Algorithms and complexity in biological pattern formation problems

D. Grigoriev [1], **S. Vakulenko** [2]

[1]IRMAR Université de Rennes, Beaulieu, 35042, Rennes, France

[2] Institute of Mechanical Engineering Problems, S. Petersburg, Russia

Abstract

In this paper we develop a new mathematical approach to the pattern formation problem in biology. First this problem was posed mathematically by A. M. Turing, however some principal questions left open (for example whether there exists an "universal" mathematical model that allows one to obtain any spatio-temporal patterns).

We consider here the pattern formation ability of some class of genetic circuits. First, we show that the genetic circuits are capable to generate arbitrary spatio-temporal patterns. Second, we give upper and lower bounds on the number of genes in a circuit generating given pattern. A connection between complexity of gene interaction and pattern complexity is found. We investigate stochastic stability of patterning algorithms. Results are consistent with experimental data.

1 Introduction. Turing approach

This paper deals with special circuits of the neural type playing a key role in contemporary biology, and our results can be applied to the pattern formation problem in biology. Mathematical approaches to this problem started with the seminal paper of A. M. Turing [1]. Turing studied how chemical patterns could emerge from spatially uniform states. His model is a system of two special partial differential equations, so-called two component reaction-diffusion system. In a more general multicomponent case, these systems have the form:

$$\frac{\partial u_i}{\partial t} = d_i \Delta u_i + f_i(u_1, u_2, ..., u_m), \quad x \in \Omega, \quad t \geq 0, \tag{1.1}$$

where unknown functions $u_i(x,t)$ can be interpreted as a reagent concentration, the term $d_i \Delta u_i$ describes the reagent diffusion and f_i are smooth (usually polynomial or rational in u_i) functions describing a nonlinear chemical interaction between the reagents. We suppose that Ω is a bounded domain and set some boundary and initial conditions. Turing introduced as well some key notions such as activator and inhibitor. He assumed that state cells are discrete and they can be modified by special chemical reagents.

Now the existence of such reagents is well known [3],[4]. Moreover, it is proved experimentally that, in multicellular organisms, the state of a cell can depend on gene expression inside this cell and on some signals from environment (electrical, chemical or pressure, [4]).

The Turing approach was developed by numerous works (see [5] for a review). Patterns obtained numerically are often similar to patterns actually observed in biology [5]. However, the equations for these models have been selected to be mathematically tractable and a priori they do not take into account experimental genetic information. Moreover, there is no direct evidence for a Turing's patterning any developing organism ([3], p.347).

Mathematically, the two main questions were open. First, whether actually model (1.1) is capable to produce any patterns, or not. Second, whether there exist algorithms that allow us to choose parameters (functions f_i and d_i) such that solution of (1.1) will approximate a given pattern.

More precisely, the first problem can be formulated as follows:

Universal pattern generation problem for Turing model (1.1).

Let $T_0 > 0$ and $T_0 < T$. Given a function $z(x,t), x \in \Omega, t \in [0,T]$ and a positive number ϵ, to find a number m, functions $f_i(u_1, ..., u_m)$ and coefficients d_i (where $i = 1, ..., m$) such that the solution of problem (1.1) with initial conditions $u_j = 0$ satisfies

$$\sup_{x,t} |z(x,t) - u_m(x,t)| < \epsilon, \quad x \in \Omega, \quad t \in [T_0, T]. \tag{1.2}$$

Below we consider a time discrete version of (1.1). (Notice that if we shall try to resolve (1.1) numerically, this version inevitably arises from (1.1)).

Using genetic circuits (a special subclass of systems (1.1)) we show that the universal pattern generation problem can be resolved. Moreover, it can be done by an algorithm, i.e., the pattern problem can be resolved constructively.

2 Genetic circuits

Genetic circuit models were proposed ([7], [8], [9], [11]-[13] among many others, see [10] for a review) to take into account theoretical ideas and experimental information on gene interaction.

Model [9] uses Boolean algebra (so-called Boolean switch network). Circuit studied by [11]-[13] is a generalization of the famous Hopfield model of attractor neural network [2]. On other hand, this circuit is a particular case of the Turing model, where f_i have a special form.

Genetic circuit approach, developed in [11]-[13], is based on two main biological ideas. The first one is to choose the gene concentrations as state variables for the description of gene regulation. The second one is to use networks similar to neural networks to describe activation or depression of one gene by another. Mathematically such model can be described as a system of partial differential equations of a special form [11, 12], namely

$$\frac{du_i}{dt} = R_i\sigma(\sum_{j=1}^{m} K_{ij}u_j + \sum_{j=1}^{m_1} M_{ij}\theta_j(x) - \eta_i) - \lambda_i u_i + d_i\Delta u_i, \tag{2.1}$$

where m is the number of genes included in the circuit, $u_i(x,t)$ the concentration of the i-th gene, λ_i the gene decay rates and d_i the gene diffusion coefficient, the parameters η_i are activation thresholds and σ is so-called sigmoidal function (see below).

The real number K_{ij} measures the influence of the j-th gene on the i-th one. The assumption that gene interactions can be expressed by a single real number per pair of genes is a simplification excluding complicated interactions between three, four and more genes. Clearly such interactions can exist, however then the problem becomes mathematically much more complicated.

In (2.1), θ_i are fixed functions. They give the densities of so-called " maternal genes" that engine pattern growth. The number of these genes is m_1. (For example, for *Drosophila Melanogaster* the key maternal gene is *bicoid*. The complete number of the maternal genes is about 50, see [3]). Also they can describe concentrations of the substrates involved in patterning. Indeed, we need some food to grow. The matrix M_{ij} describes an interaction between the genes u_i and the maternal genes.

One considers (2.1) in some open domain Ω with a regular boundary $\partial\Omega$. If $d_i > 0$ then, in addition to (1.1), one sets the standard zero Neumann conditions [33] for u_i on $\partial\Omega$:

$$\frac{\partial u_i}{\partial \mathbf{n}}(x) = 0, \quad x \in \partial\Omega,$$

where $\mathbf{n} = \mathbf{n}(x)$ is the unit vector orthogonal to the boundary $\partial\Omega$ at the point x and directed inward Ω.

If $d_i = 0$ there are no boundary conditions. The initial data equal zero

$$u_i(x, 0) \equiv 0, \quad x \in \Omega. \tag{2.2}$$

The function σ satisfies the following supposition:

Assumption 2.1. *Suppose σ is a strictly monotone increasing function satisfying*

$$\lim_{z \to -\infty} \sigma(z) = 0, \quad \lim_{z \to \infty} \sigma(z) = 1 \tag{2.3}$$

and a differential equation

$$\sigma' = P(\sigma), \tag{2.4}$$

where P is a polynomial.

The well known example can be given by $\sigma(z) = \frac{1+\tanh(z)}{2}$ (here $P = \sigma(1-\sigma)/2$). It is easy to see that the polynomial P satisfies the following properties: $P(0) = 0, P(1) = 0$ and $P(z)$ is positive for any $z \in (0, 1)$. We also observe that σ is a real analytic function satisfying estimates

$$\sigma(z) = O(\exp(-c_1|z|)), \quad z \to -\infty$$

$$\sigma(z) - 1 = O(\exp(-c_2 z)), \quad z \to +\infty, \tag{2.5}$$

where c_i are positive constants.

An example of σ playing an important role for biology is given by so-called Michaelis-Menten function. This function σ equals $x/(K + x)$ for positive x and equals 0 for $x \leq 0$, where K is a positive constant. This function satisfies (2.3) and (2.5). Relation (2.4) holds for $\sigma \in (0, 1)$ but $\sigma'(0)$ is not defined. Nonetheless, under some additional conditions, some results hold in this case as well (see Section 6).

Model (2.1) takes into account only three fundamental processes: a) the decay (degradation) of gene products (the term $-\lambda_i u_i$); b) exchange of gene products between cells (the term with Δ) and c) gene regulation and synthesis. Notice that (2.1) is a particular case of (1.1) with nonlinearities of a special form.

Another possible model is a dynamical system with discrete time, for example, defined by the following iterative process

$$u_i^{t+1}(x) = r_i \sigma(\sum_{j=1}^{m} K_{ij} u_j^t(x) + \sum_{j=1}^{m_1} M_{ij} \theta_j(x) - \eta_i) - \lambda_i u_i^t(x) + d_i \Delta u_i^t(x), \tag{2.6}$$

$$u_i^0(x) \equiv 0 \tag{2.7}$$

where $t = 0, 1, 2, ..., T$, T is an integer and $x \in \Omega$. Numerical procedures solving (2.1) lead to models similar to (2.6). A simplified variant of system (2.6) was investigated, for example, in [14].

