ParCFD'2021
33^{nd} International Conference on Parallel Computational Fluid Dynamics
May-17-19 2021, Nice FRANCE

HIPPYLIB-MUQ: SCALABLE MARKOV CHAIN MONTE CARLO SAMPLING METHODS FOR LARGE-SCALE BAYESIAN INVERSE PROBLEMS GOVERNED BY PDES

Ki-Tae Kim[*], Umberto Villa[†], Matthew Parno[‡], Noemi Petra[*], Youssef Marzouk[§], and Omar Ghattas[¶]

[*]Department of Applied Mathematics
University of California, Merced
95343, Merced, CA, USA
e-mail: kkim107@ucmerced.edu, npetra@ucmerced.edu

[†]Department of Electrical & Systems Engineering
Washington University in St. Louis
63130, St. Louis, MO, USA
e-mail: uvilla@wustl.edu

[‡]Cold Regions Research & Engineering Laboratory
The United States Army Corps of Engineers
e-mail: matthew.d.parno@erdc.dren.mil

[§]Department of Aeronautics and Astronautics
Massachusetts Institute of Technology
02139, Cambridge, MA, USA
e-mail: ymarz@mit.edu

[¶]Oden Institute for Computational Engineering and Sciences
Department of Mechanical Engineering, and Department of Geological Sciences
The University of Texas at Austin
78712, Austin, TX, USA
e-mail: omar@oden.utexas.edu

Key words: Infinite-dimensional inverse problems, adjoint-based methods, inexact Newton-CG method, low-rank approximation, Bayesian inference, uncertainty quantification, sampling, generic PDE toolkit

Abstract. Recent years have seen a massive explosion of datasets across all areas of science and engineering. The central questions are: How do we optimally learn from data through the lens of models? And how do we account for uncertainties in both data and models? These questions can be mathematically framed as Bayesian inverse problems. While powerful and sophisticated approaches have been developed to tackle these problems, such methods are often challenging to implement and typically require first and second order derivatives that are not always available in existing computational models. In this talk, we present an extensible software framework MUQ-hIPPYlib that overcomes these challenges by providing access to state-of-the-art algorithms that offer the capability to solve complex large-scale Bayesian inverse problems across a broad spectrum of scientific and engineering areas.

1 Summary

Inverse problems arise in all areas of science, engineering, technology, and medicine and are often governed by complex physics-based mathematical models. These models are often subject to considerable uncertainties stemming from unknown or uncertain inputs (e.g., coefficient fields, constitutive laws, source terms, geometries, initial and/or boundary conditions) as well as from noisy and limited observations. While many of these input parameters cannot be directly observed, they can be inferred indirectly from observations via an inverse problem. Bayesian inversion provides a framework for integration of data with complex physics-based models to quantify and reduce uncertainties in model predictions [7]. Bayesian inversion with complex forward models faces several computational challenges. First, characterizing the posterior distribution of the parameters of interest or subsequent predictions often requires repeated evaluations of large-scale partial differential equation (PDE) models. Second, the posterior distribution often has a complex structure stemming from nonlinear parameter-to-observable maps and heterogeneous sources of data. Third, the parameters often are fields, which when discretized lead to very high-dimensional posteriors.

Our objective is to create a robust and scalable software framework to tackle large-scale PDE-constrained Bayesian inverse problems across a wide range of science and engineering fields. hIPPYlib-MUQ [8] is a Python interface between two open source software packages, hIPPYlib and MUQ, which have complementary capabilities. hIPPYlib [17, 18] is an extensible software package aimed at solving deterministic and linearized Bayesian inverse problems governed by PDEs. Based on FEniCS [11, 10, 9] for the solution of forward PDE problems and on PETSc [1] for scalable and parallel linear algebra operations and solvers, hIPPYlib implements globalized inexact Newton-conjugate gradient methods, adjoint-based computation of gradients and Hessian actions, low-rank approximation of Hessians, and sampling from large-scale Gaussian fields; see [18] for the details. MUQ [13] is a collection of tools for solving uncertainty quantification problems. MUQ provides a suite of powerful uncertainty quantification algorithms including Markov chain Monte Carlo (MCMC) methods [14], transport maps [12], polynomial chaos expansions [4], Karhunen-Loeve expansions, and Gaussian process modeling [16, 6]. MUQ also provides a framework for easily combining statistical and physical models in a way that supports the efficient computation of gradients, Jacobians, and Hessian-vector products.

hIPPYlib-MUQ integrates these two libraries into a unique software framework, allowing users to implement state-of-the-art Bayesian inversion algorithms for PDE models in a seamless way. In this framework, hIPPYlib is used to define the forward model, the prior, and the likelihood, to compute the maximum a posteriori (MAP) point, and to construct a Gaussian (Laplace) approximation of the posterior distribution based on approximations of the posterior covariance as a low-rank update of the prior [3]. MUQ is employed to exploit advanced MCMC methods to fully characterize the posterior distribution in non-Gaussain/nonlinear settings. hIPPYlib-MUQ offers a set of wrappers that encapsulate the functionality of hIPPYlib in a way that various features of hIPPYlib can be accessed by MUQ. A key aspect of hIPPYlib-MUQ is that it enables the use of MCMC

methods enriched by the Hessian of the log likelihood [5], which is crucial for efficient and scalable exploration of the posterior distribution for large-scale Bayesian inverse problems. For example, the Laplace approximation of the posterior with the low-rank factorization of the Hessian can be invoked to generate high-quality proposals for MCMC methods, thereby significantly enhancing sampling efficiency [15].

hIPPYlib-MUQ also provides convergence diagnostics for MCMC samples: the potential scale reduction factor and its extension to multivariate parameter cases [2], the autocorrelation function, and the effective sample size. hIPPYlib-MUQ is designed for general large-scale Bayesian inverse problems, not only for research and application in diverse fields, but also for educational purposes.

