Particle filtering with transformed weights

Joaquín Míguez and Eugenia Koblents
Department of Signal Theory & Communications, Universidad Carlos III de Madrid.
Avenida de la Universidad 30, 28911 Leganés (Madrid), Spain.
Email: joaquin.miguez@uc3m.es, ekoblents@tsc.uc3m.es

***Abstract*—Particle filters are simulation-based algorithms for computational inference in dynamical systems that have become very popular over the years in many areas of science and engineering. They are derived from Bayes' theorem and the technique of importance sampling (IS), which entails the approximation of probability measures by way of weighted random samples in the space of interest. As a consequence, particle filters suffer from problems related to the *degeneracy* of these weights, a limitation shared with other IS-based methods. In practice, the weight degeneracy implies that in some scenarios (typically when the dimension of the state space is high or when the likelihood function of the system is sharp) classical particle filters become numerically unstable and fail to converge. In this paper we investigate the application of a recently proposed technique, termed nonlinear importance sampling (NIS), to the design of particle filters. We show how the standard particle filter can be easily modified to incorporate transformed weights computed according to the NIS scheme, then provide a concise proof of convergence of the resulting algorithm, and finally present computer simulation results to illustrate the potential improvement in performance that can be attained.**

I. Introduction

Particle filters (PFs) are simulation-based algorithms for inference in dynamical random systems that have become popular tools in various fields of science and engineering [1], [2], [3], [4].

Essentially, a PF is a recursive algorithm that approximates the sequence of posterior probability distributions of the (unobserved) dynamic variables in the system of interest given a series of observations collected sequentially. The probability distributions are approximated recursively (by exploiting Bayes' theorem) by constructing and sequentially updating a set of random samples in the space of the variables of interest with associated weights. The computation of the weights is carried out according to the *importance sampling* (IS) technique and, as a consequence, PFs suffer from the *weight degeneracy* phenomenon [5], the same as virtually all other IS-based schemes. While occasional resampling steps can keep the problem under control in many systems [5], severe weight degeneracy can occur nevertheless in some scenarios. In that case classical particle filters become numerically unstable and fail to converge. This is a typical phenomenon in high-dimensional models [6] but also, contrary to intuition, in systems where the observations present a high signal-to-noise ratio and hence the posterior probability mass is confined within a very small region of the space of the system variables [7].

We propose to tackle the problem of weight degeneracy using a recently introduced variation of the IS technique called nonlinear importance sampling (NIS) [7]. In classical IS schemes, standard importance weights (IWs) are computed proportional to the ratio of the target probability density function (pdf) and the proposal density used to generate samples, and then they are normalized (so they add up to 1). When the IWs present large variations in their magnitude, they degenerate, meaning that after normalization all IWs but the biggest one become negligible (practically zero). In a NIS scheme, transformed importance weights (TIWs) are additionally computed by applying a nonlinearity to the standard IWs before their normalization. The nonlinearity is chosen to reduce the variation of the IWs and hence mitigate their degeneracy, at the expense of introducing a certain distortion in the approximation of the target probability distribution[1]

The benefit of performing NIS has been advocated for the population Monte Carlo class of methods in [7]. In this paper, we investigate the application of this technique to PFs. Specifically, we introduce a variation of the sequential importance resampling (SIR) [5] algorithm that computes TIWs, then sketch a proof of its asymptotic convergence and finally show computer simulation results that illustrate the sort of performance improvement that can be expected from the new algorithm compared to the standard SIR filter. Namely, we provide numerical evidence of the SIR algorithm with TIWs being robust in a simulation scenario where standard IWs typically degenerate.

The rest of this paper is organized as follows. Section II contains background material on state space Markov dynamical models, the stochastic filtering problem and standard PFs. We intoduce the new algorithm, including a simple analysis of its asymptotics, in Section III. A numerical example is presented in Section IV and, finally, Section V is devoted to the conclusions.

II. Background

A. Filtering in state space Markov models

Many real-world systems of interest can be formally represented as discrete-time stochastic dynamical models in state-space form (state-space models, for conciseness, in the sequel). Let X_t be a $d_x \times 1$ random vector taking values in $\mathbb{R}^{d_x}$ that represents the state of the system of interest at discrete time $t \geq 0$. Typically, this state cannot be observed and, instead, one observes realization of the $d_y \times 1$ random vector Y_t, $t \geq 1$, that takes values in $\mathbb{R}^{d_y}$ and depends (statistically) on X_t.

