Semi-independent resampling for particle filtering

Roland Lamberti, Yohan Petetin, François Desbouvries, *Senior Member, IEEE,* and François Septier

***Abstract*—Among Sequential Monte Carlo (SMC) methods, Sampling Importance Resampling (SIR) algorithms are based on Importance Sampling (IS) and on some (resampling-based) rejuvenation algorithm which aims at fighting against weight degeneracy. However this mechanism tends to be insufficient when applied to informative or high-dimensional models. In this paper we revisit the rejuvenation mechanism and propose a class of parameterized SIR-based solutions which enable to adjust the tradeoff between computational cost and statistical performances.**

I. Introduction and background

Bayesian filtering consists in estimating some variable x_t from noisy measurements $y_{0:t} = \{y_0, \cdots, y_t\}$. We assume that $\{(x_t, y_t)\}_{t\geq 0}$ is a Hidden Markov Chain, i.e. that the joint density of $(x_{0:t}, y_{0:t})$ reads $p(x_{0:t}, y_{0:t}) = p(x_0)\prod_{s=1}^{t} f_s(x_s|x_{s-1})\prod_{s=0}^{t} g_s(y_s|x_s)$. The problem can be traced back to Kalman [1] in the context of linear and Gaussian state space models. Approximate solutions for non linear and/or non Gaussian state space models include the extended Kalman filter [2]–[4], the unscented Kalman filter [5]–[8], or SMC methods (also called particle filters (PF)) [7], [9], [10], which propagate in time a discrete approximation $\widehat{p}(x_t|y_{0:t}) = \sum_{i=1}^{N} w_t^i \delta_{x_t^i}$ of the posterior pdf $p(x_t|y_{0:t})$.

A. The classical SIR algorithm

Let $\Theta_t = \int \varphi(x_t)p(x_t|y_{0:t})\mathrm{d}x_t$ be a moment of interest of $p(x_t|y_{0:t})$. One iteration of an SMC algorithm can be decomposed in three steps. Starting at time $t-1$ from $\{w_{t-1}^i, x_{t-1}^i\}_{i=1}^N$, the first two steps consist of sampling (*S.*) N particles $\tilde{x}_t^i$ from importance densities q_i and weighting (*W.*) them so as to take into account the discrepancy between the target and importance densities; then Θ_t is estimated as $\widehat{\Theta}_t^{\mathrm{SIS},N} = \sum_{i=1}^{N} \tilde{w}_t^i \varphi(\tilde{x}_t^i)$ (superscript SIS will be justified below). Finally a third (optional) step consists of re-sampling (*R.*) the weighted particles, i.e. in re-drawing each particle with a probability equal to its weight and assigning to the resampled particles the same weight $\frac{1}{N}$. This yields the class of SIR algorithms [7], [9]–[11] described by Algorithm 1.

Let us comment this algorithm. If resampling is totally absent, each time iteration reduces to the first two steps, i.e. is based on IS only. However such a sequential IS (SIS) algorithm is well known to fail in practice since after a few iterations most weights get close to zero. The third step (which can be performed at each iteration t or depending on some criterion such as the number of effective particles [12]–[15]) discards particles with low weights (such particles are likely never to be resampled) and is considered as a traditional rescue against weight degeneracy. On the other hand, this (*R.*) step introduces local extra variance [13, section 4.2.1], which in turn affects the variance of $\widehat{\Theta}_t^{\mathrm{SIS},N}$ at subsequent iterations. It has thus been proposed to control this extra variance term via alternative resampling schemes (e.g. [16]–[18]). Yet despite many proposed refinements this generic SIR mechanism remains inefficient in informative models featuring very sharp likelihood functions (i.e., when $g_t(y_t|x_t)$ is very small for most values of x_t), and in particular in high-dimensional state-space models [19], [20].

Algorithm 1 The classical SIR algorithm

Data: $q(x_t|x_{t-1})$, y_t, $\{w_{t-1}^i, x_{t-1}^i\}_{i=1}^N$;
for $1 \leq i \leq N$ **do**
  S. $\tilde{x}_t^i \sim q(x_t|x_{t-1}^i)$;
  W. $\tilde{w}_t^i \propto w_{t-1}^i \frac{f_t(\tilde{x}_t^i|x_{t-1}^i)g_t(y_t|\tilde{x}_t^i)}{q(\tilde{x}_t^i|x_{t-1}^i)}$, $\sum_{i=1}^N \tilde{w}_t^i = 1$;
end for
$\widehat{\Theta}_t^{\mathrm{SIS},N} = \sum_{i=1}^N \tilde{w}_t^i \varphi(\tilde{x}_t^i)$;
if R. then
  for $1 \leq i \leq N$ **do**
    $l^i \sim \Pr(L = l) = \tilde{w}_t^l$, $1 \leq l \leq N$;
  end for
  Set $\{w_t^i, x_t^i\}_{i=1}^N = \{\frac{1}{N}, \tilde{x}_t^{l_i}\}_{i=1}^N$.
else
  Set $\{w_t^i, x_t^i\}_{i=1}^N = \{\tilde{w}_t^i, \tilde{x}_t^i\}_{i=1}^N$.
end if

