Advances in Water Resources 82 (2015) 124–138

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Fast computation of uncertainty quantification measures in the geostatistical approach to solve inverse problems

Arvind K. Saibaba [a,*], Peter K. Kitanidis [b,c]

[a] Department of Electrical and Computer Engineering, Tufts University, 196 Boston Avenue, Medford, MA 02155, United States
[b] Department of Civil and Environmental Engineering, Yang & Yamazaki Environment & Energy Building, 473 Via Ortega, Stanford, CA 94305, United States
[c] Institute for Computational and Mathematical Engineering, Huang Engineering Center, 475 Via Ortega, Stanford, CA 94305, United States

ARTICLE INFO

Article history:
Received 13 September 2014
Received in revised form 7 March 2015
Accepted 20 April 2015
Available online 9 May 2015

Keywords:
Inverse problems
Geostatistical approach
Uncertainty quantification
Seismic tomography
Hydraulic tomography

ABSTRACT

We consider the computational challenges associated with uncertainty quantification involved in parameter estimation such as seismic slowness and hydraulic transmissivity fields. The reconstruction of these parameters can be mathematically described as inverse problems which we tackle using the geostatistical approach. The quantification of uncertainty in the geostatistical approach involves computing the posterior covariance matrix which is prohibitively expensive to fully compute and store. We consider an efficient representation of the posterior covariance matrix at the maximum a posteriori (MAP) point as the sum of the prior covariance matrix and a low-rank update that contains information from the dominant generalized eigenmodes of the data misfit part of the Hessian and the inverse covariance matrix. The rank of the low-rank update is typically independent of the dimension of the unknown parameter. The cost of our method scales as $\mathcal{O}(m \log m)$ where m dimension of unknown parameter vector space. Furthermore, we show how to efficiently compute measures of uncertainty that are based on scalar functions of the posterior covariance matrix. The performance of our algorithms is demonstrated by application to model problems in synthetic travel-time tomography and steady-state hydraulic tomography. We explore the accuracy of the posterior covariance on different experimental parameters and show that the cost of approximating the posterior covariance matrix depends on the problem size and is not sensitive to other experimental parameters.

1. Introduction

One of the central challenges in the field of geosciences is to develop computationally efficient statistical methods for optimizing the use of limited and noisy environmental data to accurately estimate heterogeneous subsurface geological properties. In addition, it is necessary to quantify the corresponding predictive uncertainty. Mathematically, imaging can be performed using inverse problems theory, which uses measurements to make inference of system parameters. Efficient algorithms for inverse problems are necessary to solve problems of realistic sizes, quantified by the spatial resolution of the reconstructed parameters and number of measurements available for reconstruction. Using these efficient algorithms scientists can gain better knowledge of soil moisture content, the porosity of geologic formations, distributions of dissolved pollutants, and the locations of oil deposits or buried liquid contaminants. These detailed images can then be used to better locate natural resources, treat pollution, and monitor underground networks associated with geothermal plants, nuclear waste repositories, and carbon dioxide sequestration sites. We aim to solve these problems by employing the geostatistical approach that stochastically models unknowns as random fields and uses Bayes' rule to infer unknown parameters by conditioning on measurements. However, due to high computational costs in identifying small scale features, these methods are challenging. These costs occur because solving inverse problems requires multiple expensive simulations of partial differential equations as well as representing high dimensional random fields, especially on irregular grids and complicated domains. Additional details about the geostatistical approach have been provided in Section 2.

Uncertainty in the context of Bayesian inverse problems is represented by the posterior probability density function. For linear inverse problems, if the measurement noise is additive Gaussian and the prior model is specified by a Gaussian random field, then

* Corresponding author.
E-mail addresses: arvind.saibaba@tufts.edu (A.K. Saibaba), peterk@stanford.edu (P.K. Kitanidis).

http://dx.doi.org/10.1016/j.advwatres.2015.04.012
0309-1708/

the resulting posterior probability density function (PDF) is also Gaussian and is fully specified by calculating the mean (which coincides with the maximum a posteriori, or MAP, estimate), and the posterior covariance matrix. Computing the mean leads to a weighted linear least squares optimization problem, which can be tackled by several efficient numerical algorithms. For nonlinear inverse problems, a linearization of the measurement operator yields a local Gaussian for the posterior PDF. The MAP point can be computed by solving a weighted regularized nonlinear least squares problem and the posterior covariance matrix can be approximated by the inverse of the Hessian of the negative log-likelihood of the posterior PDF computed at the MAP estimate. Although considerable effort has been devoted to computing the MAP estimate (see for example [25,35,36]), relatively fewer number of works have addressed the computation of the posterior covariance matrix. In this work we focus on an efficient representation of the posterior covariance matrix which is a measure of uncertainty associated with the reconstruction of the parameters of interest.

