

PARAMETER AND STATE MODEL REDUCTION FOR LARGE-SCALE STATISTICAL INVERSE PROBLEMS*

CHAD LIEBERMAN[†], KAREN WILLCOX[†], AND OMAR GHATTAS[‡]

Abstract. A greedy algorithm for the construction of a reduced model with reduction in both parameter and state is developed for an efficient solution of statistical inverse problems governed by partial differential equations with distributed parameters. Large-scale models are too costly to evaluate repeatedly, as is required in the statistical setting. Furthermore, these models often have high-dimensional parametric input spaces, which compounds the difficulty of effectively exploring the uncertainty space. We simultaneously address both challenges by constructing a projection-based reduced model that accepts low-dimensional parameter inputs and whose model evaluations are inexpensive. The associated parameter and state bases are obtained through a greedy procedure that targets the governing equations, model outputs, and prior information. The methodology and results are presented for groundwater inverse problems in one and two dimensions.

Key words. model reduction, statistical inverse problems, Markov chain Monte Carlo, optimization

AMS subject classifications. 65K10, 37M99, 35J15

DOI. 10.1137/090775622

1. Introduction. Statistical inverse problems governed by partial differential equations (PDEs) with spatially distributed parameters pose a significant computational challenge for existing methods. While the cost of a repeated PDE solutions can be addressed by traditional model reduction techniques, the difficulty in sampling in high-dimensional parameter spaces remains. We present a model reduction algorithm that seeks low-dimensional representations of parameters and states while maintaining fidelity in outputs of interest. The resulting reduced model accelerates model evaluations and facilitates efficient sampling in the reduced parameter space. The result is a tractable procedure for the solution of statistical inverse problems involving PDEs with high-dimensional parametric input spaces.

Given a parameterized mathematical model of a certain phenomenon, the forward problem is to compute output quantities of interest for specified parameter inputs. In many cases, the parameters are uncertain, but they can be inferred from observations by solving an inverse problem. Inference is often performed by solving an optimization problem to minimize the disparity between model-predicted outputs and observations. Many inverse problems of this form are ill-posed in the sense that there may be many values of the parameters whose model-predicted outputs reproduce the observations. The set of parameters consistent with the observations may be larger still if we also admit noise in the sensor instruments. In the deterministic setting, a regularization term is often included in the objective function to make the problem well-posed. The

*Received by the editors November 2, 2009; accepted for publication (in revised form) June 29, 2010; published electronically August 19, 2010. This work was supported in part by the Department of Energy under grants DE-FG02-08ER25858 and DE-FG02-08ER25860, the Singapore-MIT Alliance Computational Engineering Programme, and the Air Force Office of Sponsored Research under grant FA9550-06-0271.

<http://www.siam.org/journals/sisc/32-5/77562.html>

[†]Massachusetts Institute of Technology, Cambridge, MA 02139 (celieber@mit.edu, kwillcox@mit.edu).

[‡]University of Texas at Austin, Austin, TX 78712 (omar@ices.utexas.edu).

form of the regularization is chosen to express preference for desired characteristics of the solution (e.g., smoothness).

While the regularized deterministic formulation leads to a single point estimate in parameter space, a statistical approach quantifies the relative likelihood of the observation-consistent parameters. The result is a probability density function over the parameters termed the *posterior distribution* [6, 25, 35]. Under assumptions on the probability distribution of sensor noise, the relative likelihood of observation-consistent parameters can be ascertained by Bayesian inference. In this setting we can also express a preference for solutions with certain characteristics in the *prior distribution*. The prior distribution expresses the relative likelihood of parameters independently of the observations. In this way we may include in the formulation any problem-specific knowledge outside of the mathematical model of the phenomena of interest.

In decision-making scenarios we require the evaluation of weighted integrals of the posterior over parameter space, e.g., mean and variance. For applications of interest where we may have millions of parameters, the associated integral computations cannot be performed analytically, nor can they be estimated by numerical quadrature. Instead, we compute approximations to the moments by generating samples from the posterior distribution and calculating the discrete analogues. Samples may be generated from an implicitly defined posterior by Markov chain Monte Carlo (MCMC) methods [2, 7, 10, 11, 17, 18, 19, 28, 30] whereby a Markov chain is established whose stationary distribution is the posterior.

The Metropolis–Hastings algorithm [11] is an MCMC method. At each step, a new sample is generated by proposing a candidate and then accepting or rejecting based on the associated Hastings ratio. Computation of the Hastings ratio requires one posterior evaluation, which, for applications of interest, corresponds to the numerical solution of a PDE. A repeated solution of a PDE is a prohibitively expensive task even for some simple model problems. In addition to the computational cost of the sampling process, an efficient sampler is difficult to design for high-dimensional parameter spaces. Applications of interest are parameterized by distributed field quantities, and when discretized, have dimensionality in the millions or more, putting them far beyond the reach of current MCMC methods.

In the present paper, we address challenges of sampling a high-dimensional parameter space and the cost of PDE solutions at each posterior evaluation by projection-based model reduction. We develop a reduced model with low-dimensional parametric input space and low-dimensional state space but whose outputs are accurate over the parameter range of interest. The reduction in state (and to a lesser degree, the parameter) accelerates PDE solutions and therefore posterior evaluations; reduction in parameter permits sampling in a much lower-dimensional space where traditional, nonadaptive Metropolis–Hastings samplers are effective and require little hand tuning.

Model reduction is the process by which one derives a low-dimensional, computationally inexpensive model that accurately predicts outputs of a high-fidelity, computationally costly model. Traditionally, the model is a map from an input space to a set of outputs through state space. Projection-based methods like moment-matching [12, 13], proper orthogonal decomposition [22, 34], and reduced basis methods [32] establish a low-dimensional subspace of the state space to which the reduced state is restricted.

Although model reduction is typically applied to state alone, we propose to reduce the parameter space as well. The extension of the projection-based model reduction to the parameter space enables the efficient solution of problems requiring the exploration

of a high-dimensional parameter space, e.g., in design optimization, distributed control problems, and statistical inverse problems. For the statistical inverse problem, we propose a reduced MCMC algorithm that samples in the reduced parameter space with a Metropolis–Hastings sampler and whose posterior evaluations are computed via the reduced model. As a result, we make tractable a class of statistical inverse problems where the forward model is a PDE with distributed parameters.

This paper is organized as follows. In section 2 we describe in detail the statistical inverse problem, emphasizing the probabilistic characterization of parameter space. Bayesian inference is introduced, and the inverse problem is formulated. We highlight the challenges of exploring the posterior using MCMC in the distributed-parameter PDE setting. Projection-based model reduction in parameter and state is presented in section 3 in anticipation of our reduced MCMC algorithm, which is described and analyzed in section 4. In section 5 we demonstrate reduced MCMC on one-dimensional (1-D) and two-dimensional (2-D) synthetic groundwater inverse problems. Finally, we make concluding remarks in section 6.

2. Statistical inverse problem. Let \mathcal{P} and \mathcal{Y} be parameter and output spaces, respectively, and consider the forward model $\mathcal{M} : \mathcal{P} \rightarrow \mathcal{Y}$. For the true parameter $\mathbf{p} \in \mathcal{P}$, the model predicts output $\mathbf{y} \in \mathcal{Y}$, and we make noisy observations $\mathbf{y}_d \in \mathcal{Y}$. The inverse problem consists of utilizing the noisy observations to infer the parameter. In the deterministic setting, this process involves regularization and optimization, and it results in a single-point estimate of the parameter with no measure of uncertainty. On the other hand, the Bayesian formulation of the statistical inverse problem yields a conditional probability density $\pi_{\mathbf{p}}(\mathbf{p}|\mathbf{y}_d)$ over parameter space from which one can compute an estimate and credibility interval.

2.1. Bayesian formulation. The statistical inverse problem is conveniently formulated as one of inference by exploiting Bayes’s rule. Define \mathbb{R}_0^+ as the set of nonnegative reals. Let $\gamma_{\mathbf{p}}(\mathbf{p}) : \mathcal{P} \rightarrow \mathbb{R}_0^+$ be the *prior* probability density over the parameter space. The prior expresses one’s knowledge of the probabilistic distribution of parameters before observations are made. Let $L(\mathbf{y}_d, \mathbf{p}) : \mathcal{Y} \times \mathcal{P} \rightarrow \mathbb{R}_0^+$ be the *likelihood* function. The likelihood embeds the map from parameter input to noisy observations by way of the forward model \mathcal{M} and a suitable error model. Provided the output space is bounded, we write the *posterior*

$$(2.1) \quad \pi_{\mathbf{p}}(\mathbf{p}|\mathbf{y}_d) \propto L(\mathbf{y}_d, \mathbf{p})\gamma_{\mathbf{p}}(\mathbf{p}),$$

which expresses our updated knowledge of the probabilistically observation-consistent parameters.

