ARTICLE IN PRESS

Mechanical Systems and Signal Processing xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

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

Learning of state-space models with highly informative observations: A tempered sequential Monte Carlo solution

Andreas Svensson *, Thomas B. Schön, Fredrik Lindsten

Department of Information Technology, Uppsala University, 751 05 Uppsala, Sweden

ARTICLE INFO

Article history:
Received 3 February 2017
Received in revised form 21 August 2017
Accepted 12 September 2017
Available online xxxx

Keywords:
Probabilistic modelling
Bayesian methods
Nonlinear system identification
Sequential Monte Carlo
Particle filter
Approximate Bayesian computations
Highly informative observations
Tempering
Wiener-Hammerstein model

ABSTRACT

Probabilistic (or Bayesian) modeling and learning offers interesting possibilities for systematic representation of uncertainty using probability theory. However, probabilistic learning often leads to computationally challenging problems. Some problems of this type that were previously intractable can now be solved on standard personal computers thanks to recent advances in Monte Carlo methods. In particular, for learning of unknown parameters in nonlinear state-space models, methods based on the particle filter (a Monte Carlo method) have proven very useful. A notoriously challenging problem, however, still occurs when the observations in the state-space model are highly informative, i.e. when there is very little or no measurement noise present, relative to the amount of process noise. The particle filter will then struggle in estimating one of the basic components for probabilistic learning, namely the likelihood $p(\text{data}|\text{parameters})$. To this end we suggest an algorithm which initially assumes that there is substantial amount of artificial measurement noise present. The variance of this noise is sequentially decreased in an adaptive fashion such that we, in the end, recover the original problem or possibly a very close approximation of it. The main component in our algorithm is a sequential Monte Carlo (SMC) sampler, which gives our proposed method a clear resemblance to the SMC^2 method. Another natural link is also made to the ideas underlying the approximate Bayesian computation (ABC). We illustrate it with numerical examples, and in particular show promising results for a challenging Wiener-Hammerstein benchmark problem.

1. Introduction

Probabilistic (or Bayesian) modeling and learning offers interesting and promising possibilities for a coherent and systematic description of model and parameter *uncertainty* based on probability theory [31,33]. The computational tools for probabilistic learning in state-space models have lately been developed. In this paper, we study probabilistic learning based on measured data $\{y_1, \ldots, y_T\} \triangleq y_{1:T}$, which we assume to be well described by a nonlinear state-space model with (almost) no measurement noise,

$$x_t|(x_{1:t-1}, \theta) \sim f(x_t|x_{t-1}, u_{t-1}, \theta), \tag{1a}$$
$$y_t = g(x_t), \tag{1b}$$

* Corresponding author.
E-mail addresses: andreas.svensson@it.uu.se (A. Svensson), thomas.schon@it.uu.se (T.B. Schön), fredrik.lindsten@it.uu.se (F. Lindsten).
URL: http://www.it.uu.se/katalog/andsv164 (A. Svensson).

https://doi.org/10.1016/j.ymssp.2017.09.016
0888-3270/

with some unknown parameters $\theta \in \Theta$ which we want to learn. The lack of measurement noise in (1b) gives a deterministic mapping $g : \mathsf{X} \mapsto \mathsf{Y}$ from the unobserved states $x_t \in \mathsf{X}$ to the measurement $y_t \in \mathsf{Y}$, on the contrary to (1a) which encodes uncertainty about x_t, mathematically represented as a probability density f over x_t conditional on x_{t-1} and possibly an exogenous input u_{t-1}. We refer to this uncertainty as *process noise*, but its origin does not have to be a physical noise, but possibly originating from lack of information or model errors. The reasoning and contributions of this paper will be applicable also to the case where the relationship (1b) does contain uncertainty, *measurement noise*, but its variance is much smaller than the process noise. As a general term, we refer to the model as having *highly informative observations*. Furthermore, g could also be allowed to depend on θ and u_t, but we omit that possibility for notational clarity.

Models on the form (1) may arise in several practical situations, for instance in a mechanical system where the measurements can be made with good precision but some unobserved forces are acting on the system. The situation may also appear if the measurements, yet again, can be made with good precision, but the user's understanding of the physical system is limited, which in the probabilistic framework can be modeled as a stochastic element in f.