Computing and storing the approximation to the posterior covariance matrix is computationally infeasible because the prior covariance matrices arising from finely discretized fields and certain covariance kernels are dense, and computing the dense measurement operator requires solving many forward PDE problems, which can be computationally intractable. In ill-posed inverse problems, the data is informative only about a low-dimensional manifold in the parameter space. This property has been exploited previously in developing efficient approximate representations to the posterior covariance matrix as the sum of the prior covariance matrix and a low-rank update that contains combined information from both the prior and the data misfit part of the Hessian (see for example, [11,12,18]). The low-rank modification is computed by the solution of a large-scale eigenvalue problem involving the prior-preconditioned Hessian. The prior covariance matrices can be modeled as discrete representations of operators of the form $\mathcal{A}^{-\alpha}$, where $\mathcal{A}$ is a partial differential operator (for e.g., the Laplacian) and $\alpha$is a parameter chosen such that the infinite dimensional formulation is well-posed [39]. Another choice for prior covariance matrices is using Spartan Gibbs random field [23]. In this work we focus on the Matérn class of covariance kernels [40].

The ability of being able to compute measures of uncertainty is extremely important for the field of Optimal Experimental Design (OED), which seeks to determine the experimental setups which maximize the amount of information that can be gained about the parameters of interest. The design variables which control the accuracy of the parameter reconstructions could be the measurements or measurement types, numbers, locations of sources and/or detectors and other experimental conditions. A prevalent approach to OED involves optimizing an objective function which involves a scalar measure of uncertainty associated with the parameter reconstruction (defined on the basis of the posterior covariance matrix) and attempts to minimize this objective function with respect to design parameters. Since during the context of optimizing the experimental setup, the inverse problem has to be solved several times and the resulting uncertainty needs to be estimated at each iteration of the optimization routine, we would like an efficient method for computing the objective function (i.e., the measure of uncertainty). We will provide an efficient method for computing a few of these uncertainty measures. A good review of optimal experimental design in the Bayesian context is provided in [13]. Common optimality criteria which can be used as objective functions include the alphabetic criteria, for example, A-, C-, D-, E-, and T-optimality criteria (these will be defined in Section 6). The definitions of the optimality criteria in the geostatistical context, along with a discussion of physical and statistical significance of these criteria and its applicability in non-Gaussian settings is available in [32].

Contributions: We model the prior covariance matrix Γ_{prior} with entries arising from covariance kernels, as is common practice [25]. Although the resulting covariance matrices are dense, in our previous work [4,36], we have shown that we can obtain the best estimate using techniques (such as FFT based methods and $\mathcal{H}$-matrix approach) to reduce the storage and computational cost from $\mathcal{O}(m^2)$ to $\mathcal{O}(m\ \log\ m)$ (m is the number of unknown parameters). However, except when the number of measurements are small, e.g., $\mathcal{O}(100)$, computing entries of the posterior covariance matrix is computationally impractical. We show how to compute an efficient representation of the posterior covariance matrix as a low-rank modification of Γ_{prior} and the low-rank update is computed efficiently using a randomized algorithm. A major advantage of our approach is that there is great flexibility in experimenting with several covariance kernels, since the prior covariance matrix computations are handled in a black-box fashion. This is a major difference in our work compared to [11,12,18], that we consider directly modeling the entries of the prior covariance matrix using the Matérn class of covariance kernels, instead of modeling the prior as the inverse of a discretized differential operator (such as the Laplacian). Although these two approaches appear disparate, their equivalence has been established in [28].

A second contribution of this paper is that we provide an algorithm for approximating the posterior covariance matrix that does not require forming the square root (or equivalently Cholesky factorization) and inverse of the prior covariance matrix Γ_{prior}. The approach in [11,12,18] considers the prior preconditioned Hessian (defined as $\Gamma_{\text{prior}}^{1/2} H_{\text{red}} \Gamma_{\text{prior}}^{1/2}$, where H_{red} is the Gauss–Newton Hessian of the data misfit term). Computing the square root of a matrix is an expensive operation for finely discretized grids arising from large-scale 3D problems. The work in [11,12,18] avoids this issue by considering priors for which the square-root is explicitly known. Since the matrix square root is not explicitly known for arbitrary covariance matrices, this assumption is very restrictive from a modeling stand point. The algorithm we propose only requires forming matrix–vector products (henceforth, referred to as matvecs). The key idea is to consider an equivalent generalized eigenvalue problem $Ax = \lambda Bx$ where $A$is the Hessian corresponding to the data-misfit and $B = \Gamma_{\text{prior}}^{-1}$ is the inverse prior covariance matrix. The randomized algorithm that we propose for approximating the posterior covariance matrix Γ_{post} is simple to implement, is computationally efficient, and comes with error bounds established in our previous work [37].

Another important contribution of our work is the efficient computation of various measures of uncertainty which leverages the efficient representation of the posterior covariance matrix, written as a low-rank correction to the prior covariance matrix. A second computational burden occurs when the number of measurements are large because of operations on a dense cross-covariance matrix, which scale as $\mathcal{O}(n^3)$, where n is the number of measurements. While some of the criteria (A- and C-) can be evaluated when the number of measurements are small, other criteria (such as D- and E-) are altogether computationally infeasible. However, using an efficient representation of the approximate posterior covariance and using matrix-free techniques, we show how several of these optimality criteria can be computed more efficiently. We note a further advantage of using covariance kernels to model Γ_{prior}: Computing the variance of the posterior covariance requires computing the diagonals of the prior covariance matrix which can be easily computed, when the covariance kernel is spec-