AFRL-AFOSR-VA-TR-2021-0102

Bayesian Optimal Experimental Design for Inverse Scattering

Ghattas, Omar
UNIVERSITY OF TEXAS AT AUSTIN
110 INNER CAMPUS DR MN 13
AUSTIN, TX, 78712
USA

08/26/2021
Final Technical Report

Air Force Research Laboratory
Air Force Office of Scientific Research
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
26-08-2021	Final	15 Mar 2019 - 14 Mar 2020

4. TITLE AND SUBTITLE
Bayesian Optimal Experimental Design for Inverse Scattering

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-17-1-0190

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Omar Ghattas

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
UNIVERSITY OF TEXAS AT AUSTIN
110 INNER CAMPUS DR MN 13
AUSTIN, TX 78712
USA

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AF Office of Scientific Research
875 N. Randolph St. Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR RTA2

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-AFOSR-VA-TR-2021-0102

12. DISTRIBUTION/AVAILABILITY STATEMENT
A Distribution Unlimited: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
Bayesian inference provides a rational framework for learning models from data under uncertainty. The optimal experimental design (OED) problem then asks: how do we acquire the \best" data, i.e., the data from which we learn the most? OED is particularly crucial for Air Force systems, since physical or numerical experiments are often highly resource intensive. The goal of Bayesian OED is to optimally design the data acquisition (e.g., sensor locations, what quantities are measured, which experiments), so that uncertainty in the inferred parameters or some predicted quantity of interest is minimized with respect to a criterion. We focus on expected information gain, i.e., the expected Kullback{Leibler divergence between the prior and posterior with respect to the data. Since OED subsumes Bayesian inversion, and Bayesian inversion presents sign cant challenges in high dimensions and with expensive forward models, this leads to challenges of the highest order. Our work addressed these challenges by advancing the state of the art in methods for Bayesian inversion, optimal experimental design, and supporting optimization and machine learning tools.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			FARIBA FAHROO
U	U	U	UU	3	**19b. TELEPHONE NUMBER** *(Include area code)* 426-8429

Standard Form 298 (Rev.8/98)
Prescribed by ANSI Std. Z39.18

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
06-10-20	Final report	15-03-2017 – 14-03-2020

4. TITLE AND SUBTITLE

Bayesian Optimal Experimental Design for Inverse Scattering

5a. CONTRACT NUMBER

5b. GRANT NUMBER

FA9550-17-1-0190

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

George Biros, Omar Ghattas, Youssef Marzouk

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Oden Institute for Computational Engineering & Sciences
The University of Texas at Austin
Stop C0200
Austin, TX 78712-1229

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Office of Scientific Research
Computational Mathematics Program
875 North Randolph Street, Suite 325
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)

AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Bayesian inference provides a rational framework for learning models from data under uncertainty. The optimal experimental design (OED) problem then asks: how do we acquire the "best" data, i.e., the data from which we learn the most? OED is particularly crucial for Air Force systems, since physical or numerical experiments are often highly resource intensive. The goal of Bayesian OED is to optimally design the data acquisition (e.g., sensor locations, what quantities are measured, which experiments), so that uncertainty in the inferred parameters or some predicted quantity of interest is minimized with respect to a criterion. We focus on expected information gain, i.e., the expected Kullback–Leibler divergence between the prior and posterior with respect to the data. Since OED subsumes Bayesian inversion, and Bayesian inversion presents significant challenges in high dimensions and with expensive forward models, this leads to challenges of the highest order. Our work addressed these challenges by advancing the state of the art in methods for Bayesian inversion, optimal experimental design, and supporting optimization and machine learning tools.

15. SUBJECT TERMS

Bayesian inverse problems; optimal experimental design; optimization under uncertainty; data assimilation; machine learning; deep learning; turbulence; wave propagation; advection-dominated transport; heat conduction; reaction-diffusion

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Omar Ghattas
UU	UU	UU	UU	13	19b. TELEPHONE NUMBER *(include area code)* 512-232-4304

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
06-10-20	Final report	15-03-2017 – 14-03-2020

4. TITLE AND SUBTITLE

Bayesian Optimal Experimental Design for Inverse Scattering

5a. CONTRACT NUMBER

5b. GRANT NUMBER

FA9550-17-1-0190

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

George Biros, Omar Ghattas, Youssef Marzouk

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Oden Institute for Computational Engineering & Sciences
The University of Texas at Austin
Stop C0200
Austin, TX 78712-1229

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Office of Scientific Research
Computational Mathematics Program
875 North Randolph Street, Suite 325
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)

AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Bayesian inference provides a rational framework for learning models from data under uncertainty. The optimal experimental design (OED) problem then asks: how do we acquire the "best" data, i.e., the data from which we learn the most? OED is particularly crucial for Air Force systems, since physical or numerical experiments are often highly resource intensive. The goal of Bayesian OED is to optimally design the data acquisition (e.g., sensor locations, what quantities are measured, which experiments), so that uncertainty in the inferred parameters or some predicted quantity of interest is minimized with respect to a criterion. We focus on expected information gain, i.e., the expected Kullback–Leibler divergence between the prior and posterior with respect to the data. Since OED subsumes Bayesian inversion, and Bayesian inversion presents significant challenges in high dimensions and with expensive forward models, this leads to challenges of the highest order. Our work addressed these challenges by advancing the state of the art in methods for Bayesian inversion, optimal experimental design, and supporting optimization and machine learning tools.

