IOP Publishing Inverse Problems
Inverse Problems **36** (2020) 105008 (27pp) https://doi.org/10.1088/1361-6420/abb2fa

Data-driven forward discretizations for Bayesian inversion

D Bigoni[1], Y Chen[2], N Garcia Trillos[3], Y Marzouk[1] and D Sanz-Alonso[2,*]

[1] Massachusetts Institute of Technology, MA, United States of America
[2] University of Chicago, IL, United States of America
[3] University of Wisconsin, Madison, WI, United States of America

E-mail: sanzalonso@uchicago.edu

Received 16 March 2020, revised 25 July 2020
Accepted for publication 26 August 2020
Published 25 September 2020

Abstract

This paper suggests a framework for the learning of discretizations of expensive forward models in Bayesian inverse problems. The main idea is to incorporate the parameters governing the discretization as part of the unknown to be estimated within the Bayesian machinery. We numerically show that in a variety of inverse problems arising in mechanical engineering, signal processing and the geosciences, the observations contain useful information to guide the choice of discretization.

Keywords: Bayesian inverse problems, data-driven discretizations, hierarchical learning

(Some figures may appear in colour only in the online journal)

1. Introduction

Models used in science and engineering are often described by problem-specific input parameters that are estimated from indirect and noisy observations. The inverse problem of input reconstruction is defined in terms of a *forward model* from inputs to observable quantities, which in many applications needs to be approximated by discretization. A broad class of examples motivating this paper is the reconstruction of input parameters of differential equations. The choice of forward model discretization is particularly important in Bayesian formulations of inverse problems: discretizations need to be cheap since statistical recovery may involve millions of evaluations of the discretized forward model; they also need to be accurate enough to enable input reconstruction. The goal of this paper is to suggest a simple data-driven framework to build forward model discretizations to be used in Bayesian inverse problems. The resulting discretizations are data-driven in that they finely resolve regions of the input space

*Author to whom any correspondence should be addressed.

where the data are most informative, while keeping the cost moderate by coarsely resolving regions that are not informed by the data.

To be concrete and explain the idea, let us consider the inverse problem of recovering an unknown u from data y related by

$$y = \mathcal{G}(u) + \eta, \tag{1.1}$$

where $\mathcal{G}$ denotes the forward model from inputs to observables, $\eta \sim N(0, \Gamma)$ represents model error and observation noise, and Γ denotes a positive definite noise covariance matrix. We will follow a Bayesian approach, viewing u as a random variable [23, 37, 39] with *prior* distribution $p_u(u)$. The Bayesian solution to the inverse problem is the *posterior* distribution $p_{u|y}(u)$ of u given the data y, which by an informal application of Bayes theorem is characterized by

$$p_{u|y}(u) \propto \exp\left(-\Phi(u; y)\right) p_u(u), \quad \Phi(u; y) := \frac{1}{2}\|y - \mathcal{G}(u)\|_\Gamma^2 \tag{1.2}$$

with $\|\cdot\|_\Gamma := \|\Gamma^{-1/2}\cdot\|$. A common computational bottleneck arises when the forward model $\mathcal{G}$ and hence the likelihood are intractable, meaning that it is impossible or too costly to evaluate. This paper introduces a framework to tackle this computational challenge by employing data-driven discretizations of the forward model. The main idea is to include the parameters that govern the discretization as part of the unknown to be estimated within the Bayesian machinery. More precisely, we consider a family $\{\mathcal{G}^a\}_{a\in\mathcal{A}}$ of approximate forward models and put a prior $q_{u,a}(u, a)$ over both unknown inputs u and forward discretization parameters $a \in \mathcal{A}$ to define a joint posterior

$$q_{u,a|y}(u, a) \propto \exp\left(-\Psi(u, a)\right) q_{u,a}(u, a), \quad \Psi(u, a; y) := \frac{1}{2}\|y - \mathcal{G}^a(u)\|_\Gamma^2. \tag{1.3}$$

While this structure underlies many hierarchical formulations of Bayesian inverse problems [23], in this paper the hyper-parameter a determines the choice of discretization of the forward model $\mathcal{G}$.

Including the learning of the numerical discretizations of the forward map as part of the inference agrees with the Bayesian philosophy of treating unknown quantities as random variables, and is also in the spirit of recent probabilistic numerical methods [7]; rather than implicitly assuming that a true hidden numerical discretization of the forward model generates the data, a Bayesian would acknowledge the uncertainty in the choice of a suitable discretization and let the observed data inform such a choice. Moreover, the Bayesian viewpoint has two main practical advantages. First, data-informed grids will typically be coarse in regions of the input space that are not informed by the data, allowing successful input reconstruction at a reduced computational cost. Second, the posterior $q_{u,a|y}(u, a)$ contains useful uncertainty quantification on the discretizations. This additional uncertainty information may be exploited to build a high-fidelity forward model to be employed within existing inverse problem solvers, either in Bayesian or classical settings.

1.1. Related work

The Bayesian formulation of inverse problems provides a flexible and principled way to combine data with prior knowledge. However, in practice it is rarely possible to perform posterior inference with the model of interest (1.2) due to various computational challenges. In this paper we investigate the construction of computable data-driven forward discretizations of intractable likelihoods arising in the inversion of differential equations. Other intertwined obstacles for posterior inference are:

- **Sampling cost.** While exact posterior inference is often intractable, approximate posterior inference can be performed by employing sampling algorithms. Markov chain Monte Carlo and particle-based methods are popular, but implementations of these algorithms require repeated evaluation of the forward model $\mathcal{G}$, which may be costly.
- **Large input dimension.** The unknown parameter u may be high, or even infinite dimensional. While the convergence rate of certain sampling schemes may be independent of the input dimension [1, 8, 15], the computational and memory cost per sample may increase prohibitively with dimension.
- **Model error.** The forward model is only an approximation of the real relationship between input and observable output variables. Model discrepancy can damage input recovery.
- **Complex geometry.** The unknown may be a function defined on a complex, perhaps unknown domain that needs to be approximated.

All these challenges have long been identified [23–25, 36], giving rise to a host of methods for sampling, parameter reduction, model reduction, enhanced model error techniques and geometric methods for inverse problems. We focus on the model-reduction problem of building forward discretizations, but the methodology proposed in this paper can be naturally combined with existing techniques that address complementary challenges. For instance, our forward model discretizations may be used within multilevel MCMC methods [18] or within two-stage sampling methods [6, 10, 13, 20, 41], and thus help to reduce the sampling cost. Also, forward model discretizations may be combined with parameter reduction and model adaptation techniques, as in [26, 28]. It is important, however, to distinguish between the parameter and model reduction problems. While the former aims to find suitable small-dimensional representations of the input u, the latter is concerned with effectively reducing the number of degrees of freedom used to compute the forward model $\mathcal{G}$. In regards to model error, our framework may be thought of as incorporating Bayesian model choice to the Bayesian solution of inverse problems by viewing each forward model discretization as a potential model. Following this interpretation, the *a posteriori* choice of forward discretization may in principle be determined using Bayes factors. Lastly, learning appropriate discretizations of forward models is particularly important for inverse problems set in complex, possibly uncertain geometries [15, 17, 22].

Many approaches to computing forward map surrogates and reduced-order models have been proposed; we refer to [14] for an extended survey, and to [32] for a broader discussion of multi-fidelity models in other outer-loop applications. Most methods fall naturally into one of three categories:

(a) Projection-based methods: the forward model equations are described in a reduced basis that is constructed using few high-fidelity forward solves (called snapshots). Two popular ways to construct the reduced basis are proper orthogonal decomposition and reduced order basis. In the inverse problem context, data-informed construction of snapshots [10] allows to approximate the posterior support with fewer high-fidelity forward runs. To our knowledge, there is little theory to guide the required number or location of snapshots to meet a given error tolerance.

(b) Spectral methods: polynomial chaos [42] is a popular method for forward propagation of uncertainty, that has more recently been used to produce surrogates for intractable likelihoods [29]. The paper [30] translates error in the likelihood approximation to Kullback–Leibler posterior error. A drawback of these methods is that they are only practical

when the random inputs can be represented by a small number of random variables. Recent interest lies in adapting the spectral approximations to observed data [27].

(c) Gaussian processes and neural networks: some of the earliest efforts to allow for Bayesian inference with complex models suggested to use Gaussian processes [33] to construct surrogate likelihood models [25, 36]. The accuracy of the resulting approximations has been studied in [40], which again requires a suitable representation of the input space. Finally, representation of the likelihood using neural networks in combination with generalized polynomial chaos expansions has been investigated in [38].

This paper focuses on grid-based discretizations and density-based discretizations of static inverse problems arising in mechanical engineering, signal processing and the geophysical sciences. However, the proposed framework may be used in conjunction with other reduced-order models, in dynamic data assimilation problems, and in other applications. Finally, we mention that for classical formulations of certain specific inverse problems, optimal forward discretization choices have been proposed [2, 5].

1.2. Outline and contributions

Section 2 reviews the Bayesian formulation of inverse problems. Section 3 describes the main framework for the Bayesian learning of forward map discretizations. We will consider two ways to parametrize discretizations: in the first, the grid points locations are learned directly, and in the second we learn a probability density from which to obtain the grid. In section 4 we discuss a general approach to sampling the joint posterior over unknown input and discretization parameters, which consists of a Metropolis-within-Gibbs that alternates between a reversible jump Markov chain Monte Carlo (MCMC) algorithm to update the discretization parameters and a standard MCMC to update the unknown input. Section 5 demonstrates the applicability, benefits, and limitations of our approach in a variety of inverse problems arising in mechanical engineering, signal processing and source detection, considering Euler discretization of ODEs, Euler–Maruyama discretization of SDEs, and finite element methods for PDEs. We conclude in section 6 with some open questions for further research.

2. Background: Bayesian formulation of inverse problems

Consider the inverse problem of recovering an unknown $u \in \mathcal{U}$ from data $y \in \mathbb{R}^m$ related by

$$y = \mathcal{G}(u) + \eta, \tag{2.1}$$

where $\mathcal{U}$ is a space of admissible unknowns and η is a random variable whose distribution is known to us, but not its realization. In many applications, the *forward model* $\mathcal{G} : \mathcal{U} \to \mathbb{R}^m$ can be written as the composition of forward and observation maps, $\mathcal{G} = \mathcal{O} \circ \mathcal{F}$. The forward map $\mathcal{F} : \mathcal{U} \to \mathcal{Z}$ represents a *complex* mathematical model that assigns outputs $z \in \mathcal{Z}$ to inputs $u \in \mathcal{U}$. For instance, u may be the parameters of a differential equation, and z may be its analytical solution. The observation map $\mathcal{O} : \mathcal{Z} \to \mathcal{Y}$ establishes a link between outputs and observable quantities, e.g. by point-wise evaluation of the solution.

In the Bayesian formulation of the inverse problem (2.1), one specifies *a prior* distribution on u and seeks to characterize the *posterior* distribution of u given y. If the input space $\mathcal{U}$ is finite dimensional, $\mathcal{U} \subset \mathbb{R}^d$, then the *prior* distribution, denoted as $p_u(u)$, can be defined through its Lebesgue density. The noise distribution of η in $\mathbb{R}^m$ gives the *likelihood* $p_{y|u}(y|u)$. In this work

we assume, for concreteness, that η is a zero-mean Gaussian with covariance $\Gamma \in \mathbb{R}^{m\times m}$, so that

$$p_{y|u}(y|u) \propto \exp\left(-\Phi(u;y)\right), \quad \Phi(u;y) := \frac{1}{2}\|y - \mathcal{G}(u)\|_\Gamma^2, \tag{2.2}$$

where $\|\cdot\|_\Gamma := \|\Gamma^{-1/2}\cdot\|$. Using Bayes' formula, the posterior density is given by

$$p_{u|y}(u) = \frac{1}{Z}\, p_{y|u}(y|u)p_u(u), \quad Z = \int_{\mathcal{U}} p_{y|u}(y|u)p_u(u)\mathrm{d}u \tag{2.3}$$

with multiplicative constant Z depending on y.

