NEW CAPABILITIES FOR LARGE-SCALE MODELS IN COMPUTATIONAL BIOLOGY

Casey S. Abbott[1], Eric L. Haseltine[2], R. Abraham Martin[1], and John D. Hedengren[1]
[1]Brigham Young University, Provo, Utah
[2]Vertex Pharmaceuticals, Cambridge, MA

Introduction

Advances in biomedical research have lead to an increase of experimental data to be interpreted in the context of reaction pathways, molecular transport, and population dynamics. Kinetic modeling is one way employed to interpret this data and is used in the pharmaceutical industry in developing clinical trials for new medications [1]. Kinetic modeling's role will only increase as large pharmaceutical companies look to scale back and be more focused, spending less on R&D while expecting more results. We propose that kinetic modeling offers one such mechanism.

One collection of kinetic models is built on the System Biology Markup Language (SBML), which includes hundreds of contributions. Many models in this collection have detailed reaction metabolic pathways that describe biological systems, including cause and effect relationships in the human body. While simulations of these biological systems have been successfully applied for many years, the alignment to available measurements continues to be a challenge. The best available solution techniques continue to limit the size of the reconciliation of models and measurements to small and medium size problems. This limits the usefulness of the models due to the many assumptions and simplifications that are required in order for the optimizer to be able to perform the parameter estimation.

Recent developments may have changed this situation: an optimization technique known as the simultaneous approach has shown promise in efficiently optimizing large models (thousands of variables and parameters) [2]. In this method, the model and optimization problem are solved simultaneously, as opposed to the traditional approach of solving the differential and algebraic equation (DAE) model sequentially. In the sequential approach, each iteration of the optimization requires the solution of the DAE model. Much of the recent development for the simultaneous approach has occurred in the petrochemical industry, where on-line process control applications require optimization of nonlinear models with many decision variables in the span of minutes. The applications drive petrochemical processes to produce more from existing processing units, stay within safe operating conditions, and minimize environmental emissions. Using the same solution techniques, major advances in the biological systems area are possible. One software that takes advantage of the simultaneous approach is APMonitor (APM).

Results

To verify that APM can accurately simulate biological models it was used to simulate a toy model and the results were compared to literature values. The model used is a basic model describing the HIV virus over thirty days with nine parameters, three variables and three differential equations [3]. APM was successfully able to replicate the results published in the literature. APM simulation was also verified with a model that describes the dynamics of HIV infection of CD4+ T cells [4]. This model is a little larger with nine parameters, four variables, and five DAE. This model was obtained from the BioModels Database and was manually converted to a format that could be used by APM. It was found that APM could accurately simulate this model and match values from literature and simulations in MATLAB.

Once it was shown that APM could simulate biological models accurately the next step was to verify the parameter estimation capabilities of APM. This was done using a HIV model similar to those mentioned above. In order to perform the parameter estimation the objective function was set to minimize the absolute error between the model and synthetic data. There was measurement noise of plus or minus 0.5 log order added to the synthetic data. All six parameters values were changed to see if APM could find the correct parameter values from several different starting points. It was found that APM was able to accurately find the correct parameter values. Figure 1 shows the concentration of HIV viruses from the synthetic data and the predicted model values with the estimated parameters. As seen in the figure APM was able to correctly find the parameters that allowed the model to fit the synthetic data. With this same model it was shown that APM allows for parallel processing, allowing multiple parameter estimations to be run simultaneously.