B. The independent SIR algorithm

Recently it has thus been proposed to revisit the SIR algorithm [21]–[23] and more precisely to come back to the rejuvenation mechanism (*R.*). The counterpart of this (*R.*) step is that it duplicates particles with high weights, which results in support degeneracy. Moreover given $\{w_{t-1}^i, x_{t-1}^i\}_{i=1}^N$ the samples $\{x_t^j\}$ produced by Algorithm 1 are marginally distributed from some compound pdf $\tilde{q}_t^N$ which takes into account the effects of the three elementary (*S.*), (*W.*) and (*R.*) steps, but are obviously dependent [21] (a single particle can be resampled more than once); by contrast, given $\{w_{t-1}^i, x_{t-1}^i\}_{i=1}^N$ the independent SIR Algorithm [21], [23] produces N i.i.d. draws from $\tilde{q}_t^N$ by resampling from N sets of N particles. Thus, it ensures the diversity of the final support compared to a classical approach where the final particles are resampled from the same set of (possibly N^2) particles.

Note that Algorithm 2 below only decribes the rejuvenation step of the independent SIR algorithm, and replaces the "**if R. then**" part of Algorithm 1.

R. Lamberti, Y. Petetin and F. Desbouvries are with Samovar, Telecom Sudparis, CNRS, Université Paris-Saclay, 9 rue Charles Fourier, 91011 Evry, France. F. Septier is with IMT Lille Douai, Univ. Lille, CNRS, UMR 9189 - CRIStAL, F-59000 Lille, France.

Algorithm 2 Indep. SIR algorithm (*resampling step only*)

Data: $q(x_t|x_{t-1})$, y_t, $\{w_{t-1}^i, x_{t-1}^i\}_{i=1}^N$, $\{\tilde{w}_t^i, \tilde{x}_t^i\}_{i=1}^N$;
for $1 \leq j \leq N$ **do**
  $\tilde{x}_t^{1,j} \leftarrow \tilde{x}_t^j$, $\tilde{w}_t^{1,j} \leftarrow \tilde{w}_t^j$.
end for
for $1 \leq i \leq N$ **do**
  R. $l^i \sim \Pr(L = l) = \tilde{w}_t^{i,l}$, $1 \leq l \leq N$;
  Rejuvenation of the support for iteration $i+1$
  if $(i < N)$ **then**
    for $1 \leq j \leq N$ **do**
      $\tilde{x}_t^{i+1,j} \sim q(x_t|x_{t-1}^j)$;
      $\overline{w}_t^{i+1,j} = w_{t-1}^j \frac{f_t(\tilde{x}_t^{i+1,j}|x_{t-1}^j) g_t(y_t|\tilde{x}_t^{i+1,j})}{q(\tilde{x}_t^{i+1,j}|x_{t-1}^j)}$,
    end for
    $\tilde{w}_t^{i+1,:} \propto \overline{w}_t^{i+1,:}$, $\sum_{j=1}^N \tilde{w}_t^{i+1,j} = 1$;
  end if
end for
Set $\{w_t^i, x_t^i\}_{i=1}^N = \{\frac{1}{N}, \tilde{x}_t^{i,l^i}\}_{i=1}^N$.

C. Scope of the paper

Algorithm 2 has displayed good results in severe situations [21] and can be combined with a post-resampling, second-stage reweigthing scheme due to its auxiliary particle filtering interpretation [22], [23]. However its rejuvenation mechanism involves the sampling of N^2 particles $\{\tilde{x}_t^{i,j}\}_{i,j=1}^N$ (plus N resampling steps), by contrast with the classical SIR algorithm which only samples N intermediate particles $\{\tilde{x}_t^j\}_{j=1}^N$ (and is also followed by N resampling steps). One can wonder whether this extra cost is indeed necessary, so the aim of this paper is to design an algorithm which is both efficient (in terms of computational cost) and effective (in terms of statistical results). The rest of this paper is organized as follows. Our algorithm is described in section II. Simulations are displayed in section III, and the paper ends with a conclusion.

II. Semi-independent resampling

A. An intermediate resampling scheme

The classical and independent SIR resampling mechanisms can be reconciled in a common framework. In both schemes, one progressively builds N weighted sets $\tilde{x}_t^{1,:}$, $\cdots$, $\tilde{x}_t^{N,:}$ (the N supports) and redraws one sample x_t^i out of each of them (see figure 1). The difference lies in the way $\tilde{x}_t^{i,:}$ is built from $\tilde{x}_t^{i-1,:}$: in the classical SIR mechanism, $\tilde{x}_t^{i,:}$ is a copy of $\tilde{x}_t^{i-1,:}$ (so the resampling step amounts to redrawing N samples from the common support $\tilde{x}_t^{1,:}$, see Algorithm 1); in the independent SIR mechanism, a whole new support $\tilde{x}_t^{i,:}$ is drawn at each iteration i. In other words, from a computational point of view both schemes resample N particles from some intermediate set $\{\tilde{x}_t^{i,j}\}_{i,j=1}^N$, but building that set requires N preliminary independent sampling steps in the classical case, while it requires N^2 independent sampling steps in the independent case.

