European Conference on Mathematical and Theoretical Biology 2014 - Gothenburg

Minisymposium: Numerical methods for high-dimensional problems in biology
Date: Monday, 16th of June
Time: 16:00 - 19:00

Organizers:

Shuohao Liao
Wolfson Centre for Mathematical Biology
University of Oxford
E-mail: liao@maths.ox.ac.uk

Tomas Vejchodsky
Wolfson Centre for Mathematical Biology
University of Oxford
E-mail: vejchodsky@maths.ox.ac.uk

Program:

16:00 - 16:25	**Shuohao Liao** (University of Oxford) *Parameter estimation, robustness, sensitivity and bifurcation of stochastic gene regulation network: Tensor-structured parametric analysis*
16:25 – 16:50	**Martin Lindén** (Stockholm University) *Benchmarking analysis methods for live cell single particle tracking using simulated microscopy*
16:50 — 17:15	**Vikram Sunkara** (The University of Adelaide) *Tackling the construction of probability distributions arising in Bio-Chemical networks*
17:15 — 17:25	Break
17:25 – 17:50	**Ivan Oseledets** (Skolkovo Institute of Science and Technology) *Effective tensor-based methods for the solution of high-dimensional non-stationary problems with the application to the chemical master equation*
17:50 – 18:15	**Vladimir Kazeev** (ETH Zürich) *Tensor-structured treatment of the chemical master equation*
18:15 – 18:40	**Sergey Dolgov** (Max Planck Institute for Mathematics) *Solution of the chemical master equation by the separation of variables and alternating optimization methods*

Shuohao Liao
Mathematical Institute, University of Oxford
e-mail: liao@maths.ox.ac.uk

Parameter Estimation, Robustness, Sensitivity and Bifurcation of Stochastic Gene Regulation Networks: Tensor-Structured Parametric Analysis

Genetically identical cells are experimentally proven to exhibit considerable cell-to-cell variations in mRNA and protein levels. Such variability of phenotypes is driven by both the intrinsic noise, characterized by gene network structure, and the extrinsic noise, which acts globally on a single cell but varies from one to another. For intrinsic noise, mathematical and computational biologists have been developing and improving simulation techniques to investigate stochastic networks under strict parameters, and discovered interesting behaviours, like noise-induced multi-stability and oscillations. However, global effect of extrinsic noise raises a few more questions: how does the internal stochastic behavior change with respect to some perturbations of biophysical parameters? How may we define, quantify and predict these changes under stochastic context? Do these changes in a single cell apply similarly to a collection of cells? If not, how to identify the role of intrinsic and extrinsic noise in generating the variations in experimental observations?

Under deterministic context, all these questions fall into the subject of bifurcation analysis, a theory describing the dependence of the steady state on continuous changes in parameters. But stochastic modeling, by contrast, is still lagging behind concerning appropriate parametric methods. The commonly used Monte Carlo formulation becomes inefficient in parametric analysis, as it requires separate simulations for different parameter combinations. On the other hand, equation-based models, like chemical master equation and its Fokker-Planck approximation, suffers from the so-called 'bless of dimensionality'. In my talk, I will introduce the newly-developed tensor-structured data format, and demonstrate its prospective in: a) parameter estimation of stochastic networks, b) robustness analysis of intrinsic stochastic system against extrinsic noise, and c) stochastic sensitivity and bifurcation analysis of complex genetic regulatory networks.

Martin Lindén
Department of Biochemistry and Biophysics, Stockholm University
e-mail: martin.linden@dbb.su.se

Benchmarking analysis methods for live cell single particle tracking using simulated microscopy

Single particle tracking in live cells is emerging as a quantitative and non-invasive tool for systems biology. A particularly promising direction is the possibility to monitor chemical reactions by exploiting the fact that small molecules diffuse slower than large ones, so that a fluorescently tagged ligand will change diffusion constant as it associates and dissociates from its binding partners. We have developed an analysis suite, vbspt.sourceforce.net (1), that uses a Bayesian treatment of hidden Markov models to learn the number of diffusive states and their interconversion rates from position trajectories of diffusing particles with random jumps in diffusion constant.

However, limitations in optical microscopy and fluorescence labeling influences the resolution of the method, and live cells are more complex than the model assumptions used by vbSPT. To learn more about how such effects influences our analysis and limits the kind of mechanisms that can be resolved, we are currently developing computational tools to simulate live cell microscopy, using a combination of reaction-diffusion kinetics and photophysics models. These simulations include more physical realism than the models on which vbSPT are based, which makes them suitable for benchmarking, and for optimizing experimental conditions.

In this talk, I will give a brief introduction to single particle tracking, describe the mathematics and models underlying the vbSPT analysis suite, and show some preliminary results and lessons from our analysis of simulated data.

REFERENCES

[1] Persson F*, Lindén M*, Unoson C, Elf J (2013). Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nat Methods*, **10**: 265–269. * Equal contributions.

Vikram Sunkara
School of Mathematical Science, The University of Adelaide
e-mail: Vikram. Sunkara@ adelaide.edu.au

Tackling the construction of probability distributions arising in Bio-Chemical networks

Modern systems biology is working towards describing the interactions of processes in biology by networks, we refer to these as Bio-Chemical Reaction Networks (Biological Networks for short). Through advances in fluorescence techniques and sequencing technologies, biologists are able to scope deeper and describe critical biological processes as paths on a complex biological network. We are interested in a particular class of biological networks, that is, biological networks with paths which behave as jump Markov processes. The probability of observing any particular state the network is in at a particular point in time is given by solving the Chemical Master Equation (CME). For these particular class of networks, a variety of Monte Carlo based simulation methods have been proposed. However computing a probability distribution over all possible features of the network is computationally difficult. It has been shown that the computational complexity grows exponentially in the number of species in the system. In this talk we discuss the major issues that arise in computing probability distributions for large biological networks. We introduce and demonstrate some modern techniques inspired by adaptive domain selection (Optimal Finite State Projection method) and dimension reduction (Hybrid Models) to help tackle our issues. We apply these techniques on real life examples to demonstrate their contribution.