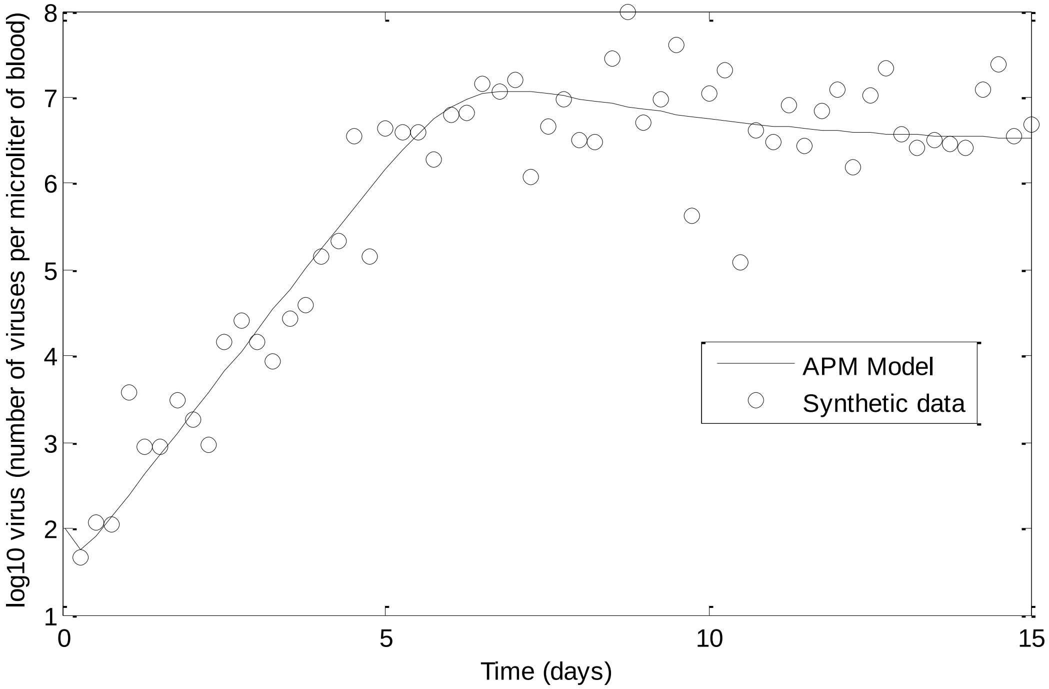

Figure 1: Shows the fit of APM model with estimated parameters to the synthetic data

Before the capability of APM's parameter estimation could be applied to large-scale biological models it was necessary to create an automatic conversion from the Systems Biology Markup Language (SBML) to a version usable by APM. This not only eliminates the human error in the conversion process, but it also allows for the quick evaluation of many publicly available models.

This conversion tool was used to automatically convert a model that describes the ErbB signaling pathways [5]. This model shall be referred to as the ErbB model. It is a large model with 225 parameters, 504 variables and 1331 DAEs. In the article the authors estimated 75 of the initial conditions and rate constants out of the 229 identified by the sensitivity analysis. This was accomplished through simulated annealing which required 100 annealing runs and 24 hours on a 100-node cluster computer on average to obtain just one good fit. APM is currently able to simulate the model and properly shows the dynamics found in the literature. It is believed that values do not match the literature values due to the limitations found in the conversion tool as it does not properly handle piecewise functions. Even if the literature values are not replicated exactly, parameter estimation can still be preformed. This is the next step in the project. The objective function will minimize the absolute error between the model and two measurable concentrations. Instead of using simulated annealing to estimate the parameters, a multi-start optimization will be used. To accomplish this, the parameter values will be randomly varied from 2.5 log order from the prior value. Many of these runs will be conducted and the results will be compared to those found using simulated annealing. If it is found that the design space is too flat or if there are too many optimums other optimizing techniques will be considered such as simulated annealing or a genetic algorithm.

In addition to the HIV and ErbB model parameter estimation, a benchmark test of Nonlinear Programming (NLP) solvers was conducted on the test set of over 400 curated models in the BioModels database. The test compares the performance of APOPT, BPOPT, IPOPT, SNOPT, and MINOS on this class of problems that is characterized by challenges unique to computation biology. This benchmark study points to a number of characteristics of interior point and active set methods for solving large-scale and sparse systems of differential and algebraic equations.

10c10 system engineering approaches in biology medicine

References

1. Adiwijaya, B. S. Herrmann, E. Hare, B. Kieffer, T. Lin, C. Kwong, A. D. Garg, Randle, J. C. R. Sarrazin, C. Zeuzem, S. and Caron, P. R. (2010), “A Multi-Variant, viral dynamic model of genotype 1 HCV to assess the in vivo evolution of protease-inhibitor resistant variants,” *PLoS Comput. Biol.*, 6(4):e1000745.
2. Biegler, L. T. (2007), “An overview of simultaneous strategies for dynamic optimization,” *Chemical Engineering and Processing: Process Intensification*, 46(11) pp. 1043 – 1053.
3. Nowark, M. and May, R. (2000), Virus Dynamics Mathematical Principles of Immunology and Virology. Oxford, New York: Oxford University Press.
4. Perelson, A. S. Kirschner, D. E. De Boer, R. (1993), “Dynamics of HIV infection of CD4 + T cells,” *Math Biosci,* March, pp. 81-125.
5. Chen, William W. Schoeberl, Birgit Jasper, Paul J. Niepel, Ulrik B. Lauffenburger, Douglas A. and Sorger, Peter K. (2009), “Input-output behavior of ErbB signaling pathways as revealed by a mass action model trainded against dynamic data,” *Molecular Systems Biology,* 5, pp. 239.