Independent Resampling Sequential Monte Carlo Algorithms

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

Abstract—Sequential Monte Carlo algorithms, or Particle Filters, are Bayesian filtering algorithms which propagate in time a discrete and random approximation of the a posteriori distribution of interest. Such algorithms are based on Importance Sampling with a bootstrap resampling step which aims at struggling against weight degeneracy. However, in some situations (informative measurements, high dimensional model), the resampling step can prove inefficient. In this paper, we revisit the fundamental resampling mechanism which leads us back to Rubin’s static resampling mechanism. We propose an alternative rejuvenation scheme in which the resampled particles share the same marginal distribution as in the classical setup, but are now independent. This set of independent particles provides a new alternative to compute a moment of the target distribution and enables to compute an estimate \( \hat{\Theta}_k \) of \( \Theta_k \):

\[
\hat{\Theta}_k = \sum_{i=1}^{N} w_i^k \varphi(x_i^k). \tag{3}
\]

More precisely, the computation of the set \( \{w_i^k, x_i^k\}_{i=1}^{N} \) is based on the sequential application of the Importance Sampling (IS) mechanism [4] [5]. This mechanism consists in sampling particles according to an importance distribution and weighting these samples in order to correct the discrepancy between the target and importance distributions. However the direct sequential application of the IS mechanism in model (1) fails in practice since after a few time steps most weights get close to zero, with only a few particles having non-negligible weights. Consequently IS alone becomes more and more ineffective since a lot of computational effort is devoted to sampling particles which will hardly contribute to the estimate \( \hat{\Theta}_k \) in (3).

As is well known, \( \Theta_k \) can be exactly computed only in very specific models, and one needs to resort to approximations in the general case. In this paper, we focus on a popular class of approximations called sequential Monte Carlo (SMC) algorithms or Particle Filters (PF), see e.g. [1]–[3]. PF propagate over time a set of \( N \) MC weighted samples \( \{w_i^k, x_i^k\}_{i=1}^{N} \) which defines a discrete approximation \( \sum_{i=1}^{N} w_i^k \delta_{x_i^k} \) of \( p(x_k|y_{0:k}) \) and enables to compute an estimate \( \hat{\Theta}_k \) of \( \Theta_k \):

\[
\hat{\Theta}_k = \int \varphi(x_k)p(x_k|y_{0:k})dx_k. \tag{2}
\]

As is well known, \( \Theta_k \) can be exactly computed only in very specific models, and one needs to resort to approximations in the general case. In this paper, we focus on a popular class of approximations called sequential Monte Carlo (SMC) algorithms or Particle Filters (PF), see e.g. [1]–[3]. PF propagate over time a set of \( N \) MC weighted samples \( \{w_i^k, x_i^k\}_{i=1}^{N} \) which defines a discrete approximation \( \sum_{i=1}^{N} w_i^k \delta_{x_i^k} \) of \( p(x_k|y_{0:k}) \) and enables to compute an estimate \( \hat{\Theta}_k \) of \( \Theta_k \):

\[
\hat{\Theta}_k = \int \varphi(x_k)p(x_k|y_{0:k})dx_k. \tag{2}
\]

More precisely, the computation of the set \( \{w_i^k, x_i^k\}_{i=1}^{N} \) is based on the sequential application of the Importance Sampling (IS) mechanism [4] [5]. This mechanism consists in sampling particles according to an importance distribution and weighting these samples in order to correct the discrepancy between the target and importance distributions. However the direct sequential application of the IS mechanism in model (1) fails in practice since after a few time steps most weights get close to zero, with only a few particles having non-negligible weights. Consequently IS alone becomes more and more ineffective since a lot of computational effort is devoted to sampling particles which will hardly contribute to the estimate \( \hat{\Theta}_k \) in (3).

As is well known, a traditional rescue against weight degeneracy consists in resampling the particles (– either at each time step or depending on some criterion such as the number of effective particles [6] [7] [8] [9]), i.e. of re-drawing each particle with a probability equal to its weight and next assigning to the resampled particles the same weight. This yields the class of Sampling Importance Resampling (SIR) algorithms [10] [1] [11] [12]. Resampling has proved to be beneficial in the long run since it recreates diversity for subsequent time steps, but also results in dependency among the resampled points and support shrinkage.

The local effect of the resampling step can be measured by comparing the additional variance of the estimate of interest deduced from the resampled particles with the variance of the estimate before resampling and has consequences on the asymptotic statistical properties of the estimate \( \hat{\Theta}_k \) [11]; thus, reducing the additional variance introduced by the resampling while keeping the advantage of this operation is critical.
and it has been proposed to reduce that additional variance by replacing the multinomial resampling step by alternative schemes such as residual, systematic or stratified resampling schemes [13] [11] [12]. However even when using such schemes particle filters based on the resampling mechanism can give poor results in some Markovian models (1), such as informative models where the likelihood $g_k(y_k|x_k)$ is sharp or high dimensional models [14].