In this paper we propose a resampling scheme which creates an intermediate set $\{\tilde{x}_t^{i,j}\}_{i,j=1}^N$ with more diversity than in the classical case, but at a reduced sampling cost as compared to the independent case. Starting from $\tilde{x}_t^{i-1,j}$, $\tilde{x}_t^{i,j}$ can now either be a copy (to save cost) or a new sample (to enhance diversity). The algorithm is as follows. Fix the number k (with $0 \leq k \leq N$) of samples which will be redrawn at each iteration. At step i, uniformly draw a subset $m^{i,1:k} = (m^{i,1}, \cdots, m^{i,k})$ of size k out of $(1, \cdots, N)$ ($m^{i,l}$ are the indices of the particles which will be redrawn). Next $\tilde{x}_t^{i,j} \sim q(x_t|x_{t-1}^j)$ if $j \in m^{i,1:k}$, and $\tilde{x}_t^{i,j} = \tilde{x}_t^{i-1,j}$ if $j \notin m^{i,1:k}$. Finally observe that the classical (resp. independent) SIR algorithm corresponds to the particular case $k = 0$ (resp. $k = N$). The resampling step associated with this scheme is summarized in Algorithm 3.

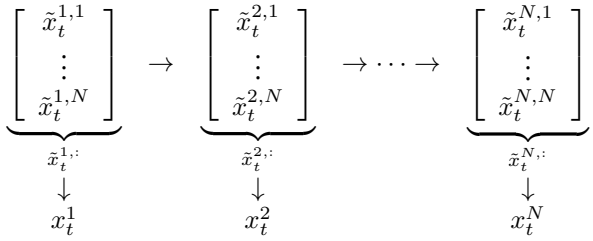

Fig. 1: The classical, independent and semi-independent resampling mechanisms. Each scheme draws N supports $\tilde{x}_t^{i,:}$ and redraws one sample x_t^i out of each support. The difference lies in the way $\tilde{x}_t^{i,:}$ is built from $\tilde{x}_t^{i-1,:}$: $\tilde{x}_t^{i,j}$ is a copy of $\tilde{x}_t^{i-1,j}$ in the classical case; is a new particle in the independent case; and can be either copied of redrawn in the intermediate, semi-independent case.

Algorithm 3 Semi-ind. SIR algorithm (*resampling step only*)

Data: $q(x_t|x_{t-1})$, y_t, $\{w_{t-1}^i, x_{t-1}^i\}_{i=1}^N$, $\{\tilde{w}_t^i, \tilde{x}_t^i\}_{i=1}^N$;
for $1 \leq j \leq N$ **do**
  $\tilde{x}_t^{1,j} \leftarrow \tilde{x}_t^j$, $\tilde{w}_t^{1,j} \leftarrow \tilde{w}_t^j$.
end for
for $1 \leq i \leq N$ **do**
  R. $l^i \sim \Pr(L=l) = \tilde{w}_t^{i,l}$, $1 \leq l \leq N$;
  Partial rejuvenation of the support for iteration $i+1$
  if $(i < N)$ **then**
    $\tilde{x}_t^{i+1,:} \leftarrow \tilde{x}_t^{i,:}$, $\overline{w}_t^{i+1,:} \leftarrow \overline{w}_t^{i,:}$;
    for $1 \leq j \leq k$ **do**
      $m^j \sim \Pr(M = n | n \in 1{:}N \backslash \{m^{1:j-1}\}) = \frac{1}{N-j+1}$;
      $\tilde{x}_t^{i+1,m^j} \sim q(x_t|x_{t-1}^{m^j})$;
      $\overline{w}_t^{i+1,m^j} = w_{t-1}^{m^j} \frac{f_t(\tilde{x}_t^{i+1,m^j}|x_{t-1}^{m^j}) g_t(y_t|\tilde{x}_t^{i+1,m^j})}{q(\tilde{x}_t^{i+1,m^j}|x_{t-1}^{m^j})}$;
    end for
    $\tilde{w}_t^{i+1,:} \propto \overline{w}_t^{i+1,:}$, $\sum_{j=1}^N \tilde{w}_t^{i+1,j} = 1$;
  end if
end for
Set $\{w_t^i, x_t^i\}_{i=1}^N = \{\frac{1}{N}, \tilde{x}_t^{i,l^i}\}_{i=1}^N$.