For many inverse problems of interest, the unknown u is a function and the input space $\mathcal{U}$ is an infinite-dimensional Banach space. In such a case, the prior cannot be specified in terms of its Lebesgue density, but rather as a measure μ_u supported on $\mathcal{U}$. Provided that $\mathcal{G}: \mathcal{U} \to \mathbb{R}^m$ is measurable and that $\mu_u(\mathcal{U}) = 1$, the *posterior* measure $\mu_{u|y}$ is still defined, in analogy to (2.3), as a change of measure with respect to the *prior*

$$\frac{\mathrm{d}\mu_{u|y}}{\mathrm{d}\mu_u}(u) \propto \exp\left(-\Phi(u;y)\right). \tag{2.4}$$

We refer to [16, 39] for further details. The posterior $\mu_{u|y}$ contains, in a precise sense [43], all the information on u available in the data y and the prior μ_u. This paper is concerned with inverse problems where $\mathcal{G} = \mathcal{O}\circ\mathcal{F}$ arises from a complex model $\mathcal{F}$ that cannot be evaluated point-wise; we then seek to approximate the *idealized* posterior $\mu_{u|y}$ finding a compromise between accuracy and computational cost.

A simple but important observation is that approximating $\mathcal{F}$ accurately is *not* necessary in order to approximate $\mu_{u|y}$ accurately. It is *only* necessary to approximate $\mathcal{G} = \mathcal{O}\circ\mathcal{F}$, since $\mathcal{F}$ appears in the *posterior* only through $\mathcal{G}$. While producing discretizations to complex models $\mathcal{F}$ has been widely studied in numerical analysis, here we investigate how to approximate $\mathcal{F}$ with the specific goal of approximating the *posterior* $\mu_{u|y}$, incorporating *prior* and data knowledge into the discretizations. For some inverse problems the observation operator $\mathcal{O}$ also needs to be discretized, leading to similar considerations.

3. Bayesian discretization of the forward model

Suppose that $\mathcal{F}$ is the solution map to a differential equation that cannot be solved in closed form, and $\mathcal{O}$ is point-wise evaluation of the solution. Standard practice in computing the Bayesian solution to the inverse problem involves using an *a priori* fixed discretization, e.g., by discretizing the domain of the differential equation into a fine grid. Provided that the grid is fine enough, the *posterior* defined with the discretized forward map can approximate well the one in (2.4). However, the discretizations are usually performed on a *fine* uniform grid which may lead to unnecessary waste of computational resources. Indeed, it is expected that the choice of discretization should be problem dependent, and should be informed both by the observation locations (which are often not uniform in space) and by the value of the unknown input parameter that we seek to reconstruct. Thus we seek to learn *jointly* the unknown input u and the discretization of the forward map.

We will consider a parametric family of discretizations. Precisely, we let

$$\mathcal{A} := \left\{a = (k,\theta) : k \in \mathcal{K} \subset \{1,2,\ldots\}, \quad \theta \in D(k) \subset \mathbb{R}^{d(k)}\right\}, \tag{3.1}$$

and each pair $a = (k, \theta) \in \mathcal{A}$ will parameterize a discretized forward model $\mathcal{G}^a$. For given $k \in \mathcal{K}$, $d(k)$ represents the degrees of freedom in the discretization, and $\theta \in D(k)$ is the $d(k)$-dimensional model parameter of the discretization, where $D(k)$ is the region containing all parameters of interest. In analogy with Bayesian model selection frameworks [19, 34], $k \in \mathcal{K}$ may be interpreted as indexing the discretization model. We focus on the model reduction rather than the parameter reduction problem, and assume that all approximation maps share the same input and output spaces $\mathcal{U}$ and $\mathcal{Z}$. We will illustrate the flexibility of this framework using grid-based approximations and density-based discretizations.

Example 3.1 (Grid-based discretizations). Here the first component of each element $a = (k, \theta) \in \mathcal{A}$ represents the number of points in a grid. The set $\mathcal{K}$ contains all allowed grid sizes. If we denote $D \subset \mathbb{R}^d$ as the temporal or spatial domain of the equation being discretized, we define $D(k) = D^k$ and $d(k) = d \times k$, where $D^k := D \times \cdots \times D$ denotes the k-fold Cartesian product of D. Then the second component $\theta = [x_1, \ldots x_k]$ encodes the locations of k grid points.

Example 3.2 (Density-based discretizations). Here the first component of each element $a = (k, \theta) \in \mathcal{A}$ represents again the number of points in a grid, and the second component parametrizes a probability density $\rho = \rho(x; \theta)$ on the temporal or spatial domain of interest, by a parameter θ of fixed dimension, independent of k. Given $a \in \mathcal{A}$ we may for instance employ MacQueen's method [12] to formulate a centroidal Voronoi tessellation, which outputs k generators $\{x_1, \ldots, x_k\}$, and then use them as grid points to generate a finite element grid by Delaunay triangulation. Intuitively θ controls the spatial density of the non-uniform grid points $\{x_1, \ldots, x_k\}$. The space $\mathcal{K}$ represents, as before, all the allowed number of grid points.

Example 3.3 (Other discretizations). As mentioned in the introduction, other discretizations and model reduction techniques could be considered within the above framework, including projection-based approximations, Gaussian processes, and graph-based methods. However, in our numerical experiments we will focus on grid-based and density-based discretizations.

We consider a product *prior* on $(u, a) \in \mathcal{U} \times \mathcal{A}$, given by

$$q_{u,a}(u, a) = q_u(u) q_a(a), \tag{3.2}$$

where $q_u(u) = p_u(u)$ is as in the original, idealized inverse problem (2.1). In general, conditioning on u may or may not provide useful information about how to approximate $\mathcal{G}(u)$. When it does, this can be infused into the *prior* by letting the conditional distribution of a given u depend on u. For simplicity we restrict ourselves to the product structure (3.2).

The examples above and the structure of the space $\mathcal{A}$ defined in equation (3.1) suggest to define hierarchically *a prior* over $a \in \mathcal{A}$

$$q_a(a) = q_{k,\theta}(k, \theta) = q_k(k) q_{\theta|k}(\theta|k), \tag{3.3}$$

where $q_k(k)$ is a probability mass function that penalizes expensive discretizations that employ large number $d(k)$ of degrees of freedom, and $q_{\theta|k}(\theta|k)$ denotes the conditional distribution of θ given k in $D(k)$.

We define the likelihood of observing data y given (u, a) by

$$q_{y|u,a}(y|u,a) \propto \exp\left(-\Psi(u,a;y)\right), \quad \Psi(u,a;y) := \frac{1}{2}\|y - \mathcal{G}^a(u)\|_\Gamma^2, \tag{3.4}$$

where $\mathcal{G}^a = \mathcal{O} \circ \mathcal{F}^a$. The discretized forward maps $\mathcal{F}^a$ will be chosen so that evaluating Ψ is possible.

We first consider the case where $\mathcal{K} = \{k\}$ is a singleton, and $\mathcal{A} := \{a = (k, \theta) : \theta \in D(k) \subset \mathbb{R}^{d(k)}\}$ has a Euclidean space structure. Then, by Bayes' formula,

$$\begin{aligned} q_{u,a|y}(u,a) &= \frac{1}{\tilde{Z}} q_{y|u,a}(y|u,a) q_u(u) q_a(a), \\ \tilde{Z} &= \int_{\mathcal{U}\times\mathcal{A}} q_{y|u,a}(y|u,a) q_u(u) q_a(a) \mathrm{d}u \ \mathrm{d}a, \end{aligned} \tag{3.5}$$

where $\tilde{Z} = \tilde{Z}(y)$ is a normalizing constant. The first marginal of $q_{u,a|y}(u,a)$, which we denote as $q_{u|y}(u)$, constitutes a data-informed approximation of the *posterior* $p_{u|y}(u)$ from the full, idealized inverse problem (2.3). We have the following result:

Proposition 3.4. *Let $p_{u|y}(u)$ be defined as in (2.3) and $q_{u|y}(u)$ be defined as above. If $\mathcal{G}$ is bounded and, for $q_{u,a}$-almost any (u,a), $\|\mathcal{G}^a(u) - \mathcal{G}(u)\| < \epsilon$, then*

$$d_{TV}\left(q_{u|y}(u), p_{u|y}(u)\right) < C\epsilon \tag{3.6}$$

for some constant C independent of ϵ.

Proof. Integrating both sides of the first equation in (3.5) with respect to a:

$$q_{u|y}(u) = \frac{1}{\tilde{Z}} \left(\int_{\mathcal{A}} q_{y|u,a}(y|u,a) q_a(a) \mathrm{d}a \right) q_u(u) =: \frac{1}{\tilde{Z}} \tilde{g}_y(u) q_u(u).$$

Compare this to equation (2.3) and write $g_y(u) = p_{y|u}(y|u)$, we have

$$\begin{aligned} \|\tilde{g}_y(u) - g_y(u)\| \leqslant \int_{\mathcal{A}} &\exp\left(-\frac{1}{2}\|y - \mathcal{G}^a(u)\|_\Gamma^2\right) \\ &- \exp\left(-\frac{1}{2}\|y - \mathcal{G}(u)\|_\Gamma^2\right) \mathrm{d}a \leqslant C\|\mathcal{G}^a(u) - \mathcal{G}(u)\|, \end{aligned}$$

where the last inequality follows from the Lipschitz continuity of e^{-w} for $w \geqslant 0$, boundedness of $\mathcal{G}$, and equivalence of norm in $\mathbb{R}^m$. This implies that $|\tilde{g}_y(u) - g_y(u)| \leqslant C\epsilon$ and hence $|\tilde{Z} - Z| \leqslant C\epsilon$. Then the statement follows from a slight modification of theorem 1.14 in [37] and the definition of TV distance.

Now we are ready to extend the above results to infinite-dimensional input space $\mathcal{U}$. We define a prior measure on $\mathcal{U} \times \mathcal{A}$ given by $\nu_{u,a}(\mathrm{d}u, \mathrm{d}a) = \nu_u(\mathrm{d}u) \times \nu_a(\mathrm{d}a)$, where $\nu_u(\mathrm{d}u) = \mu_u(\mathrm{d}u)$ is as in the idealized inverse problem. The posterior measure on $\mathcal{U} \times \mathcal{A}$ conditioning on y will still be denoted by $\nu_{u,a|y}$.

