Qualitative Simulation of Large and Complex Genetic Regulatory Systems

Hidde de Jong[1] **and Michel Page**[1,2]

[1]Institut National de Recherche en Informatique et en Automatique (INRIA) Rhône-Alpes
655 avenue de l'Europe, 38330 Montbonnot Saint Martin, France
[2]Université Pierre Mendès France, Grenoble, France
{ Hidde.de-Jong,Michel.Page}@inrialpes.fr

Abstract

Modeling and simulation techniques developed within qualitative reasoning might profitably be used for the analysis of genetic regulatory systems. A major problem of current qualitative simulation methods is their lack of upscalability. We describe a method that is able to deal with large and complex systems, and discuss its performance in simulation experiments with random regulatory networks.

Introduction

In the last few years biologists have completed the sequencing of the entire genome of model organisms like *S. cerevisiae* and *E. coli*, and the human genome is expected to follow without much delay. The analysis of these huge amounts of data involves such tasks as the identification of genes and regulatory signals, the prediction of folding structures of proteins, and the construction of phylogenetic trees. It is clear, however, that the structural analysis of sequence data needs to be complemented with a functional analysis to elucidate the role of genes in controlling fundamental biological processes. One of the central problems to be addressed by this *functional genomics* is the analysis of regulatory systems controlling gene expression through a network of interactions between DNA, RNA, and proteins. The study of these networks will contribute to our understanding of complex processes like the development of a multicellular organism from a fertilized egg, and to applications ranging from drug discovery to biological computing.

Computer tools will be indispensable for the analysis of genetic regulatory systems, as these usually involve many genes connected through regulatory cascades and feedback loops. Currently, only a few regulatory networks are well-understood on the molecular level, and quantitative knowledge about the interactions is seldom available. This has stimulated an interest in modeling and simulation techniques developed within qualitative reasoning (QR), most notably QSIM (Kuipers 1994) and QPT (Forbus 1984). QR methods have been applied to the regulation of tryptophan synthesis (Karp 1993) and λ phage growth (Heidtke & Schulze-Kremer 1998) in *E. coli*, and to the regulation of the transcription factor families AP-1 and NF-κB in different classes of animals (Trelease, Henderson, & Park 1999).

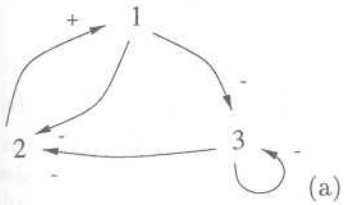
A major problem is the lack of upscalability of these approaches. As a consequence of the weak nature of qualitative constraints, and the difficulty to identify implicit constraints, behavior trees and envisionments quickly grow out of bounds. This causes the range of application of the methods to be limited to regulatory systems of modest size and complexity. Systems of even a few genes related by positive and negative feedback loops cannot be handled, unless these systems have been so well-studied already that behavior predictions can be tightly constrained.

In this paper we will show that it is possible to qualitatively analyse genetic regulatory networks of larger size and complexity. In order to achieve this, we describe the systems by a class of piece-wise linear differential equations (PLDEs) putting strong constraints on possible trajectories in the phase space. Simulation is carried out by an algorithm tailored to this class of models. The method has been implemented in Java and used for the simulation of regulatory networks of currently up to 12 genes involved in complex feedback loops. As improvements of the method and its implementation can be readily imagined, it seems possible to achieve further upscaling.

In the next two sections, we will introduce the class of PLDEs by which genetic regulatory systems can be described and review its mathematical properties. The subsequent sections introduce the qualitative simulation algorithm and present the results of simulation studies. The paper concludes with a discussion of the results and an outline of ideas for further work.

Modeling genetic regulatory systems

The regulation of gene expression is achieved by the interactions between proteins, DNA, RNA, and other molecules taking place under a variety of cellular and environmental conditions (Lewin 1997). The set of genes involved in a particular process, their regulatory sites, the products and the regulators of the genes, and their mutual interactions constitute a *genetic regulatory system*. The structure of this system can be rep-



(a)

$$\dot{x}_1 = \kappa_{12} h^+(x_2, \theta_{12}, m) - \gamma_1 x_1$$
$$\dot{x}_2 = \kappa_{21} h^-(x_1, \theta_{21}, m) \cdot \kappa_{23} h^-(x_3, \theta_{23}, m) - \gamma_2 x_2$$
$$\dot{x}_3 = \kappa_{31} h^-(x_1, \theta_{31}, m) + \kappa_{33} h^-(x_3, \theta_{33}, m) - \gamma_3 x_3 \quad \text{(b)}$$

$$0 < \theta_{21} < \theta_{31} < max_1$$
$$0 < \theta_{12} < max_2$$
$$0 < \theta_{23} < \theta_{33} < max_3 \quad \text{(c)}$$

$$0 < \kappa_{12}/\gamma_1 < \theta_{21}$$
$$\theta_{12} < \kappa_{21}\kappa_{23}/\gamma_2 < max_2$$
$$0 < \kappa_{31}/\gamma_3 < \theta_{23}$$

$$0 < \kappa_{33}/\gamma_3 < \theta_{23}$$
$$\theta_{23} < (\kappa_{31} + \kappa_{33})/\gamma_3 < \theta_{33} \quad \text{(d)}$$

Figure 1: (a) Example regulatory network modeled by the state equations in (b) the threshold inequalities in (c), and the nullcline inequalities in (d). The symbols '+' and '−' in (a) denote activating and inhibitory interactions, respectively.

resented as a *genetic regulatory network*, composed of the elements and their interactions. In Fig. 1(a) a simple example of a regulatory network is shown, involving three genes and several activating and inhibiting interactions.