B. Performances vs. computational cost

We now evaluate the performance of this procedure by comparing the variances of the estimates computed after the resampling step because they affect the variances of the estimates at subsequent iterations [17]. So let $\widehat{\Theta}_t^{:,N} = \frac{1}{N}\sum_{i=1}^N \varphi(x_t^{:,i})$, where the generic notation $x_t^{:,i}$ represents the points produced either by Algorithm 1, 2 or 3 (so we consider $\widehat{\Theta}_t^{\mathrm{SIR},N}$, $\widehat{\Theta}_t^{\mathrm{I-SIR},N}$ and $\widehat{\Theta}_t^{\mathrm{SR},N,k}$, where SR stands for semi-

resampling). We have the following proposition (the proof is given in the Appendix).

Proposition 1: Given the previous set of particles $\{x_{0:t-1}^i\}_{i=1}^N$, for all k, $0 \leq k \leq N$, we have:

$$\mathrm{E}(\widehat{\Theta}_t^{\mathrm{SR},N,k}) = \mathrm{E}(\widehat{\Theta}_t^{\mathrm{I-SIR},N}) = \mathrm{E}(\widehat{\Theta}_t^{\mathrm{SIR},N}), \quad (1a)$$

$$\mathrm{var}(\widehat{\Theta}_t^{\mathrm{I-SIR},N}) \leq \mathrm{var}(\widehat{\Theta}_t^{\mathrm{SR},N,k}) \leq \mathrm{var}(\widehat{\Theta}_t^{\mathrm{SIR},N}), \quad (1b)$$

$$\mathrm{var}(\widehat{\Theta}_t^{\mathrm{SR},N,k}) \leq \mathrm{var}(\widehat{\Theta}_t^{\mathrm{SR},N,k-1}). \quad (1c)$$

So as the number k of intermediate redrawings increases from 0 to N, the conditional variance of the semi-independent resampling estimator $\widehat{\Theta}_t^{\mathrm{SR},N,k}$ decreases from the upper bound of inequality (1b) (if $k = 0$, $\widehat{\Theta}_t^{\mathrm{SR},N,0}$ reduces to $\widehat{\Theta}_t^{\mathrm{SIR},N}$) to its lower bound (if $k = N$, $\widehat{\Theta}_t^{\mathrm{SR},N,N}$ reduces to $\widehat{\Theta}_t^{\mathrm{I-SIR},N}$). However remember from section II-A that $N + (N-1) \times k$ samples are needed for building $\widehat{\Theta}_t^{\mathrm{SR},N,k}$; so parameter k of the SR scheme enables to fix a compromise between variance reduction and computational budget.

C. A parallelized version

Furthermore, Algorithm 3 can be transformed into a parallelized version, the non-sequential SR (NSSR) algorithm. At iteration i, instead of duplicating the $N-k$ surviving particles from the previous support $\tilde{x}_t^{i-1,:}$ (see Fig. 1), we propose to duplicate the $N-k$ surviving particles directly from the initial set $\tilde{x}_t^{1,:}$ of particles. The $N-1$ new supports can thus be produced in parallel, contrary to Algorithm 3 which by nature is sequential even if the intermediate sampling steps can be parallelized. Of course, this procedure alters the diversity of the final set of particles, as is illustrated by the following proposition.

Proposition 2: Let $\widehat{\Theta}_t^{NSSR,k}$ be the estimate built from the non-sequential semi-independent resampling procedure. Then given the previous set of particles $\{x_{0:t-1}^i\}_{i=1}^N$, for all k, $0 \leq k \leq N$, we have:

$$\mathrm{E}(\widehat{\Theta}_t^{\mathrm{NSSR},N,k}) = \mathrm{E}(\widehat{\Theta}_t^{\mathrm{I-SIR},N}) = \mathrm{E}(\widehat{\Theta}_t^{\mathrm{SIR},N}), \quad (2a)$$

$$\mathrm{var}(\widehat{\Theta}_t^{\mathrm{I-SIR},N}) \leq \mathrm{var}(\widehat{\Theta}_t^{\mathrm{NSSR},N,k}) \leq \mathrm{var}(\widehat{\Theta}_t^{\mathrm{SIR},N}), \quad (2b)$$

$$\mathrm{var}(\widehat{\Theta}_t^{\mathrm{NSSR},N,k}) \leq \mathrm{var}(\widehat{\Theta}_t^{\mathrm{NSSR},N,k-1}), \quad (2c)$$

$$\mathrm{var}(\widehat{\Theta}_t^{\mathrm{SR},N,k}) \leq \mathrm{var}(\widehat{\Theta}_t^{\mathrm{NSSR},N,k}). \quad (2d)$$

So we see that $\mathrm{var}(\widehat{\Theta}_t^{\mathrm{NSSR},N,k})$ still decreases with k, but is always larger than $\mathrm{var}(\widehat{\Theta}_t^{\mathrm{SR},N,k})$. As with Proposition 1, the variance inequalities still rely on Jensen's inequality, and the proof is omitted.

