Stability Analysis of Gene Expression Patterns by Dynamical Systems and a Combinatorial Algorithm

M. Taştan[†], *T. Ergenç*[†], *S.W. Pickl*[*], *G.-W. Weber*[†]

[†]Institute of Applied Mathematics, METU, Ankara, Turkey

[*]Department of Computer Science, Universität der Bundeswehr, Munich, Germany

Abstract

The emerging field of modern biosciences and biotechnologies ask for appropriate and refined methods from applied mathematics. Modeling and prediction of gene-expression patterns have an important place here. In our paper, we deepen the analytical understanding and the algorithmical treatment by including affine terms on the right-hand side of the nonlinear differential equations and using Runge-Kutta (e.g., Heun) rather than Euler discretization. In the center of our research, there is the investigation of stability which we motivate in terms of biology and medicine, and detect by modified *Brayton and Tong algorithm* applied to the corresponding time-discrete dynamics in an extended space. This paper pays special attention to the motivating and analytical preparations of the combinatorial algorithm which based on the observation of polyhedra.

Keywords: Gene-Expression Data, Computational Biology, Mathematical Modeling, Prediction, Dynamical System, Runge-Kutta Discretization, Stability.

1 Introduction

Stability has a positive meaning in science and technology [10]; in biological sciences and technologies, and in medicine, it may has meaning of a disease coming to a rest (recovering), etc.. On the other hand, stability could also negatively be interpreted: a biosystem which is stable in the sense of insensitivity or lack of flexibility to changes in the environment is threatened in its existence. Finally, if a model behaves unstable, or unbounded, this contradicts with natural and technical conditions such as expression levels of genes lying in a certain bounded interval, then this instability can lead us to reject the hypothesis given by our mathematical model. This paper is a contribution to mathematical stability analysis applied to gene expression patterns, and based on an improved modeling compared with former approaches [7, 8, 22].

The process by which gene information is converted for producing cell structures and cell functions is called *gene* (or *protein*) *expression*. There are two main process events: *transcription and translation*. After them, steps like *folding*, *post-translational modification* and *targeting* occur up to the protein product, we leave these details to [15]. Analysis of *mRNA* being an exact copy of the DNA coding regions can be well used to explore the process in coding regions of DNA. More importantly, the measure of gene expression can be determined from the genomic analysis at the mRNA level [18]. Both genomic and environmental factors affect the expression levels. For example, environmental effects including stress, light, temperature and other signals cause some changes in hormones and in enzymatic reactions which influence the gene expression level. Because of this, mRNA analysis informs us not only about genetic aspects on an organism but also about the dynamical changes in environment of that organism. For most genes, protein levels are defined by steady state mRNA levels [20]. Thus, quantitative expressions at mRNA level provide important clues about the underlying dynamics. Peculiar changes in monitoring mRNA levels generally refer to drug treatment, shocks, disease or metabolic shifts.

1.1 Microarray Technology

The array-based *microarray technology* monitors thousands of different RNA molecules simultaneously revealing their expression patterns and perturbed subsequent cellular pathways. One of the most frequently used microarray applications [3] is to compare gene expression levels of the same cell type like healthy cell and diseased cell under two different conditions. Such application can give vital information on the reasons of diseases. Expression analysis is the recent main large-scale application

of microarrays, it is followed by DNA variation on a genome-wide scale [4]. Both applications share similar requirements but differ in some crucial aspects that have resulted in two different types of microarrays.

1.2 Evaluating the Expression Data

The expression values for large numbers of genes are quickly monitored by microarray experiments. One of the goals researchers have in mind is to clarify the precise connections of the *genetic network*: mathematically speaking a graph consisting of nodes representing genes and with the edges and their weights representing the influence which the genes mutually exercise. Here, the nodes themselves can also be viewed as a function obtained by combining basic inputs. For each gene it is aimed to find and to predict which and how much other genes influence it. Different mathematical methods have been developed for construction and analyzing such networks. In this study, we refine the model derived from differential equations by adding shift terms and by extending space. These dynamical systems will be characterized by *matrices* which are encoding our genetic networks [19]

2 Modeling Gene Networks with Ordinary Differential Equations

Easily accessible data through databases make modeling techniques popular. Based on these experimental data it is aimed to make reliable future predictions and simulations and to find the correlation between genes.

