Optimal Transportation for Bayesian Inference in Engineering

Joseph B. Nagel & Bruno Sudret
ETH Zürich, Institute of Structural Engineering, Zürich, Switzerland

KEYWORDS: Bayesian inference; inverse problems; optimal transport maps

ABSTRACT: The proper assessment of uncertainties plays a key role in reliability analysis. Bayesian inference establishes a convenient framework for the solution of inversion problems and uncertainty quantification [1,2]. Classical inverse problems are posed when unknown parameters of a forward model are inferred based on measurements of forward model response quantities. The Bayesian probabilistic approach to inversion is based on a prior distribution, representing the uncertainty of the true parameter value before the data are analyzed, which is updated into the posterior distribution by conditioning on the data.

Bayesian inference is most often accomplished by means of Markov chain Monte Carlo (MCMC) techniques [3]. The posterior distribution is sampled by constructing an appropriate Markov chain over its support. Key challenges to MCMC include an enormous number of necessary forward model runs, the presence of sample autocorrelation and the lack of a clear means to monitor and assess convergence. Nowadays the employment of cheap forward model surrogates, e.g. based on polynomial chaos expansions (PCE), is a widespread practice for accelerating Bayesian inference in inverse problems [4,5]. However, this approach still suffers from various shortcomings of sampling-based posterior representations.

Beyond advanced and highly efficient MCMC samplers, recent research work also focuses on numerical solutions to Bayesian inference that entirely avoid MCMC [6,7]. One approach is based on polynomial maps that push the prior to the posterior measure [6]. Practically the inferential map is computed by pulling-back the posterior to the prior. This formulation rests upon optimal transportation theory. As an alternative to the theory of Markov chains, the abovementioned approach establishes a fundamentally new framework for accomplishing Bayesian inference. Nevertheless there are many open questions, e.g. regarding its applicability in typical engineering applications and its competitiveness against advanced MCMC methods.

In this contribution we discuss the mapping approach in the context of variational Bayesian inference [8]. Moreover, we investigate the potential of optimal transport maps for engineering problems. An inverse heat conduction problem serves as a benchmark application to compare the approach with MCMC. More specifically, the identification of unknown thermal conductivities of inclusions in a composite material (see Fig. 1) is envisaged.

Figure 1: Heat conduction setup
1. VARIATIONAL BAYESIAN INFERENCE

Let $\pi(x)$ be the prior density of the vector of unknowns $x \in \mathbb{R}^m$. Moreover, let us denote the likelihood function as $L(x) = \pi(y|x)$ where $y \in \mathbb{R}^n$ is the actually acquired data. One obtains the posterior density by conditioning on the data via Bayes’ law

$$\pi(x|y) = \frac{1}{Z} L(x) \pi(x).$$

In variational Bayesian inference [8] a member $\tilde{q} \in Q$ from a parametric class of probability densities $Q$ is sought such that the degree of dissimilarity relative to the posterior is minimized. A commonly used measure of the difference between $\tilde{q}$ and $\pi(\cdot|y)$ is the Kullback-Leibler (KL) divergence

$$D_{KL}(\tilde{q}||\pi(\cdot|y)) = \int \log \left( \frac{\tilde{q}(x)}{\pi(x|y)} \right) \tilde{q}(x) d x = \mathbb{E}_\tilde{q} \left[ \log \left( \frac{\tilde{q}(X)}{\pi(X|y)} \right) \right].$$

The non-symmetric KL divergence is here arranged in such a way that tractable expectations with respect to the parametric distribution $\tilde{q}$ emerge. The “best” approximation of the posterior $\tilde{q}(x) \approx \pi(x|y)$ is then found by solving the optimization problem

$$\tilde{q} = \arg\min_{q \in Q} D_{KL}(q^*||\pi(\cdot|y)).$$

2. OPTIMAL TRANSPORT MAPS

In [6] the prior and the posterior are coupled as follows. The random variable $X \sim \pi$ is distributed according to the prior. The goal is to find a map $T : \mathbb{R}^m \rightarrow \mathbb{R}^m$ such that the transformed random variable $\tilde{X} = T(X)$ follows the posterior $\tilde{\pi} = \pi(\cdot|y)$. Under certain assumptions one may write the corresponding transformation of the probability density as

$$\tilde{\pi}(\tilde{x}) = \pi(T^{-1}(\tilde{x})) |\det J_{T^{-1}}(\tilde{x})|,$$

where $J_{T^{-1}} = \partial T^{-1}/\partial \tilde{x}$ is the Jacobian matrix. Conversely, in order to compute a map $T \in P_p$, e.g. in the space $P_p$ of multivariate polynomials up to degree $p$, one can solve the optimization problem

$$\tilde{T} = \arg\min_{T \in P_p} D_{KL}(\pi||Z^{-1}(L \pi) \circ T^* |\det J_{T^*}|).$$

Here, $D_{KL}$ denotes the KL divergence. It is here posed in such a way that tractable prior expectations emerge in the objective function. This optimization problem is reminiscent of variational Bayesian inference. On top of that, it is supported by optimal transportation theory. The latter ensures the existence of a solution. That the uniqueness of the solution is guaranteed under certain “cost”-considerations can assist in regularizing the optimization problem. Once a transport map is computed, the posterior is sampled by independently drawing from the prior and simply applying the map.

REFERENCES