III. Simulations

We consider a tracking problem based on range-bearing measurements. The hidden state-vector contains the position and velocity of the target in cartesian coordinates, $x_t = [c_{x,t}, \dot{c}_{x,t}, c_{y,t}, \dot{c}_{y,t}]^T$. We set $f_t(x_t|x_{t-1}) = \mathcal{N}(x_t; \mathbf{F}x_{t-1}; \mathbf{Q})$, $g_t(y_t|x_t) = \mathcal{N}(y_t; \left[\sqrt{c_{x,t}^2 + c_{y,t}^2}; \arctan \frac{c_{y,t}}{c_{x,t}}\right]^T; \mathbf{R})$, with $\mathbf{R} = \mathrm{diag}(\sigma_\rho^2, \sigma_\theta^2)$, $\mathbf{F} = \mathbf{I}_2 \otimes \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, $\mathbf{Q} = 10 \times \mathbf{I}_2 \otimes \begin{bmatrix} \frac{1}{3} & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$ where $\otimes$ is the Kronecker product. We set $q(x_t|x_{t-1}) = f_t(x_t|x_{t-1})$ and we compare the RMSEs averaged over 1000 MC runs.

A. Variance of SR procedures

We first analyze the behaviour of our algorithms as a function of k. We set $N = 100$, $\sigma_\rho = 0.1$ and $\sigma_\theta = \frac{\pi}{1800}$; all MC runs use the same measurements. Fig. 2 displays the RMSE of $\widehat{\Theta}_t^{\mathrm{SR},N,k}$, $\widehat{\Theta}_t^{\mathrm{NSSR},N,k}$, $\widehat{\Theta}_t^{\mathrm{I-SIR},N}$ deduced from our resampling schemes and $\widehat{\Theta}_t^{\mathrm{SIS},N}$. Of course, the performances of estimates based on the SR procedure improve when k increases. It is also interesting to note that $\widehat{\Theta}_t^{\mathrm{SR},N,k}$ (resp. $\widehat{\Theta}_t^{\mathrm{NSSR},N}$) has the same performance as $\widehat{\Theta}_t^{\mathrm{I-SIR},N}$ when $k \geq N/2$ (resp. $k \geq 4N/5$) (these results have also been observed in other models). With Matlab, the averaged computational times of a single iteration of the SIR, SR with $k = N/2$ and I-SIR algorithms are 0.0244s, 0.0267s and 0.0283s, respectively.

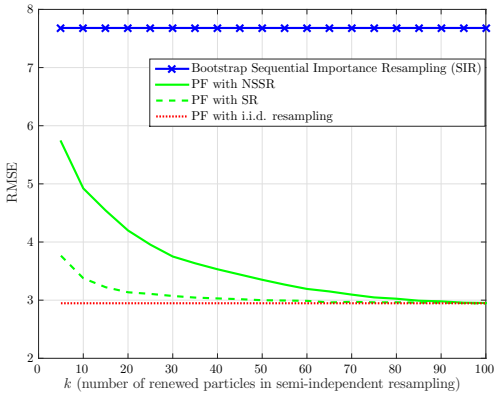

Fig. 2: RMSE as a function of k, tracking model.

B. RMSE in the informative case at equal cost

We now compare our estimates with existing improvements of the PF in informative models. In particular, the PF with MCMC resample move is a popular solution to introduce sample variety after resampling [24]. Roughly speaking, the N particles which follow the (*R.*) step of Algorithm 1 are moved via an MCMC algorithm with k iterations (here an independent Metropolis-Hasting algorithm). We also compare our estimates with those based on the classical SIR and I-SIR algorithms but with a given budget of total sampling (sampling + resampling) operations. We thus set $N = 100$ particles and $k = N/2$ for the computation of $\widehat{\Theta}_t^{\mathrm{SR},N,k}$ and the estimate based on the resample move PF, $N^{\mathrm{I-SIR}} = 72$ for that of $\widehat{\Theta}_t^{\mathrm{I-SIR},N^{\mathrm{I-SIR}}}$ and $N^{\mathrm{SIR}} = N + \frac{(N-1)k}{2} = 2575$ particles for that of $\widehat{\Theta}_t^{\mathrm{SIS},N^{\mathrm{SIR}}}$. The global sampling cost for all these algorithms is $(2N + N(k-1))$. We also compute $\widehat{\Theta}_t^{\mathrm{NSSR},N,k}$ with $k = 4N/5$; its computation does not have the same computational cost but can be totally parallelized. The results are displayed in Fig. 3.

