

A STOCHASTIC EXTENSION OF R. THOMAS REGULATORY NETWORK MODELLING

BARTEK WILCZYŃSKI

*Institute of Mathematics, Polish Academy of Sciences
Śniadeckich 8, 00-956 Warszawa, Poland
E-mail: bartek@impan.gov.pl*

Abstract. In this paper we present the extension of the kinetic logic proposed by René Thomas for analysis of genetic regulatory gene networks. We consider the case with a Gaussian noise added to the regulation function and propose a method of analyzing the resulting model with a discrete time Markov model.

1. Introduction. Understanding the regulatory mechanisms of gene expression is one of the key problems in molecular biology. Since such mechanisms are extremely hard to study *in vivo*, many mathematical models were proposed to help understanding the principles of regulatory network operation. The pioneering work in the field of regulatory network modelling was done in the 1960s by S. Kauffman [2] who showed that such fundamental phenomena of gene regulation as epigenesis and stable convergence can be modelled with a very simple mathematical framework of Boolean networks. This model was extended by René Thomas and co-workers [11, 12] leading to formulation of generalized logical description of regulatory networks.

Generalized logical modeling approach was applied to many experimentally studied biological regulatory circuits (e.g. [6, 10, 5, 4, 7]) showing that this formalism may be useful in regulatory network modelling. However, it is difficult to reconstruct such networks from experimental data. The problem with reconstruction of such networks lies in the lack of a natural scoring function for different models for a given dataset. Even though there are computational ways to effectively simulate such models (i.e. GINSim software package [3]), it doesn't help us with choosing the right model from many possibilities.

Another approach to this problem is using a stochastic dynamical system to model the dependencies of expression levels of genes. Chen et al. [1] managed to reconstruct

2000 *Mathematics Subject Classification*: Primary 92B99; Secondary 37H10.

Key words and phrases: gene networks, stochastic logical networks.

The author receives support from Foundation for Polish Science.

The paper is in final form and no version of it will be published elsewhere.

parameters for a system of stochastic differential equations from Yeast cell cycle expression data [9]. This was possible with the assumption that the expression of genes depends linearly on the expression of its regulators preventing the method to predict correct dependencies in cases where the regulation is non-linear.

In this work we present a novel modelling framework for regulatory networks, called Stochastic Logical Networks (referred to as SLNs). It is based on the formalism of generalized logical description of networks as introduced by Thomas [11] extended with the stochastic factor leading to a simple scoring function based on the likelihood of the model given the observed data.

2. Genetic network modelling

2.1. Gene networks as dynamical systems. In order to describe the formalism of SLNs, we provide a brief introduction into the way of modelling introduced by R. Thomas [11]. It is based on the assumption that a regulatory system can be accurately described as a dynamical system of ordinary differential equations. We treat the state of the cell, i.e. concentrations of all interesting gene products, as a vector of non-negative real values $\vec{v} = \langle v_1, v_2, \dots, v_n \rangle \in R_{\geq 0}^n$, dependent on time t , so the equations have the following form:

$$\frac{\partial v_i}{\partial t} = -v_i \cdot \lambda_i + \mathcal{F}_i(\vec{v}), \quad (1)$$

where $\mathcal{F}_i(\vec{v})$ denotes the production rate of gene v_i depending on the state of all genes, whereas λ_i represents the decay constant responsible for degrading the gene product proportionally to its current concentration. To account for non-linearity and combinatorial nature of the dependence of the production rate on the state of regulators, the production rate of gene i is defined by Snoussi [8] as a linear combination of products of sigmoid functions of expression of regulators:

$$\mathcal{F}_i(\vec{v}) = \sum_{G \subseteq \{1..n\}} I_{G,i} \cdot \prod_{j \in G} S_{i,j}(v_j, \theta_{i,j}), \quad (2)$$

where $I_{G,i} \in \mathbb{R}^+$ is the regulatory influence of the set of regulators G on gene i , $\theta_{i,j} \in \mathbb{R}$ are the activation thresholds and $S_{i,j}$ is the sigmoid activation function of gene i by gene j being one of the following:

$$S^+(x, \theta) = \text{sigm}_k(x - \theta) \quad (3)$$

$$S^-(x, \theta) = 1 - \text{sigm}_k(x - \theta), \quad (4)$$

where $\text{sigm}_k(x) = (1 + e^{-kx})^{-1}$ with a notable case of sigm_∞ equal to the Heaviside step function. Different forms of $S_{i,j}$ represent different possible regulatory interactions. If $S_{i,j} = S^-$ we say that j is a repressor of i , otherwise j is an enhancer of i .