There are several modeling approaches, namely, Bayesian networks, Boolean networks, models derived from ordinary or piece-wise linear differential equations, hybrid systems modeling and etc.. All these methods have both advantages and disadvantages [20, 30] concerning goodness of data fit, computation time, capturing dynamics well, stability and other qualitative or quantitative aspects.

Differential equations are one of the most widely used modeling formalisms in mathematical biology. First of all, their more detailed representation of regulatory interactions can provide a more accurate understanding of the physical systems. Secondly, there is a large body of dynamical systems theory that can be used to analyze such models. Thirdly, concerning that biological systems evolve in continuous time, we prefer to use the systems of differential equations.

A differential relation between variables of gene networks is generally represented in the form of ordinary differential equations (ODEs)

$$\dot{E}_i = f_i(E) \qquad (i = 1, 2, ..., n),$$

where $E = (E_1, E_2, ..., E_n)^T$ is the vector of positive concentrations of proteins, mRNAs, or small components, $f_i : \mathbb{R}^n \to \mathbb{R}$ are nonlinear functions and n being the number of genes.

A first differential equation or dynamical system model consisting of mRNA and protein concentrations was proposed by *Chen, He and Church* [5] in the form of $\dot{E} = ME$, where M is a constant matrix and the vector E comprises the expression level of individual genes. Later on, *De Hoon and Imoto* [12] used this linear model on mRNA data of *Bacillus subtilis* to estimate M with maximum likelihood estimation method. In 2001, *Sakamoto and Iba* [17] proposed the more flexible model

$$\dot{E}_i = f_i(E_1, E_2, ..., E_n),$$

with f_i being functions of $E = (E_1, E_2, ..., E_n)^T$ determined by genetic programming and least-squares methods.

The models described above were studied and improved by *Gebert, Lätsch, Pickl, Weber and Wünschiers* with many ideas. In [8], they regarded the model $\dot{E} = M(E)E$ in which the matrix M, not usually a constant matrix, depends on E. In the same study, for the least-squares optimization problem on finding an approximate model, the solution space is restricted by assuming that number of regulating factors for each gene is bounded.

3 The State of the Art

Let the n-column vector $E = E(t)$ consist of gene expression patterns at different times t. We denote the given finite set of experimental results as $\bar{E}_0, \bar{E}_1, ..., \bar{E}_{l-1}$, where each $\bar{E}_m \in \mathbb{R}^n$ corresponds to the gene profile taken at time $\bar{t}_m$ and the sample times are increasing.

Gebert et al. [9] refined the time-continuous model ($\mathcal{CE}$) first formulated by *Chen et al.* by taking into account that the interaction between variables is nonlinear but the number of associated regulating influences is bounded. This model was represented by the multiplicative nonlinear form

$$(\mathcal{CE}) \qquad \dot{E} = M(E)E.$$

Here, we refer to corresponding initial values $E(t_0) = E_0$. Note that ($\mathcal{CE}$) is homogeneous and autonomous (i.e., the right hand-side depends on the states E but not on time t). This implies that trajectories do not cross themselves. The matrix $M(E)$ is defined component-wise by a family of any class of functions including unknown parameters. For example, for a (2×1)-vector $E = (E_1, E_2)^T$, the matrix $M(E)$ could be

$$M_{\substack{a_1,a_2,a_3,a_4 \\ a_5,a_6,a_7,a_8}} := \begin{pmatrix} a_1E_1^2 + a_2E_1E_2 & a_3E_2\cos(E_1) + a_4 \\ a_5\cos(E_2) + a_6E_1 & a_7E_1^2 + a_8E_2 \end{pmatrix}.$$

We note that the polynomial, trigonometric, but otherwise also exponential, etc., entries represent the growth or other kinds of changes in the concentrations. In this example, there are eight parameters in total.

Now, two different stages of problem come into consideration concerning the parametrized entries of the matrices $M(E)$. Firstly, the optimization problem of discrete (least-squares) approximation which can be written as

$$\underset{\alpha}{\text{minimize}} \quad \sum_{k=0}^{l-1} ||M_\alpha(\bar{E}_k)\bar{E}_k - \dot{\bar{E}}_k||^2.$$

Here, the least-squares methods of linear and nonlinear regression are used to estimate the vector α of a first part of the parameters to fit the set of given experimental data and to characterize the statistical properties of estimates. Secondly, we investigate which components of the remaining parameter vector β produce a stable, which ones an unstable influence on the dynamics. For a closer presentation of this two-stage problem from parametric optimization, we refer to [8, 14].

