Signal Processing 142 (2018) 281–291

Contents lists available at ScienceDirect

Signal Processing

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

# Analysis of a nonlinear importance sampling scheme for Bayesian parameter estimation in state-space models

Joaquín Míguez[a,*], Inés P. Mariño[b], Manuel A. Vázquez[a]

[a] Department of Signal Theory & Communications, Universidad Carlos III de Madrid. Avenida de la Universidad 30, 28911 Leganés, Madrid, Spain
[b] Department of Biology and Geology, Physics and Inorganic Chemistry, Universidad Rey Juan Carlos. C/ Tulipán s/n, 28933 Móstoles, Madrid, Spain

ARTICLE INFO

*Article history:*
Received 22 January 2017
Revised 15 June 2017
Accepted 24 July 2017
Available online 25 July 2017

*Keywords:*
Importance sampling
Population Monte Carlo
State space models
Bayesian inference
Adaptive importance sampling
Parameter estimation

ABSTRACT

The Bayesian estimation of the unknown parameters of state-space (dynamical) systems has received considerable attention over the past decade, with a handful of powerful algorithms being introduced. In this paper we tackle the theoretical analysis of the recently proposed *nonlinear* population Monte Carlo (NPMC). This is an iterative importance sampling scheme whose key features, compared to conventional importance samplers, are (i) the approximate computation of the importance weights (IWs) assigned to the Monte Carlo samples and (ii) the nonlinear transformation of these IWs in order to prevent the degeneracy problem that flaws the performance of conventional importance samplers. The contribution of the present paper is a rigorous proof of convergence of the nonlinear IS (NIS) scheme as the number of Monte Carlo samples, $M$, increases. Our analysis reveals that the NIS approximation errors converge to 0 almost surely and with the optimal Monte Carlo rate of $M^{-\frac{1}{2}}$. Moreover, we prove that this is achieved even when the mean estimation error of the IWs remains constant, a property that has been termed *exact approximation* in the Markov chain Monte Carlo literature. We illustrate these theoretical results by means of a computer simulation example involving the estimation of the parameters of a state-space model typically used for target tracking.



## 1. Introduction

The estimation of the static unknown parameters of state-space dynamic models is a classical problem in statistical signal processing [1–6] which has also received considerable attention, very recently, from the computational statistics community [7–9] (see also [10] for a recent survey) partly because of the ubiquity of the problem in science and engineering and partly because of the availability of more powerful computational resources to address it.

The particle Markov chain Monte Carlo (pMCMC) method, originally proposed in [7], has gained popularity in the signal processing community [6,11–14]. This is a Markov chain Monte Carlo (MCMC) algorithm [15] where the target probability density function (pdf) is the posterior density of the unknown parameters conditional on the available observations. This pdf is analytically intractable and, hence, it is approximated (for each element of the chain) via particle filtering [16–20]. The most popular MCMC schemes (including Metropolis and Metropolis-Hastings algorithms) admit a pMCMC implementation. A key feature of these methods is that they have the so-called *exact approximation* property. This means that, even if the acceptance test of the MCMC algorithm is only approximate (since the true target pdf is intractable), the stationary distribution of the Markov chain is still the actual posterior density of the parameters. While popular, pMCMC procedures suffer from the same limitations as regular MCMC schemes [15,21]:

- Convergence of the chain is purely asymptotic and potentially slow: we need to generate a chain that is long enough to converge to its stationary distribution; then we need to generate a sufficiently large number of additional samples in the chain to compute any desired estimators. There are no known convergence rates, neither for the convergence of the chain to its stationary distribution nor for the convergence of the resulting Monte Carlo estimators.
- The Monte Carlo samples in the chain are correlated (hence the difficulty to obtain theoretical convergence rates). Correlation reduces the accuracy of estimators compared to methods that produce independent samples.
- If the target pdf is multimodal, MCMC algorithms may get trapped in local maxima of the function.

