The Second International Symposium on Optimization and Systems Biology (OSB'08)
Lijiang, China, October 31– November 3, 2008

Clustering Parameter Values for Differential Equation Models of Biological Pathways

Dong-Soo Kahng Doheon Lee

Dept. of Bio and Brain Engineering, KAIST 335 Gwahangno Yuseong-gu,
Daejeon, 305-701, KOREA.
{dskahng, dhlee}@biosoft.kaist.ac.kr

Abstract Dynamics of many biological systems can be modeled in the form of nonlinear differential equations, where variables represent concentrations of participating molecular species, and parameters specify dynamics coefficients such as reaction rates and activity levels. It has been one of the hardest problems to determine right parameters even after we have acceptable model equations for a particular biological pathway. In this study, we propose a parameter space clustering method based on top-down refinement. The whole parameter space of a given model is explored by means of randomized comparison and top-down stepwise refinement. After the process, we come up with clusters of parameter values, each of which shows similar dynamics of a particular model. We expect that each of the clusters may be associated to a distinct phenotypical state of a given biological pathway. A simplified model of the well-known JAK-STAT pathway is used to illustrate the clustering process, and show the applicability of this technique.

1 Introduction

Many complex systems of interactive proteins make interesting phenomena in cells such as DNA synthesis and metabolism. These cell physiological properties are represented as subtle molecular movements which are influenced by sophisticated regulatory networks. For understanding those movements, biochemical networks are represented by mathematical languages as dynamic systems and they are restructured by computer simulations [1]. Biochemical models which show diverse biological phenomena such as cell cycle in fission yeast are composed of various components such as genes, proteins and metabolites, and they can be represented as dynamical systems of nonlinear ordinary differential equations which have parameters such as rate constants and vary with time under interactions of components and external influences [2].

There were several previous studies that connect the difference of phenotypes on diseases or cell cycles to the difference of the dynamics for the corresponding dynamical systems with computer simulations and biological experiments. These studies have compared wild types with mutants: Budding yeast cell cycle [3], a signal transduction pathway in fission yeast cell [4], Calcineurin(CaN)-modulatory calcineurin-interacting protein (MCIP) signaling pathway [5], Janus-associated kinases and signal transducers and activators of transcription (JAK/STAT) and mitogen-activated protein kinases (MARK) with IL-6 signal transduction [6], and Wnt and ERK Pathways involved cancers [7].

Complex dynamical systems may have a number of nonlinear ordinary differential equations and parameters which orders are from tens to hundreds. The effective ranges of parameters in the dynamical systems might be determined by biochemical experiments, but the useful values are unknown in the most cases. Parameter estimation is needed to find a set of optimal parameters using computer simulations. Parameter estimation adjusts the model to reproduce the experimental results in the best possible way for a set of experimental data [8] The methods of parameter estimation are divided by local optimization methods and global optimization methods with search ranges. Local optimization methods finish when they find the first local minimum, then they have low computational time complexity and they do not guarantee the global optimum when they fall in a local minimum. Global optimization methods need high computational time complexity but they can solve the local minimum problem because they can find better values in the parameter space over local minimums using combination of an exploration step and a selection step [9]. The hybrid methods combining the global optimization methods and the Local optimization methods are possible. Firstly the global optimization methods are executed and the Local optimization methods are used in the narrow ranges [10].

Because of the high computational time complexity, if we search a set of optimal parameter values from the randomly generated sets of parameters in the all possible ranges of the parameter space composed of many parameters, it would be hard to obtain parameter estimation results after consuming reasonable computation time [11]. In this study, we propose a parameter space clustering methods based on top-down approach.

2 Dynamic Profiles of Biological Pathways

For the qualitative and quantitative analysis for dynamic profiles, pairwise metrics such as the Euclidean distance or the Pearson correlation with the distance between pairs of time series data could be used [12][13]. These metrics would be too sensitive with small changes so they are not appropriate to quantify the differences of dynamics among the biological pathways. More abstract metric is needed for effective analysis for dynamic profiles of biological pathways. There is alternative way to represent the changes in variables at periods qualitatively as vector representation corresponding to increase, decrease, no change, maximum and minimum, etc. [14]. For the quantitative analysis of the dynamical profiles, we choose a method that divide a time series by several intervals with same width and code increase, decrease and no change into 1, -l and 0, respectively.

