asymptotically, as $t \to \infty$. Moreover, from the definition of $\Psi(M)$ and Lemmas 1 and 3, it follows for every $j \in \{1, \ldots, n\}$ that either $\xi_j(t) \in \Psi_j(M)$ or that $\xi_j(t)$ monotonically converges towards $\Psi_j(M)$, as long as $\xi(t)$ remains in D (and thus in M). Therefore, because $\xi(t)$ reaches $x' \in D'$ asymptotically, we have $x' \in \Psi(M)$. This contradicts the assumption $\psi \neq x'$, so there does not exist a dim^- transition.

Next, we prove sufficiency. Suppose $\Psi(M) \neq \{\}$ and there exist $x \in D$, $x' \in D'$, and $\psi \in \Psi(M)$, such that (a) for all $i \in \{1, \ldots, n\}$ for which $D_i \subseteq \partial D'_i$, it holds that $(\psi_i - x'_i)(x'_i - x_i) > 0$, or (b) it holds that $\psi = x'$. Consider (a) and assume $x'_i - x_i > 0$ for all $x \in D$, $x' \in D'$ (the case $x'_i - x_i < 0$ goes analogously). As a consequence, $x_i < x'_i < \psi_i$ for all $x \in D$ and $x' \in D'$, and some $\psi \in \Psi(M)$. By Lemma 4, for all $x \in D$ there exists a solution $\xi(t) \in \Xi_\Sigma$ in D, with $\xi(0) = x$, which monotonically converges towards ψ. We choose x such that the corresponding solution enters D' from D, and thus obtain a transition $D \xrightarrow{dim^-}_{\sim_\Omega} D'$. In case (b) we find $x_i < x'_i = \psi_i$, and by the same argument there exists a solution that enters D' from D. However, in this case the solution only reaches D' asymptotically, as $t \to \infty$ (see the proof of necessity). □

References

[1] M. Adélade and G. Sutre. Parametric analysis and abstraction of genetic regulatory networks. In *Proceedings of the Second Workshop on Concurrent Models in Molecular Biology, BioCONCUR 04*, London, UK, 2004.

[2] R. Alur, T. Dang, and F. Ivancic. Progress on reachability analysis of hybrid systems using predicate abstraction. In A. Pnueli and O. Maler, editors, *Hybrid Systems: Computation and Control (HSCC 2003)*, volume 2623 of *Lecture Notes in Computer Science*, pages 4–19. Springer-Verlag, Berlin, 2003.

[3] R. Alur and D.L. Dill. A theory of timed automata. *Theoretical Computer Science*, 126(2):183–235, 1994.

[4] R. Alur, T.A. Henzinger, G. Lafferriere, and G.J. Pappas. Discrete abstractions of hybrid systems. *Proceedings of the IEEE*, 88(7):971–984, 2000.

[5] P.J. Antsaklis and X.D. Koutsoukos. Hybrid systems: Review and recent progress. In T. Samad and G. Balas, editors, *Software-enabled Control: Information Technologies for Dynamical Systems*, chapter 4. Wiley-IEEE Press, 2003.

[6] T. Ali Azam, A. Iwata., A. Nishimura, S. Ueda, and A. Ishihama. Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *Journal of Bacteriology*, 181(20):6361–6370, 1999.

[7] G. Batt. *Validation of qualitative models of genetic regulatory networks: A method based on formal verification techniques.* PhD thesis, Universit Joseph Fourier, Grenoble, France, 2006. In French.

[8] G. Batt, H. de Jong, J. Geiselmann, M. Page, D. Ropers, and D. Schneider. Symbolic reachability analysis of genetic regulatory networks using qualitative abstraction. Technical Report RR-5362, INRIA Rhône-Alpes, 2004.

[9] G. Batt, D. Ropers, H. de Jong, J. Geiselmann, R. Mateescu, M. Page, and D. Schneider. Validation of qualitative models of genetic regulatory networks by model checking: Analysis of the nutritional stress response in *Escherichia coli*. *Bioinformatics*, 21(Suppl 1):i19–i28, 2005.