15. SUBJECT TERMS

Bayesian inverse problems; optimal experimental design; optimization under uncertainty; data assimilation; machine learning; deep learning; turbulence; wave propagation; advection-dominated transport; heat conduction; reaction-diffusion

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Omar Ghattas
UU	UU	UU	UU	13	19b. TELEPHONE NUMBER *(include area code)* 512-232-4304

Bayesian Optimal Experimental Design for Inverse Scattering

George Biros (UT-Austin), Omar Ghattas (UT-Austin), Youssef Marzouk (MIT)

AFOSR Computational Mathematics grant FA9550-17-1-0190
03/15/2017–3/14/2020

1. Introduction

Bayesian inference provides the most systematic and rational framework for learning complex physical models from complex observational data under uncertainty. In the resulting Bayesian inverse problem, we employ observational data and a model mapping parameters to observables to infer unknown parameters and quantify their uncertainty, in the form of a posterior probability distribution. But this begs the question: how do we acquire the "best" data, i.e., the data from which we can learn the most? This is the optimal experimental design (OED) problem. This is particularly crucial in the setting of scientific and technological systems—including those relevant to the Air Force—since experiments on these systems, whether from physical tests or from high fidelity simulations, are typically highly resource-intensive. In the context of Bayesian inversion, the goal of OED is to find the optimal design of a data acquisition system (e.g., location of sensors, what quantities are measured and how often, what sources are used in each experiment), so that the uncertainty in the inferred parameters—or some predicted QoI derived from the parameters—is minimized with respect to a criterion. In particular, we consider the expected information gain (EIG) criterion, i.e., the expected Kullback–Leibler divergence between the prior and posterior with respect to the data. The problem is formulated as an outer optimization problem with respect to the experimental design and inner Bayesian inversion with respect to the parameters. Since OED includes Bayesian inversion as a subproblem, and Bayesian inversion presents significant challenges in high dimensions and with expensive forward models, this leads to mathematical and computational challenges of the highest order. **Yet these challenges must be overcome if we are to make the best use of data, and ultimately to best learn predictive models of complex systems.**

To address these challenges, advances are needed both on the Bayesian inversion methods that form the "inner problems" of OED and on the "outer" optimal experimental design methods, as well as on optimization and machine learning tools that support both of these tasks. The research we have conducted along these thrusts includes:

- **Advanced methods for Bayesian inversion**, including Stein variational methods (§2); Bayesian inversion with uncertain PDE forward models (§4); software frameworks for Bayesian inversion (§9); inverse scattering in noisy media (§10); sparse reconstruction of reaction-diffusion systems (§12); and coupling techniques for ensemble filtering and likelihood-free inference (§16).
- **Advanced methods for optimal experimental design**, including fast Laplace-based OED (§3), combinatorial optimization for Bayesian OED (§14); and optimal projection of observations (§15).
- **Advanced optimization and machine learning methods that support inversion and OED**, including optimal control and design under uncertainty (§5); product convolution (§6) and domain decomposition methods (§11) for Hessian approximations; scalable high order derivative tensor construction (§7); parsimonious deep neural network surrogates for parameter-to-observable maps (§8); and ODE/PDE-based machine-learning (§13).

These advances are described in the sections below.

2. Stein variational methods for high-dimensional Bayesian inverse problems

The central tasks of Bayesian inversion are to draw samples from the posterior distribution and compute various statistics of quantities of interest with respect to the posterior, e.g., mean, variance, failure probability appearing in uncertainty quantification, risk/sensitivity analysis, and stochastic optimization. Conventional methods (e.g., standard MCMC, tensor-product numerical quadrature, and dictionary-based variational methods) become intractable in the following settings: (1) when the parameters are high- or even infinite-dimensional, which results in the notorious "curse-of-dimensionality", i.e., the computational complexity of the aforementioned methods grows exponentially with respect to the dimension of the parameters; (2) the posterior is highly non-Gaussian because of the strong nonlinearity of the parameter-to-observable maps, even when the prior and the observation noise are assumed Gaussian; (3) solution of the large-scale PDE models at one parameter realization is often very expensive, i.e., taking hours if not days on the most advanced supercomputers, while we need to solve these models many times to draw posterior samples; (4) continuous availability of large-volume, dynamic, and heterogeneous data, and the goal of real-time system control pose considerable demands of fast and scalable computational methods for Bayesian inversion. These challenges are not unique to Bayesian inversion but ubiquitous in optimal data acquisition and stochastic optimal control, for which Bayesian inversion is only the "inner" problem. The key to overcoming these challenges lies in effectively exploiting the problem structure, including geometry, smoothness, and intrinsic low-dimensionality to implicitly or explicitly reduce the dimension of parameter space. The motivation for this reduction lies in the fact that the data are often informative about just a fraction of "modes" of the parameter field, due to ill-posedness of the inverse problem.

A number of current approaches to parameter and model reduction for Bayesian inverse problems, many developed by our team with AFOSR support, have shown considerable promise. These range from Gaussian process approximation of the parameter-to-observable map, to projection-type forward model reductions, to polynomial chaos approximations of the stochastic forward problem, to low-rank approximation of the Hessian of the log-likelihood, to delayed acceptance MCMC methods, to randomize-then-optimize parallel sampling methods, to adaptive sparse quadrature and Hessian-based sparse quadrature combined with reduced basis approximations, to optimal transport-based variational methods using parametric and non-parametric transport maps. Despite these advances, challenges remain in solving Bayesian inference problems in very high dimensions and governed by complex PDEs, such as is encountered in many Air Force problems. The last class of methods, based on optimal transport ideas, have recently emerged as perhaps the most promising approach to such problems in terms of scalability, accuracy, and parallelism [7, 8, 11].

