1

Positive or negative regulatory circuit inference from multilevel dynamics

Élisabeth Remy[1], Paul Ruet[1], and Denis Thieffry[2]

[1] Institut de Mathématiques de Luminy
[2] Institut de Biologie du Développement de Marseille-Luminy
CNRS - Université de la Méditerranée, 163 avenue de Luminy, Case 907, 13288 Marseille Cedex 9 (France)
Email: {remy,ruet}@iml.univ-mrs.fr, thieffry@ibdm.univ-mrs.fr

Summary. In the course of his work on the analysis of genetic regulatory networks represented by signed directed graphs, R. Thomas has proposed that the occurrence of a positive regulatory circuit is a necessary condition for the existence of multiple stable states, whereas a negative circuit is necessary to generate stable oscillations. Here, we formulate and prove a theorem establishing these rules in a multilevel discrete framework.

1.1 Introduction

This paper aims to establish formal relationships between the structure of discrete concurrent systems, in particular related to genetic regulatory networks, with notable dynamical properties. In the course of his work on the analysis of genetic regulatory networks, usually represented by signed directed graphs (*regulatory graphs*), R. Thomas has proposed two general rules stating that the occurrence of a positive regulatory circuit in the regulatory graph is a necessary condition for the occurrence of multiple stable states, whereas a negative circuit is necessary to generate stable oscillations [10]. Originally, Thomas was using a discrete approach as most experimental data available are qualitative, and non-linear regulatory effects ubiquitous. Although induced from discrete modeling applications, these rules were initially demonstrated in a continuous context (see [8] and references therein).

In the present paper, we enunciate and demonstrate two theorems establishing these rules in a multilevel discrete framework. Our approach leans on previous demonstrations of similar conjectures in the simpler Boolean case [4] and involves the definition of Boolean projections of the multilevel dynamics. Remark that a generalisation of [4] to the existence of positive circuits in multilevel dynamics has been independently proposed in [6].

We illustrate our definitions and results through the example of the bacteriophage λ.

1.2 Structure and dynamics of genetic regulatory graphs

1.2.1 Regulatory graphs and parameterisations

Let $n \in \mathbb{N}$. A *regulatory graph* $G = (\{1, \dots, n\}, E, \sigma)$ is a simple directed graph with vertex set $\{1, \dots, n\}$ and edge set E, together with a map σ providing each edge with a sign, $+1$ or -1. Intuitively, in the case of genetic networks, the vertices of this graph represent genes, whereas edges represent regulatory interactions between genes. For each $i \in \{1, \dots, n\}$, E_i denotes the set of edges with source i.

Given a regulatory graph $G = (\{1, \dots, n\}, E, \sigma)$, a *parameterisation* of G is a triple (m, ℓ, K) where:

- $m = (m_1, \dots, m_n)$ is an n-tuple of (non-zero) positive integers;
- $\ell : E \to \mathbb{N}$ is a disjoint union of surjective maps $\ell_1, \dots, \ell_n$ with $\ell_i : E_i \to \{1, \dots, m_i\}$;
- $K = (K_1, \dots, K_n)$ is an n-tuple of functions $K_j : \mathcal{F} \to \{0, \dots, m_j\}$, where $\mathcal{F}$ is the set of subsets of $\{(i, s) \text{ s.t. } i \in \{1, \dots, n\}, s \in \{1, \dots, m_i\}\}$.

For each gene i, m_i represents the maximal expression level of i. The function ℓ associates to each edge e the threshold above which the interaction is operating, in the sense that it has some effect on the dynamics defined below. A regulatory graph G and a parameterisation (m, ℓ, K) of G are *coherent* when:

- if G has a positive (resp. negative) edge e from i to j, then there exists a subset X of $\mathcal{F}$ not containing $(i, \ell(e))$ such that $K_j(X) < K_j(X \cup \{(i, \ell(e))\})$ (resp. $K_j(X) > K_j(X \cup \{(i, \ell(e))\})$);
- if there exist $s \in \{1, \dots, m_i\}$ and a subset X of $\mathcal{F}$ not containing (i, s) such that $K_j(X) < K_j(X \cup \{(i, s)\})$ (resp. $K_j(X) > K_j(X \cup \{(i, s)\})$), then G has a positive (resp. negative) edge e from i to j and $\ell(e) \leqslant s$.

If $I \subseteq \{1, \dots, n\}$, an *$I$-circuit* is a circuit $(e_1, \dots, e_k)$ whose sources are in I. The *sign of a circuit* is the product of the signs of its edges.