Gene regulation is often modeled by differential equations expressing the rate of synthesis of a gene product in terms of the balance of the amounts of product appearing and disappearing per unit time (Glass 1975; Thomas & d'Ari 1990). This leads to *state equations* of the form

$$\dot{x}_i = f_i(\boldsymbol{x}) - \gamma_i x_i, \; x_i \geq 0, \; 1 \leq i \leq n, \qquad (1)$$

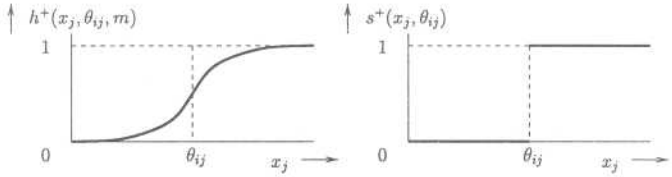
where $\boldsymbol{x}$ is a vector of cellular concentrations of gene products, $\gamma_i > 0$ the decay rate of x_i, and f_i a usually highly nonlinear function. The rate of synthesis of the ith product is dependent upon the concentrations $\boldsymbol{x}$, possibly including x_i. The term $-\gamma_i x_i$ states that the concentration x_i decreases through spontaneous destruction processes, such as degradation, diffusion, and growth dilution, at a rate proportional to the concentration itself.

The function f_i in (1) can be further specified as the composition of a number of basic *regulation functions*. A regulation function accounts for the variation in expression level of a gene i with the concentration x_j of the product of another gene j. A regulation function often found in the literature is the sigmoid Hill curve (Fig. 2):

$$h^+(x_j, \theta_{ij}, m) = x_j^m / (x_j^m + \theta_{ij}^m), \qquad (2)$$

where $\theta_{ij} > 0$ denotes the *threshold* for the influence of j on i, and $m > 1$ a parameter determining the steepness of the function around θ_{ij}. The function ranges from 0 to 1, and increases as $x_j \to \infty$, so that j positively regulates or *induces* i. In order to express that j negatively regulates or *represses* i, the regulation function $h^+(x_j, \theta_{ij}, m)$ must be replaced by $h^-(x_j, \theta_{ij}, m) = 1 - h^+(x_j, \theta_{ij}, m)$.

Individual regulation functions can be combined to account for the fact that usually several regulators combine in determining the rate of synthesis of i. In particular, regulation functions $h^+(x_j, \theta_{ij}, m)$ and $h^+(x_k, \theta_{ik}, m)$ are summed when either j or k is sufficient for efficient synthesis, and they are multiplied when both j and k are necessary. In Fig. 1(b) the equations for the example network are shown, assuming that the two interactions regulating gene 2 are multiplicative and the two interactions regulating gene 3 additive. The equations also contain positive *rate constants* κ_{ij}, determining the maximum expression level of i under the influence of j.



(a) (b)

Figure 2: Examples of regulation functions: (a) Hill function and (b) step function.

Due to the nonlinear character of the functions f_i, analytical solution of the state equations (1) is not possible. The nonlinear terms can be eliminated by replacing the continuous Hill function by the discontinuous step function (Fig. 2):

$$s^+(x_j, \theta_{ij}) = \begin{cases} 1, & x_j > \theta_{ij} \\ 0, & x_j < \theta_{ij} \end{cases} \qquad (3)$$

The resulting equations are piecewise-linear differential equations (PLDEs) of the form

$$\dot{x}_i = b_i(\boldsymbol{x}) - \gamma_i x_i, \; x_i \geq 0, \; 1 \leq i \leq n, \qquad (4)$$

where b_i is a piecewise-linear function composed of additions and multiplications of step functions. The approximation of a continuous sigmoid by a discontinuous

step function has been justified on the ground of the switch-like character displayed by genes whose expression is regulated by steep sigmoid curves (e.g., Glass & Kauffman (1973)). In what follows, we will assume that genetic regulatory systems are modeled by PLDEs of the form (4).