The probability measure that characterizes the initial state, X_0, is denoted π_0. We assume that π_0 has a density with respect to the Lebesgue measure, denoted p_0, hence $\pi_0(dx_0) =$

[1] In other words, the TIWs are not proper in the sense of [8] or [9].

978-1-4673-3146-3/13/$31.00 ©2013IEEE

$p_0(x_0)dx_0$. The sequence $\{X_t; t \geq 0\}$ is Markov and the state dynamics are governed by a transition kernel $K_t(dx_t, x_{t-1})$. As in the case of π_0, we assume that, for every $x_{t-1} \in \mathbb{R}^{d_x}$, the measure $K_t(dx_t, x_{t-1})$ has a pdf and we denote it $k_t(x_t|x_{t-1})$ (i.e., the conditional density of x_t given x_{t-1}). Finally, the dependence between the observation Y_t and the state X_t is given by the pdf $g_t(y_t|x_t)$. In particular, we assume that the observations are conditionally independent (given the states), i.e., the random vector Y_t is independent of all other observations Y_n, $n \neq t$, and all other states X_n, $n \neq t$, given X_t. To summarize, a state-space Markov model with conditionally independent observations is completely described by the triplet

$$p_0(x_0), \quad k_t(x_t|x_{t-1}) \quad \text{and} \quad g_t(y_t|x_t), \quad \text{where } t = 1, 2, ... \tag{1}$$

Let $x_{t_1:t_2}$, with $t_1 \leq t_2$, be shorthand for the set $x_{t_1:t_2} = \{x_{t_1}, x_{t_1+1}, ..., x_{t_2}\}$ (if $t_1 > t_2$ then $x_{t_1:t_2}$ is the empty set). The problem of stochastic filtering consists in computing the posterior probability distribution of the state X_t given a sequence of observations $Y_{1:t} = y_{1:t}$. This distribution is denoted $\pi_t(dx_t)$. The measure π_t has a pdf, that we denote $p_t(x_t|y_{1:t})$ and is often termed the *filter density*.

The filter density can be written recursively as

$$p_t(x_t|y_{1:t}) \propto g_t(y_t|x_t) \int k_t(x_t|x') p_{t-1}(x'|y_{1:t-1}) dx'.$$

However, the integral in the equation above does not have a closed form except in a few specific cases, namely when the model in Eq. (1) is linear and Gaussian (and p_t is then computed exactly using a Kalman filter; see, e.g., [3]) or when X_t takes values on a finite and discrete space (and the integral reduces to a finite sum).

B. Particle filtering

Particle filters are simulation-based (i.e., Monte Carlo) methods for the recursive approximation of the sequence of measures π_t, $t \geq 0$. The following procedure is often termed standard particle filter or sequential importance resampling (SIR) algorithm [5], [1], [2].

Initialization. Draw N samples $x_0^{(i)}$, $i = 1, ..., N$, from the prior pdf $p_0(x_0)$ and assign them equal weights, $w_0^{(i)} = \frac{1}{N}$.

Recursive update. Given the set $\{x_{t-1}^{(i)}, w_{t-1}^{(i)}; i = 1, ..., N\}$ and a positive integer $R \geq 1$, take the following steps:

1) Draw $\bar{x}_t^{(i)}$ from $k_t(x_t|x_{t-1}^{(i)})$, $i = 1, ..., N$.
2) Update the importance weights,

$$\begin{aligned} \hat{w}_t^{(i)} &= w_{t-1}^{(i)} g_t(y_t|\bar{x}_t^{(i)}), \\ \bar{w}_t^{(i)} &= \frac{\hat{w}_t^{(i)}}{\sum_{j=1}^N \hat{w}_t^{(j)}}, \end{aligned} \tag{2}$$

for $i = 1, ..., N$.
3) If $\frac{t}{R} \in \mathbb{N}$ then resample: for $i = 1, ..., N$ set $x_t^{(i)} = \bar{x}_t^{(j)}$ with probability $\bar{w}_t^{(j)}$, $j \in \{1, ..., N\}$. Reset the weights: $w_t^{(i)} = \frac{1}{N}$ for every i.
Otherwise, if $\frac{t}{R} \notin \mathbb{N}$, set $x_t^{(i)} = \bar{x}_t^{(i)}$ and $w_t^{(i)} = \bar{w}_t^{(i)}$ for $i = 1, ..., N$.