Our aim in this paper is not to propose a new resampling scheme, but rather to revisit this complete key rejuvenation mechanism (sampling, weighting and resampling) in order to design new PF algorithms which would keep the benefits of the resampling mechanism, while avoiding the local impoverishment of the resulting MC approximation of the filtering distribution.

To that end we begin with revisiting the SIR mechanism at one single time step $k \rightarrow k + 1$. This leads us back to an analysis of Rubin’s static SIR mechanism [15, §2] [16] [10] [17, §9.2], in which, roughly speaking, one obtains samples $x^j$ approximately drawn from a target distribution $p$ by drawing intermediate samples $\{x^i\}_{i=1}^N$ from an instrumental distribution $q$, and next selecting $x^j$ among $\{x^i\}_{i=1}^N$ with a probability proportional to $q(\delta x^i)/q(x^i)$. We first observe that for fixed $N$, the samples $\{x^j\}$ produced by this SIR mechanism are dependent and marginally distributed from some compound pdf $\tilde{q}^N = \phi(p,q,N)$ which takes into account the effects of both pdfs $p$ and $q$. Here the dependency is detrimental, because samples that would be i.i.d from $\tilde{q}^N$ would produce, whichever the number of sampled and resampled particles, a moment estimate with reduced variance; this result is further illustrated by a central limit theorem (CLT) which is compared to the existing CLTs for the static IS estimate (based on the pre-resampling samples $\{\tilde{x}^i\}_{i=1}^N$, on the one hand, and for the SIR estimate (based on the post-resampling ones $\{x^j\}_{j=1}^N$), on the other hand.

We next propose a procedure to obtain i.i.d. samples from $\tilde{q}^N$, which leads to the computation of two point estimates of $\Theta = \int \varphi(x)p(x)dx$. The first one is based on unweighted i.i.d. samples and is an improved version of the classical (i.e., dependent) SIR estimate; the second one is based on post-resampling-weighted i.i.d. samples and can be seen as a new IS estimate, based on the compound pdf $\tilde{q}^N$. Finally we adapt these results to the sequential computation of $\Theta_k$ in model (1). We thus propose two new PF algorithms. One of them has an interesting interpretation in terms of Auxiliary Particle Filter (APF); more precisely, that algorithm naturally produces an i.i.d. sampling-weighted i.i.d. samples and can indeed be seen as a compound IS density $q^N = \phi(p,q,N)$ produced by the succession of the sampling (S), weighting (W) and resampling (R) steps. We prove the benefits of drawing independent samples from $q^N$ (see section II-B), and next propose to reweight these independent samples with post-resampling weights $w^i \propto \int \frac{p(x)}{q(x)}$ (see section II-C). In all this section we assume the scalar case for simplicity. We end the section with a summary (see section II-D).

A. The dependent SIR mechanism

Let us begin with a brief review of Rubin’s classical SIR sampling mechanism described in Algorithm 1 and of the properties of the sampled and resampled particles.

1) Properties of the sampled particles $\{\tilde{x}^i\}_{i=1}^N$: In the context of this paper we first recall the principle of IS. Let $p(x)$ be a probability density function and assume that we want to compute

$$\Theta = \int \varphi(x)p(x)dx = E_p(\varphi(x)).$$

In the Bayesian framework $p(x)$ is generally only known up to a constant, i.e. $p(x) \propto p_u(x)$ (subscript u is for unnormalized) and it is not possible to obtain samples directly drawn from $p(x)$. A solution is to introduce an importance distribution $q(x)$ which satisfies $q(x) > 0$ when $p(x) > 0$ and to rewrite $\Theta$ as the ratio of two expectations w.r.t. $q$.

$$\Theta = \frac{\int \varphi(x)p_u(x)q(x)dx}{\int p_u(x)q(x)dx} = \frac{E_q(\varphi(x)p_u(x)/q(x))}{E_q(p_u(x)/q(x))}.$$  

Next, each expectation is approximated by an MC method based on $N$ i.i.d. samples $\{\tilde{x}^1, \cdots, \tilde{x}^N\}$ drawn from $q(.)$: the IS estimate of $\Theta$ is given by

$$\hat{\Theta}^{IS,N} = \sum_{i=1}^N w^i \varphi(\tilde{x}^i) = \bar{\tilde{p}}(\varphi(x))$$

where

$$\bar{\tilde{p}}(x) = \sum_{i=1}^N w^i \delta_{\tilde{x}^i}(x)$$

and $w^i$ (the $i$-th normalized importance weight) reads

$$w^i = \frac{p_u(\tilde{x}^i)}{q(\tilde{x}^i)} = \frac{p(x)}{q(x)}.$$
As is well known [5], under mild assumptions
\[ \hat{\Theta}^{\text{SIR}, N} \overset{\text{D}}{=} \Theta, \] (9)
and a CLT is available too [5] (\( \overset{\text{D}}{\rightarrow} \) denotes convergence in distribution):
\[ \sqrt{N}(\hat{\Theta}^{\text{SIR}, N} - \Theta) \overset{\text{D}}{\rightarrow} \mathcal{N}(0, \text{E}(p^2(x)/q^2(x)(\varphi(x)) - \Theta^2)). \] (10)