2.2. Qualitative approach. René Thomas observed, that qualitative behaviour of such systems can be modelled as a non deterministic discrete process whose states correspond to discretized states of the original dynamical system. This is due to the fact that the production rates of genes change substantially only around the threshold values $\theta_{i,j}$. If we consider the case with sigm_∞ step function, the hyperplanes $v_j = \theta_{i,j}$ divide the phase space of the dynamical system into a finite set of disjoint parts on which the production

rates of all genes are constant. In such case, the behavior of the system is determined by the choice of production rates of all genes in all discretized states. A simplistic example of a 2-gene negative feedback loop with its phase space, dependency graph and discrete state graph is depicted in Figure 1.

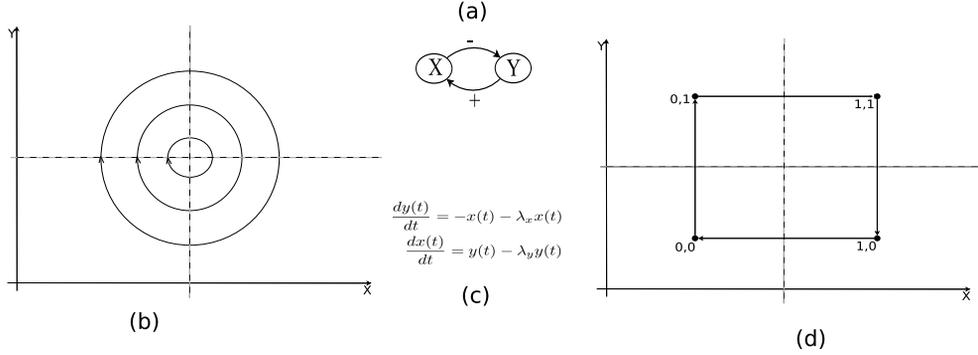


Fig. 1. An example of a dynamical system consisting of two genes: X and Y (a). Its phase space (b), ODEs (c) and state graph (d).

We use the notion of a discretization mapping $\delta(\vec{v}) = \langle \delta_1(v_1) \dots \delta_n \rangle$, where each δ_i is a mapping of the i -th variable into its discrete values using the thresholds $\{\theta_{i,j}\}_{j=1..n}$. We also denote the space of all discrete states by $\Sigma_n = \{0 \dots n\}^n$.

Evolution of the qualitative system can be derived from equations governing the dynamical system. If we assume that $S_{i,j}$ are indeed Heaviside step functions, the value of the regulation functions $\mathcal{F}_i(\vec{v})$ is constant between the thresholds $\theta_{i,j}$. Therefore for each discrete state σ , corresponding to the discretization domain $\delta^{-1}(\sigma)$, there exist constant production rates $F_i(\sigma)$ of all genes, such that:

$$\forall \vec{v} \in \delta^{-1}(\sigma) \mathcal{F}_i(\vec{v}) = F_i(\sigma).$$

However, the Heaviside regulatory functions introduce discontinuity in the right-hand side of ordinary differential equations for points in the state-space where $v_j = \theta_{i,j}$. We call such points *singular*, and exclude them from further analysis. This can be done without loss of generality since the measure of the set of singular points is 0.

After Thomas [11], we use the notion of an image function \mathcal{R} of a discrete state σ :

$$\mathcal{R}(\sigma) = \langle \delta_1(F_1/\lambda_1), \dots, \delta_n(F_n/\lambda_n) \rangle.$$

If a state is the image of itself, we call it *stable*. Otherwise, since there may be different trajectories of the dynamical system traversing this discretization domain, the state succession is non-deterministic. For each discrete non-stable state σ we define a set of *successor states* succ_σ containing all neighbouring discrete states σ' such that there exists a trajectory in the dynamical system going from $\delta^{-1}(\sigma)$ directly to $\delta^{-1}(\sigma')$. We use the notation of $\sigma \rightarrow \sigma'$ to denote the fact that $\sigma' \in \text{succ}_\sigma$. The generalization of a successor state is its transitive closure: the *reachability* relation $\sigma \rightarrow^+ \sigma'$.