This is arguably the simplest PF, where the new samples $\bar{x}_t^{(i)}$, usually called *particles*, are drawn from the kernel K_t in the model, yielding the simple weight update of Eq. (2). In general, other proposal distributions can be selected to improve the efficiency of the algorithm [1], [3], [4]. The resampling step 3) is necessary to mitigate the weight degeneracy phenomenon [5]. Here we assume a multinomial resampling scheme that is carried out periodically, every R steps, for simplicity. More sophisticated schemes can be plugged in, however, including systematic, residual or minimum variance resampling [10], without affecting the other steps of the algorithm. Hereafter we will refer to the unnormalized weights $\hat{w}_t^{(i)}$ as standard IWs (normalized IWs, correspondingly, for $\bar{w}_t^{(i)}$ and/or $w_t^{(i)}$).

The particles and normalized IWs at time t yield a random approximation of the measure π_t, namely

$$\pi_t^N(dx_t) = \sum_{i=1}^N w_t^{(i)} \delta_{x_t^{(i)}}(dx_t), \tag{3}$$

where δ_x denotes the unit delta measure located at $x \in \mathbb{R}^{d_x}$. For any integrable function $f : \mathbb{R}^{d_x} \to \mathbb{R}^d$, we can approximate the integral $(f, \pi_t) \triangleq \int f(x_t)\pi_t(dx_t)$ as the weighted sum $(f, \pi_t^N) = \int f(x_t)\pi_t^N(dx_t) = \sum_{i=1}^N w_t^{(i)} f(x_t^{(i)})$. In particular, for the identity function $I(x) = x$, (I, π_t^N) is an estimate of the posterior mean of $p_t(x_t|y_{1:t})$.

III. Particle filter with transformed weights

A. Modified SIR algorithm

The weight update of Eq. (2) results from the sequential application of the IS method. In particular, note that the joint posterior pdf of the sequence $x_{0:t}$ given the observations $y_{1:t}$, denoted $p_{0:t}(x_{0:t}|y_{1:t})$, can be decomposed recursively as

$$p_{0:t}(x_{0:t}|y_{1:t}) \propto g_t(y_t|x_t) k_t(x_t|x_{t-1}) p_{0:t-1}(x_{0:t-1}|y_{1:t-1}).$$

If we ignore the resampling steps for the moment, it is apparent that the sequence of particles $x_{0:t}^{(i)}$ is drawn from the product pdf $q_t(x_{0:t}) = p_0(x_0) \prod_{n=1}^t k_n(x_n|x_{n-1})$. Therefore, the standard IW associated to $x_{0:t}^{(i)}$ is

$$\hat{w}_t^{(i)} \propto g_t(y_t|x_t^{(i)}) w_{t-1}^{(i)} \propto \frac{p_{0:t}(x_{0:t}^{(i)}|y_{1:t})}{q_t(x_{0:t}^{(i)})}, \tag{4}$$

i.e., proportional to the ratio of the target pdf, $p_{0:t}(x_{0:t}|y_{1:t})$, and the proposal pdf, $q_t(x_{0:t})$. Note that the filter density p_t is just a marginal of $p_{0:t}$. If there is a resampling step at time $t-1$ then $w_{t-1}^{(i)} = \frac{1}{N}$ is constant and the recursion for the weights is "restarted" at time t.

The nonlinear IS (NIS) method of [7] entails the application of a nonlinear transformation to the ratio in Eq. (4), in such a way that the resulting transformed IWs (TIWs) have a smaller range of variation (and hence a smaller empirical variance) than the original IWs. While various possibilities exist, we restrict our attention here to the *clipping* transformation of [7]. Specifically, let $i_1, ..., i_N$ be a permutation of the indices $\{1, ..., N\}$ such that $\hat{w}_t^{(i_1)} > \cdots > w_t^{(i_N)}$, and choose an integer N_c such that $1 \leq N_c < N$. We compute the TIWs

by way of the nonlinear function $\psi_t^N : \{\hat{w}_t^{(i)}, i = 1, ..., N\} \times \{1, ..., N\} \to [0, 1)$ defined as

$$\psi_t^N(\{\hat{w}_t^{(j)}, j = 1, ..., N\}, i) = \begin{cases} \hat{w}_t^{(i_{N_c})}, & \text{if } \hat{w}_t^{(i)} \geq w_t^{(i_{N_c})} \\ \hat{w}_t^{(i)}, & \text{if } \hat{w}_t^{(i)} < w_t^{(i_{N_c})} \end{cases}. \quad (5)$$

