Journal of Computational Physics 301 (2015) 377–393

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

A variational Bayesian approach for inverse problems with skew-t error distributions

Nilabja Guha [a,b,*], Xiaoqing Wu [b], Yalchin Efendiev [a], Bangti Jin [c], Bani K. Mallick [b]

[a] *Department of Mathematics, Texas A&M University, College Station, TX 77843, USA*
[b] *Department of Statistics, Texas A&M University, College Station, TX 77843, USA*
[c] *Department of Computer Science, University College London, Gower Street, London WC1E 6BT, UK*

ARTICLE INFO

Article history:
Received 27 December 2014
Received in revised form 5 June 2015
Accepted 29 July 2015
Available online 3 August 2015

Keywords:
Bayesian inverse problems
Hierarchical Bayesian model
Variational approximation
Kullback–Leibler divergence

ABSTRACT

In this work, we develop a novel robust Bayesian approach to inverse problems with data errors following a skew-*t* distribution. A hierarchical Bayesian model is developed in the inverse problem setup. The Bayesian approach contains a natural mechanism for regularization in the form of a prior distribution, and a LASSO type prior distribution is used to strongly induce sparseness. We propose a variational type algorithm by minimizing the Kullback–Leibler divergence between the true posterior distribution and a separable approximation. The proposed method is illustrated on several two-dimensional linear and nonlinear inverse problems, e.g. Cauchy problem and permeability estimation problem.

1. Introduction

Mathematical models are frequently used in science and engineering, with applications in weather forecasting, climate prediction, chemical kinetics and oil reservoir forecasting. In these mathematical models, there are often model parameters or inputs that have to be estimated from indirect observational data, which constitutes an inverse problem. In practice, observations are inevitably noisy, due to limited precision of measurement sensors. Often the noises exhibit both heavy tail and skewness behavior, hence flexible non-Gaussian distributions are needed to adequately accommodate these features and to fully extract all relevant information. Further, inverse problems are often ill-posed in the sense that the solution lacks a stable dependence on data perturbations, which necessitates the use of regularization techniques [16]. Hence, obtaining a stable and accurate numerical solution is generally a daunting task.

In this work, we shall develop a robust hierarchical Bayesian model which provides a principled yet very flexible framework for solving inverse problems. We incorporate regularization through a suitable prior distribution. Moreover, we allow a heavy-tailed distribution for the error via the likelihood function. The posterior distribution is obtained by using Bayes' theorem. In this way, it yields an ensemble of inverse solutions consistent with the given data to various extents. In particular, it enables uncertainty quantification of a specific inverse solution within the ensemble. Furthermore, it provides a flexible regularization technique by selecting nuisance parameters, e.g., regularization parameter and noise level, adaptively and automatically, through hierarchical Bayesian modeling, via e.g., the full or empirical Bayesian treatment.

* Corresponding author at: Department of Mathematics, Texas A&M University, College Station, TX 77843, USA.
E-mail address: nguha@math.tamu.edu (N. Guha).

http://dx.doi.org/10.1016/j.jcp.2015.07.062
0021-9991/

Inference based on Bayesian hierarchical models provides an attractive tool for solving inverse problems due to its inherent ability to jointly estimate the regularizing parameters, noise level and inverse solution as well as to calibrate their uncertainties. The Gaussian error model is the most popular tool used in the existing Bayesian inverse problem setup. However, in practice, the normality assumption is usually violated because of the presence of skewness and kurtosis in real data [12]. Thus, one may seek more flexible parametric families that are capable of capturing such features of the data. The family of skew-normal distributions to capture the skewness in the data has been widely studied due to its mathematical tractability and appealing probabilistic properties [2,6,4,3]. One further extension of the skew-normal distribution is the skew-t distribution which allows both nonzero skewness and heavy tails in the distribution [8]. For a general background on the skew-normal and related distributions, see [15] for an overview.