The prior $\gamma_{\mathbf{p}}(\mathbf{p})$ is selected by incorporating problem-specific information. Akin to regularization of deterministic inverse problems, the prior can be used to represent a preference for certain types of parameters. Let $\mathcal{N}(\boldsymbol{\mu}, \mathbf{G})$ be the multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix \mathbf{G} . In the present work, we employ a Gaussian process prior with a Gaussian kernel to preferentially treat smooth parameter fields, i.e., $\gamma_{\mathbf{p}}(\mathbf{p}) \sim \mathcal{N}(\mathbf{0}, \mathbf{S})$ where the ij th element of the covariance matrix is given by

$$(2.2) \quad S_{ij} = a \exp \left\{ \frac{-\|\vec{x}_i - \vec{x}_j\|_2^2}{2b^2} \right\} + c\delta_{ij},$$

a formulation that expresses correlation between the discretized parameter at \vec{x}_i and \vec{x}_j according to their Euclidian separation distance. Here, δ_{ij} is the Kronecker delta

and a , b , and c are positive scalar parameters of the kernel. If discontinuities are to be admitted, an outlier process (e.g., gamma or inverse gamma distribution) should be chosen instead. In problems for which no expertise can be drawn on, a uniform prior is usually employed [15], although a maximum entropy principle may be more consistent [24].

The likelihood function establishes the relationship between observations \mathbf{y}_d and model-predicted output $\mathbf{y}(\mathbf{p})$. It is convenient, and often representative of the physical process, to consider an additive error model $\mathbf{y}_d = \mathbf{y}(\mathbf{p}) + \mathbf{e}$, where \mathbf{e} is the output error usually associated with sensor measurements. Furthermore, we often assume the errors are unbiased, uncorrelated, and normally distributed with variance σ^2 , i.e., $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. In this work we do not consider the uncertainty associated with our model's potentially inadequate representation of the physical system. These assumptions result in the likelihood function

$$(2.3) \quad L(\mathbf{y}_d, \mathbf{p}) = \exp \left\{ \frac{-\|\mathbf{y}_d - \mathbf{y}(\mathbf{p})\|_2^2}{2\sigma^2} \right\}.$$

For these particular choices of the prior and likelihood, we obtain the posterior

$$(2.4) \quad \pi_{\mathbf{p}}(\mathbf{p} | \mathbf{y}_d) \propto \exp \left\{ \frac{-\|\mathbf{y}_d - \mathbf{y}(\mathbf{p})\|_2^2}{2\sigma^2} - \frac{1}{2} \|\mathbf{p}\|_{\mathbf{S}^{-1}}^2 \right\},$$

where $\|\mathbf{p}\|_{\mathbf{S}^{-1}}^2 = -2 \log \gamma_{\mathbf{p}}(\mathbf{p})$. Note that under the aforementioned assumptions, the solution to the statistical inverse problem, i.e., the posterior $\pi_{\mathbf{p}}(\mathbf{p} | \mathbf{y}_d)$, is known up to a normalizing constant. If the parameter space is very low-dimensional, approximations to moments of the posterior could be computed by quadrature. For distributed parameters, this is not feasible; instead, samples must be generated from the posterior indirectly by MCMC, and moments must be estimated by their discrete analogues.

2.2. MCMC. MCMC was first introduced by Metropolis et al. [30] and was later generalized by Hastings [19]. A Markov chain with the posterior as its stationary distribution is constructed via a random walk. A transition from one state to the next in the chain is achieved by generating a candidate from the proposal distribution. The proposal is accepted with certain probability and rejected otherwise. In the Metropolis–Hastings algorithm, the proposal distribution is subject to very mild restrictions—any proposal distribution yielding an ergodic Markov chain is acceptable and automatically has the target as its stationary distribution when the acceptance ratio is defined appropriately. This generalization opened the door to adaptive algorithms.

Adaptive methods include those that use the chain to modify the proposal distribution [17, 18] and adaptive direction samplers [7, 11] that maintain multiple points in parameter space. While adaptive methods speed convergence by more efficiently sampling the parameter space, other methods accelerate the posterior evaluations required to compute the Hastings ratio at each step. Arridge et al. recently proposed mesh-coarsening for solving the linear inverse problem [2]. They utilize the Bayesian formulation to quantify the statistics of the error associated with discretization. Polynomial chaos expansions have also been used in this context [28]. The associated stochastic spectral methods are used to obtain a surrogate for the posterior which can be evaluated by computing the terms in a series. The number of terms in this series, however, scales exponentially with the number of parameters; therefore, stochastic spectral methods have only been proven for inverse problems with a handful of parameters [4]. Efendiev, Hou, and Luo introduced a preconditioned MCMC in which

coarse grid solutions of the underlying PDE were used in a two-stage process to guide sampling to reduce the number of full-scale computations [10]. Their procedure relies on a Karhunen–Loève expansion in parameter space to reduce dimensionality.

We are not aware of instances of MCMC algorithms scaling to even thousands of parameter dimensions for general posteriors. Many problems of interest are three-dimensional and require multiscale resolution; therefore, the discretized parameter input space will typically have dimension in the millions. In high-dimensional parameter spaces, sufficient exploration and maintaining a minimum acceptance rate have proven challenging for even the most adaptive MCMC schemes. On the other hand, samplers can be efficiently tuned in low-dimensional parameter spaces. We will exploit the structure of our problem to systematically identify a parametric subspace on which to run the MCMC process.

The two computational challenges—sampling in high-dimensional parameter space and costly forward model evaluations—are addressed simultaneously by parameter and state model reduction, as we now describe.

3. Parameter and state model reduction. In this section, we classify the large-scale models of interest and present an algorithm to construct a parameter- and state-reduced model that maintains fidelity in observable outputs. The steady parameterized PDEs of interest are discretized and result in a system of algebraic equations we refer to as the *full* model. The full model depends on a high-dimensional parametric input; see section 3.1. In section 3.2 we propose a *reduced* model that takes parameter inputs in a low-dimensional subspace of the full parameter space and whose state must reside in a low-dimensional subspace of the full state space. While a state basis is necessary for projection in the traditional model reduction framework, here we require also a basis for the parameter. In section 3.3 we present an algorithm for the simultaneous construction of these bases.

3.1. Full model. Although we think of our forward model as a map from parameter space to output space, typical formulations yield models with state space \mathcal{U} . Thus, the forward model may be written more completely $\mathcal{M} : \mathcal{P} \rightarrow \mathcal{U} \rightarrow \mathcal{Y}$. We focus here on models that are linear in the state variables. The output space \mathcal{Y} can be any linear functional of the state; in some cases, we formulate the model such that the outputs are a subset of the states.

Our interest is in steady linear PDEs discretized in space, e.g., by finite elements, resulting in a system of algebraic equations of the form

$$(3.1) \quad \mathbf{A}(\mathbf{p})\mathbf{u} = \mathbf{f}, \quad \mathbf{y} = \mathbf{C}\mathbf{u},$$

where $\mathbf{A}(\mathbf{p}) \in \mathbb{R}^{N \times N}$ is the forward operator depending on the parameter $\mathbf{p} \in \mathbb{R}^{N_p}$, $\mathbf{u} \in \mathbb{R}^N$ is the state, $\mathbf{f} \in \mathbb{R}^N$ is the source, $\mathbf{C} \in \mathbb{R}^{N_o \times N}$ is the observation operator, and $\mathbf{y} \in \mathbb{R}^{N_o}$ is the vector of observable outputs.

In this case, the number of states scales with the number of grid points; therefore, three-dimensional problems typically have $N > 10^6$ discrete states. Furthermore, we have particular interest in distributed parameters, which also reside on the grid, i.e., $N_p > 10^6$ discrete parameters as well. Although the equations in (3.1) are linear in state, it should be noted that the map from parameter to state can be highly nonlinear, as the state depends on the parameter through the inverse of the forward operator. While the parameter and state are high-dimensional, the outputs are typically few in number, e.g., $N_o < 10^2$.

For models of this type, projection-based model reduction is well-established for the acceleration of forward model evaluations by reduction in state. Others have

applied model reduction to the posterior evaluation process in the Bayesian inference of a heat source in radiation [37], an application to real-time Bayesian parameter estimation [31], and in optical diffusion tomography [2]. In addition to reduction in state, reducing in the parameter space is essential for efficient exploration of the parameter space in many settings including design optimization, distributed control, and statistical inverse problems.

Next, we define the form of the reduced model and then present an algorithm for its construction.

