On the Bias of the SIR Filter in Parameter Estimation of the Dynamics Process of State Space Models

Tiancheng Li[1,2], Sara Rodríguez[1], Javier Bajo[3],
Juan M. Corchado[1], and Shudong Sun[2]

[1] BISITE group, Faculty of Science, University of Salamanca, Salamanca 37008, Spain
[2] School of Mechanical Engineering, Northwestern Polytechnical University, 710072, China
[3] Department of Artificial Intelligence, Technical University of Madrid, 28660, Spain
{t.c.li,srg,corchado}@usal.es, jbajo@fi.upm.es,sdsun@nwpu.edu.cn

Abstract. As a popular nonlinear estimation tool, the sampling importance re-sampling (SIR) filter has been applied with the expectation–maximization (EM) principle, including the typical maximum a posteriori (MAP) estimation and maximum likelihood (ML) estimation, for estimating the parameters of the state space model (SSM). This paper concentrates on an inevitable bias existing in the EM-SIR filter for estimating the dynamics process of the SSM. It is analyzed that the root reason for the bias is the sample impoverishment caused by the resampling procedure employed in the filter. A process noise simulated for the particle propagation that is larger than the real noise involved with the true state will be helpful to counteract sample impoverishment, thereby providing better filtering result. Correspondingly, the EM-SIR filter tends to yield a biased (larger-than-the-truth) estimate of the process noise if it is unknown and needs to be estimated. The bias is elaborated via a straightforward roughening approach by means of both qualitative logical deduction and quantitative numerical simulation. However, it seems hard to fully remove this bias in practice.

Keywords: Particle filter, expectation–maximization, parameter estimation.

1   Introduction

Nonlinear state space models (SSMs) widely exist in the field of control and signal processing, often in which the observed data may be used to estimate the parameter(s) of the model (in the process of estimating the state) if it is unknown. The goal of parameter estimation is to compute an estimate of the true parameter that can provide best match to the observations [1, 2, 3], on a specific principle such as the most typical expectation–maximization (EM) estimation including maximum a posteriori (MAP) estimation and maximum likelihood (ML) estimation. The EM could be implemented in both batch and online manners [4]. This is also the main content of system identification for the state estimation, for which the Bayes filter provides a useful solution. A powerful approximation of the Bayes filter for nonlinear SSMs is based on random sampling, namely Sequential Monte Carlo, which is often known as the particle filter. A variety of particle filters have been developed in the last two decades.

This paper concerns with estimating the state process noise by using the sampling importance resampling (SIR, also referred to as sequential importance sampling (SIS) and resampling, SISR) filter, which is the most commonly used type of particle filters and is the basis of most of the other types of particle filters. Based on SIS and resampling, the SIR filter and its further extension have been applied for various parameter estimation for the state space model since they relaxes linearity and Gaussian assumptions; see some very recent work e.g. [5-7].

However, unwanted bias has been observed in experiments e.g. [1, 3, 8, 9] which could not be ruled out completely; see also [10]. This bias was abstractly attributed to the sample degeneracy [9, 10, 11] of the PF simulation for on-line Bayesian estimation. That is, most of the particles will have negligible weights after a few iterations and, as a result, only few particles really work. The sample degeneracy is an inherent drawback of SIS and to combat it, resampling is often applied that is apparently replicating high-weighted particles to replace low-weighted particles. However, this often comes at the price of generating a large number of particles of the same state (as they are replicated from the same patent particles) and small-weighted particles are discarded, reducing particle diversity and causing sample impoverishment. The more serious the sample degeneracy, the more serious the sample impoverishment after resampling; see further explanation given in [12, 13]. It is fairer to attribute the bias of the SIR filter for parameter estimation to sample impoverishment rather than degeneracy. This is a critical problem for the SIR filter's application for parameter estimation, in addition to the state estimation, but a specific discussion on this problem seems still missing.

In this paper, we clarify the problem from a new perspective based on an in-depth analysis of the side-effect of resampling, with particularly regards to parameter estimation of the dynamics noise of the SSM. It is formally demonstrated that the EM parameter estimation of the state process noise implemented by the SIR filter [3, 8, 9, 14, 15, 16] suffer from a bias, i.e. the estimation of the process noise given by the EM-SIR filter is larger, statistically, than the truth. It is not the intention of this paper to go into details of any specific EM estimator or even the content of the SIR filter but instead, we focus on the statistical property of the SIR filter.

The rest of the paper is organized as follows. Section 2 describes the qualitative logical deduction model and Section 3 presents the quantitative simulation evidences model. The simulation results and discussion are shown in Section 4. Finally, Section 5 concludes the paper.