An important advantage of (2.6) with respect to (2.1) is that, if $d_i = 0$, the Khovanskii [15] results can be applied to this model. In fact, we shall see below that (2.6) defines a Pfaffian chain if the functions θ_i are Pffafian.

Of course models (2.1) and (2.6) are rough simplifications. Actually many other processes can be taken into account. In fact, the number of involved genes is of order of many thousands, even a reasonable approximation of this process is not known [3]. There is no single universal strategy of patterning ([3], p. 10). Nonetheless it is clear that this rough approximation (2.6) has a connection with actual biology. There are no doubts that threshold mechanisms are important and complicated circuits of interacting proteins and genes actually exist [17],[18].

To investigate (2.1) and (2.6), the most of the previous works used numerical simulations. For example, the paper [13] analyzes complicated patterns occurring under a random choice of the matrix K.

In this paper we focus our attention on model (2.6). We show that model (2.6) is mathematically tractable. First, we show, in a purely analytical way and without any numerical calculations, that any time sequence of any space patterns can be approximated by genetic circuit (2.6). Second, we examine a connection between "the complexity of a genetic circuit" and the "pattern complexity". Naturally, both complexities should be defined in a reasonable way. Third, we are going to investigate stability of morphogenesis process with respect to random perturbations.

Let us formulate now the pattern generation problem for system (2.6).

Let us fix some function σ satisfying Assumption 2.1. On the contrary, we consider $N, K_{ij}, M_{ij}, \lambda_i, d_i, r_i$ and θ_i as "control" parameters. We denote the set of these parameters by $\mathcal{P}$. The morphogenesis problem for (2.6) can be described as follows. Given a spatio-temporal pattern and a number $\epsilon > 0$, the problem is to adjust parameters $\mathcal{P}$ of (2.6) such that network (2.6) would approximate the given target pattern. The target pattern is defined by a time sequence of functions $z^t(x)$ where $x \in \Omega \subset \mathbf{R}^n, t \in [0, T]$ with the values z from $[0, 1]$.

Pattern generation problem for gene circuits

Let $T_0 > 0$ and $T_0 < T$, where T_0, T are integers. Given functions $z^t(x) \in [0,1], x \in \Omega, t = 0, 1, ..., T$ and a positive ϵ, to find parameters $\mathcal{P}$ such that the functions generated by relations (2.6), (2.7) satisfy

$$\sup_{x,t} |z^t(x) - u^t_m(x)| < \epsilon, \quad x \in \Omega, \quad t = T_0, ..., T. \tag{2.8}$$

Remark. We cannot satisfy (2.8) for $t = 0$ since initial functions u^0_j are equal zero.

Let us give a biological interpretation of this formulation. Among the genes u_i, we select a special gene, say u_m. The cell states depend on the expression of this gene. Other genes $u_1, u_2, ..., u_{m-1}$ are "hidden genes". They are involved in a cell biochemical machinery, but they do not act directly on the cell states. Such an approach is in a good accordance with experimental facts (see [3, 4]). It reminds classical approaches of neural network theory [20, 22, 25], where, similarly, we distinguish "input", "output" neurons and "hidden" ones.

3 Main results and organization of paper

Let us formulate now main mathematical results, ideas of proofs and give their biological interpretation (see also [40]).

Results

A Under some conditions on $\theta_i(x)$ and T_0, problem (2.8) always has a solution. Any sequences of the patterns $z^t(x)$ can be approximated, within an arbitrarily small error, by gene circuits (2.6). Notice that our conditions are necessary and sufficient (see Section 4 for details).

B Parameters of a circuit that approximates a given sequence $z^t(x)$, can be found by an algorithm.

C Given a final pattern $z^T(x)$, one can estimate the minimal number of genes in a network that generates this pattern. We give definitions of "complexity" of the circuits and pattern "complexity". We show, by the Khovanski theory [15], that there exists a connection between these complexities: it is impossible to obtain a "complex" pattern by a "simple" circuit.

We introduce and apply the different measures of the pattern complexity. Basic biological concepts on gene expression [3, 4], lead, in a natural way, to the definition of pattern complexity as the number of connectivity components of some sets D defined by the pattern $z^T(x)$. These sets can be defined in different ways. We consider here two cases. In the first case we define D

as a level set,

$$D_{c,t} = \{x : \ z^t(x) = c\}.$$

In the second case

$$D_{c_1,c_2,t} = \{x : \ c_1 \leq z^t(x) \leq c_2\}.$$

Here $0 \leq c \leq 1, 0 \leq c_1 < c_2 \leq 1$. These definitions admit a biological interpretation [4, 13]. The sets $D_{c,t}$ and $D_{c_1,c_2,t}$ are boundaries of a domain, where the gene u_m (that defines "structure" of "organism") is expressed.

In the first case, in order to connect the pattern complexity and the circuit parameter, we use estimates following from the fundamental results of Khovanskii [15]. These estimates are independent of the diameter of the domain $\Omega \subset \mathbf{R}^n$ and of the maximum of the absolute values of the entries $|K_{ij}|$. In this case the pattern complexity can be estimated via $(r_\theta + mT + n)$, where parameter r_θ is a complexity of inputs $\theta(x,t)$, the number mT characterizes the complexity of gene interactions.

In the second case we obtain essentially stronger estimates, in a quite elementary inductive way. However, in opposite to the previous ones, these estimates depend on the diameter of the domain Ω and on the maximum of the absolute values of the entries $|K_{ij}|$.

It is not sufficient to have a patterning algorithm; actually, algorithms have to be stable under random errors and perturbations. In particular, they must be stable under random noise and sharp changes of ecological conditions. Indeed, ecological catastrophes can eliminate a food, mutations can change properties of some genes. Mathematically this means that actually the functions $\theta_i(x)$ depend randomly on time t.

We consider the question on the stochastic stability of genetic circuits (2.6). We define stochastic stability of system (2.6) on time interval $[0, T]$ as the probability that the gene densities $u_i^t(x)$ stay inside some fixed bounded domain for all t from $[0, T]$. Notice that such a definition follows standard ideas of the theory of random perturbations of dynamical systems [27]. This probability can be called the survival probability.

Simple estimates allow to conclude that

D the more is the valency of a node the stabler is the circuit with respect to perturbations in this node. (The valency of the node is the number of links connecting this node with other ones; in our case the valency of i-th gene is the number of non-zero entries K_{ij}).

M. Gromov and A. Carbone formulated the following important problem: "Homeostasis of an individual cell cannot be stable for a long time as it would be destroyed by random fluctuations within and without cell. There is no adequate mathematical formalism to express the intuitively clear idea of replicative stability of dynamical systems" ([26], p.40).

Recall that homeostasis here means supporting of life functions of the cell. Namely, it is well known that biological moleculas amd chemical mechanisms in the cell are fragile [4]. Thus, in order to support their functioning, main parameters of the cell medium (temperature, pressure, pH, reagent concentrations) must be within some (sometimes narrow) intervals independtly of external medium oscillations [4].

This problem can be formulated within the framework of model (2.6). We use here a classical measure of stability from the theory of dynamic systems under random perturbations [27]. We prove that the survival probability of each circuit of a fixed structure tends to zero as $T \to \infty$. Therefore, "homeostasis" generated by a fixed circuit will be broken as time tends to infinity.

E To answer Gromov-Carbone's question by means of model (2.6), we show that although a fixed isolated circuit is always stochastically unstable, a chain of circuits can be stable. In this chain, each circuit is obtained from the previous one by some algorithm modifying the circuit parameter (replication algorithm). Roughly speaking, to survive, it is necessary to evolve.

However, the replication algorithm leading to "eternal" evolution cannot be arbitrary. We show that, for example, the mean valency must increase during evolution.

Outline of the proofs

The key point of the proof of **A** is Lemma 4.2. This Lemma can be interpreted as a Superposition principle. Namely, if circuits $C_1, C_2, ..., C_k$ generate chains of functions $z_i^t(x)$, where $t \in [0, T]$ and $i = 1, 2, ..., k$ then, for any continuous function $F(u_1, u_2, ..., u_k)$, we can find a new circuit that generate the superposition $F(z_1^t, z_2^t, ..., z_k^t)$. We show how the matrix $\mathbf{K}$ of the new circuit can be obtained from the matrices of given circuits. To this end, we use a special decomposition of the matrix $\mathbf{K}$.

