Bi-fidelity Gradient-Based Approach for Nonlinear Well Logging Inverse Problems

Han Lu, Qiuyang Shen, Jiefu Chen, Xuqing Wu, Xin Fu, Mohammad Khalil, Cosmin Safta, Yueqin Huang

***Abstract*—Solving a non-linear inverse problem is challenging in computational science and engineering. Sampling based methods require a large number of model evaluations; gradient-based methods require fewer model evaluations but only find the local minima. Multifidelity optimization combines the low-fidelity model and the high-fidelity model to achieve both high accuracy and high efficiency. In this paper, we present a bi-fidelity approach to solve non-linear inverse problems. In the bi-fidelity inversion method, the low-fidelity model is used to acquire a good initial guess, and the high-fidelity model is used to locate the global minimum. Combined with a multi-start optimization scheme, the proposed approach significantly increases the possibility of finding the global minimum for non-linear inverse problems with many local minima. The method is tested with two toy problems and then applied to an electromagnetic well logging inverse problem, which is difficult to solve using traditional gradient-based methods. The bi-fidelity method provides promising inversion results and can be easily applied to traditional gradient-based methods.**

***Index Terms*—bi-fidelity, multi-fidelity, gradient-based inversion, inverse problem, polynomial chaos expansion, well logging**

I. Introduction

INVERSE problems are ubiquitous in scientific and engineering fields, such as geophysics, astronomy, computer vision and medical imaging. Most of the time, these inversion problems are ill-posed and non-linear thus can not be well-solved using deterministic inversion methods [1]. The Bayesian inference approaches, such as Markov Chain Monte Carlo (MCMC) sampling [2], make it possible to find the global optimum. The major challenge of sampling-based methods is the computational burden induced by a large number of evaluations of the forward model. When the forward simulation is CPU-intensive, sampling-based methods become computationally prohibitive. The forward model producing outputs that satisfy the accuracy requirement of the task at hand is normally a high-fidelity model. The low-fidelity model or the surrogate model, constructed as an approximation of the high-fidelity model, is cheaper to evaluate but not necessarily accurate. Fast surrogate models are also valuable for sampling-based methods to dramatically reduce computational costs. Many surrogate modeling methods such as polynomial chaos expansion (PCE) [3], Gaussian process regression [4], support vector machines (SVMs) [5], [6] and other simplified models [7], [8] have been developed and successfully employed for many optimization tasks [9], [10]. However, simply replacing the high-fidelity model with the low-fidelity model can lead to biased parameter estimations. A multifidelity optimization method that takes advantage of both low-fidelity and high-fidelity models has drawn great attention because of its capability of achieving desired accuracy and efficiency.

Many previous works implemented the multifidelity model by adapting the low-fidelity model during the inversion process. These methods rely on a large number of low-fidelity evaluations along with a small number of high-fidelity evaluations to build a correction for the low-fidelity model. The adjusted model then serves as a multifidelity approximation to the high-fidelity model. For example, in [11] the polynomial chaos expansion (PCE) surrogate modeling error is modeled by another PCE during the MCMC sampling process. The low-fidelity model can also be corrected by a multiplicative or/and additive term, see [12], [13], [14] for example.

There are also some works using a model management strategy based on filtering, where the high-fidelity model is invoked following the evaluation of a low-fidelity filter. This strategy is mainly employed under the context of statistical inference [15], [16], [17], [18], [19]. However, these methods may have limited applications, because they are sensitive to large modeling errors introduced by the low-fidelity model [11] and the computational budget might still be unacceptable for high-dimensional ($d > 10$) problems [20]. For example, the logging-while-drilling (LWD) electromagnetic (EM) resistivity measurement inverse problem shown in *Section 5* is high-dimensional (usually larger than 10) with strong non-linearity and requires real-time inversion. Our tests demonstrate that a 9-parameter LWD forward model is too hard to approximate at an affordable computational cost in the context of a multi-fidelity statistical inference method.

Different from previous works, in this work we develop a bi-fidelity approach that is designed for gradient-based methods. The motivation for this work is twofold. First, an accurate surrogate is not always available. Second, sampling-based methods are not computationally efficient enough for inverse problems in some real-time applications even with low-fidelity models. To address the above problems, our bi-fidelity approach is summarized as follows:

- The surrogate model is used in a gradient-based optimizer. The result is taken as the initial guess for the next gradient-based optimization process using the high-fidelity model. The two-step approach is less sensitive to the modeling error of the low-fidelity model. The surrogate model smooths out local minima of the high-fidelity model and is much cheaper to construct.
- To reduce the probability of being stuck in local minima, we use a multi-start framework for the deterministic inversion. The multi-start framework is thoroughly described in [21], it is efficient for problems where solutions

can be easily constructed but have multiple local minima.

Though the surrogate modeling methods have been widely explored in recent years, there is no surrogate model that works well for all kinds of applications. In this work, we use PCEs as low-fidelity models of the EM resistivity LWD forward model for the purpose of bi-fidelity inversion. The reason of choosing PCE as the surrogate model is two-fold: (a). PCE is good at capturing global trends in the high-fidelity model; (b) the input dimensionality is too high for many other surrogate models. The original PCE model also suffers from the *curse of dimensionality*, which means that a prohibitively large number of sample points are required when the problem dimensionality is high. To combat this challenge, Sargsyan et al. developed a methodology to construct a sparse PCE surrogate by learning and retaining the most relevant polynomial basis terms with the aid of sparse Bayesian learning [22]. We adopt this approach to construct a sparse PCE surrogate for the EM resistivity LWD forward model for multi-layer earth models. The proposed approach is tested on 2D and 3D Shekel functions as well as multi-layer LWD inverse problems. LWD inverse problems usually have many local optimal solutions due to the non-linearity of the forward model. Given the high dimensionality and time requirement of LWD inverse problems, deterministic inversion methods based on the gradient is still widely used [23], [24], [25], [26]. To our best knowledge, it is the first work that uses multi-fidelity optimization for LWD inverse problems. Testing results show that our approach significantly improves the inversion accuracy with negligible computational overheads.

II. Background

In this section, we first describe the gradient-based optimization methods for inverse problems. Then we introduce a PCE surrogate construction method using sparse learning in a Bayesian framework.

A. Inverse Problems

In this paper, we are interested in the problem of estimating a set of unknown model parameters $\boldsymbol{x} \in \mathbb{R}^n$ from indirect measurements $\boldsymbol{y} \in \mathbb{R}^m$. With a non-linear forward function $f(\boldsymbol{x})$ that yields responses given model parameters, the inverse problem is defined as follows:

$$\underset{\boldsymbol{x}\in\mathbb{R}^n}{\operatorname{argmin}} \|f(\boldsymbol{x}) - \boldsymbol{y}\|_2^2 \tag{1}$$

This is an unconstrained non-linear least-square optimization problem and many numerical methods are developed to solve such a problem by iteratively updating the model parameters to minimize the misfit between the observation $\boldsymbol{y}$ and the model prediction $f(\boldsymbol{x})$. To determine the direction of the model updating, derivative of the objective function needs to be calculated during each iteration. For example, given the objective function $F(\boldsymbol{x}) = \|f(\boldsymbol{x}) - \boldsymbol{y}\|_2^2$, at the i-th iteration the gradient-descent algorithm [27] updates the model parameters $\boldsymbol{x}$ to the direction of the negative gradient of F at $\boldsymbol{x}_{i-1}$, i.e. $\nabla F(\boldsymbol{x}_{i-1})$, as follows:

$$\boldsymbol{x}_i = \boldsymbol{x}_{i-1} - \gamma \nabla F(\boldsymbol{x}_{i-1}) \tag{2}$$

where the positive step size γ controls the converge speed and the accuracy of the algorithm. For multi-variate functions, the derivative $\nabla F(\boldsymbol{x}_{i-1})$ is the Jacobian matrix $\boldsymbol{J}$ of F at $\boldsymbol{x}_{i-1}$. With certain assumptions on the function F and particular choices of γ, the algorithm is guaranteed to converge to a local minimum. In case that the function F is twice-differentiable, one can use Newton's method [28] to minimize the objective function more efficiently by the following updating rule:

$$\boldsymbol{x}_i = \boldsymbol{x}_{i-1} - (\nabla^2 F(\boldsymbol{x}_{i-1}))^{-1} \nabla F(\boldsymbol{x}_{i-1}) \tag{3}$$

where $\nabla^2 F(\boldsymbol{x}_{i-1})$ is the Hessian matrix of F at $\boldsymbol{x}_{i-1}$ and automatically decides the step size. However, if the second-order derivative is computationally expensive, it will be less efficient to compute it at each iteration. As a modification of Newton's method, the Gauss-Newton algorithm replaces the second derivative with an approximation:

$$\nabla^2 F(\boldsymbol{x}_i) \approx \boldsymbol{J}(\boldsymbol{x}_i)^T \boldsymbol{J}(\boldsymbol{x}_i) \tag{4}$$

The Levenberg-Marquadt algorithm (LMA) [29], [30], also known as the damped least-squares (DLS) method, is a hybrid approach. Comparing to the Gauss-Newton algorithm, a damping term is added to the updating rule:

$$\boldsymbol{x}_i = \boldsymbol{x}_{i-1} - (\boldsymbol{J}(\boldsymbol{x}_{i-1})^T \boldsymbol{J}(\boldsymbol{x}_{i-1})) + \lambda \boldsymbol{I})^{-1} \nabla F(\boldsymbol{x}_{i-1}) \tag{5}$$

where small values of λ result in a Gauss-Newton update and large values of λ result in a gradient descent update. The damping factor is adjusted during each iteration, as a result, the solution typically reaches the local minimum faster. In this work, we use LMA to solve the inverse problems.