3.2. Reduced model. Consider the full model (3.1). We propose the construction of a reduced model $\mathcal{M}_r : \mathcal{P}_r \rightarrow \mathcal{U}_r \rightarrow \mathcal{Y}$ whose outputs are accurate but parameter and state reside in low-dimensional subspaces $\mathcal{P}_r \subset \mathcal{P}$ and $\mathcal{U}_r \subset \mathcal{U}$, respectively. We assume that the parameter \mathbf{p} and state \mathbf{u} can be adequately approximated in the span of parameter and state bases $\mathbf{P} \in \mathbb{R}^{N_p \times n_p}$ and $\mathbf{V} \in \mathbb{R}^{N \times n}$, respectively. We obtain by Galerkin projection a reduced model of the form

$$(3.2) \quad \mathbf{A}_r(\mathbf{p}_r)\mathbf{u}_r = \mathbf{f}_r, \quad \mathbf{y}_r = \mathbf{C}_r\mathbf{u}_r,$$

where

$$\mathbf{A}_r(\mathbf{p}_r) = \mathbf{V}^T \mathbf{A}(\mathbf{P}\mathbf{p}_r)\mathbf{V}, \quad \mathbf{f}_r = \mathbf{V}^T \mathbf{f}, \quad \mathbf{C}_r = \mathbf{C}\mathbf{V},$$

where $\mathbf{A}_r \in \mathbb{R}^{n \times n}$ is the reduced forward operator depending on the reduced parameter $\mathbf{p}_r \in \mathbb{R}^{n_p}$, $\mathbf{u}_r \in \mathbb{R}^n$ is the reduced state, $\mathbf{f}_r \in \mathbb{R}^n$ is the projected source, $\mathbf{C}_r \in \mathbb{R}^{N_o \times n}$ is the reduced model observation operator, and $\mathbf{y}_r \in \mathbb{R}^{N_o}$ are the reduced model outputs. The reduction in parameter space is enforced directly by assuming $\mathbf{p} = \mathbf{P}\mathbf{p}_r$.

In traditional state reduction, a key challenge is identification of a low-dimensional subspace \mathcal{U}_r such that full and reduced model outputs are consistent. In a typical forward problem setting, reduced model accuracy may be desired for a finite set of parameters. In that case, one should sample those parameters, compute the corresponding states, and utilize the span of the resulting sets to form the basis. For inverse problems in particular, the parameters over which we desire reduced model accuracy are unknown—it is precisely these parameters that we wish to infer. In the absence of additional information, black box methods such as random sampling, Latin hypercube sampling, and centroidal Voronoi tessellations have been used to sample the parameter space and derive the state basis. If the parameter space has more than a handful of dimensions, however, [5] demonstrates greater efficiency over the black box samplers using a greedy approach [16, 36].

For the parameter- and state-reduced model (3.2), we must also build a basis for the parameter separately from the basis constructed for the state. We use the simplest approach, but perhaps also the most reasonable one; we derive the parameter basis from the set of parameters sampled to construct the state basis. Therefore, for the parameter basis vectors, we are guaranteed that the reduced model (3.2) will be as accurate as a reduced model without parameter reduction. This extension comes at a cost of only the orthogonalization process for the parameter basis. If the parameter vector is not associated with the discretization of a field quantity and an uncertainty estimate for a particular element is required, then the reduction may have to be orchestrated to maintain the structure of the problem. In some cases, reduction of that parameter may not be advisable. This treatment is problem dependent and is not addressed in the current paper. Here we are interested in global uncertainty estimates

of a scalar parameter field that has been represented by a vector of modal coefficients in a linear nodal basis.

We have described the form of the reduced model, but we have not yet discussed how to obtain the parameter samples which will define the reduced bases. In the next section, we describe a goal-oriented, model-constrained greedy approach to sampling the parameter space to build the reduced model.

3.3. Greedy sampling. A sequence of reduced models of increasing fidelity results from iteratively building up the parameter and state bases. At each step, we find the field in parameter space that maximizes the error between full and current reduced model outputs, subject to regularization by the prior. Although this approach is heuristic, it accounts for the underlying mathematical model and observable outputs. Furthermore, it is tractable even for models (3.1) with high-dimensional parameter and state spaces.

At each iteration of the greedy algorithm, we must evaluate the full and reduced models for members of the high-dimensional parameter space. Given that our parameter and state reduced model (3.2) accepts only reduced parameters as inputs, we need an additional map from high-fidelity parameters to their reduced counterparts. Let Ω be the computational domain. Since the parameter in this case is a distributed quantity, we choose a discretized $L_2(\Omega)$ projection such that

$$(3.3) \quad \mathbf{P}^T \mathbf{M} \mathbf{P} \mathbf{p}_r = \mathbf{P}^T \mathbf{M} \mathbf{p},$$

where \mathbf{M} is the mass matrix arising from the finite element discretization. With the addition of this constraint and the specification of a regularization parameter β , the k th greedy optimization problem

$$(3.4) \quad \mathbf{p}^k = \arg \max_{\mathbf{p} \in \mathcal{P}^k} \mathcal{J} = \frac{1}{2} \|\mathbf{y}(\mathbf{p}) - \mathbf{y}_r(\mathbf{p}_r)\|_2^2 - \frac{1}{2} \beta \|\mathbf{p}\|_{\mathbf{S}^{-1}}^2$$

subject to (3.1), (3.2), and (3.3) is completely defined. The objective function $\mathcal{J} : \mathcal{P}^k \rightarrow \mathbb{R}$ is a weighted sum of two terms. The first term measures the disparity between observable outputs of the full and current reduced model. The second term penalizes parameters of low probability in a manner consistent with the prior belief in the statistical inverse problem.

We search for the parameter field in a restricted set $\mathcal{P}^k \subset \mathcal{P}$ which may change from greedy cycle to greedy cycle. This restriction allows enforcement of additional constraints on the parameter. For example, take $\mathcal{P}^k = \mathbf{P}^\perp$ to be the orthogonal complement of the current parameter basis. Then, in the limit $\mathbf{y}_r(\mathbf{p}_r) \rightarrow \mathbf{y}(\mathbf{p})$, the misfit term goes to zero, and we sample the eigenvectors of the prior covariance in \mathbf{P}^\perp . Those eigenvectors are precisely the Karhunen–Loève modes typically used in practice when parameter reduction is performed in the statistical sampling setting. Please refer to section 4.2 for further discussion.

Optimization problem (3.4) is a nonconvex PDE-constrained nonlinear optimization problem over a high-dimensional parameter space. Since the constraints are linear, however, we may rewrite (3.4) as an unconstrained problem where we find the \mathbf{p}^k that maximizes

$$(3.5) \quad \mathcal{J} = \frac{1}{2} \|\mathbf{C} \mathbf{A}^{-1}(\mathbf{p}) \mathbf{f} - \mathbf{C} \mathbf{V} (\mathbf{V}^T \mathbf{A} (\mathbf{P} (\mathbf{P}^T \mathbf{M} \mathbf{P})^{-1} \mathbf{P}^T \mathbf{M} \mathbf{p}) \mathbf{V})^{-1} \mathbf{V}^T \mathbf{f}\|_2^2 - \frac{1}{2} \beta \|\mathbf{p}\|_{\mathbf{S}^{-1}}^2.$$

We solve (3.5) with a trust-region Newton method where we provide analytical gradients and a subroutine for the Hessian-vector product. At each outer loop iteration,

the Newton direction is computed using conjugate gradients (CG) [20, 23]. The objective function is nonconvex, and we are not guaranteed to find the global optimum at each iteration. Grid continuation methods are the usual combatant for this issue in high-dimensional PDE-constrained optimization problems [3]. By solving the optimization problem on a sequence of refined meshes, it is often the case that we gradually approach the basin of attraction of the global optimum. The solution of (3.4) is a significant challenge; however, the similarity between (3.4) and a deterministic inverse problem formulation can be exploited. We have a plethora of knowledge and methodology from optimization problems of similar form; see, e.g., [1, 20, 23] and the references therein. Further, as shown in [5], the computational cost of solving (3.4) via these methods has an attractive scalability with the dimension of the parameter space.

To summarize, we present the model reduction procedure in Algorithm 3.1. We typically terminate the greedy sampling process once a reduction of several orders of magnitude is achieved in the objective function.

ALGORITHM 3.1.

Greedy parameter and state model reduction.

1. Initialize parameter basis to a single constant field $\mathbf{P} = \mathbf{p}_c$, solve (3.1), and initialize state basis to $\mathbf{V} = \mathbf{u}(\mathbf{p}_c)$; set $k = 2$.
2. Solve the greedy optimization problem (3.4) subject to (3.1), (3.2), and (3.3) for an appropriate regularization parameter β to find \mathbf{p}^k and compute the corresponding forward solution $\mathbf{u}(\mathbf{p}^k)$ using (3.1).
3. Update the reduced model. Incorporate \mathbf{p}^k and $\mathbf{u}(\mathbf{p}^k)$ into the parameter and state bases, respectively, by Gram–Schmidt orthogonalization.
4. If converged, stop. Otherwise, increment k and loop to 2.

In the next section, we describe how a reduced model derived using Algorithm 3.1 can be exploited in a reduced MCMC approach to the statistical inverse problem.

4. Reduced MCMC. Motivated by the need for uncertainty quantification in inverse problem solutions, the difficulty of sampling in high-dimensional parameter spaces, and the excessive computational cost of forward model solutions, we propose a reduced MCMC; see Algorithm 4.1. Sampling takes place in the reduced parameter space, and posterior evaluations are performed by the reduced model. The forthcoming analysis assumes that the forward operator is linear in the parameter, a property of our target groundwater problem. In that case, online reduced MCMC computations scale with the reduced dimensions n and n_p instead of N and N_p . The model underlying the reduced MCMC is derived using Algorithm 3.1.

