Stability Analysis of Time Series Gene Expression Data

J. Gebert, M. Lätsch, S. W. Pickl, N. Radde [1], G. - W. Weber [2], Röbbe Wünschiers [3], Bob Veroff [4]

Keywords: Gene Expression Analysis, Stability Analysis

Abstract

The analysis of gene expression data is a challenging field of computational biology. We represented the behavior of time series gene expression patterns by a system of differential equations, which we analytically and algorithmically investigate. Therefore we use an algorithm which computes regions of stability and instability. It takes advantage of extremal points of polyhedra, which grow step by step, up to a possible stopping. The overall goal is to correlate the stability pattern with metabolic states. New numerical results will be presented.

1 Mathematical Modelling of Gene Expression Data

Several approaches to analyze time-series gene expression data have been presented. Chen *et al.* [3] proposed to use the differential equation $\dot{E} = ME$ to model gene expression data. M represents a matrix with constant entries and $E(t)$ the vector of mRNA and protein concentrations at time t. The vector $\dot{E}$ represents the change of concentrations in time. The modelling approach suggested by De Hoon *et al.* [5] is based on Chen *et al.*. However, only mRNA measurements and different preconditions are used. In 2001, Sakamoto and Iba [6] proposed a more flexible approach $\dot{E}_i = f_i(E_1, ..., E_n)\, \forall i = 1, ..., n$ with n being the number of genes and f_i being a function in $E_1, ..., E_n$. Genetic programming is used to construct the model which has also been proposed by Cao *et al.* To be more effective Sakamoto and Iba also used the least mean square method along with the genetic programming. The method worked very well for small samples. For a large-sized network they will search for methods of pre-processing to reduce the given networks to small-sized sub-networks. Our approach is a further extension, since we firstly make M dependent on E and secondly analyze the system for stability.

2 Data Sets

Our initial experiments were based on artificial data sets. Thereby, we could for example control the innode degrees. Currently, we are working on the time-series gene expression data set from a diauxic shift of yeast. This data-set and the experimental background can be found in [4].

[1]Institute of Mathematics, Center for Applied Computer Science, University of Cologne, Weyertal 80, 50931 Cologne, Germany. E-mail: gebert@zpr.uni-koeln.de, laetsch@zpr.uni-koeln.de, pickl@zpr.uni-koeln.de, radde@zpr.uni-koeln.de

[2]Institute of Applied Mathematics, Middle East Technical University, 06531 Ankara, Turkey. E-mail: gweber@metu.edu.tr

[3]Institute of Genetics, University of Cologne, Weyertal 121, 50931 Cologne, Germany. E-mail: robbe.wunschiers@uni-koeln.de

[4]Compputer Science Department, University of Mexico, 153 Farris Engineering Center Albuquerque, NM 87131. E-mail: veroff@cs.unm.edu

3 Mathematical Modelling

Step 1 Based on the data of microarray experiments we want to develop a system of differential equations

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

which describes the cell's process. The expression levels of the genes at time t are described by the vector $E(t) = E_1(t), ..., E_n(t)$, n being the number of genes in the microarray experiment, and $M(E)$ being a matrix depending on E.
How can $M(E)$ be calculated? For the time being we will cluster the data and use least squares to solve this problem.

Step 2 In the second step we want to analyze the system of differential equations for stability using an algorithm of Brayton and Tong [2, 1].
i) We apply Euler's discretization on the time-continuous system of differential equations $(\mathcal{CE})$ by which we describe the metabolic process. Based on our good differentiable approximation of the biochemical processes, the time-discretization by Euler's method supports and simplifies the analysis and approximate resolution of the continuous process $(\mathcal{CE})$. This method is not a very refined approximation but convenient and a strong tool, here. We get a set of matrices $\mathcal{M} = \{M_0, ..., M_{m-1}\}$ which we analyze for stability.
ii)We apply the algorithm of Brayton and Tong on the set $\mathcal{M}$ of matrices. This set is **stable**, if for every neighborhood of the origin $U \subset \mathbb{C}^n$ there exists another neighborhood of the origin $\tilde{U}$ such that, for each $M \in \mathcal{M}'$ it lasts $M\tilde{U} \subseteq U$. Brayton and Tong proved that $\mathcal{M}$ is stable, iff $\breve{B}$ is bounded, where

$$\breve{B} := \bigcup_{k=0}^{\infty} B_k\,, \text{ with } B_k := \mathcal{H}\left(\bigcup_{i=0}^{\infty} M_{k'}^{i} B_{k-1}\right) \quad \text{and} \quad k' = (k-1) \; \textit{modulo}\; m\,,$$

for $k \in \mathbb{N}$ and $B_0 \in \mathbb{C}^n$ a bounded neighborhood of the origin.
The algorithm constructs polytopes B_i consecutively and checks whether $\breve{B}$ is bounded and therefore whether the set $\mathcal{M}$ is stable.

References

[1] R. Brayton and C. Tong. Constructive stability and asymptotic stability of dynamical systems. *IEEE Transactions on Circuits and Systems*, 27(11):1121–1130, 1980.

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

[3] T. Chen, H. L. He, and G. M. Church. Modeling gene expression with differential equations. *Pac Symp Biocomput*, pages 29–40, 1999.

[4] I. V. B. P. DeRisi, J.L. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, 278(5338):680–6, 1997.

[5] M. D. Hoon, S. Imoto, and S. Miyano. Inferring gene regulatory networks from time-ordered gene expression data using differential equations. In C. H. S. Steffen Lange, Ken Satoh, editor, *Lecture Notes in Computer Science*, volume 2534, pages 267–274. Springer-Verlag, 2002.

[6] E. Sakamoto and H. Iba. Inferring a system of differential equations for a gene regulatory network by using genetic programming. *Proc. Congress on Evolutionary Computation*, pages 720–726, 2001.