B. Polynomial Chaos Expansion Surrogate

In this section, we describe the polynomial chaos approximation to the forward model $y = f(\boldsymbol{x})$, where $\boldsymbol{x} \in \mathbb{R}^n$ is an n-dimensional input vector and y is a scalar output. In polynomial chaos (PC) theory [31], [32], both input parameters and the output of interest are represented as a series of orthogonal polynomials $\Psi_k(\boldsymbol{\xi})$ of standard, i.i.d. random variables $\boldsymbol{\xi} \in \mathbb{R}^{\tilde{n}}$. As such, the input parameters can be written as follows:

$$x_i \cong \sum_{k=0}^{K_{in}-1} x_{i,k} \Psi_k(\boldsymbol{\xi}) \tag{6}$$

where $x_{i,k}$ for $i = 1, 2, ..., n, k = 0, 1, ..., K_{in} - 1$ are the expansion coefficients corresponding to the polynomial chaos expansion for the input parameter $\boldsymbol{x}$. K_{in} is the number of basis terms in the input PC expansions which can be fixed at 2 so that $\boldsymbol{x}$ and $\boldsymbol{\xi}$ has a linear relationship. We go through this exercise to construct a probabilistically consistent functional form for the model outputs dependency on the input parameters. The output is written as

$$y \cong \sum_{k=0}^{K-1} c_k \Psi_k(\boldsymbol{\xi}) \tag{7}$$

where c_k for $k = 0, 1, ..., K - 1$ are the PCE coefficients for the output y, the value of K is chosen according to the modeling accuracy requirements.

The type of polynomials $\Psi_k(\boldsymbol{\xi})$ is chosen to keep consistency with the distribution of $\boldsymbol{\xi}$. For example, Hermite polynomials are used for normally distributed random variables and Legendre polynomials are used for uniformly distributed random variables. For the surrogate construction purpose, one can use the first-order polynomial of ξ_i to represent the input parameters x_i. Subsequently, an available data set of input-output pairs $\{\boldsymbol{x}_i, y_i\}_{i=1}^N$ can be linearly transferred to $\mathcal{D} = \{\boldsymbol{\xi}_i, y_i\}_{i=1}^N$ and what remains is determining c_k in eq. (7) with the given data.

There exist intrusive and non-intrusive methods [33] for the calculation of the polynomial coefficients c_k. The intrusive approach requires a reformulation of the solution method and rewriting of the code and it is not always practical. Alternatively, the non-intrusive method does not require an explicit representation of the forward model but treats it as a black-box. Non-intrusive methods attempt to solve the following explicit problem:

$$\sum_{l=0}^{K-1} c_k \Psi_k(\boldsymbol{\xi}) \cong f(\sum_{k=0}^{K_{in}-1} x_{i,k} \Psi_k(\boldsymbol{\xi})) \tag{8}$$

which can be solved using non-intrusive spectral projection (NISP) [34] or regression-based methods [35] to compute the coefficients c_k. NISP provide the coefficients as

$$c_k = \langle y, \psi_k \rangle / \langle \psi_k, \psi_k \rangle \tag{9}$$

where $\langle X(\boldsymbol{\xi}), Y(\boldsymbol{\xi}) \rangle$ is the inner product of two functions X and Y with respect to the probability density function of $\boldsymbol{\xi}$. $\langle \psi_k, \psi_k \rangle$ is in practice known exactly, $\langle y, \psi_k \rangle$ can be calculated by numerical integration

$$\langle y, \psi_k \rangle = \int y \psi_k(\boldsymbol{\xi}) p(\boldsymbol{\xi}) d\boldsymbol{\xi} \tag{10}$$

Deterministic (Gauss quadrature) or random (Monte Carlo) sampling can be used to compute the integration. However, for high-dimensional problems, both methods require a large number of simulation runs even with sparse sampling techniques, which renders those approaches impractical for the inversion problem under consideration.

Alternatively, eq. (7) can be treated as a regression model, the regression coefficients c_k can then be computed using regression methods. In the following section, we will demonstrate how to solve this regression problem efficiently with Bayesian compressive sensing (BCS).

C. Bayesian Compressive Sensing for Polynomial Regression

From the Bayesian point of view, the solution to the problem of determining the coefficients c_k in eq. (7) is a posterior probability density function $q(\boldsymbol{c})$:

$$q(\boldsymbol{c}) \propto L_{\mathcal{D}}(\boldsymbol{c}) p(\boldsymbol{c}) \tag{11}$$

where $L_{\mathcal{D}}(\boldsymbol{c})$ is the likelihood of $\boldsymbol{c}$, a measure of a goodness-of-fit of the corresponding surrogate with respect to the given data $\mathcal{D}$, and $p(\boldsymbol{c})$ is the prior distribution of $\boldsymbol{c}$. Given a zero-mean normal distributed noise model ϵ with standard deviation σ, we can write the likelihood:

$$L_{\mathcal{D}}(\boldsymbol{c}) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp(-\sum_{i=1}^{N} \frac{(y_i - y_c(\boldsymbol{\xi}_i)^2}{2\sigma^2}) \tag{12}$$

This problem becomes intractable for high-dimensional problems since the number of unknown coefficients will grow rapidly with increasing dimensionality. In many applications, most of the basis functions in PCE have negligible impact, i.e. the vector $\boldsymbol{c}$ is sparse. It is efficient and reasonable to only compute the most significant terms of the PCE, both in the construction and evaluation of the PCE surrogates. To this end, in [22] the authors proposed to use Bayesian compressive sensing (BCS) to find a sparse representation of $\boldsymbol{c}$ given available data [36], [37]. The key for inferring a sparse PCE is to impose a prior distribution on $\boldsymbol{c}$ that induces sparsity. A commonly used sparsity-inducing prior is the Laplace prior

$$p(\boldsymbol{c}) = (\frac{\alpha}{2})^{k+1} \exp(-\alpha \sum_{k=0}^{K} \|\boldsymbol{c}\|) \tag{13}$$

The vector $\boldsymbol{c}$ that maximize the posterior $q(\boldsymbol{c})$ in eq. (7) coincides with the solution of the classical compressive sensing problem

$$\arg\max_{\boldsymbol{c}} (\log L_{\mathcal{D}}(\boldsymbol{c}) - \alpha \|\boldsymbol{c}\|_1) \tag{14}$$

where the regularization term $\alpha\|\boldsymbol{c}\|_1$ corresponds to the sparsity-inducing prior distribution. The positive parameter α is a user-defined value that controls the level of sparsity. However, the Laplace distribution is not conjugate to the Gaussian likelihood and thus does not allow a tractable Bayesian analysis. This issue was addressed in sparse Bayesian learning, particularly with the relevance vector machine (RVM) [38]. Instead of directly using the Laplace prior, a hierarchical prior distribution is constructed with a Gaussian prior distribution on $\boldsymbol{c}$

$$p(c_k | s_k^2) = \frac{1}{\sqrt{2\pi s_k^2}} \exp(-\frac{c_k^2}{2s_k^2}) \tag{15}$$

and a gamma prior to the hyperparameter s_k^2

$$p(s_k^2 | \alpha^2) = \frac{\alpha^2}{2} \exp(-\frac{\alpha^2 s_k^2}{2}) \tag{16}$$

with a resulting (marginalized) Laplace prior density

$$p(\boldsymbol{c}|\alpha^2) = \int_0^\infty \prod_{k=0}^{K-1} p(c_k|s_k^2) p(s_k^2|\alpha^2) ds_k^2 = \prod_{k=0}^{K-1} \frac{\alpha}{2} e^{-\alpha|c_k|} \tag{17}$$