[10] G. Batt, D. Ropers, H. de Jong, J. Geiselmann, M. Page, and D. Schneider. Qualitative analysis and verification of hybrid models of genetic regulatory networks: Nutritional stress response in *Escherichia coli*. In M. Morari and L. Thiele, editors, *Hybrid Systems: Computation and Control (HSCC 2005)*, volume 3414 of *Lecture Notes in Computer Science*, pages 134–150. Springer-Verlag, Berlin, 2005.

[11] C. Belta, J.M. Esposito, J. Kim, and V. Kumar. Computational techniques for analysis of genetic network dynamics. *International Journal of Robotics Research*, 24(2):219–235, 2005.

[12] O. Bernard and J.-L. Gouzé. Transient behavior of biological loop models with application to the Droop model. *Mathematical Biosciences*, 127(1):19–43, 1995.

[13] O. Bernard and J.-L. Gouzé. Global qualitative description of a class of nonlinear dynamical systems. *Artificial Intelligence*, 136(1):29–59, 2002.

[14] K. Bettenbrock, S. Fischer, A. Kremling, K. Jahreis, T. Sauter, and E.D. Gilles. A quantitative approach to catabolite repression in *Escherichia coli*. *Journal of Biological Chemistry*, 281(5):2578–2584, 2005.

[15] R. Casey, H. de Jong, and J.-L. Gouzé. Piecewise-linear models of genetic regulatory networks: Equilibria and their stability. *Journal of Mathematical Biology*, 52(1):27–56, 2006.

[16] A. Chutinan and B.H. Krogh. Verification of infinite-state dynamic systems using approximate quotient transition systems. *IEEE Transactions on Automatic Control*, 46(9):1401–1410, 2001.

[17] E.M. Clarke, O. Grumberg, and D.A. Peled. *Model Checking*. MIT Press, Boston, MA, 1999.

[18] R. Coutinho, B. Fernandez, R. Lima, and A. Meyroneinc. Discrete time piecewise affine models of genetic regulatory networks. *Journal of Mathematical Biology*, 52(4):524–570, 2006.

[19] H. de Jong. Modeling and simulation of genetic regulatory systems: A literature review. *Journal of Computational Biology*, 9(1):67–103, 2002.

[20] H. de Jong, J. Geiselmann, G. Batt, C. Hernandez, and M. Page. Qualitative simulation of the initiation of sporulation in *B. subtilis*. *Bulletin of Mathematical Biology*, 66(2):261–299, 2004.

[21] H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, and J. Geiselmann. Hybrid modeling and simulation of genetic regulatory networks: A qualitative approach. In A. Pnueli and O. Maler, editors, *Hybrid Systems: Computation and Control (HSCC 2003)*, volume 2623 of *Lecture Notes in Computer Science*, pages 267–282. Springer-Verlag, Berlin, 2003.

[22] 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.

[23] H. de Jong and D. Ropers. Qualitative approaches towards the analysis of genetic regulatory networks. In Z. Szallasi, V. Periwal, and J. Stelling, editors, *System Modeling in Cellular Biology: From Concepts to Nuts and Bolts*, pages 125–148. MIT Press, Cambridge, MA, 2006.

[24] R. Edwards. Analysis of continuous-time switching networks. *Physica D*, 146(1-4):165–199, 2000.

[25] R. Edwards, H.T. Siegelmann, K. Aziza, and L. Glass. Symbolic dynamics and computation in model gene networks. *Chaos*, 11(1):160–169, 2001.

[26] E. Farcot. Geometric properties of a class of piecewise affine biological network models. *Journal of Mathematical Biology*, 52(3):373–418, 2006.

[27] A.F. Filippov. *Differential Equations with Discontinuous Righthand Sides*. Kluwer Academic Publishers, Dordrecht, 1988.

[28] R. Ghosh and C.J. Tomlin. Symbolic reachable set computation of piecewise affine hybrid automata and its application to biological modelling: Delta-Notch protein signalling. *Systems Biology*, 1(1):170–183, 2004.

[29] L. Glass. Classification of biological networks by their qualitative dynamics. *Journal of Theoretical Biology*, 54(1):85–107, 1975.

[30] 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.

[31] L. Glass and J.S. Pasternack. Stable oscillations in mathematical models of biological control systems. *Journal of Mathematical Biology*, 6:207–223, 1978.