Transport-based variational Bayesian inference is a powerful approach to variational inference that constructs deterministic couplings induced by transport maps between probability distributions, and is particularly well-suited to dynamic data and complex models of evolving systems. It seeks a transport map $T : \mathbb{R}^n \to \mathbb{R}^n$ by minimizing the Kullback–Leibler (KL) divergence $D_{\mathsf{KL}}(T_*\pi_{\mathsf{prior}}|\pi_{\mathsf{post}})$ over a given class of maps $\mathcal{T} \ni T$, where T_* denotes the pushforward of the prior π_{prior} under T. Then, given independent samples $\{\vec{x}_i\}$ from the prior distribution, the transported samples $\{T\vec{x}_i\}$ can be viewed as (approximately) independent samples drawn from the posterior distribution, whose approximation errors are determined by the divergence between $T_*\pi_{\mathsf{prior}}$ and π_{post}. To minimize this divergence, several classes of representations of the transport maps have been investigated, including radial basis function (RBF) kernels in the Stein variational approach and globally supported polynomials structured as Knothe–Rosenblatt lower triangular maps. To address the challenge of the curse-of-dimensionality in the RBF kernel representation, we developed a projected Stein variational Newton (pSVN) method [7], a variant of Stein variational Newton [11], by projecting high-dimensional parameters to an intrinsic low-dimensional subspace informed by the Hessian of the parameter-to-observable map, which is embedded

in the Stein variational framework and demonstrated to be fast and scalable with respect to the number of parameters, samples, and cores in a massively parallel implementation.

3. Fast Laplace–approximation–based OED

We addressed the OED objective of maximizing the expected information gain (EIG), i.e., the expectation (over the data) of the Kullback—Leibler (KL) divergence from posterior to prior [23]. Naive evaluation of EIG is intractable for large-scale problems due to the large number of samples (and thus forward PDE solves) required by double loop Monte Carlo sampling of the KL divergence and its expectation.

To overcome the prohibitive nature of double loop Monte Carlo, we have constructed an approximation of the EIG based on the Laplace approximation of the posterior, which is a Gaussian distribution centered at the point that maximizes the posterior probability (the MAP point) with covariance operator given by the inverse of the Hessian at that point. This permits the KL divergence to be expressed in terms of the log-determinant and trace of the preconditioned Hessian of the log posterior. Rapid spectral decay and a randomized eigensolver allow us to estimate these invariants at a cost—measured in number of forward PDE solves—that is independent of the number of uncertain parameters and experimental design variables. This results in an optimization problem to find the optimal experimental design that includes not only the PDEs governing the inverse problem as constraints, but also the adjoint PDEs and the gradient equation (to define the MAP point), as well as second order forward and adjoint PDEs and an eigenvalue problem for the Hessian operator (to define the trace and determinant). Despite the formidable appearance of this multi-PDE constrained optimization problem, we show that it can be solved at a cost, measured in PDE solves, that is independent of the dimensions of both the (discretized) random parameter field and the observational data, and only weakly dependent on the experimental design dimension. The specific application problem is the optimal placement of sensors to maximize information gain in the inference of an (infinite-dimensional) thermal conductivity field from temperature measurements in a heat-conducting medium.

4. Bayesian inversion with uncertain forward models

We have addressed the problem of Bayesian inversion governed by forward PDEs with random parameters, i.e., uncertain forward models [3]. This is a critical class of problems within the "learning-from-data-through-the-lens-of-models" paradigm: uncertain or inadequate models abound in all areas of science and engineering, and in particular those that characterize complex problems within the Air Force portfolio.

Specifically, given a prior probability of the inversion parameter m, a statistical model of observations d, and a forward model in the form of PDEs with a random parameter field k (representing model uncertainty), we wish to find the posterior probability of the inversion parameter. Besides the usual difficulties of Bayesian inversion, the random PDE forward model creates significant additional challenges. Finding the MAP point alone is a PDE-constrained optimization under uncertainty problem, and fully characterizing the posterior formally requires nested Monte Carlo sampling in random parameter and inversion parameter spaces. For complex forward problems, the inverse problem is thus intractable.

To address this challenge, we linearize the random-parameter-to-observable map, which leads to an explicit formula for the likelihood. New terms, involving the Fréchet derivative (J) of d with respect to k, appear in the likelihood, in particular in the noise covariance operator, where they represent model uncertainty. The explicit computation of J is prohibitive, requiring as many PDE solves as the lesser of the data and random parameter dimensions. Instead, a low rank approximation of J is made efficiently via randomized SVD, typically requiring a small and (k) dimension-independent number of PDE solves.

Finding the MAP point (with respect to m) then requires solving a deterministic many-PDE-constrained optimization problem, with an objective involving the log posterior (enhanced with terms involving J), and with PDE constraints representing the forward problem, the generalized SVD problem for J, and incremental forward and adjoint problems representing actions of J and J^*. Since m is an explicit parameter within this operator, the gradient of the objective with respect to m can be found efficiently via an adjoint method, leading to an efficient MAP point solver. The Hessian of this functional evaluated at the MAP point provides a Laplace approximation to the posterior covariance. All of this can be accomplished at a cost, measured in forward PDE solves, that is independent of the dimensions of the discretized unknown parameter field m and the discretized uncertain parameter field k.

Examples involving a heterogeneous heat conduction PDE and an advection–diffusion–reaction PDE demonstrate both the need to account for model uncertainty (if only to first order) and the tractability and scalability of doing so. Finally, the method we have developed can be employed within a variance reduction framework—specifically, the approximate posterior serves as a control variate—to compute the exact posterior. Future work will investigate this approach and quantify the value of the linearized approximation as a control variate.

5. Fast algorithms for optimal control under uncertainty

We addressed optimization problems with uncertain models—in particular, models governed by PDEs with infinite dimensional random parameter fields [9, 10]. Such problems arise in numerous applications: optimal design and control of systems with stochastic forcing or uncertain material properties or geometry; inverse problems with stochastic forward problems; or Bayesian optimal experimental design problems with the goal of minimizing the uncertainty or maximizing the information gain in the inferred parameters.