Figure 1 shows a coding example of a time series for a variable in a differential equation The graph made of the time series is divided by intervals with same width. Up, down, and stay (increase/decrease/no change) in each interval is checked. And the coding vector values 1, -1, and 0 corresponding up, down, and stay, respectively, are recoded.

3 The Process of Parameter Space Clustering

The process of parameter space clustering is implemented as matlab modules. These steps show the process of parameter space clustering:

1) Initially in the p-dimensional parameter space which represents p parameters, choose k parameters for an investigation, take the ranges of the parameters, and divide each coordinate corresponding to the each chosen parameter by two equal sections. The values of remained p-k parameters are fixed as arbitrary numbers.

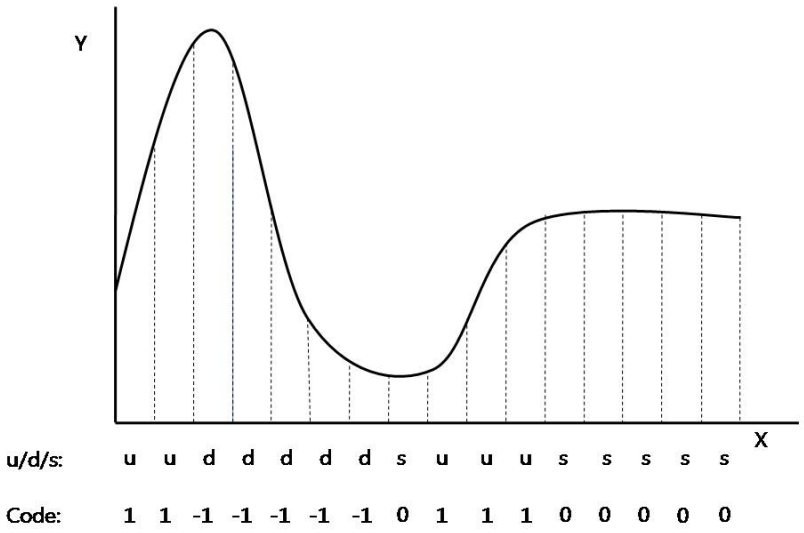

Figure 1: A coding example of a time series. X axis is the time, and Y axis is concentration of a protein. The graph is divided by intervals with same width. There are 16 intervals. u/d/s means up, down, and stay (increase/decrease/no change) in each interval. Code mean the coding vector values 1, -1, and 0 corresponding up, down, and stay, respectively.

2) For the set of 2^k parameter space sections, make randomly n pairs of parameters in the range of each section, obtain time series of the variables from numerical simulation of the dynamical systems of nonlinear ordinary differential equations with the given pairs of parameters, take changes of time series among simulation intervals as increase/decrease/no change, and make a comparison of similarities like the Euclidean distance corresponding to the time series change of each pair of variables.

3) Choose the one section having the minimum similarity from the all compared parameter space sections.

4) Divide the range of the selected parameter space section by two equally, and iterated the step 2 and the step 3 several times. The similarity obtained in the each iteration is compared with the similarities obtained in the previous iterations, and the minimum similarities are determined.

5) After given number of iterations, the sets of the parameter space sections are clustered with corresponding similarities as indices, and validate the separation of the sets of the parameter space sections with close values of similarities as a few clusters.

4 Simplified JAK-STAT Equation

These are the simplified JAK-STAT Equations [15] for the test of matlab implementation of parameter space clustering:

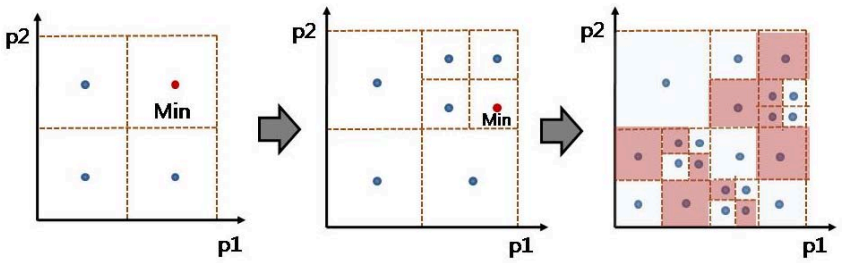

Figure 2: Example of parameter space clustering.

