Topics for Bachelor Thesis

The focus of the groups of Dr. Clemens Kreutz and Prof. Jens Timmer is

(1) application of theoretical methods from physics and mathematical statistics for analyzing experimental data generated by collaborators in biomedical research,
(2) establishing mathematical models of dynamic systems, especially in the field of cellular signal transduction and gene regulation,
(3) understanding of biochemical mechanisms in living cells and of dysfunction in diseases,
(4) improvement of existing methods and establishment new approaches for statistical analysis and mathematical modelling,
(5) assessing the performance of state of the art methodology.

The following topics for Bachelor thesis are offered

1) Improved parameter estimation by duplicating the model equations

Background

Understanding of complex biochemical networks as they occur in living cells requires the combination of experimental work with mathematical modelling. Ordinary differential equation models (ODEs) can be used as mathematical representation for analyzing known biochemical interaction networks.

A major goal is the calibration of such models, i.e. to estimate the parameters like initial concentrations or rate constants based on experimental data. For parameter estimation, numerical optimization algorithm e.g. for minimizing $\chi^2(\theta)$ are applied. However, parameter optimization of ODE models is challenged because of the nonlinear dependency on parameters and by the high-dimension of the parameter space. Therefore, application examples occur where numerical optimization fails.

One aspect is that numerical integration of ODEs involves algorithms for adaptive step-size control. This means that if ODEs are integrated twice, e.g. for different sets of parameters, the step size control differs. This results in incomparable accuracy of $\chi^2(\theta)$ or other objective functions.

State of the art

- In our group, a comprehensive implementation of the parameter estimation methodology is available [1,2]. This implementation has been awarded twice within scientific benchmark challenges [3].
- For the calculation of derivatives using difference quotient $[\ \chi^2(\theta+\Delta\theta) - \chi^2(\theta)\ ] / \ \Delta\theta$ is not applicable because of the impact of the step size control. Alternative approaches have been established to circumvent this issue [4] for small $\Delta\theta$.
- However, there are no available methods to circumvent this issue within iterative parameter optimization. Here, $\Delta\theta$ is typically large and methods utilizing the Taylor approximation are inaccurate.

Bachelor thesis topic

In this Bachelor thesis, a trick is utilized to circumvent the numerical error. The idea behind this new approach is to duplicate the state equations, using parameter sets θ_{i-1} and θ_i for each set of state equations, and then jointly integrate the ODEs for both parameter sets. Then, the objective function χ^2 can be calculated for both parameter sets by a common step-size control.

This new method for getting more reliable comparisons of the objective function $\chi^2(\theta_{i-1}) - \chi^2(\theta_i)$ is implemented and compared to the standard approach where both parameter sets are integrated independently.

2) Comparison of two methods for assessing prediction uncertainty

The major tasks in establishing dynamic models of living systems are model discrimination, parameter estimation and model predictions. For all these tasks, it is essential to assess and control uncertainties. In contrast to a regression setting, this is a nontrivial task for ODE models because of nonlinearity and the absence of analytical solutions.

For model predictions, statistically valid confidence intervals can be derived using the prediction profile likelihood [5,6]. Another method which directly translates parameter uncertainties to predictions requires less demanding calculation but underestimates the size of the confidence intervals [7].

The task of this Bachelor thesis is compare both approaches using the data2Dynamics modelling framework [1,2]. For obtaining generally valid results, several models as well as many data settings will be analyzed.

Programming will be performed in MATLAB. The models as well as the two approaches for assessing prediction uncertainty are already implemented.

Primary supervisor and contact person: Dr. Clemens Kreutz, ckreutz@fdm.uni-freiburg.de

3) Identification of context-specific parameters using group-LASSO

Many projects within Systems Biology address the detection of drivers of dysregulation in cellular signaling that lead to disease or drug resistance. If measurements in a healthy control and one or more mutated cell-lines are available, the L1 approach can be used [8]. Starting from distinct parameters for each sample, their deviation is penalized to achieve a sparse solution of model parameters that would lead to significantly worse description of measurements if they would not be cell-line specific.

In this Bachelor thesis, the classic L1 approach will be compared to group-LASSO, where parameters are penalized in groups to end up with a sparse solution of groups [9]. These combine e.g. parameters for a specific signaling pathway or group parameters of the whole set of cell-lines. The task will be to implement the group-LASSO approach into our existing modeling framework Data2Dynamics, and monitor and interpret the results of both approaches for a toy model with known parameter deviations.

Programming will be performed in MATLAB. A toy model for which data can be simulated for several distinct cell lines will be available, and the approach can be extended to a published model of Erythropoietin receptor signaling.

Primary supervisor and contact person: Helge Hass, helge.hass@fdm.uni-freiburg.de

References

[1] A. Raue, et al. Data2Dynamics: a modeling environment tailored to parameter estimation in dynamical systems. *Bioinformatics* **31**, 2015, 3558-3560

[2] Data2Dynamics Software Website www.data2dynamics.org

[3] P. Meyer, et al. Network topology and parameter estimation: from experimental design methods to gene regulatory network kinetics using a community based approach. *BMC Systems Biology*, 8(1), 1, 2014.

[4] A. Raue, et al. Lessons Learned from Quantitative Dynamical Modeling in Systems Biology. *PLOS One* 8(9), e74335, 2013.

[5] C. Kreutz, A. Raue, J. Timmer Likelihood based observability analysis and confidence intervals for predictions of dynamic models. BMC Systems Biology 2012; 6; doi:10.1186/1752-0509-6-120

[6] Kreutz C, Raue A, Kaschek D, and Timmer J. Profile Likelihood in Systems Biology. *FEBS Journal*, 2013

[7] Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, Timmer J. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*. 2009 25(15):1923-9.

[8] Steiert, B., Timmer, J. and Kreutz, C. L1 regularization facilitates detection of cell type-specific parameters in dynamical systems. *Bioinformatics.* 2009 32(17):i718-i726.

[9] Yuan, M. and Lin, Y. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B,* 2006 68(1):49-67.