Complexity of gene circuits, Pfaffian functions and morphogenesis problem

Sergey VAKULENKO [1], **Dmitry GRIGORIEV** [2]

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

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

Abstract. We consider a model of gene circuits. We show that these circuits are capable to generate any spatio-temporal patterns. We give lower bounds on the number of genes required to create a given pattern.

Complexité de réseaux de gènes, fonctions de Pfaff et le problème de morphogenèse

Résumé. On considère un modèle de réseaux de gènes. Nous démontrons que ces réseaux peuvent engendrer toutes les structures spatio-temporelles et nous obtenons des bornes inférieures du nombre de gènes du réseau qui engendrent une structure prescrite.

Version française abrégée

Le problème de morphogenèse biologique a été étudié par de nombreux travaux, à partir du papier célèbre de A. M. Turing [1]. Deux questions fondamentales ouvertes, qui se posent naturellement, sont comme suit. Premièrement, existent-ils des modèles mathématiques engendrant toutes les structures spatio-temporelles possibles? De plus, de trouver des algorithmes qui résolvent le problème suivant: étant donnée une structure, déterminer des paramètres du modèle qui engendre cette structure.

Ce papier considère un système dynamique important en biologie [3, 4, 5]. C'est un circuit de gènes qui est une version simplifiée, en temps discret, du modèle [4, 5].

Nous démontrons que :

A Toutes les structures spatio-temporelles peuvent être obtenues par tels circuits;

B Il existe des algorithmes universels qui nous permettent de définir de paramètres du circuit qui produit une structure donnée;

C Etant donnée une structure, on peut majorer le nombre de gènes du circuit qui engendre cette structure.

La formulation stricte et la démonstration du dernier résultat sont basées sur les méthodes de Khovanski [6]. Les algorithmes **B** utilisent les résultats et les algorithmes classiques de la théorie des réseaux en couche (voir [8]) et une idée, empruntée de biologie : les circuits génétiques réels ont une structure modulaire [7].

1 Introduction

Mathematically, the biological morphogenesis problem (how complicated patterns emerge from homogeneous or almost homogeneous structures) was posed first by A. M. Turing [1]. Turing's approach was developed in many works (for example, [2]). However, the following fundamental question was open: whether there exist mathematical models generating any spatio-temporal patterns? In the framework of these models, whether there are algorithms "programming" these patterns?

This paper deals with special circuits of the neural type playing a key role in biology [3, 4, 5]. These circuits are dynamical systems with discrete time and can be considered as a simplified version of the gene networks [4, 5]. We show that

A Any patterns can be obtained by these gene circuits;

B Parameters of a circuit generating given pattern, can be found by an algorithm;

C Given pattern, one can evaluate the minimal number of genes in a network that generates this pattern. There exists a connection between "complexity" of the circuits and pattern complexity.

Below we give a strict formulation and proofs of these results. The time discrete model has an important advantage: in this case we can use the methods [6]. It allows us to obtain **C**.

2 Model

We consider the following chain of functions $u_i^t(x)$ defined iteratively by

$$u_i^{t+1}(x) = \sigma(\sum_{j=1}^{m} K_{ij} u_j^t(x) + \mu_i \theta_i(x) - \eta_i), \quad u_j^0(x) = 0, \tag{1.1}$$

where $t = 0, 1, 2, ..., T$, $i = 1, 2, ..., m$ and $x \in \Omega$, $\Omega \in \mathbf{R}^d$ is a bounded domain, $d = 1, 2, 3$. Here m is the number of genes involved in the circuit, $u_i(x, t)$ the concentration of the i-th gene, the matrix K_{ij} defines a gene interaction, η_i are activation thresholds, μ_i are coefficients. Assume σ is a strictly monotone increasing function satisfying $\lim_{z\to-\infty} \sigma(z) = 0$, $\lim_{z\to\infty} \sigma(z) = 1$ and a differential equation $\sigma' = P(\sigma)$, where P a polynom. The well known example is $\sigma(z) = \frac{1+\tanh(z)}{2}$ (here $P = \sigma(1-\sigma)/2$). Functions $\theta_i(x)$ define concentrations of so-called maternal genes [9, 5].