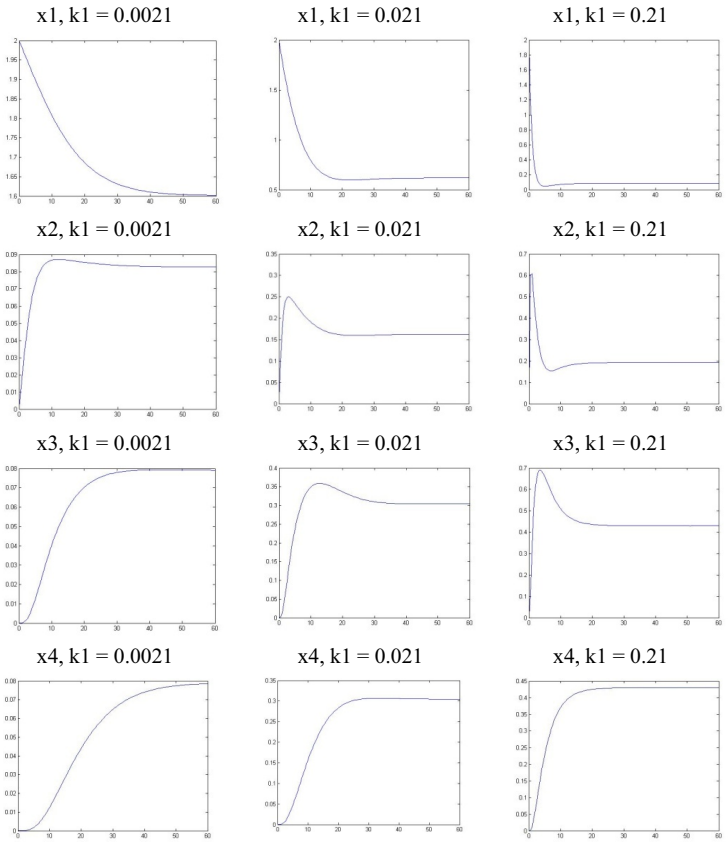

Figure 3: The graphs of x_1, x_2, x_3 and x_4 with different k_1 values. The horizontal axis is time and the vertical axis is the concentration. The k_1 values at 1^{st} , 2^{nd}, and 3^{rd} column are 0.0021, 0.021, and 0.21, respectively. k_2 = 2.46/min/mol, k_3 = 0.1066/min, k_4 = 0.10658/min, initial values $[x_1 x_2 x_3 x_4 x_5]$ = [2 0 0 0 5].

$$\begin{aligned}
\dot{x_1} &= -k_1 x_1 x_5 + 2k_4 x_4 \\
\dot{x_2} &= -k_2 x_2^2 + k_1 x_1 x_5 \\
\dot{x_3} &= -k_3 x_3 + 0.5 k_2 x_2^2 \\
\dot{x_4} &= -k_4 x_4 + k_2 x_2
\end{aligned}$$

where x_1, x_2, x_3 and x_4 are variables that indicate the concentration of unphosphorylated cytoplasmic STAT5, tyrosine phosphorylated monomeric cytoplasmic STAT5, tyrosine phosphorylated dimeric cytoplasmic STAT5, and nuclear STAT5, respectively. x5 indicates the concentration of the EpoRA protein which is regarded as constant value in this model. Figure 3 shows the graphs of x_1, x_2, x_3 and x_4 with different k_1 values.

The parameters are given in the study for the simulation; k_1 varies in the range of 0.021∼0.21/min, k_2 varies in the range of 0.246∼2.46/min/mol, k_3 = 0.1066/min and k_4 = 0.10658/min are fixed. The given initial values of the variables are $[x_1 x_2 x_3 x_4 x_5]$ = [2 0 0 0 5].

For the time series of x_1, x_2, x_3 and x_4 in the simulation of the given equations with matlab for 60 minutes, simulation time intervals are equally divided by 120, and one of the numbers 1/-1/0 is recorded for the increase/decrease/no change in each divided interval, respectively. The sum of the Euclidean distances corresponding to the randomly generated (k_1, k_2) pairs in each parameter space section is defined as the similarity, and the divided parameter space sections are clustered by 2 classes with various thresholds of the similarity.