1.2.2 Dynamics

Given $m = (m_1, \dots, m_n)$ an n-tuple of positive integers, a state of the system consisting in the n genes $\{1, \dots, n\}$ is an element of the *phase space* $S_m = \prod_{i=1}^{n} \{0, \dots, m_i\}$. In the case of genetic networks, $x = (x_1, \dots, x_n) \in S_m$ denotes the expression levels of the genes.

Given a parameterisation (m, ℓ, K) of a regulatory graph G, we need to specify the rules to compute the dynamics of the system, i.e., the allowed transitions between states. Given a state $x \in S_m$ and $i \in \{1, \dots, n\}$, the *operating input set* $X(x,i)$ of i at x is the set of pairs $(j,\ell(e))$ where e is a (unique) edge from j to i such that $x_j \geqslant \ell(e)$.

Consider the map $f : S_m \to S_m$, $f(x) = (f_1(x),\dots,f_n(x))$, defined by $f_i(x) = K_i(X(x,i))$. This map is used to define the nondeterministic *asynchronous dynamics*, which is a graph with vertex set S_m and with an edge from x to x' when there exists i such that $f_i(x) \neq x_i$ and

$$x'_j = \begin{cases} x_j & \text{for any } j \neq i \\ x_i + 1 & \text{if } i = j \text{ and } f_i(x) > x_i \\ x_i - 1 & \text{if } i = j \text{ and } f_i(x) < x_i. \end{cases}$$

Consequently, the expression level of at most one gene is updated at each step. Other dynamics can be considered, including the (deterministic) synchronous dynamics where all the expression levels x_i are simultaneously updated towards $f_i(x)$ in one step. However the synchronous dynamics is not biologically realistic (see [4] and the discretisation of piecewise-linear differential equations [11]), and we therefore consider only the case of the asynchronous dynamics.

Observe that for any $i,j \in \{1,\dots,n\}$, there is at most one edge from j to i, hence there are two possible situations:

1. If there is an edge e from j to i, then for some $x \in \{0,1\}^n$, the function $\chi \mapsto f_i(x^{j,\chi})$, where
$$x^{j,\chi}_k = \begin{cases} x_k & \text{for } k \neq j \\ \chi & \text{if } k = j, \end{cases}$$
takes a constant value v for $\chi < \ell(e)$ and a constant value $w \neq v$ for $\chi \geqslant \ell(e)$. Intuitively, f_i is "Heaviside-like" in x_j with jump at $\ell(e)$.
2. If there is no edge from j to i, then the function $\chi \mapsto f_i(x^{j,\chi})$ is constant.

We are particularly interested in the delineation of asymptotic attractors. A *stable state* is an $x \in S_m$ such that $f(x) = x$. For $x \in S_m$ and $I \subseteq \{1,\dots,n\}$, let $x[I]$ be the subset of S_m consisting in those y such that $y_i = x_i$ for $i \notin I$. Such an $x[I]$ is called a *stable subspace* when $y \in x[I]$ implies $f(y) \in x[I]$. A *cycle* is a sequence of states $(x^1,\dots,x^r)$ such that for each $i = 1,\dots,r$, the dynamics contains an edge from x^i to x^{i+1}. Indices are taken here modulo r, i.e., $r+1 = 1$. A cycle $(x^1,\dots,x^r)$ is completely described by one of its vertices, say x^1, and its *strategy*, which is the map $\varphi : \{1,\dots,r\} \to \{1,\dots,n\}$ such that for each i, $(x^{i+1})_j = (x^i)_j$ for $j \neq \varphi(i)$ and $(x^{i+1})_{\varphi(i)} = (f(x^i))_{\varphi(i)}$. A cycle $(x^1,\dots,x^r)$ with strategy φ is said to be a *trap cycle* when, once in the cycle, one cannot escape any more, i.e., for all $i = 1,\dots,r$ and $j \neq \varphi(i)$, $f_j(x^i) = (x^i)_j$.

1.2.3 Example of the bacteriophage λ

The bacteriophage lambda is a virus infecting E. coli bacteria. Upon infection, it can integrate the bacterial genome to be duplicated with it through normal cell divisions (lysogenic state). Alternatively, it can divert bacterial resources to rapidly multiply and finally provoke the lysis of the host, leading to the release of a hundred of infectious bacteriophage (lytic cycle). The decision between lysis and lysogeny is controlled by a complex regulatory network, involving six viral regulatory genes, as well as several bacterial products (mainly proteases). A multilevel discrete model for Lambda lysis/lysogeny decision has been proposed in [9], focusing on the regulatory graph involving the genes CI and Cro. This model relies on [3], a nice synthesis of available experimental data. (The reader may consult [1] for a stochastic model of the same system.) The main interactions between these regulatory products are schematized in Figure 1.1, where genes are represented by their names CI, Cro instead of numbers $1, 2$.