4.1. Algorithm. Once the reduced model is constructed in the offline phase, it can be employed at little cost in the online phase, i.e., when we use MCMC to generate samples from the posterior. Reduced MCMC yields samples in the reduced parameter space and utilizes a random walk based on the Metropolis–Hastings sampler. Each posterior evaluation required to compute the Hastings ratio is computed using the reduced model exclusively. The cost of an MCMC sample scales with the reduced dimensions n and n_p , as opposed to the high-fidelity dimensions N and N_p . The cost is independent of N because the Galerkin matrix has dimension n -by- n . The cost is independent of N_p because the forward operator’s dependence on the parameter is linear; therefore, we can precompute the dependence on each parameter basis vector and sum the contributions for a given reduced parameter \mathbf{p}_r . This is a particular example of the more general offline/online decomposition procedure of reduced basis methods [36].

We summarize the reduced MCMC algorithm for a desired number of samples N_s . Consider a point in the chain \mathbf{p}_r . The proposal distribution $\xi(\mathbf{q}_r|\mathbf{p}_r)$, which may be a multivariate normal with mean \mathbf{p}_r and projected prior covariance $\mathbf{S}_r = \mathbf{P}^T\mathbf{S}\mathbf{P}$, is sampled to generate a candidate \mathbf{q}_r . In the low-dimensional parameter space, it is not difficult to find a proposal distribution that yields an optimal acceptance rate. We suspect that a multivariate normal proposal distribution with independent components can also be utilized successfully in the reduced MCMC.

ALGORITHM 4.1.

Reduced MCMC.

1. Given a full model (3.1), construct a reduced model (3.2) using Algorithm 3.1.
2. Initialize the Markov chain at \mathbf{p}_r^0 ; set $i = 1$.
3. Generate a candidate from the proposal distribution $\xi(\mathbf{p}_r|\mathbf{p}_r^{i-1})$. We recommend $\xi \sim \mathcal{N}(\mathbf{p}_r^{i-1}, \mathbf{S}_r)$, where $\mathbf{S}_r = \mathbf{P}^T\mathbf{S}\mathbf{P}$ is the projection of the prior covariance onto the reduced parameter space.
4. Compute the acceptance ratio

$$(4.1) \quad \alpha = \min \left[1, \frac{\pi(\mathbf{p}_r|\mathbf{y}_d)\xi(\mathbf{p}_r^{i-1}|\mathbf{p}_r)}{\pi(\mathbf{p}_r^{i-1}|\mathbf{y}_d)\xi(\mathbf{p}_r|\mathbf{p}_r^{i-1})} \right]$$

by evaluating $\pi(\mathbf{p}_r|\mathbf{y}_d) = \exp\{-\frac{1}{2\sigma^2}\|\mathbf{y}_d - \mathbf{y}_r(\mathbf{p}_r)\|_2^2 - \frac{1}{2}\|\mathbf{P}\mathbf{p}_r\|_{\mathbf{S}^{-1}}^2\}$ using the reduced model (3.2).

5. With probability α , accept the new candidate, i.e., $\mathbf{p}_r^i = \mathbf{p}_r$; otherwise, reject the candidate and take $\mathbf{p}_r^i = \mathbf{p}_r^{i-1}$.
6. If $i < N_s$, loop to 3; otherwise, stop.

In the next section, we analyze the effect of utilizing a parameter- and state-reduced model in MCMC.

4.2. Analysis. Although rigorous error bounds exist for projection-based reduced models in very specific settings (see, e.g. [16, 21, 31, 36]), such results do not exist in general. In light of this, error analysis and convergence results are not available in the statistical setting. On the other hand, we are in a position to provide a detailed complexity analysis.

Our analysis breaks down into offline and online components. In the offline stage, we build the reduced model by Algorithm 3.1. To obtain an additional pair of parameter and state vectors, we must solve the greedy optimization problem (3.4). We utilize a reduced-space matrix-free inexact Newton-CG algorithm. The gradient \mathbf{g} is computed via adjoint computations which are linear in the number of parameters N_p . Never is the Hessian \mathbf{H} constructed and stored; instead, we require only the action of the Hessian on a vector to compute the Newton direction $\hat{\mathbf{p}}$ as the solution to $\mathbf{H}\hat{\mathbf{p}} = -\mathbf{g}$. This matrix-vector product requires a sequence of forward and adjoint solves. We assume that both the number of linear and number of nonlinear iterations are independent of the problem size due to the special structure in the problem [1, 20, 23]. The full-order MCMC does not have an offline stage.

In the online stage of the full-order implementation, we cannot use traditional Metropolis–Hastings samplers. Instead, we use the delayed rejection adaptive Metropolis (DRAM) algorithm, which utilizes a limited-history sample covariance in the proposal distribution [17]. Since the sample covariance is dense, the dominant cost is shared by updating the factorization and evaluating the proposal probability, both of which cost $\mathcal{O}(N_p^2)$. In the online stage of the reduced MCMC, we are able to utilize the traditional Metropolis–Hastings sampler whose dominant cost $\mathcal{O}(n_p n^2)$ is given by the posterior evaluation when we must solve for the outputs of the reduced model.

TABLE 4.1

Complexity analysis for solution to the statistical inverse problem by MCMC sampling at full and with reduced MCMC. Here, N_p and N are the dimensions of the full parameter and state, respectively; n_p and n are the parameter and state dimensions for the corresponding reduced model. We assume that no greedy samples are redundant such that we require n greedy cycles to build a reduced model of size n . On average, we take m nonlinear iterations (iters.) to converge the greedy optimization problem during each greedy cycle. All linear systems are solved by iterative Krylov methods and take advantage of structure in the spectrum of the matrix operator through appropriate preconditioning such that the number of iterations is independent of the dimension of the system.

	Operation	Full	Reduced
Offline	greedy cycles		n
	nonlinear iters.		m
	gradient		$\mathcal{O}(N_p n_p n^2)$
	linear iters. forward solve		$\mathcal{O}(N_p n_p + Nn)$ $\mathcal{O}(N_p + N)$
Subtotal			$\mathcal{O}(mN_p n_p n^3)$
Online	MCMC samples	$\mathcal{O}(N_p^2)$	$\mathcal{O}(n_p^2)$
	proposal	$\mathcal{O}(N_p^2)$	$\mathcal{O}(n_p^2)$
	proposal evaluation	$\mathcal{O}(N_p^2)$	$\mathcal{O}(n_p^2)$
	posterior evaluation	$\mathcal{O}(N_p + N)$	$\mathcal{O}(n_p n^2)$
Subtotal		$\mathcal{O}(N_p^4)$	$\mathcal{O}(n_p^3 n^2)$
Total		$\mathcal{O}(N_p^4)$	$\mathcal{O}(mN_p n_p n^3 + n_p^3 n^2)$

In both the full-order and reduced implementations we assume that the number of samples required to achieve the desired level of accuracy scales with the square of the number of parameters [33].

A complexity analysis comparing full and reduced MCMC instantiations is provided in Table 4.1. Asymptotic cost is presented for the offline and online portions. It is our assumption that we may construct a reduced model with $n_p \ll N_p$ and $n \ll N$ that maintains the integrity of the parameter-output map of the original model. Therefore, the dominant costs are given by N_p and N for each approach. While the cost of the full-order implementation scales like N_p^4 , the reduced MCMC scales only linearly with N_p . Furthermore, the online portion of the reduced MCMC is independent of the original parameter and state space dimensions. If one has run reduced MCMC for N_s samples and then decides to collect more samples, the cost of obtaining higher accuracy in the statistical results will not depend on the full dimensionality of the problem. Indeed, this is the advantage of the offline/online decomposition.

The reduced MCMC sacrifices accuracy by requiring samples to reside within a low-dimensional subspace and approximating the output by reduction in state; the full case, on the other hand, is completely intractable for large N_p and N . We now turn to some discussion regarding the treatment of uncertainty by the reduced MCMC approach.

Our formulation of the greedy sampling problem represents an attempt to reduce the maximum error in the outputs between full and reduced models while avoiding unnecessary sampling of parameters excluded by the prior information. This goal-oriented approach produces a reduced model that accurately matches the outputs of the full model over a subset of the parameters with appreciable prior probability. The greedy objective function (3.4) is knowledgeable about the prior of the statistical inverse problem in the following manner. If there is a set of parameters that produce the same output error, the one that will be sampled will be the one with largest prior probability.

In section 3.3 we discussed how the regularized greedy formulation (3.4) can be chosen so that in the limit $\mathbf{y}_r(\mathbf{p}_r) \rightarrow \mathbf{y}(\mathbf{p})$, the optimization results in sampling the Karhunen–Loève modes. However, we comment here that sampling past this limit may be inefficient. Samples not driven by the misfit in our representation of the physics amount to parameters for which the data is uninformative in the statistical inverse setting. The additional directions (i.e., those arising as eigenvectors of the prior covariance) can be sampled more efficiently using a method such as Rao–Blackwellization [14, 27].

