Synthesis of identifiable chemical reaction networks

Daniel Georgiev

***Short Abstract*—Identification of predictive system models for chemical reaction networks is considered during the design and experimental stages of synthesis. General analytical tools for testing the identifiability of a system and designing its identification experiments were derived. The predictive quality of the models is ensured by using an invalidation based identification approach. The tools will be applied to study genetic and epigenetic changes in populations.**

***Keywords*—model discrimination, model invalidation, model equivalency, model identification**

I. Introduction

Recent advancements in analytical technology are alone not enough to sufficiently reduce the time, cost, and feasibility of system identification. Reverse transcription qPCR technology has produced some promising results in predictive model identification [1, 2]. However, these experimental tools are expensive, require significant setup, and, most importantly, only work with cell populations. Identification of individual cells must primarily rely on fluorescence signals from reporter genes, where the number of simultaneous measurements is limited by the number of available proteins with different excitation or emission wavelengths. To deal with the feasibility and measurement constraints, advanced identification methods applicable in all parts of the synthesis process are needed.

General analytical tools for both system and experimental design are developed [3, 4]. The purpose of these tools is to produce systems for which predictive models can be identified. Model identification is ensured by minimizing the set of corresponding equivalent models. Prediction capability of resulting models is ensured by basing all identification methods on invalidation instead of parameter matching.

II. Methods

Considered systems are entirely described by a collection of controlled variables u (e.g., metabolite concentrations) and unknown variables w (e.g., reaction rates) so that the measured variables y (e.g., reporter protein concentrations) are defined by a model $y = M(u, w)$. Model identification is performed by invalidating as many possible values of w as possible. Systems are tested through a series of discrimination experiments and those values of w that contradict the experimental data are discarded. All values of w that make the same predictions for all values of u are said to be equivalent.

This work was supported in part by the grant WAMS-MPO FR-TT1/462.

D. Georgiev is with the Department of Cybernetics in the College of Applied Sciences, University of West Bohemia, Pilsen, 30000, Czech Republic. georgiev@kky.zcu.cz

A. Synthesis Procedure

A procedure for building systems with identifiable predictive models is proposed. The synthesis procedure begins by creating an identifiable design that has no equivalent values of w. Proposed designs are checked for identifiability by solving the Model Equivalence Problem. The synthesis procedure continues with a series of discriminating experiments whose inputs are computed by solving the Model Discrimination Problem [3, 4]. The synthesis procedure concludes by invalidating all values of w that fail to reproduce the experimental data [5]. This is done by solving the Model Invalidation Problem. All of the above steps are formulated in a scalable way and cast as convex optimization problems using either relaxation or approximation methods.

B. Technical Details

The mapping M, referred to above, is given implicitly by solving a system of polynomial differential equations representing the biochemical reaction network. Identifiability for such a system implies the entire state at any given time can be estimated from the initial conditions. The underlying assumptions of the experimental and design tools are consistent with available measurement technology. The entire initial state of the system (e.g., the molecular concentrations of important substrates) is assumed to be measurable (e.g., using RT qPCR or mass spectrometry). The subsequent measurements are assumed to be limited so that, for instance, individual cells can be identified using reporter gene methods.

III. Future work

The above tools will be fully integrated to study individual cell differences at the new Cell Cybernetics Lab at the University of West Bohemia. For a given functional module of interest, a suitable reporting network will be designed and the state fully measured under standard conditions. Genetic and epigenetic changes (as well as mechanisms of effecting these changes) will then be studied through ongoing identification of the corresponding models.

References

[1] A. Julius, M. Zavlanos, S. Boyd, and G. Pappas, "Genetic network identification using convex programming," *IET Systems Biology*, vol. 3, 2009.

[2] T. S. Gardner, D. di Bernardo, D. Lorenz, and J. J. Collins, "Inferring genetic networks and identifying compound mode of action via expression profiling," *Science*, vol. 301, 2003.

[3] D. Georgiev and E. Klavins, "Model discrimination of polynomial systems via stochastic inputs," in *Proc. IEEE Conf. on Decision and Control*, 2008.

[4] D. Georgiev, M. Fazel, and E. Klavins, "Model discrimination of chemical reaction networks by linearization," in *Proc. American Control Conf.*, 2010.

[5] S. Prajna, "Barrier certificates for nonlinear model validation," *Automatica*, vol. 42, no. 1, 2006.