4 Model with Quadratic Polynomials

An extension to ($\mathcal{CE}$) was considered by *Yılmaz* [22] and *Yılmaz et al.* by proposing

$$\dot{E} = F(E),$$

where $F = (F_1, F_2, ..., F)^T$ is a tuple of functions depending on $E \in \mathbb{R}^n$. More specifically, for representing the influence of gene i to gene j the authors considered the quadratic (constant, linear) functions $f_{j,i}(x) = a_{j,i}x^2 + b_{j,i}x + c_{j,i}$, where $x = E_i$ denotes the concentration of $gene_i$ and $a_{j,i}, b_{j,i}, c_{j,i} \in \mathbb{R}$. Please note that in comparison to the model ($\mathcal{CE}$) with its multiplicative form $\mathbb{M}(\mathbb{E})$, now the vector $C \in \mathbb{R}^n$ coming from the absolute effects $c_{j,i}$ means a parametrical enrichment, an additive "shift" on the right-hand side. In [22], the least-squares approximation errors of linear and nonlinear, in fact quadratic, models are compared.

5 Our Generalized Model

The model extended by *Yılmaz et al.* [22] allows the nonlinear interactions and uses affine linear terms as shifts. However, the recursive iteration idea mentioned in [7] is lost by these shift terms, at the first glance. Thus, we again turn to ($\mathcal{CE}$) by making following *affine* addition:

$$(\mathcal{ACE}) \qquad \dot{E} = M(E)E + C(E).$$

Here, we defend that additional column vector $C(E)$ can represent the environmental perturbations and provide us better least-squares approximations which has already been guaranteed by regarding the important and the basic case where $C(E)$ is constant, i.e., $C(E) \equiv C$. Differently from $M(E)E$, the second term (*shift*) $C(E)$ does not need to reveal E as a factor, e.g., exp or cos. In case where $M(E)$ and $C(E)$ are polynomial, component-wise understood, $M(E)E$ may have a higher degree than $C(E)$.

Our approach in overcoming the more complex form of ($\mathcal{ACE}$) algorithmically is that $C(E)$ can be written as

$$C(E) = \check{M}(E)\check{E},$$

where

$$\check{M}(E) := \mathrm{diag}(C^T(E)) = \begin{pmatrix} C_1(E) & & & & 0 \\ & C_2(E) & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ 0 & & & & C_n(E) \end{pmatrix} \quad \text{and} \quad \check{E} := \begin{pmatrix} \check{E}_1 \\ \check{E}_2 \\ \cdot \\ \cdot \\ \cdot \\ \check{E}_n \end{pmatrix}.$$

In fact, we shall see by means of the corresponding initial value $\check{E}(t_0) = e$ ($e := (1, 1, ..., 1)^T$) that the time depending variable $\check{E}$ is constant (i.e., $\dot{\check{E}} = 0$) and identically $\check{E} \equiv e$. In this sense, $(\mathcal{ACE})$ is equivalent to

$$\dot{E} = M(E)E + \check{M}(E)\check{E}.$$

Let us define the vector and the matrix

$$\mathbb{E} := \begin{pmatrix} E \\ \check{E} \end{pmatrix} \quad \text{and} \quad \mathbb{M}(\mathbb{E}) := \begin{pmatrix} M(E) & \check{M}(E) \\ 0 & 0 \end{pmatrix}.$$

so that we end up with the following form of an extended initial value problem

$$(\mathcal{CE})_{ext} \qquad \dot{\mathbb{E}} = \mathbb{M}(\mathbb{E})\mathbb{E} \qquad \text{and} \qquad \mathbb{E}_0 := \mathbb{E}(t_0) = \begin{pmatrix} E(t_0) \\ \check{E}(t_0) \end{pmatrix} = \begin{pmatrix} E_0 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{pmatrix}.$$

In this study, we (without further drawbacks) combine and benefit from both the affine term structure to model gene expression patterns for better least-squares approximations and more accurate future predictions, and the time-continuous iterative matrix multiplication approach by means of higher dimension $2n$.

5.1 Time Discretization

Discretization concerns the process of transferring continuous models and equations into discrete counterparts. Numerical solution generated by simulating the behavior of system governed by ODEs, initiated at t_0 with given initial value E_0, is an approximation to the solution at discrete set of points. We follow trajectories with approximate solution values. Hence, choosing a suitable numerical method applied on the time-continuous model is an extremely important task. Euler's method, the simplest case of time-discretization, have been used for gene expression patterns, but we know that it is slow and inaccurate. Thus, on our way of using more refined and convincing techniques, we use a Runga-Kutta discretization method.