Mathematical analysis

Eqs. (4) have been well-studied in mathematical biology (Glass 1975; Glass & Pasternack 1978; Thomas & d'Ari 1990; Snoussi 1989; Snoussi & Thomas 1993; Lewis & Glass 1991; Plahte, Mestl, & Omholt 1994; Mestl, Plahte, & Omholt 1995a; 1995b). Consider an n-dimensional (hyper)box of the phase space defined as follows:

$$0 \leq x_i \leq max_i = \max_{\boldsymbol{x} \geq \mathbf{0}} b_i(\boldsymbol{x})/\gamma_i, \ 1 \leq i \leq n. \quad (5)$$

It can be shown that all trajectories starting inside the n-box will remain in it, while trajectories starting outside will enter the box at some time or approach it asymptotically as $t \to \infty$. We assume that $\theta_{ji} < max_i$ for all genes j regulated by the product of gene i. The $n-1$-dimensional threshold (hyper)planes $x_i = \theta_{ji}$ divide the n-box into *volumes*. The volumes of the n-box are determined by the *threshold inequalities*

$$0 < \sigma_i^{(1)} < \ldots < \sigma_i^{(p_i)} < max_i, \quad (6)$$

obtained by ordering and renaming the p_i thresholds θ_{ji} of gene i. The threshold equations for the example of the previous section are shown in Fig. 1(c). Since the step function is not defined at its threshold, Eqs. (4) are not defined in the threshold planes separating the volumes. Fig. 3(a) displays the phase space box corresponding to the example network.

In each volume of the n-box, (4) reduces to *volume state equations* with a constant production term μ_i composed of rate parameters in b_i:

$$\dot{x}_i = \mu_i - \gamma_i x_i, \ x_i \geq 0, \ 1 \leq i \leq n. \quad (7)$$

Notice that Eqs. (7) are linear and orthogonal. Fig. 3(b) gives an example of the state equations corresponding to the volume $0 \leq x_1 < \theta_{21}, \theta_{12} < x_2 \leq max_2$, and $\theta_{33} < x_3 \leq max_3$. It can be easily shown that within a volume all trajectories evolve towards the *focal state* $\boldsymbol{\mu}/\boldsymbol{\gamma}$. The focal state is the single, stable steady of (7). It lies at the intersection of the *nullcline* (hyper)planes $x_i = \mu_i/\gamma_i$ defined by $\dot{x}_i = 0$. As the nullclines are assumed not to coincide with the threshold planes, the focal state will be located at some distance from the threshold planes.

The focal state of a volume may lie inside or outside that volume. Whether the focal state lies inside or outside the volume is determined by the *nullcline inequalities*, which locate the nullclines $x_i = \mu_i/\gamma_i$ between two subsequent thresholds of x_i:

$$\sigma_i^{(l_i)} < \mu_i/\gamma_i < \sigma_i^{(l_i+1)}, \ 1 \leq l_i < p_i, \quad (8)$$

and the special cases $0 < \mu_i/\gamma_i < \sigma_i^{(1)}$ and $\sigma_i^{(p_i)} < \mu_i/\gamma_i < max_i$. If for every i, μ_i/γ_i lies between the threshold boundaries of the volume, then the focal state lies inside the volume. If not, it lies outside the volume (Fig. 3(c)). The nullcline inequalities for the example regulatory system are shown in Fig. 1(d). Notice that several nullcline inequalities have been specified for x_3, as a consequence of the fact that μ_3 changes between different volumes. More generally, the set of possible nullclines in the ith dimension is given by $\{b_i(\boldsymbol{x})/\gamma_i \mid \mathbf{0} \leq \boldsymbol{x} \leq \boldsymbol{max}\}$.

If the focal state lies outside the volume, the trajectories will tend towards one or several of the threshold planes bounding the volume. Since (4) is not defined in the threshold planes, special attention should be given to the behavior of the system as it approaches the threshold planes. Following Plahte, Mestl & Omholt (1998), the behavior of the piecewise-linear equations (4) at the threshold planes is defined as the behavior of the original nonlinear equations (1) in the limit $n \to \infty$ (see also Plahte, Mestl & Omholt (1994)). The definition is motivated by the observation that, as n goes to ∞, the sigmoid function (2) approaches the step function (3).

Given this definition, basically two different things can happen when a trajectory approaches a threshold plane $x_i = \theta_{ji}$. First, the trajectory may be continued by a trajectory in the neighbouring volume, which moves towards a different focal state determined by the volume state equations of the new volume. In this case a *transition* from the volume to its neighbouring volume takes place and the threshold plane is *transparent*. Second, if the focal state of the neighbouring volume is such that the trajectories in that volume also approach the threshold plane $x_i = \theta_{ji}$, no transition between the volumes is possible and the threshold plane is *non-transparent*.[1] Non-transparent threshold planes may indicate additional steady states of the PLDEs located in the threshold planes. Techniques to identify these steady states are presented by Snoussi & Thomas (1993) and by Plahte, Mestl & Omholt (1994; 1998), and will not be considered here.

The global behavior of the PLDEs may be quite complex and is not well understood. Continuations of trajectories in several volumes may give rise to oscillations towards a steady state located at the intersection of threshold planes, cycles, limit cycles, or even chaotic oscillations (for $n \geq 4$) (Glass & Pasternack 1978; Lewis & Glass 1991; Mestl, Plahte, & Omholt 1995b; Mestl, Lemay, & Glass 1996). Numerical simulation studies have shown that in many cases the global behavior of the piecewise-linear systems (4) and nonlinear systems (1) with steep sigmoids exhibit the

[1] Non-transparency of threshold planes is possible when the regulatory system involves autoregulation, a regulatory interaction of a gene with itself (like gene 3 in Fig. 1(a)).