2   Problem Statement

In the context of the general non-linear SSM, it is often of great interest and significance to estimate the parameter(s) of the model that can be time-varying during the process of estimating the state. The SSM consists of two recursive equations as shown in the following:

$$x_k = f_{k|k-1}\left(x_{k-1}, \theta_k\right) \quad \text{(state transition equation)} \tag{1}$$

$$y_k = g_{k|k}\left(x_k, \beta_k\right) \quad \text{(observation equation)} \tag{2}$$

where k indicates time, x_k denotes the state, y_k denotes the observation, θ_k and β_k denote noises affecting the state Markov process equation $f_{k|k-1}(\cdot)$, and observation equation $g_{k|k}(\cdot)$, respectively and this paper focuses particularly on estimating the static process noise θ_k by the SIR filter, i.e. for $t=1, 2,\ldots, k$, $\theta_t = \theta*$, where $\theta*$ is a static parameter.

Available solutions for the parameter estimation can be classified into two groups: MAP estimation (also called Bayesian estimation or approximate Bayesian computation [5]) and ML estimation. There are several well-established MAP estimation filters for sequentially learning both x_t and θ, such as [1, 6, 17, 18], see the discussion and comparison in [19]. Bias issue has been pointed out in the online MAP estimation and is attributed to the degeneracy problem of the particle filter [9-11]. On the contrary, a different group of work [3, 5-7, 14] particularly focuses on the ML estimation and points out that the bias issue also exists in ML parameter estimation by using the SIR filter. It is demonstrated that this bias is independent of the Bayesian recursion path but instead it can occur at the very first step.

Before we go into the logical deduction and simulation demonstration of the bias, the basic content of ML and MAP estimation is provided in the following context for clarification. Assuming that true noise $\theta*$ is involved with the state, generating observations $y_{1:k}$ and its value is unknown, our goal is to compute point estimates of $\theta*$ from the observations. In general ML principles, the estimate of $\theta*$ is the maximizing argument of the marginal likelihood of the observed data, which can be termed as:

$$\hat{\theta} = \arg\max_{\theta \in \Omega} p\left(y_{1:k} \mid \theta\right) \tag{3}$$

where Ω is a specified parameter space, $y_{1:k} \triangleq (y_1, y_2, ..., y_k)$ denotes the history path of the observation process. For the detail of various implementations of the ML estimation, the reader is refer to [3, 8, 9, 14] and the references therein.

Given a prior distribution of $\theta*$, the MAP estimate is

$$\hat{\theta} = \arg\max_{\theta \in \Omega} \frac{p\left(y_{1:k} \mid \theta\right) g(\theta)}{\int_{\vartheta} p\left(y_{1:k} \mid \vartheta\right) g(\vartheta)} = \arg\max_{\theta \in \Omega} p\left(y_{1:k} \mid \theta\right) g(\theta) \tag{4}$$

As shown, the difference between MAP and the ML estimates is the use of a prior distribution. However, the choice of prior will heavily influence quality of result of MAP as well as the Bayes posterior.

3   Logical Deduction

3.1   EM Estimation of the State Dynamics Noise

Remark 1. $\forall$ two parameters θ_1 and θ_2 that are close-enough to each other in the monotonic domain of Eq. (3): if a better particle approximation of the posterior is obtained by using θ_1 as the particle propagation noise parameter in the SIR filter than by using θ_2, then the estimate of the true state dynamics noise parameter θ^* obtained

by the SIR filter will be more likely closer to θ_1 than to θ_2, i.e. parameters θ_1 is more likely to match the real observations than θ_2 in the sense of likelihood or the posterior

$$p\left(y_{1:k}\,|\theta_1\right) > p\left(y_{1:k}\,|\theta_2\right), \quad \text{or}$$
$$p\left(y_{1:k}\,|\theta_1\right)g\left(\theta\right) > p\left(y_{1:k}\,|\theta_2\right)g\left(\theta\right) \tag{5}$$

This Remark is just for the content of the ML or MAP principle and proof seems unnecessary here. By saying a better particle approximation of the posterior, it means that the underlying particle approximation of the state is closer to the real state distribution and is therefore more likely to match the real observation. θ_1 and θ_2 are limited to be close enough with each other so as to eliminate any local maximum point between them, for monotonicity.