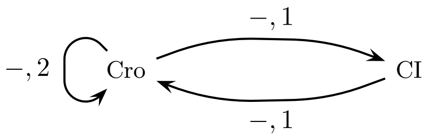

Fig. 1.1. Bacteriophage λ regulatory graph.

A coherent parameterisation (m, ℓ, K) of this 2-genes regulatory graph G has been proposed in [9]. This parameterisation associates a Boolean variable with CI and a ternary variable with Cro: $m_{\mathrm{CI}} = 1, m_{\mathrm{Cro}} = 2$. The values of ℓ can be read on Figure 1.1. The non-zero values of K are $K_{\mathrm{CI}}(\varnothing) = 1$, $K_{\mathrm{Cro}}(\varnothing) = 2, K_{\mathrm{Cro}}(\{(\mathrm{Cro}, 2)\}) = 1$. The asynchronous dynamics associated to this parameterised regulatory graph is:

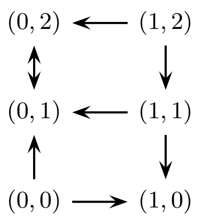

Fig. 1.2. Asynchronous dynamics of the parammeterised graph given in Figure 1.1. The horizontal axis represents the expression level of CI, and the vertical axis the expression level of Cro.

It encompasses two attractors:

- a stable state $(1,0)$, with CI as the only product expressed, a situation corresponding to the lysogenic pathway;
- a trap cycle between the states $(0,1)$ and $(0,2)$, corresponding to an oscillation of the expression level of Cro.

1.3 The Boolean case

In the Boolean case, it is natural to associate a regulatory graph to each x in the phase space S_m, which reduces to $\{0,1\}^n$: for $x \in \{0,1\}^n$, $x_i = 1$ when gene i is expressed, 0 otherwise. For $\beta \in \{0,1\}$, we define $\overline{\beta}$ by $\overline{0} = 1$ and $\overline{1} = 0$. For $x \in \{0,1\}^n$ and $I \subseteq \{1,\dots,n\}$, $\overline{x}^I \in \{0,1\}^n$ is defined by $(\overline{x}^I)_i = x_i$ for $i \notin I$ and $(\overline{x}^I)_i = \overline{x_i}$ for $i \in I$. When $I = \{i\}$ is a singleton, $\overline{x}^{\{i\}}$ is denoted by $\overline{x}^i$.

A cycle $(x^1,\dots,x^r)$ with strategy φ is such that $x^{i+1} = \overline{x^i}^{\varphi(i)}$. It is a trap cycle when $f(x^i) = \overline{x^i}^{\varphi(i)}$.

To $f : \{0,1\}^n \to \{0,1\}^n$ and $x \in \{0,1\}^n$, we associate a regulatory graph $\mathcal{G}(f)(x)$ with an edge from j to i when

$$f_i(\overline{x}^j) \neq f_i(x),$$

with positive sign when $x_j = f_i(x)$, and negative sign otherwise. Note that the adjacency matrix of $\mathcal{G}(f)(x)$ (forgetting the signs) corresponds to the *discrete Jacobian matrix* of f at x, as defined in [7]. The second condition states that the edge is positive when the values x_j and $f_i(x)$ either both increase or both decrease.

If G is a regulatory graph and $I \subseteq \{1,\dots,n\}$, the *restriction of* G *to* I is the regulatory graph obtained from G by removing any vertex not in I and any edge whose source or target is not in I.

In this Boolean context, Thomas' rules can be reformulated as follows:

Theorem 1. *Let* $f : \{0,1\}^n \to \{0,1\}^n$.

1. *If f has at least two stable states, then there is an $x \in \{0,1\}^n$ such that $\mathcal{G}(f)(x)$ has a positive circuit. More precisely, if f has two stable states a and b, and if I is such that $b = \overline{a}^I$, then there is an $x \in \{0,1\}^n$ such that $\mathcal{G}(f)(x)$ has a positive I-circuit.*

1'. *More generally, if f has at least two stable subspaces, then there is an $x \in \{0,1\}^n$ such that $\mathcal{G}(f)(x)$ has a positive circuit.*

2. *If f has a trap cycle $(x^1,\dots,x^r)$ with strategy φ, then $\mathcal{G}(f)(x^1) \cup \cdots \cup \mathcal{G}(f)(x^r)$ has a negative I-circuit with $I = \{\varphi(1),\dots,\varphi(r)\}$.*

These results have been proved in [4] (cases 1 and 2) and [5] (case 1').