Although it is not clear how the reduced model treats uncertainty in the nonlinear setting, in the linear case, a reduced model constructed by Algorithm 3.1 will be a basis for the parameters about which we are most certain. In this way, the reduced model is foremost established to approximate the full model in a parameter–output map. As the dimension of the reduced model grows, its approximation of the likelihood becomes more accurate, and the reduced-fidelity posterior approaches the full posterior. It is a significant challenge to obtain a reduced model that both matches full model output while also spanning the parameters that contribute most significantly to the uncertainty in the posterior. This topic is the subject of ongoing research.

In the next section, we present numerical results from inverse problems in one and two spatial dimensions.

5. Results. The governing equations are those of steady flow in porous media. We specify the problem description in section 5.1. In section 5.2, a parameter- and state-reduced model is constructed by means of the greedy sampling procedure outlined in section 3.3. The reduced MCMC results are compared with inversions based on the full model in section 5.3. The largest mesh contains 501 degrees of freedom. We select a modest size in order to make the full inversion tractable, so that we may compare results with the reduced MCMC algorithm.

It should be noted immediately that the full MCMC results presented in this section represent the first author’s greatest effort to develop an efficient sampler in the high-dimensional parameter space with the use of DRAM. In order to obtain a nonzero acceptance rate, the MCMC chain has to be started near the actual solution, and the candidates must remain nearby. This process may produce results which are misleading in accuracy: The full posteriors are not well-explored; in fact, the samples in the chain remain very close to the maximum a posteriori estimate. The result is an underprediction of the true variance. The algorithm DRAM should not be blamed for this inadequacy; sampling in such a high-dimensional parameter space is an open problem in this area—precisely the motivation for our reduced MCMC methodology.

5.1. Problem description. The governing equations are those of steady flow in porous media. Let $u(\vec{x})$ be the pressure head, $K(\vec{x})$ be the hydraulic conductivity, and $f(\vec{x})$ a recharge term. Given a field K , the pressure head u is given by

$$\begin{aligned} -\vec{\nabla} \cdot (K \vec{\nabla} u) &= f \quad \text{in } \Omega, \\ K \vec{\nabla} u \cdot \vec{n} &= 0 \quad \text{on } \Gamma_{\text{N}}, \\ u &= 0 \quad \text{on } \Gamma_{\text{D}}, \end{aligned}$$

where Ω is the computational domain with boundary $\partial\Omega = \Gamma_{\text{D}} \cup \Gamma_{\text{N}}$ and $\Gamma_{\text{D}} \cap \Gamma_{\text{N}} = \emptyset$, Γ_{D} and Γ_{N} are Dirichlet and Neumann boundaries, respectively, and \vec{n} is the outward-pointing unit normal. For well-posedness of the forward problem, we require $K > 0$ in Ω ; thus it is convenient to work with $\log K$ as the parameter. Our forward model also includes a set of outputs $y_i = u(\vec{x}_i)$, $i = 1, 2, \dots, N_o$, corresponding to the pressure

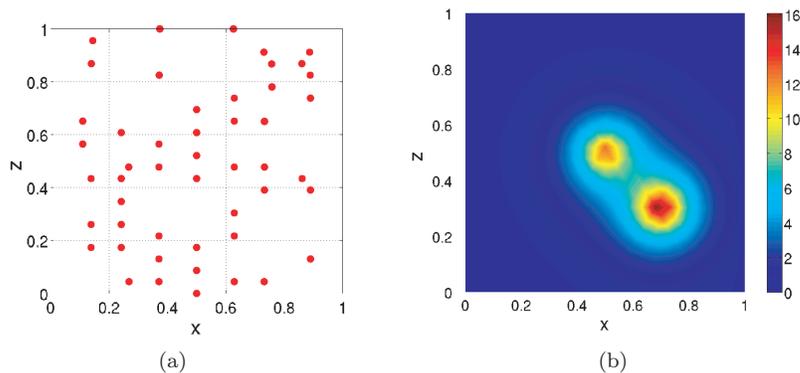


FIG. 5.1. For the $N = 494$ test case in 2-D, the (a) sensor array and (b) recharge term f .

head at a set of sparsely distributed sensors. When discretized, e.g., by finite elements, the forward model can be expressed in the form (3.1) where \mathbf{u} and \mathbf{p} are the discretized forms of u and $\log K$, respectively.

We present results for test problems in one and two spatial dimensions. In one dimension, $\Omega = (0, 1]$, $\Gamma_D = \{x|x = 0\}$, and $\Gamma_N = \{x|x = 1\}$. The sensors are distributed evenly throughout the domain. The source term is a superposition of three exponentials

$$f(x) = \sum_{i=1}^3 \alpha_i \exp \left\{ -\frac{(x - \mu_i)^2}{\beta_i^2} \right\},$$

where $\alpha = (1900, 5100, 2800)$, $\mu = (0.3, 0.6, 0.9)$, and $\beta = (0.01, 0.05, 0.02)$. In two dimensions, we have $\Omega = [0, 1]^2$, $\Gamma_D = \{(x, z)|x \in \{0, 1\}\}$, and $\Gamma_N = \{(x, z)|z \in \{0, 1\}\}$. For each of the test problems, we assume that the sensor array and recharge term f are fixed. In Figure 5.1 we show the sensor array in the domain along with the recharge term

$$f(x, z) = 15 \exp \left\{ -\frac{\|(x, z) - (0.5, 0.5)\|_2}{0.3^2} \right\} + 19 \exp \left\{ -\frac{\|(x, z) - (0.7, 0.3)\|_2}{0.3^2} \right\}$$

for the $N = 494$ 2-D test case. In each case, we specify less than 10% of the nodes as sensors, and in the 2-D problems, they are chosen to be roughly aligned vertically to simulate a set of boreholes in the geophysical setting.

5.2. Reduced model performance. We now demonstrate the construction of the reduced model by the greedy sampling procedure described above. The greedy sampling problem (3.4) is solved at every iteration of the procedure using a trust-region, interior-reflective, inexact Newton method [8, 9]. We provide the gradient and the Hessian-vector product as required by the optimizer with a series of forward and adjoint solves. The reader is referred to [26] for details. In this case, the Gaussian process prior adds sufficient regularity to the optimization problem; however, for a different choice of the prior, one may need to employ grid continuation [3].

For the sake of brevity, we present results from the greedy sampling procedure only for the 2-D case with $N = 494$. In Figure 5.2, we plot the parameter and state basis vectors obtained by Algorithm 3.1. We initialize the procedure by first sampling the

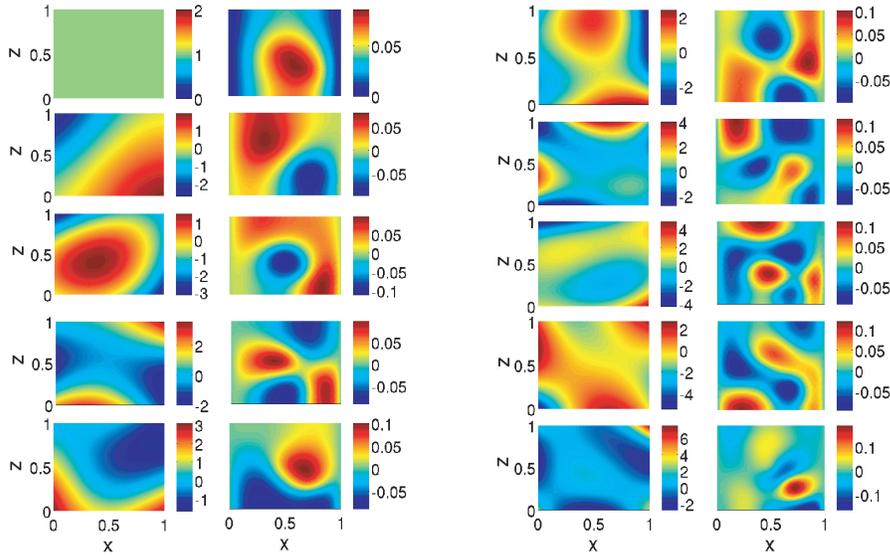


FIG. 5.2. Ten orthogonalized parameter and ten orthogonalized state basis vectors derived using the greedy sampling algorithm (3.4). In the first two columns, the first five pairs; in the last two columns, the last five pairs. In each pair, the parameter is shown on the left, and the state on the right.

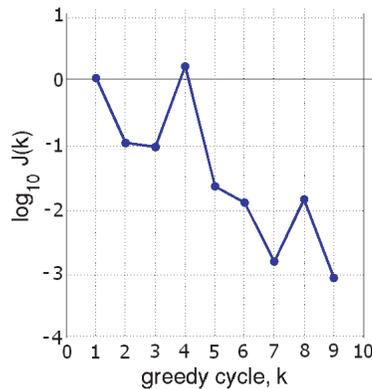


FIG. 5.3. The greedy objective function value (3.4) versus the greedy cycle index for the 2-D problem with $N = 494$ variables. We observe a decreasing trend. The procedure is stopped when the objective function has decreased by three orders of magnitude; see Algorithm 3.1.

constant parameter $\mathbf{p} = \mathbf{1}$. When we do so, we obtain the first parameter basis vector. Then we solve the full model to obtain $\mathbf{u}(\mathbf{p})$, which is the first state basis vector. On each subsequent iteration we follow the same process with \mathbf{p} determined by solving the greedy sampling problem (3.4). The new basis vectors are incorporated into the basis via Gram–Schmidt orthogonalization. As the iteration proceeds, we observe a decreasing trend in the objective function; see Figure 5.3. When the objective function has decreased by three orders of magnitude, we conclude the process.