3.2   Direct Roughening

A critical step for the SIR filter is resampling [13] that is designed to reduce the weight variance. As a result of this, many particles may constitute the very similar or the same state i.e. they are replications of the same particle, leading to the so-called sample impoverishment problem. To counteract this problem, one effective solution is to spread the replicated particles by introducing additional noise, namely the roughening. This can be realized in two equivalent ways. One way is to increase the dynamics noise for particle propagation directly which is known as   direct roughening [20], and the other way is to apply roughening separately after resampling similar to the separate roughening scheme proposed in [21]; with similar idea called the 'move step' in [22].

In contrast to the state dynamics given in (1), the Markov process (called propagation) of the ith particle that is perturbed by a roughening noise r in the direct roughening approach, i.e. the proposal function, can be written as:

$$x_k^{(i)} = f_{k|k-1}\left(x_{k-1}^{(i)}, \theta^* + r\right) \tag{6}$$

where θ^* is the dynamics noise involved with the state and the roughening noise r is normally a zero-mean Gaussian $N(0, \Sigma r)$ distribution. In the case of sample impoverishment, the direct roughening helps to improve the approximation quality of the posterior by spreading particles in the state space. It is vital to note that, over roughening (too significant r) however will lead to very dispersive distribution of particles and will conversely reduce the estimation accuracy. As long as the SIR filter suffers from sample impoverishment, we have

Remark 2. $\forall \theta$ that is 'slightly' larger than θ^*: the SIR filter that uses θ as the particle propagation noise will obtain better approximation of the posterior than uses θ^*.

This Remark is no more than a re-statement of the validity of the direct roughening strategy. It is worth nothing that θ is limited to be 'slightly' larger than θ^* to eliminate any local peak between them, for monotonicity.

The theory suggests that the particle filter benefits from a sampling proposal function that that has a 'heavier' tail so that the filter is sensitive to the outliers [12]. This also indicates that the SIR filter can benefit from a comparably large state dynamics noise. Combining Remark 1 and 2, we can arrive at the assertion that the SIR filter will tend to yield a larger-than-the-truth estimate of the state dynamics noise in the forward-only filtering when sample impoverishment occurs. In fact, existing experiments e.g. [1, 8, 9] have observed the bias but attributed the reason to the sample degeneracy without providing clear explanation. Furthermore, the effect of roughening has been verified in e.g. [21, 22]. In the following, further simulations are provided to demonstrate the bias of the EM-SIR filter for estimating the state Markov process noise.

4   Simulation Demonstration

As indicated by remark 1 and 2, a sufficient condition for the occurrence of the bias of the EM-SIR filter in estimation of the dynamics noise is that the filter benefits from the direct roughening approach, which will be demonstrated quantitatively below. Without the loss of generality, we consider estimating the static dynamics noise in a classical 1-dimensional SSM as follows:

$$x_k = 0.5x_{k-1} + \frac{25x_{k-1}}{\left(1+x_{k-1}^2\right)} + 8\cos\left(1.2\left(k-1\right)\right) + \theta \tag{7}$$

$$y_k = 0.05x_k^2 + \beta \tag{8}$$

where Gaussian noise $\theta \sim N(0, Q)$, $\beta \sim N(0, 1)$, Q is the unknown variance of the zero-mean Gaussian Markov process noise to be estimated. Without the loss of generality, the state x_k evolves with the dynamics noise with variance $Q^*=1$.

In order to evaluate the estimation accuracy, the RMSE (root mean square error) is used for evaluation, which is defined as follows:

$$\text{RMSE} = \left(\frac{1}{T}\sum_{k=1}^{T}\left(x_k - \hat{x}_k\right)^2\right)^{1/2} \tag{9}$$

where $\hat{x}_k$ is the estimate of the state x_k which is the mean state of all particles. Nevertheless, the RMSE is unavailable in practice since the true state x_k is unknown. Hence, the RMSD (root mean square discrepancy) between the estimated observation $\hat{y}_k$ and the real observation y_k are defined as a measurement of the likelihood for batch parameter estimation as:

$$\text{RMSD} = \left(\frac{1}{T}\sum_{k=1}^{T}\left(y_k - \hat{y}_k\right)^2\right)^{1/2} \tag{10}$$

where $\hat{y}_k = 0.05\hat{x}_k$.

In order to capture the average performance, the simulation length is set to T=1000 steps and each simulation runs 500 trials.

In the first simulation, four bootstrap SIR filters are designed that apply Q=1, 1.1, 1.2 and 1.5 respectively for particle propagation. $Q=Q^*$=1 is the basic SIR filter and Q=1.1, 1.2 and 1.5 are roughening-enhanced. Their average RMSEs are given in Fig.1, which show that roughening-enhanced SIR filters perform better than the basic SIR filters. Especially when the number of particles is small (e.g. between 20~60), the sample impoverishment is getting more serious and therefore roughening is more helpful. Just as the impoverishment is case specific, the effectiveness of the roughening approach for the SIR filter also varies on case by case basis see the discussion given in [20] therein, which is a multi-dimensional SSM. This indicates that the bias of the SIR filter is also case specific.