Notice that the proof is constructive and gives us an algorithm. This algorithm exploits a modular structure of the circuits. The key tool is the well studied multilayered approximations [22]-[25]. It gives an explicit upper estimate of gene number m via the target pattern. Suppose that for any $t = 0, 1, ..., T$ the functions $z^t(x)$ are Lipshitzian. Then the number m of the genes participating in a circuit generating a sequence $z^t(x)$, $t = 1, ...T$, can be estimated through

$\max_t Lip(z^t)$. Here $Lip(z)$ is the Liphitz constant of $z(x)$.

In one-dimensional case ($dim\Omega = 1$), to approximate any $z^t(x)$ by (2.6), it is sufficient to have only one strictly monotone function $\theta_i(x)$ ($m_1 = 1$).

To demonstrate results **C** we show that, under our assumptions on σ, this function is Pffafian. Under the assumption that θ_i are Pffafian, it is easy to prove, by an induction, that circuit (2.6) gives rise to a Pfaffian chain of functions. The Khovanskii estimates allow now to connect topological properties of final pattern $z^T(x)$ with some circuit parameter and to obtain the result **C**.

The proof of **D** and **E** is quite straightforward.

Comments and interpretations

Result **A** can be considered as a generalization of previous results on multilayered neural networks and the Hopfield circuits. It is well known that any pattern $z(x)$ can be approximated, within arbitrary precision, by a multilayered neural network with sufficiently large number of neurons [22]-[25]. On other hand, it was shown that the Hopfield model produces, within arbitrary precision, any time trajectories [20] and even any structurally stable attractors [21]. To obtain a complex time trajectory or a complex attractor we must take a sufficient number of neurons.

Result **A** generalizes, for system (2.6), simultaneously the both previous results . This shows that any time sequences of any patterns $z(x)$ can be approximated. Of course, such a result is quite evident if we consider a sufficiently large circuit with parameters $\theta_i(x)$ and if we can adjust these $\theta_i(x)$. However, in our case the functions $\theta_i(x)$ are subject to some conditions and are fixed as well as the whole structure of our dynamical systems (2.6).

It is interesting to note that the main idea in proving **A** and **B** is connected with contemporary ideas of molecular and developmental biology [3, 4, 16, 19]. It is well known now that the genes are organized in blocks and their interaction has a modular structure [3, 4]. Mathematically this means that the matrix **K** is decomposed in some blocks (see Section 4).

Let us observe that conclusion **D** is in a good accordance with experimental results of [18]. This work investigated protein networks in 43 microorganisms. It was shown that the most connected proteins in the cell are the most important for its survival.

Conclusion **E** also is confirmed by experimental data (see [17]). It is known that, for biological networks, the averaged valency increased during evolution process.

Organization of the paper

We state the results **A** and **B** in Section 4. In Sections 5-7 we state results **C**. In Section 5 we introduce different measures of complexities. Section 6 studies the Khovanskii estimates of network complexity via pattern complexity. Section 7 is focused on simpler non-uniform estimates. Section 8 considers stability under random perturbations and Section 9 concerns with Gromov-Carbone's problem.

4 Pattern generation and patterning algorithm

We simplify model (2.6) removing the terms describing the gene diffusion and degradation (i.e., we put $\lambda_i = d_i = 0$). We set $M_{ij} = \delta_{ij}$, where δ_{ij} is the Kronecker symbol, and $m_1 = m$. We also suppose $r_i = 1$. Let us denote m_0 the number of non-trivial functions θ_i, i.e., such that $\theta_i(x) \neq const$ on Ω.

As a result, we obtain the following iterative model

$$u_i^{t+1}(x) = \sigma(\sum_{j=1}^{m} K_{ij} u_j^t(x) + \theta_i(x) - \eta_i), \tag{4.1}$$

where

$$u_i^0(x) = 0, \quad t = 0, 1, 2, ..., T, \quad x \in \Omega. \tag{4.2}$$

We show that the universal pattern generation problem can be resolved even for this simplified model. Notice that this system is a particular case of circuits considered in [29]-[32].

For (4.1), the pattern generation problem can be formulated as above (see (2.8)), but now the parameters $\mathcal{P}$ are the integer number m, the matrix $\mathbf{K}$ and the numbers η_i, $i = 1, ..., m$. Recall that $\theta_j(x)$ are fixed.

Our main result is

Theorem 4.1. *Suppose $T_0 > 2$ and that there exist continuous functions $\phi_l(\theta)$, $l = 1, ..., d$ defined on $\mathbf{R}^m$ such that $x_l = \phi_l(\theta_1(x), ..., \theta_m(x))$ for each $x \in \Omega \subset \mathbf{R}^d$. Then the pattern generation problem for (4.1) has a solution.*

Remark 1. The assumption of the theorem implies that $m_0 \geq d$ (at least d functions θ_i are non-trivial). In the one-dimensional case $d = 1$ this assumption holds if at least one function θ_i is strictly monotone. Moreover, under the condition of Theorem 4.1, any function $f(x_1, ..., x_d)$

can be represented as a function of $\theta = (\theta_1, ..., \theta_m)$. Indeed, $f(x_1, ..., x_d) = f(\phi_1(\theta), ..., \phi_d(\theta)) = \tilde{f}(\theta)$.

Remark 2. We also observe that the assumption on θ_i is necessary to approximate any sequences $z^t(x)$ by (4.1). In fact, chain (4.1) can generate only such sequences z^t, where each $z^t(x)$ depend on x through $\theta(x) = (\theta_1(x), ..., \theta_m(x))$. This means that for each z^t must exist a function $G^t(\theta)$ such that $z^t(x) = G^t(\theta)$. If our assumption does not hold, the trivial target sequence $z^t = x_k$ cannot be approximated by (4.1). Consequently, we conclude that the assumption of the theorem is sufficient and necessary in order to resolve by (4.1) the pattern generation problem for any outputs z^t.

A brief proof of Theorem 4.1 can be obtained by the following lemma.

Superposition Lemma 4.2. *Consider a family consisting of p circuits (4.1) generating functions $u^t_{i,s}$, where $t = 0, ..., T_1$, $s = 1, ..., p$ and $i = 1, 2, ..., m_s$ (here the index s marks the functions generated by s-th circuit, m_s is the number of the genes involved in s-th circuit). Denote by $\mathbf{u}^t$ the vector with the components $u^t_{1,1}, u^t_{2,1}, ..., u^t_{m_1,1}, u^t_{1,2}, ..., u^t_{m_2,2}, ..., u^t_{1,p}, ...u^t_{m_p,p}$.*

Suppose that $z^t(x) = F(\mathbf{u}^t(x))$, where F is a continuous function of N variables defined on N -dimensional cube $Q_N = [0,1]^N$ and $N = \sum_{s=1}^p m_s$ is the complete number of functions involved in the circuits. (This means that the target pattern can be expressed through the patterns generated by our family). Then for any $\epsilon > 0$, there exists a circuit (4.1) satisfying (2.8) with $T_0 = 2$ and $T = T_1 + 2$.

The main idea of the proof is based on the well known fact: the gene networks have modular structure and are organized in blocks [19]. We use as well the following well known approximation result (see [22, 25, 24, 21]): for $\kappa > 0$ there exist such M and coefficients A_{kjs}, b_k, η_k such that

$$|\sigma^{-1}(F(\mathbf{u})) - \sum_{k=1}^{M} b_k \sigma(\sum_{s=1}^{p} \sum_{j=1}^{m_s} A_{kjs} u_{j,s} - \eta_k)| < \kappa, \quad \mathbf{u} \in Q^N. \tag{4.3}$$

Let us construct now a large circuit including given networks and additional variables v_k, w, where $k = 1, ..., M$. The time evolution is defined by

$$v_k^{t+1} = \sigma(\sum_{s=1}^{p} \sum_{j=1}^{m_s} A_{kjs} u^t_{j,s} - \eta_k), \quad w^{t+1} = \sigma(\sum_{k=1}^{M} b_k v_k^t). \tag{4.4}$$

This means that w^{t+2} is determined through $\mathbf{u}^t$. We renumerate all set of the functions $u_{j,s}, v_k, w$ in such a way that $u_{m'} = w$, where m' is the complete number of these functions, i.e., w defines "the output pattern". Now relations (4.3) and (4.4) yield (2.8) if $\kappa = \kappa(\epsilon)$ is sufficiently small and M is large enough.

