

Open PhD position: Scientific computing in Machine learning

Stefan Engblom
stefane@it.uu.se

Application deadline August 31, 2016.
--

THE PROJECT

The overall target of the project is to devise, develop, and study machine learning algorithms and implementations. The emphasis for the project is on the *scientific computing* viewpoint of machine learning algorithms and methods; specifically set-ups of compute-intensive character requiring advanced tools from high-performance computing. Concretely we aim to solve problems in one or several of the suggested research/application areas described below.

We offer an interdisciplinary project which to a high degree can be formed by the selected candidate. Suitable targets of research within the project include high performance algorithms on multi- and manycore computers, but also design and theoretical studies of algorithms for compute intensive machine learning or data assimilation applications.

Suggested areas and directions include:

Epidemics: Estimate infectious state given measured data; evaluate intervention strategies; assimilate measurements; actively suggest and monitor counter measures. *Further reading:* [1].

Chemical kinetics: Find rate parameters in chemical networks given observations; estimate uncertainties and evaluate goodness-of-fit; determine optimal reaction networks. *Further reading:* [2, 3].

Nested SMC: This line of research involves a recently proposed algorithm with potential for strong scaling thanks to its distributable nature. In particular, it may be suitable to spatio-temporal problems (including also, for example, the suggested areas above). This project draws on recent output from Thomas Schön's group. *Further reading:* [4].

THE PARALLEL ALGORITHMS GROUP @ UPMARC

The work will be done within the Division of Scientific Computing and the Parallel Algorithms group at the UPMARC Linnaeus center of Excellence. UPMARC is a long-term interdisciplinary research program in parallel computer systems. The work within the PA group is concerned with high performance scientific computing on multicore computing platforms. Methods, implementations, and applications in partial differential equations

(PDEs), numerical linear algebra, N-body formulations, event-based simulations, and large-scale inverse problems have been targeted previously.

With today's fast computers and efficient implementations, the increased collection of data and cheap high-bandwidth memories, the possibilities for 'backward mode' computations have increased tremendously. That is, given data plus model, infer the parameters of the model and make intelligent use of this knowledge by feeding it back into the running system. Such set-ups are inherently challenging as they rely on being able to evolve the underlying model in 'forward mode' many times.

Most learning algorithms work in terms of iterations that sweep over data, and often over *all* of data in each sweep. The complexity of the actual computations – “the quality of the flops” – depend on the mathematical model to be inferred over, and how this model is handled numerically. *For sufficiently performance demanding applications, the correct use of cores may have a very large impact on performance, second only to selecting the best numerical method.*

APPLICATION

We wish to come in contact with students with a background in Scientific Computing at broad; including for example Applied Mathematics, Computer Science, or in one of the application areas pointed out above. Prior knowledge in Computational Statistics, Machine Learning, Data Assimilation, and High Performance Programming are considered meriful. *In your application letter, please point out what parts of your background fit with the above description.*

The research will be centered on formulations and methods in one or in a combination of the suggested areas as outlined above. *In your application letter, explain which area(s) interest you the most, why, and what your background would bring to the project.*

If you have a strong research case yourself, please feel invited to present a suitable research idea completely of your own, so long as it matches the general description of the call.

<p>Interested candidates with a suitable background are more than welcome to contact me for further information.</p>

<p>Application deadline August 31, 2016.</p>

REFERENCES

- [1] P. Bauer, S. Engblom, and S. Widgren. Fast event-based epidemiological simulations on national scales, 2016. Accepted for publication in *Int. J. High Perf. Comput. Appl.* [doi:10.1177/1094342016635723](https://doi.org/10.1177/1094342016635723).

- [2] A. Golightly and D. J. Wilkinson. Bayesian parameter inference for stochastic biochemical network models using particle Markov chain Monte Carlo. *Interface Focus*, 1: 807–820, 2011. doi:10.1098/rsfs.2011.0047.
- [3] G. Lillacci and M. Khammash. The signal within the noise: efficient inference of stochastic gene regulation models using fluorescence histograms and stochastic simulations. *Bioinformatics*, 29(18):2311–2319, 2013. doi:10.1093/bioinformatics/btt380.
- [4] C. A. Naesseth, F. Lindsten, and T. B. Schön. Nested sequential Monte Carlo methods. In *Proceedings of the 32nd International Conference on Machine Learning*, 2015.

(Stefan Engblom) DIVISION OF SCIENTIFIC COMPUTING, DEPARTMENT OF INFORMATION TECHNOLOGY, UPPSALA UNIVERSITY, SWEDEN.

E-mail address: stefane@it.uu.se, <http://user.it.uu.se/~stefane>