Efficient Bayesian Parameter Inversion Facilitated by Multi-Fidelity Modeling

Yaning Liu
Department of Mathematical & Statistical Sciences
University of Colorado Denver
Denver, Colorado, USA
yaning.liu@ucdenver.edu

***Abstract*—We propose an efficient Bayesian parameter inversion technique that utilizes the implicit particle filter to characterize the posterior distribution, and a multi-scale surrogate modeling method called the proper orthogonal decomposition mapping method to provide high-fidelity solutions to the forward model by conducting only low-fidelity simulations. The proposed method is applied to the nonlinear Burgers equation, widely used to model electromagnetic waves, with stochastic viscosity and periodic solutions. We consider solving the equation with a coarsely-discretized finite difference scheme, of which the solutions are used as the low-fidelity solutions, and a Fourier spectral collocation method, which can provide high-fidelity solutions. The results demonstrate that the computational cost of characterizing the posterior distribution of viscosity is greatly reduced by utilizing the low-fidelity simulations, while the loss of accuracy is unnoticeable.**

***Index Terms*—Bayesian parameter inversion, implicit particle filters, proper orthogonal decomposition mapping method, multi-fidelity modeling, surrogate modeling.**

I. Introduction

Bayesian parameter inversion (BPI) techniques are desirable for estimating model parameters, since they consider the uncertainty associated with the inversion in the form of a posterior distribution. However, BPI is computationally expensive because it involves a number of forward model simulations to sample the posterior distribution. If it is expensive to obtain the solutions to the forward model, such as the fine-resolution solutions to a physical model, the BPI can become intractable.

As a widely used BPI technique, the Markov Chain Monte Carlo (MCMC) method is robust for estimating posterior information; however, the efficiency of MCMC can be low due to its acceptance-rejection nature. The particle filtering (PF) method, as an alternative to MCMC, has many advantages, such as being "embarrassingly parallel" and applicable to high dimensional systems. Recently, a particular PF method, the implicit particle filter (IPF), was developed in [1] and [2], which can alleviate or prevent the so-called "weight collapse" phenomenon. It has been shown, by [3] for instance, that IPF can outperform state-of-the-art MCMC for some applications.

Surrogate models, also known as reduced order models, serving as efficient approximations for computationally expensive models, have been used in BPI, including IPF [3], to improve efficiency. Surrogate models are constructed in the training period, and can then be used to approximate the original models with negligible computational cost. Most surrogate modeling techniques address the approximation of the mapping from the model input parameters to a scalar model output (e.g., [4], [5]), while an efficient method that can approximate a solution field (see [6] for example) is more attractive, since the solutions of many problems in electromagnetics take the form of a field of solutions, such as the solutions to the three dimensional Maxwell equations. One particular technique, called the proper orthogonal decomposition mapping method (PODMM) [7], can approximate a fine-resolution solution from a coarse-resolution solution, which is significantly more inexpensive to obtain. The PODMM technique has been applied to climate models and shown good accuracy when used to construct surrogates.

In this paper, we consider BPI with IPF for the one-dimensional Burgers equation with stochastic viscosity. To further improve the efficiency, we couple PODMM with IPF by building a multi-scale surrogate mapping to accurately approximate the solutions to the Burgers equation required by the IPF. In the following sections, we first show the technical details of IPF and PODMM, and then present the numerical results for the stochastic Burgers equation.

II. Methodology

A. Implicit Particle Filters

Let $\boldsymbol{D}$ be the k-dimensional observations, $\boldsymbol{f}$ be the forward model and $\boldsymbol{\theta}$ be the d-dimensional unknown input parameters. They are related by:

$$\boldsymbol{D} = \boldsymbol{f}(\boldsymbol{\theta}) + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{\epsilon}$ is some random noise. In Bayesian parameter inversion, we aim to obtain the conditional probability distribution of the parameters given the observations, $p(\boldsymbol{\theta}|\boldsymbol{D})$, known as the posterior density. Combining the knowledge of the parameters prior to data being available, $p(\boldsymbol{\theta})$, the prior distribution, and the measure of how likely a set of model output values are given the parameters $p(\boldsymbol{D}|\boldsymbol{\theta})$, the likelihood function, the Bayes' theorem infers the posterior by:

$$p(\boldsymbol{\theta}|\boldsymbol{D}) \propto p(\boldsymbol{\theta})p(\boldsymbol{D}|\boldsymbol{\theta}). \tag{2}$$

In general, $p(\boldsymbol{\theta}|\boldsymbol{D})$ takes a complex form and therefore cannot be sampled directly, for instance by the inverse-transform method or the acceptance-rejection sampling method [8].