Theorem 4.1 follows from Lemma 4.2. To show it, we construct the circuit defined by the following relations. We can suppose without any loss of generality that all θ_i are not constants in Ω. Furthermore, we set $\mu_i = 1$. To apply Lemma 4.2, we define a family of networks consisting of a single circuit, where the number of the genes $m = m_0 + 1$. We define this circuit by the relations $u_m^{t+1} rew = \sigma(u_m^t - \eta_m), u_i^{t+1} = \sigma(\theta_i)$, where $i = 1, ..., m_0$. We observe now that u_m^t is a strictly monotone increasing sequence of constants, i.e., $u_m^t = q^t$, where q^t are independent of x. For $i \leq m_0$ and $t \geq 1$ we have $u_i^t = \sigma(\theta_i(x)) = \rho_i(x)$. Then Lemma 4.2 entails that any sequence of the functions z^t of the form $z^t(x) = F(\rho_1(x), \rho_2(x), ..., \rho_{m_0}(x), q^t)$ can be approximated by a circuit (4.1). Since the sequence q^t is strictly monotone in t, this means that any sequences of the functions of the form $G^t(\rho_1, ..., \rho_{m_0})$ can be approximated as well. Now we use that $\sigma(z)$ is strictly monotone in z. This entails that circuits (4.1) can approximate any sequences of functions $\tilde{G}^t(\theta_1, ..., \theta_{m_0})$ and thus, according to remark 1 (see above), any sequences $f(x_1, ..., x_d)$. This completes the proof.

This proof gives moreover an algorithm to resolve the universal pattern generation problem. Namely, the key step of the proof (approximation (4.3)) can be performedL by a constructive procedure (see [21]). With little modifications of the proof, a simple explicit estimate of the gene number M can be obtained under some supplementary assumptions on F from Lemma 4.2 and on z^t from Theorem 4.1. Namely, we suppose that the functions $F(\mathbf{u})$ and $z^t(x)$ are Lipshitzian, with the Lipshitz constants $Lip(F)$ and $Lip(z^t)$.

Then the function F can be approximated as follows. First, for any $\kappa > 0$ we can approximate F by a sum of characteristic functions

$$|F(\mathbf{u}) - \sum_{k=1}^{M_1} f_k \chi_{\pi_k}(\mathbf{u})| < \kappa, \tag{4.5}$$

where π_k are the N-dimensional boxes $\pi_k = \{a_i < u_i < b_i\}$ and u_k are components of the vector $\mathbf{u}$. The number M_1 can be evaluated by

$$M_1 < const(\kappa^{-1} N^{1/2} Lip(F))^N. \tag{4.6}$$

Each χ_{π_k} can be approximated by the sigmoidal functions:

$$|\chi_{\pi_k} - \sigma(\alpha(\sum_{k=1F}^{N} \sigma(\alpha(b_k - u_k)) + \sigma(\alpha(u_k - a_k))) - \alpha(2N - 1/2))| < \kappa_1, \tag{4.7}$$

where $\alpha(\kappa_1)$ is a positive number large enough. Relations (4.4) can be modified in the following way. We introduce a network consisting of the old genes u_k and new ones $v_k, \tilde{v}_k$ and w. We set

$$v_k^{t+1} = \sigma(\alpha(u_k^t - a_k)), \quad \tilde{v}_k^{t+1} = \sigma(\alpha(b_k - u_k^t)),$$

$$w^{t+1} = \sigma(\alpha(\sum_{k=1}^{N} v_k^t + \tilde{v}_k^t) - \alpha(2N - 1/2)). \tag{4.4a}$$

In contrast to (4.4), Lemma 4.2 holds now with $T_0 = 3$ since w^{t+3} can be expressed through $\mathbf{u}^t$.

The network generating z^t can be constructed as above in the proof of Theorem 4.1 under condition that $T_0 > 3$.

Inequalities (4.6),(4.7) and arguments from the proof of Theorem 4.1 give the following upper estimate of the number m of the genes in the chain generating a given sequence z^t:

$$m < const(\max_{t \in [0,T]} \epsilon^{-1} Lip(z^t))^{m_0}. \tag{4.8}$$

Let us find conditions on the matrix $\mathbf{K}$ guaranteeing that the pattern sequences $u_i^t(x)$ converge as $t \to \infty$. Iterations (4.1) can be considered as a dynamical system with the discrete time. Such convergence property holds for so-called monotone systems preserving some (partial) order in an appropriate Banach phase space [35, 36, 37].

For mappings acting in $\mathbf{R}^n$ we can introduce such a partial order $u < v$ by

$$u < v \text{ if } u_j < v_j \quad \text{for} \quad each\ j. \tag{4.9}$$

Let $u \to F(u)$ be a smooth map. This map F conserves order (4.9) if

$$\frac{\partial F_i}{\partial u_j} > 0, \quad i \neq j. \tag{4.10}$$

In the case of dynamics (4.1) this condition holds for matrices $\mathbf{K}$ such that

$$K_{ij} > 0, \quad (i \neq j). \tag{4.11}$$

The theory of monotone dynamical systems have been pioneered by the seminal work of M. Hirsch [35], afterwards developed P. Polácik et al. (for example, [36], for a review see [37]).

If (4.11) is satisfied, the functions $u_i^t(x)$ converge to functions $U_i(x)$ ("final pattern"). This final pattern is the solution of the system

$$U_i(x) = \sigma(\sum_{j=1}^{m} K_{ij} U_j(x) + \mu_i \theta_i(x) - \eta_i). \tag{4.8}$$

The properties of this pattern can be investigated in some cases (see below).

To conclude this section, let us notice that the universal pattern generation problem for Turing model (1.1) (formulated above, see Section 1) can be studied by means of an analogous approach (see work [41]).

5 Complexity of a pattern and complexity of a network

In this section we consider the following problem. Suppose we observe some sequence of patterns $z^t(x),\ x \in \Omega,\ t \in [0,T]$. We would like to estimate the number of the genes required to create this sequence.

To resolve this problem we can use different characteristics of pattern complexity. In this paper we employ the following three quantities: $C_1(z^t(\cdot), c)$, $C_2(z^t(\cdot), c_1, c_2)$, $E(z^t(\cdot))$. They are functions of the discrete time t.

The quantity C_1 is the number of the connected components of the set

$$D_{c,t} = \{x : \ z^t(x) = c\}. \tag{5.1}$$

To define C_2, let us consider a set $D_{c_1,c_2,t}$ depending on two parameters c_1, c_2 and t. Namely, let us define

$$D_{c_1,c_2,t} = \{x : \ c_1 \leq z^t(x) \leq c_2\}. \tag{5.2}$$

Then C_2 is the number of the connected components of this set.

Both complexity measures are discrete, whereas E is a continuous quantity defined by

$$E(t) = \int_\Omega |\nabla z^t|^2 dx. \tag{5.3}$$

Let us discuss now the biological sense of C_1, C_2 and E and relations between them.

Organisms consist of cells and these cells can be in different states. Following the ideas stated in Introduction (see as well [3, 4, 13] , we assume that different cell states appear as a result of expression of different genes. We consider here the case of one gene. Let u_m be such a gene.

Then we can study structures consisting of two kinds of cells: modified and the usual ones. If u_m is expressed at x then we have here a modified cell at x, otherwise the cell remains in a usual state.

Following the threshold approach (see Introduction) we suppose that the gene u_m is expressed if $u_m > c$ and it is not expressed in the opposite case ($u_m \le c$). In this case we obtain, as a natural measure of complexity, the quantity C_1.

The measure C_2 admits a similar interpretation. Here we assume that u_m is expressed if $u_m > c_2$ and it is not expressed if $u_m < c_1$. In the case $c_1 < u_m < c_2$ we deal with an intermediate (transient) state.

Thus both measures C_1 and C_2 relate to the number of transitions between the cells of different types.

Notice that using Sard' theorem, we can choose c, c_1, c_2 in definitions (5.1) and (5.2) such that at least locally the boundaries of the connected components will be smooth submanifolds of Ω of the codimension 1. In particular, if Ω is an interval, these components will be isolated points.

Example. For a periodical in x function $z^t(x)$ ("layered structure") C_1=C_2= number of layers (for appropriate c, c_1, c_2).

The third measure, the quantity E, can be interpreted as a mean value of "oscillations" of z.

The results for C_1 and C_2 are quite different. To estimate m through C_1 we use so-called Pfaffian chains [15], under some additional assumptions on σ. It allows us to obtain rough estimates of C_1 by Khovanski's results. Estimates of C_2 and E can be derived in a simpler way and appear to be essentially better.

