Scalable Algorithms for Bayesian Inference of Large-Scale Models from Large-Scale Data

Omar Ghattas[1(✉)], Tobin Isaac[2], Noémi Petra[3], and Georg Stadler[4]

[1] Institute for Computational Engineering and Sciences,
Departments of Geological Sciences and Mechanical Engineering,
The University of Texas at Austin, Austin, USA
omar@ices.utexas.edu
[2] Computation Institute, University of Chicago, Chicago, USA
[3] School of Natural Sciences, University of California, Merced, Merced, USA
[4] Courant Institute for Mathematical Sciences, New York University, New York, USA

One of the greatest challenges in computational science and engineering today is how to combine complex data with complex models to create better predictions. This challenge cuts across every application area within CS&E, from geosciences, materials, chemical systems, biological systems, and astrophysics to engineered systems in aerospace, transportation, structures, electronics, biomedicine, and beyond. Many of these systems are characterized by complex nonlinear behavior coupling multiple physical processes over a wide range of length and time scales. Mathematical and computational models of these systems often contain numerous uncertain parameters, making high-reliability predictive modeling a challenge. Rapidly expanding volumes of observational data—along with tremendous increases in HPC capability—present opportunities to reduce these uncertainties via solution of large-scale inverse problems.

In an inverse problem, we infer unknown model parameters (e.g., coefficients, material properties, source terms, initial or boundary conditions, geometry, model structure) from observations of model outputs. The need to quantify the uncertainty in the solution of such inverse problems has attracted widespread attention in recent years. This can be carried out in a systematic manner by casting the inverse problem within the framework of Bayesian inference. In this framework, uncertain observations and uncertain models are combined with available prior knowledge to yield a probability density in the model parameters as the solution of the inverse problem, thereby providing a rational and systematic means of quantifying uncertainties in the inference of these parameters. The resulting uncertainties in model parameters are then propagated forward through models to yield predictions with associated uncertainty. Finally, given this capability to quantify uncertainties in inverse problems, one can determine the design of the observational system (e.g., location of sensors, nature of

measured quantities) that maximizes the information gain from the observations (or minimizes the uncertainty in the inferred model or subsequent prediction). This is the optimal experimental design (OED) problem, which wraps an optimization problem around the Bayesian inverse problem.

The Markov chain Monte Carlo (MCMC) method has emerged as the method of choice for solving Bayesian inverse problems. Unfortunately, when the forward model is large and complex (e.g., when the model takes the form of an expensive-to-solve system of partial differential equations), and when the parameters are high-dimensional (as results from discretization of an infinite dimensional field such as an initial condition or heterogeneous material property), solution of Bayesian inverse problems via conventional MCMC is intractable. Moreover, addressing the meta-question of how to optimally obtain experimental data for such problems via solution of an OED problem is completely out of the question.

However, a number of advances over the past decade have brought the goal of Bayesian inference of large-scale complex models from large-scale complex data much closer. First, improvements in scalable forward solvers for many classes of large-scale models have made feasible numerous evaluations of model outputs for differing inputs. Second, sustained growth in HPC capabilities has multiplied the effects of the advances in solvers. Third, the emergence of MCMC methods that exploit problem structure (e.g., curvature of the posterior probability) has radically improved the prospects of sampling posterior distributions for inverse problems governed by expensive models. And fourth, recent exponential expansions of observational capabilities have produced massive volumes of data from which inference of large computational models can be carried out.

To overcome the prohibitive nature of Bayesian methods for high-dimensional inverse problems governed by expensive-to-solve PDEs, we exploit the fact that, despite the large size of observational data, they typically provide only sparse information on model parameters. This implicit dimension reduction is provided by low rank approximations of the Hessian of the data misfit functional, which is typically a compact operator due to ill-posedness of the inverse problem. A low rank approximation of the Hessian can be extracted efficiently in a matrix-free manner (without forming the Hessian) by a Lanczos [8,14] or randomized SVD [4,5,12,15,21] method, requiring a number of matrix-vector products that scales only with the rank of the Hessian, and not the parameter dimension. Moreover, the rank reflects how informative the data are, i.e., how many directions in parameter space are informed by the data. Finally, each Hessian-vector product can be computed using just a pair of linearized forward/adjoint PDE solves [4,5,8,9,12,14–17,21,22].