Submitted On: September 30, 2018
Accepted On: October 3, 2018

One widely used method that facilitates such sampling is importance sampling, in the framework of which, we instead sample from a new, easy-to-sample probability distribution $\pi(\boldsymbol{\theta})$, called importance distribution. The resulting N samples $\boldsymbol{\theta}_i$, $i = 1, \ldots, N$, are then weighted by:

$$w_i = \frac{p(\boldsymbol{\theta_i})p(\boldsymbol{D}|\boldsymbol{\theta_i})}{\pi(\boldsymbol{\theta}_i)}, \quad i = 1, 2, \ldots, N \tag{3}$$

to correct the bias from taking samples from the importance distribution rather than the original distribution, so that the samples and the weights collectively characterize the posterior distribution. To achieve this, a resampling procedure [9] is used to eliminate samples with small weights and obtain a set of samples that discretely represents the posterior distribution of $\boldsymbol{\theta}$. Unlike MCMC, represented by the well-known Metropolis-Hastings algorithm and Gibbs sampling, the samples in importance sampling are independent and thus can be embarrassingly parallel. Nonetheless, the importance function must be chosen carefully, or else the sampling can be inefficient due to the appearance of a large portion of samples having weights that are small enough to be negligible.

IPFs are based on importance sampling, where an importance distribution is constructed by computing the maximizer of the posterior $p(\boldsymbol{\theta}|\boldsymbol{D})$, i.e., the maximum a posteriori (MAP) of the parameters $\boldsymbol{\theta}$ given the data D, which can be found by minimizing the objective function:

$$F(\boldsymbol{\theta}) = -\log\left(p(\boldsymbol{\theta})p(\boldsymbol{D}|\boldsymbol{\theta})\right),$$

where $p(\boldsymbol{\theta})$ and $p(\boldsymbol{D}|\boldsymbol{\theta})$ are the prior distribution and the likelihood function, respectively. The idea is to construct a importance distribution so that it has large values where the posterior is large. Once the minimization problem is solved, one generates samples in the neighborhood of the minimizer $\boldsymbol{\mu} = \arg\min F$ as follows. A sequence of reference variables $\{\boldsymbol{\xi}_i\}_{i=1}^N$ are first sampled from a reference probability distribution $g(\boldsymbol{\xi})$, and subsequently each target posterior sample $\boldsymbol{\theta_i}$ is obtained by solving,

$$F(\boldsymbol{\theta}_i) - \phi = G(\boldsymbol{\xi}_i) - \gamma, \tag{4}$$

where $\phi = \min_{\boldsymbol{\theta}} F$, $G(\boldsymbol{\xi}) = -\log\left(g(\boldsymbol{\xi})\right)$ and $\gamma = \min_{\boldsymbol{\xi}} G$. The sample weights are:

$$w_i = J(\boldsymbol{\theta}_i), \tag{5}$$

where J is the Jacobian of the bijective map $\boldsymbol{\xi} \to \boldsymbol{\theta}$ [2]. Note that the sequence of samples $\boldsymbol{\theta}_i$ obtained by solving Eq. (4) are in the neighborhood of the MAP $\boldsymbol{\mu}$, since the right-hand side is small if $\boldsymbol{\xi}_i$'s are sampled close to the minimizer of G. Thus, Eq. (4) maps a likely $\boldsymbol{\xi}$ to a likely $\boldsymbol{\theta}$.