Up to now, nobody knows whether the Khovanskii bounds can be improved. The key difference between estimates of C_1 on the one hand, and C_2, E on the other is that the estimates of C_2 and E depend, in particular, on the diameter $diam(\Omega)$ of domain Ω whereas the ones of C_1 are independent of this diameter.

6 An estimate of m via C_1

Let us introduce the key notion of a Pfaffian chain [15], [28].

Definition. A Pfaffian chain of the length r and degree $d \geq 1$ is a sequence of real analytic functions $f_1(x), f_2(x), ...f_T(x)$ in $\mathbf{R}^n$ with the following property: every f_j, $1 \le j \le T$ satisfies a Pfaffian equation

$$\frac{\partial f_j}{\partial x_k} = g_{kj}(x, f_1(x), ..., f_j(x)), \tag{6.1}$$

where g_{kj} are polynomials of degrees $\leq d$. Then T is called the length and d the degree of the Pfaffian chain.

Pffafian functions are well studied. They enjoy the following properties: the sum and the product of two Pfaffian functions f_1 and f_2 of lengths r_i and degrees d_i are again Pffafian functions of length $r_1 + r_2$ and degree $d_1 + d_2$ for both the sum and the product. Superpositions of Pfaffian functions also are Pfaffian (see [28] for details).

Consider some elementary examples. The exponent $\exp(ax)$, $x \in \mathbf{R}$ is a Pfaffian function of length 1 and degree 2. More generally, any real analytic function $f(z)$, $z \in \mathbf{R}$ satisfying an equation

$$\frac{df}{dz} = P(z, f) \tag{6.2}$$

is a Pfaffian of degree $degP$. We observe thus that many classical sigmoidal functions are Pfaffian. For example, $f = (1 + \exp(z))^{-1}$ satisfies (6.2) with $P = f^2 - f$. Superposition $\sigma(\exp(ax))$ also is a Pfaffian, etc.

Let us show first that under Assumption 2.1 chain (4.1) can be considered as a Pfaffian chain. Let us introduce complexity of chain (4.1) as the tuple of integers

$$Comp = \{\ m,\ T,\ r_\theta,\ d_\theta,\ degP\}, \tag{6.3}$$

where r_θ is the sum of the lengths of Pfaffian chains for θ_i, d_θ is the maximum of the degrees of Pfaffian chains determining θ_i, $degP$ is the degree of the polynomial from (6.2) that defines σ.

Using induction, let us consider now the functions u_i^1. By differentiating, one has

$$\frac{\partial u_i^1}{\partial x_l} = \sigma'(\mu_i\theta_i - \eta_i)\mu_i\frac{\partial \theta_i(x)}{\partial x_l}.$$

Consequently by assumption 2.1 one obtains

$$\frac{\partial u_i^1}{\partial x_l} = P(\mu_i\theta_i - \eta_i)\mu_i P_{i,l}(x, v_1^i, v_2^i, ..., \theta_i), \tag{6.4}$$

where $P_{j,l}$ are appropriate polynomials, v_k^j are functions of chains determining θ_j. Thus, u_i^1 and θ_j form a chain of the degree $d_\theta + degP$ and the length $r_\theta + m$. Repeating these calculations, we conclude that $u_i^t, u_i^{t-1}, ...\theta_i$ form a chain of the degree $d_\theta + tdegP$ and the length $r_t = r_\theta + tm$.

Now the complexity of the pattern $u_m^T(x)$ can be estimated applying known results ([15], see also [28], Proposition A4).

Theorem 6.1. *The number C_1 of the connected components of the pattern $u_m^T(x)$ generated by (4.1) can be bounded from above by*

$$C_1 < 2^{(r_\theta+Tm)^2}(d_\theta + TdegP)^{O(r_\theta+Tm+n)}. \tag{6.5}$$

Thus given C_1 we can bound from below $R = r_\theta + Tm$ roughly as $(\log_2 C_1)^{1/2}$, provided that $\log(degP), \log(d_\theta), n^{1/2}$ are less than $r_\theta + Tm$. The quantity R can be interpreted as a "complexity" of gene circuit (4.1).

This estimate does not look optimal but in general case up to now there exist no methods that could improve it.

However, if we consider rational σ, for example, the Michaelis-Menten case, then this estimate can be improved.

Recall that matrices K_{ij}, which actually meet in biological applications, are "sparse", i.e., each gene interacts only with few other genes. To describe this situation, we introduce the following characteristics: the valency V of the circuit. For each i we define V_i as the number of entries K_{ij} such that $K_{ij} \neq 0$. Then V is the maximum of V_i over i.

We first consider u_m^T as a function of variables $\theta_1, \theta_2, ..., \theta_s$. (We suppose after permuting subscripts that u_m^T actually depends only on s functions θ_i among $\theta_1, \theta_2, ..., \theta_m$, i.e., $\mu_1 = \mu_2 = ... = \mu_s = 1, \mu_{s+1} = ... = \mu_m = 0$).

Finally, for Michaelis -Menten circuits we consider the following set as a complexity of the circuit

$$Comp_M = \{\ m, s,\ T,\ r_\theta,\ d_\theta\}. \tag{6.6}$$

We shall show now that, under suitable suppositions, the final pattern u_m^T is a rational function in $\theta_1, \theta_2, ..., \theta_s$ and calculate the degrees of the numerator and the denominator of this function. It allows us to evaluate C_1 through $Comp_M$.

Assumption 6.2. *Suppose that the chain u_i^t consists of strictly positive functions.*

(This assumptions is natural from biological point of view and means that the concentrations $u_i^t(x)$ stay positive for any x).

Again we apply an inductive procedure. Let us consider $u_i^1(\theta)$. We see that

$$u_i^1 = \frac{\mu_i\theta_i - \eta_i}{1 + \mu_i\theta_i - \eta_i} = R_i^1/Q_i^1,$$

where R^1 and Q^1 are polynomials in θ_k of degree 1. At the second step, we have

$$u_i^2 = \frac{\sum_j K_{ij} R_j^1/Q_j^1 + \mu_i\theta_i - \eta_i}{1 + \sum_j K_{ij} R_j^1/Q_j^1 + \mu_i\theta_i - \eta_i}. \tag{6.7}$$

By elementary transformations we find from (6.7) that

$$u_i^2 = R_i^2/Q_i^2,$$

where $degR_i^2,\ degQ_i^2 \le V + 1$.

Repeating this procedure for the final pattern we find

$$u_i^T = R_i^T/Q_i^T, \quad degR_i^T, degQ_i^T \le (V+1)^{T-1}. \tag{6.8}$$

Applying the Khovanski's bound [15] to the polynomials R_m^T we conclude with the following proposition:

Proposition 6.1. *Under Assumption 6.2, the complexity C_1 of the pattern $u_m^T(x)$ of the Michaelis-Menten circuit does not exceed*

$$2^{r_\theta^2}(V^T + d_\theta)^{r_\theta + n}. \tag{6.9}$$

7 Estimates of E and C_2

The estimates of the previous section were independent of $\max_{i,j}|K_{ij}|$ and the diameter $diam\Omega$. Throughout this section we assume that the domain Ω is open and topologically trivial (contractable). In this section the bounds on E and C_2 are stronger than the ones on C_1 from the previous section, but hold under the conditions that

$$\max_{i,j}|K_{ij}| \le K_*, \quad diam\ \Omega = \delta > 0. \tag{7.1}$$

Other parameters involved in our estimates are V (the circuit valency defined above) and

$$\rho = sup_{i,k}|\frac{\partial\theta_i}{\partial x_k}|. \tag{7.2}$$

Let us denote

$$\sup \sigma'(z) = C_\sigma. \tag{7.3}$$

Now we can estimate ∇u_i^t inductively. Indeed, denote $\sup_{i,x} |\nabla u_i^t| = \mu^t$. Then

$$\mu^{t+1} \leq C_\sigma (V K_* \mu^t + \rho), \quad t = 0, 1, ..., \tag{7.4}$$

where $\mu^0 = 0$. Therefore,

$$\mu^t \leq \rho C_\sigma \frac{(C_\sigma V K_*)^t - 1}{C_\sigma V K_* - 1} \tag{7.5}$$

if $a = C_\sigma V K_* \neq 1$ and

$$\mu^t \leq t \rho C_\sigma, \tag{7.6}$$

if $a = 1$. We can suppose without any loss of generality that $a \neq 1$.