Although we have included the prior information from the statistical inverse problem as a regularization penalty, the greedy process is deterministic. We now consider a statistical measurement of the accuracy of the reduced model. For the 2-D test case with $N = N_p = 494$, we select 1,000 conductivity fields at random from our

TABLE 5.1

Performance statistics for the full model and reduced models of dimension five and dimension ten for the $N = N_p = 494$, 2-D test case. The output vectors \mathbf{y} , \mathbf{y}_{r_5} , and $\mathbf{y}_{r_{10}}$ correspond to the full model, the reduced model of dimension five, and the reduced model of dimension ten. For 1,000 random samples from the prior, we present the sample mean and sample variance for the ℓ_2 -norm of the outputs and the output errors. As expected, the reduced model with more basis vectors more accurately replicates the statistics of the full model predicted outputs.

	$\mathbb{E}(\cdot)$	$\text{var}(\cdot)$
$\ \mathbf{y}\ _2$	1.5754	0.1906
$\ \mathbf{y}_{r_5}\ _2$	1.5756	0.1943
$\ \mathbf{y}_{r_{10}}\ _2$	1.5754	0.1906
$\ \mathbf{y} - \mathbf{y}_{r_5}\ _2$	2.3986×10^{-2}	2.2284×10^{-4}
$\ \mathbf{y} - \mathbf{y}_{r_{10}}\ _2$	3.2796×10^{-3}	3.9419×10^{-6}

Gaussian process prior and compute the predicted outputs using the full model and the reduced models of dimension five and dimension ten. Each model estimates the pressure head at the 49 sensors in the domain; see Figure 5.1(a). Let \mathbf{y} , \mathbf{y}_{r_5} , and $\mathbf{y}_{r_{10}}$ be the vectors of outputs for the full model, the reduced model of dimension five, and the reduced model of dimension ten, respectively. In Table 5.1, we present the sample mean and sample variance for $\|\mathbf{y}\|_2$, $\|\mathbf{y}_{r_5}\|_2$, $\|\mathbf{y}_{r_{10}}\|_2$, and the error in outputs $\|\mathbf{y} - \mathbf{y}_{r_5}\|_2$ and $\|\mathbf{y} - \mathbf{y}_{r_{10}}\|_2$. In this case, the statistics of the outputs of the reduced model adequately match those of the full model; however, it is important to note that there may be parameter values (e.g., values in the tails of the prior distribution) for which the reduced model is inaccurate. As expected, the larger reduced model is more accurate—as the number of parameter and state basis vectors increases, the reduced model tends toward the full model.

5.3. Inverse problem solution. The following is the experimental procedure. Given a certain discretization of the domain, we construct the prior probability density $\gamma_p(p) \sim \mathcal{N}(0, S)$ using the Gaussian kernel (2.2) for the choices $a = 0.1$, $b = 0.8$, and $c = 10^{-8}$. We employ greedy parameter and state model reduction as described above to determine a reduced model. Then we generate synthetic data.

First, we select a hydraulic conductivity field from the prior at random. We solve the forward model on the given discretization using Galerkin finite elements with piecewise linear polynomial interpolation. The output data are generated by selecting the values of the states at a small subset of the mesh nodes and then corrupting the data with noise drawn from a multivariate normal distribution $\mathcal{N}(0, \sigma^2 I)$. In these cases, we choose $\sigma = 0.01$.

Once the data are generated, we solve the statistical inverse problem in two ways. Samples are generated from the posterior $\pi_{\mathbf{p}}(\mathbf{p}|\mathbf{y}_d)$ using the full model and adaptive sampling. The starting location for the chain is given by a weighted combination of the true hydraulic conductivity and another sample drawn from the prior. Although this is unfairly biased in the direction of the performance of the full instantiation, it is necessary for a successful sampling process—without this, sampling with MCMC at full order is nearly impossible even for modest problem sizes. Our interest is primarily in evaluating the accuracy of the reduced MCMC results; selecting an appropriate initial sample is necessary to complete that task. Note that we are able to solve the full statistical inverse problem in this case due to our choice of problem size. We obtain a benchmark against which we may test the performance of the reduced MCMC approach.

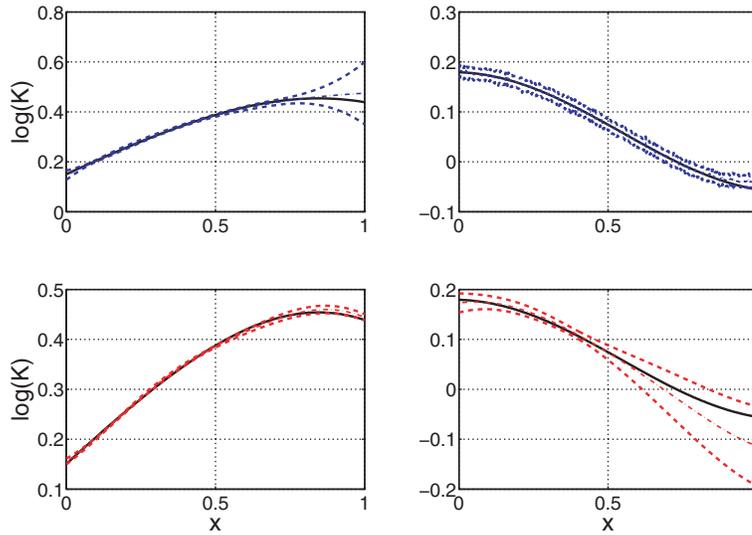


FIG. 5.4. Statistical inversions for the 1-D model problem for two mesh discretizations, (left) $N = 51$ and (right) $N = 501$. On the top we present results from the full inversion; the reduced MCMC results are on the bottom. The solid line is the parameter used to generate the data, the dash-dot line is the sample mean, and the dashed lines denote the sample mean plus and minus two pointwise standard deviations.

With reduced MCMC, we do not require that the seed of the chain be near the true parameter because the burn-in time is minimal in the reduced parameter space for a well-tuned sampler. We initialize by drawing at random from the projected prior distribution. The posterior evaluations are given by solutions to the reduced model (3.2). The proposal distribution is defined on the reduced parameter space, and therefore all samples in the chain are of reduced dimension. For the test cases presented herein, we use a multivariate normal distribution whose mean is the previous sample and whose covariance is the product of a scaling factor and the projected prior covariance. We tune the scaling factor, with negligible effort, to provide an acceptance rate between 20% and 30% [6, 33].

We utilize the sample mean and pointwise variance (diagonal of the sample covariance) as the metric to assess our inversion. In the following figures, we plot the log hydraulic conductivity utilized to generate the data, the sample mean, and the sample mean plus or minus two pointwise standard deviations. In Figure 5.4, we present results from the 1-D model problem for two mesh discretizations, $N = N_p = 51$ and $N = N_p = 501$.

In each case, the reduced MCMC credible interval envelopes the true parameter. For the 1-D $N = N_p = 51$ test case, reduced MCMC underestimates, with respect to the full inversion, the uncertainty in the parameter near the right boundary. In the 1-D $N = N_p = 501$ case, the reverse is true—reduced MCMC appears to overestimate the uncertainty. However, it must be noted that the full model solution appears to predict an unrealistically small credible interval. This may be a result of a combination of two factors: (1) the starting location's proximity to the true parameter, and (2) the small length-scale required to achieve acceptances in MCMC for high-dimensional cases. Together, these factors may have resulted in a short-sighted sampler; that is, one that does not sufficiently explore the posterior. The increase in uncertainty near the right boundary, where we apply Neumann conditions, is consistent with results

TABLE 5.2

Number of full model degrees of freedom, number of outputs, reduced model degrees of freedom (– indicates the absence of a reduced model) and offline, online, and total time required to generate the results for the 1-D test cases on a DELL Latitude D530 with Intel Core 2Duo at 2 GHz. The offline time denotes the CPU time required to solve five iterations of the greedy sampling problem with MATLAB's `fmincon`. In the MCMC simulation, the first 10% of the samples were discarded as burn-in.

$N = N_p$	N_o	$n = n_p$	Offline time (sections)	Samples	Online time (seconds)	Total time (seconds)
51	5	–	–	500,000	1.05×10^3	1.05×10^3
51	5	5	3.98×10^2	150,000	1.60×10^2	5.58×10^2
501	25	–	–	500,000	2.43×10^4	2.43×10^4
501	25	5	7.24×10^3	150,000	2.53×10^2	7.50×10^3

TABLE 5.3

Number of full model degrees of freedom, number of outputs, reduced model degrees of freedom (– indicates the absence of a reduced model) and offline, online, and total time required to generate the results for the 2-D model problems on a DELL Latitude D530 with Intel Core 2Duo at 2 GHz. The offline time denotes the CPU time required to solve $n - 1$ iterations of the greedy sampling problem with MATLAB's `fmincon`. In the MCMC simulation, the first 10% of the samples were discarded as burn-in.