Monte Carlo evaluation of the objective as per the popular Sample Average Approximation (SAA) algorithm results in an optimization problem that is constrained by N PDE systems, where N is the number of samples (which typically must number in the tens of thousands, if not more). This results in an optimization problem that is prohibitive to solve, especially when the PDEs are "complex" (e.g., large-scale, nonlinear, coupled) and discretization of the infinite-dimensional parameter field results in a high-dimensional parameter space.

To overcome these challenges, we introduced high-order derivative-based approximations of the parameter-to-objective maps that exploit the structure of these maps, in particular their smoothness, geometry, and low effective dimensionality. Their use as a basis for variance reduction, in combination with randomized linear algebra algorithms, is demonstrated to accelerate Monte Carlo sampling by up to three orders of magnitude and permit efficient solution of large scale stochastic PDE-constrained optimization problems with up to with $O(10^6)$ uncertain parameters and $O(10^6)$ optimization variables. The mathematical/computational framework for optimization under uncertainty we developed was applied to optimal control of turbulent flow [10] and wave propagation-based optimal design of acoustic metamaterials [9].

6. An adaptive product-convolution Hessian approximation for highly data-informed Bayesian inverse problems and OED

We developed an adaptive grid matrix-free operator approximation scheme based on a "product-convolution" interpolation of convolution operators [1]. This scheme is appropriate for operators that are locally translation-invariant, even if these operators are high rank or full rank. Such operators arise in Schur complement methods for solving partial differential equations (PDEs), as Hessians in PDE-constrained optimization and (Bayesian) inverse problems, as integral operators, as covariance operators,

and as Dirichlet-to-Neumann maps. Of particular interest are Bayesian inverse, optimal experimental design, and optimization under uncertainty problems that are in the "highly data-informed" regime. For such problems, one cannot invoke a low-rank approximation of the Hessian, since the eigenvalues of the Hessian represent the information gain from prior to posterior (in eigenvector directions in parameter space). Thus inability to invoke a low-rank approximation means we can no longer exploit advances over the past decade aimed at breaking the curse of dimensionality by exploiting the property that the data inform the model in just a low dimensional manifold—and this manifold does not depend on the parameter dimension. Thus we are back to having to compute the Hessian in a number of PDE solves that scales with the actual parameter dimension.

When do we find ourselves in this highly data-informative regime? Generally, (1) when we have vast amounts of data, such as numerous sensors; (2) when the forward models have little dissipation (and thus information loss), such as high frequency wave propagation and advection-dominated flow and transport; and (3) when numerous sources are used to interrogate a medium or numerous experiments are conducted to acquire data. All three of these settings are critical to AFOSR problems.

Constructing the product-convolution approximation requires computing the impulse responses of the operator to point sources centered on nodes in an adaptively refined grid of sample points. A randomized a posteriori error estimator drives the adaptivity. Once constructed, the approximation can be efficiently applied to vectors using the fast Fourier transform. The approximation can be efficiently converted to hierarchical matrix ($\mathcal{H}$-matrix) format, then inverted or factorized using scalable $\mathcal{H}$-matrix arithmetic. The quality of the approximation degrades gracefully as fewer sample points are used, allowing cheap lower quality approximations to be used as preconditioners. This yields an automated method to construct preconditioners for locally translation-invariant Schur complements. We have directly addressed issues related to boundaries and have proven that our scheme eliminates boundary artifacts. We tested the scheme on a spatially-varying blurring kernel, on the nonlocal component of an interface Schur complement for the Poisson operator, and on the data misfit Hessian for an advection dominated advection-diffusion inverse problem. Numerical results show that the scheme substantially outperforms existing methods.

7. Tensor train construction from tensor actions for compression of large high order derivative tensors

Much of the machinery for making Bayesian inverse problems and optimal experimental design tractable for high-dimensional, expensive-to-evaluate posteriors has been predicated on compact representations (e.g., low rank approximations) of the Hessian of the negative log likelihood (i.e., of the data misfit or loss functional). Recently, however, we have begun to employ higher-order derivatives to accelerate exploration of the posterior by exploiting the geometry of parameter space, for example in the context of Riemannian manifold MCMC methods. Such methods require the efficient manipulation of higher-order derivative tensors, which in turn require their compact representation in a manner that does not require construction of the tensor (much as low rank approximation of Hessians can be constructed from adjoint-based matrix-free actions and randomized SVD algorithms).

We have developed a method for converting tensors into tensor train format (an efficient representation) based on actions of the tensor as a vector-valued multilinear function [2]. Existing methods for constructing tensor trains require access to "array entries" of the tensor and are therefore inefficient or computationally prohibitive if the tensor is accessible only through its action, especially for high order tensors. Our method permits efficient tensor train compression of large high order derivative tensors for nonlinear mappings that are implicitly defined through the solution of a system of equations, as arise in Bayesian inverse problems and optimal experimental design. Array entries of these derivative tensors are not efficiently computable, but we have developed an algorithm to compute their *actions*

efficiently. Such tensors are often amenable to tensor train compression in theory, but until now no practical algorithm existed to convert them into tensor train format. We demonstrated our method by compressing a Hilbert tensor of size $41 \times 42 \times 43 \times 44 \times 45$, and by forming high order (up to 5^{th} order derivatives/6^{th} order tensors) Taylor series surrogates of the noise-whitened parameter-to-output map for a stochastic partial differential equation with boundary output.

8. Parsimonious structure-exploiting deep neural network surrogates for Bayesian inverse problems and optimal experimental design

In both Bayesian inversion and optimal experimental design, the forward model must be solved numerous times—as many as millions—to characterize the uncertainty in the parameters. Bayesian inversion and OED problems governed by large-scale complex models in high parameter dimensions (such as nonlinear PDEs with uncertain infinite dimensional parameter fields) quickly become prohibitive.