5.1.1 Runga-Kutta Method

While solving ODEs numerically, we face with two kinds of errors, namely, the rounding error as a result of finite precision of floating-point arithmetic and, secondly, the truncation error associated with the method used. For example, in Euler's method the truncation error is far larger because the curve $E(t)$ is approximated by a straight-line between the end-points t_k and t_{k+1} of time intervals. In addition, Euler's method evaluates derivatives at the beginning of the interval, i.e., at t_k which makes the method asymmetric with respect to the beginning and the end of the interval. Hence, more symmetric integration methods like *Runge-Kutta method* (RK), which takes into account that the midpoint of the interval can be applied on the the system $(\mathcal{CE})_{ext}$. Runge-Kutta methods have the advantage of stability which is closer to the stability of the given time-continuous model.

RK methods use only the information at time t_k, which makes them self-starting at beginning of integration, and also makes methods easy to program, which accounts in part for their popularity [11].

A central idea of applying RK methods to model of gene expression patterns was first introduced by *Ergenç and Weber* [6]. Here, we illustrate the application of a different RK method, called *Heun's method*. Heun's method is a modified version of Euler's method, more illustrative, explicit and the simplest case of the Runge-Kutta approach. In our extended space it is formulated as follows:

$$\mathbb{E}_{k+1} = \mathbb{E}_k + \frac{h_k}{2}(k_1 + k_2), \tag{1}$$

where

$$\begin{aligned} k_1 &= \mathbb{M}(\mathbb{E}_k)\mathbb{E}_k, \quad \text{and} \\ k_2 &= \mathbb{M}(\mathbb{E}_k + h_k k_1)(\mathbb{E}_k + h_k k_1). \end{aligned}$$

More explicitly, instead of (1) we write

$$\begin{aligned} \mathbb{E}_{k+1} &= \mathbb{E}_k + \frac{h_k}{2}\mathbb{M}(\mathbb{E}_k)\mathbb{E}_k + \frac{h_k}{2}\mathbb{M}(\mathbb{E}_k + h_k\mathbb{M}(\mathbb{E}_k)\mathbb{E}_k)(\mathbb{E}_k + h_k\mathbb{M}(\mathbb{E}_k)\mathbb{E}_k), \\ \Leftrightarrow \quad \mathbb{E}_{k+1} &= [I + \frac{h_k}{2}\mathbb{M}(\mathbb{E}_k) + \frac{h_k}{2}\mathbb{M}(\mathbb{E}_k + h_k\mathbb{M}(\mathbb{E}_k)\mathbb{E}_k)(I + h_k\mathbb{M}(\mathbb{E}_k))]\mathbb{E}_k. \end{aligned}$$

Defining

$$\mathbb{M}_k := I + \frac{h_k}{2}\mathbb{M}(\mathbb{E}_k) + \frac{h_k}{2}\mathbb{M}(\mathbb{E}_k + h_k\mathbb{M}(\mathbb{E}_k)\mathbb{E}_k)(I + h_k\mathbb{M}(\mathbb{E}_k)),$$

we get the following discrete equation

$$(\mathcal{DE})_{ext} \quad \mathbb{E}_{k+1} = \mathbb{M}_k\mathbb{E}_k.$$

Thus, we iteratively approximate the next state from the previous one. We note that since the experimental results are represented as $\bar{E}_0, \bar{E}_1, ..., \bar{E}_{l-1}$, we can represent the approximations by $\widehat{\mathbb{E}}_l, \widehat{\mathbb{E}}_2, ..., \widehat{\mathbb{E}}_{l-1}$ in the following way: setting $\widehat{\mathbb{E}}_0 = \bar{E}_0$, the k^{th} approximation is calculated as

$$\widehat{\mathbb{E}}_k = \mathbb{M}_{k-1}(\mathbb{M}_{k-2}...(\mathbb{M}_1(\mathbb{M}_0\bar{E}_0))) \quad (k \in \mathbb{N}_0).$$

Having a multiplicative formula for predictions has a great analytical and numerical advantage. Now, according to our motivations of stability analysis given in Section 1, these iterative matrix multiplications in front of the given initial state $\bar{E}_0$ force us to consider the stability and boundedness of the solution. Thus, we investigate the questions concerning how products of matrices $\mathbb{M}_k$ look like, what is the product structure and what does the block structure say about boundedness or unboundedness of the products of finitely many matrices.