Proposition 3.5. *Suppose that $\mathcal{U}$ is a separable Banach space with $\nu_u(\mathcal{U}) = 1$, $\Psi : \mathcal{U} \times \mathcal{A} \to \mathbb{R}$ is continuous. Then the* posterior *measure $\nu_{u,a|y}$ of (u, a) given y is absolutely continuous with respect to the* prior *$\nu_{u,a}$ on $\mathcal{U} \times \mathcal{A}$ and has Radon–Nikodym derivative*

$$\frac{\mathrm{d}\nu_{u,a|y}}{\mathrm{d}\nu_{u,a}}(u, a) \propto \exp\left(-\Psi(u, a; y)\right). \tag{3.7}$$

Proof. By the disintegration theorem (which holds for arbitrary Radon measures on separable metric spaces—see [11] chapter 3, page 70) for all measurable subsets $U' \subseteq \mathcal{U}$, $A' \subseteq \mathcal{A}$ and $Y' \subseteq \mathcal{Y}$, we can write $\nu_{u,a,y}(U' \times A' \times Y')$ in two different ways:

$$\begin{aligned}\int_{Y'} \nu_{u,a|y}(U' \times A'|y)\mathrm{d}\nu_y(y) &= \nu_{u,a,y}(U', A', Y') \\ &= \int_{U' \times A'} \nu_{y|u,a}(Y'|u, a)\mathrm{d}\nu_{u,a}(u, a).\end{aligned}$$

In particular,

$$\begin{aligned}\nu_y(Y') &= \int_{\mathcal{U}\times\mathcal{A}} \nu_{y|u,a}(Y'|u, a)\mathrm{d}\nu_{u,a}(u, a) \\ &= \int_{\mathcal{U}\times\mathcal{A}} \int_{Y'} Z_\Gamma^{-1} \exp\left(-\frac{1}{2}\|y - \mathcal{G}^a(u)\|_\Gamma^2\right) \mathrm{d}y\mathrm{d}\nu_{u,a}(u, a),\end{aligned}$$

given our assumptions on the noise model, where Z_Γ is a constant depending on the noise covariance Γ. We can then use Tonelli's theorem to swap the order of the integrals and obtain

$$\nu_y(Y') = \int_{Y'} \left(\int_{\mathcal{U}\times\mathcal{A}} Z_\Gamma^{-1} \exp\left(-\frac{1}{2}\|y - \mathcal{G}^a(u)\|_\Gamma^2\right) \mathrm{d}\nu_{u,a}(u, a) \right) \mathrm{d}y.$$

Given that Y' is arbitrary, we conclude that ν_y is absolutely continuous with respect to the Lebesgue measure with density:

$$\frac{\mathrm{d}\nu_y(y)}{\mathrm{d}y} = \int_{\mathcal{U}\times\mathcal{A}} Z_\Gamma^{-1} \exp\left(-\frac{1}{2}\|y - \mathcal{G}^a(u)\|_\Gamma^2\right) \mathrm{d}\nu_{u,a}(u, a).$$

On the other hand,

$$\begin{aligned}&\int_{U' \times A'} \nu_{y|u,a}(Y'|u, a)\mathrm{d}\nu_{u,a}(u, a) \\ &\quad= \int_{U' \times A'} \int_{Y'} Z_\Gamma^{-1} \exp\left(-\frac{1}{2}\|y - \mathcal{G}^a(u)\|_\Gamma^2\right) \mathrm{d}y \; \mathrm{d}\nu_{u,a}(u, a) \\ &\quad= \int_{U' \times A'} \int_{Y'} Z_\Gamma^{-1} \exp\left(-\frac{1}{2}\|y - \mathcal{G}^a(u)\|_\Gamma^2\right) \left(\frac{\mathrm{d}\nu_y(y)}{\mathrm{d}y}\right)^{-1} \mathrm{d}\nu_y(y)\mathrm{d}\nu_{u,a}(u, a) \\ &\quad= \int_{Y'} \left(\int_{U' \times A'} Z_\Gamma^{-1} \exp\left(-\frac{1}{2}\|y - \mathcal{G}^a(u)\|_\Gamma^2\right) \left(\frac{\mathrm{d}\nu_y(y)}{\mathrm{d}y}\right)^{-1} \mathrm{d}\nu_{u,a}(u, a) \right) \mathrm{d}\nu_y(y),\end{aligned}$$

Algorithm 4.1. Metropolis-within-Gibbs Core Structure.

Choose $(u^{(1)}, a^{(1)}) \in \mathcal{U} \times \mathcal{A}$.
for $n = 1 : N$ **do**
1. Sample $u^{(n+1)} \sim \mathbb{K}^{a^{(n)},y}(u^{(n)}|\cdot)$.
2. Sample $a^{(n+1)} \sim \mathbb{L}^{u^{(n+1)},y}(a^{(n)}|\cdot)$.

end for

applying Tonelli's theorem once again to obtain the last equality. Since Y' was arbitrary, it follows that for ν_y-a.e. y we have

$$\nu_{u,a|y}(U' \times A'|y) = \int_{U' \times A'} Z_\Gamma^{-1} \exp\left(-\frac{1}{2}\|y - \mathcal{G}^a(u)\|_\Gamma^2\right) \times \left(\frac{\mathrm{d}\nu_y(y)}{\mathrm{d}y}\right)^{-1} \mathrm{d}\nu_{u,a}(u,a).$$

In turn, from the arbitrariness of U', A' it follows that, for ν_y-a.e. y the measure $\nu_{u,a|y}(\cdot|y)$ is absolutely continuous with respect to $\nu_{u,a}$ and its Radon–Nykodym derivative satisfies

$$\frac{\mathrm{d}\nu_{u,a|y}}{\mathrm{d}\nu_{u,a}}(u,a) \propto \exp\left(-\frac{1}{2}\|y - \mathcal{G}^a(u)\|_\Gamma^2\right)$$

as claimed, where the constant of proportionality depends on y. □

As in the finite dimensional case, we have the following result:

Proposition 3.6 (Well-posedness of posterior). *Under the same assumption as in proposition 3.5, suppose further that for $q_{u,a}$-almost any (u,a), $\|\mathcal{G}^a(u) - \mathcal{G}(u)\| < \epsilon$, and $\mathcal{G}$ is bounded. Then we have*

$$d_{\mathrm{TV}}(\nu_{u|y}, \mu_{u|y}) < C\epsilon \tag{3.8}$$

for some constant C independent of ϵ.

Remark 3.7. In the context of grid-based forward approximations, the condition '$\|\mathcal{G}^a(u) - \mathcal{G}(u)\| < \epsilon$ $q_{u,a}$-almost surely' can be interpreted as 'almost any draw from the approximation parameter space $\mathcal{A}$ can produce an approximation of the forward model with error at most ϵ'. This is often the case, for example, when the grids are finer than some threshold under regularity conditions on the input space.

4. Sampling the posterior

The structure of the joint posterior $\nu_{u,a|y}$ over unknowns $u \in \mathcal{U}$ and approximations $a \in \mathcal{A}$ suggests using a Metropolis-within-Gibbs sampler, which constructs a Markov chain $(u^{(n)}, a^{(n)})$ by alternatingly sampling each coordinate (algorithm 4.1):

In the above, $\mathbb{K}^{a,y}$ and $\mathbb{L}^{u,y}$ are Metropolis–Hastings Markov kernels that are reversible with respect to $u|(a,y)$ and $a|(u,y)$. We remark that the kernel $\mathbb{K}^{a,y}$ involves evaluation of the forward model approximation $\mathcal{G}^a$ but not of the intractable full model $\mathcal{G}$. While the choice and design of the kernels $\mathbb{K}^{a,y}$ and $\mathbb{L}^{u,y}$ will clearly be problem-specific, and here we consider a

standard method appropriate for the case where the input space $\mathcal{U}$ is a space of functions to define $\mathbb{K}^{a,y}$.

Before describing how to sample the full conditionals $\nu_{u|a,y}$ and $\nu_{a|u,y}$ of $u|(a,y)$ and $a|(u,y)$ it is useful to note that they satisfy the following expressions:

$$\frac{\mathrm{d}\nu_{u|a,y}}{\mathrm{d}\nu_u}(u) \propto \exp\left(-\Psi(u,a;y)\right), \qquad \frac{\mathrm{d}\nu_{a|u,y}}{\mathrm{d}\nu_a}(a) \propto \exp\left(-\Psi(u,a;y)\right). \tag{4.1}$$

4.1. Sampling the full conditional u|y, a

For given a and y, we can sample from $\nu_{u|a,y}$ using pCN [3], with proposal

$$\tilde{u} := \sqrt{1-\beta^2}u + \beta\xi, \quad \xi \sim \mu_u,$$

and acceptance probability

$$\alpha(u,\tilde{u}) := \min\left\{1, \exp\left(-\Psi(\tilde{u},a;y) + \Psi(u,a;y)\right)\right\}.$$

Other discretization-invariant MCMC samplers [9, 35] could also be used to update $u|y,a$, but pCN is a straightforward and effective choice in the examples considered here.

4.2. Sampling the full conditional a|y, u

4.2.1. Sampling grid-based discretizations. We will use a Markov kernel $\mathbb{L}^{u,y}(a|\cdot)$ written as a mixture of two kernels, i.e.

$$\mathbb{L}^{u,y}(a|\cdot) = \zeta\mathbb{L}_1^{u,y}(a|\cdot) + (1-\zeta)\mathbb{L}_2^{u,y}(a|\cdot),$$

each of which is induced by a different Metropolis–Hastings algorithm. ζ determines the mixture weight. The proposal mechanism for each of the kernels corresponds to a different type of movement, described next:

(a) For $\mathbb{L}_1^{u,y}(a|\cdot)$ we use Metropolis–Hastings to sample from the distribution $\nu_{a|u,y}$ using the following proposal: given $a = (k,\theta)$ with $\theta = [\theta_1,\ldots,\theta_k]$ we set $\tilde{k} = k$ (i.e. the number of grid points stays the same) and let $\tilde{\theta}$ be defined by

$$\tilde{\theta}_i = \theta_i, \quad i = 1,\ldots,k-1,$$

and sample $\tilde{\theta}_k$ from a distribution on D with density (w.r.t. Lebesgue measure on D) τ_θ. In principle the density used to sample $\tilde{\theta}_k$ may depend on θ.

(b) For $\mathbb{L}_2^{u,y}(a|\cdot)$ we use Metropolis–Hastings to sample from the distribution $\nu_{a|u,y}$ using the following proposal: given $a = (k,\theta)$ we sample $\tilde{k} \sim \sigma(k|\cdot)$ where $\sigma(k|\cdot)$ is a Markov kernel on $\mathbb{N}$, and then generate $\tilde{\theta}$ according to

- If $\tilde{k} > k$ let $\tilde{\theta}_i = \theta_i$ for all $i = 1,\ldots,k$ and then sample $\tilde{\theta}_{k+1},\ldots,\tilde{\theta}_{\tilde{k}}$ independently from the density τ_θ.
- If $\tilde{k} \leqslant k$ let $\tilde{\theta}_i = \theta_i$ for all $i = 1,\ldots,\tilde{k}$.

Remark 4.1. We notice that the proposals described above are particular cases of the ones used in reversible jump Markov chain Monte Carlo [19].

For the Metropolis–Hastings algorithm associated to $\mathbb{L}_1^{u,y}(a|\cdot)$ the acceptance probability takes the form

$$\alpha_1(a,\tilde{a}) = \min\left\{1, \exp\left(-\Psi(u,\tilde{a};y) + \Psi(u,a;y)\right)\frac{\tau_{\tilde{\theta}}(\theta_k)}{\tau_{\theta}(\tilde{\theta}_k)}\right\},$$

where recall $a = (k,\theta)$ and $\theta = (\theta_1,\ldots,\theta_k)$ and $\tilde{a}$ is defined similarly.