2) Properties of the resampled particles \( \{x^i\}_{i=1}^{M_N} \). From (9) and (10), \( \hat{\rho} \) can be seen as a discrete approximation of the target density \( p \), and one expects that for large \( N \), (re)sampling from \( \hat{\rho} \) would produce samples approximately drawn from \( p \). This is the rationale of Rubin’s SIR mechanism [15, 52, 16, 10, 17, §9.2]. More precisely, let us as above draw \( N \) i.i.d. samples \( \tilde{x}^i \) from \( q \), and next \( M_N \) i.i.d samples \( x^i \) from \( \tilde{p} \) in (7). It is indeed well-known (see [10] [15]) that when \( N \rightarrow \infty \), each r.v. \( x^i \) produced by this mechanism converges in distribution to \( p \), so Rubin’s technique can be seen as a two-step sampling mechanism which transforms samples drawn from \( q \) into samples (approximately) drawn from \( p \).

This convergence result can be completed by a CLT which involves the estimate of \( \Theta \) based on the unweighted set \( \{((1/M_N, x^i))\}_{i=1}^N \):
\[ \hat{\Theta}^{\text{SIR}, M_N} = \frac{1}{M_N} \sum_{i=1}^{M_N} \varphi(x^i). \] (11)

Let \( N \rightarrow \infty \), let \( M_N \) be a non-decreasing sequence with \( M_N \rightarrow \infty \), and let \( \lim_{N \rightarrow \infty} \frac{M_N}{N} = \alpha > 0 \) (possibly \( \infty \)); then under mild conditions (see e.g. [17, §9])
\[ \sqrt{M_N}(\hat{\Theta}^{\text{SIR}, M_N} - \Theta) \overset{\text{D}}{\rightarrow} \mathcal{N}(0, \text{E}((p^2(x)/q^2(x)(\varphi(x)) - \Theta^2)). \] (12)

If \( \alpha \rightarrow \infty \) then the asymptotic variance tends to \( \text{var}_{p}(\varphi(x)) \), which shows that the SIR estimate asymptotically has the same behavior as a crude Monte Carlo estimate directly deduced from \( M_N \) samples drawn according to the target distribution \( p(.) \), provided the number \( N \) of intermediate samples is large as compared to \( M_N \).

However, for computational reasons, the numbers of samples \( N \) and \( M_N \) should not be too large in practice. Consequently we now focus on the samples produced by the SIR procedure from a non asymptotical point of view and we have the following result.

**Proposition 1.** Let us consider the samples \( \{x^i\}_{i=1}^{M_N} \) produced by the SIR mechanism described above. Then these samples are identically distributed according to a pdf \( \hat{q}^N \), with
\[ \hat{q}^N(x) = Nh^N(x)q(x), \] (13)
\[ h^N(x) = \frac{\int \frac{p(x)}{q(x)} \sum_{l=1}^{N-1} \frac{p(x_l)}{q(x_l)} q(x_l)dx_l}{\int \frac{p(x)}{q(x)} \sum_{l=1}^{N-1} \frac{p(x_l)}{q(x_l)} q(x_l)dx_l}. \] (14)

**Proof.** The proof is given in the Appendix.

So for fixed sample size \( N \), the SIR mechanism produces dependent samples \( \{x^i\}_{i=1}^{M_N} \) distributed according to \( \hat{q}^N \) (these samples are independent given the intermediate set \( \{\tilde{x}^i\}_{i=1}^N \), but become dependent when this conditioning is removed). In practice, this dependency results in support shrinkage since, by construction, an intermediate sample \( \tilde{x}^i \) can be resampled several times, and \( \{x^i\}_{i=1}^{M_N} \) is a subset of \( \{\tilde{x}^i\}_{i=1}^N \). For instance let \( M_N = N \). If we assume that \( w^i = 1 \) for some \( j \) and \( w^i = 0 \) for \( i \neq j \), then \( x^i = \tilde{x}^j \) for all \( i \). By contrast, if \( w^i = 1/N \) for all \( i \), then the average number of different samples \( \{x^i\}_{i=1}^N \) is approximately \( 2N/3 \) [18]. Nevertheless the resampling step remains useful in a dynamic setup (see section III): even though locally it leads to an impoverishment of the diversity, this step is critical for recreating diversity at the next time step.