An alternative to pMCMC methods is to employ schemes based on importance sampling (IS) [21]. This class of techniques includes

* Corresponding author.
*E-mail addresses:* joaquin.miguez@uc3m.es, jmiguez@ieee.org (J. Míguez), ines.perez@urjc.es (I.P. Mariño), mvazquez@tsc.uc3m.es (M.A. Vázquez).

http://dx.doi.org/10.1016/j.sigpro.2017.07.030

population Monte Carlo (PMC) [22], the sequential Monte Carlo square (SMC$^2$) of [23] or the nested particle filter of [9]. In general, IS methods aim at approximating a complicated, or directly intractable, *target* probability distribution by generating Monte Carlo samples from a simpler *proposal* distribution (different from the target). The samples are assigned importance weights (IWs) in order to account for the mismatch between the target and the proposal. Note that, in the setup of interest in this paper, the target is the posterior distribution of the unknown parameters of the state-space model.

The family of PMC methods includes adaptive IS schemes in which the proposal functions used to generate the samples are improved across a number of iterations [24–26]. The intuition behind this approach is rather straightforward: if we are able to produce an initial approximation of the target probability via IS, using some starting proposal distribution, then we should be able to use that approximation in order to design an improved proposal (e.g., closer to the target) that we can use to apply IS again and obtain an improved approximation. See [27–30] for recent applications, and new developments, of this methodology in statistical signal processing.

The SMC$^2$ method is a generalisation of the iterative batch importance sampling (IBIS) algorithm of [31]. It mimics the standard particle filter, but the Monte Carlo samples are drawn from the space of the (static) parameters and they are sequentially updated using a pMCMC kernel. All these methods, including SMC$^2$, are batch, meaning that the whole record of observations is typically processed many times. A purely recursive version of the SMC$^2$ algorithm has been proposed in [9]. The reduction in computational complexity, however, is obtained at the expense of a reduction in the convergence rate of the algorithm. It is worth mentioning that all these techniques (including pMCMC) can be fit within the theoretical framework of sequential Monte Carlo samplers introduced in [32].

The key feature of IS-based methods is the use of almost-arbitrary proposal functions to generate Monte Carlo samples and the computation of IWs for these samples. While this is a very flexible approach, it suffers from the well-known problem of degeneracy of the IWs [8,18,21,33]: when the target pdf is concentrated in a very small region of the space of the unknowns, the largest IW tends to be orders of magnitude greater than all other IWs. As a result the IS-based scheme practically yields a degenerate one-sample approximation.

In this paper we address the analysis of the nonlinear population Monte Carlo (NPMC) algorithm proposed in [8]. This is a PMC-type method, in which the proposal functions are adapted (intuitively, to be closer to the target) through an iterative scheme. The key feature of the NPMC algorithm is that the IWs undergo a nonlinear transformation to control their variance and, in this way, mitigate the degeneracy problem. In [8] it was proved that the approximation of the target distribution produced *at each iteration* of the NPMC method converges asymptotically, with the number of Monte Carlo samples $M$, and almost surely (a.s.). Therefore, the weight transformation preserves asymptotic convergence, while it has been shown through numerical examples that performance for finite $M$ is consistently improved compared to conventional PMC procedures. The analysis in [8], however

- relies on the exact computation of the IWs, which is not feasible for general state-space models,
- and does not provide explicit convergence rates[1]