In the second simulation, the SIR filter uses different parameters Q (from 0.5 to 4 with interval 0.1) and the same 50 particles. The average RMSE and RMSD results are plotted in Fig.2, which provide more details of the bias of the SIR filter in terms of estimating Q. As indicated, the RMSE result compared with the red dotted line, the SIR filter benefits from a state dynamics noise that is larger, but not too much to prevent overshooting, than that involves with the true state.

It is worth noting that, RSMD is not monotonically proportional with RMSE in the whole domain but instead, the larger the Q used for the particle propagation, the smaller the RMSD. Now that RMSD can be used as a measure for forward-only offline/batch estimation of Q^*, while the maximum likelihood estimate of Q shall constitute a value that is larger than the real Q^*=1 in the most basic sense. This directly demonstrates that the SIR filter can yield a larger-than-the-truth estimate of (the variance of) the state dynamics noise, although here we did not specify the gradient used to search for the optimal parameter.

As stated, more experimental evidences of online parameter estimation can be found in e.g. [1, 8, 9] which are consistent with our assertion. As the importance sampling theory suggest, the simulation result can be interpreted as that the particle filter benefits from a sampling proposal with a heavy tail that is insensitive to the outliers. This in turn can be taken as that a smoother distribution enjoying better particle diversity to alleviate the sample impoverishment. Based on this, we conclude that the sample impoverishment is the primary cause of the bias of the SIR-like filter in parameter estimation of the SSM. Therefore, attentions shall be paid to the use of the SIR-like filter for estimating the dynamics parameters of the SSM, which includes not only the process noise but also the state transition function.
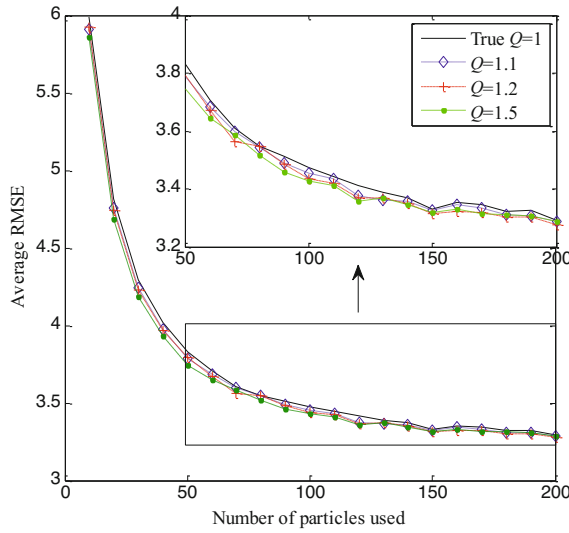
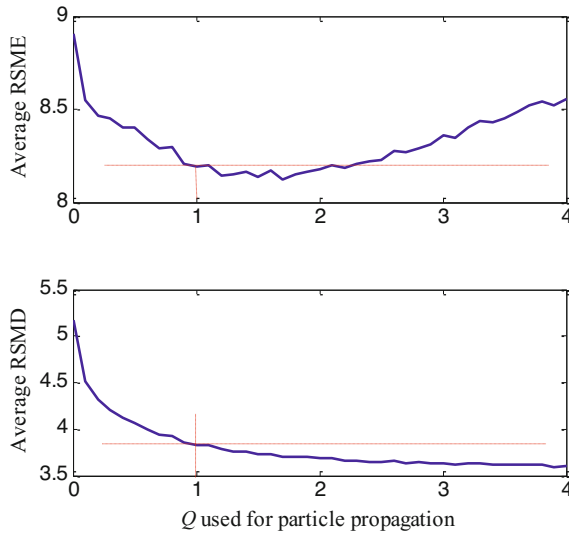**Fig. 1.** RMSE against different number of particles

Fig. 2. RMSE and RMSD against different Q used in the filter

5   Conclusion

Based on the findings that a comparably large noise used for the particle propagation helps to alleviate the sample impoverishment caused by resampling and will therefore produce better approximation of the posterior, we assert that the EM-SIR filter tends to yield a biased (larger than the truth) estimate of the state dynamics noise. This has

been elaborated by means of both qualitative deduction and quantitative simulations. However, the sample degeneracy and impoverishment can hardly be ruled out completely and so is the bias of the EM-SIR filter for the estimation of the dynamics noise. Therefore, one must be careful with the use of the SIR filter for parameter estimation. More concerns on the use and misuse of Bayes filters can be found in [23].