For the Metropolis–Hastings algorithm associated to $\mathbb{L}_2^{u,y}(a|\cdot)$ the acceptance probability takes the form

$$\alpha_2(a,\tilde{a}) = \min\left\{1, \frac{\sigma(k|\tilde{k})\nu_k(\tilde{k})}{\sigma(\tilde{k}|k)\nu_k(k)}\ \exp\left(-\Psi(u,\tilde{a};y) + \Psi(u,a;y)\right)H(k,\theta,\tilde{k},\tilde{\theta})\right\},$$

where

$$H(k,\theta,\tilde{k},\tilde{\theta}) := \begin{cases} \displaystyle\prod_{i=1}^{k-\tilde{k}} \tau_{\tilde{\theta}}(\theta_{\tilde{k}+i}) & \text{if } k > \tilde{k}, \\ \displaystyle\left(\prod_{i=1}^{\tilde{k}-k} \tau_{\theta}(\tilde{\theta}_{k+i})\right)^{-1} & \text{if } \tilde{k} \geqslant k. \end{cases}$$

We notice that since each of the kernels $\mathbb{L}_1^{u,y}(a|\cdot)$ and $\mathbb{L}_2^{u,y}(a|\cdot)$ is defined by a Metropolis–Hastings algorithm, they leave the target $\nu_{a|u,y}$ invariant, and hence so does the kernel $\mathbb{L}^{u,y}(a|\cdot)$.

Remark 4.2. If in the above the distribution τ_θ is, regardless of θ, the uniform distribution on the domain D, then the acceptance probabilities reduce, respectively, to

$$\alpha_1(a,\tilde{a}) = \min\left\{1, \exp\left(-\Psi(u,\tilde{a};y) + \Psi(u,a;y)\right)\right\},$$

and

$$\alpha_2(a,\tilde{a}) = \min\left\{1, \frac{\sigma(k|\tilde{k})\nu_k(\tilde{k})}{\sigma(\tilde{k}|k)\nu_k(k)}\ \exp\left(-\Psi(u,\tilde{a};y) + \Psi(u,a;y)\right)\right\}.$$

4.2.2. Sampling density-based discretizations. Since in this case the dimension of θ is fixed, the calculation of the acceptance probabilities is straightforward and the details are omitted. We refer to subsection 5.3 for a numerical example.

5. Numerical examples

In this section we demonstrate the applicability of our framework and sampling approach in a variety of inverse problems. Our aim is illustrating the benefits and potential limitations of the methods; for this reason we consider inverse problems for which we have intuitive understanding of where the discretizations should concentrate, thus validating the performance of the proposed approach. Before discussing the numerical results, we summarize the main goals and outcomes of each set of experiments:

- In subsection 5.1 we consider an inverse problem in mechanics [4], for which some observation settings highly influence the best choice of discretization while others inform it mildly. Our numerical results show that the gain afforded by grid learning is most

clear whenever the observation locations highly influence the choice of discretization. We employ grid-based discretizations as described in example 3.1 with an Euler discretization of the forward map. We also illustrate the applicability of the method in both finite and infinite-dimensional representations of the unknown parameter, showing a more dramatic effect in the latter.

- In subsection 5.2 we consider an inverse problem in signal processing [21], with a choice of observation locations that determine where the discretization should concentrate. Our numerical results show that the grids adapt to the expected region, and that the degrees of freedom in the discretization necessary to reconstruct the unknown is below that necessary to satisfy stability of the numerical method with uniform grids. We employ grid-based discretizations as described in example 3.1 with an Euler–Maruyama discretization of the forward map.
- In subsection 5.3 we consider an inverse problem in source detection, where the true hidden unknown determines how best to discretize the forward model. Our numerical results show that the grids adapt as expected. We employ density-based discretizations as described in example 3.2 with a finite element discretization of the forward model.

5.1. Euler discretization of ODEs: estimation of the Young's modulus of a cantilever beam

We consider an inhomogeneous cantilever beam clamped on one side ($x = 0$) and free on the other ($x = L$). Define $D = [0, L]$. Let $u(x)$ denote its Young's modulus and let $M(x)$ be a load applied onto the beam. Timoshenko's beam theory gives the displacement $z(x)$ of the beam and the angle of rotation $\varphi(x)$ through the coupled ordinary differential equations

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}x}\left[\dfrac{u(x)}{2(1+r)}\left(\varphi(x) - \dfrac{\mathrm{d}}{\mathrm{d}x}\, z(x)\right)\right] = \dfrac{M(x)}{\kappa A}, \\ \dfrac{\mathrm{d}}{\mathrm{d}x}\left(u(x)I\, \dfrac{\mathrm{d}}{\mathrm{d}x}\varphi(x)\right) = \kappa A \dfrac{u(x)}{2(1+r)}\left(\varphi(x) - \dfrac{\mathrm{d}}{\mathrm{d}x}\, z(x)\right), \end{cases} \tag{5.1}$$

where r, κ, A, I are physical constants. Following [4], we consider the inverse problem of estimating the Young's modulus $u(x)$ from sparse observations of the displacement $z(x)$, where both u and z are functions from D to $\mathbb{R}$.

Let $\mathcal{F} : u \mapsto z$ be the solution map to equation (5.1). Let $\{s_i\}_{i=1}^m \subset D$ be the locations of the observation sensors, leading to the observation operator $\mathcal{O} : z \mapsto y \in \mathbb{R}^m$ defined coordinate-wise by

$$O_i(z) := \int_0^L z\varphi_i \, \mathrm{d}x, \quad \varphi_i(x) := \frac{1}{\gamma_i} \exp\left(-(s_i - x)^2/(2\delta^2)\right), \quad 1 \leqslant i \leqslant m,$$

where $\delta = 10^{-4}$ and γ_i is the normalizing constant such that $\int_0^L \varphi_i \, \mathrm{d}x = 1$. Data are generated according to the model

$$y = \mathcal{O} \circ \mathcal{F}(u) + \eta =: \mathcal{G}(u) + \eta,$$

where η denotes the observation error, which is assumed to follow a Gaussian distribution $N(0, \gamma_{\text{obs}}^2 I)$. Notice that for system (5.1) with proper boundary conditions specified at $x = 0$, the displacement $z(x^\star)$ at any point $0 < x^\star < L$ depends only on the values $\{u(x) : x < x^\star\}$. Thus, we expect suitable discretizations of the forward model to refine finely only the region $\{0 < x < s_m\}$, where s_m is the right-most observation location. We will discuss this in detail in section 5.1.2.

5.1.1. Forward discretization. To solve system (5.1) we employ a finite difference method. A family of numerical solutions can be parameterized by the set

$$\mathcal{A} := \left\{ a = (k, \theta) : k \in \mathcal{K} \subset \{1, 2, \dots\}, \quad \theta = [x_1, \dots, x_k] \in [0, L]^k \right\},$$

where k is the number of grid points and θ are the grid locations. Precisely, for $a = (k, \theta) \in \mathcal{A}$, we first reorder θ so that

$$0 =: x_0 \leqslant x_1 \leqslant \dots \leqslant x_k \leqslant x_{k+1} := L$$

and we let

$$\mathcal{F}^a : u \mapsto z^a$$

be the linearly interpolated explicit Euler finite difference solution to (5.1), discretized using the ordered grid θ. We also discretize the observation operator $\mathcal{O}$ using an Euler forward method, defined by

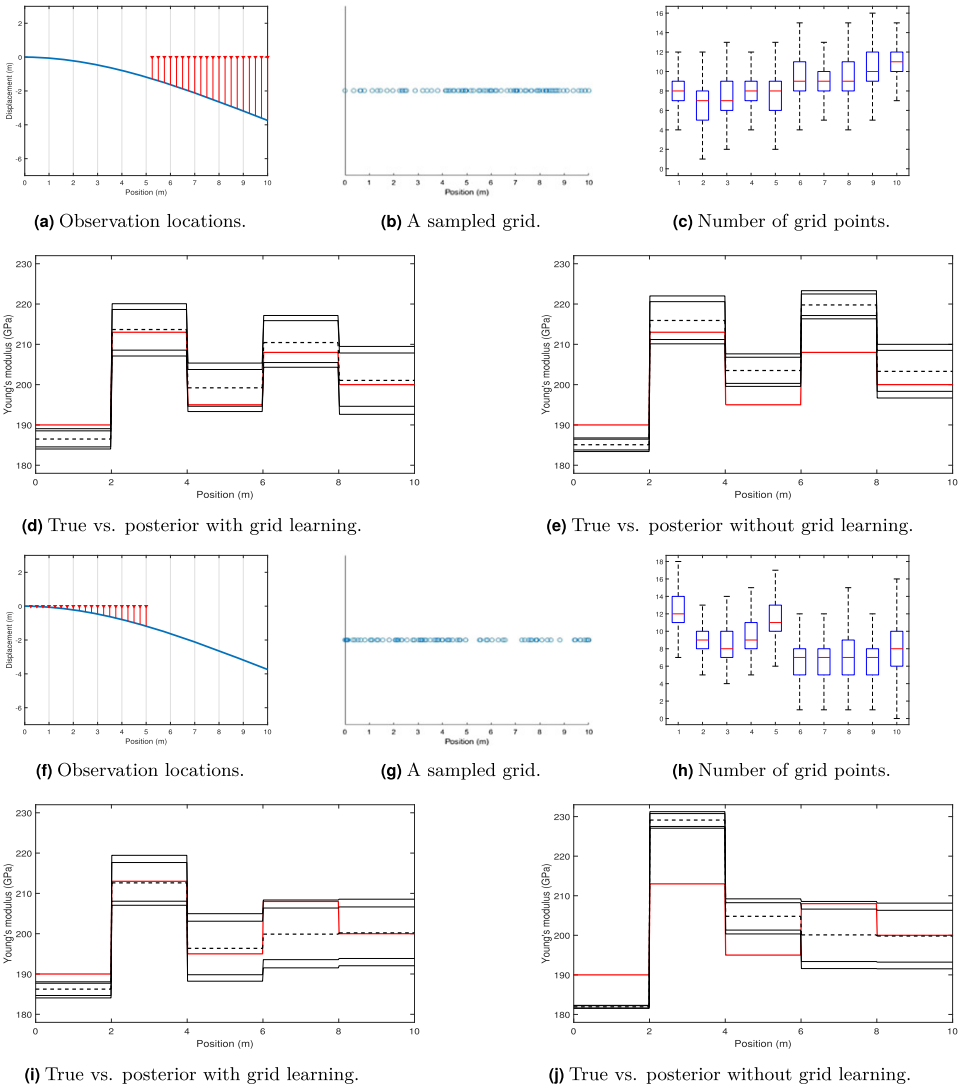
$$\mathcal{O}_i^a(z^a) = \sum_{j=0}^{k} z^a(x_j)\varphi_i(x_j)(x_{j+1} - x_j).$$

Finally $\mathcal{G}$ is approximated by $\mathcal{G}^a := \mathcal{O}^a \circ \mathcal{F}^a$.

5.1.2. Implementation details and numerical results. For our numerical experiments we consider a beam of length $L = 10\,\text{m}$, width $w = 0.1\,\text{m}$ and thickness $h = 0.3\,\text{m}$. We use a Poisson ratio $r = 0.28$ and Timoshenko shear coefficient $\kappa = 5/6$. $A = wh$ represents the cross-sectional area of the beam and $I = wh^3/12$ is the second moment of inertia. We run a virtual experiment of applying a point mass of 5 kg at the end of the beam, as seen in blue in figures 1(a) and 2(a). We assume that the observations are gathered with error $\gamma_{\text{obs}}^2 = 10^{-3}$.