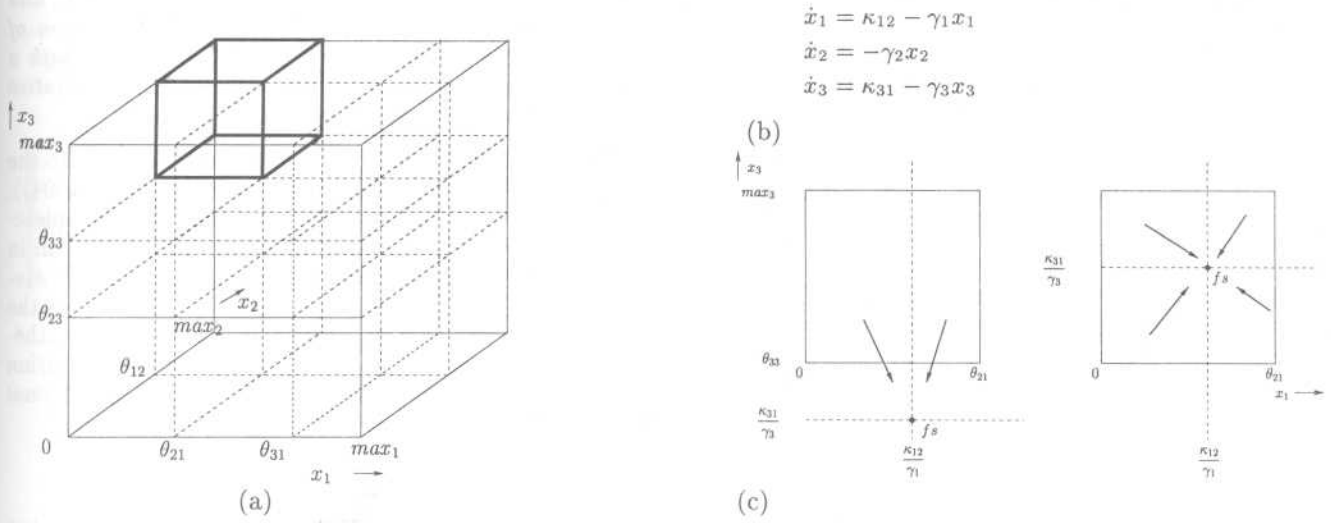


Figure 3: (a) The phase space box of the model in Fig. 1, divided into $2 \cdot 3 \cdot 3 = 18$ volumes by the threshold planes. (b) The state equations for volume $0 \leq x_1 < \theta_{21}$, $\theta_{12} < x_2 \leq max_2$, and $\theta_{33} < x_3 \leq max_3$ (the volume demarcated by bold lines). (c) The focal state of this volume projected on the $x_1 - x_3$ plane. Depending on whether $\kappa_{31}/\gamma_3 > \theta_{33}$ or $\kappa_{31}/\gamma_3 < \theta_{33}$, the focal state lies inside or outside the volume.

same qualitative properties (Glass & Kauffman 1973; Thomas & d'Ari 1990; Plahte, Mestl, & Omholt 1994).

Qualitative simulation method

Our method performs a qualitative simulation of regulatory systems described by state equations of the form (4). The basic idea underlying the method is to determine, in an iterative way, all volumes that are reachable from an initial volume through successive volume transitions. Which volumes are reachable depends on the threshold inequalities (6) and nullcline inequalities (8), as these parameter constraints determine the qualitative dynamics of the system. The state equations, threshold inequalities, nullcline inequalities, and initial volume together form a *qualitative simulation problem (QSP)*. In what follows we assume that it has been possible to specify the elements of a QSP, on the basis of prior knowledge or educated guesses. In the concluding section we will reconsider this assumption.

Consider a volume defined in the ith dimension by two consecutive thresholds $\sigma_i^{(l_i)}$ and $\sigma_i^{(l_i+1)}$. The inequalities

$$\sigma_i^{(l_i)} < x_i < \sigma_i^{(l_i+1)} \tag{9}$$

form the *qualitative value* of x_i, denoted by qv_i. In addition to the qualitative value for x_i, we have a qualitative value $\dot{qv}_i$ for $\dot{x}_i$, being one of the following three inequalities

$$\dot{x}_i > 0, \ \dot{x}_i < 0, \text{ or } \dot{x}_i \lesseqgtr 0. \tag{10}$$

If the nullcline plane for x_i lies outside the volume, i.e., $\mu_i/\gamma_i < \sigma_i^{(l_i)}$ or $\mu_i/\gamma_i > \sigma_i^{(l_i+1)}$, the qualitative value will be $\dot{x}_i > 0$ or $\dot{x}_i < 0$, respectively, everywhere in the volume. If the nullcline runs through the volume, i.e., $\sigma_i^{(l_i)} < \mu_i/\gamma_i < \sigma_i^{(l_i+1)}$, it holds that $\dot{x}_i < 0$ on one side of the nullcline plane, $\dot{x}_i > 0$ on the other side, and $\dot{x}_i = 0$ in the nullcline plane. The qualitative value of $\dot{x}_i$ in the volume is then written as $\dot{x}_i \lesseqgtr 0$.