However, solving the mapping Eq. (4) is nontrivial, noting that it is in general nonlinear. One strategy, named "linear maps", is inspired by the approximation of F by its second-order Taylor expansion around the MAP $\boldsymbol{\mu}$:

$$F_0(\boldsymbol{\theta}) = \phi + \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\mu})^T H(\boldsymbol{\theta} - \boldsymbol{\mu}), \tag{6}$$

where H is the Hessian matrix at $\boldsymbol{\mu}$. For an uncorrelated standard Gaussian reference variable, Eq. (4) is transformed to:

$$F(\boldsymbol{\theta}_i) - \phi = \frac{1}{2}\boldsymbol{\xi}_i^T\boldsymbol{\xi}_i. \tag{7}$$

Substituting $F(\boldsymbol{\theta}_i)$ in Eq. (7) with $F_0(\boldsymbol{\theta}_i)$, the mapping now takes the simple form:

$$\boldsymbol{\theta}_i = \boldsymbol{\mu} + L^{-T}\boldsymbol{\xi}_i, \tag{8}$$

where L is a lower triangular matrix obtained from the Cholesky decomposition of H. Accounting for the error of approximating the objective function F by F_0, the weights of the samples are computed as:

$$w_i \propto \exp\left(F_0(\boldsymbol{\theta_i}) - F(\boldsymbol{\theta_i})\right). \tag{9}$$

The posteriors are then represented by the ensemble of weighted samples $\{(\boldsymbol{\theta}_i, w_i)\}_{i=1}^N$.

B. Proper Orthogonal Decomposition Mapping Method

PODMM maps $\boldsymbol{g} = [g_1, \ldots, g_{\mathcal{N}_g}]^T$, the coarse-resolution solutions, to fine-resolution solutions $\boldsymbol{f} = [f_1, \ldots, f_{\mathcal{N}_f}]^T$, where $\mathcal{N}_g$ and $\mathcal{N}_f$ are the respective degrees of freedom. Suppose N coarse- and fine-resolution solutions, $\{\boldsymbol{g}_1, \ldots, \boldsymbol{g}_N\}$ and $\{\boldsymbol{f}_1, \ldots, \boldsymbol{f}_N\}$ are available for the training process. We first form the data matrix $\boldsymbol{W}^{\text{PODMM}}$ given by:

$$\boldsymbol{W}^{\text{PODMM}} = \begin{bmatrix} \boldsymbol{f}_1 - \bar{\boldsymbol{f}} & \cdots & \boldsymbol{f}_N - \bar{\boldsymbol{f}} \\ \boldsymbol{g}_1 - \bar{\boldsymbol{g}} & \cdots & \boldsymbol{g}_N - \bar{\boldsymbol{g}} \end{bmatrix},$$

where $\bar{\boldsymbol{f}}$ and $\bar{\boldsymbol{g}}$ are the averages of $\{\boldsymbol{f}_i\}_{i=1}^N$ and $\{\boldsymbol{g}_i\}_{i=1}^N$ respectively. We then determine M right singular vectors, $\boldsymbol{V} = \{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_M\}$ corresponding to the M largest singular values for $\boldsymbol{W}^{\text{PODMM}}$. The proper orthogonal decomposition (POD) bases are $\boldsymbol{\xi}_i = \boldsymbol{W}^{\text{PODMM}}\boldsymbol{v}_i, i = 1, \ldots, M$. Split $\boldsymbol{\xi}_i$ into,

$$\boldsymbol{\xi}_i = \begin{bmatrix} \boldsymbol{\xi}_i^f \\ \boldsymbol{\xi}_i^g \end{bmatrix}.$$

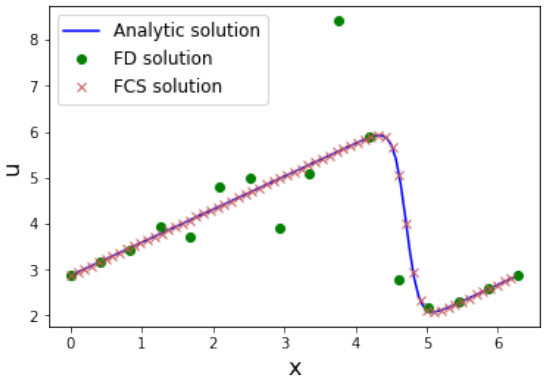

Fig. 1. Analytic, FD and FSC solutions for Burgers equation with $\nu = 0.2$.