This procedure has been implemented in the Bayesian LASSO method [39]. For details of the implementation, see [22], [37], [40]. The task of finding PCE coefficients $\boldsymbol{c}$ now has became an optimization problem that finds hyperparameters σ^2, $\boldsymbol{s}^2$ and α that maximizes the *evidence* or the integrated likelihood

$$\begin{aligned} E(\sigma^2, \boldsymbol{s}^2, \alpha) &= \int_{\mathbb{R}^K} L_{\mathcal{D}}(\boldsymbol{c}; \sigma^2) p(\boldsymbol{c}|\boldsymbol{s}^2) p(\boldsymbol{s}^2|\alpha) p(\alpha) p(\sigma^2) \mathrm{d}\boldsymbol{c} \\ &\propto p(\alpha) p(\sigma^2) p(\boldsymbol{s}^2|\alpha) \sigma^{-1} |\boldsymbol{C}|^{-\frac{1}{2}} \exp(-\frac{1}{2\sigma^2} y^T \boldsymbol{C}^{-1} y) \end{aligned} \tag{18}$$

where $\boldsymbol{C} = \boldsymbol{I} + \boldsymbol{\Psi} \boldsymbol{S}^{-1} \boldsymbol{\Psi}$. Here $\boldsymbol{\Psi}$ is an $N \times K$ projection matrix with entries $\boldsymbol{\Psi}_{ik} = \Psi_k(\boldsymbol{\xi}_i)$ and $\boldsymbol{S} = \mathrm{diag}(\sigma/s_0^2, ..., \sigma/s_{K-1}^2)$. In practice it turns out that for many basis terms, the inverse variance $\frac{1}{s_k^2}$ that maximizes eq. (18) grows indefinitely, i.e., $s_k^2 \to 0$. These terms will be purged from the basis set.

Let us denote by K' the number of retained basis functions, and reindex them using k, for $k = 0, 1, ..., K'-1$. One obtains a Gaussian posterior distribution for $\boldsymbol{c}$ with mean and variance

$$\begin{aligned} \boldsymbol{\mu} &= \sigma^{-2}\boldsymbol{\Sigma}\boldsymbol{\Psi}^T y \\ \boldsymbol{\Sigma} &= \sigma^2(\boldsymbol{\Psi}^T\boldsymbol{\Psi} + \boldsymbol{S})^{-1} \end{aligned} \tag{19}$$

where $\boldsymbol{\Psi}$ is an $N \times K'$ projection matrix with entries $\boldsymbol{\Psi}_{ik} = \Psi_k(\boldsymbol{\xi}_i)$ and $\boldsymbol{S} = \mathrm{diag}(\sigma/s_0^2, ..., \sigma/s_{K'-1}^2)$. Finally, $\boldsymbol{\mu}$ is used as coefficients to form a sparse PCE surrogate

$$y \simeq \sum_{i=0}^{K'-1} \mu_i \boldsymbol{\Psi}_k(\boldsymbol{\xi}) \tag{20}$$

III. Bi-fidelity Approach for Gradient-Based Optimization

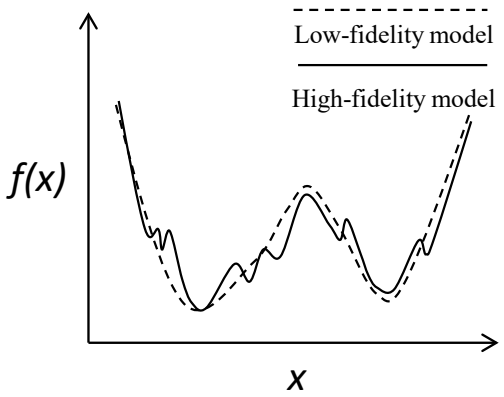


Fig. 1: An example of surrogate model that have large modeling error at some points but are helpful for global search.

The major drawback of gradient-based methods is that they are inherently local methods. However, as pointed out in [41], for differentiable problems, one should first consider using multi-start gradient-based optimization because of their ease of implementation as well as their advantages in using derivative. With many randomly distributed initial guesses, it is likely that some models are close to the global minimum and will finally converge to the global minimum. For high-dimensional problems, the required number of initial guesses grows exponentially to obtain the global minimum, which has put big pressure on the computational resources. Based on the empirical observation, a low-order surrogate with large modeling error may not mislead the global search [42] since it can smooth out the high-fidelity model as illustrated in Figure 1. In this paper, we propose a bi-fidelity approach that utilizes a sparse PCE surrogate to help find initial guesses that may converge to the global minimum, then uses the high-fidelity model for an accurate inversion. PCE is a spectral expansion approach that represents quantities-of-interest as a series of orthogonal polynomials of standard, i.i.d. random variables and has already been introduced in *Section 2.2*. We use LMA for the inversion task in this work due to its good performance on non-smooth objective functions, one can switch to any other gradient-based algorithm according to the behavior of the objective function. The workflow of the proposed method is shown in Figure 2.

Before performing the bi-fidelity inversion, the low-fidelity model should be constructed based on the high-fidelity model. As shown in Figure 2, the surrogate replaces the high-fidelity model in the inversion process at the first stage. The surrogate used in step 1 is to capture the global trends in the response surface and the solution at this stage is located closer to the global optimum. At step 2, the LMA inversion is performed with the high-fidelity model initialized with the solutions from the previous stage. Finally, the solution with the smallest data misfit will be selected as the final result. Comparing to the single-fidelity inversion that only uses the high-fidelity model, this bi-fidelity optimization strategy has a better chance of escaping from local minima and reaching the global minimum. Multi-start optimization scheme is used in this approach, the number of required initial models is greatly reduced with the help of the low-fidelity model.

IV. Numerical Tests

For the purpose of visualization, we test the performance of the proposed approach on 2D and 3D Shekel functions. Shekel function is a multi-dimensional, multi-modal, continuous function and is commonly used as a test function for optimization algorithms [43]. As shown in Figure 3(a), there are 10 local minima in the Shekel function. The global minimum has a value of -11.03 and is located at (4, 4). We construct PCE surrogates as the low-fidelity model of the shekel functions for the purpose of bi-fidelity inversion. The 2D shekel function is cheap and only contains 2 input variables, so we construct full PCE surrogate models of order 30 and 15 using the non-intrusive method introduced in II-B. The samples used for PCE construction are generated using the uniform-Legendre quadrature rule [44]. The surrogate models are shown in Figure 3(b) and Figure 3(c) respectively. When the PCE order is increased, more local minimum in the high-fidelity model is captured. Though the approximation error is large at some points for both surrogate models, the response surface is smoothed.

We first perform LMA inversion for the 2D Shekel problem with only the high-fidelity model, which will be referred to as the single-fidelity inversion. 10 different initial guesses are generated by the Latin hypercube sampling (LHS) algorithm and both input parameters are within the range [-15,15]. The inversion results are shown in Figure 4(a). Only 1 model converges to the global minimum and the rests converge to the other 4 local minima. Figure 4(b) shows the inversion results of using the proposed bi-fidelity method with a 30th order PCE surrogate model. Compare to the high-fidelity case, more initial models have converged to the global minimum. However, there are still some initial models trapped in local minima. The result of bi-fidelity with a PCE surrogate of 15th order is shown in Figure 4(c), where all initial models converge to the global minimum. This example demonstrates that the fidelity of the surrogate model can have great effect to the bi-fidelity inversion result. The best choice of low-fidelity model can be different for different applications. For PCE surrogates,

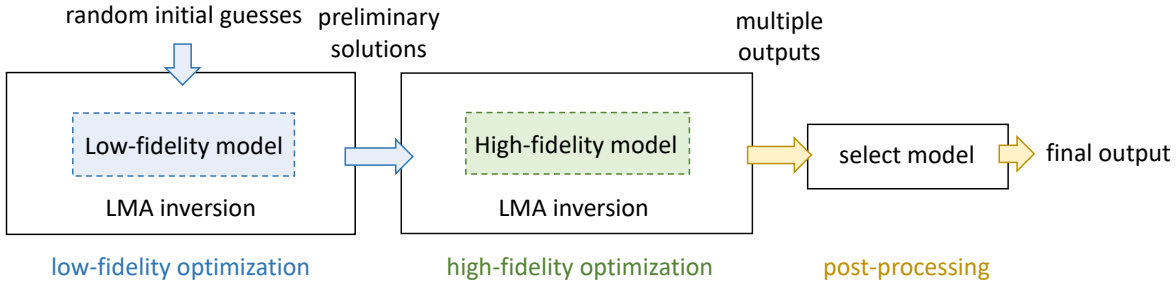
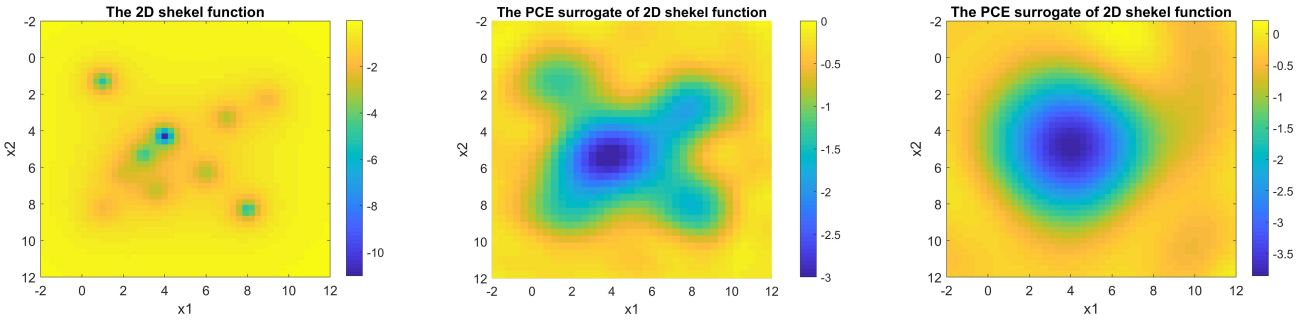


Fig. 2: The bi-fidelity LMA inversion workflow.



