PhD project

Asynchronous MCMC algorithms for fast Bayesian inference

2022 – 2025

Pierre-Antoine THOUVENIN, Pierre CHAINAIS
École Centrale de Lille, UMR 9189 CRIStAL

January 18, 2022

1 Project overview

This project is aimed at accelerating MCMC algorithms for fast Bayesian inference in large scale problems. Applications in astronomy (*e.g.*, hyperspectral imaging) or in remote sensing (*e.g.*, multimodal multi-temporal source separation) could be considered. The project is part of the *ANR Chaire IA SHERLOCK* led by Pierre Chainais (co-funded by ISITE, Centrale Lille Institut and Région Haut-de-France).

Many signal and image processing applications, ranging from astronomy (Abdulaziz et al. 2019; Cai et al. 2018) to remote sensing (Borsoi et al. 2021; Ghamisi et al. 2019), involve large datasets. In absence of ground truth, fast parameter inference under controlled uncertainty is critical to guarantee the quality of the resulting predictions.

Asynchronous (parallel or distributed) optimization algorithms have recently regained interest due to their potential of acceleration, in comparison with their synchronous counterparts (Hannah et al. 2017). However, optimization algorithms only bring a point estimate, such as the *maximum a posteriori* (MAP) estimator. Markov-chain Monte Carlo (MCMC) methods bring a richer information by sampling the posterior distribution of the model. MCMC methods are known to induce larger computational costs compared to optimization algorithms. Nevertheless, recent works at the interface between deterministic and stochastic optimization have introduced efficient samplers to address larger datasets (Durmus et al. 2018; Vono et al. 2020). With the exception of (Simsekli et al. 2018; Terenin et al. 2020), asynchronous MCMC algorithms largely remain to be investigated.

This PhD project is aimed at studying the potential of asynchronous MCMC algorithms for fast Bayesian inference in high dimensional problems.

Keywords. Bayesian inference, MCMC algorithms, asynchronous algorithms.

2 Detailed description

An inverse problem consists in estimating a collection of parameters involved in a physical model from incomplete, degraded and noisy observations. In many signal and image processing applications, especially in astronomy (Abdulaziz et al. 2019; Cai et al. 2018) and remote sensing (Ghamisi et al. 2019), no ground truth is available. This is also often the case for predictors based on machine learning. Therefore, it is critical to provide estimates with quantified uncertainty.

Beyond the point estimate offered by an optimization algorithm, Markov-chain Monte Carlo (MCMC) algorithms give access to credibility intervals to quantify the estimation uncertainty. This usually comes at a higher computational cost. However, recent works at the interface between the deterministic and stochastic optimization literature (Durmus et al. 2018; Vono et al. 2020) paved the way to fast distributed MCMC algorithms.

When dealing with big datasets, Bayesian inference algorithms may benefit from divide to conquer strategies. Distributed inference – involving multiple computing units referred to as *workers* – comes as a solution to face both data storage limitations and reduce the estimation runtime. Adding more workers reduces the costs on a per worker basis, and allows the algorithm to scale to large datasets.

The workers involved in a distributed inference task need to exchange information to update parameters. A tradeoff between computation and communication loads needs to be found. To reduce communication costs, the amount of data and the frequency of communications need to be reduced as much as possible. One solution consists in relaxing synchronization requirements between the workers. This allows the performance of the workers to be better exploited, while offering more robustness against load balancing issues. Asynchronous optimization algorithms have been more specifically analyzed (Cannelli et al. 2019; Mishchenko et al. 2020), calling for specific modifications to ensure their convergence, in comparison with equivalent synchronous procedures. To the best of our knowledge, only (Simsekli et al. 2018; Terenin et al. 2020) have analyzed asynchronous MCMC algorithms. This topic largely remains to be investigated.

This PhD project is aimed at studying the potential of asynchronous MCMC algorithms for fast Bayesian inference in high dimensional problems. Performance assessment could be conducted on inverse problems in astronomy (Abdulaziz et al. 2019; Cai et al. 2018) or remote sensing (multimodal multi-temporal source separation under spectral variability (Borsoi et al. 2021; Ghamisi et al. 2019)). At the interface between deterministic and stochastic optimization, this project is expected to overcome the limitations of existing methods to deal with otherwise oversized problems. Note that the proposed methods will also be applied to Bayesian machine learning problems.

The project will directly benefit from interactions with physicists and astrophysicists from the ORION-B consortium (giant molecular clouds in astrophysics) and the ANR RICOCHET (gravitational waves) projects. Applications to real data will be possible.