I. V. Oseledets
Skolkovo Institute of Science and Technology
e-mail: i.oseledets@ skolkovotech.ru

Denis Kolesnikov
Skolkovo Institute of Science and Technology

Effective tensor-based methods for the solution of high-dimensional non-stationary problems with the application to the chemical master equation

Solving non-stationary problems is hard, even if they are linear. There exist numerous techniques for the efficient solution of such kind of problems. The task becomes much harder, when the problem in question is high-dimensional. Chemical master equation for stochastic modelling of chemical kinetics is of the most vivid examples of such kind. Using Finite State Projection (FSP) the problem is reduced to a linear non-stationary system of equations, that has very high dimensionality, i.e. the solution can be represented as a multidimensional array (tensor) with large mode sizes. Application of tensor decompositions to such kind of problems is very promising, and that has been shown by recent work of Dolgov and Khoromskij, and also Kazeev, Schwab, Khammash and Nip. Tensor techniques allow to break the curse of dimensionality and reduce the complexity. Still, there is a lot of to on the algorithmic side. In this talk I will present several new approaches for solving high-dimensional non-stationary problem by means of tensor techniques. The new approach is related to the so-called global time-stepping scheme and we will show that it boils down to a simple Krylov-type approximation scheme. Finally we will show how the Tensor Train Toolbox (TT-Toolbox, http://github.com/oseledets/TT-Toolbox and Python version http://github.com/oseledets/ttpy) can be applied to routinely solve high-dimensional chemical master equations.

Vladimir Kazeev
Seminar for Applied Mathematics, ETH Zürich
e-mail: Vladimir.kazeev@sam.math.ethz.ch

Tensor-structured treatment of the chemical master equation

The chemical master equation (CME) is a cornerstone of the stochastic analysis and simulation of models of biochemical reaction networks. Nevertheless, the direct numerical treatment of the CME has remained elusive due to the curse of dimensionality, i.e. the exponential growth of the storage cost and computational complexity with respect to the dimensionality of the problem.

We consider a novel approach based on the use of the quantized tensor train (QTT) decomposition for the low-parametric representation of tensors. As a measure of the storage cost and computational complexity, the size of the state space considered is replaced with the QTT ranks of the operator and solution. Those are intrinsic characteristics, which govern the number of parameters in the corresponding QTT representations and depend on the desired accuracy.

First, we discuss how the CME can be recast in QTT format efficiently and present bounds on the QTT ranks of the operator depending on the desired accuracy. We consider the classical mass-action and single-enzyme Michaelis–Menten kinetics, which correspond to two widely used classes of propensity functions. Arbitrary combinations of these two types of kinetics are also covered by our results.

Second, for evolution problems, we consider the hp-discontinuous Galerkin discretization in time to reduce the CME evolution problem to a sequence of QTT-structured system of linear equations, which are solved in the course of time marching. For solving the linear systems, we use an algorithm of the TT Toolbox inspired by an approach from quantum chemistry. We demonstrate the efficiency of the approach in three different examples from systems biology: independent birth-death process, an enzymatic futile cycle, and a stochastic switch model. The numerical results demonstrate dramatic speedups and storage savings over standard approaches with elementwise data representation.

Third, for the types of kinetics mentioned above, we consider systems of reacting species with a weakly-reversible reaction network of zero deficiency in the sense of Feinberg. We bound the QTT ranks of stationary distributions of such systems depending on the desired accuracy. We show that the complexity of the approximation scales linearly with respect to the number of species and logarithmically in the maximum copy numbers and desired accuracy. This rigorous theoretical result partly justifies the experimental observations of the efficiency of the QTT representation.

The talk is based on joint works with Christoph Schwab and Mustafa Khammash (ETH Zurich) and Michael Nip (UC Santa Barbara).

Sergey Dolgov
Max Planck Institute for Mathematics
e-mail: sergey.v.dolgov@gmail.com

Solution of the chemical master equation by the separation of variables and alternating optimization methods

The chemical master equation (CME) serves as an accurate model of stochastic fluctuations in chemical kinetics, which are ubiquitous in systems with small amounts of reacting species, such as cells, viruses and others. However, the solution of the CME is the probability distribution function, which is defined on all states of the system, and may require enormously large storage, growing exponentially with the number of species.

In this talk, we employ algebraic methods for separation of variables to reduce the complexity, namely, the so-called Matrix Product States (MPS)/Tensor Train format. This format is an elegant generalization of the low-rank matrix factorization to high dimensions, and we demonstrate that the CME operator may be efficiently represented and the solution sought directly in the MPS form, avoiding exponential cost dependence on the number of species. Besides, the spectral time discretization may be simply incorporated into the scheme, by considering time as an additional dimension. The resulting large linear system is solved in the MPS format by the recently developed Alternating Minimal Energy (AMEn) algorithm. The latter supports the alternating optimization technique (so-called DMRG in quantum physics) by the classical residual direction, similarly to the steepest descent method. This scheme avoids traps of spurious optima and manifests rapid convergence even for non-symmetric systems, arising from the CME.

We present a couple of convincing examples, such as the cascade gene regulatory network and the phage-lambda model.