3 Main Results

A. Consider the following problem:

Pattern generation problem on $[T_0, T]$**.** *Let T_0, T be integers such that $0 \le T_0 < T$. Given a sequence of continuous functions $z^t(x) \in [0, 1], x \in \Omega, t = 0, 1, ..., T$ and a positive ϵ, to find m, K_{ij}, η_i, μ_i such that the solution of problem (1.1) satisfies*

$$\sup_{x,t} |z^t(x) - u_1^t(x)| < \epsilon, \quad x = (x_1, ..., x_d) \in \Omega, \quad t = T_0, ..., T. \tag{1.2}$$

Notice that σ and θ_i are fixed.

Theorem 1.1. Suppose $T_0 \geq 2$ and 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 has a solution.

B. Let us introduce a measure of the complexity of the pattern $z(x)$. The complexity $C_1(z(\cdot), c)$ is the number of connected components of the set $D_{c,t} = \{x : \ z(x) = c\}$.

Theorem 1.2. Consider chain (1.1). The number C_1 of the connected components of the pattern $u_1^T(x)$ can be bounded from above by

$$C_1 < 2^{(r_\theta + Tm)^2}(d_\theta + TdegP)^{O(r_\theta + Tm + n)}. \tag{1.3}$$

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 (1.1), where r_θ is the sum of the lengths of Pfaffian chains (see [6] for all necessary definitions and bounds on 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 that defines σ. Analogous estimates can be obtained for other complexity measures (see [11]).

4 Outline of proofs

A brief proof of Theorem 1.1 can be obtained by the following Lemma.

Superposition Lemma. Consider a family consisting of p circuits (1.1) generating functions $u_{i,s}^t$, where $t = 0, ..., T_1$, $s = 1, ..., p$ and $i = 1, ..., m_s$ (where the index s marks the s-th circuit). Denote by $\mathbf{u}^t$ a vector function composed of all the functions $u_{i,s}^t$, i.e., $\mathbf{u}^t = (u_{1,1}^t, u_{2,1}^t, ...u_{m_1,1}^t, ..., u_{m_p,p}^t)$.

Suppose $z^t(x) = F(\mathbf{u}^t(x))$, where F is a continuous function defined on N -dimensional cube $Q = [0,1]^N$ and $N = \sum_{s=1}^p m_s$ is the number of the functions involved in the circuits. (In other words, the target pattern can be expressed through the functions generated by the family).

Then for any $\epsilon > 0$ there exists a circuit (1.1) satisfying (1.3) with $T_0 = 2$ and $T = T_1 + 2$.

The main idea of the proof is based on the well known biological fact: networks have a modular structure [9, 7]. Moreover, we use the following approximation result ([8, 12]): for any $\kappa > 0$ there exist M and A_{kjs}, b_k, η_k such that

$$|\sigma^{-1}(z) - \sum_{k=1}^{M} b_k \sigma(\sum_{j,s} A_{kjs} u_{j,s} - \eta_k)| < \kappa, \quad u_{j,s} \in Q^N. \tag{1.4}$$

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

$$v_k^{t+1} = \sigma(\sum_{j,s} A_{kjs} u_{j,s}^t - \eta_k), \quad w^{t+1} = \sigma(\sum_{k=1}^{M} b_k v_k^t). \tag{1.5}$$

Relations (1.4) and (1.5) imply approximation (1.2) (we set $u_1 = w$) if $\kappa = \kappa(\epsilon)$ is a sufficiently small and M is large enough.

Theorem 1.1 follows from this Lemma. To show it, we define the circuit by $u^t_{m+1} = \sigma(u_{m+1}), u^t_i = \sigma(\theta_i), i = 1, 2, ..., m$.

This proof gives an effective algorithm solving the pattern generation problem. Indeed, the key step of the proof (approximation (1.4)) can be performed by constructive procedures ([8, 12]).

To prove Theorem 1.2, we notice that, under our hypothesis on σ (the derivative of σ is a polynomial of σ), iterations (1.1) defines a Pfaffian chain. Thus, Khovanskii estimates [6] can be applied and we obtain (1.3) by an induction (see [10, 11]).

References

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

[2] H. Meinhardt, in Mathematical Models for Biological Pattern Formation, eds. P. K. Maini, H. G. Othmer, IMA Volumes in Mathematics and its Applications, 121, Springer, 2000.

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

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

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

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

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

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

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

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

[11] S. Vakulenko and D. Grigoriev, Complexity of patterns generated by genetic circuits and Pfaffian functions, Preprint IHES, 2003.

[12] Vakulenko S., Dissipative systems generating any structurally stable chaos, Adv. Differential Equations. 5 (2000) 1139-1178.