**B. The independent SIR mechanism**

Observe that the two factors in (13) reflect the effects of the sampling and resampling step: pdf \( q \) is used in the S step, while \( h^N(x) \), which can be interpreted as the conditional expectation of a normalized importance weight when its associated particle is \( x \), results from the (W,R) steps. So particles drawn from \( q^N \) are likely to be in regions where both \( q(.) \) and the ratio \( p(.) / q(.) \) are large. Now our objective is to propose an alternative mechanism which, in the sequential case, will produce the same positive effect as the classical SIR mechanism (i.e. fighting against weight degeneracy by eliminating the samples with weak importance weights), while ensuring the diversity of the final support. Such a support diversity is ensured if we draw samples independently from the continuous pdf \( q^N(.) \). We first study the potential benefits of this sampling mechanism (see section II-B1) and next discuss its implementation (see section II-B2).

1) **Statistical properties:** Let us now assume that we have at our disposal a set of \( M_N \) i.i.d samples \( \{\tilde{x}^i\}_{i=1}^{M_N} \) drawn from \( q^N(.) \) defined in (13) (14). Before addressing the practical computation of such a set (see section II-B2), let us study its properties by considering the crude estimate of \( \Theta \) based on these \( M_N \) i.i.d samples:
\[ \hat{\Theta}^{I-\text{SIR}, M_N} = \frac{1}{M_N} \sum_{i=1}^{M_N} \varphi(x^i). \] (15)

(1 in notation I-SIR stands for independent). Our aim is to compare \( \hat{\Theta}^{I-\text{SIR}, M_N} \) to \( \hat{\Theta}^{\text{SIR}, M_N} \), and more generally \( \hat{\Theta}^{\text{IS}, N}, \hat{\Theta}^{\text{SIR}, M_N} \) and \( \hat{\Theta}^{I-\text{SIR}, M_N} \). We first have the following result.

**Proposition 2.** Let us consider the three estimates \( \hat{\Theta}^{\text{IS}, N}, \hat{\Theta}^{\text{SIR}, M_N} \) and \( \hat{\Theta}^{I-\text{SIR}, M_N} \) defined in (6), (11) and (15) respectively. Then
\[ E(\hat{\Theta}^{\text{IS}, N}) = E(\hat{\Theta}^{\text{SIR}, M_N}) = E(\hat{\Theta}^{I-\text{SIR}, M_N}), \] (16)
\[ \text{var}(\hat{\Theta}^{\text{SIR}, M_N}) = \text{var}(\hat{\Theta}^{I-\text{SIR}, M_N}) + \frac{M_N - 1}{M_N} \text{var}(\hat{\Theta}^{\text{IS}, N}). \] (17)

**Proof.** The proof is given in the Appendix.

Equation (17) ensures that an estimate based on independent samples obtained from \( \hat{q}^N \) would outperform the classical SIR estimate (at this step, we have not discussed on how to obtain such samples); the gain of \( \hat{\Theta}^{I-\text{SIR}, M_N} \) w.r.t. \( \hat{\Theta}^{\text{SIR}, M_N} \) depends on \( \text{var}(\hat{\Theta}^{\text{IS}, N}) \). On the other hand it is
well known (see e.g. [17, p. 213]) that \( \text{var} (\hat{\Theta}^{\text{SIR}, MN}) = \text{var}(\hat{\Theta}^{\text{IS}, N}) + E[\text{var}(\hat{\Theta}^{\text{SIR}, MN}|\{\hat{X}_i\}_{i=1}^N)] \); so both \( \hat{\Theta}^{\text{SIR}, MN} \) and \( \hat{\Theta}^{\text{IS}, N} \) are preferable to \( \hat{\Theta}^{\text{SIR}, M_N} \).

Now, comparing the variance of \( \hat{\Theta}^{\text{IS}, N} \) to that of \( \hat{\Theta}^{\text{SIR}, M_N} \) is more difficult, because we have to compare \( \frac{1}{M_N} \text{var}_N(\varphi(X)) \) to \( \text{var}(\sum_{i=1}^N w^i(x^1, \ldots, x^N)\varphi(x^i)) \) where \( x^i \sim_d q(x) \). However, we have the following CLT.

**Theorem 1.** Let us consider the independent SIR estimate defined in (15). Let us assume that \( N \to \infty, M_N \to \infty \) and \( \lim_{N \to \infty} \frac{M_N}{N} = \alpha > 0 \) (possibly \( \infty \)). We also assume that \( E \left( \sum_{i=1}^N w^i(x^i) \right)^2 \) is finite. Then \( \hat{\Theta}^{\text{SIR}, M_N} \) satisfies

\[
\sqrt{M_N} (\hat{\Theta}^{\text{SIR}, M_N} - \Theta) \xrightarrow{D} N(0, \text{var}_R(\varphi(x))) \tag{18}
\]

**Proof.** The proof is given in the Appendix.

Let us comment this result. First Theorem 1 again enables to compare \( \hat{\Theta}^{\text{SIR}, M_N} \) to \( \hat{\Theta}^{\text{IS}, N} \). Comparing (12) and (18) confirms (17), since the asymptotic variance of \( \hat{\Theta}^{\text{SIR}, M_N} \) is always lower than that of \( \hat{\Theta}^{\text{IS}, N} \). Also note that in the independent case the asymptotic variance of \( \hat{\Theta}^{\text{SIR}, M_N} \) no longer depends on \( \alpha > 0 \).