Given a volume with a vector $\boldsymbol{qv}$ of qualitative values for $\boldsymbol{x}$, $\dot{\boldsymbol{qv}}$ can be easily inferred from the equations and inequalities (5)-(9) by means of basic algebraic rules. As a consequence of the orthogonality of the volume state equations, this can be done separately in each dimension, thus requiring only $\mathcal{O}(n)$ inferences. For the volume emphasized in Fig. 3(a), for example, we find the vector $[\dot{x}_1 \lesseqgtr 0, \dot{x}_2 < 0, \dot{x}_3 < 0]$. Take $\dot{x}_3$. From the volume state equation $\dot{x}_3 = \kappa_{31} - \gamma_3 x_3$ and the nullcline inequality $\kappa_{31}/\gamma_3 < \theta_{23}$ we obtain $\dot{x}_3 < \gamma_3\theta_{23} - \gamma_3 x_3$. As $x_3 > \theta_{33}$ in the volume, and $\theta_{33} > \theta_{23}$, it follows that $\gamma_3\theta_{23} - \gamma_3 x_3 < 0$, and hence $\dot{x}_3 < 0$.

The trajectories in a volume move towards the focal state lying at the intersection of the nullclines $x_i = \mu_i/\gamma_i$. Knowing the qualitative values $\boldsymbol{qv}$ and $\dot{\boldsymbol{qv}}$ for a volume allows one to determine the *a priori* possible transitions to neighbouring volumes. The function *succ* in Table 1 defines the possible successor qualitative values for x_i. The successor relations are motivated by basic contuinity restrictions, as in Kuipers (1994). By combining the successor qualitative values for all x_i, we obtain a set of *candidate successor volumes*. More formally, a volume v' defined by $\boldsymbol{qv'}$ is a candidate successor of v, if $\boldsymbol{qv'} \neq \boldsymbol{qv}$ and $qv_i' \in succ(qv_i)$, $1 \leq i \leq n$.

Concentrations of gene products may reach their thresholds simultaneously, since more than one x_i can change its qualitative value in a transition. As a conse-

quence, a volume may have as much as $2^n - 1$ candidate successors.

qv_i and $\dot{qv}_i$	$succ(qv_i, \dot{qv}_i)$
$\sigma_i^{(l_i)} < x_i < \sigma_i^{(l_i+1)}$, $\dot{x}_i > 0$	$\{\sigma_i^{(l_i)} < x_i < \sigma_i^{(l_i+1)}$, $\sigma_i^{(l_i+1)} < x_i < \sigma_i^{(l_i+2)}\}$
$\sigma_i^{(l_i)} < x_i < \sigma_i^{(l_i+1)}$, $\dot{x}_i < 0$	$\{\sigma_i^{(l_i)} < x_i < \sigma_i^{(l_i+1)}$, $\sigma_i^{(l_i-1)} < x_i < \sigma_i^{(l_i)}\}$
$\sigma_i^{(l_i)} < x_i < \sigma_i^{(l_i+1)}$, $\dot{x}_i \lesseqgtr 0$	$\{\sigma_i^{(l_i)} < x_i < \sigma_i^{(l_i+1)}\}$

Table 1: The function *succ* maps qualitative values for x_i and $\dot{x}_i$ to possible successor qualitative values for x_i. It is here assumed that $1 < l_i < p_i - 1$, but generalization of the table to other cases is straightforward.

For a candidate successor v' to be actually *reachable* from v, the threshold plane separating v and v' must be transparent. This implies that the qualitative values of $\dot{x}_i$ before and after the transition, $\dot{qv}_i$ and $\dot{qv}'_i$, should not be opposite for those x_i changing their qualitative value in the transition. That is, if $qv'_i \neq qv_i$, then $\dot{qv}'_i = \dot{qv}_i$, $1 \leq i \leq n$. In order to test the reachability of a candidate successor volume, it suffices to determine $\dot{\boldsymbol{qv}}'$ and compare the qualitative values before and after the transition.

A volume may thus have no reachable successors for one of two reasons. It either contains a steady state, defined by $\dot{x}_i \lesseqgtr 0$ for all i, or the outgoing trajectories approach non-transparent threshold plane(s).

The simulation algorithm iteratively generates, in a depth-first manner, all volumes that are reachable from an initial volume v_{init} defined by qualitative values $\boldsymbol{qv}_{init}$.

```
push(stack, v_init)
determine qv̇_init
while not stack is empty do
    current volume v ← pop(stack)
    generate candidate successor volumes of v
    for all candidate successors v' do
        determine qv̇'
        if v' is reachable
        then if not v' has been reached and
                not v' on stack
             then push(stack, v')
```

The volumes and their reachable successors form a directed *transition graph*. The graph may contain volumes without successors and volume cycles, which will be together referred to as *attractors*. In the worst case, the algorithm will generate $\mathcal{O}((p+1)^n)$ reachable volumes, where p is the maximum number of genes influenced by a single gene.

An initial value problem (IVP), consisting of a model of the form (4) with numerical values for the parameters and initial values $\boldsymbol{x}(t_0)$, can be transformed into a QSP. In addition, the numerical solution of an IVP on an interval $[t_0, t]$ can be abstracted into a sequence of volumes. These abstraction relations underlie the following guarantee on the output of the algorithm.