Efficient evaluation of the parameter-to-observable (p2o) map, defined by solution of the forward model, is the key to making Bayesian inversion and OED tractable. Surrogate approximations of p2o maps have the potential to greatly accelerate Bayesian inversion and OED, provided that the p2o map can be accurately approximated using (far) fewer forward model solves than would be required for solving the Bayesian inverse or OED problem using the full p2o map. Unfortunately, constructing such surrogates presents significant challenges when the parameter dimension is high and the forward model is expensive. There are really no viable methods that are available for this class of problems—Gaussian processes, radial basis functions, and orthogonal polynomial approximations, which work well in low dimensions, all eventually encounter the curse of dimensionality.

On the other hand, deep neural networks (DNNs) have emerged as leading contenders for overcoming these challenges. We have demonstrated that black box application of DNNs for problems with infinite dimensional parameter fields leads to poor results, particularly in the common situation when training data are limited due to the expense of solving the forward model. However, by constructing a network architecture that is adapted to the geometry and intrinsic low-dimensionality of the p2o map as revealed through adjoint PDEs, one can construct a "parsimonious" DNN surrogate with superior approximation properties with only limited training data [19]. For training the DNN, we introduce the low rank saddle-free Newton method for stochastic optimization, and show that it outperforms first order methods such as Adam and stochastic gradient descent [16, 17].

9. hIPPYlib: An extensible software framework for large-scale deterministic and Bayesian inverse problems governed by PDEs

We have developed an extensible, open-source software framework, `hIPPYlib` (https://hippylib.github.io/) for solution of large-scale deterministic and Bayesian inverse problems governed by partial differential equations (PDEs) with (possibly) infinite-dimensional parameter fields (which are high-dimensional after discretization) [24]. `hIPPYlib` overcomes the prohibitively expensive nature of Bayesian inversion for this class of problems by implementing state-of-the-art scalable algorithms for PDE-based inverse problems that exploit the structure of the underlying operators, notably the Hessian of the log-posterior. The key property of the algorithms implemented in `hIPPYlib` is that the solution of the inverse problem is computed at a cost, measured in linearized forward PDE solves, that is independent of the parameter dimension. The mean of the posterior is approximated by the MAP point, which is found by minimizing the negative log-posterior with an inexact matrix-free Newton-CG method. The posterior covariance is approximated by the inverse of the Hessian of the negative log posterior evaluated at the MAP point. The construction of the posterior covariance is made tractable

by invoking a low-rank approximation of the Hessian of the log-likelihood. Scalable tools for sample generation are also discussed. `hIPPYlib` makes all of these advanced algorithms easily accessible to domain scientists and provides an environment that expedites the development of new algorithms.

10. Inverse scattering with noisy background medium

Inverse scattering finds applications in remote sensing, ocean acoustics, nondestructive testing, geophysics, and sonar and radar imaging. We are interested in recovering a compactly supported unknown scatterer in the presence of a compactly supported random background medium using multi-angle measurements of the scattered far field. We are interested in the low to medium frequency regime (one to forty wavelengths) with multiple illuminations and large (measurement) signal-to-noise ratio . In this regime, the inverse problem is in fact reasonably well posed. The difficulty arises from the presence of the noisy background medium. In our problem setup the L^2 norm of the noisy background can be up to 200% of the target scatterer.

In [5], we presented efficient algorithms for reconstructing such an unknown compact scatterer embedded in a random noisy background medium. We presented and analyzed six different methods for the solution of this inverse problem using different amounts of scattered data and prior information about the random background medium and the scatterer. The different inversion algorithms are defined by a combination of stochastic programming methods and a Bayesian inversion formulation. Our results showed that if we have data for just one instance of the random background medium the best strategy is, perhaps surprisingly, to invert for both random medium and unknown scatterer with appropriate regularization. However, if we have data for multiple instances of the medium it may be worth solving a coupled set of multiple inverse problems. We presented several numerical results for inverting for various scatterer geometries under different inversion scenarios. The main take-away of our study is that (when the inverse problem is well posed) one should invert for both unknown scatterer and random medium, with appropriate, prior-information based regularization. Furthermore, if data from multiple realizations of the background medium is available, then it may be beneficial to combine results from multiple inversions.

11. Domain decomposition preconditioners for Hessian operators in inverse scattering

A key algorithmic component required for constructing fast solvers for OED and Bayesian inversion is the efficient approximation of the Hessian operator and its inverse. Although there is a lot of work on such approximations for discretization schemes based on stencil-like discretizations, there is less work on integral equation formulations—despite the popularity of the latter for scattering problems.

In [6], we proposed domain decomposition preconditioners for the solution of an integral equation formulation of the acoustic forward and inverse scattering problems. We studied both forward and inverse volume problems and propose preconditioning techniques to accelerate the iterative solvers. For the forward scattering problem, we extended the domain decomposition based preconditioning techniques presented for partial differential equations in Cai and Sarkis (1999 SIAM J. Sci. Comput. 21 792–797) to integral equations. We combined this domain decomposition preconditioner with a low-rank correction, which is easy to construct, forming a new preconditioner. For the inverse scattering problem, we used the forward problem preconditioner as a building block for constructing a preconditioner for the Gauss–Newton Hessian. We presented numerical results that demonstrated good performance of both preconditioning strategies and showed that it is possible to construct effective preconditioners for the Hessian operator for problems up to 50 wavelengths.

12. Sparse reconstruction for reaction-diffusion systems

In a series of papers [20, 22] we considered the problem of inverting for unknown initial conditions and coefficients for semi-linear parabolic partial differential equations (in the context of tumor growth modeling). Applications of this model extend to materials science, combustion, non-destructive evaluation, and biophysical transport phenomena.