6 Algebra of Matrix Products

Let us remember that the matrix in the time-continuous model has the *canonical* form

$$\mathbb{M}(\mathbb{E}) = \begin{pmatrix} M(E) & \check{M}(E) \\ 0 & 0 \end{pmatrix}.$$

These matrices help us for defining relation between genes and understanding the structure of gene networks. The product of two matrices having this block form is again a matrix in the same structure, because for any $X, Y \in \mathbb{R}^n$ it holds:

$$\begin{aligned} \begin{pmatrix} M(X) & \check{M}(X) \\ 0 & 0 \end{pmatrix}\begin{pmatrix} M(Y) & \check{M}(Y) \\ 0 & 0 \end{pmatrix} &= \begin{pmatrix} M(X)M(Y) & M(X)\check{M}(Y) \\ 0 & 0 \end{pmatrix} \\ &:= \begin{pmatrix} \widetilde{M}(X,Y) & \widetilde{\check{M}}(X,Y) \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

Matrix multiplication is not needed in the case of the time-continuous model, but we try to understand whether our matrices $\mathbb{M}_k$ and their products in the time-discrete iterative system have some "canonical" block form or not. After some simplifications and by definition of $\mathbb{M}_k$, we find that

$$\mathbb{M}_k = \mathbb{I} + \frac{h_k}{2}\begin{pmatrix} M(E_k) & \check{M}(E_k) \\ 0 & 0 \end{pmatrix} + \frac{h_k}{2}\begin{pmatrix} A & \widetilde{A} \\ 0 & 0 \end{pmatrix} + \frac{h_k^2}{2}\begin{pmatrix} B & \widetilde{B} \\ 0 & 0 \end{pmatrix},$$

where $\mathbb{I} = I_{2n}$ (unit matrix of type $(2n) \times (2n)$) and

$$\begin{aligned} A &:= M\left(E_k + h_k\left(M(E_k)E_k + \check{M}(E_k)\check{E}_k\right)\right), \\ \widetilde{A} &:= \check{M}\left(E_k + h_k\left(M(E_k)E_k + \check{M}(E_k)\check{E}_k\right)\right), \\ B &:= M\left(E_k + h_k\left(M(E_k)E_k + \check{M}(E_k)\check{E}_k\right)\right)M(E_k) \quad \text{and} \\ \widetilde{B} &:= M\left(E_k + h_k\left(M(E_k)E_k + \check{M}(E_k)\check{E}_k\right)\right)\check{M}(E_k). \end{aligned}$$

We conclude that $\mathbb{M}_k$ has its final *canonical* block form

$$\begin{pmatrix} \widehat{M(E_k)} & \widehat{\check{M}(E_k)} \\ 0 & I_n \end{pmatrix}.$$

Here, one of our main questions concerns *iterative multiplication* of matrices having the same form with model $\mathbb{M}_k$. In the next section, for our stability analysis we have to study these matrices $\mathbb{M}_k$ in detail. What a form has the product of two and, by induction, finitely many matrices $\mathbb{M}_k$? By using $\hat{A}, \hat{B}, \hat{C}, \hat{D}$ to represent the corresponding block matrices; we calculate:

$$\begin{pmatrix} \widehat{A} & \widehat{B} \\ 0 & I_n \end{pmatrix} \begin{pmatrix} \widehat{C} & \widehat{D} \\ 0 & I_n \end{pmatrix} = \begin{pmatrix} \widehat{A}\widehat{C} & \widehat{A}\widehat{D} + \widehat{B} \\ 0 & I_n \end{pmatrix} =: \begin{pmatrix} \widehat{K} & \widehat{L} \\ 0 & I_n \end{pmatrix}.$$

We observe that any finite product of matrices in the extended space preserves the same structure as a single matrix $\mathbb{M}_k$. In fact, multiplying any canonical matrix $\mathbb{M}_k$ by a vector $(E^T, e^T)^T$ reproduces a vector $(\widetilde{E}^T, e^T)^T$ of the same type. For this reason, there is no restriction if we focus our attention on the first n coordinates of the vectors and on the first n rows of our matrices.

By linear algebra it is easy to see that the matrices

$$\begin{pmatrix} \widehat{K} & \widehat{L} \\ 0 & I_n \end{pmatrix} \quad \text{and} \quad \widehat{K}$$

have the same eigenvalues if we disregard the ones coming from the $n \times n$ unit matrix I. In fact, the additional eigenvalue 1 has its algebraic multiplicity equal to its geometrical multiplicity as it is being requested for the eigenvalue λ with $|\lambda| = 1$ to ensure stability (cf. [10]). This enables us for doing a similar, n-dimensional stability analysis performed for $(\mathcal{CE})$ in [7].