[32] J.-L. Gouzé and T. Sari. A class of piecewise linear differential equations arising in biological models. *Dynamical Systems*, 17(4):299–316, 2002.

[33] R. Hengge-Aronis. The general stress response in *Escherichia coli*. In G. Storz and R. Hengge-Aronis, editors, *Bacterial Stress Responses*, pages 161–177. ASM Press, Washington, DC, 2000.

[34] T.A. Henzinger. Hybrid automata with finite bisimulations. In Z. Fülöp and F. Gécseg, editors, *22nd International Colloquium on Automata, Languages and Programming, ICALP'95*, volume 944 of *Lecture Notes in Computer Science*, pages 324–335. Springer-Verlag, Berlin, 1995.

[35] H. Hong. An improvement of the projection operator in cylindrical algebraic decomposition. In *International Symposium of Symbolic and Algebraic Computation, ISSAC'90*, pages 261–264, New York, 1990. ACM Press.

[36] J. Hu, W.-C. Wu, and S. Sastry. Modeling subtilin production in *Bacillus subtilis* using hybrid stochastic systems. In R. Alur and G.J. Pappas, editors, *Hybrid Systems: Computation and Control (HSCC 2004)*, volume 2993 of *Lecture Notes in Computer Science*, pages 417–431. Springer-Verlag, Berlin, 2004.

[37] G.W. Huisman, D.A. Siegele, M.M. Zambrano, and R. Kolter. Morphological and physiological changes during stationary phase. In F.C. Neidhardt, R. Curtiss III, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, M. Schaechter, and H.E. Umbarger, editors, *Escherichia coli and Salmonella: Cellular and Molecular Biology*, pages 1672–1682. ASM Press, Washington, DC, 2nd edition, 1996.

[38] M. Kawamukai, J. Kishimoto, R. Utsumi, M. Himeno, T. Komano, and H. Aiba. Negative regulation of adenylate cyclase gene (*cya*) expression by cyclic AMP-cyclic AMP receptor protein in *Escherichia coli*: Studies with *cya-lac* protein and operon fusion plasmids. *Journal of Bacteriology*, 164(2):872–877, 1985.

[39] B.J. Kuipers. *Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge*. MIT Press, Cambridge, MA, 1994.

[40] J. Lunze. Qualitative modelling of linear dynamical systems with quantised state measurements. *Automatica*, 30(3):417–432, 1994.

[41] J. Lygeros, G.J. Pappas, and S. Sastry. An introduction to hybrid system modeling, analysis, and control. First Nonlinear Control Network (NCN) Pedagogical School, Athens, Greece, 1999.

[42] T. Mestl, E. Plahte, and S.W. Omholt. A mathematical framework for describing and analysing gene regulatory networks. *Journal of Theoretical Biology*, 176(2):291–300, 1995.

[43] E. Plahte and S. Kjóglum. Analysis and generic properties of gene regulatory networks with graded response functions. *Physica D*, 201(1):150–176, 2005.

[44] D. Ropers, H. de Jong, M. Page, D. Schneider, and J. Geiselmann. Qualitative simulation of the carbon starvation response in *Escherichia coli*. *BioSystems*, 84(2):124–152, 2006.

[45] E.H. Snoussi. Qualitative dynamics of piecewise-linear differential equations: A discrete mapping approach. *Dynamics and Stability of Systems*, 4(3-4):189–207, 1989.

[46] E.D. Sontag. Molecular systems biology and control. *European Journal of Control*, 11(5):396–435, 2005.

[47] R. Thomas and R. d'Ari. *Biological Feedback*. CRC Press, Boca Raton, FL, 1990.

[48] R. Thomas, D. Thieffry, and M. Kaufman. Dynamical behaviour of biological regulatory networks: I. Biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bulletin of Mathematical Biology*, 57(2):247–276, 1995.

[49] A. Tiwari and G. Khanna. Series abstractions for hybrid automata. In C.J. Tomlin and M.R. Greenstreet, editors, *Hybrid Systems: Computation and Control (HSCC 2002)*, volume 2289 of *Lecture Notes in Computer Science*, pages 465–478. Springer-Verlag, Berlin, 2002.

Contents