5 Summary

We have proposed a parameter space clustering method based on the binary top-down approach with the similarity. The range of parameters in the parameter space has been equally divided by two. In each parameter space section, the sets of parameters have been randomly generated. The nonlinear ordinary differential equations connecting with the sets of parameters have been simulated. The similarities calculated from the simulation results of the nonlinear ordinary differential equations have been obtained. The divide parameter sections have been clustered by the corresponding similarity values.

It would be possible to construct simple dynamical systems artificially which show distinct changes following the change of parameter values and have a few separate states or to find dynamical systems which are well known from biological experiments. The validation of the implemented nonlinear parameter space clustering algorithms might use those dynamical systems. If the sets of parameter space sections of the new dynamical systems to represent complex biological phenomena which have not been known experimentally yet could be clustered using the validated nonlinear parameter space clustering algorithms, it would be possible to propose the range of parameters which are appropriate to the results of experiments and to show the connection between the set of parameters and the distinct biological state.

Acknowledgements

This work was supported by the Korean Systems Biology Program from the Ministry of Education, Science and Technology through the Korea Science and Engineering Foundation (No. M10309020000-03B5002-00000). This work was supported by the second

stage of the Brain Korea 21 Project in 2008. The authors would like to acknowledge the support from KISTI (Korea Institute of Science and Technology Information).

References

[1] Tyson JJ, Chen K, Novak B: Network dynamics and cell physiology. Nat Rev Mol Cell Biol. 2001, 2(12):908-16.

[2] Novak B, Tyson JJ: Modeling the Control of DNA replication in fission yeast, PNAS 1997, 94(17): 9147-9152

[3] Chen KC, Csikasz-Nagy A, Gyorffy B, Val J, Novak B, Tyson JJ: Kinetic Analysis of a Molecular Model of the Budding Yeast Cell Cycle, Mol Biol Cell. 2000, 11(1):369-91.

[4] Csikász-Nagy A, Kapuy O, Gyorffy B, Tyson JJ, Novák B: Modeling the Septation Initiation Network in Fission Yeast Cell, Curr Genet. 2007, 51(4):245-55.

[5] Switching feedback mechanisms realize the dual role of MCIP in the regulation of calcineurin activity, FEBS Lett. 2006, 580(25):5965-73.

[6] Singh A, Jayaraman A, Hahn J: Modeling Regulatory Mechanisms in IL-6 Signal Transduction in Hepatocytes, Biotechnology and Bioengineering 2006 ;95(5):850-62.

[7] Kim D, Rath O, Kolch W, Cho KH: A hidden oncogenic positive feedback loop caused by crosstalk between Wnt and ERK Pathways, Oncogene 2007, 26(31):4571-9

[8] Moles CG, Mendes P, Banga JR: Parameter Estimation in Biochemical Pathways: A Comparison of Global Optimization Methods, Genome Research 2003, 13(11):2467-2474.

[9] Allen NA, Chen KC, Shaffer CA, Tyson JJ, Watson LT: Computer Evaluation of Network Dynamics Models with application to Cell Cycle Control in Budding Yeast, Syst Biol (Stevenage). 2006, 153(1):13-21

[10] Zwolak JW, Tyson JJ, Watson LT: Globally optimised parameters for a model of mitotic control in frog egg extracts. Syst Biol (Stevenage). 2005 152(2):81-92.

[11] Duda RO, Hart PE: Pattern Classification and Scene Analysis. John Wiley, London, 1973.

[12] Brody CD: Correlations without synchrony, Neural Comput. 1999, 11(7):1537-1551.

[13] Pereda E, Quiroga RQ, Bhattacharya J: Nonlinear multivariate analysis of neurophysiological signals, Prog Neurobiol. 2005, 77(1-2):1-37

[14] Lovrics A, Csikasz-Nagy A, Zsely IG, Zador J, Turanyi T, Novak B: Time scale and dimension analysis of a budding yeast cell cycle model, BMC bioinformatics 2006, 7:494.

[15] Swameye I, Muller TG, Timmer J, Sandra O, Klingmuller U: Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. PNAS 2003, 100(3):1028-33.