(a) Original response (b) PCE order = 30 (c) PCE order = 15

Fig. 3: (a) The 2D Shekel function with 10 local minima. (b) The 30th order PCE surrogate of the 2D Shekel function. (c) The 15th order PCE surrogate of the 2D Shekel function.

we suggest one to start with a surrogate model with higher fidelity, then gradually reduce the fidelity to find a proper surrogate model for the bi-fidelity inversion.

Next, we increase the number of initial models to 100 and the results are shown in Figure 5. In the single-fidelity test results shown in Figure 5(a), only 7% of the models converge to the global minimum and many models converge to the local minima located at (3, 6), (5, 3), (6, 6) and (7, 3.5). The bi-fidelity inversion result with a 30th order PCE surrogate is shown in Figure 5(b). In this case, 25% of the models finally converge to the global minimum. When the 15th order PCE surrogate is used for the bi-fidelity inversion, all of the models converge to the global minimum, as shown in Figure 5(c). The examples shown in Figure 4 and Figure 5 demonstrate that a proper low-fidelity surrogate model can help to avoid some local minimum in the bi-fidelity inversion method.

Given the observations from the 2D Shekel example, we construct full PCE surrogate of order 15 for the 3D Shekel function. The 3D Shekel problem becomes more difficult to solve since the dimensionality is increased, thus more initial models are needed to cover the parameter space. Figure 6 shows the model response of the 3D Shekel function along the second and the third coordinate when fixing the first coordinate at 4. With 100 initial models generated by the LHS algorithm, only 2 models converge to the global minimum when only using the high-fidelity model. However, with the help of the low-fidelity model, 83 models converge to the global minimum. As shown in Figure 7, most of the local minima are smoothed out by the low-fidelity model and only the one located at (5, 3, 5) is left.

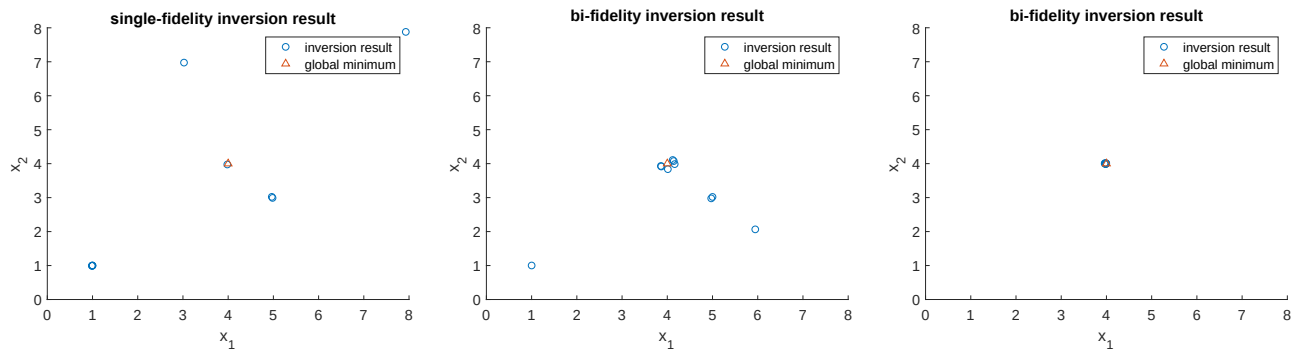
V. Application to the LWD Inverse Problem

In this section, we demonstrate the bi-fidelity gradient-based inversion approach by solving resistivity LWD inverse problems where the objective is to infer the earth model parameters (e.g. resistivity of each formation layer, distances from the logging tool to formation interfaces) based on downhole LWD measurements. The recent development of azimuthal EM resistivity LWD tool has greatly extended the depth of investigation, thus increases the number of unknown parameters of the LWD inverse problem. Meanwhile, EM responses are highly non-linear due to multiple transmissions and reflections between formation interfaces. Being inherently high-dimensional and ill-posed, one needs to perform a large number of independent gradient-based optimization with different initial models to obtain an acceptable result. The problem can also be solved by statistical inference methods [45], [46]. However, the computation cost is too high to be used in real-time.

The remainder of this section describes the implementation details for the bi-fidelity LWD inversion, which includes the forward model of the ultra-deep EM resistivity LWD tool, the sparse PCE surrogate construction, and the bi-fidelity inversion process.

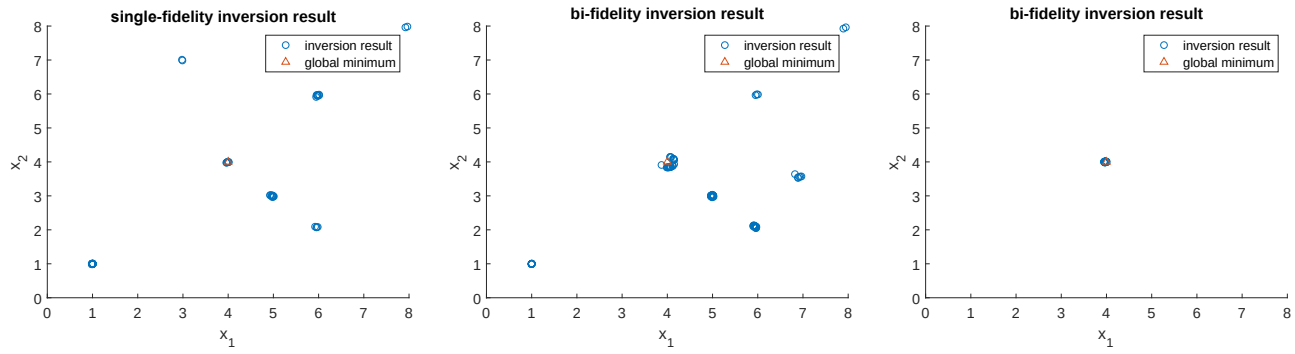
A. The High-fidelity and Low-fidelity Models

In the LWD inversion application, the high-fidelity model is the forward model that simulates the responses of az-



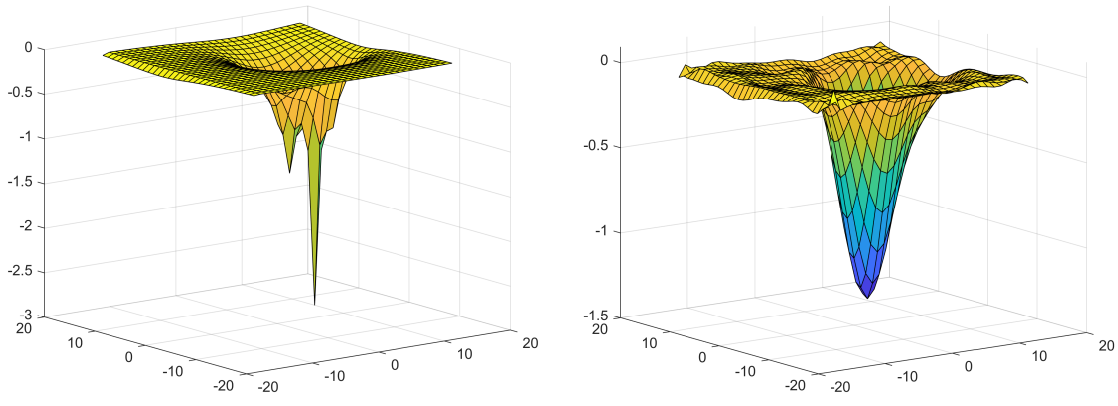
(a) Inversion with the high-fidelity model. (b) Bi-fidelity inversion, PCE order = 30. (c) Bi-fidelity inversion, PCE order = 15.

Fig. 4: Inversion results of the 2D Shekel function with 10 initial models.



(a) Inversion with the high-fidelity model. (b) Bi-fidelity inversion, PCE order = 30. (c) Bi-fidelity inversion, PCE order = 15

Fig. 5: Inversion results of the 2D Shekel function with 100 initial models.