$N = N_p$	N_o	$n = n_p$	Offline time (seconds)	Samples	Online time (seconds)	Total time (seconds)
59	5	–	–	500,000	1.10×10^3	1.10×10^3
59	5	5	2.04×10^1	150,000	1.57×10^2	1.77×10^2
494	49	–	–	500,000	1.67×10^4	1.67×10^4
494	49	5	4.69×10^3	200,000	3.32×10^2	5.02×10^3
494	49	10	1.04×10^4	200,000	4.92×10^2	1.09×10^4

in [29]. There, Neumann conditions were applied on either end of a 1-D interval and increases in the standard deviation were observed at both ends.

Table 5.2 shows the computing time for each case. We record the offline, online, and total time for each run. The offline time corresponds to the computational effort required for the greedy parameter and state model reduction. The performance benefit we obtain is due to a dramatic decrease in the online time, which corresponds to the MCMC sampling procedure. In the full MCMC, we utilize the DRAM algorithm to achieve a reasonable acceptance rate. For reduced MCMC, we can achieve optimal acceptance rates [33] using a Metropolis–Hastings sampler. We achieve a factor of two speedup in the $N = 51$ test problem and a factor of three in the $N = 501$ test problem. For these 1-D results and the 2-D results presented in Table 5.3, our observed computing times are not consistent with the asymptotic analysis in Table 4.1 for two reasons. First, to make feasible comparisons with the full implementation, the problem dimensions are moderate, and therefore are not in the asymptotic regime. Secondly, the overhead offline cost of optimization for the greedy sampling problem in MATLAB does not scale efficiently with problem dimension, as it would, e.g., for a C implementation. In that case, we expect that the savings will increase dramatically with the dimension of the problem.

We present similar results for the 2-D model problem. We present timings for two cases, $N = N_p = 59$ and $N = N_p = 494$, but we show results for the $N = N_p = 494$ case only. The true hydraulic conductivity field is plotted in Figure 5.5. In Figure 5.6, we plot the lower bound of the credible interval, the mean, and the upper bound of the credible interval from left to right.

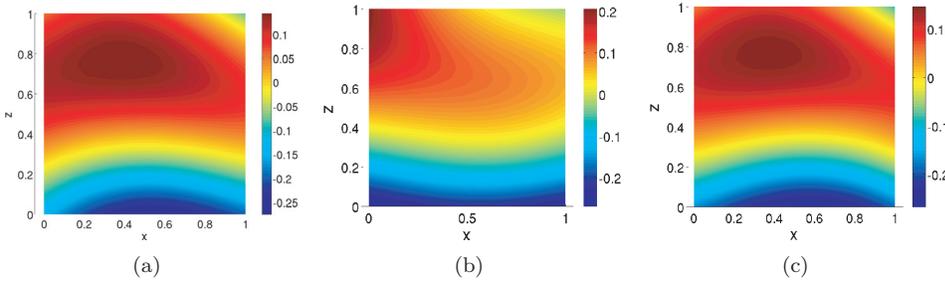


FIG. 5.5. For the 2-D test case $N = N_p = 494$, (a) the parameter field used to generate the data and its projection onto the reduced parameter space of dimension, (b) $n_p = 5$, and (c) $n_p = 10$.

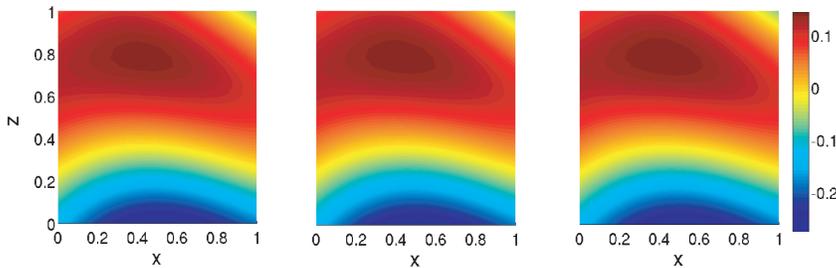


FIG. 5.6. The results of the MCMC solution to the 2-D $N = N_p = 494$ test case. On the left and right, respectively, we plot the lower and upper bounds of the $\pm 2\sigma$ credible interval. The mean is shown in the center. The actual parameter field is shown in Figure 5.5.

Consider reduced MCMC results for reduced models with parameter and state dimensions of $n = n_p = 5$ and $n = n_p = 10$. We plot the $L_2(\Omega)$ projections of the true parameter to these reduced parameter spaces in Figures 5.5(b) and 5.5(c). It is clear that the first five basis vectors are insufficient to properly capture the true parameter, but with ten basis vectors, we match the true parameter almost exactly. The corresponding results from reduced MCMC are shown in Figures 5.7 and 5.8, respectively.

For each set of results, we calculate the relative $L_2(\Omega)$ error in each of the three statistics of interest: the mean, and the upper and lower bounds of the credibility interval. The error is calculated with respect to the full MCMC solution. We find that our reduced model of size $n = n_p = 5$ produces relative errors in the range 35%–37%, whereas the higher fidelity reduced model $n = n_p = 10$ has relative errors in the range 12%–17%. We expect that these errors will continue to decrease as the dimension of the reduced model increases.

From these results, we have demonstrated that as the basis becomes richer, we are better able to characterize moments of the posterior with reduced MCMC. In this case, we have reduced the problem from $N = 494$ to $n_p = 10$ parameters and $n = 10$ states. We expect the number of parameter and state dimensions required to achieve an adequate solution to scale with the complexity of the physics, but not with the dimension of the underlying discretization.

In Table 5.3 we present the CPU timings for the runs in 2-D. We achieve about one order of magnitude speedup in the $N = 59$ case in total time. In the $N = 494$ case, the greedy sampling problem becomes significantly more challenging. A speedup of

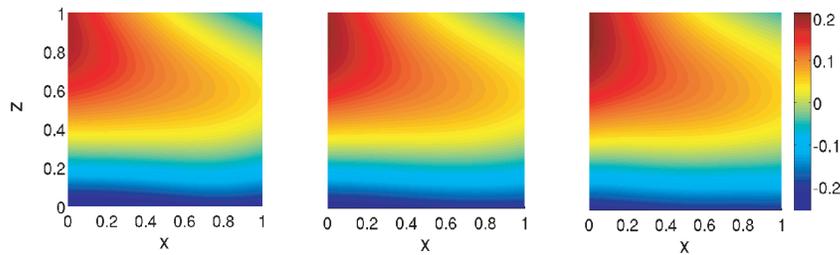


FIG. 5.7. The results of the MCMC solution to the 2-D $N = N_p = 494$ test case with reduction to $n = n_p = 5$. On the left and right, respectively, we plot the lower and upper bounds of the $\pm 2\sigma$ credible interval. The mean is shown in the center. The actual parameter field is shown in Figure 5.5(a). For reference, we show the projected actual parameter field below in Figure 5.5(b). In comparison to the full MCMC solution, these results have a relative $L_2(\Omega)$ error of 35.4%, 35.7%, and 36.9%, from left to right.

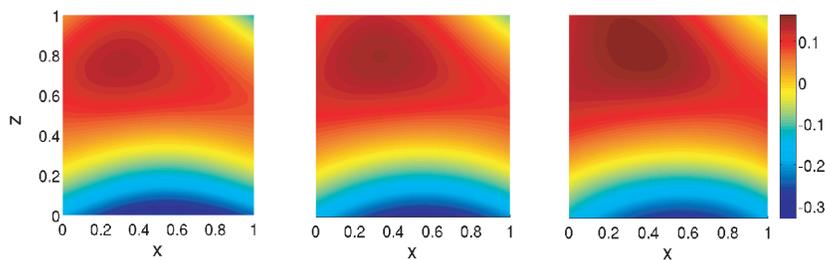


FIG. 5.8. The results of the MCMC solution to the 2-D $N = N_p = 494$ test case with reduction to $n = n_p = 10$. On the left and right, respectively, we plot the lower and upper bounds of the $\pm 2\sigma$ credible interval. The mean is shown in the center. The actual parameter field is shown in Figure 5.5(a). For reference, we show the projected actual parameter field below in Figure 5.5(c). In comparison to the full MCMC solution, these results have a relative $L_2(\Omega)$ error of 16.6%, 12.2%, and 16.3%, from left to right.

almost two orders of magnitude is observed in online time, but only a 35% reduction in total time.

It is important to note that our method targets moments of the posterior, i.e., integrals that may be estimated by their discrete analogues. It is possible to obtain uncertainty estimates for individual parameters by projecting reduced MCMC samples back up to the full-dimensional parameter space by premultiplication of the parameter basis \mathbf{P} . However, the estimate of such localized uncertainty may not be predicted well by our approach, which is designed for inferring global quantities. If such a localized quantity is required, the reduced model construction should be modified (see section 3.2 for related discussion).