2 Acknowledgements

This work was supported by the U.S. National Science Foundation (NSF), Software Infrastructure for Sustained Innovation (SI2: SSE & SSI) Program under grants ACI-1550487, ACI-1550593, and ACI-1550547. In addition, the authors acknowledge computing time on the Multi-Environment Computer for Exploration and Discovery (MERCED) cluster at UC Merced, which was funded by NSF grant ACI-1429783.

REFERENCES

[1] S. Balay, S. Abhyankar, M. F. Adams, J. Brown, P. Brune, K. Buschelman, L. Dalcin, A. Dener, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, D. A. May, L. C. McInnes, R. T. Mills, T. Munson, K. Rupp, P. Sanan, B. F. Smith, S. Zampini, and H. Zhang, *PETSc Web page*. `http://www.mcs.anl.gov/petsc`, 2018, `http://www.mcs.anl.gov/petsc`.

[2] S. P. Brooks and A. Gelman, *General methods for monitoring convergence of iterative simulations*, Journal of Computational and Graphical Statistics, 7 (1998), pp. 434–455.

[3] T. Bui-Thanh, O. Ghattas, J. Martin, and G. Stadler, *A computational framework for infinite-dimensional Bayesian inverse problems Part I: The linearized case, with application to global seismic inversion*, SIAM Journal on Scientific Computing, 35 (2013), pp. A2494–A2523.

[4] P. R. Conrad and Y. M. Marzouk, *Adaptive Smolyak pseudospectral approximations*, SIAM Journal on Scientific Computing, 35 (2013), pp. A2643–A2670.

[5] T. Cui, K. J. Law, and Y. M. Marzouk, *Dimension-independent likelihood-informed MCMC*, Journal of Computational Physics, 304 (2016), pp. 109–137.

[6] J. Hartikainen and S. Särkkä, *Kalman filtering and smoothing solutions to temporal Gaussian process regression models*, in 2010 IEEE international workshop on machine learning for signal processing, IEEE, 2010, pp. 379–384.

[7] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, vol. 160 of Applied Mathematical Sciences, Springer-Verlag New York, 2005.

[8] K.-T. Kim, U. Villa, M. Parno, N. Petra, Y. Marzouk, and O. Ghattas, *hIPPYlib2MUQ: Version 0.1.0*, Nov. 2020, https://doi.org/10.5281/zenodo.4270739, https://doi.org/10.5281/zenodo.4270739.

[9] H. P. Langtangen and A. Logg, *Solving PDEs in Python*, Springer, 2017.

[10] A. Logg, K.-A. Mardal, and G. Wells, *Automated Solution of Differential Equations by the Finite Element Method: The FEniCS book*, vol. 84, Springer Science & Business Media, 2012.

[11] A. Logg and G. N. Wells, *DOLFIN: Automated finite element computing*, ACM Transactions on Mathematical Software, 37 (2010), pp. 1–28.

[12] Y. Marzouk, T. Moselhy, M. Parno, and A. Spantini, *Sampling via Measure Transport: An Introduction*, Springer International Publishing, 2016, pp. 1–41.

[13] M. Parno, A. Davis, P. Conrad, and Y. Marzouk, *MIT Uncertainty Quantification (MUQ) Library*, 2014, https://muq.mit.edu.

[14] M. D. Parno and Y. M. Marzouk, *Transport map accelerated Markov chain Monte Carlo*, SIAM/ASA Journal on Uncertainty Quantification, 6 (2018), pp. 645–682.

[15] N. Petra, J. Martin, G. Stadler, and O. Ghattas, *A computational framework for infinite-dimensional Bayesian inverse problems: Part II. Stochastic Newton MCMC with application to ice sheet inverse problems*, SIAM Journal on Scientific Computing, 36 (2014), pp. A1525–A1555, https://doi.org/10.1137/130934805.

[16] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, 2005.

[17] U. Villa, N. Petra, and O. Ghattas, *hIPPYlib: An extensible software framework for large-scale inverse problems*, Journal of Open Source Software, 3 (2018), p. 940, https://doi.org/10.21105/joss.00940, https://doi.org/10.21105/joss.00940.

[18] U. Villa, N. Petra, and O. Ghattas, *hIPPYlib: an extensible software framework for large-scale inverse problems; Part I: Deterministic inversion and linearized Bayesian inference*, ACM Transactions on Mathematical Software, (To appear.).