Journal of Computational Physics 265 (2014) 1–15

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

Minimization for conditional simulation: Relationship to optimal transport

Dean S. Oliver

Centre for Integrated Petroleum Research, Uni Research, Allegaten 41, Box 7800, N-5020 Bergen, Norway

ARTICLE INFO

Article history:
Received 12 February 2013
Received in revised form 21 January 2014
Accepted 26 January 2014
Available online 6 February 2014

Keywords:
Ensemble Kalman filter
Ensemble smoother
Optimal transport
Iterative EnKF
Randomized maximum likelihood
Quasi-linear
Conditional simulation

ABSTRACT

In this paper, we consider the problem of generating independent samples from a conditional distribution when independent samples from the prior distribution are available. Although there are exact methods for sampling from the posterior (e.g. Markov chain Monte Carlo or acceptance/rejection), these methods tend to be computationally demanding when evaluation of the likelihood function is expensive, as it is for most geoscience applications. As an alternative, in this paper we discuss deterministic mappings of variables distributed according to the prior to variables distributed according to the posterior. Although any deterministic mappings might be equally useful, we will focus our discussion on a class of algorithms that obtain implicit mappings by minimization of a cost function that includes measures of data mismatch and model variable mismatch. Algorithms of this type include quasi-linear estimation, randomized maximum likelihood, perturbed observation ensemble Kalman filter, and ensemble of perturbed analyses (4D-Var).

When the prior pdf is Gaussian and the observation operators are linear, we show that these minimization-based simulation methods solve an optimal transport problem with a nonstandard cost function. When the observation operators are nonlinear, however, the mapping of variables from the prior to the posterior obtained from those methods is only approximate. Errors arise from neglect of the Jacobian determinant of the transformation and from the possibility of discontinuous mappings.

1. Introduction

Ensemble-based methods of data assimilation such as the ensemble Kalman filter (EnKF) and the ensemble smoother (ES) rely on an ensemble of samples to provide a Monte Carlo representation of the probability density. Each ensemble member is ideally an independent sample from the appropriate probability density. The objective of the analysis step in data assimilation is to update each ensemble member so that the resulting ensemble provides a representation of the posterior pdf, i.e. after updating each ensemble member is an independent sample from the posterior. Although Bayes' rule specifies how to update the pdf when data are assimilated, it does not specify how individual ensemble members should be updated, so while it is often relatively easy to sample from the prior, it is much more difficult to sample from the posterior. In this paper, we examine the problem of transforming samples from the prior to samples from the posterior, focusing on the connection between minimization-based methods and the optimal transport problem.

E-mail address: dean.oliver@uni.no.

http://dx.doi.org/10.1016/j.jcp.2014.01.048

We consider three distinct probability distributions characterized by a prior probability density, p, a posterior probability density, q, and a probability density q_s for a variable obtained by a transformation of variables distributed according to the prior. Let X be a multivariate normal random variable with probability density p,

$$p(x) = c_p \exp\left(-\frac{1}{2}(x-\mu)^{\mathrm{T}} C_x^{-1}(x-\mu)\right),$$

and let Y be a random vector defined by the transformation $s(X) = Y$. If the transformation is invertible, the probability density of Y is

$$\begin{aligned} q_s(y) &= p\big(s^{-1}(y)\big) \det S^{-1}(y) \\ &= c_p \exp\left(-\frac{1}{2}\big(s^{-1}(y)-\mu\big)^{\mathrm{T}} C_x^{-1}\big(s^{-1}(y)-\mu\big)\right) \det S^{-1}(y) \end{aligned} \tag{1}$$

where

$$S^{-1}(y) = \nabla s^{-\mathrm{T}}.$$

The posterior pdf, for a variable Y with prior distribution $p(y)$ and observation relation $d^o = h(y) + \epsilon_d$ for $\epsilon_d \sim N(0, C_d)$ is

$$q(y) = c_q \exp\left(-\frac{1}{2}(y-\mu)^{\mathrm{T}} C_x^{-1}(y-\mu) - \frac{1}{2}\big(h(y)-d^o\big)^{\mathrm{T}} C_d^{-1}\big(h(y)-d^o\big)\right). \tag{2}$$

In a Monte Carlo approach to uncertainty analysis, the objective is often to sample from the posterior q, but this is difficult in high dimensions, especially when evaluation of h is costly. If, however, sampling from the prior, p, is relatively simple then a straightforward approach to sampling from the posterior, q, is via the transformation $s(X) = Y$, as long as the transformed variables are distributed correctly, or that $q_s(y) = q(y)$. If the mapping is only approximate, importance sampling could potentially be used to correct the discrepancy caused by sampling from $q_s(y)$ [1]. Computation of the weights requires evaluation of $q_s(y)$, however, which can be difficult for high-dimensional nonlinear problems. In Section 3.1, we discuss approximation of the importance weights.

The possibility of computing a map that transforms random variables that are distributed as the prior pdf to random variables that are distributed as the posterior pdf has recently been discussed [2]. In that approach, the transformation is described by multivariate orthogonal polynomials whose coefficients are computed by minimizing the penalized difference between the prior pdf and the approximate transformation-dependent pdf. The method has been applied to the problem of estimating a 2D permeability field from spatially distributed measurements of steady state pressure. Although the approach avoids Markov chain simulation, in the example, the cost of generating an essentially arbitrary number of independent samples using the transformation approach was similar to the cost of generating 200 independent samples from MCMC [2].

Morzfeld and Chorin [3] describe a method of conditional simulation that they describe as a random map. In their approach, a sample Y_j from a posterior distribution, is obtained from a sample $\eta_j \sim N(0, I)$ via the relationship

$$Y_j = \mu + \lambda_j L \eta_j$$

where μ is the set of model variables that maximizes the posteriori probability density. L is recommended to be chosen from a Cholesky decomposition of the inverse Hessian at μ. One then solves

$$F(Y_j) - F(\mu) = \frac{1}{2}\eta_j^{\mathrm{T}}\eta_j \tag{3}$$

for λ_j, where $F(Y)$ is the argument of the exponential of the posteriori probability density, i.e.,

$$F(Y_j) = \frac{1}{2}(Y_j-\mu)^{\mathrm{T}} \Sigma_y^{-1}(Y_j-\mu) + \frac{1}{2}\big(h(Y_j)-d^o\big)^{\mathrm{T}} \Sigma_d^{-1}\big(h(Y_j)-d^o\big).$$

The Jacobian of the transformation is used to weight the particles. A modification of the method to deal with multimodal posterior probability density functions has been discussed [4].

Reich [5] and Cotter and Reich [6] introduce the concept of coupling of probability measures with examples including optimal transport maps for Gaussian distributions. They provide a simplified proof that the optimal transference plan with maximal covariance of variables corresponds to a deterministic coupling with a transfer map that is the gradient of a convex potential. They also make the connection between optimal transport and the data assimilation problem of updating of samples from the prior to samples of the posterior, with explicit discussion of ensemble square-root filters.

The approach that we focus on in this paper is most closely associated with minimization-based conditional simulation methods that have appeared in a wide variety of geoscience fields, including the quasi-linear methodology [7], randomized maximum likelihood (RML) [8,9], perturbed observation ensemble Kalman filter [10–12], iterative ensemble-based methods [13–15], and ensemble of perturbed analyses (4D-Var) [16–18]. In each case, samples, x, of the model variables are drawn