2.3. Stochastic logical networks. We propose to introduce a stochastic factor to the differential equations describing regulatory systems. To this aim, we follow the common

practice of adding white noise in the form of an independent Wiener processes $W_i(t)$ scaled for each variable by ϵ_i to the right-hand side of differential equations (1) obtaining the following

$$dv_i = (\mathcal{F}_i(\vec{v}) - v_i \cdot \lambda_i)dt + \epsilon_i dW_i(t). \tag{5}$$

If we use the regulation functions \mathcal{F}_i as defined in equation (2) and apply the same discretization mapping δ we obtain a stochastic system consisting of a finite set of disjoint domains. In each of these domains we observe a multivariate Brownian motion with linear drift – a very well studied mathematical model. What we are interested in, is a discrete stochastic process representing the movements of the Brownian motion between domains through time that correspond to the changes of qualitative behaviour of the whole system.

Because of the singularities in the right-hand side of stochastic equations at domain borders, we cannot analyze the exact dynamics of such systems in general. However, if we limit our analysis to the qualitative behaviour of the discrete system, we may make a simplifying assumption that the discretized process is Markovian. This is analogous to the definition of kinetic logic which assumes that successor states depend only on the current qualitative state of the system. It should be noted here, that at this point we trade the ability to analyze the subtle behaviour that may occur at the boundaries of domains for the simplicity of the discrete model.

For our considerations it is important to find the relationship between the parameters of the discrete process and the dynamical system. We need to answer the question: given a multidimensional Brownian motion with drift and a discretization of its state space into a set of disjoint domains what is the natural probability distribution of exiting from any of the domains to each of its neighbours. It is clear that the probability of Brownian motion exiting from any finite domain is equal to 1 and that the probability of moving directly to any non-adjacent domain is 0. We need to calculate the probabilities of travelling to all adjacent states for any given state.

Let us first examine the simple case of one dimensional Brownian motion with drift

$$\frac{\partial x}{\partial t} = F(x) + \epsilon \frac{\partial w}{\partial t}$$

and a finite closed interval $[a, b]$. We are interested in the probability $p_a(x_0)$ of exiting from $[a, b]$ through the barrier a starting from x_0 . In general, we can determine this by solving the following differential equation

$$\frac{\partial^2 p_a}{\partial x^2} \cdot \frac{\epsilon^2}{2} + F(x) \frac{\partial p_a}{\partial x} = 0$$

assuming that $p_a(a) = 1$ and $p_a(b) = 0$. The solution of this equation depends heavily on F . In our case in each domain of continuity of the regulation function, $F(x) = C - x \cdot \lambda$. In this case we can obtain the following solution:

$$p_a(x) = \frac{\operatorname{erf}\left(\frac{\sqrt{2}(b\lambda - C)}{\epsilon^2\sqrt{-\lambda/\frac{\epsilon^2}{2}}}\right) - \operatorname{erf}\left(\frac{\sqrt{2}(\lambda x - C)}{\epsilon^2\sqrt{-\lambda/\frac{\epsilon^2}{2}}}\right)}{\operatorname{erf}\left(\frac{\sqrt{2}(b\lambda - C)}{\epsilon^2\sqrt{-\lambda/\frac{\epsilon^2}{2}}}\right) - \operatorname{erf}\left(\frac{\sqrt{2}(a\lambda - C)}{\epsilon^2\sqrt{-\lambda/\frac{\epsilon^2}{2}}}\right)}$$

where erf is the error function:

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt.$$

After simplifications we obtain the formula

$$p_a(x) = \frac{\int_x^b e^{-(\lambda(\frac{it}{\epsilon\sqrt{\lambda}} - C))^2} dt}{\int_a^b e^{-(\lambda(\frac{it}{\epsilon\sqrt{\lambda}} - C))^2} dt}.$$

Since the result is not elementary and we would like to have a reversible function we assume that the drift is constant and equal to c (it is reasonable in cases of small intervals). In this case our solution is

$$p_a(x) = \frac{e^{-\frac{2bc}{\epsilon^2}} - e^{-\frac{2xc}{\epsilon^2}}}{e^{-\frac{2bc}{\epsilon^2}} - e^{-\frac{2ac}{\epsilon^2}}}.$$

Now in order to have a solution not depending on the starting point, we integrate it over all possible starting points:

$$\int_a^b p_a(x) dx = \frac{\epsilon^2}{2c} - \frac{b-a}{e^{\frac{2c \cdot (b-a)}{\epsilon^2}} - 1}.$$

2.4. Multidimensional case. Now as we have the solution for one-dimensional case, we shall take multiple dimensions into account. As the stochastic components for different variables are independent, we can consider them as independent in each discretization domain.