From the point of view of mathematics, stability is a condition on the behavior of dynamical systems under initial perturbations around equilibrium points. This can be thought as a characterization of environmental changes (perturbation) given to the system, of disease or of the treatment of the cell by some medicine or radiation. Since gene expression values lie in a bounded region, stable solutions can refer to a better goodness of data fit (see Figure 1).

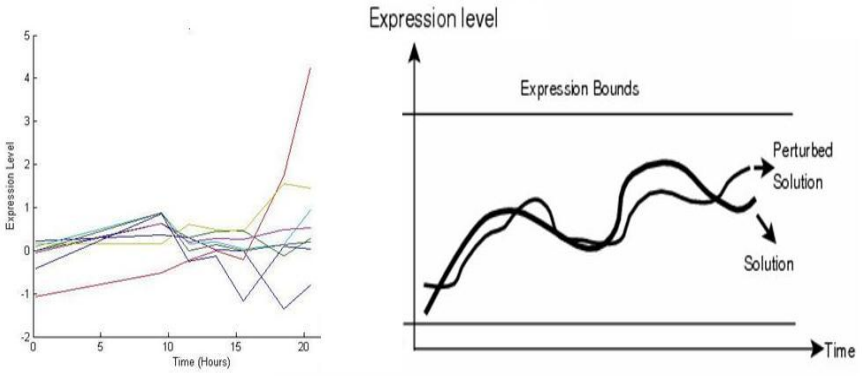

Figure 1: Real data [19] and model stability .

Here we start with mathematical definition of stability of a time-continuous system:

Definition 1 *A point $E^* \in \mathbb{R}^n$ is called an equilibrium point of system (S) $\dot{E} = f(t, E)$ where $(t, E) \in \mathbb{R} \times \mathbb{R}^n$ if $f(t, E^*) = 0$ for all $t \in \mathbb{R}$. An equilibrium E^* of (S) is called* stable *(in the Lypunov sense) if for every $\varepsilon > 0$ there exists a $\delta = \delta(\varepsilon) > 0$ such that it satisfies at time $t = t_0$ $||E(t) - E^*|| \leq \delta$ and for all $t > t_0$ $||E(t) - E^*|| < \varepsilon$.*

A common method for demonstration of stability is to find a Lyapunov function for that system. However, the problem of finding a suitable Lyapunov function arises because there is no general rule for establishing such functions [2]. Therefore, an algorithmic method which studies stability and introduces Lyapunov functions in the time-discrete case has first been introduced by *Brayton and Tong* [2]. In this paper, we are focussing more and the analytical side of our research. Herewith, we prepare the algorithm, our insight in its working. The algorithmical theory is in detail explained in [1, 2] for the case of Euler discretization, and in [6, 19], corresponding with our extended model and Runge-Kutta discretization used. In any of these cases, the algorithm bases on the study of a sequence of polyhedra by which we observe the virtue on matrices applied, i.e., stability or instability of the dynamics considered to become detected.

The stability of time-continuous model $(\mathcal{CE})_{ext}$ describing gene expression profiles is strongly related with the stability of time-discrete system $(\mathcal{DE})_{ext}$ introduced in the next section by the following theorem.

Theorem 2 *[2] Let the map $\mathbb{E} \mapsto \mathbb{M}(\mathbb{E})$ be Lipschitzian. If the time-discrete system $\mathbb{E}_{k+1} = \mathbb{E}_k + h_k\mathbb{M}(\mathbb{E}_k)\mathbb{E}_k \quad (k \in \mathbb{N}_0),\ \mathbb{E}_0 \in \mathbb{R}^{2n}$ some appropriate $h_{max} > 0$ being given, is stable for all values $h_k \in [0, h_{max}]$, then the continuous system $\dot{\mathbb{E}} = \mathbb{M}(\mathbb{E})\mathbb{E}$ is also stable.*