In particular, we considered numerical schemes for solving an inverse problem for parameter estimation for reaction-diffusion equations from a single time snapshot. We used a PDE-constrained optimization formulation for the inverse problem. The unknown parameters are the reaction coefficient (proliferation), the diffusion coefficient (infiltration), and the initial condition field. The data are noisy and partial observations at a single time point. Like most cases in clinical practice, we used data from a single time snapshot. Moreover, the precise time relative to the initiation of the tumor is unknown, which poses an additional difficulty for inversion. We performed a frozen-coefficient spectral analysis and show that the inverse problem is severely ill-posed. We introduced a regularization on the structure and magnitude of the initial condition. In particular, we assume that the field starts at a few locations (enforced with a sparsity constraint on the initial condition) and that the initial condition magnitude in the maximum norm is equal to one. We solved the resulting optimization problem using an inexact quasi-Newton method combined with a compressive sampling algorithm for the sparsity constraint. Our implementation uses PETSc and AccFFT libraries. We conducted numerical experiments on synthetic and clinical images to highlight the improved performance of our solver over a previously existing solver that uses standard two-norm regularization for the calibration parameters. The existing solver is unable to localize the initial condition. Our new solver can localize the initial condition and recover infiltration and proliferation. In clinical datasets (for which the ground truth is unknown), our solver results in qualitatively different solutions compared to the two-norm regularized solver.

13. Machine learning and PDEs

In a series of papers [13, 26], we considered the intersection of time-dependent PDEs and ODEs, machine learning, and inverse problems.

Residual neural networks can be viewed as the forward Euler discretization of an Ordinary Differential Equation (ODE) with a unit time step. This has recently motivated researchers to explore other discretization approaches and train ODE based networks. However, an important challenge of neural ODEs is their prohibitive memory cost during gradient backpropogation. A recently proposed method claimed that this memory overhead can be reduced from $O(LN_t)$, where N_t is the number of time steps, down to $O(L)$ by solving forward ODEs backwards in time, where L is the depth of the network. In [13], we showed that this approach may lead to several problems: (i) it may be numerically unstable for ReLU/non-ReLU activation and general convolution operators, and (ii) the proposed optimize-then-discretize approach may lead to divergent training due to inconsistent gradients for small time step sizes. We discussed the underlying problems, and to address them we proposed ANODE, an adjoint based Neural ODE framework that avoids the numerical instability related problems noted above, and provides unconditionally accurate gradients. ANODE has a memory footprint of $O(L) + O(N_t)$, with the same computational cost as reversing the ODE solve. We furthermore discussed a memory efficient algorithm that can further reduce this footprint with a trade-off of additional computational cost. We showed results on Cifar-10/100 datasets using ResNet and SqueezeNext neural networks.

In [26], we proposed ANODEV2, which is an extension of neural ODEs that also allows evolution of the neural network parameters, in a coupled ODE-based formulation. The Neural ODE method introduced earlier is in fact a special case of this new more general framework. We presented the formulation of ANODEV2, derived optimality conditions, and implemented a coupled reaction-diffusion-

advection version of this framework in PyTorch. We presented empirical results using several different configurations of ANODEV2, testing them on multiple models on CIFAR-10. We reported results showing that this coupled ODE-based framework is indeed trainable, and that it achieves higher accuracy, as compared to the baseline models as well as the recently-proposed Neural ODE approach.

14. Combinatorial optimization for optimal Bayesian experimental design

Evaluating Bayesian OED objectives—in particular, the expected information gain, equivalent to the mutual information between parameters and observations—remains a challenging task in nonlinear/non-Gaussian problems; in this project we have developed new Monte Carlo methods for this purpose [12]. Yet this evaluation remains only an inner element of the overall optimization process: one must then select the experimental conditions, or sensor locations, or any other "control" parameters of the observation process, to maximize this objective.

The combinatorial perspective on this optimization problem is particularly flexible: candidate sets of observations can easily be selected to satisfy complex constraints, and theoretical guarantees for the *observation selection* problem can be derived in certain settings. In our recent paper [14], we propose and analyze *batch greedy* heuristics for cardinality constrained maximization of non-submodular non-decreasing set functions. Our theoretical guarantees are characterized by the combination of submodularity and supermodularity ratios. We argue how these parameters define tight modular bounds based on incremental gains, and provide a novel reinterpretation of the classical greedy algorithm using the minorize-maximize (MM) principle. Based on that analogy, we also propose a new class of methods exploiting generic modular bounds, and show how the approximation guarantee can be given in terms of the tightness of the bound.

In the context of optimal experimental design for linear Bayesian inverse problems, we then develop bounds for the submodularity and supermodularity ratios of the underlying mutual information objective, in a general setting that allows for correlated observation noise. We also develop novel modular bounds for the mutual information in the sequential design setting, as an alternative to computing incremental gains, and describe certain connections to polyhedral combinatorics. Algorithms using these modular bounds have interesting and important relationships to established statistical notions such as leverage scores and to more recent efforts such as volume sampling. Overall, using these bounds within a batch greedy approach yields excellent performance, as demonstrated on a range of synthetic and realistic examples in [14].

15. Optimal projection of observations

A complementary perspective on OED is that of data compression: finding low-dimensional functionals of the data that preserve mutual information with model parameters or with other quantities of interest. We have shown that optimal linear compression of the data in linear Bayesian inverse problems is tightly linked to the solution of certain generalized eigenvalue problems: in analogy with the parameter-space reduction methods, these data compressions yield optimal approximations of the posterior covariance and posterior mean under the Rao distance between Gaussian distributions. Moreover, for any fixed projection dimension, they maximize the mutual information between the parameters and the projected data. We have now extended this perspective to the goal-oriented setting, where the quantity of interest (QoI) is a function of the inversion parameters [15]. Here, one seeks low-dimensional projections of the data that are optimal in the sense of preserving mutual information between the QoI and the projected data. Again, the solution can be found via the "data-space" analog of certain generalized eigenvalue problems previously used for goal-oriented parameter dimension reduction in Bayesian inverse problems.