When the observations are very informative, the classical solution tends to degenerate (we have obtained the same results with alternative resampling schemes such as the residual or stratified resampling), while our solutions are robust and particularly present better performances than the resample move algorithm which also uses k extra samples. As the variance of the measurement noise increases, the different estimates tend to behave similarly; the classical SIR algorithm

performs slightly better because it no longer suffers from the degeneracy phenomenon and the number of final samples used is far superior to the other solutions. Additional simulations have also shown the superiority of our SR scheme (RMSE of 2.671) compared to a classical SIR with $N + k(N-1)$ intermediate particles and N final particles (RMSE of 3.375) when $\sigma_\rho = 0.03$ and $\sigma_\theta = \frac{\pi}{6000}$. Finally, we have observed that the I-SIR and SR procedures improve the diversity of complete trajectories compared to the SIR algorithm and could potentially be used for building relevant smoothing estimators.

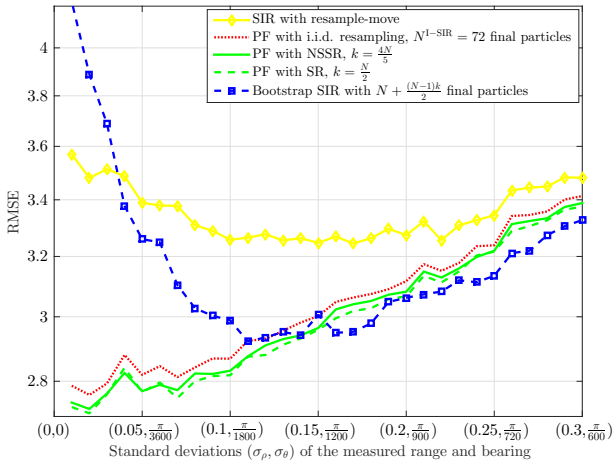

Fig. 3: Tracking model, $\sigma_\rho \in [0.01, 0.3]$ and $\sigma_\theta \in [\frac{\pi}{18000}, \frac{\pi}{600}]$.

IV. Conclusion

In this paper we revisited the resampling step of PF algorithms, and proposed a resampling scheme where each new final particle is resampled from a support which is partially rejuvenated with k new particles. This yields a class of parameterized solutions which encompasses the classical multinomial resampling technique ($k = 0$) and the independent resampling one ($k = N$), enabling to tune the balance between variance and computational cost. Simulations showed that choosing $k = N/2$ leads to similar performances to the fully independent resampling procedure. Moreover, in very informative models our algorithm is not affected by the degeneration phenomenon, contrary to the classical SIR algorithm.

Appendix
Proof of Proposition 1

Let us consider a PF with resampling at time t. (1a) holds because the SIR, I-SIR and SR procedures all produce resampled particles which, given $\{x^i_{0:t-1}\}_{i=1}^N$, are sampled from the same distribution $\tilde{q}_t^N$; and (1b) is straightforward from (1c) and the fact that SR reduces to SIR (resp. I-SIR) when $k = 0$ (resp. $k = N$). Let us address (1c). Since $\widehat{\Theta}_t^{\text{SR},N,k} = \frac{1}{N}\sum_{i=1}^N \varphi(x_t^{\text{SR},i})$, given $\{x^i_{0:t-1}\}_{i=1}^N$

$$N^2 \text{var}_k(\widehat{\Theta}_t^{\text{SR},N,k}) = \sum_{i=1}^N \text{var}(\varphi(x_t^i)) + 2\sum_{\substack{i_1,i_2=1 \\ i_1<i_2}}^N \text{cov}_k(\varphi(x_t^{i_1}),\varphi(x_t^{i_2}));$$

here index k in a (co)variance emphasizes the fact that it depends on k. The first term of the r.h.s. is independent of k so the difference between different values of k stems from the covariance terms. Next $\text{cov}_k(\varphi(x_t^{i_1}),\varphi(x_t^{i_2})) = \text{E}_k[\varphi(x_t^{i_1})\varphi(x_t^{i_2})] - \text{E}[\varphi(x_t^{i_1})]\text{E}[\varphi(x_t^{i_2})]$, and again, the second term of the r.h.s. is independent of k. Finally for $i_1 < i_2$,

$$\begin{aligned}
\text{E}_k[\varphi(x_t^{i_1})\varphi(x_t^{i_2})] &= \text{E}[\text{E}[\varphi(x_t^{i_1})\varphi(x_t^{i_2})|\tilde{x}_t^{i_1:i_2,:}]] \\
&= \text{E}[\text{E}[\varphi(x_t^{i_1})|\tilde{x}_t^{i_1,:}]\text{E}[\varphi(x_t^{i_2})|\tilde{x}_t^{i_2,:}]] \\
&= \text{E}[\widehat{\Theta}_t^{\text{SIS}}(\tilde{x}_t^{i_1,:})\widehat{\Theta}_t^{\text{SIS}}(\tilde{x}_t^{i_2,:})] \\
&= \text{E}[\text{E}[\widehat{\Theta}_t^{\text{SIS}}(\tilde{x}_t^{i_1,:})\widehat{\Theta}_t^{\text{SIS}}(\tilde{x}_t^{i_2,:})|m_{i_1+1}^{i_2}(1{:}k)]]
\end{aligned}$$

where $m_{i_1+1}^{i_2}(1{:}k)$ represents all the indices redrawn from iterations i_1+1 to i_2 (the third equality holds because x_t^i is resampled from support $\tilde{x}_t^{i,:}$ (see Fig. 1), so $\text{E}(\varphi(x_t^i)|\tilde{x}_t^{i,:}) = \widehat{\Theta}_t^{\text{SIS},N}(\tilde{x}_t^{i,:})$ where $\widehat{\Theta}_t^{\text{SIS},N}$ was defined in section I-A).