For simplicity, in the sequel we write $\breve{w}_t^{(i)} = \psi^N(\hat{w}_t^{(i)})$ as shorthand for $\breve{w}_t^{(i)} = \psi_t^N(\{\hat{w}_t^{(j)}, j = 1, ..., N\}, i)$. For the class of models in Section II, all the weights are different with probability 1 and this transformation guarantees a decrease the range of variation, i.e., $\frac{\max_i \breve{w}_t^{(i)}}{\min_i \breve{w}_t^{(i)}} < \frac{\max_i \hat{w}_t^{(i)}}{\min_i \hat{w}_t^{(i)}}$.

The SIR algorithm with TIWs using the clipping function of Eq. (5) is identical to the original SIR algorithm except for the weight update step 2), which is extended as follows:

2) Update the importance weights: for $i = 1, ..., N$, first compute the unnormalized IWs

$$\hat{w}_t^{(i)} = w_{t-1}^{(i)} g_t(y_t|\bar{x}_t^{(i)}),$$

and then obtain the clipped TIWs, $\breve{w}_t^{(i)} = \psi^N(\hat{w}_t^{(i)})$. Finally, normalize the TIWs,

$$\bar{w}_t^{(i)} = \frac{\breve{w}_t^{(i)}}{\sum_{j=1}^N \breve{w}_t^{(j)}}, \quad i = 1, ..., N.$$

We hereafter refer to the SIR algorithm with clipped TIWs as ClipSIR for conciseness.

B. Asymptotic convergence

Consider the ClipSIR algorithm with resampling at every time step (i.e., $R = 1$) for simplicity, in such a way that $w_t^{(i)} = 1/N$ for every t and i. The extension of the analysis to the case $R > 1$ is straightforward but notationally cumbersome. Let $f : \mathbb{R}^{d_x} \to \mathbb{R}$ be an arbitrary bounded real function of x_t, namely $\|f\|_\infty = \sup_{x \in \mathbb{R}^{d_x}} |f(x_t)| < \infty$, and define the following discrete random measures,

$$\xi_t^N = \frac{1}{N} \sum_{i=1}^N \delta_{\bar{x}_t^{(i)}}, \ \hat{\pi}_t^N = \sum_{i=1}^N \frac{\hat{w}_t^{(i)} \delta_{\bar{x}_t^{(i)}}}{\sum_{j=1}^N \hat{w}_t^{(j)}}, \ \bar{\pi}_t^N = \sum_{i=1}^N \bar{w}_t^{(i)} \delta_{\bar{x}_t^{(i)}}$$

and π_t^N like in Eq. (3). The measure ξ_t^N is an approximation of $\xi_t = K_t \pi_{t-1}$, which is defined, in turn, as

$$(f, \xi_t) = \int \left(\int f(x_t) K_t(dx_t|x') \right) \pi_{t-1}(dx') = (\bar{f}_t, \pi_{t-1}),$$

where $\bar{f}_t(x_{t-1}) = \int f(x_t) K_t(dx_t|x_{t-1})$ and $\|\bar{f}_t\|_\infty \leq \|f\|_\infty < \infty$. All $\hat{\pi}_t^N$, $\bar{\pi}_t^N$ and π_t^N are approximations of π_t. Additionally, let us denote $g_t^{y_t}(x_t) \triangleq g_t(y_t|x_t)$ for the likelihood function. We have the following result on the asymptotic convergence of the ClipSIR algorithm.

Theorem 1: Let $f : \mathbb{R}^{d_x} \to \mathbb{R}$ be a bounded function, let $T < \infty$ be an arbitrary but finite time horizon and let $Y_{1:T} = y_{1:T}$ be an arbitrary but fixed (deterministic) sequence of observations. Assume that $\forall t \leq T$ $g_t^{y_t} > 0$ is bounded, $(g_t^{y_t}, \xi_t) > 0$ and $\lim_{N \to \infty} \frac{N_c}{N} = 0$. Then, $\lim_{N \to \infty} (f, \pi_t^N) = (f, \pi_t)$ almost surely (a.s.) for every $t = 0, ..., T$.