We first assume that the beam is made of 5 segments of different kinds of steel, each of length 2 m, with corresponding Young's moduli $u^* = \{u_i^*\}_{i=1}^5 = \{190, 213, 195, 208, 200\,\text{GPa}\}$. The prior on $u \in \mathcal{U} = \mathbb{R}^5$ is given by $p_u(u) = \mathcal{N}(u; 2001, 25I_5)$ where $\mathbf{1}$ denotes the all-ones vector. For this case we assume that the number of grid points k is fixed to be $k = 85$, i.e., the prior $p_k(k)$ is a point mass. The grid locations θ are assumed to be *a priori* uniformly distributed in $[0, L]^k$. Results are reported in figure 1. We next assume that the Young's modulus $u(x) \in \mathcal{U} = C([0, L]; \mathbb{R})$ varies continuously with x. We set a Gaussian process prior on u defined by $\mu_u = \mathcal{GP}(200, c)$ with $c(x, x') = 50\ \exp\left(-(x - x')^2/0.5\right)$. For this case we assume that the prior on the number of grid points k follows a Poisson distribution with mean 60, i.e., $\nu_k(k) = \text{Poisson}(60)$, and the grid locations θ still have a uniform prior given k. The true Young's modulus underlying the data and the reconstruction results are reported in figure 2. Sampling is performed, both in the discrete and continuous settings, updating u and a alternately for a total number of $N = 1.2 \times 10^5$ iterations, with $\beta = 0.08$, $\zeta = 0.5$.

In both figures 1 and 2 two settings are considered. In the first one observations are concentrated on the right side of the beam and in the second on the left. For reference, figure 3 shows idealized posteriors considered, obtained with *very* fine discretizations $k = 500$, for each of the settings. Notice that for system (5.1) with proper boundary conditions specified at $x = 0$, the displacement $z(x_0)$ at any point $0 < x_0 < L$ depends only on the values $u(x)$ of Young's moduli with $x < x_0$. This implies that when observations are gathered on the left side of the beam, the posterior on $u(x)$ agrees with the prior on the right-side, and no resources should be



(a) Observation locations. **(b)** A sampled grid. **(c)** Number of grid points.

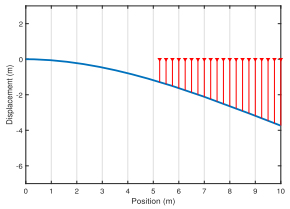
(d) True vs. posterior with grid learning. **(e)** True vs. posterior without grid learning.

(f) Observation locations. **(g)** A sampled grid. **(h)** Number of grid points.

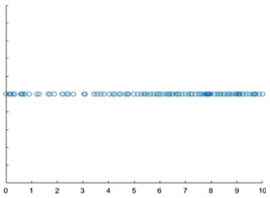
(i) True vs. posterior with grid learning. **(j)** True vs. posterior without grid learning.

Figure 1. Reconstruction of a piece-wise constant Young's modulus. Two settings for the observation locations are considered, shown in figures (a) and (f). For each setting, figures (b) and (g) show one sample from the marginal distribution $q_{a|y}(a)$ simulated by MCMC. Figures (c) and (h) report box-plots with the number of grid points that fall in each subinterval $[i-1, i]$, $i = 1, \ldots, 10$. Figures (d) and (i) show the mean (dashed black) and the 5, 10, 90, 95-percentiles (thin black) of the marginal $q_{u|y}(u)$, versus the true value (red), with data-driven forward discretization. Figures (e) and (j) show the same results with a fixed uniform-grid discretization.

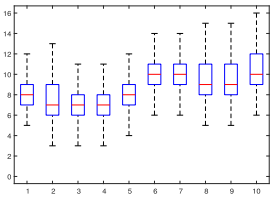
on discretizing the forward map on that region. In that case our adaptive data-driven discretizations are strongly concentrated on the left, as shown in figures 1(g) and (h) and 2(b) and 2(c). However, when observations are gathered on the right side of the beam, the data is informative on $u(x)$ for all $0 < x < L$. In such case, figures 1(b) and (c) and 2(b) and 2(c) show that the data-driven discretizations are concentrated on the right, but less heavily so. See tables B.1 and B.2 in the appendix B for a more detailed description of the grid points distribution in both



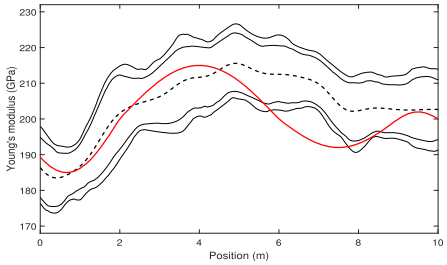
(a) Observation locations.



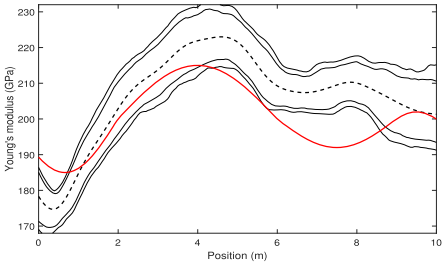
(b) A sampled grid.



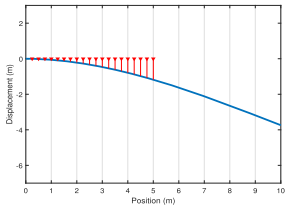
(c) Number of grid points.



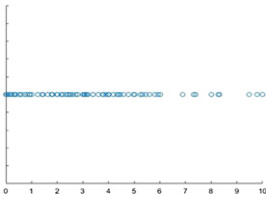
(d) True vs. posterior with grid learning.



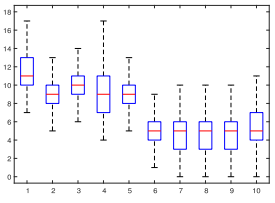
(e) True vs. posterior without grid learning.



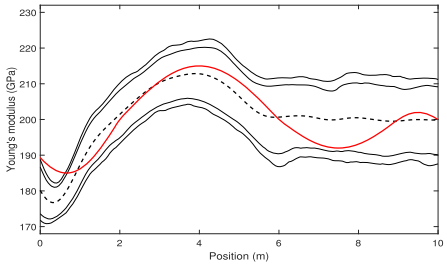
(f) Observation locations.



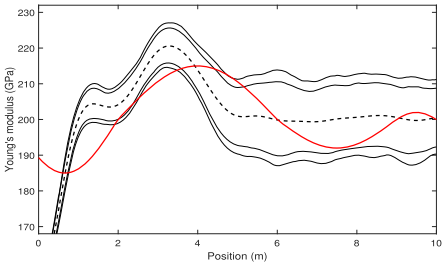
(g) A sampled grid.



(h) Number of grid points.



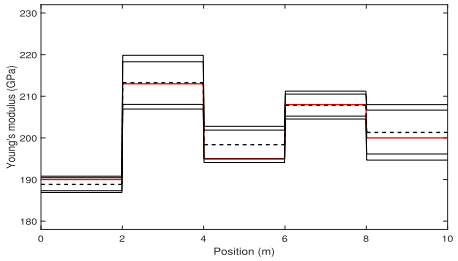
(i) True vs. posterior with grid learning.



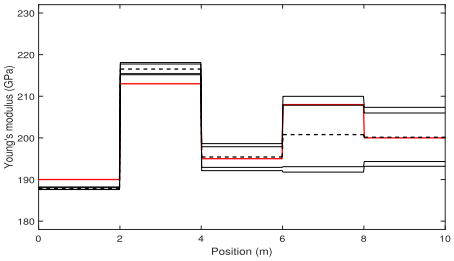
(j) True vs. posterior without grid learning.

Figure 2. Reconstruction of a continuous Young's modulus. Two settings for the observation locations are considered, shown in figures (a) and (f). For each setting, figures (b) and (g) show one sample from the marginal distribution $q_{a|y}(a)$ simulated by MCMC. Figures (c) and (h) report box plots with number of grid points that fall in each subinterval $[i-1, i]$, $i = 1, \ldots, 10$. Figures (d) and (i) show the mean (dashed black) and the 5, 10, 90, 95-percentiles (thin black) of the marginal $q_{u|y}(u)$, versus the true value (red), with data-driven forward discretization. Figures (e) and (j) show the same results with a fixed uniform-grid discretization.

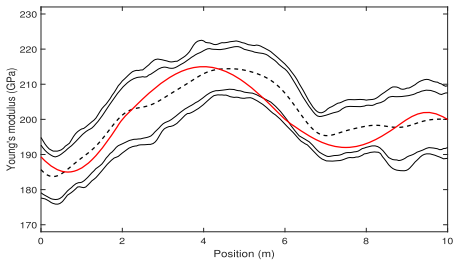
cases. Also, our results indicate that using data-driven discretizations will lead to a better estimation of the true Young's modulus, compared to fixed-grid discretizations. Additional results in the continuous Young's modulus setting are provided in the appendix. See table B.3 and figure 8 for the averaged acceptence probability for u and a, and history of MCMC samples



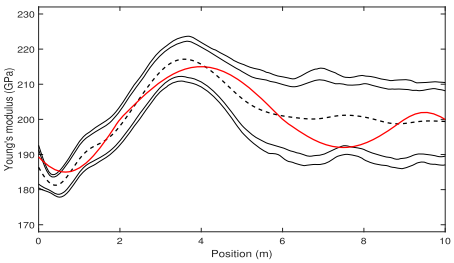
(a) Piece-wise constant, right observations.



(b) Piece-wise constant, left observations.



(c) Continuous, right observations.



(d) Continuous, left observations.

Figure 3. Idealized posterior $p_{u|y}(u)$, with mean (dashed black) and the 5, 10, 90, 95-percentiles (thin black), versus the true value (red).

of the high-dimensional u at some fixed locations, indicating the stationarity of the Markov chain.

Let $(u^{(n)}, a^{(n)})$ be the output of the Gibbs sampling algorithm at iteration n. The *reconstruction error* is defined as follows:

$$e_r = \sqrt{\sum_{n=1}^{N} \left| \mathcal{G}^{a^{(n)}}(u^{(n)}) - \mathcal{G}(u) \right|^2}, \tag{5.2}$$

where $\mathcal{G}(u)$ is approximately calculated on a very fine grid. In figure 4 we plot the reconstruction error for the second experiment where the Young's moduli is continuous and observations are gathered on the right-side. With fixed-grid discretization, the reconstruction error is small where the discretization matches the observation points. With adaptive data-driven discretizations the grid points will adaptively match the observation points in order to produce less error.

5.2. Euler–Maruyama discretization of SDEs: a signal processing application

Let $f : \mathbb{R}^d \to \mathbb{R}^d$ be globally Lipschitz continuous and consider the SDE

$$\mathrm{d}z(t) = f(z)\mathrm{d}t + \mathrm{d}u, \quad 0 < t \leqslant T, \quad z(0) = 0, \tag{5.3}$$

where u denotes d-dimensional Brownian motion. We aim to recover u from observations of the solution z. We suppose that the observations $y = [y_1, \ldots, y_m]$ are given by

$$y_i = z(t_i) + \eta_i, \quad i = 1, \ldots, m, \tag{5.4}$$

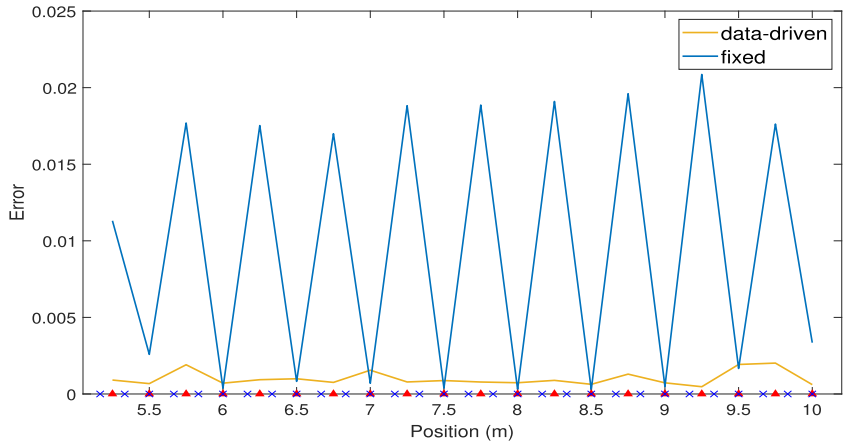


Figure 4. The reconstruction error with fixed-grid discretization (blue) and with data-driven grid discretization (orange). Red triangles are observations locations, while blue crosses are the grid points used in the fixed-grid discretization.

where $\eta = [\eta_1, \dots, \eta_m]$ is assumed to follow a centered Gaussian distribution with covariance Γ and

$$0 < t_1 < \cdots < t_m < T$$

are given observation times. Following [21], we cast the problem in the setting of section 2. First note that the solution to the integral equation

$$z(t) = \int_0^t f\,(z(s))\,\mathrm{d}s + u(t), \quad 0 \leqslant t \leqslant T, \tag{5.5}$$

defines a map

$$\mathcal{F} : C([0,T], \mathbb{R}^d) \to C([0,T], \mathbb{R}^d) \tag{5.6}$$

$$u \mapsto z. \tag{5.7}$$

Thus we set the input and output space to be $\mathcal{U} = \mathcal{Z} = C([0,T], \mathbb{R}^d)$.

