Journal of Computational Mathematics
Vol.39, No.6, 2021, 848–864.
http://www.global-sci.org/jcm
doi:10.4208/jcm.2102-m2020-0339

AN ACCELERATION STRATEGY FOR RANDOMIZE-THEN-OPTIMIZE SAMPLING VIA DEEP NEURAL NETWORKS*

Liang Yan
School of Mathematics, Southeast University, Nanjing, 210096, China;
Nanjing Center for Applied Mathematics, Nanjing, 211135, China
Email: yanliang@seu.edu.cn
Tao Zhou[1)]
LSEC, Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, China
Email: tzhou@lsec.cc.ac.cn

Abstract

Randomize-then-optimize (RTO) is widely used for sampling from posterior distributions in Bayesian inverse problems. However, RTO can be computationally intensive for complexity problems due to repetitive evaluations of the expensive forward model and its gradient. In this work, we present a novel goal-oriented deep neural networks (DNN) surrogate approach to substantially reduce the computation burden of RTO. In particular, we propose to drawn the training points for the DNN-surrogate from a local approximated posterior distribution – yielding a flexible and efficient sampling algorithm that converges to the direct RTO approach. We present a Bayesian inverse problem governed by elliptic PDEs to demonstrate the computational accuracy and efficiency of our DNN-RTO approach, which shows that DNN-RTO can significantly outperform the traditional RTO.

Mathematics subject classification: 35R30, 62F15, 65C60, 68T05.
Key words: Bayesian inverse problems, Deep neural network, Markov chain Monte Carlo.

1. Introduction

The Bayesian approach provides a systematic framework for quantifying the uncertainty in parameter estimations of inverse problems [14, 26]. In the Bayesian approach, one combine a prior knowledge of the unknown parameters and the forward model to yield a *posterior* probability distribution, form which the statistic information of the unknown parameters can be characterized. The main task of the Bayesian approach is to draw samples from the posterior distributions, and then evaluate the associate statistic information, e.g., expectation, variance, etc. Since analytical formulas of the posterior are in general not available, many numerical sampling approaches such as Markov chain Monte Carlo (MCMC) methods [7] have been developed.

Nevertheless, the MCMC sampling scheme can be computationally challenging. Firstly, each evaluation of the system output involves a forward model evaluation, and this is infeasible if the model is expensive to evaluate. Secondly, the geometry of the posterior distribution may admits complex features in the parametric space (such as local concentration). One polular

* Received March 12, 2020 / Revised version received January 27, 2021 / Accepted February 22, 2021 / Published online September 15, 2021 /
1) Corresponding author

way for reducing the computational complexity is to use the so-called surrogate approach: one constructs a surrogate to the true model and samples from the posterior distribution induced by the surrogate. In case the surrogate is computationally less expensive, one can dramatically speed up the MCMC algorithms. Different surrogate approaches have been investigated in recent years, for instance, projection-type reduced order models [1, 9, 18], polynomial chaos (PC) based surrogates [20–22, 31, 32], and Gaussian process regression [15, 25, 27], to name a few. Although the surrogate approach can be very effective when exploring low-dimensional distributions, it can still be inefficient for high-dimensional distributions with local concentration [7]. In these cases, the effective sample size (ESS) tends to be very low. To address this challenge, several strategies that combine the geometry information of the posterior (such as the gradient and Hessian information) have been exploited to accelerate the convergence of MCMC, see, e.g. [4, 11, 13, 17, 19]. We also mention recent progresses on adaptive multi-fidelity surrogate approaches [33–35].

In this work, we propose a new approach that combines a deep neural networks (DNN) surrogate and an optimization-based sampling approach for large-scale PDE constrained Bayesian inverse problems. In particular, we focus on the randomize-then-optimize (RTO) approach [2, 3, 30] which uses repeated solutions of a randomly perturbed optimization problem to produce samples from a non-Gaussian distribution (used as a Metropolis independence proposal). Compared to the classical Metroplis-Hastings (MH) random walk algorithm, RTO admits higher acceptance probability and lower sample auto-correlation even for high-dimensional problems [2]. The main drawback of the original RTO lines in the repetitive evaluation of the forward model and its gradient. To this end, we construct a DNN-based surrogate which makes the optimization problems rather efficient to solve. Moreover, to obtain an accurate and efficient DNN-surrogate, we propose to generate the training points from a local approximated posterior distribution, and this makes the training procedure very efficient.

We next summarize the main features of our DNN-RTO approach:

- A new approach that combines RTO and a DNN-surrogate. The new approach is expected to be promising for high dimensional problems.

- With the DNN-surrogate, the gradient information can be obtained efficiently by the backward propagation, so that the associated optimization procedure can be efficiently solved.

- We choose the training points from a local approximated posterior distribution, and this makes the training procedure (for the DNN-surrogate) very efficient.

- We present numerical examples to demonstrate that the DNN-RTO outperforms the traditional RTO.

The rest of the paper is organized as follows. In the next section, we give a brief introduction to the Bayesian inverse problems. In Section 3, we introduce the RTO algorithm. Details of our new approach are presented in Section 4. In Section 5, we use two nonlinear inverse problems to demonstrate the accuracy and efficiency of the new approach. Finally, we give some concluding remarks in Section 6.