Let us introduce, for each variable x_i , the probability $p_{i,a,b}(x, t)$ that the process starting from x will not leave the interval (a, b) in time t . What we want to know is, for each variable, what is the probability that it will be the first to leave the domain,

$$q_i(x) = \int_0^\infty \hat{p}_{i,a_i,b_i}(x, t) \prod_{j \neq i} p_{j,a_j,b_j}(x, t) dt,$$

where $\hat{p}_{i,a,b}(x, t)$ denotes the probability of the i th process reaching one of the points $\{a_i, b_i\}$ at time t starting from x .

In our case, in which the influence of drift is stronger than the influence of stochastic noise, $p_{i,a,b}(x, t)$ can be approximated by the probability $p'_{i,a,b}(x, t)$ of the i th process being in the interval (a, b) at time t . This can be computed from normal cumulative distribution function.

3. Conclusion. We present an approach extending the kinetic logic model of regulatory networks. We describe the continuous regulatory function as Langevin equations and under the simplifying assumption that the discrete process is Markovian, we are able to calculate the discrete state change probabilities using the Fokker-Planck equation. Even though such a simplified model is not suitable for quantitative analysis it may be useful in applications requiring more qualitative description. One such application is the task of network topology reconstruction from data. We have used this model for this task [13] with satisfactory results.

References

- [1] K. Chen, T. Wang, H. Tseng, C. Huang, and C. Kao, *A stochastic differential equation model for quantifying transcriptional regulatory network in Saccharomyces cerevisiae*, *Bioinformatics* 21 (2005), 2883–90.
- [2] S. Kauffman, *Homeostasis and differentiation in random genetic control networks*, *Nature* 224 (1969), 177–178.
- [3] A. Larrinaga, A. Naldi, L. Sanchez, D. Thieffry, and C. Chaouiya, *GINsim: A software suite for the qualitative modelling, simulation and analysis of regulatory networks*, *Biosystems* 84 (2006), 91–100.
- [4] L. Mendoza, D. Thieffry, and E. Alvarez-Buylla, *Genetic control of flower morphogenesis in arabidopsis thaliana: a logical analysis*, *Bioinformatics* 7 (1999), 593–606.
- [5] L. Sanchez and D. Thieffry, *A logical analysis of the Drosophila gap-gene system*, *Journal of Theoretical Biology* 211 (2001), 115–141.
- [6] L. Sanchez and D. Thieffry, *Segmenting the fly embryo: a logical analysis of the pair-rule cross-regulatory module*, *Journal of Theoretical Biology* 224 (2003), 517–537.
- [7] L. Sanchez, J. van Helden, and D. Thieffry, *Establishment of the dorso-ventral pattern during embryonic development of drosophila melanogaster: a logical analysis*, *Journal of Theoretical Biology* 189 (1997), 377–389.
- [8] E. H. Snoussi, *Qualitative dynamics of piecewise-linear differential equations: a discrete mapping approach*, *Dynamics and Stability of Systems* 4 (1989), 189–207.
- [9] P. T. Spellman, G. Sherlock, M. Q. Zhang, V. R. Iyer, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein, and B. Futcher, *Comprehensive identification of cell cycle regulated genes of the yeast saccharomyces cerevisiae by microarray hybridization*, *Molecular Biology of the Cell* 9 (1998), 3273–3297.
- [10] D. Thieffry and L. Sanchez, *Dynamical modelling of pattern formation during embryonic development*, *Current Opinions on Genetic Development* 13 (2003), 326–330.
- [11] R. Thomas, *Boolean formalization of genetic control circuits*, *Journal of Theoretical Biology* 42 (1973), 563.
- [12] R. Thomas and R. D’Ari, *Biological Feedback*, CRC Press, 1990.
- [13] B. Wilczyński and J. Tiuryn, *Regulatory network reconstruction using stochastic logical networks*, *Lecture Notes in Computer Science* 4210 (2006), 142–154. http://dx.doi.org/10.1007/11885191_10.