The outer expectation in this last expression corresponds to a uniformly weighted sum over all possible values of $m_{i_1+1}^{i_2}(1{:}k)$, i.e. over $(A_N^k)^{i_2-i_1}$ terms where A_N^k is the number of arrangements of k among N. Given $m_{i_1+1}^{i_2}(1{:}k)$, the general term of this sum reads

$$\begin{aligned}
&\text{E}[\widehat{\Theta}_t^{\text{SIS}}(\tilde{x}_t^{i_1,:})\widehat{\Theta}_t^{\text{SIS}}(\tilde{x}_t^{i_2,:})|m_{i_1+1}^{i_2}(1{:}k)] = \\
&\text{E}[\text{E}[\widehat{\Theta}_t^{\text{SIS}}(\tilde{x}_t^{i_1,:})\widehat{\Theta}_t^{\text{SIS}}(\tilde{x}_t^{i_2,:})|\tilde{x}_t^{i_1,1:N\setminus m_{i_1+1}^{i_2}(1:k)}]|m_{i_1+1}^{i_2}(1{:}k)]
\end{aligned}$$

where $\tilde{x}^{i_1,1:N\setminus m_{i_1+1}^{i_2}(1:k)}$ are the particles shared by supports $\tilde{x}^{i_1,:}$ and $\tilde{x}^{i_2,:}$. Under this conditioning, $\widehat{\Theta}_t^{\text{SIS}}(\tilde{x}_t^{i_1,:})$ and $\widehat{\Theta}_t^{\text{SIS}}(\tilde{x}_t^{i_2,:})$ are independent so the general term is

$$\begin{aligned}
&\text{E}[\text{E}[\widehat{\Theta}_t^{\text{SIS}}(\tilde{x}_t^{i_1,:})|\tilde{x}_t^{i_1,1:N\setminus m_{i_1+1}^{i_2}(1:k)}] \\
&\quad\times \text{E}[\widehat{\Theta}_t^{\text{SIS}}(\tilde{x}_t^{i_2,:})|\tilde{x}_t^{i_1,1:N\setminus m_{i_1+1}^{i_2}(1:k)}]|m_{i_1+1}^{i_2}(1{:}k)] \\
&\quad= \text{E}[\text{E}^2[\widehat{\Theta}_t^{\text{SIS}}(\tilde{x}_t^{i_1,:})|\tilde{x}^{i_1,1:N\setminus m_{i_1+1}^{i_2}(1:k)}]|m_{i_1+1}^{i_2}(1{:}k)] \\
&\quad= h(m_{i_1+1}^{i_2}(1{:}k)) = h(m_{i_1+1}^{i_2}(1{:}k-1), m_{i_1+1}^{i_2}(k))
\end{aligned}$$

because given the trajectories from the previous time steps, particles from different supports are all marginally drawn from the same densities. Finally

$$\text{E}_k(\varphi(x_t^{i_1})\varphi(x_t^{i_2})) = \frac{1}{(A_N^k)^{i_2-i_1}} \sum_{m_{i_1+1}^{i_2}(1:k)} h(m_{i_1+1}^{i_2}(1{:}k)). \quad (3)$$

It remains to compare (3) with the same expression with $k \leftarrow k-1$. We observe that (3) can be rewritten as

$$\begin{aligned}
\text{E}_k(\varphi(x_t^{i_1})\varphi(x_t^{i_2})) = \frac{1}{(A_N^{k-1})^{i_2-i_1}} &\sum_{m_{i_1+1}^{i_2}(1:k-1)} \\
\frac{1}{N-k+1} &\sum_{m_{i_1+1}^{i_2}(k)} h(m_{i_1+1}^{i_2}(1{:}k-1), m_{i_1+1}^{i_2}(k)),
\end{aligned}$$

where the second line coincides with the conditionnal expectation $\text{E}[h(m_{i_1+1}^{i_2}(1{:}k))|m_{i_1+1}^{i_2}(1{:}k-1)]$. Given $m_{i_1+1}^{i_2}(1{:}k-1)$, the set $\tilde{x}^{i_1,1:N\setminus m_{i_1+1}^{i_2}(1:k)}$ is included in $\tilde{x}^{i_1,1:N\setminus m_{i_1+1}^{i_2}(1:k-1)}$; consequently, the Rao-Blackwell decomposition ($\text{E}(\text{E}^2(X|Y)) \le \text{E}(\text{E}^2(X|Y,Z))$) ensures that

$$h(m_{i_1+1}^{i_2}(1{:}k-1), m_{i_1+1}^{i_2}(k)) \le h(m_{i_1+1}^{i_2}(1{:}k-1))$$

for all $m_{i_1+1}^{i_2}(k)$, and so that $\text{E}[h(m_{i_1+1}^{i_2}(1{:}k))|m_{i_1+1}^{i_2}(1{:}k-1)] \le h(m_{i_1+1}^{i_2}(1{:}k-1))$, whence (1c).