Outline of the proof: We proceed by induction in t, using the same type of argument as, e.g., in [10]. At $t = 0$, the particles are independent and identically distributed (i.i.d.) samples from the true distribution π_0, hence it is straightforward to show that $\lim_{N \to \infty} (f, \pi_0^N) = (f, \pi_0)$ a.s. As an induction hypothesis, we assume that $\lim_{N \to \infty} (f, \pi_{t-1}^N) = (f, \pi_{t-1})$ a.s., for arbitray t, and try to establish the convergence of (f, π_t^N).

We proceed in four steps. Let us first consider (f, ξ_t^N). A simple triangle inequality yields

$$\begin{aligned} |(f, \xi_t^N) - (f, \xi_t)| \quad \leq \quad & |(f, \xi_t^N) - (f, K_t \pi_{t-1}^N)| \\ & + |(f, K_t \pi_{t-1}^N) - (f, K_t \pi_{t-1})|, \end{aligned} \quad (6)$$

where $\xi_t = K_t \pi_{t-1}$. We note that the second term in (6) is $|(f, K_t \pi_{t-1}^N) - (f, K_t \pi_{t-1})| = |(\bar{f}_t, \pi_{t-1}^N) - (\bar{f}_t, \pi_{t-1})|$, hence $|(f, K_t \pi_{t-1}^N) - (f, K_t \pi_{t-1})| \to 0$ a.s. using the induction hypothesis. As for the first term in (6), let $\mathcal{F}_t$ be the σ-algebra generated by the variables $x_{0:t}^{(i)}$ and $\bar{x}_{1:t}^{(i)}$ for $1 \leq i \leq N$, and define the random variables $\theta_t^{(i)} \triangleq f(\bar{x}_t^{(i)}) - \bar{f}_t(x_{t-1}^{(i)})$, which are centered and conditionally independent given $\mathcal{F}_{t-1}$. It can be shown that $|(f, \xi_t^N) - (f, K_t \pi_{t-1}^N)| = |\frac{1}{N} \sum_{i=1}^N \theta_t^{(i)}| \to 0$ a.s. using the Marcinkiewicz-Zygmund inequality and the Borel-Cantelli lemma. Since both terms on the right-hand side of (6) converge to 0, we obtain $|(f, \xi_t^N) - (f, \xi_t)| \to 0$ a.s.

To analyze the error $|(f, \hat{\pi}_t^N) - (f, \pi_t)|$ we note that $(f, \hat{\pi}_t^N) = (f g_t^{y_t}, \xi_t^N)/(g_t^{y_t}, \xi_t^N)$ and, similarly, $(f, \pi_t) = (f g_t^{y_t}, \xi_t)/(g_t^{y_t}, \xi_t)$. After some manipulations we obtain

$$\begin{aligned} |(f, \hat{\pi}_t^N) - (f, \pi_t)| \quad \leq \quad & \|f\|_\infty \frac{|(g_t^{y_t}, \xi_t) - (g_t^{y_t}, \xi_t^N)|}{(g_t^{y_t}, \xi_t)} \\ & + \frac{|(f g_t^{y_t}, \xi_t^N) - (f g_t^{y_t})|}{(g_t^{y_t}, \xi_t)}. \end{aligned} \quad (7)$$

Since $\|f g_t^{y_t}\|_\infty \leq \|f\|_\infty \|g_t^{y_t}\|_\infty < \infty$, the induction hypothesis combined with (7) yields $|(f, \hat{\pi}_t^N) - (f, \pi_t)| \to \infty$ a.s.

In the third step we bring in the error introduced by the TIWs, namely $|(f, \bar{\pi}_t^N) - (f, \pi_t)|$. Since

$$|(f, \bar{\pi}_t^N) - (f, \pi_t)| \leq |(f, \bar{\pi}_t^N) - (f, \hat{\pi}_t^N)| + |(f, \hat{\pi}_t^N) - (f, \pi_t)|$$

and $|(f, \hat{\pi}_t^N) - (f, \pi_t)| \to 0$ a.s. we only have to show that $|(f, \bar{\pi}_t^N) - (f, \hat{\pi}_t^N)| \to 0$ a.s. However, the latter is immediately obtained from [7, Lemma 3] under the assumption that $\lim_{N \to \infty} N_c/N = 0$.