1.4 The multilevel case

In this context, Thomas' rules can be restated as follows:

Theorem 2. *Let (m, ℓ, K) be a coherent parameterisation of a regulatory graph G.*

1. *If the associated dynamics has at least two stable states, then G has a positive circuit. More precisely, if it has two stable states a and b and if I is such that $b_i = a_i$ for any $i \notin I$, then G has a positive I-circuit.*

1'. *More generally, if the dynamics has at least two stable subspaces, then G has a positive circuit.*

2. *If the dynamics has a trap cycle $(x^1, \ldots, x^r)$ with strategy φ, then G has a negative I-circuit with $I = \{\varphi(1), \ldots, \varphi(r)\}$.*

To prove this theorem, we first define Boolean projections of the multilevel dynamics, i.e., maps from S_m to $\{0,1\}^n$. The graphs obtained from the projected Boolean dynamics as in Section 1.3 are subgraphs of the multilevel regulatory graph G.

Definition 1. *Given $m = (m_1, \ldots, m_n)$ an n-tuple of (non-zero) positive integers and $s \in \prod_{i=1}^n \{1, \ldots, m_i\}$, define $\pi_s : S_m \to \{0,1\}^n$ by*

$$(\pi_s(x))_i = \begin{cases} 0 & \text{if } x_i < s_i \\ 1 & \text{if } x_i \geqslant s_i. \end{cases}$$

Given such an s, let t_i^0, t_i^1 be such that $0 \leqslant t_i^0 < s_i \leqslant t_i^1 \leqslant m_i$ for each $i = 1, \ldots, n$, and let $\sigma_t : \{0,1\}^n \to S_m$ be defined by $(\sigma_t(z))_i = t_i^{z_i}$; σ_t is then a section of π_s: $\pi_s \circ \sigma_t$ is the identity.

Definition 2. *If (m, ℓ, K) is a coherent parameterisation of a regulatory graph G and s, t are as in Definition 1, define the* projected Boolean dynamics *$f_{s,t} = \pi_s \circ f \circ \sigma_t : \{0,1\}^n \to \{0,1\}^n$.*

Lemma 1. *If (m, ℓ, K) is a coherent parameterisation of G, s, t are as in Definition 1 and $z \in \{0,1\}^n$, then $G \supseteq \mathcal{G}(f_{s,t})(z)$.*

Proof — If $\mathcal{G}(f_{s,t})(z)$ has an egde from j to i, then $(f_{s,t}(z))_i \neq (f_{s,t}(\overline{z}^j))_i$. This implies that $f_i(\sigma_t(z)) \neq f_i(\sigma_t(\overline{z}^j))$. Besides, $\sigma_t(z)_k = t_k^{z_k}$ for $k = 1, \ldots, n$ and

$$\sigma_t(\overline{z}^j)_k = \begin{cases} t_j^{\overline{z_j}} & \text{for } k = j \\ t_k^{z_k} & \text{for } k \neq j, \end{cases}$$

hence $\sigma_t(z)$ and $\sigma_t(\overline{z}^j)$ differ only in their j-coordinate. Therefore f_i is not constant in the j-coordinate, we are in situation 1 observed in Section 1.2.2 with f_i "Heaviside-like", and there is an edge from j to i in G. The fact that the sign of this edge equals the sign of the edge in $\mathcal{G}(f_{s,t})(z)$ is a direct consequence of the hypotheses that G is simple and that the parameterisation is coherent with G. $\square$

Now let a, b be stable states of f. Let s be chosen so that $a_i \neq b_i$ implies $\pi_s(a)_i \neq \pi_s(b)_i$ for $i = 1, \dots, n$. Let t_i^0 be the minimum of a_i and b_i and t_i^1 be the maximum. With these choices, $f_{s,t}$ has two stable states $\pi_s(a)$ and $\pi_s(b)$. Then, case 1 of Theorem 1 give the existence of a positive circuit in $G \supseteq \mathcal{G}(f_{s,t})(z)$ for some z, hence in G by Lemma 1. In presence of several stable subspaces, such an appropriate choice of s, t gives rise to several stable subspaces in the projected Boolean dynamics, and the conclusion follows from case 1'.