For any given coarse-resolution $\boldsymbol{g} \notin \{\boldsymbol{g}_1, \ldots, \boldsymbol{g}_N\}$, we can predict the corresponding fine-resolution $\boldsymbol{f}$ by,

$$\boldsymbol{f} = \bar{\boldsymbol{f}} + \sum_{i=1}^{M} \gamma_i \boldsymbol{\xi}_i^f.$$

Here $\boldsymbol{\gamma} = \{\gamma_1, \ldots, \gamma_m\}$ is the solution to the least-squares problem,

$$\arg\min_{\boldsymbol{\gamma}} \left\| \boldsymbol{g} - \bar{\boldsymbol{g}} - \sum_{i=1}^{M} \gamma_i \boldsymbol{\xi}_i^g \right\|_2,$$

where

$$\|\boldsymbol{h}\|_2 = \left(\frac{1}{\mathcal{N}_h} \sum_{i=1}^{\mathcal{N}_h} h_i^2 \right)^{1/2},$$

is the root mean square for a given vector $\boldsymbol{h}$ with degree of freedom $\mathcal{N}_h$.

III. Results and Discussion

To illustrate the proposed method, we consider the one-dimensional Burgers equation,

$$\frac{\partial u}{\partial t} + \frac{1}{2}\frac{\partial u^2}{\partial x} - \nu \frac{\partial^2 u}{\partial x^2} = 0,$$

where the viscosity ν is a random variable. We consider a periodic solution with an infinite number of N-wave solutions. The analytical periodic solution we use is:

$$u = -2\nu \frac{\phi_x}{\phi}$$

$$\phi(x,t) = \frac{1}{4\pi\nu t} \sum_{n=-\infty}^{\infty} \exp -(x - 2\pi n)^2 / 4\nu t.$$

The coarse-resolution solutions are obtained from a second-order finite difference (FD) spatial discretization on a grid of 16 points, while the fine-resolution solutions are computed by a Fourier spectral collocation (FSC) method [10] with 64 spatial collocation points. The Runge-Kutta method is used for the temporal discretization in both cases.

Fig. 1 shows the analytical, FD, and FSC solutions at $t = \frac{\pi}{8}$ to the Burgers equation corresponding to $\nu = 0.2$. It can be seen that the FD solution oscillates around the shock, while the FSC solution can resolve the transition zone successfully.

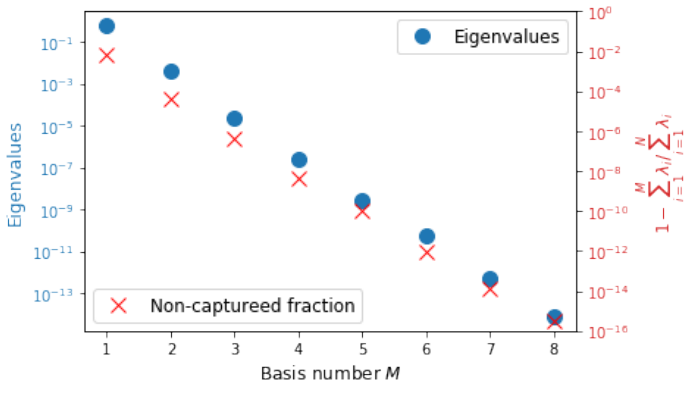

Fig. 2. Eigenvalues (left y axis) and the noncaptured fraction of the variability (right y axis) of the training data.