Next we define an observation operator

$$\mathcal{O} : C([0,T], \mathbb{R}^d) \to \mathbb{R}^m$$

$$z \mapsto [z(s_1), \dots, z(s_m)]$$

and set $\mathcal{G} = \mathcal{O} \circ \mathcal{F}$. We put as prior on u the standard d-dimensional Wiener measure, that we denote μ_u. Then the posterior distribution $\mu_{u|y}$ is given by equation (2.4), which if $\Gamma = \gamma^2 I_m$ may be rewritten as

$$\frac{\mathrm{d}\mu_{u|y}}{\mathrm{d}\mu_u}(u) \propto \exp\left(\frac{1}{2\gamma^2}\sum_{i=1}^{m} |y_i - \mathcal{G}(u)(s_i)|^2\right). \tag{5.8}$$

Note that the likelihood does not involve evaluation of $\mathcal{F}(u)$ at times $t > s_m$, and hence changing the definition of $\mathcal{F}(u)(t)$ for $t > s_m$ does not change the posterior measure. Thus, we expect

suitable discretizations of the forward model to refine finely only times up to the right-most observation.

5.2.1. Forward discretization. For most nonlinear drifts f, the integral equation (5.6) cannot be solved in closed form and a numerical method needs to be employed. A family of numerical solutions may be parameterized by the set

$$\mathcal{A} := \left\{ a = (k, \theta) : k \in \mathcal{K} \subset \{1, 2, \ldots\}, \quad \theta = [t_1, \ldots, t_k] \in [0, T]^k \right\}.$$

Precisely, for $a = (k, \theta) \in \mathcal{A}$ we define an approximate, Euler–Maruyama solution map

$$\begin{aligned} \mathcal{F}^a : C([0, T], \mathbb{R}^d) &\to C([0, T], \mathbb{R}^d) \\ u &\mapsto z^a \end{aligned}$$

as follows. First, we reorder the t_j's so that

$$0 =: t_0 \leqslant t_1 \leqslant \ldots \leqslant t_k \leqslant t_{k+1} := T.$$

Then we define $z^a_j := z^a(t_j)$ as $z^a_0 = 0$, and

$$z^a_{j+1} = z^a_j + (t_{j+1} - t_j) f(z^a_j) + u(t_{j+1}) - u(t_j), \quad 1 \leqslant j \leqslant k. \tag{5.9}$$

Finally, for $t \in (t_j, t_{j+1})$ we define $z^a(t)$ by linear interpolation of z^a_j and z^a_{j+1}.

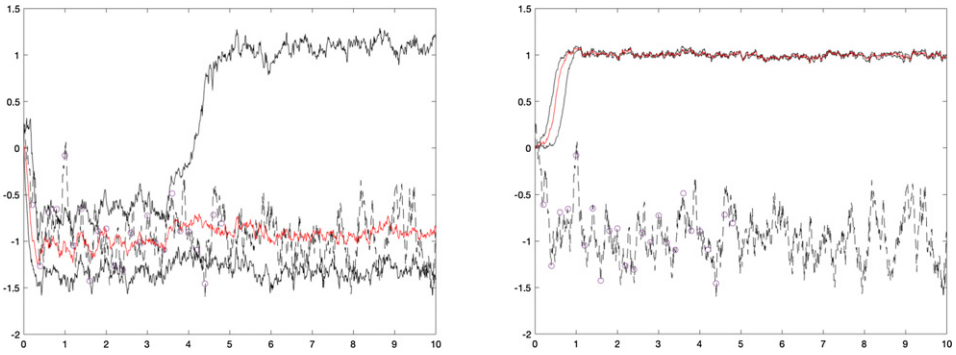
Having defined the parameter space $\mathcal{A}$ we now describe a choice of prior distribution ν_a on $\mathcal{A}$ and the resulting combined prior $\nu_{u,a}$ on $(u, a) \in \mathcal{U} \times \mathcal{A}$. First we choose a prior ν_k on the number k of grid-points. Given the number of grid points k, assuming no knowledge on appropriate discretizations for the SDE (5.3) we put a uniform prior on grid locations $\theta = [t_1, \ldots, t_k]$.

Remark 5.1. More information could be put into the prior. In particular it seems natural to impose that grids are finer at the beginning of the time interval.

5.2.2. Implementation details and numerical results. For our numerical experiments we considered the SDE (5.3) with $T = 10$ and double-well drift

$$f(t) = 10t(1 - t^2)/(1 + t^2). \tag{5.10}$$

We generated synthetic observation data y by solving (5.3) on a very fine grid, and then perturbing the solution at uniformly distributed times $t_i = 0.2i, i = 1, \ldots, 24 = m$ so that the last observation corresponds to time $t = 4.8$. The observation noise was taken to uncorrelated, $\Gamma = \gamma^2 I_m$, with $\gamma = 0.1$. The motivation for choosing this example is that there is certain intuition as to where one would desire the discretization grid-points to concentrate. Indeed, since all the observations t_i are in the interval $[0.2, 4.8]$ it is clear from equation (5.8) that any discretization points $t_j \in (4.8, 10]$ will not contribute to better approximate $\mu_{u|y}$. In other words, those grid points would help in approximating $\mathcal{F}$ but not in approximating $\mathcal{G} = \mathcal{O} \circ \mathcal{F}$.



(a) Data-driven discretization. **(b)** Uniform discretization.

Figure 5. Recovered SDE trajectory in the time-interval $t \in [0, 10]$. The true trajectory is shown in dashed black line. The posterior median is shown in red, and 5 and 95-percentiles are shown in black. The small circles denote the locations of the observations.

We report our results for a small grid-size $k = 24$. Similar but less dramatic effect was seen for larger grid size. Precisely, we chose our set of admissible grids to be given by

$$[t_0, \ldots, t_{25}] : 0.01 = t_0 \leqslant t_1 \leqslant \ldots \leqslant t_{25} = 10.$$

For implementation purposes, elements in the space $\mathcal{U} = C([0, T], \mathbb{R})$ were represented as vectors in $\mathbb{R}^{1000}$ containing their values on a uniform grid of step-size 0.1. We run these algorithms with parameter choices $N = 10^5$, $\beta = 0.1$, $\zeta = 0.5$.

The experiments show a successful reconstruction of the SDE path. Moreover, the grids concentrate in $[0, 4.8]$ in agreement with our intuition and the uncertainty quantification is satisfactory. In contrast, we see that when using the same number of grid points but on a uniform grid the Euler–Maruyama scheme is unstable, leading to a collapse of the MCMC algorithm. Then, the posterior constructed with a uniform grid completely fails at reconstructing the SDE path, and the uncertainty quantification is overoptimistic due to poor mixing of the chain (figure 5).

5.3. *Finite element discretization: source detection*

Consider the boundary value problem

$$\begin{cases} -\Delta z(x) = \delta(x - u), & x \in D, \\ z(x) = 0, & x \in \partial D, \end{cases} \tag{5.11}$$

where $D = (0, 1) \times (0, 1) \subset \mathbb{R}^2$ is the unit square and δ is the Dirac function at the origin. We aim to recover the source location u from sparse observations

$$y_i = z(s_i) + \eta_i, \quad i = 1, \ldots, m, \tag{5.12}$$

where $\eta = [\eta_1, \ldots, \eta_m]$ follows a centered Gaussian distribution with covariance Γ and $s_1, \ldots, s_m \in D \backslash \{u\}$ are observation locations. To cast the problem in the setting of section 2, we let $\mathcal{F}$ be given by Green's function for the Laplacian on the unit square (which does not

admit an analytical formula but can be computed e.g. via series expansions [31]) and $\mathcal{O}$ be defined by point-wise evaluation at the observation locations. The prior on u is the uniform distribution in the unit square D, which we denote $p_u(u)$. Since $\mathcal{U} = D$ is finite dimensional, the posterior has Lebesgue density given by equation (2.3). We find this problem to be a good test, as there is a clear understanding that the data-driven mesh should concentrate around the source.

5.3.1. Forward discretization. To solve equation (5.11) numerically we employ the finite element method. The use of uniform grid is here wasteful, as the mesh should ideally concentrate around the unknown source u.

We will use grids obtained as the Delauny triangulation of central Voronoi tessellations $\{V_i\}_{i=1}^k$ and generators $\{x_i\}_{i=1}^k$, where each $x_i \in D$ and $V_i \subset D$. This can be calculated as the solution of the optimization problem, parameterized by a probability density ρ on D:

$$\min_{\{x_i\}\subset D,\{V_i\}} \sum_{i=1}^{k} \int_{V_i} \rho(x)\|x - x_i\|^2 \, \mathrm{d}x, \tag{5.13}$$

subject to the constraint that $\{V_i\}_{i=1}^k$ is a tessellation of D. One can refer to [12] for more details. For a fixed density ρ and integer k we denote the optimal grid points by $\{x_{\rho,i}\}_{i=1}^k$. Then the approximated solution map is defined as

$$\mathcal{F}^a : D \to H_0^1(D) \tag{5.14}$$

$$u \mapsto z^a \tag{5.15}$$

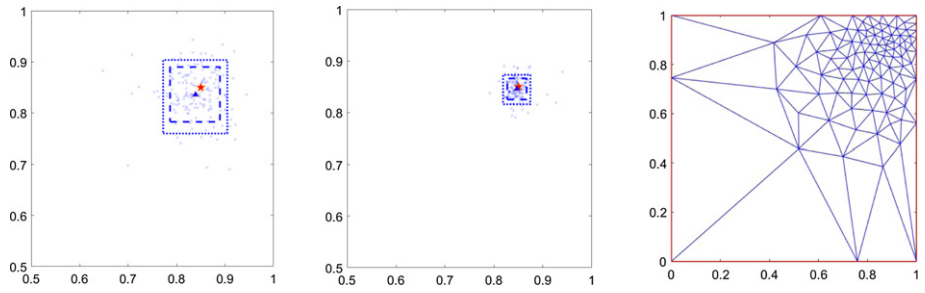
where z^a is given by the finite element solution of equation (5.11) with respect to (the Delauny triangulation of) the grid points $\{x_{\rho,i}\}_{i=1}^k$. Details on the creation of grid for prescribed parameters ρ and k will be discussed below.