(a) (b)

Fig. 6: Model responses when fixing the first coordinate of (a) the 3D Shekel model and (b) the PCE surrogate of the 3D Shekel model.

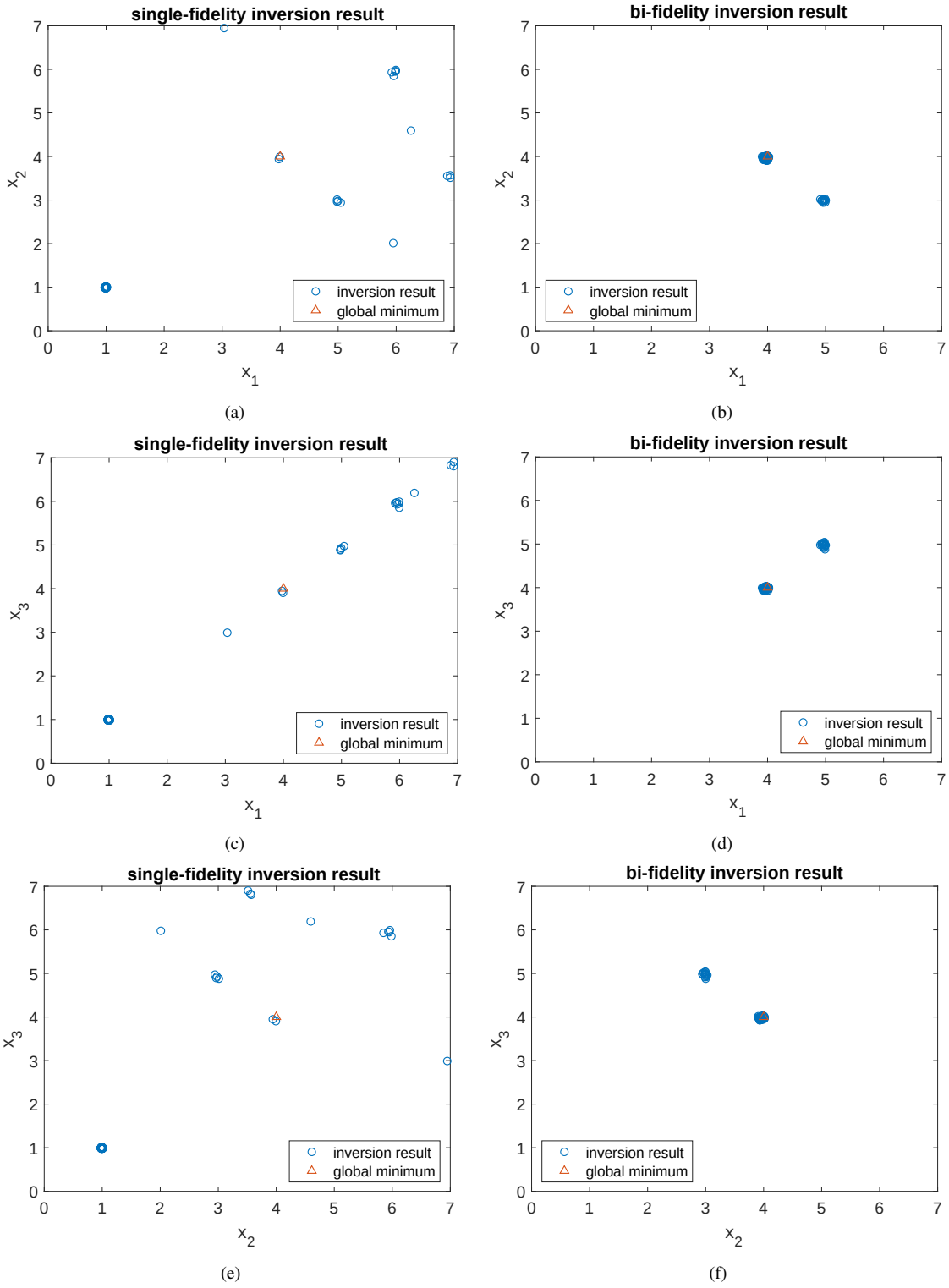


Fig. 7: Inversion results of the 3D Shekel function with 100 initial models. (a-b) Inversion results along the first and the second coordinate. (c-d) Inversion results along the second and the third coordinate. (e-f) Inversion results along the second and the third coordinate.

imuthal EM LWD tools in 1D earth models. Given an n-layer earth model, the input consists of n resistivity values and $n-1$ depth-to-the-boundaries. Multiple transmitter-receiver pairs are set at different locations of the tool working at multiple frequencies (2 kHz, 6 kHz, and 24 kHz used here). In this paper, we synthesize an azimuthal EM LWD tool generating 72 measurements at each logging point. We build a surrogate model for each of the 72 signals, i.e. the low-fidelity model consists of 72 independent single-output PCEs. 7-layer earth models are assumed in the following numerical studies. With 13 input parameters, the full representation of PCE is unfeasible at an acceptable computational cost, for example, to construct a 5-order PCE surrogate model one needs to generate $(5+1)^{13} \cong 1.3e10$ quadrature points. In this work, the Bayesian sparse learning is used to construct sparse PCEs for the high-dimensional LWD inverse problems.

The resistivity value varies from 0.1 $\Omega \cdot m$ to 300 $\Omega \cdot m$ and is first transformed into the logarithmic scale then re-scale to [0,1] for the surrogate modeling. The outputs are linearly re-scaled to [0,1].

Considering the detection scope of the ultra-deep directional logging tool, we tested earth models with 7 layers in this investigation. As the drilling tool penetrates bed boundaries, the model response can be highly nonlinear and difficult to capture with a global PCE approximation. To avoid this situation, we split the training data into 7 subsets, corresponding to 7 scenarios (i.e. the tool is located in the first layer, the tool in located in the second layer, and so forth). In the following, we describe the steps to build a piece-wise PCE surrogate for the LWD forward model.

- Split the training data set into non-overlapping subsets $\mathcal{D}_i, i = 1, 2, ..., 7$. $\boldsymbol{x} \in \mathcal{D}_i$ means that the tool is in the i-th layer.
- Construct sparse PCE $g_{ij}(\boldsymbol{x})$ using each data set for each output individually ($i = 1, 2, ..., 7$, $j = 1, 2, ..., 72$).
- Declare piece-wise PCE surrogates

$$\boldsymbol{g}_i(\boldsymbol{x}) = (g_{i1}(\boldsymbol{x}), g_{i2}(\boldsymbol{x}), ..., g_{i72}(\boldsymbol{x})) \quad \text{if} \quad \boldsymbol{x} \in \mathcal{D}_i (i = 1, 2, ..., 7) \tag{21}$$

The data in each subset consists of 5×10^4 samples generated by the LHS algorithm. The surrogate construction takes 8 hours in total using 6 64-bit Intel(R) Xeon(R) CPU E5-2650 v4 @ 2.20GHz processors. The high-fidelity model takes 4×10^{-2} seconds for each evaluation and the corresponding surrogate takes only 3×10^{-3} seconds.

B. The LWD Inverse Problem

TABLE I: Layer resistivities and thicknesses of the synthetic earth model.

layer #	1	2	3	4	5	6	7
resistivity($\Omega \cdot m$)	1	20	2	100	3	50	3
thickness(ft)	-	10	7	40	15	20	-

The synthetic earth model has 7 layers and the parameter values are shown in Table 1. Consider a drilling process where the tool keeps drilling down with a dip angle of 82 degrees and travels through 6 boundaries. The LWD data is collected every 10 ft and the total working region extends to 800 ft horizontally. The problem consists of 80 continuous 1D LWD inverse problems and we solve them independently. Figure 8 shows the structure of the earth model as well as the drilling trajectory represented by the black dot line.

We perform LMA inversion for each of the 80 LWD inverse problems using single-fidelity and bi-fidelity inversion methods, the initial models are randomly generated by LHS and the parameter range is the same as the range of the surrogate training data. For the sake of fairness, the initial models for the single-fidelity and bi-fidelity inversion approach are the same, and the results are shown in Figure 9. In the single-fidelity inversion, the resistivity and the locations of layer boundaries near the borehole can be well inferred in most of the cases. However, the parameters of layers away from the borehole can not be accurately estimated due to the existence of local minima. The reason is twofold: first, signals are reflected by many boundaries so there may exist multiple solutions that can cause similar responses; second, signals become weak due to attenuation. As shown in Figure 9 (left column), with the number of initial models increasing from 50 to 500, the single-fidelity inversion results are improved because more models are initialized around the global minimum. In the bi-fidelity inversion, tool responses are smoothed out by the low-fidelity model so that many local minima are skipped. The bi-fidelity inversion with only 50 initial models performs better than the single-fidelity inversion with 500 initial models. Though there is an added one-time cost for the surrogate construction before inversion, the surrogate models can be used for similar inverse problems in the future.