We have applied the methodology described above (for exploiting the geometric structure of the posterior) to geophysical inverse problems arising in ice sheet flow, seismic wave propagation, mantle convection, atmospheric transport, poromechanics, and subsurface flow. We are able to substantially reduce the effective parameter dimension (often by three orders of magnitude) at a cost, measured in (linearized) forward/adjoint PDE solves, that is independent of both the parameter and data dimensions [4,5,8,9,12,14,15,20,21].

For linearized Bayesian analysis of nonlinear inverse problems, the Hessian evaluated at the point in parameter space that maximizes the posterior (i.e., the MAP point) completely characterizes the uncertainty in inferred parameters. One can build on this idea to solve optimal experimental design problems at a cost that also does not scale with the parameter or data dimensions [1–3]. For nonlinear Bayesian inverse problems, the Hessian varies from point to point. However the low rank Hessian approximation machinery described above can still be exploited to accelerate MCMC sampling, by serving as an inverse covariance approximation for a Gaussian proposal that is tailored to the local curvature of the posterior [14,15] (this is known as the stochastic Newton method).

The most complex inverse problem for which we have carried out Bayesian inversion involves ice sheet flow [12,15,16,22]. The flow of ice from polar ice sheets such as Antarctica and Greenland is the primary contributor to projected sea level rise in the 21st century. The ice is modeled as a creeping, viscous, incompressible, non-Newtonian, shear-thinning fluid, for which we have developed custom scalable parallel solvers [13,18,19] on adaptively refined forest-of-octree meshes [6,7,10,11], the combination of which has scaled to hundreds of billions of unknowns on up to 1.6 million cores [4,6,18]. One of the main difficulties faced in modeling ice sheet flow is the unknown spatially-varying Robin boundary condition that describes the resistance to sliding at the base of the ice. Satellite observations of the surface ice flow velocity can be used to infer this uncertain basal boundary condition. We have solved this ill-posed inverse problem using the (linearized) Bayesian inference machinery described above, which allows us to infer not only the unknown basal sliding parameters, but also the associated uncertainty [12]. We have demonstrated that the number of required forward solves is independent of the parameter dimension, data dimension, and number of processor cores. The largest Bayesian inverse problem solved has over one million uncertain parameters.

References

1. Alexanderian, A., Petra, N., Stadler, G., Ghattas, O.: A-optimal design of experiments for infinite-dimensional Bayesian linear inverse problems with regularized ℓ_0-sparsification. SIAM J. Sci. Comput. **36**(5), A2122–A2148 (2014)
2. Alexanderian, A., Petra, N., Stadler, G., Ghattas, O.: A fast and scalable method for A-optimal design of experiments for infinite-dimensional Bayesian nonlinear inverse problems. SIAM J. Sci. Comput. **38**(1), A243–A272 (2016)
3. Alexanderian, A., Gloor, P., Ghattas, O.: On Bayesian A- and D-optimal experimental designs in infinite dimensions. Bayesian Anal. **11**(3), 671–695 (2016)
4. Bui-Thanh, T., Burstedde, C., Ghattas, O., Martin, J., Stadler, G., Wilcox, L.C.: Extreme-scale UQ for Bayesian inverse problems governed by PDEs. In: Proceedings of IEEE/ACM SC12 (2012)
5. Bui-Thanh, T., Ghattas, O., Martin, J., Stadler, G.: A computational framework for infinite-dimensional Bayesian inverse problems. Part I: The linearized case, with applications to global seismic inversion. SIAM J. Sci. Comput. **35**(6), A2494–A2523 (2013)
6. Burstedde, C., Ghattas, O., Gurnis, M., Isaac, T., Stadler, G., Warburton, T., Wilcox, L.C.: Extreme-scale AMR. In: Proceedings of ACM/IEEE SC 2010 (2010)