In the spirit of adapting the grid to favor the ones maximizing the model evidence, we constrain ρ to belong to a family of parametric densities $\Pi = \{\rho(x;\theta)|\theta \in \mathbb{R}^P\}$ where $\rho(x;\theta) = \mathrm{Beta}(\alpha_1,\beta_1) \times \mathrm{Beta}(\alpha_2,\beta_2)$ is the product measure of two Beta distributions. Therefore in this case $\theta = (\alpha_1,\beta_1,\alpha_2,\beta_2)$ and $P = 4$. Each pair (k,θ) describes a member in the discretization family $\mathcal{A}$, where k controls the number of grid points, while θ controls how these grid points are distributed in the spatial dimension.

5.3.2. Implementation details and numerical results. We solved equation (5.11) on a fine grid $k = 2000$ with the true point source $u^* = (0.85, 0.85)$. The observation locations were $\{s_1,\ldots,s_{25}\} = \{0.5,0.6,0.7,0.8,0.9\} \times \{0.5,0.6,0.7,0.8,0.9\}$. Observation noise was uncorrelated with $\Gamma = \gamma^2 I_{25}$, $\gamma = 0.05$.

In this example the prior of u is the uniform distribution on $D = (0,1) \times (0,1)$ and u is initialized at $(0.2, 0.2)$. The parameters θ are also set to have a uniform prior $\theta = (\alpha_1,\beta_1,\alpha_2,\beta_2) \sim \mathrm{Uniform}\left([1,10]^4\right)$. We initialize $(\alpha_1,\beta_1,\alpha_2,\beta_2) = (1,1,1,1)$, which corresponds to (near) uniform grid in D. For simplicity, in this experiment we set a point mass prior on k, with $k = 100$. We run the algorithm with $N = 10^4$.

We compare our algorithm to the traditional method where we fix a uniform grid in D and run the MCMC algorithm only on u. We found out that with the same number of grid points, our data-driven approach gives a posterior distribution $q_{u|y}$ that is more concentrated around the true location of the point source, as shown in figures 6(a) and (b). Also, figure 6(c) shows that the adaptive discretization is concentrated at the top right corner of the region, where the hidden point source u^* is located.



(a) Fixed grid. **(b)** Data-driven grid. **(c)** Grid generated in the last iteration.

Figure 6. Figures (a) and (b) show the posterior distribution $q(u|y)$, where the grid is fixed and uniform in (a), and data-driven in (b). Red star indicates the true location of the source, blue dots are random samples from the posterior, blue triangle is the posterior mean, and dash (resp. dotted) lines correspond to the 90% (resp. 95%) coordinate-wise credible regions. Figure (c) shows the grid generated in the last iteration of the MCMC update.

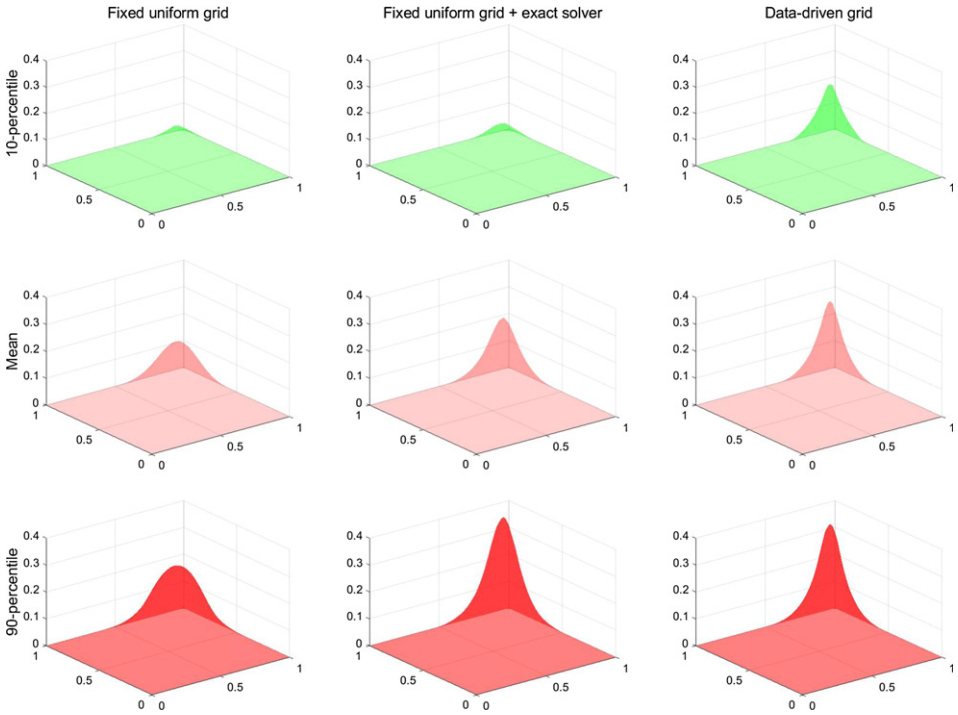


Figure 7. The mean, 10 and 90-percentile of the pushforward distribution $\mathcal{F}_\sharp(q_{u|y})$ under three different settings: (1) both the posterior $q_{u|y}$ and its pushforward $\mathcal{F}_\sharp(q_{u|y})$ are computed on a fixed and uniform grid; (2) the posterior $q_{u|y}$ is computed on a fixed and uniform grid, and its pushforward $\mathcal{F}_\sharp(q_{u|y})$ is calculated using a (nearly) exact solver; (3) both the posterior and its pushforward are computed on a data-driven grid.

Next we show that data-driven discretizations of the forward map can be employed to provide improved uncertainty quantification of the PDE solution, and not only to better reconstruct the unknown input. To illustrate this, we approximate the pushforward distribution $\mathcal{F}_\sharp(q_{u|y})$ in three different ways, as shown in figure 7. Let $(u^{(n)}, a^{(n)})$ denote the output of the Gibbs sampling algorithm at iteration n. We first consider the traditional method where the grid is fixed and uniform, that is, $a^{(n)} = a$ is fixed. Then the pushforward distribution can be well-approximated by $\{\mathcal{F}^a(u^{(n)})\}_{n=1}^N$, for N large enough. We then consider the same setting except that $\mathcal{F}^a$ is replaced by a forward map $\mathcal{F}$ computed in a fine grid $k = 2000$ and the pushforward is approximated by $\{\mathcal{F}(u^{(n)})\}_{n=1}^N$. Finally we consider a data-driven setting stemming from our algorithm, where the pushforward distribution is approximated by $\{\mathcal{F}^{a^{(n)}}(u^{(n)})\}_{n=1}^N$. We see that our algorithm reconstructs well the solution to the PDE, with a more accurate mean and a smaller variance.

6. Conclusions and open directions

- We have shown that, in a variety of inverse problems, the observations contain useful information to guide the discretization of the forward model, allowing a better reconstruction of the unknown than using uniform grids with the same number of degrees of freedom. Despite these results being promising, it is important to note that updating the discretization parameters may be costly in itself, and may result in slower mixing of the MCMC methods. For this reason, we envision that the proposed approach may have more potential when the computational solution of the inverse problem is very sensitive to the discretization of the forward map and discretizing it is expensive. We also believe that density-based discretizations may help in alleviating the cost of discretization learning.
- An interesting avenue of research stemming from this work is the development of prior discretization models that are informed by numerical analysis of the forward map $\mathcal{F}$, while recognizing the uncertainty in the best discretization of the forward model $\mathcal{G}$. Moreover, more sophisticated prior models beyond the product structure considered here should be investigated.
- Topics for further research include the development of new local proposals and sampling algorithms for grid-based discretizations, and the numerical implementation of the approach in computationally demanding inverse problems beyond the proof-of-concept ones considered here.

Acknowledgments

The work of NGT and DSA was supported by the NSF Grant DMS-1912818/1912802. The work of DB and YMM was supported by NSF Grant DMS-1723011.

Appendix A. Algorithm pseudo-code

See (algorithm A.1).

Appendix B. Additional results for section 5.1

See (tables B.1–B.3 and figure 8).

Algorithm A.1. Metropolis within-Gibbs.

Input parameters: β (pCN step-size), ζ (probability of location moves), N (sample size).
Choose $(u^{(1)}, a^{(1)}) \in \mathcal{U} \times \mathcal{A}$.
for $n = 1 : N$ **do**

Stage I. Do a pCN move to update u given a, y:

(i) Propose $\tilde{u}^{(n)} = \sqrt{1-\beta^2}\, u^{(n)} + \beta v^{(n)}, \quad v^{(n)} \sim \mu_u$.

(ii) Set $u^{(n+1)} = \tilde{u}^{(n)}$ with probability

$$a(u^{(n)}, \tilde{u}^{(n)}) = \min\left\{1, \exp\left(\Psi(u^{(n)}, a^{(n)}; y) - \Psi(\tilde{u}^{(n)}, a^{(n)}; y)\right)\right\}$$

(iii) Set $u^{(n+1)} = u^{(n)}$ otherwise.

Stage II. Update $a = (k, \theta)$ given u and y.
Stage IIa. With probability ζ, update θ given u, y with a grid re-location step:

(i) Propose $\tilde{a}^{(n)}$ by picking one of the k interior grid points of $a^{(n)}$ uniformly at random, and replacing it by a uniform draw in D.

(ii) Set $a^{(n+1)} = \tilde{a}^{(n)}$ with probability

$$\alpha(a^{(n)}, \tilde{a}^{(n)}) = \min\left\{1, \exp\left(\Psi(u^{(n+1)}, a^{(n)}; y) - \Psi(u^{(n+1)}, \tilde{a}^{(n)}; y)\right)\right\}.$$

(iii) Set $a^{(n+1)} = a^{(n)}$ otherwise.

Stage IIb. Otherwise, (with probability $1 - \zeta$) update k with a birth/death step:

(i) Propose a new number $\tilde{k}^{(n)}$ of grid-points.

(ii) If $\tilde{k}^{(n)} \leqslant k^{(n)}$ remove uniformly chosen grid-points.

(iii) If $\tilde{k}^{(n)} > k^{(n)}$ draw required number of new grid points uniformly at random in D.

(iv) Set $a^{(n+1)} = \tilde{a}^{(n)}$ with probability

$$\alpha(a^{(n)}, \tilde{a}^{(n)}) = \min\left\{1, \tfrac{\nu_k(\tilde{k}^{(n)})}{\nu_k(k^{(n)})} \ \exp\left(\Psi(u^{(n+1)}, a^{(n)}; y) - \Psi(u^{(n+1)}, \tilde{a}^{(n)}; y)\right)\right\}.$$

(v) Set $a^{(n+1)} = a^{(n)}$ otherwise.

end for

Table B.1. Distribution of grid points when observations are concentrated on the right, in the piecewise-constant Young' modulus case. Element on *i*th row and *j*th column represents the posterior probability of having *i* grid points in the subinterval *j*.