To further examine the feasibility of the proposed method, we perform single-fidelity and bi-fidelity inversion with LWD data contaminated by synthetic zero-mean Gaussian noises. The noise standard deviation $\sigma_{ps} = 0.375$ and $\sigma_{att} = 0.0625$ are used for phase-shift and attenuation measurements respectively. Figure 10 shows 3 example measurements with and without noise. We run the inversion algorithms using the same configurations as those for the clean data example. Reconstructed 2D earth models are shown in Figure 11. Compare to the clean data case, both the results for the single-fidelity and the bi-fidelity inversion are less accurate due to the existence of noises. However, the bi-fidelity inversion algorithm still exhibits good capability of inverting noisy data. Similar to the case without noise, the bi-fidelity inversion result with only 50 initial values achieves similar results as the single-fidelity inversion with 500 initial values. These observations are also reflected by the inverted data misfit defined as follows:

To better quantify the inversion accuracy, we compare data misfits of the two approaches which is defined as follows:

$$e = \frac{1}{80} \sum_{i=1}^{80} \frac{\|f(\boldsymbol{x}^i_{real}) - f(\boldsymbol{x}^i_{inv}))\|_2}{\|f(\boldsymbol{x}^i_{real})\|_2} \tag{22}$$

Figure 12 shows the average normalized data misfit of the 80 inverse problems by the single-fidelity and bi-fidelity approach for the clean data and noisy data tests. For both the clean data and noisy data inversion, the bi-fidelity inversion

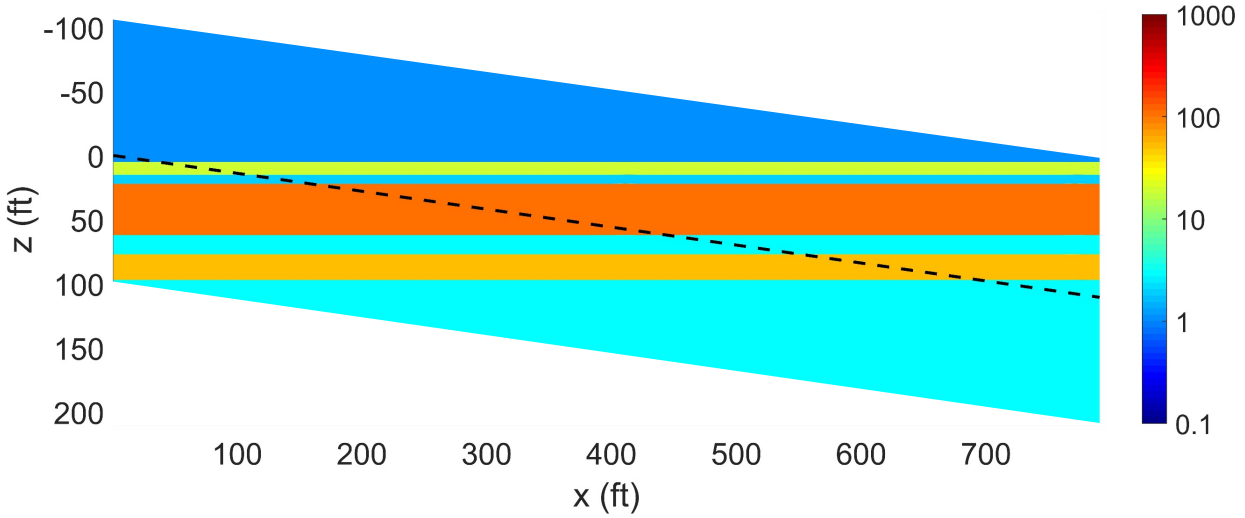
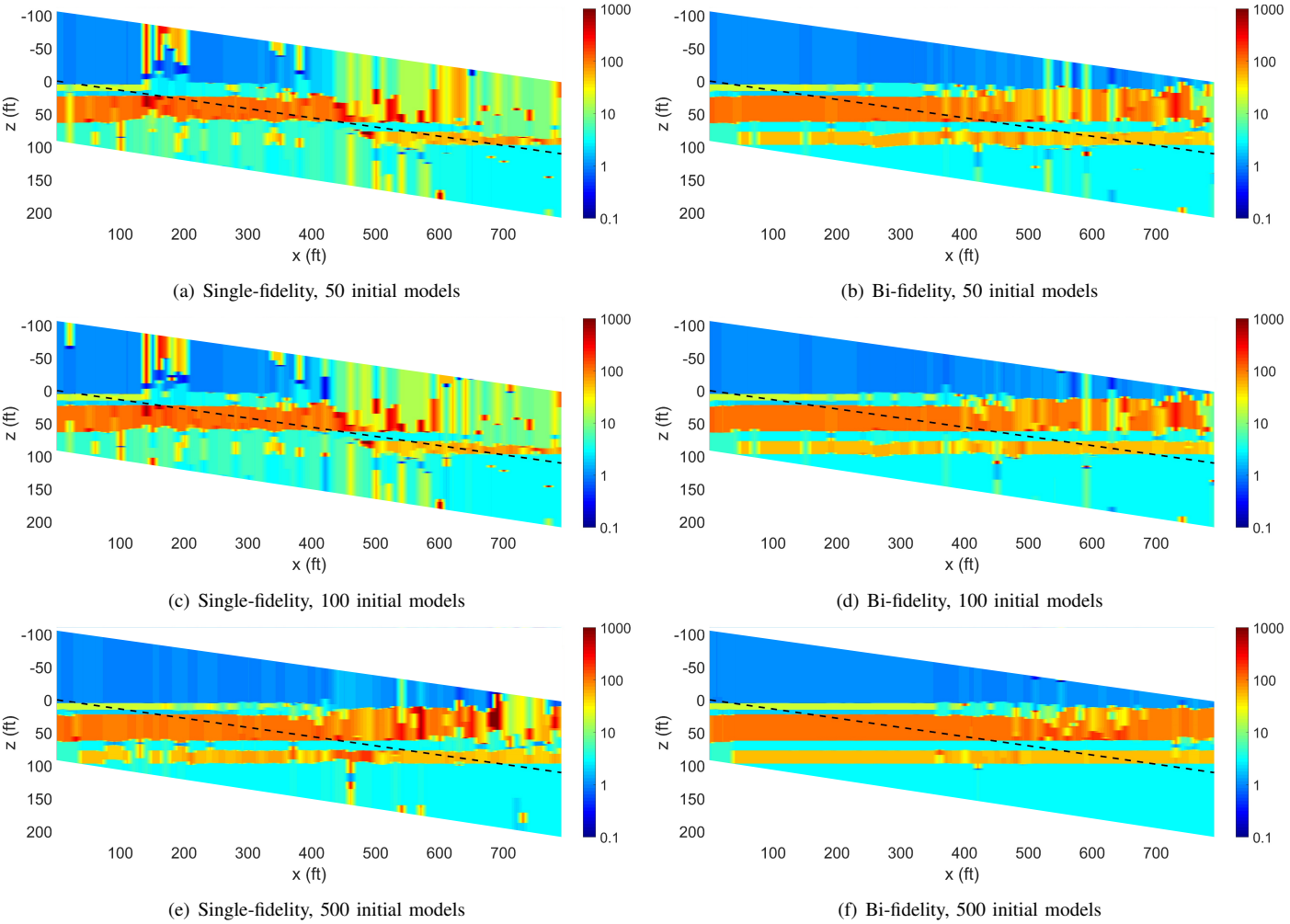


Fig. 8: The tested 7-layer model. The color bar represents the resistivity value of each layer.



(a) Single-fidelity, 50 initial models

(b) Bi-fidelity, 50 initial models

(c) Single-fidelity, 100 initial models

(d) Bi-fidelity, 100 initial models

(e) Single-fidelity, 500 initial models

(f) Bi-fidelity, 500 initial models

Fig. 9: Curtain plot of the inversion results of the single-fidelity inversion and the bi-fidelity inversion with different number of initial models. The number of initial models are increased from 50 to 500.