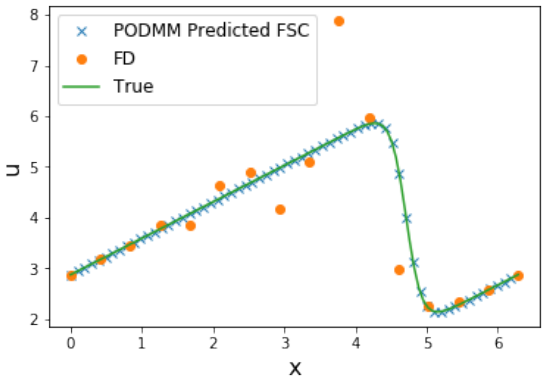

Fig. 3. Analytic, FD and PODMM predicted FSC solutions for Burgers equation with $\nu = 0.304$.

To test the performance of PODMM surrogate model in IPF, we consider a synthetic test case, where the FSC solution corresponding to $\nu = 0.2$ is perturbed by a Gaussian noise with mean 0 and 10% of the solution as the standard deviation. The prior distribution for the viscosity is also taken as Gaussian: $\mathcal{N}(0.4, 0.2^2)$. The coarse-resolution solutions $\boldsymbol{g}$'s are the FD solutions and the fine-resolution solutions $\boldsymbol{f}$'s are the FSC solutions. A set of 64 training data $\boldsymbol{g}_i, \boldsymbol{f}_i, i = 1, \ldots, 64$ is generated to construct the PODMM surrogate.

Fig. 2 exhibits the fast decay of the eigenvalues of the training data matrix $\boldsymbol{W}^{\text{PODMM}}$. For the eighth largest eigenvalue, the magnitude already drops below 10^{-13}. Also plotted is the noncaptured fraction of the variance, which is the variance not explained by using only a smaller number of eigenvalues. The noncaptured fraction also decreases very quickly. Note the noncaptured fraction can be used to determine the number of eigenvalues to be used in PODMM by prescribing a threshold value. The number will be the smallest number such that the noncaptured fraction of the total variance is smaller than the threshold. Here we will use only 8 eigenvalues, as they can capture a large enough portion of the total variance.

Fig. 3 displays the analytical, FD and PODMM predicted FSC solutions at $t = \frac{\pi}{8}$ to the Burgers equation corresponding to $\nu = 0.304$. Note that only the FD solution is computed, which is computationally inexpensive, to obtain the PODMM solution. It is demonstrated that although the FD solution does not resolve the transition zone of the shock, the PODMM-predicted FSC fine-resolution solution is in good agreement with the analytical solution. Therefore PODMM can serve as a reliable surrogate in the process of the following IPF for the characterization of the posterior distribution.

We now show the performance of PODMM in IPF. The linear map is used so that one PODMM predicted solution is needed to obtain one posterior sample. To measure the quality of the samples, we estimate the effective sample size by:

$$N_{\text{eff}} = \frac{1}{\sum_{i=1}^{N} (w_i)^2}.$$

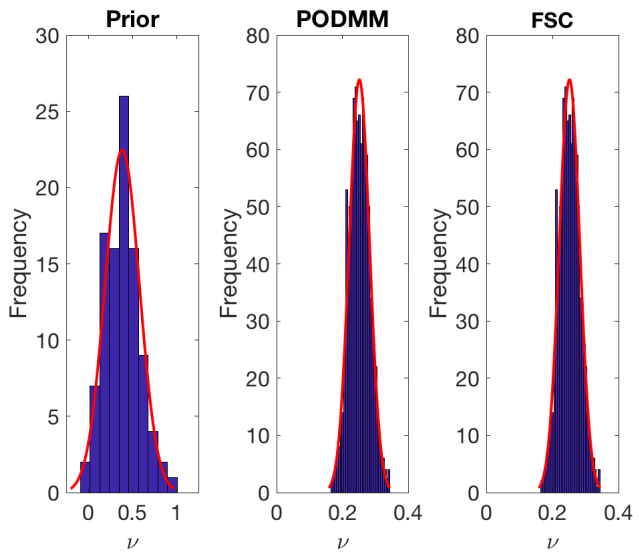

Fig. 4. Prior density and posterior densities estimated by IPF using PODMM and FSC.