Theorem 1 *Given a QSP abstracted from an IVP, and a sequence of volumes abstracted from the solution of the IVP. The sequence of volumes corresponds with a path in the transition graph generated by the qualitative simulation algorithm applied to the QSP.*

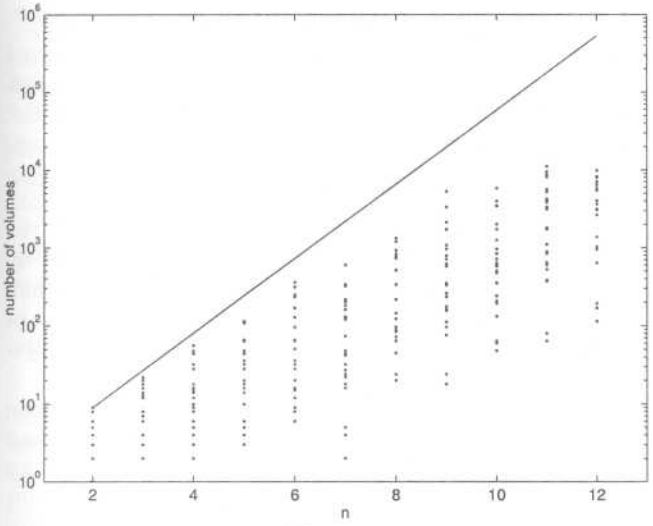
The proof will be omitted, as it is much similar to the soundness proof of QSIM (Theor. 6 in Kuipers (1994)). As in qualitative reasoning more generally, completeness cannot be guaranteed. That is, not every path in the transition graph generated by the algorithm corresponds with a sequence of volumes abstracted from the solution of some IVP. In some cases, additional mathematical constraints will rule out paths in the transition graph as being impossible (Olivier Bernard, personal communication).

Experimental results

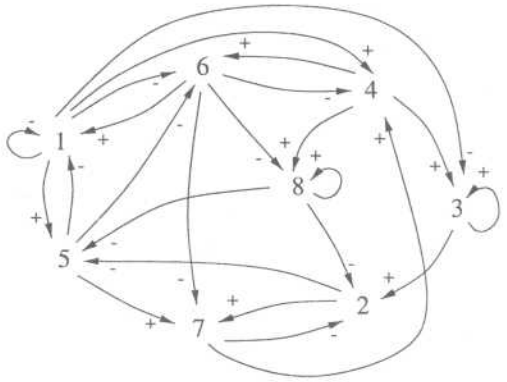
The simulation algorithm predicts the attractors that may be reached from an initial volume following a sequence of volume transitions. From a biological point of view, this means that possible functional states of the regulatory system are predicted given certain initial gene expression levels (Kauffman 1993; Thomas & d'Ari 1990). As shown in the previous section, the number of reachable volumes theoretically grows in an exponential fashion. This compromises the objective to deal with larger-scale regulatory systems. In order to test whether the dynamics of PLDEs of the form (4) exhibits more favorable average-case behavior, we have performed a series of computer experiments.

The experiments have been carried out with an implementation of the simulation method in Java 1.2. The program reads and parses input files with the equations and inequalities specifying a QSP. From this information it produces tables containing all possible qualitative values for x_i and $\dot{x}_i$, and successor relations between qualitative values. The tables are employed in the generation of successor volumes in the main loop of the algorithm. The core of the program consists of an inequality reasoner for determining the qualitative value of every $\dot{x}_i$ in a volume. We have developed a version of Simmons' (1986) Quantity Lattice, adapted to the particularities of the class of PLDEs we are dealing with. The output produced by the program consists of a tabular representation of the volume transition graph, a list of attractors, and run-time statistics. The simulations reported below were run on a SUN Ultra 10 workstation with 128 Mb of RAM.

As a first test, we have simulated three genetic regulatory networks described in the literature. The networks concern λ phage growth control (Thieffry & Thomas 1995) and the regulation of arginine synthesis (Thomas, Thieffry, & Kaufman 1995) ($n = 2, 3, 4$). (In fact, the model in Fig. 1 is a slightly more complex variant of the latter model.) We have compared our results with those obtained by the logical method of Thomas and colleagues and found good agreement, bearing in mind the differences between the methods (see below).



(a)



(b)

Figure 4: (a) The number of reachable volumes from an initial volume for models with $k = 2$ and $n = 2, \ldots, 12$. Each dot in the plot represents a simulation. (b) An example regulatory network for $n = 8$ and $k = 3$.

In order to study the upscaling properties of the method in a systematic way, we have carried out simulation experiments with random regulation networks. For each of the n genes in a network, k inputs were randomly chosen among the other genes. Next, the functions f_i were randomy selected from the set of possible functions with k inputs. No restrictions were put on the form of the network, which allows complex and interlocking feedback structures to be generated. Further, a random order between the thresholds of the regulation functions was generated, as well as a lower and upper threshold bound for the possible nullcline terms. Each of the models thus obtained was simulated from a randomly-selected initial volume in the phase space.