It is obvious that

$$E(u_m^t) < c\delta^n (\rho C_\sigma \frac{(C_\sigma V K_*)^t - 1}{C_\sigma V K_* - 1})^2, \quad n = dim \Omega. \tag{7.7}$$

Now we proceed to an estimate of C_2 and begin with the one-dimensional case. The inequality $C_2 > k$ where k is an integer, entails that there are two points x_1, x_2 such that

$$|x_1 - x_2| < \delta / k, \quad u_m^t(x_1) = c_1, \quad u_m^t(x_2) = c_2. \tag{7.8}$$

Thus there is a point ξ such that

$$|\frac{du_m^t}{dx}(\xi)| > \frac{(c_2 - c_1) C_2}{\delta}. \tag{7.9}$$

But by (7.5) we obtain then

Proposition 7.1. *If Ω is an interval, the following estimate of the pattern complexity via the circuit complexity holds:*

$$C_2(u_m^t, c_1, c_2) < diam \Omega (c_2 - c_1)^{-1} \rho C_\sigma \frac{(C_\sigma V K_*)^t - 1}{C_\sigma V K_* - 1}. \tag{7.10}$$

It gives us the required estimate. Let us notice that an analogue of this estimate also holds for continuous model (2.1). Its deduction is similar, and we leave it to a reader.

Let us turn now to the case $n = dim \Omega > 1$.

Theorem 7.2. *If Ω is a topologically trivial domain with a smooth boundary, for generic c_1 and c_2 we have*

$$C_2(u_m^T, c_1, c_2) < const \ mes \Omega \ (\rho C_\sigma \frac{(C_\sigma V K_*)^T - 1}{C_\sigma V K_* - 1})^n. \tag{7.11}$$

We start with an elementary assertion: if each connected component contains a ball of a radius r, then the number of connected components

$$C_2 < const\ mes\Omega\ r^{-n}, \tag{7.12}$$

where the factor $const$ depends on n.

Now, to prove Theorem, we are going to estimate r.

First we, using Sard's Theorem, choose c_1, c_2 such that they are regular values of a smooth function u_m^T.

Consider a connected component D_k of the set defined by (4.2). Then the boundary ∂D_k is a union of two disjoint smooth manifolds B_i of the codimension 1, $B_i = \{x : u_m^T(x) = c_i\}, i = 1, 2,$ herein we employ the theorem on a regular value, see [34]. Since the boundaries are compact, there are two points $x^1 \in B_1, x^2 \in B_2$ such that

$$dist(x^1, x^2) = \inf_{x\in B_1,\ y\in B_2} dist(x, y). \tag{7.13}$$

Let us set $2r = dist(x^1, x^2)$ and show that the open ball $\mathcal{B}$ which have the interval $[x^1, x^2]$ with the endpoints x^1, x^2 as a diameter is contained in D_k.

Indeed, we have just two possibilities: either $\mathcal{B}$ lies completely in D_k or completely outside of D_k. Otherwise, $\mathcal{B}$ would contain some points of the boundary ∂D_k, for example, a point z where $u_m^T(z) = c_1$. But then $dist(z, x^2) < r$ that gives us the contradiction with (7.13).

Let us check now that the second possibility ($\mathcal{B}$ is outside of D_k) also leads to a contradiction.

Let us denote by W the unique connected component of B_1 which contains the point $x^1 \in W$. Since W is a smooth submanifold of the codimension 1, due to the Alexander's duality [38] the complement $\Omega \setminus W$ consists of two connected components U_0, U_1 (taking into account the topological triviality of Ω). Then D_k lies completely in one of U_0, U_1, let $D_k \subset U_0$ for definiteness. The interval $(x^1, x^2]$ (with deleted endpoint x^1) does not intersect W (due to (6.13)), therefore this interval is contained completely either in U_0 or in U_1. On the other hand, the point $x^2 \in D_k \subset U_0$, hence the whole interval $(x^1, x_2] \subset U_0$.

For a small enough ball $\mathcal{B}_{x^1}(e)$ centered at x^1 the complement $\mathcal{B}_{x^1}(e) \setminus W$ has two connected components (again we make use of that W being a smooth submanifold of the codimension 1 and a connected component of the boundary of D_k). One of these two components coincides with $\mathcal{B}_{x^1}(e)\cap D_k$ and another one with $\mathcal{B}_{x^1}(e)\setminus\overline{D_k}$. This partition is the same as the partition of $\mathcal{B}_{x^1}(e)\setminus W$ into two connected components $\mathcal{B}_{x^1}(e)\cap U_0$ and $\mathcal{B}_{x^1}(e)\cap U_1$. Because we have $D_k \subset U_0$ we conclude that $\mathcal{B}_{x^1}(e)\cap D_k = \mathcal{B}_{x^1}(e)\cap U_0$. Therefore, a suitable beginning $(x^1,x^3]\subset(x^1,x^2]$ of the interval $(x^1,x^2]$ is contained in $\mathcal{B}_{x^1}(e)\cap D_k$ (see the previous paragraph). Taking into account that the open interval (x^1,x^2) does not intersect the boundary of D_k thanks to (7.13), this implies finally that $(x^1,x^2)\subset D_k$ which is a contradiction with that $\mathcal{B}$ is outside of D_k.

To conclude the proof, it is sufficient now to estimate r. Using the Lagrange theorem, we obtain

$$c_2 - c_1 = 2r|(\mathbf{n}\cdot\nabla u_m)|,$$

where $\mathbf{n}$ is a unit vector directed along the diameter $[x^1,x^2]$. This relation entails

$$r^{-n} \le C\sup|\nabla u_m|^n.$$

Applying estimates (7.5) and (7.12), we obtain (7.11).

Notice that the complexities C_1 and C_2 are stable under small perturbations.

Lemma 7.3. *For generic c and c_i the complexities C_1, C_2 of the pattern $u_m^t(x)$ are conserved under small smooth perturbations: the complexities of the pattern u_m^t coincide with the corresponding complexities of $u_m^t+\tilde{z}(x)$ if $|\tilde{z}_{C^1}|<\epsilon$ and ϵ is small enough.*

Proof. Consider the case C_2. The connected components are disjoint. Since they are compact, the distances d_k between these components are positive. If c_1, c_2 are regular values of u_m, their boundaries are smooth submanifolds of the codimension 1. If ϵ is sufficiently small, the perturbation of these level submanifolds are small, due to the regularity of the values c_i.

Thus, since $\inf d_k > 0$, the perturbed connected components rest disjoint.

An interesting particular case is given by the Michaelis-Menten dynamics. Suppose all the entries K_{ij} are positive. Then the patterns converge (see Section 4). Final patterns $u_i(x)$ satisfy

$$u_i(1+\sum_{j=1}^m K_{ij}u_j + \mu_i\theta_i - \eta_i) = \sum_{j=1}^m K_{ij}u_j + \mu_i\theta_i - \eta_i. \tag{7.14}$$

From Khovanski's bounds we get for the solutions of (7.14) the bounds on their complexities

$$C_1, C_2 < 2^{r_\theta^2}(m+d_\theta)^{r_\theta+n}.$$

8 Stochastic Stability

The important meaning has the problem of stability of networks under random perturbations of different parameters. This problem attracts a great attention of biologists (see [17]-[18], [19]).

We prove here some estimates on stability of (4.1) under noise leading to important biological consequences. Moreover, we develop an approach to the replicator stability answering the question of M. Gromov and A. Carbone, formulated in the Introduction.

Consider a perturbed problem (4.1):

$$u_i^{t+1}(x) = \sigma(\sum_{j=1}^{m} K_{ij} u_j^t(x) + h_i(x) - \xi_i(t)), \tag{8.1}$$

where $h_i = \mu_i \theta_i - \eta_i$. Here $\xi_i(t)$ are some random processes with the discrete time. We assume that they are independent for different i. The random quantities $\xi_i(t)$ can be distributed, for example, according to gaussian laws $\mathcal{N}(e_i, \kappa_i)$ with average e_i and deviations $\kappa_i > 0$. Different choices of the values ξ_i may correspond to different "ecological conditions". We introduce two functions

$$Prob(\xi_i(t) \geq a, \text{ for some } t \in [T_1, T_2]) = \Phi_i(a, T_1, T_2) \tag{8.2}$$

and

$$Prob(\xi_i(t) < a, \text{ for all } t \in [T_1, T_2]) = \Psi_i(a, T_1, T_2). \tag{8.3}$$

It is clear that $1 - \Phi_i = \Psi_i$. The following assumption plays an important role in what follows. Suppose

$$\Psi_i(a, T_1, T_2) > 0, \quad (T_2 > T_1), \quad \Psi_i(a, T_1, T_2) \to 0 \ as \ T_2 \to \infty \tag{8.4}$$

for fixed T_1. This means roughly speaking that ξ_k can take any large values with non-zero probabilities. This assumption holds for the gaussian probability distribution. It is clear that $\Phi_i(a, T_1, T_2)$ are increasing functions of T_2 for any fixed a while $\Psi_i(a, T_1, T_2)$ are decreasing.