Markov chain Monte Carlo (MCMC) methods work particularly well in this setup and is the major engine that has fueled the development and application of Bayesian hierarchical models [14]. Despite the popularity of MCMC based methods, they can be computationally expensive, and its convergence might not be easy to diagnose [10]. In this paper we investigate an alternative approach based on the variational method [20,19,24]. In spite of its wide popularity in the machine learning community, the application of variational methods to inverse problems seems largely unexplored [23,18,17,13]. Tipping and Lawrence (2005) [23] and Jin (2012) [17] developed Bayesian approaches to inverse problems with a heavy-tailed t model to cope data with outliers. Our proposed approach generalizes the method developed in [17] by a robust Bayesian formulation of the inverse problem using the skew-t distribution and a sparse prior structure. The attractive features of this approach are (i) uncertainty quantification of the computed solution, (ii) robustness to data outliers, and (iii) general applicability to both linear and non-linear inverse problems. We shall illustrate the efficiency of our proposed method on several ill-posed inverse problems.

The present work extends prior work [17] in two aspects. First, this work considers the skew-t distribution for the skewness of data errors, whereas [17] considers only the t-distribution. The skewness in the error distribution introduces an extra layer of the computational complexity in developing efficient inference algorithms. Second, this work studies a sparse prior distribution, which is far more complicated than the smoothness prior analyzed in [17]. It is noteworthy that the hierarchical Bayesian model to be developed is generally applicable to linear and nonlinear inverse problems.

The rest of the paper is structured as follows. In Section 2, we formulate the inverse problem and construct the hierarchical model for our case. Then we derive the variational solution and discuss its theoretical properties in Section 3. Later, in Section 4 we illustrate the approach on two ill-posed inverse problems, i.e., the Cauchy problem and the permeability estimation in reservoir simulation, and compare its performance with the more conventional Markov chain Monte Carlo.

2. Methodology

Consider the following finite-dimensional linear inverse problem

$$\mathbf{y} = \mathbf{K}(\mathbf{u}) + \boldsymbol{\epsilon}, \tag{1}$$

where $\mathbf{K}: \mathbb{R}^m \to \mathbb{R}^n$ denotes the forward model, $\mathbf{u} \in \mathbb{R}^m$ is the unknown solution of interest, $\mathbf{K}(\mathbf{u})$ represents the model output from the forward model, and $\boldsymbol{\epsilon}$ is the additive error to the data. Thus, the vector $\mathbf{y} \in \mathbb{R}^n$ represents the noisy data that is observed or measured. Such a problem setup arises in various physical applications. One example is the Cauchy type problem for the Laplace equation, where an elliptic partial differential equation (PDE) is satisfied over a region with some over-specified boundary conditions on a part of the boundary. For example, in case of a re-entrance space shuttle, the temperature field $\mathbf{u}$ on the outer surface is to be estimated from the temperature and the flux measured at the inner surface, while an underlying PDE (steady/transient heat equation) is satisfied. This inverse problem is severely ill-posed and a regularized solution is often sought for. In a Bayesian framework, the data is modeled statistically, and the statistical description is given by the likelihood function $p(\mathbf{y}|\mathbf{u})$, which in turn is dictated by the error distribution of the additive noise $\boldsymbol{\epsilon}$. Furthermore, we need to specify a prior distribution $p(\mathbf{u})$ on the unknown quantity $\mathbf{u}$, reflecting the prior knowledge before collecting the data. Using Bayes' theorem, we obtain the posterior distribution $p(\mathbf{u}|\mathbf{y})$ of the unknown $\mathbf{u}$

$$p(\mathbf{u}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{u})p(\mathbf{u}),$$

where $\propto$ denotes up to a multiplicative normalizing constant. This is the complete Bayesian solution of the inverse problem (1). Hence, we have to specify the likelihood function $p(\mathbf{y}|\mathbf{u})$ and the prior distribution $p(\mathbf{u})$, which constitute the two essential components of constructing the Bayesian solution. In the following two subsections, we describe the likelihood function $p(\mathbf{y}|\mathbf{u})$ and the prior distribution $p(\mathbf{u})$.

2.1. Likelihood function

In order to cope with the presence of outliers and skewness in the observational data $\mathbf{y}$, we choose to model the noisy data by a very flexible class of distributions, i.e., the skew-t distribution. The skew-t distribution, with the scale parameter, skewness parameter, and degrees of freedom, includes Gaussian, centered-t, and skew-normal distribution as special cases. It has been intensively studied since 2001, as an extension of the skew normal family, which was first introduced by Azzalini [2]. There are several different but mathematically equivalent parameterizations of skew-t distributions; see, e.g., Branco and