Journal of Computational Physics 229 (2010) 7317–7343

Contents lists available at ScienceDirect

Journal of Computational Physics

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

Hierarchical Bayesian inference for Ill-posed problems via variational method

Bangti Jin [a,*], Jun Zou [b]

[a] Center for Industrial Mathematics, University of Bremen, D-28334 Bremen, Germany
[b] Department of Mathematics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, PR China

ARTICLE INFO

Article history:
Received 6 June 2009
Received in revised form 30 May 2010
Accepted 9 June 2010
Available online 17 June 2010

Keywords:
Variational method
Hierarchical Bayesian inference
Inverse problem
Uncertainty quantification
Cauchy problem

ABSTRACT

This paper investigates a novel approximate Bayesian inference procedure for numerically solving inverse problems. A hierarchical formulation which determines automatically the regularization parameter and the noise level together with the inverse solution is adopted. The framework is of variational type, and it can deliver the inverse solution and regularization parameter together with their uncertainties calibrated. It approximates the posteriori probability distribution by separable distributions based on Kullback–Leibler divergence. Two approximations are derived within the framework, and some theoretical properties, e.g. variance estimate and consistency, are also provided. Algorithms for their efficient numerical realization are described, and their convergence properties are also discussed. Extensions to nonquadratic regularization/nonlinear forward models are also briefly studied. Numerical results for linear and nonlinear Cauchy-type problems arising in heat conduction with both smooth and nonsmooth solutions are presented for the proposed method, and compared with that by Markov chain Monte Carlo. The results illustrate that the variational method can faithfully capture the posteriori distribution in a computationally efficient way.

1. Introduction

In this paper, we are interested in a novel numerical method of Bayesian type for solving inverse problems, especially those related to heat conduction. Inverse problems arise in many disciplines, such as heat conduction [1], mechanics and geophysics, and play an important role in revealing the underlying physical mechanisms. Typically, inverse problems are ill-posed in the sense that the solution lacks a stable dependence on the data. Therefore, their stable and accurate numerical solutions are very challenging. One of the most popular approaches is Tikhonov regularization, which solves a nearby well-posed problem and takes its solution as an approximation. Iterative type methods, such as Landweber method and conjugate gradient method, equipped with a suitable stopping criterion can also be applied.

Bayesian inference approach provides another principled and flexible framework for inverse problems, and has distinct features over classical deterministic regularization methods. Firstly, it yields an ensemble of inverse solutions consistent with the given data, and thus it enables uncertainty quantification of a specific solution. This contrasts sharply with above-mentioned deterministic inverse techniques that content with singling out one solution out of the ensemble. Secondly, it provides a flexible regularization in that the difficult problem of choosing a regularization parameter is resolved through hierarchical modeling. Therefore, it has attracted considerable attention in a wide variety of applied disciplines,

* Corresponding author. Tel.: +49 421 218 63807.
E-mail addresses: kimbtsing@yahoo.com.cn (B. Jin), zou@math.cuhk.edu.hk (J. Zou).

0021-9991/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2010.06.016

e.g. geophysics [2,3], image processing [4] and transient heat conduction [5–7]. For a comprehensive overview of methodological developments, we refer to monographs [2,4].

Hierarchical Bayesian inference has been applied to inverse heat conduction problems [5–8]. The numerical results presented in these studies are very encouraging in that the regularization parameter, noise level and inverse solution can be simultaneously estimated with their uncertainties calibrated. Despite the popularity of hierarchical Bayesian formulations in practical applications and demonstrated performances, the choice of the prior parameter pairs for the hyper-parameters was carried out in a rather ad hoc manner in existing studies. It remains unclear why these formulations work in practice, and no guidelines for their choice were available. Also the Bayesian solution, i.e., posterior probability density function (PPDF) is often numerically sampled, e.g. by Markov chain Monte Carlo (MCMC). However, the MCMC can be computationally expensive, and its convergence might be not easy to diagnose. To circumvent the computational problem, the authors [8] proposed considering the joint *maximum a posteriori* (MAP), and derived an augmented Tikhonov (a-Tikhonov for short) functional that determines the regularization parameter and the noise level along with the solution. Recently some mathematical underpinnings were also provided [9]. However, it yields only one solution like above-mentioned deterministic inverse techniques and does not calibrate the associated uncertainties, and thus it is not completely satisfactory from the point of view of Bayesian analysis.