16. Coupling techniques for ensemble filtering and likelihood-free inference

In parallel with the OED efforts described above, we have been developing new approaches to inference, initially motivated by sequential/online nonlinear filtering applications [21], but more generally applicable to the problem of likelihood-free inference—i.e., Bayesian inference when evaluations of the likelihood function are intractable or unavailable, but simulation from the conditional distribution of the observations is feasible. This setting is important not only to nonlinear filtering, but also to parameter inference in stochastic forward models and in generative models quite broadly.

We focus on methods that use transport maps to push prior samples, or samples from the joint parameter–data prior, to the desired posterior. These maps can be used for both posterior simulation and conditional density estimation. One of our key results is the development of a new composition-of-maps approach, introduced in [21], that improves finite-sample performance and weakens requirements on the map parameterization, relative to previous conditional sampling approaches that use polynomial maps or, e.g., normalizing flows. This approach can be understood as the natural generalization of the ensemble Kalman filter (EnKF) to nonlinear updates, using stochastic or deterministic couplings. Key issues in this construction center on (i) the estimation of these transport maps from samples, and the choice of estimation objective; and (ii) the parameterization of the transport. We have completed some analysis of both the estimation [4] and approximation [25] problems in the non-composed setting, and extensions to the full composition-of-maps construction are in progress.

Publications

[1] N. Alger, V. Rao, A. Meyers, T. Bui-Thanh, and O. Ghattas. Scalable matrix-free adaptive product-convolution approximation for locally translation-invariant operators. *SIAM Journal on Scientific Computing*, 41(4):A2296–A2328, 2019. https://arxiv.org/abs/1805.06018.

[2] Nick Alger, Peng Chen, and Omar Ghattas. Tensor train construction from tensor actions, with application to compression of large high order derivative tensors. *SIAM Journal on Scientific Computing*, 2020. arXiv preprint arXiv:2002.06244.

[3] Ilona Ambartsumyan and Omar Ghattas. Fast and scalable methods for Bayesian inverse problems with uncertain PDE forward models. *In preparation*, 2020.

[4] R. Baptista, O. Zahm, and Y. M. Marzouk. An adaptive transport framework for joint and conditional density estimation. *arXiv:2009.10303*, 2020.

[5] Carlos Borges and George Biros. Reconstruction of a compactly supported sound profile in the presence of a random background medium. *Inverse Problems*, pages 1–41, 2018.

[6] Carlos Borges and George Biros. A domain decomposition preconditioning for an inverse volume scattering problem. *Inverse Problems*, 36(3), 2020. Also arXiv:1901.09162.

[7] P. Chen, K. Wu, J. Chen, T. O'Leary-Roseberry, and O. Ghattas. Projected Stein variational Newton: A fast and scalable Bayesian inference method in high dimensions. *NeurIPS*, 2019. https://arxiv.org/abs/1901.08659.

[8] Peng Chen and Omar Ghattas. Projected Stein variational gradient descent. *NeurIPS 2020, to appear*, 2020. arXiv preprint arXiv:2002.03469.

[9] Peng Chen, Michael Haberman, and Omar Ghattas. Optimal design of acoustic metamaterial cloaks under uncertainty. *Submitted*, 2020. https://arxiv.org/abs/2007.13252.

[10] Peng Chen, Umberto Villa, and Omar Ghattas. Taylor approximation and variance reduction for PDE-constrained optimal control under uncertainty. *Journal of Computational Physics*, 385:163–186, 2019. https://arxiv.org/abs/1804.04301.

[11] Gianluca Detommaso, Tiangang Cui, Youssef Marzouk, Alessio Spantini, and Robert Scheichl. A Stein variational Newton method. In *Advances in Neural Information Processing Systems*, pages 9187–9197, 2018.

[12] C. Feng and Y. M. Marzouk. A layered multiple importance sampling scheme for focused optimal Bayesian experimental design. *arXiv:1903.11187*, 2019. Submitted to *SIAM/ASA Journal on Uncertainty Quantification*.

[13] Amir Gholaminejad, Kurt Keutzer, and George Biros. ANODE: Unconditionally accurate memory-efficient gradients for neural ODEs. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 730–736. International Joint Conferences on Artificial Intelligence Organization, 7 2019.

[14] J. Jagalur-Mohan and Y. M. Marzouk. Batch greedy maximization of non-submodular functions: guarantees and applications to experimental design. *arXiv:2006.04554*, 2020. Submitted to *Journal of Machine Learning Research*.

[15] J. Jagalur-Mohan and Y. M. Marzouk. Optimal data compression for goal-oriented Bayesian inference. *in preparation*, 2020.

[16] T. O'Leary-Roseberry, N. Alger, and O. Ghattas. Inexact Newton methods for stochastic non-convex optimization with applications to neural network training. *submitted*, 2019. https://arxiv.org/abs/1905.06738.

[17] T. O'Leary-Roseberry, N. Alger, and O. Ghattas. Low rank saddle free Newton: Algorithm and analysis. *submitted*, 2020. https://arxiv.org/abs/2002.02881.

[18] T. O'Leary-Roseberry and O. Ghattas. Ill-posedness and optimization geometry for nonlinear neural network training. *submitted*, 2020. https://arxiv.org/abs/2002.02882.

[19] T. O'Leary-Roseberry, U. Villa, P. Chen, and O. Ghattas. Parsimonious structure-exploiting deep neural network surrogates for parameter-to-output maps governed by PDEs. *In preparation*, 2020.