Next Theorem 1 also gives some elements for comparing \( \hat{\Theta}^{\text{SIR}, M_N} \) to \( \hat{\Theta}^{\text{IS}, N} \). Let for simplicity \( M_N = N \). Then the comparison of both estimates relies on that of the asymptotic variances in (10) and (18):

\[
\sigma^2_{\text{IS}, \infty}(q) = E_q \left( \frac{p^2(x)}{q^2(x)} (\varphi(x) - \Theta)^2 \right), \tag{19}
\]

\[
\sigma^2_{\text{SIR}, \infty}(q) = \text{var}_p(\varphi(x)). \tag{20}
\]

For a given target pdf \( p(\cdot) \) and function \( \varphi(\cdot) \), note that even though \( \hat{\Theta}^{\text{SIR}, M_N} \) is obtained from samples drawn from the importance pdf \( q(\cdot) \), \( \sigma^2_{\text{SIR}, \infty}(q) \) no longer depends on \( q(\cdot) \), by contrast with \( \sigma^2_{\text{IS}, \infty}(q) \). Now, \( \sigma^2_{\text{IS}, \infty}(q) \) is well known [19, §2.9] [5, Theorem 3] to be minimum for \( q^*(x) \propto p(x)|\varphi(x)| - \Theta \); for that \( q^*, \sigma^2_{\text{IS}, \infty}(q^*) = \left( \int |\varphi(x)| - \Theta |p(x)dx| \right)^2 \leq \text{var}_p(\varphi(x)) \), so \( \hat{\Theta}^{\text{IS}, N}(q^*) \) outperforms \( \hat{\Theta}^{\text{SIR}, \infty} \) for large values of \( N \). However note that for other importance distributions, \( \sigma^2_{\text{SIR}, \infty}(q) \) may become lower than \( \sigma^2_{\text{IS}, \infty}(q) \), by contrast with traditional resampling techniques where resampling always introduces extra variance. Also note that the variances in (19) and (20) depend on function \( \varphi(\cdot) \); on the other hand, for large \( N \), \( \hat{\Theta}^{\text{SIR}, \infty} \) has the same behavior as a crude estimate built from samples drawn from \( p(\cdot) \) and so is adapted for a large class of functions \( \varphi(\cdot) \).

2) **Sampling procedure:** It remains to describe a procedure to obtain i.i.d. samples from \( \hat{q}^N \). Algorithm 2 ensures that the final samples \( \{x^1, \ldots, x^M_N\} \) are drawn independently from \( \hat{q}^N \).

Compared to the classical SIR procedure based on \( N \) intermediate samples from which are deduced the \( M_N \) final samples, the independent SIR algorithm described in Algorithm 2 relies on a sampling step of \( N \times M_N \) intermediate samples \( \tilde{x}^{i,j} \) and \( M_N \) independent resampling steps. Consequently, for a given budget of sampling and resampling steps, the independent procedure should be compared with a classical SIR one in which we sample \( N \times M_N \) points and resample \( M_N \) of them. In this last case, we obtain \( M_N \) dependent samples drawn from \( \tilde{q}^N \times \hat{q}^N \). First, using (12) with \( \alpha = \lim_{N \to \infty} \frac{N \times M_N}{M^{'}} = \infty \), we see that both estimates \( \hat{\Theta}^{\text{SIR}, M_N} \) and \( \hat{\Theta}^{\text{SIR}, MN} \) with \( M_N \times M_N \) intermediate samples have the same asymptotic behavior. However the independent procedure can be easily parallelized because the resampling steps are by nature independent contrary to the SIR procedure where the \( N \times M_N \) intermediate samples are directly resampled. Comparing both estimates in the case where the number of particles is the same before and after the resampling step of the SIR dependent algorithm is more difficult. Indeed, in this case \( \alpha = 1 \) in (12) and the rates in (12) and in (18) are not necessarily the same. So the comparison also depends on E_q \left( (\frac{p^2(x)}{q^2(x)} (\varphi(x) - \Theta)^2) \right).

C. **Reweighting the independent samples?**

We finally discuss the final weights which are attributed to the resampled particles. In the SIR procedure, each final sample is weighted by \( 1/M_N \). From an IS point of view, this weighting traduces the fact that the final samples become drawn from the target distribution \( p(\cdot) \) and independent when \( N \to \infty \) [15]. Moreover the convergence results of \( \hat{\Theta}^{\text{SIR}, MN} \) to \( \Theta \) (see e.g. [20] [17]) confirm that these weights are valid from an asymptotical point of view. In the independent SIR procedure, the only difference is that the final samples are independent, even from a non-asymptotical point of view.