Suppose that an "organism" (a gene circuit (8.1)) "survives" (supports homeostasis) if the concentrations u_i stay at some closed domain Π in the u -phase space.

Notice that Assumption 2.1 entails

$$u_i^t(x) \in (0, 1). \tag{8.5}$$

It is natural thus to suppose that Π is contained inside the cube $[0, 1]^m$.

As a measure of the stochastic stability of the circuit homeostasis, we consider the probability

$$P(\mathcal{P}, \Pi, \Omega, T_1, T_2) = Prob\{u_i^t(x) \in \Pi \text{ for each } x \in \Omega, \text{ and } t \in [T_1, T_2]\}. \tag{8.6}$$

This probability depends on the circuit parameters $\mathcal{P}$, the homeostasis domain Π and Ω. We shall name it the survival probability on the time interval $[T_1, T_2]$ and denote by $P(T_1, T_2)$ omitting the dependence on the parameters $\mathcal{P}$, Π and Ω. Such measure of the stability is standard in the theory of dynamical systems [27]. However, one can introduce other important measures of stability, for example, with respect to random eliminations of some genes (proteins) or vanishing of some entries of the matrix $\mathbf{K}$. This kind of stability is under great attention in recent works connected with the random graph theory (see the review [39] and references in it). We shall not consider this kind of stability here.

We estimate the stability via the following parameters: the valency, the maximum $|K_*|$ of absolute values of the entries K_{ij}, the maximum b of $|\theta_i(x)|$ and some parameter N_{key} that we introduce below. It is important to take into account the valency since it is well known that biological circuits are not completely connected: for each fixed node i we have a valency $V_i < m$: only V_i of the entries K_{ij} are non-zero.

To introduce N_{key}, let us observe that

$$\inf_{\mathbf{u} \in \Pi} u_i = W_i \geq 0, \quad \mathbf{u} = (u_1, ...u_m). \tag{8.7}$$

Denote $U_i = \sigma^{-1}(W_i)$. Some W_i and U_i could be positive. The corresponding indices $i_1, ..., i_s \in [m]$ we name key indices and the corresponding genes we name the key ones. In fact, if $W_i > 0$, this means that the organism cannot survive if the concentration of i -th gene is small enough at some points. The number s of the key genes is denoted by N_{key}. We denote I the set of key indices corresponding to the key genes.

Consider (8.1). Let us take some key index $i \in I$. We have the following simple inequality

$$\sum_{j=1}^{m} K_{ij} u_j^t(x) + \theta_i - \xi_i \leq S_i = V_i K_* + b - \xi_i. \tag{8.8}$$

Thus, if

$$\xi_i(t) > V_i K_* + b - U_i, \tag{8.9}$$

the concentration $u_i^{t+1}(x)$ is less than the critical value W_i. Moreover, if at least one $u_i^t(x)$ is less than W_i at some point x, the state $\mathbf{u}^t(x)$ is outside of this domain Π. Hence, we have

$$Prob\{\mathbf{u}^t(x) \in \Pi,\ t \in [T_1+1, T_2],\ x \in \Omega\} < \prod_{i \in I} \Psi_i(V_i K_* + b - U_i, T_1, T_2 - 1). \tag{8.10}$$

Therefore, we have proved

Proposition 8.1. *The survival probability satisfies*

$$P(T_1, T_2) < \prod_{i \in I} \Psi_i(V_i K_* + b - U_i, T_1 - 1, T_2 - 1) = P_+(T_1, T_2). \tag{8.11}$$

This estimate yields interesting biological consequences. Notice that the function P_+ is a monotone increasing function of the valency. It is decreasing as the number N_{key} of the key genes increases. Moreover, the sharper is the sigmoidal function σ, the larger is P_+.

The most interesting conclusion is the following. The more is the valency of a node the stabler is the circuit with respect to perturbations in this node. It is in an accordance with experimental results of the work [18]. They show that the most connected proteins in the cell are the most important for its survival.

Moreover, we notice that all circuits are unstable, more precisely, they are stochastically unstable as the time T goes to infinity. In fact, assumption (8.4) and estimate (8.11) imply that

$$P(0, T) \to 0 \quad as\ T \to \infty. \tag{8.12}$$

Then there arises a natural question: how to stabilize the circuits. We shall consider it in the next section.

9 Replicator Stability

We show in this section that a periodic renovation (replication) of the circuit parameters $\mathcal{P}$ can transform stochastically unstable systems to the stable ones. We can consider these transformations as an algorithm of "evolution". The key question is about algorithm properties providing the stability.

We consider circuits (4.1) under the assumptions of the previous section. We also suppose that $\xi_i(t)$ are identical independent random processes, which, in a certain sense, are homogeneous in time. More precisely, let us assume

$$\Phi_i(a, T_1, T_2) = \Phi_i(a, 0, T_2 - T_1). \tag{9.1}$$

Consider possible schemes of renovation. They can be described as follows.

Each T_r time steps we change the circuit parameters $\mathcal{P}$ following some rule. For example, each T_L time steps we can add to the network a new link, and each T_n steps, we include a new node (gene). Here T_n and T_L are some positive integers. We can also use more sophisticated schemes. For example, one can add new nodes with many links. In the case of graphs, different schemes of graph evolution were studied by numerous works, see the review [39].

Let us calculate the survival probability. Let $P_n = P(\mathcal{P}_n, [nT_r, nT_r + T_r])$ be the probability to survive within the time interval $[nT_r, (n+1)T_r]$. Here $\mathcal{P}_n$ are the circuit parameters in this time interval.

The probability to survive on the interval $(0, \infty)$ is then the infinite product

$$P(0, \infty) = P_1 P_2 P_3 ... = \prod_{n \in \mathbf{N}} P_n.$$

Consequently, the quantity $P(0, \infty)$ is non-zero if the series $\log P_1 + \log P_2 + ... + \log P_n + ...$ converges. We have obtained thus the following assertion.

Proposition 9.1. *The survival probability $P(0, T)$ remains positive as $T \to \infty$ if and only if the series*

$$\log P(\mathcal{P}_0, [0, T_r]) + \log P(\mathcal{P}_1, [T_r, 2T_r]) + ... + \log P(\mathcal{P}_n, [nT_r, (n+1)T_r]) + \tag{9.2}$$

converges. If this series disverges to $-\infty$, the survival probability tends to zero as time tends to infinity.

Propositions 8.1 and 9.1 yield an elementary consequence that gives us a sufficient condition for stochastic stability in infinite time. Notice that it is more precisely to say about stochastic stability of the pair (circuit, replication algorithm) rather than about stochastic stability of the circuits.

Proposition 9.2. *The survival probability $P(0, T)$ tends to zero as $T \to \infty$ if the series*

$$\sum_{i \in I} \log(\Psi(V_i^0 K_* + b - U_i, 0, T_r)) + \sum_{i \in I} \log \Psi(V_i^1 K_* + b - U_i, T_r, 2T_r)) + ...$$
$$+ \sum_{i \in I} \log \Psi(V_i^n K_* + b - U_i, nT_r, (n+1)T_r)) + ... \tag{9.3}$$

disverges. Here V_i^n are the valencies at the n-th renovation step.

To prove it, let us notice that, due to Proposition 8.1, $-\log P(T_1, T_2) > -\sum_{i \in I} \log \Psi(V_i K_* + b - U_i, T_1 - 1, T_2 - 1)$.

Although these results look quite elementary, nonetheless they allow to analyze the different evolution algorithms and lead to interesting biological consequences. Consider some examples.

Example 1.

Let us suppose that all the genes are key ones. Suppose that their stability is a priori bounded:

$$\inf_i U_i > \bar{U} > 0. \tag{9.4}$$

Biologically, this means that the gene stability is a priori bounded during evolution. Let us suppose that the renovation algorithm is, in a certain sense, simple. This means that the renovation procedure either adds to the circuit a node (gene), with a link, or only a link connecting some existing nodes.