REFERENCES

[1] R.E.Kalman, "A new approach to linear filtering and prediction problems," *J.Basic Eng., Trans.ASME, Series D*, vol. 83, pp. 35–45, 1960.

[2] A. H. Jazwinski, *Stochastic Processes and Filtering Theory*, vol. 64 of *Mathematics in Science and Engineering*. San Diego: Academic Press, 1970.

[3] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*. Prentice Hall, Englewood Cliffs, New Jersey, 1979.

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

[5] S. Julier and J. Uhlmann, "Unscented filtering and nonlinear estimation," in *Proceedings of the IEEE*, vol. 92, pp. 401–422, March 2004.

[6] H. Tanizaki, *Nonlinear Filters, Estimation and Applications*. Berlin: Springer, 2nd ed., 1996.

[7] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear / non-Gaussian Bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, pp. 174–188, February 2002.

[8] H. M. T. Menegaz, J. Y. Ishihara, G. A. Borges, and A. N. Vargas, "A systematization of the unscented Kalman filter theory," *IEEE Transactions on Automatic Control*, vol. 60, pp. 2583–2598, October 2015.

[9] N. J. Gordon, D. J. Salmond, and A. F. M. Smith, "Novel approach to nonlinear/ non-Gaussian Bayesian state estimation," *IEE Proceedings-F*, vol. 140, pp. 107–113, April 1993.

[10] A. Doucet, N. de Freitas, and N. Gordon, eds., *Sequential Monte Carlo Methods in Practice*. Statistics for Engineering and Information Science, New York: Springer Verlag, 2001.

[11] A. F. M. Smith and A. E. Gelfand, "Bayesian statistics without tears : a sampling-resampling perspective.," *The American Statistician*, vol. 46, no. 2, pp. 84–87, 1992.

[12] A. Kong, J. S. Liu, and W. H. Wong, "Sequential imputations and bayesian missing data problems," *Journal of the American Statistical Association*, vol. 89, pp. 278–88, March 1994.

[13] J. S. Liu and R. Chen, "Blind deconvolution via sequential imputation," *Journal of the American Statistical Association*, vol. 90, pp. 567–76, June 1995.

[14] J. S. Liu, "Metropolized independent sampling with comparisons to rejection sampling and importance sampling," *Statistics and Computing*, vol. 6, pp. 113–119, 1996.

[15] J. Cornebise, É. Moulines, and J. Olsson, "Adaptive methods for sequential importance sampling with application to state-space models," *Statistics and Computing*, vol. 18, no. 4, pp. 461–480, 2008.

[16] G. Kitagawa, "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models," *Journal of Computational and Graphical Statistics*, vol. 5, no. 1, pp. 1–25, 1996.

[17] R. Douc, O. Cappé, and É. Moulines, "Comparison of resampling schemes for particle filtering," in *Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis (ISPA)*, (Zagreb, Croatia), September 2005.

[18] T. Li, M. Bolić, and P. M. Djuric, "Resampling Methods for Particle Filtering: Classification, implementation, and strategies," *IEEE Signal Processing Magazine*, vol. 32, pp. 70–86, May 2015.

[19] C. Snyder, T. Bengtsson, P. Bickel, and J. Anderson, "Obstacles to high-dimensional particle filtering," *Monthly Weather Review*, vol. 136, pp. 4629–4640, 2008.

[20] C. Snyder, "Particle filters, the optimal proposal and high-dimensional systems," in *Seminar on Data assimilation for atmosphere and ocean, 6-9 September 2011*, (Shinfield Park, Reading), pp. 161–170, ECMWF, ECMWF, 2012.

[21] R. Lamberti, Y. Petetin, F. Septier, and F. Desbouvries, "Particle filters with independent resampling," in *41st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Shanghai, China), March 20-25 2016.

[22] R. Lamberti, Y. Petetin, F. Septier, and F. Desbouvries, "An improved SIR-based sequential Monte Carlo algorithm," in *19th IEEE workshop on statistical signal processing (SSP)*, (Palma de Majorca, Spain), June 26-29 2016.

[23] R. Lamberti, Y. Petetin, F. Desbouvries, and F. Septier, "Independent resampling sequential Monte Carlo algorithms," *accepted for publication, IEEE Transactions on Signal Processing*, June 2017.

[24] W. Gilks and C. Berzuini, "Following a moving target – Monte Carlo inference for dynamic Bayesian models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, pp. 127–146, 2001.