Now, if \( N \) is finite, one can wonder if weights \( 1/M_N \) are optimal. In Algorithm 2, samples \( \{x^i\}_{i=1}^M \) are i.i.d. from \( \tilde{q}^N \). Consequently, for a given \( N \), \( \tilde{q}^N \) can be seen as a post-resampling compound importance distribution \( \tilde{q}^N = \phi(p,q,N) \), and a final sample \( x^i \) should be weighted by a post-resampling weight proportional to \( p_u(x^i)/\tilde{q}^N(x^i) \). This
yields a new estimate \( \hat{\Theta}^{1-\text{SIR} - w} \) of (4) (superscript \( w \) stands for weighted)
\[
\hat{\Theta}^{1-\text{SIR} - w,M_N} = \frac{M_N}{\sum_{j=1}^{M_N} \varphi_j(x)^2} \varphi_j(x),
\]  
(21)
which coincides with the IS estimate (6) with importance distribution \( \tilde{q}^N(.) \). It is difficult to compare \( \hat{\Theta}^{1-\text{SIR} - w,M_N} \) and \( \hat{\Theta}^{1-\text{SIR} - w,M_N} \) because the expression of the weights in this last case depends on \( M_N \). However, it is interesting to note that contrary to \( \hat{\Theta}^{1-\text{SIR} - w,M_N} \), \( M_N \) impacts the bias of the estimate \( \hat{\Theta}^{1-\text{SIR} - w,M_N} \). For example, if we set \( N = 1 \) (so \( \tilde{q}^N = q \)) and \( M_1 \) is arbitrary then \( \hat{\Theta}^{1-\text{SIR} - w,M_N} \) coincides with the IS estimate with \( M_1 \) i.i.d. samples drawn from \( q \) while the unweighted estimate \( \hat{\Theta}^{1-\text{SIR} - w} \) is a crude estimate of \( \int \varphi(x)q(x)dx \) and is not adapted to the estimation of \( \Theta \). More generally, using the delta method to approximate \( E(\hat{\Theta}^{1-\text{SIR} - w,M_N}) \) and \( E(\hat{\Theta}^{1-\text{SIR} - w,M_N}) \) [21] we observe that
\[
E(\hat{\Theta}^{1-\text{SIR} - w,M_N}) = E(\hat{\Theta}^{\text{IS} - w,N}) \approx \Theta - \frac{1}{M_N} \sum_{i=1}^{M_N} \varphi_i(x)^2 \varphi(x) - \Theta. \]
(22)
\[
E(\hat{\Theta}^{1-\text{SIR} - w,M_N}) \approx \Theta - \frac{1}{M_N} \sum_{i=1}^{M_N} \varphi_i(x)^2 \varphi(x) - \Theta. \]
(23)

So for a fixed number of sampled points \( N \), we see that in the unweighted case the bias of \( \hat{\Theta}^{1-\text{SIR} - w,M_N} \) is independent of \( M_N \). By contrast, whichever \( N \) the bias of \( E(\hat{\Theta}^{1-\text{SIR} - w,w,M_N}) \) tends to 0 as \( M_N \to \infty \).

Finally, it remains to compute \( p_u(x_i)/\tilde{q}^N(x_i) \) in practice. In general, \( \tilde{q}^N \) in (13) is not available in closed form because it relies on the integral \( h^N(x) \) in (14). However, the \( N \times M_N \) intermediate samples which have been used in Algorithm 2 can be recycled to approximate the conditional expectation \( h^N(x) \). For a given \( x \) and using the intermediate samples \( x^j \) of Algorithm 2, a crude Monte Carlo estimate of \( h^N(x) \) reads
\[
\hat{h}^N(x) = \sum_{i=1}^{M_N} \frac{p_u(x_i)}{\tilde{q}^N(x_i)} \varphi(x_i),
\]  
(24)
Importance weights \( \frac{p_u(x_i)}{\tilde{q}^N(x_i)} \) in (21) can be approximated by \( \frac{p_u(x_i)}{\tilde{q}^N(x_i)} \hat{h}^N(x) \). Note that the computation of these approximated weights does not require extra cost since \( p_u(\tilde{x}^i) \) has already been computed in Algorithm 2 to obtain i.i.d. samples.