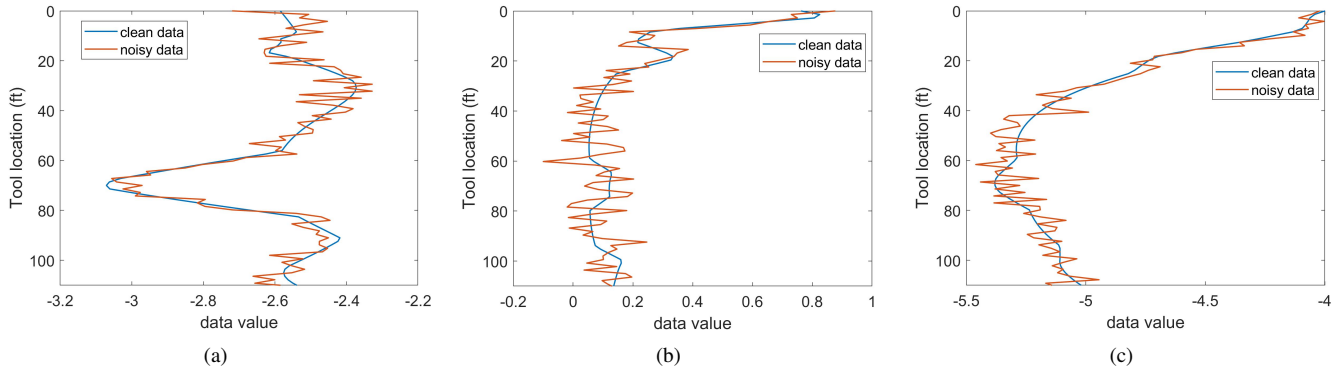


Fig. 10: Three example measurements with and without noise.

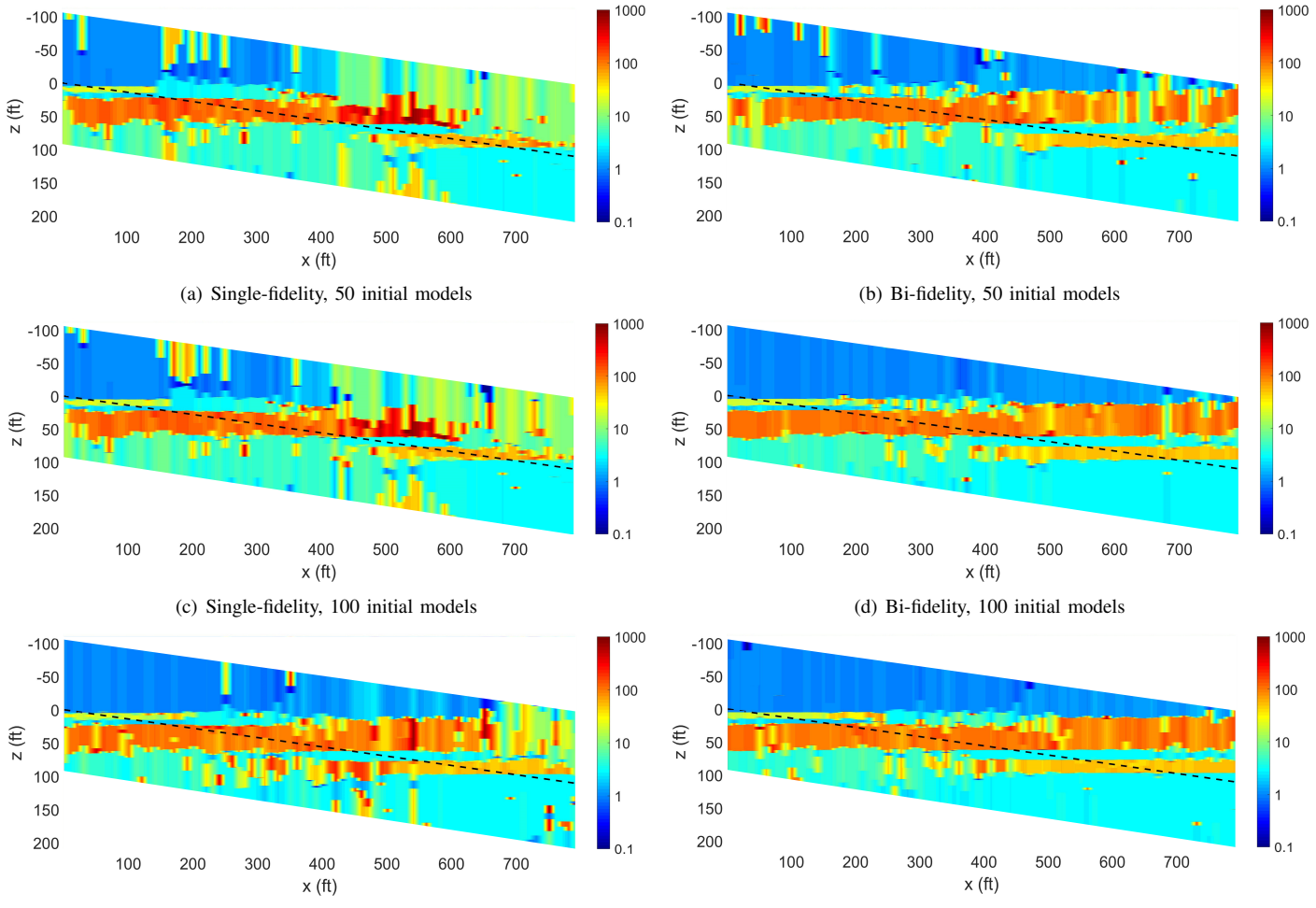


Fig. 11: Curtain plot of the inversion results of the single-fidelity inversion and the bi-fidelity inversion where the measurements are contaminated by synthetic noise. The number of initial models are increased from 50 to 500.

method has a much lower data misfit than the single-fidelity approach. The bi-fidelity approach with only 50 initial models obtains a similar inversion accuracy as that from the single-fidelity method with 1000 initial models, which means that the bi-fidelity method only requires 5% of the computational resources to achieve the similar accuracy.

In the single-fidelity approach, the high-fidelity model was evaluated for 417 times on average for each optimization; in the bi-fidelity approach, the low-fidelity and high-fidelity model was evaluated for 35 and 370 times respectively. The average time cost of the two approaches is shown in Figure 13(a), where the average run time of the bi-fidelity approach with an extra low-fidelity inversion step is even less than the single-fidelity approach. Figure 13(b) shows that in the bi-fidelity inversion, the low-fidelity model evaluation only takes 4% of the total run time, which means the bi-fidelity approach improves the inversion accuracy with negligible computational overhead.

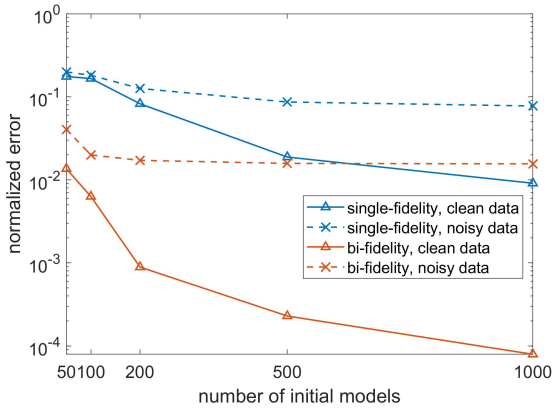


Fig. 12: The average normalized errors of the single-fidelity and the bi-fidelity inversion methods for both clean data and noisy data.

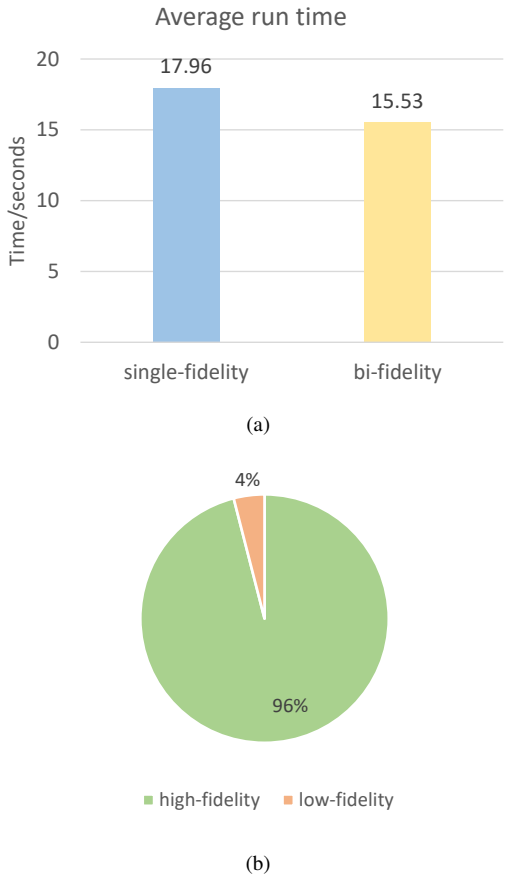


Fig. 13: (a) The average run time for each LMA inversion. (b) The percentage of the time used for the high-fidelity evaluation and the low-fidelity evaluation in the bi-fidelity inversion.

VI. Conclusions

We have presented a bi-fidelity gradient-based inversion method aiming at solving high-dimensional and non-linear inverse problems. The key idea is to use a low-fidelity model that smooths out the forward model responses to find initial models that are close to the global minimum, the high-fidelity model is then used to refine the inversion result.