7. Burstedde, C., Wilcox, L.C., Ghattas, O.: p4est: Scalable algorithms for parallel adaptive mesh refinement on forests of octrees. SIAM J. Sci. Comput. **33**(3), 1103–1133 (2011)
8. Flath, H.P., Wilcox, L.C., Akcelik, V., Hill, J., van Bloemen, B., Ghattas, O.: Fast algorithms for Bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial Hessian approximations. SIAM J. Sci. Comput. **33**(1), 407–432 (2011)
9. Hesse, M., Stadler, G.: Joint inversion in coupled quasistatic poroelasticity. J. Geophys. Res. Solid Earth **119**, 1425–1445 (2014)
10. Isaac, T., Burstedde, C., Ghattas, O.: Low-cost parallel algorithms for 2:1 octree balance. In: International Parallel and Distributed Processing Symposium (IPDPS 2012), pp. 426–437. IEEE Computer Society (2012)
11. Isaac, T., Burstedde, C., Wilcox, L.C., Ghattas, O.: Recursive algorithms for distributed forests of octrees. SIAM J. Sci. Comput. **37**(5), C497–C531 (2015)
12. Isaac, T., Petra, N., Stadler, G., Ghattas, O.: Scalable and efficient algorithms for the propagation of uncertainty from data through inference to prediction for large-scale problems, with application to flow of the Antarctic ice sheet. J. Comput. Phys. **296**(1), 348–368 (2015)
13. Isaac, T., Stadler, G., Ghattas, O.: Solution of nonlinear Stokes equations discretized by high-order finite elements on nonconforming and anisotropic meshes, with application to ice sheet dynamics. SIAM J. Sci. Comput. **37**(6), B804–B833 (2015)
14. Martin, J., Wilcox, L.C., Burstedde, C., Ghattas, O.: A Stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion. SIAM J. Sci. Comput. **34**(3), A1460–A1487 (2012)
15. Petra, N., Martin, J., Stadler, G., Ghattas, O.: A computational framework for infinite-dimensional Bayesian inverse problems: Part II: Stochastic Newton MCMC with application to ice sheet flow inverse problems. SIAM J. Sci. Comput. **36**(4), A1525–A1555 (2014)
16. Petra, N., Zhu, H., Stadler, G., Hughes, T.J.R., Ghattas, O.: An inexact Gauss-Newton method for inversion of basal sliding and rheology parameters in a nonlinear Stokes ice sheet model. J. Glaciol. **58**(211), 889–903 (2012)
17. Ratnaswamy, V., Stadler, G., Gurnis, M.: Adjoint-based estimation of plate coupling in a non-linear mantle flow model: theory and examples. Geophys. J. Int. **202**(2), 768–786 (2015)
18. Rudi, J., Malossi, A.C.I., Isaac, T., Stadler, G., Gurnis, M., Staar, P.W.J., Ineichen, Y., Bekas, C., Curioni, A., Ghattas, O.: An extreme-scale implicit solver for complex PDEs: highly heterogeneous flow in earth's mantle. In: Proceedings of IEEE/ACM SC 2015 (2015)
19. Rudi, J., Stadler, G., Ghattas, O.: Weighted BFBT Preconditioner for Stokes Flow Problems with Highly Heterogeneous Viscosity (submitted) (2016)
20. Worthen, J., Stadler, G., Petra, N., Gurnis, M., Ghattas, O.: Towards adjoint-based inversion for rheological parameters in nonlinear viscous mantle flow. Phys. Earth Planet. Inter. **234**, 23–34 (2014)
21. Zhu, H., Li, S., Fomel, S., Stadler, G., Ghattas, O.: A Bayesian approach to estimate uncertainty for full waveform inversion using a priori information from depth migration. Geophysics **81**(5), R307–R323 (2016)
22. Zhu, H., Petra, N., Stadler, G., Isaac, T., Hughes, T.J.R., Ghattas, O.: Inversion of geothermal heat flux in a thermomechanically coupled nonlinear Stokes ice sheet model. Cryosphere **10**, 1477–1494 (2016)