Assume that the dynamics has a trap cycle $(x^1, \dots, x^r)$ with strategy φ. Enrich the strategy by including for each $i = 1, \dots, r$ the direction of the dynamics: this gives an r-tuple of pairs $(\varphi(i), \epsilon_i)$ with $\epsilon_i = +1$ when $x^i_{\varphi(i)} < x^{i+1}_{\varphi(i)}$, $\epsilon_i = -1$ otherwise. Take an arbitrary $i \in \{1, \dots, r\}$. Let j be the least integer greater than i (modulo r) such that $\varphi(i) = \varphi(j)$ and $\epsilon_i \neq \epsilon_j$: such a j exists because $(x^1, \dots, x^r)$ is a cycle. Since this cycle is a trap cycle, each $k = i, \dots, j-1$ gives an edge from k to $k+1$ in $\mathcal{G}(f_{s_k,t_k})(\pi_{s_k}(x^k))$ for an appropriate choice of s_k, t_k. The sign of this edge is $\epsilon_k \epsilon_{k+1}$: the argument is indeed the same as in the Boolean case [4]. The induced (non-necessarily elementary) circuit $\varphi(i) \to \varphi(i+1) \to \cdots \to \varphi(j-1) \to \varphi(j) = \varphi(i)$ in the union of $\mathcal{G}(f_{s_i,t_i})(\pi_{s_i}(x^i)), \dots, \mathcal{G}(f_{s_{j-1},t_{j-1}})(\pi_{s_{j-1}}(x^{j-1}))$ is clearly negative, hence it contains an elementary negative subcircuit. Lemma 1 then gives the existence of a negative circuit in G. This completes the proof of Theorem 2.

1.5 Application to the bacteriophage λ

Let us apply Theorem 2 to the example of the bacteriophage λ introduced in Section 1.2.3.

$$\begin{tikzcd}
(0,1) & (1,1) \arrow[l] \arrow[d] \\
(0,0) \arrow[u] \arrow[r] & (1,0)
\end{tikzcd}
\qquad\qquad
\begin{tikzcd}
(0,1) \arrow[d, leftrightarrow] & (1,1) \arrow[l] \arrow[d] \\
(0,0) & (1,0) \arrow[l]
\end{tikzcd}$$

Fig. 1.3. Representation of $f_{s,t}$. Left: $s = (1,1)$, $t^0 = (0,0)$, $t^1 = (1,1)$. Right: $s = (1,2)$, $t^0 = (0,1)$, $t^1 = (1,2)$.

We condider two Boolean projections of the multilevel dynamics, represented in Figure 1.3. The left one preserves the presence of two stable subspaces: it contains two stable states $(1,0)$ and $(0,1)$; this projection therefore enables to recover the positive circuit between CI and Cro. The right one preserves the trap cycle and therefore enables to recover the negative self-loop on Cro. Hence, these two projections suffice to construct the whole original multilevel regulatory graph.

1.6 Conclusion and prospects

The theorems presented here have been proved in the framework of simple regulatory graphs. It is likely that they could be extended to multi-edge regulatory graphs, where a gene i can have several effects on a gene j at different thresholds.

This approach also provides interesting hints for a solution of the inverse problem, namely the problem of the induction of the regulatory graph on the basis of kinetic expression data. Indeed, using reasonably good discrete time series, it should be possible to induce local regulatory graphs, whose union should match (part of) that of the underlying regulatory network (see, e.g., [2]).

Finally, still in the context of the discrete formalism defined here, we are presently working on the delineation of sufficient conditions connecting the presence of specific regulatory circuits with specific dynamical properties, building on the notion of feedback circuit functionality proposed in [11]. Tentatively, this will lead to the definition of specific constraints on the parameter values associated with the elements involved in the generation of specific multistationary or oscillatory behaviour, in conjunction with the delineation of the feedback circuits at the basis of these properties.

Acknowledgements. We are particularly grateful to Heike Siebert for her remarks and corrections. This work has been supported by French ANR project JC05-53969.

References

1. Arkin A, Ross J, McAdams HH (1998) Genetics 149: 1633–1648.
2. Perkins TJ, Hallett M, Glass L (2004) J Theor Biol 230: 289–299.
3. Ptachne M (1992) A Genetic Switch. Phage Lambda and Higher Organisms (2nd ed.) Blackwell Science.
4. Remy E, Ruet P, Thieffry D (2005) Technical report IML 2005-8.
5. Remy E, Ruet P (2005) To appear in Transactions on Computational Systems Biology, Springer LNCS.
6. Richard A, Comet J-P (2005) Technical report LAMI 123-2005.
7. Robert F (1995) Les systèmes dynamiques discrets. Springer.
8. Soulé C (2005) ComPlexUs 1: 123–133.
9. Thieffry D, Thomas R (1995) Bull Math Biol 57: 277–295.
10. Thomas R (1980) Springer Series in Synergics 9: 180-193.
11. Thomas R, Thieffry D, Kaufman M (1995) Bull Math Biol 57: 247–276.