6. Conclusion. Bayesian statistical inverse problems are an outstanding challenge for forward models consisting of PDEs with distributed parameters. While the complexity of a posterior evaluation can be reduced by a traditional state-reduced model, efficient sampling in the high-dimensional parameter space remains an issue. To address these issues, we propose an extension to a parameter- and state-reduced model which maintains the accuracy of output predictions. In the reduced parameter space, traditional, nonadaptive Metropolis–Hastings samplers can be utilized successfully.

This reduced MCMC approach provides a systematic method for solving statistical inverse problems involving PDEs with high-dimensional parametric input spaces—a class of problems for which the full statistical inverse problem is beyond our current means. In theory, the method scales independently of the fineness of the PDE discretization, but instead depends only on the complexity of the physics. Our method can be applied to problems with many more degrees of freedom, but we cannot compare the results to full calculations because MCMC becomes prohibitively expensive in that setting.

We have presented promising results for prediction of the mean and credibility interval for some simple test problems. An important direction for future research is the establishment of error bounds on the entire posterior probability density function, which may be required in the decision-making, engineering design, or optimal control settings.

Acknowledgment. The authors would like to thank Youssef Marzouk for helpful discussions regarding MCMC.

REFERENCES

- [1] V. AKCELIK, G. BIROS, O. GHATTAS, J. HILL, D. KEYES, AND B. VAN BLOEMAN WAANDERS, *Parallel PDE constrained optimization*, in *Parallel Processing for Scientific Computing*, M. Heroux, P. Raghaven, and H. Simon, eds., SIAM, Philadelphia, 2006.
- [2] S. ARRIDGE, J. KAIPIO, V. KOLEHMAINEN, M. SCHWEIGER, E. SOMERSALO, T. TARVAINEN, AND M. VAUHKONEN, *Approximation errors and model reduction with an application in optical diffusion tomography*, *Inverse Problems*, 22 (2006), pp. 175–195.
- [3] U. ASCHER AND E. HABER, *Grid refinement and scaling for distributed parameter estimation problems*, *Inverse Problems*, 17 (2001), pp. 571–590.
- [4] I. BABUSKA, F. NOBILE, AND R. TEMPONE, *A stochastic collocation method for elliptic partial differential equations with random input data*, *SIAM J. Numer. Anal.*, 45 (2007), pp. 1005–1034.
- [5] T. BUI-THANH, K. WILLCOX, AND O. GHATTAS, *Model reduction for large-scale systems with high-dimensional parametric input space*, *SIAM J. Sci. Comput.*, 30 (2008), pp. 3270–3288.
- [6] D. CALVETTI AND E. SOMERSALO, *Introduction to Bayesian Scientific Computing: Ten Lectures on Subjective Computing*, *Surv. Tutor. Appl. Math. Sci.* (2), Springer, New York, 2007.
- [7] J. CHRISTEN AND C. FOX, *A general-purpose sampling algorithm for continuous distributions (the t-walk)*, *Bayesian Analysis*, 5 (2010), pp. 263–282.
- [8] T. COLEMAN AND Y. LI, *On the convergence of reflective Newton methods for large-scale nonlinear minimization subject to bounds*, *Math. Program.*, 67 (1994), pp. 189–224.
- [9] T. COLEMAN AND Y. LI, *An interior trust region approach for nonlinear minimization subject to bounds*, *SIAM J. Optim.*, 6 (1996), pp. 418–445.
- [10] Y. EFENDIEV, T. HOU, AND W. LUO, *Preconditioning Markov chain Monte Carlo simulations using coarse-scale models*, *SIAM J. Sci. Comput.*, 28 (2006), pp. 776–803.
- [11] J. EIDSVIK AND H. TJELMELAND, *On Directional Metropolis-Hastings Algorithms*, *Statist. Comput.* 16, Springer, New York, 2006, pp. 93–106.
- [12] P. FELDMANN AND R. FREUND, *Efficient linear circuit analysis by Padé approximation via the Lanczos process*, *IEEE Trans. Computer-Aided Des. Integr. Circuits Syst.*, 14 (1995), pp. 639–649.
- [13] K. GALLIVAN, E. GRIMME, AND P. V. DOOREN, *Padé approximation of large-scale dynamic systems with Lanczos methods*, in *Proceedings of the 33rd IEEE Conference on Decision and Control*, 1994.
- [14] A. GELFAND AND A. SMITH, *Sampling-based approaches to calculating marginal densities*, *J. Amer. Statist. Assoc.*, 85 (1990), pp. 398–409.
- [15] A. GELMAN, *Prior distributions for variance parameters in hierarchical models*, *Bayesian Anal.*, 1 (2006), pp. 515–533.
- [16] M. GREPL, Y. MADAY, N. NGUYEN, AND A. PATERA, *Efficient reduced-basis treatment of nonaffine and nonlinear partial differential equations*, *M2AN Math. Model.*, 41 (2007), pp. 575–605.

- [17] H. HAARIO, M. LAINE, A. MIRA, AND E. SAKSMAN, *DRAM: Efficient Adaptive MCMC*, *Statist. Comput.* 16, Springer, New York, 2006, pp. 339–354.
- [18] H. HAARIO, E. SAKSMAN, AND J. TAMMINEN, *Componentwise adaptation for high dimensional MCMC*, *Comput. Statist.*, 20 (2005), pp. 265–273.
- [19] W. HASTINGS, *Monte Carlo sampling methods using Markov chains and their applications*, *Biometrika*, 57 (1970), pp. 97–109.
- [20] M. HINZE, R. PINNAU, M. ULBRICH, AND S. ULBRICH, *Optimization with PDE Constraints*, Springer, New York, 2009.
- [21] M. HINZE AND S. VOLKWEIN, *Proper orthogonal decomposition surrogate models for nonlinear dynamical systems: Error estimates and suboptimal control*, in *Dimension Reduction of Large-Scale Systems*, Vol. 45, P. Benner, D. Sorensen, and V. Mehrmann, eds., Springer, New York, 2005, pp. 261–306.
- [22] P. HOLMES, J. LUMLEY, AND G. BERKOOZ, *Turbulence, Coherent Structures, and Dynamical Systems and Symmetry*, Cambridge University Press, London, 1996.
- [23] K. ITO AND K. KUNISCH, *Lagrange Multiplier Approach to Variational Problems and Applications*, *Adv. Des. Control* 15, SIAM, Philadelphia, 2008.
- [24] E. JAYNES, *On the rationale of maximum-entropy methods*, *Proceedings of the IEEE*, 70 (2006), pp. 939–952.
- [25] J. KAIPIO AND E. SOMERSALO, *Statistical and Computational Inverse Problems*, *Appl. Math. Sci.* 160, Springer, New York, 2005.
- [26] C. LIEBERMAN, *Parameter and State Model Reduction for Bayesian Statistical Inverse Problems*, Master's thesis, Massachusetts Institute of Technology, School of Engineering, Cambridge, MA, 2009.
- [27] J. LIU, W. WONG, AND A. KONG, *Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes*, *Biometrika*, 81 (1994), pp. 27–40.
- [28] Y. MARZOUK, H. NAJM, AND L. RAHN, *Stochastic spectral methods for efficient Bayesian solution of inverse problems*, *J. Comput. Phys.*, 224 (2007), pp. 560–586.
- [29] Y. MARZOUK AND H. NAJM, *Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems*, *J. Comput. Phys.*, 228 (2009), pp. 1862–1902.
- [30] N. METROPOLIS, A. ROSENBLUTH, M. ROSENBLUTH, A. TELLER, AND E. TELLER, *Equations of state calculations by fast computing machines*, *J. Chem. Phys.*, 21 (1953), pp. 1087–1092.
- [31] N. NGUYEN, G. ROZZA, D. HUYNH, AND A. PATERA, *Reduced basis approximation and a posteriori error estimation for parametrized parabolic PDEs; Application to real-time Bayesian parameter estimation*, in *Computational Methods for Large-Scale Inverse Problems and Quantification of Uncertainty*, L. Biegler, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B. Mallick, Y. Marzouk, L. Tenorio, B. van Bloemen Waanders, and K. Willcox, eds., Wiley, New York, 2010.
- [32] A. NOOR, C. ANDERSEN, AND J. PETERS, *Reduced basis technique for collapse analysis of shells*, *AIAA J.*, 19 (1981), pp. 393–397.
- [33] G. ROBERTS AND J. ROSENTHAL, *Optimal scaling for various Metropolis-Hastings algorithms*, *Statist. Sci.*, 16 (2001), pp. 351–367.
- [34] L. SIROVICH, *Turbulence and the dynamics of coherent structures. Part 1: Coherent structures*, *Quart. Appl. Math.*, 45 (1987), pp. 561–571.
- [35] A. TARANTOLA, *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM, Philadelphia, 2005.
- [36] K. VEROY, C. PRUD'HOMME, D. ROVAS, AND A. PATERA, *A posteriori error bounds for reduced-basis approximation of parametrized noncoercive and nonlinear elliptic partial differential equations*, in *Proceedings of the 16th AIAA Computational Fluid Dynamics Conference*, Orlando, FL, 2003.
- [37] J. WANG AND N. ZABARAS, *Using Bayesian statistics in the estimation of heat source in radiation*, *Internat. J. Heat Mass Transfer*, 48 (2004), pp. 15–29.