Finally, in order to prove that $|(f, \pi_t^N) - (f, \pi_t)| \to 0$ a.s. it is sufficient to prove that $|(f, \pi_t^N) - (f, \bar{\pi}_t^N)| \to 0$ a.s. and then apply a triangle inequality again. This is similar to the second step. Conditional on the σ-algebra $\bar{\mathcal{F}}_t$ generated by $x_{0:t-1}^{(i)}$ and $\bar{x}_{1:t}^{(i)}$, $1 \leq i \leq N$, the random variables $\bar{\theta}_t^{(i)} = f(x_t^{(i)}) - (\bar{f}, \bar{\pi}_t^N)$ are centered and independent, hence we combine again the Marcinkiewicz-Zygmund inequality and the Borel-Cantelli lemma to obtain

$$|(f, \pi_t^N) - (f, \bar{\pi}_t^N)| = \left| \frac{1}{N} \sum_{i=1}^N \bar{\theta}_t^{(i)} \right| \to 0 \text{ a.s.} \quad \blacksquare$$

IV. EXAMPLE

We consider a simple problem of navigation in a 2D region using radio signal strength (RSS) and acceleration observations. The state vector at time t is $X_t =$

$(P_{1,t}, P_{2,t}, V_{1,t}, V_{2,t})^\top \in \mathbb{R}^4$, where $(P_{1,t}, P_{2,t})^\top$ is the position and $(V_{1,t}, V_{2,t})^\top$ is the velocity, respectively, of the platform. The prior distribution of X_0 is zero-mean Gaussian with a 4×4 identity covariance matrix $\mathbf{I}_4$. This is denoted as $p_0(x_0) = N(x_0; 0, \mathbf{I}_4)$.

At every time t the acceleration of the platform is assumed measured and denoted $a_t = (a_{1,t}, a_{2,t})^\top$. For the simulation, we generate $\{a_t; t \geq 1\}$ as an i.i.d. sequence with pdf $N(a_t; 0, \sigma_a^2 \mathbf{I}_2)$ and variance $\sigma_a^2 = \frac{1}{5}$. The transition kernel K_t is also Gaussian, namely

$$k_t(x_t|x_{t-1}) = N\left(x_t|\mathbf{A}_\tau x_{t-1} + \mathbf{B}_\tau a_t, \sigma_x^2 \mathbf{B}_\tau \mathbf{B}_\tau^\top\right),$$

where the variance $\sigma_x^2 = 1$, the matrices $\mathbf{A}_\tau = \begin{pmatrix} \mathbf{I}_2 & \tau \mathbf{I}_2 \\ \mathbf{0}_2 & \mathbf{I}_2 \end{pmatrix}_{4\times 4}$ and $\mathbf{B}_\tau = \begin{pmatrix} \frac{\tau^2}{2}\mathbf{I}_2 \\ \tau \mathbf{I}_2 \end{pmatrix}_{4\times 2}$ and the time-increment $\tau = 1/2$ are known. Note that the process noise in this model represents the error in the measurement of a_t.

RSS observations are collected from $d_y = 4$ beacons, hence $Y_t = (Y_{1,t}, ..., Y_{4,t})^\top$, and the density g_t is assumed Gaussian as well. Specifically,

$$g_t(y_t|x_t) = \prod_{i=1}^{4} N\left(y_{i,t}; 10\log_{10}\left(\frac{S_0}{\left\|\begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix} - b_i\right\|^\alpha}, \sigma_y^2\right)\right),$$

where $S_0 = 1$ is the power of the signal transmitted from the beacons, $b_i \in \mathbb{R}^2$ is the position of the i-th beacon, $\alpha = 2$ is the path-loss coefficient (we assume open space) and σ_y^2 is the variance of the observational noise variables, assumed i.i.d. for the four beacons. The beacons are located in the positions $b_1 = (600, 0)^\top$, $b_2 = (0, 600)^\top$, $b_3 = (-600, 0)^\top$ and $b_4 = (0, -600)^\top$, with units in meters.

We have selected three different values of the observation variance, namely $\sigma_y^2 = 5 \times 10^{-3}, 1, 20$, and for each variance we have run 400 independent simulations with the SIR and ClipSIR algorithms with parameters $N = 800$ particles, $N_c = \lceil N^{1/3} \rceil = 10$ clipped weights and resampling period $R = 10$. For each run, we have generated an independent platform trajectory and an independent collection of RSS data. Then, we have applied each filter with the same data and initial particles $x_0^{(i)}$, $i = 1, ..., N$. As an outcome, for each PF and each simulation, we have recorded the square-errors

$$\varepsilon_t^2(j) = \left\| \left(\sum_{i=1}^{N} \bar{w}_t^{(i)}(j) \bar{x}_t^{(i)}(j) \right) - x_t(j) \right\|^2, \quad j = 1, ..., 400,$$

for each discrete-time step $t = 0, 1, ..., 800$, where $\bar{x}_t^{(i)}(j)$ denotes the i-th particle of the filter before resampling in the j-th simulation (with associated weight $\bar{w}_t^{(i)}(j)$), and $x_t(j)$ is the true state vector at time t in the j-th simulation.