D. Summary

In summary, we now have at our disposal four estimates to compute \( \Theta \) in (4) from an importance distribution \( q(.) \). \( \hat{\Theta}^{\text{IS} - w,N} \) and \( \hat{\Theta}^{\text{IS} - w,M_N} \) are deduced from the IS and Rubin’s SIR mechanisms, respectively. \( \hat{\Theta}^{\text{SIR} - w,M_N} \) relies on unweighted dependent samples from \( \tilde{q}^N \). Using unweighted independent samples from \( \tilde{q}^N \) produces the estimate \( \hat{\Theta}^{1-\text{SIR} - w,M_N} \) which outperforms \( \hat{\Theta}^{\text{SIR} - w,M_N} \) and possibly \( \hat{\Theta}^{\text{IS} - w,N} \); it also becomes asymptotically independent of the choice of the initial importance distribution \( q(.) \) according to Theorem 1. This estimate does not suffer from the support impoverishment caused by the resampling step. On the other hand it requires a larger computational cost which, however, can be exploited in order to associate to the i.i.d. samples post-resampling importance weights based on the \( \tilde{q}^N(x) \). We thus obtain a weighted estimate \( \hat{\Theta}^{1-\text{SIR} - w,M_N} \) which can be seen as the estimate deduced from the IS mechanism based on the compound IS distribution \( \tilde{q}^N(x) \). We will compare these estimates via simulations taking into account their computational costs in Section IV-A.

III. INDEPENDENT RESAMPLING BASED PF

We now adapt these results to the Bayesian filtering problem. In section III-A we briefly recall the principle of classical SIR algorithms which are based on dependent resampling. Our SIR algorithm with independent resampling and unweighted samples is proposed in section III-B. However, computing the post-resampling weights is more challenging here than in the static case because the pdf \( \tilde{q}^N \) of the static case becomes a sum of \( N \) terms which should be computed for each final sample. So in section III-C we revisit the algorithm of section III-B in terms of APF. We first observe that the independent SIR algorithm can be seen as the first step of an APF algorithm since it implicitly draws samples from a mixture pdf. Making full use of the APF methodology enables us to weight our final samples.

A. Classical SIR algorithms (based on dependent resampling)

We now assume that we are given some HMC model (1) and we briefly recall how \( \Theta_k \) in (2) can be computed recursively via PF. PF relies on the sequential application of the normalized IS mechanism described in Section II-A for the target distribution \( p(\theta_{0:k}|y_{0:k}) \) which is known up to a constant according to (1). Let \( q(\theta_{0:k}) \) be an importance distribution \( q(\theta_{0:k}) \) can depend on \( y_{0:k} \) but this dependency is not written here to avoid notational burden). Starting from \( \Theta \) weighted trajectories \( \theta_{0:k-1} \) sampled from \( q(\theta_{0:k-1}) \), we first extend each trajectory \( \theta_{0:k-1} \) with a particle \( \tilde{x}_k^i \) sampled from \( q(x_k|\tilde{x}_{0:k-1}^i) \) and next update the old weights \( w_{k-1}^i \ via
\[
w_k^i \propto w_{k-1}^i \frac{f_k(\tilde{x}_k^i|\theta_{0:k-1})}{f_k(\tilde{x}_k^i|\theta_{0:k-1})} q(\theta_{0:k}|\tilde{x}_{0:k-1}^i) \]
(25)
Unfortunately, it is well-known that this direct sequential application of IS leads to weight degeneracy: after a few iterations only few weights \( w_k^i \) have a non null value [22]. A traditional rescue consists in resampling, either systematically or according to some criterion such as the Effective Sample Size [6] [7] which is approximated by \( 1/\sum_{i=1}^{N} (w_k^i)^2 \). The corresponding algorithm is given in Algorithm 3 and we shall assume that the size \( N \) of the MC approximation remains constant throughout the iterations. Finally Algorithm 3 enables to compute two estimates of \( \Theta_k \):
\[
\hat{\Theta}_k^{\text{SIR} - N} = \sum_{i=1}^{N} w_k^i \varphi(x_k^i),
\]  
(26)
\[
\hat{\Theta}_k^{\text{SIR} - N} = \frac{1}{N} \sum_{i=1}^{N} \varphi(x_k^i).
\]  
(27)
As is well known, the pre-resampling estimator \( \hat{\Theta}_k^{\text{SIS}} \) is preferable to the post-resampling one \( \hat{\Theta}_k^{\text{SIR}} \) and should be used in practice; but \( \hat{\Theta}_k^{\text{SIR}} \) is recalled here because it will be compared below to the independent resampling estimator (32).

Algorithm 3 The classical SIR algorithm (based on dependent resampling)

```
Input: \( q(x_k|x_{0:k-1}), y_k, \{w_{k-1}^l, x_{0:k-1}^l\}_{l=1}^N \)
Output: \( \{w_k^l, x_{0:k}^l\}_{l=1}^N \)
for \( 1 \leq i \leq N \) do
    \( S_i \sim q(x_k|x_{0:k-1}) \)
    \( w_k^i \propto w_{k-1}^i \frac{f_k(x_k|x_{0:k-1}) y_k}{q(x_k|x_{0:k-1})} \), \( \sum_{i=1}^N w_k^i = 1 \)
end for
if Resampling then
    for \( 1 \leq i \leq N \) do
        \( R_i \sim \text{Pr}(L = l | \{x_{0:k-1}^l, \tilde{x}_{i,j}^l\}_{j=1}^N) = w_k^i \)
        Set \( x_{0:k}^i = (x_{0:k-1}^i, \tilde{x}_{i,j}^i) \), \( w_k^i = \frac{1}{N} \)
    end for
else
    Set \( \{x_{0:k}^i\}_{i=1}^N = \{\tilde{x}_{i,j}^l\}_{i,j=1}^N \)
end if
```