The model (1) defines, together with priors on θ, a joint probabilistic model $p(y_{1:T}, x_{1:T}, \theta)$. Probabilistic learning of the parameters θ amounts to computing the parameter posterior $p(\theta|y_{1:T})$, where we have conditioned on data $y_{1:T}$ and marginalized over all possible states $x_{1:T}$ (we omit the known $u_{1:T}$ to ease the notation). Although conceptually clear, the computations needed are typically challenging, and almost no cases exist that admit closed-form expressions for $p(\theta|y_{1:T})$.

For probabilistic learning, Monte Carlo methods have proven useful, as outlined in the accompanying paper [37]. The idea underlying these Monte Carlo methods is to represent the distributions of interest, such as the posterior $p(\theta|y_{1:T})$, with samples. The samples can later be used to estimate functions of the parameters θ, such as their mean, variance, etc., as well as making predictions of future outputs y_{T+1}, etc. For state-space models, the particle filter is a tailored algorithm for handling the unknown states x_t, and in particular to compute an unbiased estimate z of the likelihood

$$p(y_{1:T}|\theta) = \int p(y_{1:T}, x_{1:T}|\theta) dx_{1:T}, \tag{2}$$

which is a central object in probabilistic learning, see the accompanying paper [37] for a more thorough introduction (or, e.g., [36,24]). The peculiarity in the problem studied in this paper is the (relative) absence of measurement noise in (1) compared to the process noise level. This seemingly innocent detail is, as we will detail in Section 2.2, a show-stopper for the standard algorithms based on the particle filter, since the quality of the likelihood estimate z tends to be very poor if the model has highly informative observations.

The problem with highly informative observations has a connection to the literature on approximate Bayesian computations (ABC, [3]), where some observations y are available, as well as a model (not necessarily a state-space model) with some unknown parameters θ. In ABC problems, however, the model is only capable of *simulating* new synthetic observations $\hat{y}(\theta)$ and the likelihood $p(y|\theta)$ cannot be evaluated. The ABC idea is to construct a distance metric between the real observations y and the simulated synthetic observations $\hat{y}(\theta)$, and take this distance (which becomes a function of y and θ) as a substitute for $p(y|\theta)$. The accuracy of the approximation is controlled by the metric with higher accuracy corresponding to more informative observations, providing a clear link to the present work.

We propose in this paper a novel algorithm for the purpose of learning θ in (1). Our idea is to start the algorithm by assuming that there *is* a substantial amount of measurement noise which mitigates the computational problems, and then gradually decrease this artificial measurement noise variance simultaneously as the parameters θ are learned. The assumption of artificial measurement noise resembles the ABC methodology. The sequence of gradually decreasing measurement noise variance can be seen as tempering, which we will combine with a sequential Monte Carlo (SMC) sampler [13] to obtain a theoretically sound algorithm which generates samples from the posterior $p(\theta|y_{1:T})$.

In a sense, our proposed algorithm is a combination of the work by [11] on ABC for state-space models and the use of SMC samplers for ABC by [14], resulting in a SMC^2-like algorithm [9].

2. Background on particle filtering and tempering

In this section we will provide some background on particle filters, Markov chain Monte Carlo (MCMC) and related methods. For a more elaborate introduction, please refer to, e.g., [37,10,34]. We will in particular discuss why models on the form (1) are problematic for most existing methods, and also introduce the notion of tempering.

2.1. Particle filtering, PMCMC and SMC^2

The bootstrap particle filter was presented in the early 1990's [20,17] as a solution to the state filtering problem (computing $p(x_t|y_{1:t})$) in nonlinear state-space models. The idea is to propagate a set of N_x Monte Carlo samples $\{x_t^n\}$ along the time dimension $t = 1, 2, \ldots, T$, and for each t the algorithm follows a 3-stage scheme with resampling (sampling ancestor indices a_t^n based on weights w_{t-1}^n), propagation (sampling x_t^n from $x_{t-1}^{a_t^n}$ using (1a)) and weighting (evaluate the 'usefulness' of x_t^n using (1b) and store it as the weight w_t^n). This algorithm will be given as Algorithm 2, and a more elaborate introduction can be found in [37]. The samples are often referred to as particles, and provide an empirical approximation