Proof. See [2]. ■

In [2], it is shown that the stability of time-discrete model is determined by stability of a set of matrices, $\mathcal{M} = \{\mathbb{M}_0, \mathbb{M}_1, ..., \mathbb{M}_{l-1}\}$, derived from discretely approximating the set

$$\{\mathbb{M}(\mathbb{E}, h)|\ \ \mathbb{E} \in \mathbb{R}^{2n},\ h \in [0, h_{max}]\},$$

where $\mathbb{M}(\mathbb{E}, h) := \mathbb{I} + h\mathbb{M}(\mathbb{E})$.
Since, however, we are using RK method, in our case, $\mathbb{M}(\mathbb{E}, h)$ takes the form :

$$\mathbb{M}(\mathbb{E}, h) := I + \frac{h}{2}\mathbb{M}(\mathbb{E}) + \frac{h}{2}\mathbb{M}(\mathbb{E} + h\mathbb{M}(\mathbb{E})\mathbb{E})(I + h\mathbb{M}(\mathbb{E})).$$

In fact, we are discretizing the function $\mathbb{M}(\mathbb{E}, h)$ in a way that the values of the implied matrix entries are taken at their maximal or minimal values, and h (by h_{max}) chosen extremally as well. When iteratively applying the resulting entire matrices to polyhedral sets, then we represent and understand the worst-case growth behavior of any finite matrix multiplication, i.e., whether instability is holding.

7 Stability of a Set of Matrices

Let $\mathcal{M} = \{\mathbb{M}_0, \mathbb{M}_1, ..., \mathbb{M}_{l-1}\}$ be a set of given real matrices. (There should not be any confusion with the usage of $\mathbb{M}_k$ for the k^{th} iterate of the time-discrete dynamics, here.) We will consider the larger multiplicative semigroup $\mathcal{M}'$ containing all finite products of matrices produced from $\mathcal{M}$. In other words,

$$\mathcal{M}' = \{\prod_{s=1}^{k} \mathbb{M}_s^{l_s} :\ \mathbb{M}_s \in \mathcal{M},\ \ l_s \in \mathbb{N}\ (s \in \{1, 2, ..., k\}),$$
$$\mathbb{M}_s \neq \mathbb{M}_{s+1}\ \ \forall s \leq k-1,\ \ k \in \mathbb{N},\ \sum_{s=1}^{k} l_s = p,\ \ p \in \mathbb{N}\}.$$

Since our dynamical analysis bases on the linear algebra of matrices, especially, on the spectral study of eigenvalues, we have to locate our study over the complex numbers rather than the reals.

Definition 3 *The set $\mathcal{M}$ is stable if for every neighborhood of the origin $U \subseteq \mathbb{C}^n$ there exists another neighborhood of the origin $\tilde{U}$ such that, for each $\mathbb{M} \in \mathcal{M}'$ it holds: $\mathbb{M}\tilde{U} \subseteq U$.*

Brayton and Tong proved that $\mathcal{M}$ is stable iff $\mathbb{B}^*$ is bounded, where

$$\mathbb{B}^* := \bigcup_{j=0}^{\infty} \mathbb{B}_j, \quad \text{with} \quad \mathbb{B}_k := \mathcal{H}\left(\bigcup_{j=0}^{\infty} \mathbb{M}_{k'}^j \mathbb{B}_{k-1}\right) \quad \text{and} \quad k' \equiv k-1 \ \ (mod\ m).$$

The algorithm of Brayton and Tong has a great advantage: We can analyze a set of matrices, derived as explained above. and we can decide for which combination of these matrices the underlying dynamical system is stable or not. As the matrices represent biological information (which is influenced by errors, estimations, pollution, etc.) the approach can support biological work in a very comfortable way. The question of stability is answered by automatically generated Lyapunov functions. There are also other procedures possible. With the proposed algorithm of Brayton and Tong which is implemented by Pickl, Taştan and Weber, the frontier line between stability and instability regions can be analyzed in detail. Biological effects can be studied in a very fine way. It might be very interesting to deepen the insights about the effect of the additional "shift term" $C(E)$, representing a (deterministic) error function or even a stochastic error variable.

8 Conclusion

In this study, from the viewpoints of statistical learning and dynamical system theory, for making models more realistic, approximative and better prepared for stability analysis, we improved the mathematical model and stability analysis. Here, we analyzed the system by considering advantages of the Runge-Kutta methods. Thus, our study may help these techniques to achieve new insights by means of mathematical modeling, dynamical systems, optimization and combinatorial algorithms. This paper focused very much to a motivation and on analytical preparation of our algorithm. Please

find this refined and discrete combinatorial procedure in detail explained and by an example illustrated in [7, 19]. In fact, when applying it on real-world data, we will see it serve.

