

An Adaptive Surrogate Modeling Based on Deep Neural Networks for Large-Scale Bayesian Inverse Problems

Liang Yan^{1,2} and Tao Zhou^{3,*}

¹ School of Mathematics, Southeast University, Nanjing 210096, P.R. China.

² Nanjing Center for Applied Mathematics, Nanjing 211135, P.R. China.

³ LSEC, Institute of Computational Mathematics, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, P.R. China.

Received 22 September 2020; Accepted 26 September 2020

Abstract. In Bayesian inverse problems, surrogate models are often constructed to speed up the computational procedure, as the parameter-to-data map can be very expensive to evaluate. However, due to the curse of dimensionality and the non-linear concentration of the posterior, traditional surrogate approaches (such as the polynomial-based surrogates) are still not feasible for large scale problems. To this end, we present in this work an adaptive multi-fidelity surrogate modeling framework based on deep neural networks (DNNs), motivated by the facts that the DNNs can potentially handle functions with limited regularity and are powerful tools for high dimensional approximations. More precisely, we first construct offline a DNN-based surrogate according to the prior distribution, and then, this prior-based DNN-surrogate will be adaptively & locally refined online using only a few high-fidelity simulations. In particular, in the refine procedure, we construct a new shallow neural network that view the previous constructed surrogate as an input variable – yielding a composite multi-fidelity neural network approach. This makes the online computational procedure rather efficient. Numerical examples are presented to confirm that the proposed approach can obtain accurate posterior information with a limited number of forward simulations.

AMS subject classifications: 35R30, 62F15, 65N21, 65C60

Key words: Bayesian inverse problems, deep neural networks, multi-fidelity surrogate modeling, Markov chain Monte Carlo.

*Corresponding author. *Email addresses:* yanliang@seu.edu.cn (L. Yan), tzhou@lsec.cc.ac.cn (T. Zhou)

1 Introduction

Inverse problems arise when one is interested in determining model parameters or inputs from a set of indirect observations [5, 12]. Typically, inverse problems are ill-posed in the sense that the solution may not exist or may not be unique. More importantly, the parameters may not depend continuously on the observations – meaning that a small perturbation in the data may cause an enormous deviation in the solution. The Bayesian approach [12, 28] is a popular approach for inverse problems which casts the solution as a *posterior distribution* of the unknowns conditioned on observations, and introduces regularization in the form of *prior* information. By estimating statistic moments according to the posterior distribution, one not only gets point estimates of the parameters, but also obtains a complete description of the uncertainty in model predictions. However, in practice, the analytical treatment for the posterior is not feasible in general due to the complexity of the system. Consequently, the posterior is often approximated with numerical approaches such as the Markov chain Monte Carlo (MCMC) method.

In the standard MCMC approach, one aims at generating samples directly from the posterior distribution over the parameters space by using the unnormalized posterior, i.e., the product of the prior and likelihood. However, the cost of evaluating the likelihood in the sampling procedure can quickly become prohibitive if the forward model is computationally expensive. One popular way to reduce the computational cost in the sampling procedure is to replace the original forward model with a cheap surrogate model [7, 11, 13, 16, 17, 20–22, 29, 33]. Using a computationally less expensive, offline constructed, surrogate model can make the online computations very efficient. Furthermore, theoretical analysis shows that if the surrogate converges to the true model in the prior-weighted L_2 norm, then the posterior distribution generated by the surrogate converges to the true posterior [22, 29, 33, 34].

Although the surrogate approach can provide significant empirical performance improvements, there are however many challenges for practical applications. First, constructing a sufficiently accurate surrogate over the whole domain of the prior distribution may not be possible for many practical problems. Especially, the posterior distribution often concentrates on a small fraction of the support of the prior distribution, and a globally prior-based surrogate may not be accurate for online computations [15]. To improve this, posterior-focused approaches have been suggested recently, where one constructs a sequence of local surrogates in the important region of the posterior distribution, to alleviate the effect of the concentration of posterior [4, 15].

In our previous work [36], we also presented an adaptive multi-fidelity surrogate modelling procedure based on PCEs to speed up the online computations via MCMC. The idea is to begin with a low fidelity PCE-surrogate, and then correct it adaptively using online high fidelity data. Empirical studies on problems of moderate dimension show that the number of high-fidelity model evaluations can be reduced by orders of magnitude, with no discernible loss of accuracy in posterior expectations. Nevertheless, the approaches in [36] also admit some limitations: (i) the PCE surrogate has limitations