Transport-Based Variational Bayesian Methods for Learning from Data

Daniele Bigoni1, Joshua Chen2, Peng Chen2, Omar Ghattas2, Youssuf Marzouk2,2, Tom O’Leary-Rosebury2,2, Keyi Wu2

1Center for Computational Engineering, Massachusetts Institute of Technology
2Institute for Computational Engineering & Sciences, The University of Texas at Austin

Abstract and Motivation

Inverse problems—in which we learn from data through the lens of models—arise across numerous fields of science, engineering, technology, and medicine, and in particular our driving applications in advanced manufacturing and materials. Bayesian inference provides a systematic statistical framework for learning from data—e.g., fusing data with models and quantifying uncertainty in the results. Yet inference in large-scale settings—incorporating large multitudinal data sets, complex physics-based models, and high-dimensional parameter spaces—remains an enormous computational challenge. Moreover, inference is often only an element of “outer loop” analyses such as optimization under uncertainty or optimal experimental design, and hence must be performed repeatedly and quickly. While advanced structure-exploiting sampling methods (e.g., Markov chain Monte Carlo or sequential Monte Carlo methods) that significantly accelerate sampling have been developed in recent years (e.g., [1, 2]), many large-scale complex problems remain out of reach. To overcome these barriers, we are developing new scalable inference strategies that replace sampling with optimization. In particular, we are advancing variational inference methodologies based on transportation of measures, which describe conditioning via the action of a nonlinear map. Transport maps offer a rich and flexible representation of complex posterior distributions in non-Gaussian settings, along with the ability to continuously trade off accuracy and computational cost. We propose (1) adaptive (semi-)parametric approaches and (2) completely nonparametric approaches for representing maps, each coupled with suitable optimization methods. In both cases, our methods exploit low-dimensional structure: low-dimensional data-informed subspaces, approximate independence, or approximate conditional independence. We demonstrate inference across a spectrum of problems, including inverse problems arising in PDEs, state-space models, and statistical models in machine learning.

Approach

We consider parametric and nonparametric transport methods for the solution of Bayesian inference problems. Given an intractable target/posterior distribution \( p_* \) on \( \mathbb{R}^n \) with unnormalized density \( p \) and a tractable reference distribution \( \nu \), we seek a map \( T: \mathbb{R}^n \rightarrow \mathbb{R}^n \) that pushes forward \( \nu \) to \( p_* \), denoted \( T_* \nu = p_* \). This map renders challenging integration problems tractable:

\[
\int f(x) \nu(dx) = \int f(T(x)) \nu(dx).
\]

The map can be identified as the minimizer of the following variational problem

\[
\arg\min_{T} \mathbb{E}_{\nu} \left[ \log \left( \frac{\nu}{T_* \nu} \right) \right].
\]

- A parametric formulation [3] approximates the Knothe–Rosenblatt rearrangement within a space \( T_* \nu \) of lower triangular monotone maps:

\[
T_{ij}(x_1, \ldots, x_n) = \begin{cases} x_i & \text{if } i < j, \\ x_{ij} & \text{if } i = j, \\ 0 \end{cases} \quad \forall \Omega \in \mathbb{R}^n.
\]

- A nonparametric formulation can be obtained as the composition

\[
T(x) = S \circ T_{(1)} \circ \ldots \circ T_{(n)}, \quad T_{(i)} : x_i \mapsto x_{i+1} + Q_i, \quad Q_i = \text{a reproducing kernel Hilbert space}.
\]

Both parametric and nonparametric formulations encounter difficulties as the dimension of the problem increases:

- The expected approximation is quadratically increasing with dimension (deterministic or random) whose accuracy deteriorates with dimension.
- Parametric maps can involve bases of exponentially increasing cardinality; this can be mitigated with sparse approximations.
- Nonparametric maps can exhibit “mode collapse” in the pushforward distribution, which can be avoided by projecting the map to low-dimensional subspaces [5].

Potential Impact

Inverse and inference problems arise across numerous scientific and technological areas, and address the foundational problem of how we learn from data through the lens of models. In particular, numerous problems of DOE interest fall in this category. For example, in the case of PDEs or ODEs, we can infer: subsurface permeability and contaminant concentrations from well measurements; ice basal boundary conditions from satellite observations of surface velocities; state estimation of oceans from altimeter and ocean probes; material microstructural properties from X-ray scattering data; ast-built interior geometry of accelerators from measured EM fields; neutron star merger dynamics from measurements of gravitational fields; biromolecular potentials from dynamics; combustion reaction mechanisms from species concentrations; and so on. Beyond differential equation models, machine learning of graph-based, kernel-based, agent-based, Gaussian process-based, or neural network-based models is fundamentally an inference problem. In all of these, a Bayesian framework is attractive since it is capable of rigorously accounting for uncertainties, given uncertainties in observations, parameters, and the models themselves. The methods we are developing offer the hope of tackling large-scale instances of these, and many more, problems. To make our developments more accessible, they are being incorporated into our open source libraries for inverse problems and uncertainty quantification [7, 8].

Synergy

As mentioned above, the methods developed here are applicable to a broad spectrum of model-based inference-from-data problems. In many cases, we seek to infer infinite-dimensional fields, such as initial conditions, boundary conditions, sources, heterogeneous material properties, or geometry. Upon discretization, these lead to very high dimensional parameter spaces. The scalability of the methods we are developing relies on exploiting the underlying intrinsic low-dimensional structure. One technique we employ to achieve this is through low-rank approximations of Hessians of the log posterior. However some Hessians may not admit a global low rank structure. There is an opportunity to develop other compressed representations—requiring few forward model solves—using for example analytical representations, hierarchal matrices, or product-convolution approximations, which are currently active areas of research.

References