In this paper we analyse the performance of NPMC methods for the Bayesian estimation of the unknown parameters of state space models. In the vein of [8], we focus on the convergence of the IS estimators with transformed weights, for a fixed iteration, as the number of samples is increased (we do not analyse the convergence of the iterative process for a fixed number of samples). Based on an unbiasedness property of particle filters, we prove that IS with nonlinearly-transformed IWs also yields asymptotic convergence when the weights are approximate, i.e., computed via a particle filter with a fixed computational budget that introduces non-vanishing errors. In other words, we prove that the nonlinear importance sampler enjoys the same exact approximation property as pMCMC and SMC$^2$ algorithms. Moreover, the analysis of this paper also extends considerably the results of [8] by obtaining an explicit (and almost sure) estimation error rate of order $M^{-\frac{1}{2}+\epsilon}$, where $\epsilon > 0$ is an arbitrarily small constant. This result holds for approximate weights and under mild assumptions typical of classic IS analyses. It is worth mentioning that the analytical approach developed in this paper can be applied, in a rather natural way, to the study of recently proposed PMC-like algorithms [28,34] when the target distribution is the posterior density of the parameters of a state space model.

The rest of the paper is organised as follows. The necessary background material, including notation, state-space models and particle filters, is presented in Section 2. The nonlinear IS scheme and its iterative implementation (the NPMC algorithm) are detailed in Section 3 for the case in which the target probability distribution is the posterior distribution of the unknown parameters of a state-space model. In Section 4 we introduce the new analytical results on the convergence of nonlinear importance samplers, which is the main contribution of the paper. We illustrate the exact approximation property, and numerically compare the NPMC algorithm with a pMCMC scheme through computer simulations for a target tracking model in Section 5. Finally, some brief concluding remarks are made in Section 6.

## 2. Background and problem statement

### 2.1. State-space model

A Markov state-space model consists of two sequences of random variables (r.v.'s), $\{\mathbf{x}_n\}_{n\geq 0}$ and $\{\mathbf{y}_n\}_{n\geq 1}$. The first sequence, $\{\mathbf{x}_n\}$, is termed the system state. We assume it takes values on some space $\mathcal{X} \subseteq \mathbb{R}^{d_x}$, hence $\mathbf{x}_n$ is a random $d_x \times 1$ vector. The state dynamics are described by a prior probability measure $\mathcal{K}_0(\mathrm{d}\mathbf{x}_0)$ and a sequence of Markov kernels $\mathcal{K}_{n,\theta}(\mathrm{d}\mathbf{x}_n|\mathbf{x}_{n-1})$ that depend on a parameter vector $\theta \in \mathsf{S} \subset \mathbb{R}^{d_\theta}$. In this paper, $\theta$ is assumed unknown and modelled as a random vector, with prior pdf $p_0(\theta)$ with respect to (w.r.t.) the Lebesgue measure. The support set of the parameter vector, $\mathsf{S}$, is assumed to be compact.

The state $\mathbf{x}_n$ cannot be observed directly. Instead, some noisy observations $\mathbf{y}_n \in \mathcal{Y} \subseteq \mathbb{R}^{d_y}$, $n = 1, 2, \ldots$, are collected. We note that $\mathbf{y}_n$ is a $d_y \times 1$ vector, with $d_y \neq d_x$ in general.

We assume that the observations are conditionally independent given the system states and the parameter vector $\theta$, with a conditional pdf w.r.t. the Lebesgue measure, denoted $l_{n,\theta}(\mathbf{y}_n|\mathbf{x}_n) > 0$, which depends on the parameter vector $\theta$ as well.

### 2.2. The optimal filter and its Monte Carlo approximation

Let $\mathbf{y}_{1:n} = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$ denote the sequence of observations collected up the time $n$. The posterior probability measure of the state $\mathbf{x}_n$ conditional on the observations $\mathbf{y}_{1:n}$ and the parameter vector $\theta$ is denoted $\pi_{n,\theta}$, i.e., for any Borel set $A \subset \mathcal{X}$,

$$\pi_{n,\theta}(A) = \int_A \pi_{n,\theta}(\mathrm{d}\mathbf{x}) \quad (1)$$

is the posterior probability of the event "$\mathbf{x}_n \in A$", given $\theta$ and $\mathbf{y}_{1:n}$.

[1] Error rates are found in [8] for convergence in probability (not for almost sure convergence) when the IWs are computed exactly.