The results of experiments with $2 \leq n \leq 12$ and $k = 2$ are shown in Fig. 4(a). For each n, 25 simulations were carried out, each with a different model and initial volume. The number of volumes reachable from the initial volume is displayed as a function of the size of the network. The most important observation to be made is that the average-case behavior is much more favorable than the worst-case behavior, shown as the drawn line in the figure (notice the logarithmic scale of the y-axis). For $n = 12$, the median number of volumes reachable is 4016, less than 1% of the total number of volumes in the phase space.

The figure shows a large spread in the simulation results. Simulations for $n = 10$ give results varying from 10 to 5840 reachable volumes, although most of the time between 10^2 and 10^3 volumes are generated. A number of factors contribute to these differences, in particular the distance of the initial volume to the reachable attractors and the number and the size of the attractors. The number of attractors reached also strongly varies between simulations. For $n = 8$, about 9 attractors are reached on average, mostly cycles. This number is somewhat misleading, though, as two simulations yield 102 and 42 cycles.

For $n > 12$, over one third of the simulations take more than two hours to complete and some end with a memory overflow due to the large number of successors generated. This seldom happens when $9 \leq n < 12$, and never for lower n. We have also carried out simulations for $k = 3$ and $k = 4$, that is, for more densely connected networks. In these cases, the number of reachable volumes theoretically grows as $\mathcal{O}(4^n)$ and $\mathcal{O}(5^n)$. As for $k = 2$, the average-case behavior tended to be more favorable. However, for $n > 9$ simulations start to become intractable with the current implementation.

Fig. 4(b) shows an example of a network with $n = 8$ and $k = 3$ and a large number of positive and negative feedback loops. The model is defined by a total of 120 equations and inequalities. Simulated from a random initial volume, 3892 volumes turn out to be reachable. The trajectories either end in the single volume with a steady state or in one of the 14 cycles.

Discussion

The method for qualitative simulation of genetic regulatory systems presented in this paper has been shown capable of dealing with networks of larger size and complexity than possible with existing QR methods. We have modeled regulatory systems by a class of differential equations putting strong constraints on the behavior in volumes of the phase space, in combination with a simulation algorithm adapted to these equations. Currently we are able to deal with networks of up to 12 genes with 2 to 4 regulators per gene and complicated feedback structures. The simulation stud-

ies described here present one of the first attempts to systematically investigate upscaling of QR methods in the context of a realistic application (see also Struss (1997)). The simulation method has been tailored to one class of models, relevant to genetic and other biological regulatory systems (Plahte, Mestl, & Omholt 1995; Muraille *et al.* 1996). The principles underlying our approach, however, might be applicable in other domains as well.

Adaptation to a specific class of models is the principal respect in which the approach presented in this paper differs from well-known QR methods like QSIM and QPT (Kuipers 1994; Forbus 1984). The expressivity and generality of the formalism have been traded for the capability to deal with larger and more complex systems. For instance, the description of the state of a regulatory system is achieved on a higher level of abstraction. The basic element in our formalism is a volume, defined by a vector of qualitative values $\sigma_i^{(l_i)} < x_i < \sigma_i^{(l_i+1)}$. In QSIM one would have to distinguish individual states inside a volume as well, such as boundary states defined by $x_i = \sigma_i^{(l_i)}$ or $x_i = \sigma_i^{(l_i+1)}$, and nullcline states defined by $\dot{x}_i = 0$. The method presented here thus abstracts from trajectories inside a volume, which among other things allows a more compact representation of the behavior of the system.

Qualitative methods for the analysis of genetic regulatory systems have been developed in mathematical biology as well, the best-known example being Boolean networks (Kauffman 1993). Simulation of Boolean networks rests on the assumption that a gene is either active or inactive, and that genes change their activation state synchronously. Translated to the formalism of this paper, this means that there is only one threshold per gene and that thresholds are reached simultaneously. For many purposes these assumptions are too strong. The use of random networks to study the upscaling properties of the method has been stimulated by Kauffman's (1993) simulation studies with Boolean networks. The observation that trajectories remain localized in a small part of the phase space agrees with the results obtained for Boolean networks.

Thomas and colleagues (Thomas & d'Ari 1990; Thomas, Thieffry, & Kaufman 1995) have proposed a generalized logical formalism that permits multivalued activation states and asynchronic transitions. In fact, Snoussi has demonstrated that their formalism can be seen as an abstraction of a special case of (4), allowing additions but not multiplications of regulation functions. The logical method of Thomas focuses on the identification of steady states, including those located on threshold planes, rather than on simulation from an initial volume. The use of logical equations abstracting from differential equations makes it difficult to integrate (semi-)quantitative information (Berleant & Kuipers 1997). With the advent of cDNA microarrays and other new measurement technologies, (semi-)quantitative gene expression data is becoming available in large amounts (Brown & Botstein 1999).

