Evaluation of Parallel Tempering to Accelerate Bayesian Parameter Estimation in Systems Biology

Sanjana Gupta[1], Liam Hainsworth[1], Justin S. Hogg[1], Robin E.C. Lee[1], and James R. Faeder[1]

***Short Abstract* — Bayesian parameter estimation (BPE) is popular in systems biology, where often a large number of correlated model parameters have to be estimated from limited experimental data. Commonly-used Markov chain Monte Carlo (MCMC) methods for BPE often suffer from slow convergence. Here[1] we evaluate the performance of parallel tempering (PT), a physics-based MCMC method designed to accelerate convergence by swapping between multiple MCMC chains run in parallel at different temperatures.**

I. Introduction

Computational models are used to describe biological systems and make testable predictions[2]. Parameter estimation is the calibration of a model to data by searching for parameterizations that minimize the discrepancy between the data and model output. MCMC is a Bayesian parameter estimation method commonly used in systems biology, but standard algorithms such as the Metropolis-Hastings (MH) suffer from slow convergence. Parallel tempering (PT) is a physics-based method that accelerates sampling of probability distributions by swapping between parallel MCMC chains run at different temperatures[3]. While PT has been commonly used in molecular dynamics simulations to accelerate sampling the conformational space of biomolecules[3], it has sparsely been used in systems biology[2].

In this work[1] we evaluate the performance of PT relative to MH on six biological models of increasing complexity. We include a comparison with Approximate Bayesian Computation – Sequential Monte Carlo (ABC-SMC), another common Bayesian parameter estimation method.

II. Methods

We performed all the MCMC fits using pTempEst, our MATLAB-based tool for parameter estimation using PT. The models were specified in the BioNetGen language (BNGL)[4], and exported as ODE models in MATLAB's MEX-file format that are called by pTempEst, which invokes the CVODE library for efficient integration of high dimensional models. We used the tool ABC-SysBio to perform fits using ABC-SMC[5].

In our analyses we fit ODE models to synthetic data generated using known parameters. For smaller models (3-6 parameters), both MH and PT found the global minimum and we compared the algorithms using convergence time and sampling efficiency. For more complex models (11-25 parameters) we did not always obtain parameter sets that fit the data. In this case we compared the algorithms using the likelihood of the best-fit parameters.

We compared fits from ABC-SMC and PT by allowing each algorithm a specified number of model integrations, fixing the total amount of computational resource used.

III. Results and Future work

For simple models with 3-6 parameters (Michaelis-Menten model, mRNA self-regulation[5], simple negative feedback loop), PT accelerated convergence and improved sampling over MH. For bigger models with 12-25 parameters (calcium signaling, negative feedback oscillator[6], growth factor signaling[7]) PT more consistently found the global optimum, while MH frequently got trapped in local optima. Finally, we found that for a fixed number of integrations, PT outperformed ABC-SMC for parameter sampling on a relatively simple ODE model of mRNA self-regulation.

A current limitation of PT is that it is only moderately parallel across a small number of chains and does not fully leverage the large number of nodes on typical modern day clusters. We are currently investigating ways to increase the parallelizability of the algorithm, for example by running multiple chains at each temperature level. Another area of improvement that we are pursuing is in the proposal function, which currently does not leverage known parameter correlations to improve sampling efficiency. We will investigate whether previous work in this area[8,9] would benefit from a parallel tempering approach.

[1]Department of Computational and Systems Biology, University of Pittsburgh, School of Medicine

Acknowledgement: This work was funded by NIH grant R35-GM119462 to RECL, and by JRF via the NIGMS-funded (P41-GM103712) National Center for Multiscale Modeling of Biological Systems (MMBioS)

References

1. Gupta, S., Hogg, J. S., Lee, R. E. C. & Faeder, J. R. Evaluation of Parallel Tempering to Accelerate Bayesian Parameter Estimation in Systems Biology. *26th Euromicro Int. Conf. Parallel, Distrib. Network-based Process.* (2018).
2. Malkin, A. D. *et al.* A Neutrophil Phenotype Model for Extracorporeal Treatment of Sepsis. *PLoS Comput. Biol.* **11,** 1–30 (2015).
3. Earl, D. J. & Deem, M. W. Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.* **7,** 3910 (2005).
4. Harris, L. A. *et al.* BioNetGen 2.2: Advances in rule-based modeling. *Bioinformatics* **32,** 3366–3368 (2016).
5. Liepe, J. *et al.* A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. *Nat. Protoc.* **9,** 439–456 (2014).
6. Tyson, J. J., Chen, K. C. & Novak, B. Sniffers , buzzers , toggles and blinkers : dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* **15,** 221–231 (2003).
7. Barua, D., Faeder, J. R. & Haugh, J. M. Structure-based kinetic models of modular signaling protein function: focus on Shp2. *Biophys. J.* **92,** 2290–300 (2007).
8. Eydgahi, H. *et al.* Properties of cell death models calibrated and compared using Bayesian approaches. *Mol. Syst. Biol.* **9,** 644–644 (2014).
9. Zhang, L. A. *et al.* APT-MCMC, a C++/Python implementation of Markov Chain Monte Carlo for parameter identification. *Comput. Chem. Eng.* **110,** 1–12 (2018).