References

1. Storvik, G.: Particle filter for state-space models with the presence of unknown static parameters. IEEE Transactions on Signal Processing 50(2), 281–289 (2002)
2. Schön, T.B., Wills, A., Ninness, System, B.: identification of nonlinear state-space models. Automatica 47(1), 39–49 (2011)
3. Doucet, A., Tadic, V.B.: Parameter estimation in general state-space models using particle methods. Annals of the Institute of Statistical Mathematics 55, 409–422 (2003)
4. Yildirim, S., Jiang, L., Singh, S.S., Dean, T.A.: Calibrating the Gaussian multi-target tracking model. Statistics and Computing, 1–14 (2014)
5. Dean, T.A., Singh, S.S., Jasra, A., Peters, G.W.: Parameter estimation for hidden Markov models with intractable likelihoods. Scand. J. Statist. 41, 970–987 (2014)
6. Lundquist, C., Karlsson, R., Ozkan, E., Gustafsson, F.: Tire radii estimation using a marginalized particle filter. IEEE Transactions on Intelligent Transportation Systems 15(2), 663–672 (2014)
7. Zikmundová, M., Helisová, K.S., Beneš, V.: On the use of particle Markov Chain Monte Carlo in parameter estimation of space-time interacting discs. Methodology and Computing in Applied Probability 16(2), 451–463 (2014)
8. Andrieu, C., Doucet, A., Tadic, V.B.: Online parameter estimation in general state-space models. In: Proc. IEEE CDC/ECC, Seville, Spain, pp. 332–337 (2005)
9. Poyiadjis, G., Doucet, A., Singh, S.S.: Parameter approximations of the score and observed information matrix in state space models with application to parameter estimation. Biometrika 98, 65–80 (2011)
10. Kantas, N., Doucet, A., Singh, S.S., Maciejowski, J.M., Chopin, N.: On particle methods for parameter estimation in state-space models. Statistical Science (2015)
11. Olsson, J., Cappe, O., Douc, R., Moulines, E.: SMC Smoothing with application to parameter estimation in nonlinear state-space models. Bernoulli 14(1), 155–179 (2008)
12. Li, T., Sun, S., Sattar, T.P., Corchado, J.M.: Fight sample degeneracy and impoverishment in particle filters: A review of intelligent approaches. Expert Systems with Applications 41(8), 3944–3954 (2014)
13. Li, T., Bolic, M., Djuric, P.: Resampling methods for particle filtering, IEEE Signal Processing Magazine (to appear, May 2015), doi:10.1109/MSP.2014.2330626
14. Poyiadjis, G., Doucet, A., Singh, S.S.: Maximum likelihood parameter estimation in general state-space models using particle methods. In: Proc. American Stat. Assoc. (2005)

15. Hong, M., Bugallo, M.F., Djurić, P.M.: Joint model selection and parameter estimation by population Monte Carlo simulation. Journal of Selected Topics in Signal Processing 4(3), 526–539 (2010)
16. Míguez, J., Crisan, D., Djurić, P.M.: On the use of sequential Monte Carlo methods for maximum a posteriori sequence estimation and stochastic global optimization. Statistics and Computing 23(1), 91–107 (2013)
17. Liu, J., West, M.: Combined parameters and state estimation in simulation based filtering. In: Doucet, A., de Freitas, N., Gordon, N. (eds.) Sequential Monte Carlo Methods in Practice, pp. 197–223. Springer, New York (2001)
18. Carvalho, C.M., Johannes, M., Lopes, H.F., Polson, N.: Particle learning and smoothing. Statistical Science 25(1), 88–106 (2010)
19. Lopes, H.F., Tsay, R.S.: Particle filters and Bayesian inference in financial econometrics. J. Forecasting 30, 168–209 (2011)
20. Li, T., Sattar, T.P., Han, Q., Sun, S.: Roughening methods to prevent sample impoverishment in the particle PHD filter. In: 16th International Conference on Information Fusion, Istanbul, Turkey, July 10-12 (2013)
21. Gordon, N., Salmond, D., Smith, A.: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. IEE Proc. F Radar Signal Processing 140(2), 107–113 (1993)
22. Fulop, A., Li, J.: Efficient Learning via Simulation: A marginalized resample-move approach. Journal of Econometrics 176, 146–161 (2013)
23. Li, T., Corchado, J.M., Bajo, J., Sun, S., De Paz, J.F.: Do we always need a filter, ar-Xiv:1408.4636 [stat.AP]