[20] Klaudius Scheufele, Shashank Subramanian, Andreas Mang, George Biros, and Miriam Mehl. Image-driven biophysical tumor growth model calibration. *SIAM Journal on Scientific Computing*, 42(3):B549–B580, 2020.

[21] A. Spantini, R. Baptista, and Y. M. Marzouk. Coupling techniques for nonlinear ensemble filtering. *Preprint*, 2020. Submitted to *SIAM Review*.

[22] Shashank Subramanian, Klaudius Scheufele, Miriam Mehl, and George Biros. Where did the tumor start? an inverse solver with sparse localization for tumor growth models. *Inverse Problems*, 36(4):045006, 2020.

[23] Umberto Villa and Omar Ghattas. Scalable optimal experimental design for Laplace approximation of expected informaton gain criteria. *In preparation*, 2020.

[24] Umberto Villa, Noemi Petra, and Omar Ghattas. hIPPYlib: An extensible software framework for large-scale inverse problems governed by PDEs; Part I: Deterministic inversion and linearized Bayesian inference. *ACM Transactions on Mathematical Software*, 2020. http://arxiv.org/abs/1909.03948, to appear.

[25] J. Zech and Y. M. Marzouk. Sparse approximation of triangular transports on bounded domains. *arXiv:2006.06994*, 2020. Submitted to *Constructive Approximation*.

[26] Tianjun Zhang, Zhewei Yao, Amir Gholami, Joseph E Gonzalez, Kurt Keutzer, Michael W Mahoney, and George Biros. ANODEV2: A coupled neural ODE framework. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 5152–5162. Curran Associates, Inc., 2019.

AFOSR Deliverables Submission Survey

Response ID:13969 Data

1. Report Upload

Report Type

Final Research Performance Progress Report

Primary Contact Email

Contact email if there is a problem with the report.

omar@oden.utexas.edu

Primary Contact Phone Number

Contact phone number if there is a problem with the report

512-232-4304

Organization / Institution name

The University of Texas at Austin

Grant/Contract Title

The full title of the funded effort.

Bayesian Optimal Experimental Design for Inverse Scattering

Grant/Contract Number

AFOSR assigned control number. It must begin with "FA9550" or "F49620" or "FA2386".

FA9550-17-1-0190

Principal Investigator Name

The full name of the principal investigator on the grant or contract.

Omar Ghattas

Program Officer

The AFOSR Program Officer currently assigned to the award

Fariba Fahroo

Reporting Period Start Date

03/15/2017

Reporting Period End Date

03/14/2020

Abstract

Bayesian inference provides a rational framework for learning models from data under uncertainty. The optimal experimental design (OED) problem then asks: how do we acquire the "best" data, i.e., the data from which we learn the most? OED is particularly crucial for Air Force systems, since physical or numerical experiments are often highly resource intensive. The goal of Bayesian OED is to optimally design the data acquisition (e.g., sensor locations, what quantities are

measured, which experiments), so
that uncertainty in the inferred parameters or some predicted
quantity of interest is minimized with respect to a criterion. We
focus on expected information gain, i.e., the expected
Kullback--Leibler divergence between the prior and posterior with
respect to the data. Since OED subsumes Bayesian inversion,
and Bayesian inversion presents significant challenges in
high dimensions and with expensive forward models, this leads to
challenges of the highest order. Our
work addressed these challenges by advancing the state of the art
in methods for Bayesian inversion, optimal experimental design, and
supporting optimization and machine learning tools.

Distribution Statement

This is block 12 on the SF298 form.

Distribution A - Approved for Public Release

Explanation for Distribution Statement

If this is not approved for public release, please provide a short explanation. E.g., contains proprietary information.

SF298 Form

Please attach your SF298 form. A blank SF298 can be found here. Please do not password protect or secure the PDF The maximum file size for an SF298 is 50MB.

afosr_final_rpt_2020-documentation_page_1.pdf

Upload the Report Document. File must be a PDF. Please do not password protect or secure the PDF . The maximum file size for the Report Document is 50MB.

afosr_final_rpt_2020.pdf

Upload a Report Document, if any. The maximum file size for the Report Document is 50MB.

Archival Publications (published) during reporting period:

See list of publications in report appendix.

New discoveries, inventions, or patent disclosures:

Do you have any discoveries, inventions, or patent disclosures to report for this period?

No

Please describe and include any notable dates

Do you plan to pursue a claim for personal or organizational intellectual property?

Changes in research objectives (if any):

None

Please list personnel associated with the period of performance for this grant (to include any past and present collaborators, students and AFOSR program officers):

Carlos Borges (UT postdoc, now asst. prof, U. of Central Florida
Shashank Subramanian (UT PhD student)
Amir Gholami (UT PhD student, now postdoc at UC Berkeley)
Jayanth Jagalur-Mohan (MIT research scientist)

Robert Ren (MIT PhD student)
Nick Alger (UT postdoc)
Ilona Ambartsumyan (UT postdoc, now at Citadel Securities)
Peng Chen (UT research scientist)
Umberto Villa (UT res scientist, now research asst professor, WUSTL
Keyi Wu (UT PhD student)

Extensions granted or milestones slipped, if any:

None

AFOSR LRIR Number

LRIR Title

Reporting Period

Laboratory Task Manager

Program Officer

Research Objectives

Technical Summary

Funding Summary by Cost Category (by FY, $K)

	Starting FY	FY+1	FY+2
Salary			
Equipment/Facilities			
Supplies			
Total			

Report Document

Report Document - Text Analysis

Report Document - Text Analysis

Appendix Documents

2. Thank You

E-mail user

Oct 07, 2020 15:32:29 Success: Email Sent to: omar@oden.utexas.edu