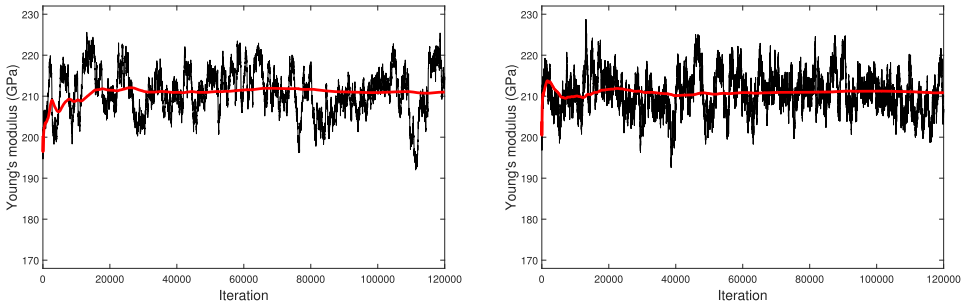
	(0, 1)	(1, 2)	(2, 3)	(3, 4)	(4, 5)	(5, 6)	(6, 7)	(7, 8)	(8, 9)	(9, 10)
2–3	0.0015	0.0047	0.0116	0.0094	0.0164	0	0	0	0	0
4–5	0.0791	0.1089	0.2003	0.1721	0.1875	0.0260	0.0967	0.0363	0.0554	0
6–7	0.3345	**0.3940**	**0.4002**	**0.4017**	**0.3956**	0.2309	0.2817	0.1789	0.3529	0.0578
8–9	**0.3956**	0.3412	0.2785	0.2967	0.2907	**0.3548**	**0.3011**	**0.4835**	**0.3944**	0.3808
10–11	0.1548	0.1268	0.0920	0.0983	0.0924	0.2579	0.2612	0.2565	0.1635	**0.3886**
12–13	0.0289	0.0230	0.0153	0.0199	0.0159	0.0988	0.0550	0.0425	0.0312	0.1524
14–15	0.0052	0.0012	0.0016	0.0019	0.0016	0.0255	0.0044	0.0020	0.0027	0.0195
16–17	0.0004	0.0003	0.0004	0	0	0.0051	0	0.0002	0	0.0009
18–19	0	0	0	0	0	0.0009	0	0	0	0

Table B.2. Distribution of grid points when observations are concentrated on the left, in the piecewise-constant Young' modulus case. Element on *i*th row and *j*th column represents the posterior probability of having *i* grid points in the subinterval *j*.

	(0, 1)	(1, 2)	(2, 3)	(3, 4)	(4, 5)	(5, 6)	(6, 7)	(7, 8)	(8, 9)	(9, 10)
2–3	0	0	0	0	0	0.1712	0.1481	0.1435	0.1259	0.0605
4–5	0	0.0496	0.0076	0	0.0325	**0.3420**	**0.3139**	**0.3183**	**0.3162**	0.2362
6–7	0	0.2488	0.1263	0.0257	0.2238	0.2860	0.2995	0.3133	0.3087	**0.3360**
8–9	0.0497	**0.3785**	0.3236	0.2174	**0.4048**	0.1341	0.1542	0.1508	0.1615	0.2316
10–11	0.2055	0.2390	**0.3357**	**0.4114**	0.2623	0.0383	0.0540	0.0483	0.0554	0.0954
12–13	**0.3384**	0.0717	0.1573	0.2492	0.0705	0.0062	0.0118	0.0097	0.0131	0.0274
14–15	0.2516	0.0111	0.0424	0.0789	0.0061	0.0012	0.0016	0.0009	0.0017	0.0053
16–17	0.1163	0.0013	0.0070	0.0155	0	0	0	0	0	0.0009
18–19	0.0321	0	0.0009	0.0018	0	0	0	0	0	0
20–21	0.0053	0	0	0	0	0	0	0	0	0
22–23	0.0007	0	0	0	0	0	0	0	0	0

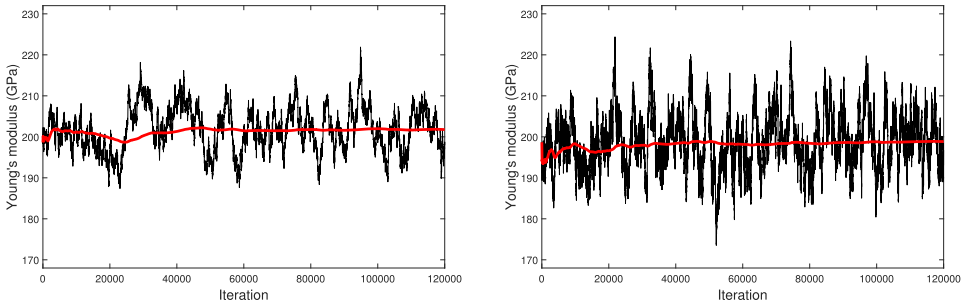
Table B.3. Averaged acceptance probability of *u* and *a* respectively, in the continuous Young' modulus case.

	Right obs.	Left obs.
u	0.2728	0.4590
a	0.1953	0.2314



(a) Continuous, right observations, $x = 4$.

(b) Continuous, left observations, $x = 4$.



(c) Continuous, right observations, $x = 8$.

(d) Continuous, left observations, $x = 8$.

Figure 8. History of MCMC samples (black line) and running sample averages (red line) of continuous Young's modulus $u(x)$, at fixed locations $x = 4$ and $x = 8$ repectively, suggesting stationarity of the Markov chain.

ORCID iDs

D Sanz-Alonso https://orcid.org/0000-0002-5022-864X

References

[1] Agapiou S, Papaspiliopoulos O, Sanz-Alonso D and Stuart A M 2017 Importance sampling: intrinsic dimension and computational cost *Stat. Sci.* **32** 405–31
[2] Becker R and Vexler B 2005 Mesh refinement and numerical sensitivity analysis for parameter calibration of partial differential equations *J. Comput. Phys.* **206** 95–110
[3] Beskos A, Roberts G, Stuart A and Voss J 2008 MCMC methods for diffusion bridges *Stoch. Dyn.* **08** 319–50
[4] Bigoni D, Zahm O, Spantini A and Marzouk Y M 2019 Greedy inference with layers of lazy maps (arXiv:1906.00031)
[5] Borcea L, Druskin V and Knizhnerman L 2005 On the continuum limit of a discrete inverse spectral problem on optimal finite difference grids *Commun. Pure Appl. Math.* **58** 1231–79
[6] Christen J A and Fox C 2005 Markov chain Monte Carlo using an approximation *J. Comput. Graph. Stat.* **14** 795–810
[7] Cockayne J, Oates C J, Sullivan T J and Girolami M 2019 Bayesian probabilistic numerical methods *SIAM Rev.* **61** 756–89
[8] Cotter S L, Roberts G O, Stuart A M and White D 2013 MCMC methods for functions: modifying old algorithms to make them faster *Stat. Sci.* **28** 424–46

[9] Cui T, Law K J H and Marzouk Y M 2016 Dimension-independent likelihood-informed MCMC *J. Comput. Phys.* **304** 109–37
[10] Cui T, Marzouk Y M and Willcox K E 2015 Data-driven model reduction for the Bayesian solution of inverse problems *Int. J. Numer. Method. Eng.* **102** 966–90
[11] Dellacherie C and Meyer P-A 2011 *Probabilities and Potentials: Potential Theory for Discrete and Continuous Semigroups* (Amsterdam: Elsevier)
[12] Du Q and Gunzburger M 2002 Grid generation and optimization based on centroidal Voronoi tessellations *Appl. Math. Comput.* **133** 591–607
[13] Efendiev Y, Hou T and Luo W 2006 Preconditioning Markov chain Monte Carlo simulations using coarse-scale models *SIAM J. Sci. Comput.* **28** 776–803
[14] Frangos M, Marzouk Y M, Willcox K and van Bloemen Waanders B 2010 Surrogate and reduced-order modeling: a comparison of approaches for large-scale statistical inverse problems *Large-Scale Inverse Problems and Quantification of Uncertainty* (New York: Wiley) pp 123–49 ch 7
[15] Garcia Trillos N, Kaplan Z, Samakhoana T and Sanz-Alonso D 2017 On the consistency of graph-based Bayesian learning and the scalability of sampling algorithms (arXiv:1710.07702)
[16] Trillos N G and Sanz-Alonso D 2017 The Bayesian formulation and well-posedness of fractional elliptic inverse problems *Inverse Problems* **33** 065006
[17] García Trillos N and Sanz-Alonso D 2018 Continuum limits of posteriors in graph bayesian inverse problems *SIAM J. Math. Anal.* **50** 4020–40
[18] Giles M B 2008 Multilevel Monte Carlo path simulation *Oper. Res.* **56** 607–17
[19] Green P J 1995 Reversible jump Markov chain Monte Carlo computation and Bayesian model determination *Biometrika* **82** 711–32
[20] Green P J and Mira A 2001 Delayed rejection in reversible jump Metropolis–Hastings *Biometrika* **88** 1035–53
[21] Hairer M, Stuart A M and Voss J 2011 Signal processing problems on function space: Bayesian formulation, stochastic PDEs and effective MCMC methods *The Oxford handbook of nonlinear filtering* pp 833–73
[22] Harlim J, Sanz-Alonso D and Yang R 2019 Kernel methods for Bayesian elliptic inverse problems on manifolds (arXiv:1910.10669)
[23] Kaipio J and Somersalo E 2006 *Statistical and Computational Inverse Problems* vol 160 (Berlin: Springer)
[24] Kaipio J and Somersalo E 2007 Statistical inverse problems: discretization, model reduction and inverse crimes *J. Comput. Appl. Math.* **198** 493–504
[25] Kennedy M C and O'Hagan A 2001 Bayesian calibration of computer models *J. R. Stat. Soc.* B **63** 425–64
[26] Li H, Garg V V and Willcox K 2018 Model adaptivity for goal-oriented inference using adjoints *Comput. Methods Appl. Mech. Eng.* **331** 1–22
[27] Li J and Marzouk Y M 2014 Adaptive construction of surrogates for the Bayesian solution of inverse problems *SIAM J. Sci. Comput.* **36** A1163–86
[28] Lieberman C, Willcox K and Ghattas O 2010 Parameter and state model reduction for large-scale statistical inverse problems *SIAM J. Sci. Comput.* **32** 2523–42
[29] Marzouk Y M, Najm H N and Rahn L A 2007 Stochastic spectral methods for efficient Bayesian solution of inverse problems *J. Comput. Phys.* **224** 560–86
[30] Marzouk Y M and Xiu D 2009 A stochastic collocation approach to Bayesian inference in inverse problems *Commun. Comput. Phys.* **6** 826–47
[31] Melnikov Y A and Melnikov M Y 2006 Computability of series representations for Green's functions in a rectangle *Eng. Anal. Bound. Elem.* **30** 774–80
[32] Peherstorfer B, Willcox K and Gunzburger M 2018 Survey of multifidelity methods in uncertainty propagation, inference, and optimization (arXiv:1806.10761)
[33] Rasmussen C E and Williams C K I 2006 *Gaussian Processes for Machine Learning* vol 1 (Cambridge, MA: MIT Press)
[34] Robert C 2007 *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation* (Berlin: Springer)
[35] Rudolf D and Sprungk B 2015 On a generalization of the preconditioned Crank–Nicolson metropolis algorithm *Found. Comput. Math.* **6** 309–43
[36] Sacks J, Welch W J, Mitchell T J and Wynn H P 1989 Design and analysis of computer experiments *Stat. Sci.* **4** 409–23

[37] Sanz-Alonso D, Stuart A M and Taeb A 2018 Inverse problems and data assimilation (arXiv:1810.06191)
[38] Schwab C and Zech J 2019 Deep learning in high dimension: neural network expression rates for generalized polynomial chaos expansions in UQ *Anal. Appl.* **17** 19–55
[39] Stuart A M 2010 Inverse problems: a Bayesian perspective *Acta Numer.* **19** 451–559
[40] Stuart A M and Teckentrup A 2017 Posterior consistency for Gaussian process approximations of Bayesian posterior distributions *Math. Comput.* **87** 721–53
[41] Tierney L and Mira A 1999 Some adaptive Monte Carlo methods for Bayesian inference *Stat. Med.* **18** 2507–15
[42] Xiu D and Karniadakis G E 2002 The Wiener–Askey polynomial chaos for stochastic differential equations *SIAM J. Sci. Comput.* **24** 619–44
[43] Zellner A 1988 Optimal information processing and Bayes's theorem *Am. Stat.* **42** 278–80