Then such evolution is always unstable. To prove it, let us consider series (9.3). First we notice that if the gene number m is bounded as $T \to \infty$ then the valency is bounded by m and it is unstable due to (8.4) and (8.10). Thus, we can assume that $m \to \infty$ as $T \to \infty$. Then series (9.3) contains infinitely many of the terms that are negative and less than

$$\mu_n = \log \Psi(K_* + b - \bar{U}, nT_r - 1, (n+1)T_r - 1), \tag{9.5}$$

since the valency of new genes is $V = 1$. Due to the time homogeneity hypothesis (9.1) we observe that $\mu_n = \mu$ is independent of n. Also μ is non zero, according to assumptions (8.4). Thus series (9.3) disverges. We obtain analogous negative results even if each new gene enters for the circuit with many links but under the condition that the valency of this new gene stays a priori bounded.

Example 2.

Let us suppose that only a part of all the genes are key ones. Suppose that (9.3) holds. Assume that the renovation procedure adds to the circuit a node (gene), with a link, and this gene is not the key gene. (Therefore, the number of the key genes conserves).

Then such evolution can be stable or unstable depending on the properties of the processes ξ_k. To see it, let us consider series (9.3). For large n we can use the asymptotics

$$\log \Psi(V_i^n K_* + b - U_i, nT_r, (n+1)T_r)) = \log(1 - \Phi(V_i^n K_* + b - U_i, nT_r, (n+1)T_r)) \approx$$

$$\approx \Phi(V_i^n K_* + b - U_i, nT_r, (n+1)T_r). \tag{9.4}$$

Let us consider the case of gaussian random processes, with a constant deviation $\kappa_i(t) = \kappa$ and the zero means. Then, for example, if V_i^n grows as $O(\log n)$ as $n \to \infty$, this series converges.

Finally, we can perform a stable evolution (i.e. to have $\lim P_T > 0$ as $T \to \infty$) only if the renovation algorithm is complicated itself. Namely, the key protein enters for the circuit together with many links, and the number of new links increases unboundedly.

10 Ackowledgements

The authors are thankful to Dr. S. Genieys and Dr. J. Reinitz for useful remarks. The second author was supported by the grant PAST (Lyon University, France).

References

[1] A. M. Turing, The chemical basis of morphogenesis, Phil. Trans. Roy. Sco. B, **237** (1952) 37-72.

[2] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, Proc. of Natl. Acad. USA, **79** (1982) 2554-2558.

[3] L. Wolpert, R. Beddington, T. Jessell, P. Lawrence, E. Meyerowitz and J. Smith, Principles of Development (Oxford University Press, 2002).

[4] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, P. Walter, Molecular Biology of the Cell, 4nd ed. (Garland Publishing, Inc., New York, 2002)

[5] H. Meinhardt, in: Mathematical Models for Biological Pattern Formation, edited by P. K. Maini, H. G. Othmer, IMA Volumes in Mathematics and its Applications, **121**, (Springer, New York, Berlin, Heidelberg, 2000).

[6] M. Ridley, Evolution, 2nd ed. (Blokwell Scientific Publications Ltd, Oxford, 1996)

[7] L. Glass and S. Kauffman, The logical analysis of continuous, nonlinear biochemical control networks, J. Theor. Biology, **34**, (1973), 103-129.

[8] D. Thieffry and R. Thomas, Dynamical behaviour of biological regulatory networks, II.Immunity control in bacteriophage lambda, Bull. Math. Biology, **57**, (1995) 277-295.

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

[10] P. Smolen, D. Baxter, J. H. Byrne, Mathematical modelling of gene networks, Review in Neuron, **25**, (2000) 247-292.

[11] E. Mjolness,D. H. Sharp and J. Reinitz, A connectionist Model of Development, J. Theor. Biol. **152**, (1991) 429-453.

[12] J. Reinitz and D. H. Sharp, Mechanism of formation of eve stripes Mechanisms of Development, **49**, (1995) 133-158.

[13] I. Salazar-Ciudad, J. Garcia-Fernadez and R.V. Solé, Gene Networks Capable of Pattern Formation: From Induction to Reaction-Diffusion, J. Theor Biology, **205** (2000), 587-603.

[14] L. Mendoza and E. R. Alvarez-Buylla, Dynamics of Genetic Regulatory Networks for *Arabodopsis thaliana* Flower Morphogenesis, J. Theor. Biol. (1998), **193**, 307-319.

[15] A. Khovanskii, Fewnomials, Translations of Mathem. Monographs, Amer. Math. Soc., **88**, 1991.

[16] M.K. S. Yeung, J. Tegner, and J. J. Collins, Reverse engineering gene networks using singular decomposition and robust regression, Proceeding of Nat. Acad. Sci, , **99**, (2002) 6163-6168.

[17] H. Jeong, B. Tombor, R. Albert, Z. N. Otvai and A.L. Barabási, The large-scale organisation of metabolic networks, Nature (London), **407**, (2000) 651-654.

[18] H. Jeong, S. P. Mason, A.L. Barabási and Z. N. Otvai, Lethality and centrality in protein networks, Nature (London), **411**, (2000) 41-42.

[19] L. H. Hartwell, J. J. Hopfield, S. Leibler and A. W. Murray, From molecular to modular cell biology, Nature(London), **402**, (1999) C47-C52.

[20] K. Funahashi and Y. Nakamura, Approximation of Dynamical Systems by Continuous Time Recurrent Neural Networks, Neural Networks, **6**, (1993) 801-806.

[21] S. A. Vakulenko, Dissipative systems generating any structurally stable chaos, Advances in Diff. Equations, **5**, (2000), 1139-1178

[22] R. Hecht-Nielsen, Kolmogorov's mapping neural network existence theorem, IEEE International Conference on Neural Networks, San Diego: SOS Printing (1987), 11-14.

[23] G. Cybenko, Approximation by superpositions of a sigmoidal function, Mathematics of Control, Signals and System, **2**, (1989) 303-314.

[24] A. Barron, Universal Approximation Bounds for superpositions of a sigmoidal functions, IEEE Trans. on Inf. theory, **39**, (1993) 930-945.

[25] K. Hornik, M. Stinchcombe and H. White, Multilayer feedforward networks are universal approximators, Neural Networks, **2**, (1989) 359-366.

[26] M. Gromov and A. Carbone, Mathematical slices of molecular biology, Preprint IHES/M/01/03, 2001.

[27] A. D. Ventsel and M. I. Freidlin, Random Perturbations of Dynamic systems (Springer, New York, 1984).

[28] D. Grigoriev and N. Vorobjov, Complexity Lower Bounds for Computation Trees with Elementary Transcendental Functions Gates, Theoret.Comput.Sci., **157**, (1996) 185-214.

[29] D. Grigoriev, Deviation theorems for solutions of linear ordinary differential equations and applications to parallel complexity of sigmoids, St.Petersburg Math. J., **6**, (1995) 89-106.

[30] D. Grigoriev, Deviation theorems for pfaffian sigmoids, St.Petersburg Math. J., **6**, (1995) 107-112.

[31] D.Grigoriev, Approximation and complexity: Liouvillean type theorems for linear differential equations on an interval, Foundations of Computational Mathematics, **1**, (2001) 289-295.

[32] D.Grigoriev, Approximation and complexity II: iterated integration, Foundations of Computational Mathematics, **2**, (2002), 295-304.

[33] J. D. Murray, Mathematical Biology, (Springer, New York, Berlin, Heidelberg, 1993).

[34] M. W. Hirsch, Differential topology, (New-York, Heidelberg, Berlin, 1976).

[35] M. W. Hirsch, Stability and convergence in strongly monotone dynamical systems, J. Reine. Angew. Math. **383**, (1988) 1-58.

[36] P. Polácik and T. Terescák, Convergence to Cycles as a Typical Asymptotic Behaviour in Smooth Strongly Monotone Dicrete -Time Dynamical Systems, Arch. Rat. Mech. Anal.**116**, (1991) 339-360.

[37] H. L. Smith and H. R. Thieme, Convergence for strongly order preserving semiflows, SIAM J. Math. Anal., **22**, (1991), 1081-1101.

[38] W. S. Massey, A basic course in algebraic topology, (Springer, 1991)

[39] R. Albert and A. L. Barabási, Statistical mechanics of complex networks, Rev. Modern Physics, **74**, (2002) 47-97

[40] S. Vakulenko, D. Grigoriev, Complexity of gene circuits, Pfaffian functions and the morphogenesis problem, C. R. Acad. sci, Ser I. **337** (2003) 721-724

[41] S. Vakulenko and S. Genieys, Pattern programming by genetic networks, Patterns and Waves, Collection of papers, eds. A. Abramian, S. Vakulenko and V. Volpert, S. Petersburg, 2003, pp. 346-366.