Table I displays the results of the simulations. The mean error $\bar{\varepsilon}^2$ is the average (over $j = 1, ..., 400$ and $t = 0, ..., 800$) of the $\varepsilon_t^2(j)$'s. The empirical variance of the $\varepsilon_t^2(j)$'s, denoted $\mathrm{Var}(\varepsilon^2)$, is also shown. The ClipSIR algorithm yields a large performance improvement when $\sigma_y^2 = 5 \times 10^{-3}$. This is the scenario where the IWs are affected by severe degeneracy. For the other two cases the algorithms attain similar performance.

TABLE I. EMPIRICAL ERRORS: MEAN ($\bar{\varepsilon}^2$) AND VARIANCE ($\mathrm{VAR}(\varepsilon^2)$).

	$\bar{\varepsilon}^2$ SIR	$\bar{\varepsilon}^2$ ClipSIR	$\mathrm{Var}(\varepsilon^2)$ SIR	$\mathrm{Var}(\varepsilon^2)$ ClipSIR
$\sigma_y^2 = 5 \times 10^{-3}$	18.69	7.36	3,029.30	79.40
$\sigma_y^2 = 1$	282.34	273.71	5,698.03	2,8264.23
$\sigma_y^2 = 20$	2,723.90	2,720.10	7.33×10^6	5.63×10^6

V. CONCLUSIONS

We have investigated a new particle filtering scheme that incorporates a nonlinear transformation of the importance weights in order to reduce the weight variance and, hence, mitigate the well-known weight degeneracy phenomenon. Modifying the importance weights introduces a bias in the estimators computed from the particles. However, we have also analyzed the asymptotics of the resulting algorithm and proved that, under mild assumptions, the estimates of integrals of bounded functions with respect to the target posterior distributions converge a.s. to the true integrals as the number of particles increases (i.e., the filter is *asymptotically* unbiased). Finally, we have conducted computer simulations to show the improvement provided by the new methodology, specially in a scenario which is prone to severe weight degeneracy.

ACKNOWLEDGMENT

E. K. acknowledges the support of *Ministerio de Educación* of Spain (*Programa de Formación de Profesorado Universitario*, ref. AP2008-00469). This work has been partially supported by *Ministerio de Economía y Competitividad* of Spain (program Consolider-Ingenio 2010 CSD2008-00010 COMONSENS and project COMPREHENSION TEC2012-38883-C02-01).

REFERENCES

[1] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. New York (USA): Springer, 2001.

[2] P. M. Djurić, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Míguez, "Particle filtering," *IEEE Signal Processing Magazine*, vol. 20, no. 5, pp. 19–38, September 2003.

[3] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Boston: Artech House, 2004.

[4] O. Cappé, S. J. Godsill, and E. Moulines, "An overview of existing methods and recent advances in sequential Monte Carlo," *Proceedings of the IEEE*, vol. 95, no. 5, pp. 899–924, 2007.

[5] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo Sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.

[6] T. Bengtsson, P. Bickel, and B. Li, "Curse of dimensionaluty revisited: Collapse of particle filter in very large scale systems," *Probability and statistics: Essay in honour of David A. Freedman*, vol. 2, pp. 316–334, 2008.

[7] E. Koblents and J. Míguez, "A population monte carlo scheme with transformed weights and its application to stochastic kinetic models," *Arxiv preprint*, vol. arXiv:1208.5600v1 [stat.CO], August 2012.

[8] J. S. Liu and R. Chen, "Sequential Monte Carlo methods for dynamic systems," *Journal of the American Statistical Association*, vol. 93, no. 443, pp. 1032–1044, September 1998.

[9] A. Doucet, N. de Freitas, and N. Gordon, "An introduction to sequential Monte Carlo methods," in *Sequential Monte Carlo Methods in Practice*, A. Doucet, N. de Freitas, and N. Gordon, Eds. Springer, 2001, ch. 1, pp. 4–14.

[10] A. Bain and D. Crisan, *Fundamentals of Stochastic Filtering*. Springer, 2008.