3 Scientific context

The project is part of the *ANR Chaire IA SHERLOCK* (*Fast inference with controlled uncertainty: application to astrophysical observations*) led by Pierre Chainais (co-funded by Agence Nationale de la Recherche (ANR), ISITE, Centrale Lille Institut and Région Haut-de-France). The succesful candidate will be jointly supervised by Pierre Chainais and Pierre-Antoine Thouvenin in the CRIStAL lab (UMR 9189), Lille, France.

This work will be conducted in the continuity of an ongoing collaboration initiated by Pierre-Antoine Thouvenin with Audrey Repetti – research associate at Heriot-Watt University – and Pierre Chainais. There will be opportunities for short or longer stays at Heriot-Watt University.

The successful candidate will have access to the medium scale computing center from the Universtiy of Lille, and the national flagship Jean Zay supercomputer.

4 Objectives and planning

The main objectives of the thesis are summarized below, with an indicative planning.

- September 2022 - February 2023
  - review of state-of-the-art algorithms from the asynchronous optimization literature
  - review of state-of-the-art analysis Langevin MCMC algorithms;
  - review on parallel and distributed programming techniques (MPI, OpenMP, GPU programming, hybrid programming, parallel data formats such as HDF5);
  - review of useful parallel libraries (Python).
- March 2023 - August 2023:
  - design and theoretical study of a first distributed asynchronous sampling algorithm;
  - application to a representative imaging inverse problem (to be discussed with the PhD advisors);
  - core library development (in Python);
  - performance assessment on a large computer grid.
- September 2023 - February 2024:
  - writing of a first journal paper;
  - writing of a first conference paper;

- (optional) extend the implementation of the algorithm to accommodate multiple parallelization modalities (both CPU and GPU parallelization);
  - (optional) port the functionalities developed thus far to C++.

- March 2024 - August 2024:
  - publication of the first version of the library implementing the algorithm from the 1st publication. The code produced should be sufficiently generic to ensure it can be consistently reused for the following publications;
  - investigate an application to a 2nd inverse problem (to be discussed with the advisors);
  - code development associated with the 2nd application;
  - writing of a second journal paper;
  - writing of a second conference paper.

- September 2024 - February 2025:
  - finalize the code associated with the 2nd journal article;
  - submit 2nd journal article;
  - ensure the complete library is fully documented, available online by the acceptance of the 2nd journal article.

References

Abdulaziz, Abdullah et al. (Oct. 11, 2019). "Wideband Super-Resolution Imaging in Radio Interferometry via Low Rankness and Joint Average Sparsity Models (HyperSARA)". In: *Monthly Notices of the Royal Astronomical Society* 489.1, pp. 1230–1248. arXiv: 1806.04596.

Borsoi, Ricardo et al. (2021). "Spectral Variability in Hyperspectral Data Unmixing: A Comprehensive Review". In: *IEEE Geoscience and Remote Sensing Magazine*, pp. 2–49. arXiv: 2001.07307.

Cai, Xiaohao et al. (Nov. 1, 2018). "Uncertainty Quantification for Radio Interferometric Imaging – I. Proximal MCMC Methods". In: *Mon Not R Astron Soc* 480.3, pp. 4154–4169.

Cannelli, Loris et al. (June 15, 2019). "Asynchronous Parallel Algorithms for Nonconvex Optimization". In: *Math. Program.*

Durmus, Alain et al. (Jan. 2018). "Efficient Bayesian Computation by Proximal Markov Chain Monte Carlo: When Langevin Meets Moreau". In: *SIAM J. Imaging Sci.* 11.1, pp. 473–506.

Ghamisi, Pedram et al. (Mar. 2019). "Multisource and Multitemporal Data Fusion in Remote Sensing: A Comprehensive Review of the State of the Art". In: *IEEE Geoscience and Remote Sensing Magazine* 7.1, pp. 6–39.

Hannah, Robert et al. (Aug. 17, 2017). "More Iterations per Second, Same Quality – Why Asynchronous Algorithms May Drastically Outperform Traditional Ones". In: arXiv: 1708.05136 [cs, math, stat].

Mishchenko, Konstantin et al. (Jan. 1, 2020). "A Distributed Flexible Delay-Tolerant Proximal Gradient Algorithm". In: *SIAM J. Optim.* 30.1, pp. 933–959. arXiv: 1806.09429.

Simsekli, Umut et al. (July 3, 2018). "Asynchronous Stochastic Quasi-Newton MCMC for Non-Convex Optimization". In: *International Conference on Machine Learning*. International Conference on Machine Learning, pp. 4674–4683.

Terenin, Alexander et al. (Feb. 29, 2020). "Asynchronous Gibbs Sampling". In: arXiv: 1509.08999 [stat].

Vono, Maxime et al. (Sept. 28, 2020). "Asymptotically Exact Data Augmentation: Models, Properties, and Algorithms". In: *Journal of Computational and Graphical Statistics* 0.0, pp. 1–14. arXiv: 1902.05754.