Further upscaling of our method might be achieved by optimizing the code of the implementation, something that has not been seriously undertaken thus far. More fundamentally, additional simplifying assumptions could be introduced that reduce the complexity of the algorithm. When synchronous transitions of qualitative values are not allowed, for instance, the number of candidate successor volumes generated by the algorithm would be of the order $\mathcal{O}(n)$ rather than $\mathcal{O}(2^n)$. This assumption implies that concentrations of gene products will never reach their threshold simultaneously.

In order to simulate a genetic regulatory system, parameter constraints in the form of threshold and nullcline inequalities need to be available. The paucity of information on model parameters may not permit such constraints to be specified. This brings a related, and in many cases more relevant problem to the fore: given observed gene expression patterns, is it possible to find threshold and nullcline inequalities such that the model of a hypothesized regulatory network yields predictions consistent with expression data? The simulation method presented in this paper forms the core of a system currently under development called the *Genetic Network Analyzer (GNA)*, which will address such model validation questions.

Acknowledgements The authors would like to thank Olivier Bernard, Jacques Cohen, Hans Geiselmann, Sébastien Maza, François Rechenmann, Denis Thieffry, and Ivayla Vatcheva for their contributions to this paper.

References

Berleant, D., and Kuipers, B. 1997. Qualitative and quantitative simulation: Bridging the gap. *Artif. Intell.* 95:215–256.

Brown, P., and Botstein, D. 1999. Exploring the new world of the genome with DNA microarrays. *Nat. Genet.* 21(supplement):33–37.

Forbus, K. 1984. Qualitative process theory. *Artif. Intell.* 24:85–168.

Glass, L., and Kauffman, S. 1973. The logical analysis of continuous non-linear biochemical control networks. *J. Theor. Biol.* 39:103–129.

Glass, L., and Pasternack, J. 1978. Stable oscillations in mathematical models of biological control systems. *J. Math. Biol.* 6:207–223.

Glass, L. 1975. Combinatorial and topological methods in nonlinear chemical kinetics. *J. Chem. Phys.* 63(4):1325–1335.

Heidtke, K., and Schulze-Kremer, S. 1998. Design and implementation of a qualitative simulation model of λ phage infection. *Bioinformatics* 14(1):81–91.

Karp, P. 1993. Design methods for scientific hypothe-

sis formation and their application to molecular biology. *Machine Learning* 12:89–116.

Kauffman, S. 1993. *The Origins of Order.* Oxford University Press.

Kuipers, B. 1994. *Qualitative Reasoning.* MIT Press.

Lewin, B. 1997. *Genes VI.* Oxford University Press.

Lewis, J., and Glass, L. 1991. Steady states, limit cycles, and chaos in models of complex biological networks. *Int. J. Bifurcation Chaos* 1(2):477–483.

Mestl, T.; Lemay, C.; and Glass, L. 1996. Chaos in high-dimensional neural and gene networks. *Physica D* 98:33–52.

Mestl, T.; Plahte, E.; and Omholt, S. 1995a. A mathematical framework for describing and analysing gene regulatory networks. *J. Theor. Biol.* 176:291–300.

Mestl, T.; Plahte, E.; and Omholt, S. 1995b. Periodic solutions in systems of piecewise-linear differential equations. *Dyn. Stab. Syst.* 10(2):179–193.

Muraille, E.; Thieffry, D.; Leo, O.; and Kaufman, M. 1996. Toxicity and neuroendocrine regulation of the immune response: A model analysis. *J. Theor. Biol.* 183:285–305.

Plahte, E.; Mestl, T.; and Omholt, S. 1994. Global analysis of steady points for systems of differential equations with sigmoid interactions. *Dyn. Stab. Syst.* 9(4):275–291.

Plahte, E.; Mestl, T.; and Omholt, S. 1995. Stationary states in food web models with threshold relationships. *J. Biol. Syst.* 3(2):569–577.

Plahte, E.; Mestl, T.; and Omholt, S. 1998. A methodological basis for description and analysis of systems with complex switch-like interactions. *J. Math. Biol.* 36:321–348.

Simmons, R. 1986. 'Common sense' arithmetic reasoning. In *Proc. AAAI-86*, 528–532. Morgan Kaufmann.

Snoussi, E., and Thomas, R. 1993. Logical identification of all steady states. *Bull. Math. Biol.* 55(5):973–991.

Snoussi, E. 1989. Qualitative dynamics of piecewise-linear differential equations. *Dyn. Stab. Syst.* 4(3-4):189–207.

Struss, P. 1997. Fundamentals of model-based diagnosis of dynamic systems. In *Proc. IJCAI-97*, 480–485. Morgan Kaufmann.

Thieffry, D., and Thomas, R. 1995. Dynamical behaviour of biological networks: II. Immunity control in bacteriophage lambda. *Bull. Math. Biol.* 57(2):277–297.

Thomas, R., and d'Ari, R. 1990. *Biological Feedback.* CRC Press.

Thomas, R.; Thieffry, D.; and Kaufman, M. 1995. Dynamical behaviour of biological regulatory networks: I. Biological rule of feedback loops. *Bull. Math. Biol.* 57(2):247–276.

Trelease, R.; Henderson, R.; and Park, J. 1999. A qualitative process system for modeling NF-κB and AP-1 gene regulation in immune cell biology research. *Artif. Intell. Med.* 17:303–321.