We first apply the proposed method to the 2D and 3D Shekel optimization problems for the convenience of visualization. The results show that most of the local minima are avoided by the PCE surrogate. We then demonstrate the performance of the bi-fidelity approach with a 13-parameter LWD inverse problem. Comparing to the single-fidelity gradient-based inversion, the proposed method significantly improves inversion accuracy. It can be easily applied to other applications that require gradient-based inversion.

Acknowledgment

This material is based upon work supported by the U.S. Department of Energy, Office of Science, and Office go Advanced Science Computing Research, under Award Numbers DE-SC0017033. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

References

[1] D. H. Rothman, Nonlinear inversion, statistical mechanics, and residual statics estimation, Geophysics 50 (12) (1985) 2784–2796.

[2] W. R. Gilks, S. Richardson, D. Spiegelhalter, Markov chain Monte Carlo in practice, Chapman and Hall/CRC, 1995.

[3] D. Xiu, G. E. Karniadakis, The wiener–askey polynomial chaos for stochastic differential equations, SIAM journal on scientific computing 24 (2) (2002) 619–644.

[4] I. Kaymaz, Application of kriging method to structural reliability problems, Structural Safety 27 (2) (2005) 133–151.

[5] C. Cortes, V. Vapnik, Support-vector networks, Machine learning 20 (3) (1995) 273–297.

[6] A. J. Majda, B. Gershgorin, Quantifying uncertainty in climate change science through empirical information theory, Proceedings of the National Academy of Sciences 107 (34) (2010) 14958–14963.

[7] P. Piperni, A. DeBlois, R. Henderson, Development of a multilevel multidisciplinary-optimization capability for an industrial environment, AIAA journal 51 (10) (2013) 2335–2352.

[8] Z.-H. Han, S. Görtz, R. Zimmermann, Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function, Aerospace Science and technology 25 (1) (2013) 177–189.

[9] B. Peherstorfer, K. Willcox, M. Gunzburger, Survey of multifidelity methods in uncertainty propagation, inference, and optimization, Siam Review 60 (3) (2018) 550–591.

[10] M. G. Fernández-Godino, C. Park, N.-H. Kim, R. T. Haftka, Review of multi-fidelity models, arXiv preprint arXiv:1609.07196.

[11] L. Yan, T. Zhou, Adaptive multi-fidelity polynomial chaos approach to bayesian inference in inverse problems, Journal of Computational Physics 381 (2019) 110–128.

[12] P. Perdikaris, D. Venturi, J. O. Royset, G. E. Karniadakis, Multi-fidelity modelling via recursive co-kriging and gaussian–markov random fields, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences 471 (2179) (2015) 20150018.

[13] C. C. Fischer, R. V. Grandhi, P. S. Beran, Bayesian low-fidelity correction approach to multi-fidelity aerospace design, in: 58th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, 2017, p. 0133.

[14] J. Zheng, X. Shao, L. Gao, P. Jiang, Z. Li, A hybrid variable-fidelity global approximation modelling method combining tuned radial basis function base and kriging correction, Journal of Engineering Design 24 (8) (2013) 604–622.

[15] J. A. Christen, C. Fox, Markov chain monte carlo using an approximation, Journal of Computational and Graphical statistics 14 (4) (2005) 795–810.

[16] E. Laloy, B. Rogiers, J. A. Vrugt, D. Mallants, D. Jacques, Efficient posterior exploration of a high-dimensional groundwater model from two-stage markov chain monte carlo simulation and polynomial chaos expansion, Water Resources Research 49 (5) (2013) 2664–2682.

[17] B. Peherstorfer, B. Kramer, K. Willcox, Combining multiple surrogate models to accelerate failure probability estimation with expensive high-fidelity models, Journal of Computational Physics 341 (2017) 61–75.

[18] B. Peherstorfer, T. Cui, Y. Marzouk, K. Willcox, Multifidelity importance sampling, Computer Methods in Applied Mechanics and Engineering 300 (2016) 490–509.

[19] M. Razi, R. M. Kirby, A. Narayan, Fast predictive multi-fidelity prediction with models of quantized fidelity levels, Journal of Computational Physics 376 (2019) 992–1008.

[20] A. H. Elsheikh, I. Hoteit, M. F. Wheeler, Efficient bayesian inference of subsurface flow models using nested sampling and sparse polynomial chaos surrogates, Computer Methods in Applied Mechanics and Engineering 269 (2014) 515–537.

[21] R. Martí, M. G. Resende, C. C. Ribeiro, Multi-start methods for combinatorial optimization, European Journal of Operational Research 226 (1) (2013) 1–8.

[22] K. Sargsyan, C. Safta, H. N. Najm, B. J. Debusschere, D. Ricciuto, P. Thornton, Dimensionality reduction for complex models via bayesian compressive sensing, International Journal for Uncertainty Quantification 4 (1).

[23] O. Ijasan, C. Torres-Verdín, W. E. Preeg, Inversion-based petrophysical interpretation of logging-while-drilling nuclear and resistivity measurements, Geophysics 78 (6) (2013) D473–D489.

[24] D. Pardo, C. Torres-Verdín, Fast 1d inversion of logging-while-drilling resistivity measurements for improved estimation of formation resistivity in high-angle and horizontal wells, Geophysics 80 (2) (2015) E111–E124.

[25] S. A. Bakr, D. Pardo, C. Torres-Verdín, Fast inversion of logging-while-drilling resistivity measurements acquired in multiple wells, Geophysics 82 (3) (2017) E111–E120.

[26] B. I. Anderson, Modeling and inversion methods for the interpretation of resistivity logging tool response.

[27] A. A. Goldstein, On steepest descent, Journal of the Society for Industrial and Applied Mathematics, Series A: Control 3 (1) (1965) 147–151.

[28] R. Fletcher, Practical methods of optimization, John Wiley & Sons, 2013.

[29] J. J. Moré, The levenberg-marquardt algorithm: implementation and theory, in: Numerical analysis, Springer, 1978, pp. 105–116.

[30] A. Ranganathan, The levenberg-marquardt algorithm, Tutoral on LM algorithm 11 (1) (2004) 101–110.

[31] N. Wiener, The homogeneous chaos, American Journal of Mathematics 60 (4) (1938) 897–936.

[32] R. G. Ghanem, P. D. Spanos, Stochastic finite elements: a spectral approach, Courier Corporation, 2003.

[33] M. Eldred, J. Burkardt, Comparison of non-intrusive polynomial chaos and stochastic collocation methods for uncertainty quantification, in: 47th AIAA aerospace sciences meeting including the new horizons forum and aerospace exposition, 2009, p. 976.

[34] R. G. Ghanem, J. Red-Horse, A. Sarka, in: 8th ASCE Specialty Conference of Probabilistic Mechanics and Structural Reliability, 2000.

[35] A. O'Hagan, Polynomial chaos: A tutorial and critique from a statistician's perspective, SIAM/ASA J. Uncertainty Quantification 20 (2013) 1–20.

[36] S. Ji, Y. Xue, L. Carin, et al., Bayesian compressive sensing, IEEE Transactions on signal processing 56 (6) (2008) 2346.

[37] S. D. Babacan, R. Molina, A. K. Katsaggelos, Bayesian compressive sensing using laplace priors, IEEE Transactions on image processing 19 (1) (2009) 53–63.

[38] M. E. Tipping, Sparse bayesian learning and the relevance vector machine, Journal of machine learning research 1 (Jun) (2001) 211–244.

[39] C. Hans, Bayesian lasso regression, Biometrika 96 (4) (2009) 835–845.

[40] M. E. Tipping, A. C. Faul, et al., Fast marginal likelihood maximisation for sparse bayesian models., in: AISTATS, 2003.

[41] R. T. Haftka, D. Villanueva, A. Chaudhuri, Parallel surrogate-assisted global optimization with expensive functions–a survey, Structural and Multidisciplinary Optimization 54 (1) (2016) 3–13.

[42] Y. Jin, Surrogate-assisted evolutionary computation: Recent advances and future challenges, Swarm and Evolutionary Computation 1 (2) (2011) 61–70.

[43] M. Molga, C. Smutnicki, Test functions for optimization needs, Test functions for optimization needs 101.

[44] A. O'Hagan, Polynomial chaos: A tutorial and critique from a statistician's perspective, SIAM/ASA J. Uncertainty Quantification 20 (2013) 1–20.

[45] H. Lu, Q. Shen, J. Chen, X. Wu, X. Fu, Parallel multiple-chain dram mcmc for large-scale geosteering inversion and uncertainty quantification, Journal of Petroleum Science and Engineering 174 (2019) 189–200.

[46] Q. Shen, X. Wu, J. Chen, Z. Han, Y. Huang, Solving geosteering inverse problems by stochastic hybrid monte carlo method, Journal of Petroleum Science and Engineering 161 (2018) 9–16.