In the future, it is our aim to find the regions of stability algorithmically based on real data. Within the regions of stability, we accept the "hypothesis" of our mathematical model, i.e., parameter estimation has been done in a satisfactory way. There are still a lot of challenges in the optimization of biosystems.

References

[1] Bechmann, E., *Analyse und Konstruktion eines Algorithmus zur Untersuchung der Stabilität dynamischer Systeme*, Mathematischen Institut der Universität zu Köln, diploma thesis 2005.

[2] Brayton, R.K., and Tong, C.H., Stability of dynamical systems: A constructive approach, IEEE Transactions on Circuits and Systems 26, 4 (1979) 224-234.

[3] Causton, C.H., Quackenbush J., and Brazma, A., *A Beginner's Guide Microarray Gene Expression Data Analysis*, Blackwell Publishing (2003).

[4] Carbayo, M.S, Bornman, W., and Cardo C.C., DNA Microchips: technical and practical considerations, Current Organic Chemistry 4, 9 (2000) 945-971.

[5] Chen, T., He, H.L., and Church, G.M., Modeling gene expression with differential equations, Proc. Pacific Symposium on Biocomputing (1999) 29-40.

[6] Ergenç, T., and Weber, G.-W., Modeling and prediction of gene-expression patterns reconsidered with Runge-Kutta discretization, special issue at the occasion of seventith birthday of Prof. Dr. Karl Roesner, TU Darmstadt, of Journals of Computational Technologies 9, 6 (2004) 40-48.

[7] Gebert, J., Lätsch, M., Pickl, S. W., Weber, G.-W., and Wünschiers, R., An algorithm to analyze stability of gene-expression pattern, Discrete Appl. Math., accepted for publication (2005).

[8] Gebert, J., Lätsch, M., Pickl, S.W., Weber, G.-W., and Wünschiers, R., Genetic networks and anticipation of gene expression patterns, *Computing Anticipatory Systems: CASYS03 - Sixth International Conference*, AIP Conference Proceedings 718 (2004) 474-485.

[9] Gebert, J., Radde, N., and Weber, G.W., Modeling gene regulatory networks with piecewise linear differential equations, Preprint 37, Middle East Technical University, Institute of Applied Mathematics (2005).

[10] Guckenheimer, J., and Holmes, P., *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Springer, 1997.

[11] Heath, M., *Scientific Computing: An Introductory Survey*, McGraw-Hill (2002).

[12] Hoon, M.D., Imoto, S., Kobayashi, K., Ogasawara, N., and Miyano, S. Inferring gene regulatory networks from time-ordered gene expression data of Bacillus subtilis using differential equations, Proc. Pacific Symposium on Biocomputing (2003) 17-28.

[13] Jong, H.D., Modeling and simulation of genetic regulatory systems: a literature review, J. Comput. Biol. 9 (2002) 103-129.

[14] Jongen, H.T., and Weber, G.-W., On parametric nonlinear programming, Annals of Operations Research 27 (1990) 253-284.

[15] Klug, W.S., and Cummings, M.R., *Concepts of Genetics,* Prentice Hall (2003).

[16] Pickl, S. W., *Der τ-value als Kontrollparameter - Modellierung und Analyse eines Joint - Implementation Programmes mithilfe der kooperativen dynamischen Spieltheorie und der diskreten Optimierung*, Shaker-Verlag (1999).

[17] Sakamoto, E., and Iba, H., Inferring a system of differential equations for a gene regulatory network by using genetic programming, Proc. Congress on Evolutionary Computation (2001) 720-726.

[18] Schena, M., *DNA Microarrays*, Oxford University Press (2000).

[19] Taştan, M., *Analysis and Prediction of Gene Expression Patterns by Dynamical Systems, and by a Combinatorial Algorithm*, Institute of Applied Mathematics, METU, MSc Thesis (2005).

[20] Yamamoto, K.R., Steroid receptor regulated transcription of specific genes and gene networks, Ann. Rev. Genetics. 19 (1985) 209-252.

[21] Weber, G.-W., Özoğur, S., and Karasözen, B., Challanges in the optimization of biosystems I: parameter estimation of enzymatic reactions with genetic algorithm, Preprint 39, Middle East Technical University, Institute of Applied Mathematics (2005).

[22] Yılmaz, F.B., *A Mathematical Modeling and Approximation of Gene Expression Patterns by Linear and Quadratic Regulatory Relations and Analysis of Gene Networks*, Institute of Applied Mathematics, METU, MSc Thesis (2004).