The larger the estimated effective sample size is, the less severely the sample quality degenerates. In the ideal case where the weights are uniformly $1/N$, $N_{\text{eff}} = N$, which means all the samples are effective, while in the worst scenario where only one sample is weighted nonzero, $N_{\text{eff}} = 1$. In our case, the ratio of the effect sample size to the total sample size is 0.993, meaning 99.3% of the samples are effective and the IPF is very effective in alleviating sample degeneration.

The posterior distributions of ν generated by the FSC and PODMM-predicted FSC solutions in IPF are shown in Fig. 4. It can be seen that the posterior densities estimated from the samples for the two solutions are nearly indistinguishable (middle and right), indicating PODMM is capable of accurately characterizing the posterior density in IPF. Also shown is the prior normal distribution (left), which is significantly different from the posteriors. Note that the range of the posterior is also smaller compared to that of the prior distribution. To confirm it, the statistical moments of these distributions are given in Table I.

As a result, we have achieved high accuracy in estimating the posterior using the PODMM surrogate model, while saving significant computational time. The computational cost of the fine-resolution FSC solution in our example is about 10 times that of the FD solution, while the cost of PODMM prediction is negligible.

TABLE I
STATISTICAL MOMENTS OF THE PRIOR AND POSTERIOR DENSITIES

	Mean	Variance	Skewness	Kurtosis
Prior	0.4	0.04	0.0	0.0
PODMM posterior	0.2517	9.5645E-4	0.1168	2.8127
FSC posterior	0.2517	9.5645E-4	0.1168	2.8127

IV. CONCLUSIONS

In this work, we implemented an efficient Bayesian parameter inversion technique, the implicit particle filtering, to determine the stochastic viscosity of the Burgers equation. The proportion of the effective sample size from implicit particle filtering can be significantly higher than that from a traditional particle filtering method. We also demonstrated a multi-fidelity modeling technique, the proper orthogonal decomposition mapping method, can construct fine-resolution model solutions from computationally inexpensive coarse-resolution solutions. The resulting surrogate model serves as an efficient and accurate model that can lead to posterior distribution distinguishable from that obtained using the high-fidelity model.

REFERENCES

[1] A. J. Chorin and X. Tu, "Implicit sampling for particle filters," Proc. Natl. Acad. Sci. 106 (41), pp. 17249–17254, 2009.
[2] M. Morzfeld, X. Tu, E. Atkins, and A. Chorin, "A random map implementation of implicit filters," J. Comput. Phys. 231 (4), pp. 2049–2066, 2012.
[3] Y. Liu, G. S. H. Pau, and S. Finsterle, "Implicit sampling combined with reduced order modeling for the inversion of vadose zone hydrological data," Comput. Geosci., vol. 108, pp. 21–32, 2017.
[4] Y. Liu, M.Y. Hussaini, and G. Ökten, "Accurate construction of high dimensional model representation with applications to uncertainty quantification," Reliab. Eng. Syst. Safe., vol. 152, pp. 281–295, 2016.
[5] Y. Zhang, Y. Liu, G. S. H. Pau, S. Oladyshkin, and S. Finsterle, "Evaluation of multiple reduced-order models to enhance confidence in global sensitivity analyses," Int. J. Greenh. Gas Con., vol. 49 (2), pp. 217–226, 2016.
[6] Y. Liu, G. Bisht, Z. M. Subin, W. J. Riley, and G. S. H. Pau, "A hybrid reduced-order model of fine-resolution hydrologic simulations at a polygonal tundra site," Vadose Zone J., vol. 15 (2), 2016.
[7] G. S. H. Pau, C. Shen, W. J. Riley, and Y. Liu, "Accurate and efficient prediction of fine-resolution hydrologic and carbon dynamic simulations from coarse-resolution models," Water Resour. Res., vol. 52, pp. 791–812, 2016.
[8] R. Y. Rubinstein and D. P. Kroese, Simulation and the Monte Carlo Method, 3rd ed.. Wiley, 2016.
[9] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," Trans. Signal. Proc., vol. 50 (2), pp. 174–188, 2002.
[10] C. Canuto, M. Y. Hussaini, A. Quarteroni, and T. A. Zang, Spectral Methods, Springer, 2006.