This paper investigates an alternative framework based on the variational method. The new approach can quantify the uncertainties of the computed solution, thereby overcoming the drawback of the a-Tikhonov method. The approach was first developed in machine learning community [10–12], however, its application to inverse problems seems largely unexplored. This paper will offer some new theoretical results, e.g. its properties in the context of classical inverse theory and convergence properties of the algorithms, to shed some lights on the practical performance. Analyzing the properties of these approximations also provide one means to interrogate the properties of hierarchical formulations. Some heuristic guidelines for the choice of prior parameter pairs in hierarchical Bayesian formulations will be derived, and thus the study sheds new insights on hierarchical Bayesian formulations. The approach is generally applicable to both linear and nonlinear inverse problems with suitable extensions. We shall examine its applicability on severely ill-posed linear and nonlinear Cauchy problems, and carry out a detailed comparison of the new method with the true PPDF explored by the MCMC.

The rest of the paper is structured as follows. Fundamentals of Bayesian inference, hierarchical modeling and associated computational challenge are recalled in Section 2. The variational method for linear inverse problems is described in Section 3, two approximations of the PPDF are derived, and their theoretical properties are analyzed. Algorithms for computing the approximations together with their convergence properties are also discussed. Two generalizations, i.e., ℓ^r prior and nonlinear forward models, are briefly discussed in Section 4. Numerical results for the Cauchy-type problems with smooth and nonsmooth solutions to illustrate their features are presented in Section 5, and compared with that by the MCMC. Finally, we conclude the paper with Section 6.

2. Bayesian inference approach

This section describes the Bayesian framework for a finite-dimensional linear inverse problem

$$\mathbf{Hm} = \mathbf{d}, \tag{1}$$

where $\mathbf{H} \in \mathbb{R}^{n\times m}, \mathbf{m} \in \mathbb{R}^m$ and $\mathbf{d} \in \mathbb{R}^n$ represent system matrix, sought-for solution and given data, respectively. We shall denote $\mathbf{d}^\dagger$ the noise-free data, and assume that $\mathbf{d} = \mathbf{d}^\dagger + \boldsymbol{\omega}$ with $\boldsymbol{\omega}$ being a random vector with mean zero and variance $\sigma_0^2\mathbf{I}$. We shall focus on hyper-parameter treatment within hierarchical models and the associated computational challenge of exploring the posterior state space.

The primary goal of Bayesian inference is to deduce the distribution of the unknown parameters $\mathbf{m}$ conditioned on the data $\mathbf{d}$, i.e., the PPDF $p(\mathbf{m}|\mathbf{d})$. According to Bayes' rule, it is related to $\mathbf{d}$ by

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m})p(\mathbf{m})}{\int p(\mathbf{d}|\mathbf{m})p(\mathbf{m})d\mathbf{m}}.$$

The functions $p(\mathbf{d}|\mathbf{m})$ and $p(\mathbf{m})$ are known as likelihood function and prior probability density, respectively, and they are two basic building blocks of Bayesian inference. Intuitively, it provides a mechanism to integrate the prior knowledge $p(\mathbf{m})$ with the information contained in the data $p(\mathbf{d}|\mathbf{m})$ to achieve the current state of knowledge, the PPDF $p(\mathbf{m}|\mathbf{d})$. The normalizing constant $\int p(\mathbf{d} \mid \mathbf{m})p(\mathbf{m})d\mathbf{m}$ is needed for estimating the credible interval [13], however its computation can be highly nontrivial, especially in high-dimensions. Fortunately, it is often unnecessary to compute the normalizing constant, e.g. the MCMC and optimization, and the PPDF $p(\mathbf{m}|\mathbf{d})$ may be simply evaluated as

$$p(\mathbf{m}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m})p(\mathbf{m}). \tag{2}$$

The PPDF $p(\mathbf{m}|\mathbf{d})$ constitutes a complete description of the inverse problem, and it contains all the information available about $\mathbf{m}$. However, it is not directly informative, and various summarizing statistics, e.g. point estimates and credible intervals, have to be computed. Typical point estimates include posterior mean $\hat{\mathbf{m}}_{\text{pm}}$ and MAP $\hat{\mathbf{m}}_{\text{map}}$. However, we caution that point estimates may not be representative of the PPDF [5,6].