In practice, it remains to design the conditional importance distribution \( q(x_k|x_{0:k-1}) \). A popular solution consists in choosing \( q(x_k|x_{0:k-1}) = f_k(x_k|x_{0:k-1}) \), since this pdf is part of model (1) and is generally easy to sample from; another one is the so-called optimal conditional importance distribution \( q(x_k|x_{0:k-1}) = p(x_k|x_{0:k-1}, y_k) \) which takes into account the new observation \( y_k \) and for which weights \( w_k^i \) no longer depend on the sampled particles \( \{\tilde{x}_{i,j}^l\}_{i,j=1}^N \). The optimal conditional importance distribution is generally not available in closed form but some approximation techniques have been proposed, see e.g. \cite{22} \cite{23} \cite{24} \cite{25}. The choice of the importance distribution will not be discussed in this paper but all the existing improvements cited above can be used in the algorithms that we propose in the next sections: they do not impact the proposed methodology. Finally, let us mention that convergence results are also available for the PF presented in Algorithm 3, see e.g. \cite{26} \cite{27} \cite{20} \cite{17}. Some of them are based on an extension of the CLTs recalled in Section II for conditionally independent triangular arrays and next on their recursive application.

B. An alternative SIR algorithm (based on independent resampling)

Let us first adapt Proposition 1 to the sequential context. So we address the conditional distribution given \( \{x_{0:k-1}^i\}_{i=1}^N \) of the resampled particles \( x_k^i \) and we have the following result (the proof is omitted).

**Proposition 3.** Let us consider the samples \( \{x_k^i\}_{i=1}^N \) produced by the SIR mechanism of Algorithm 3. Let

\[
p_k^i(x) = w_{k-1}^i f_k(x|x_{k-1}^i) g_k(y_k|x),
\]

\[
q_k^i(x) = q(x|x_{0:k-1}^i).
\]

Then given the initial trajectories \( \{x_{0:k-1}^i\}_{i=1}^N \), the new samples \( \{x_k^i\}_{i=1}^N \) are identically distributed according to a pdf \( \tilde{q}_k^N \) which reads

\[
\tilde{q}_k^N(x) = \sum_{i=1}^N h_k^i(x) q_k^i(x),
\]

where \( h_k^i(x) \) coincides with the conditional expectation (given \( x_k^i = x \)) of the \( i \)-th importance weight at time \( k \),

\[
h_k^i(x) = \int \int \frac{p_k^i(x)}{q_k^i(x)} + \sum_{l \neq i} \frac{p_k^l(x)}{q_k^l(x)} (x^l) \cdot dx^l.
\]

Proof. The proof follows the same line of reasoning as the Proof of Proposition 1. \( \Box \)

Note that in this proposition we focus on the distribution of \( x_k^i \) given \( \{x_{0:k-1}^j\}_{j=1}^N \). Given \( \{x_{0:k-1}^l, \tilde{x}_{i,j}^l\}_{j=1}^N \), the new samples \( \{x_{0:k}^l\}_{i=1}^N \) are independent; when we remove the dependency in \( \{\tilde{x}_{i,j}^l\}_{j=1}^N \), \( \{x_{0:k}^l\}_{i=1}^N \) become identically distributed according to \( \tilde{q}_k^N \) but are dependent (a same particle can be resampled several times).

Since \( \tilde{q}_k^N \) is a pdf, a procedure which would produce samples conditionally i.i.d. from \( \tilde{q}_k^N \) would enable us to keep the advantage of the resampling step, i.e. to recreate diversity for the next time iteration, while avoiding local impoverishment of the support. Except in a particular case which will be described later, sampling directly from \( \tilde{q}_k^N \) is difficult for an arbitrary conditional importance distribution \( q(x_k|x_{0:k-1}) \). We thus propose a procedure similar to Algorithm 2 but adapted to the dynamical context. The SIR algorithm with independent resampling is given by Algorithm 4. Note that a difference with Algorithm 3 is that the distribution of the discrete index \( L^i \) now depends on \( i \). In addition, note also that the proposed mechanism is different from that proposed in \cite{28} in the context of Population MC methods. Here, the trajectories \( \{x_{0:k}^i\} \) still remain dependent: some of them can disappear at the end of the procedure or be partially replicated but the last components \( \{x_{k}^i\} \) of these trajectories are necessarily different.

Algorithm 4 A SIR algorithm based on independent resampling

```
Input: \( q(x_k|x_{0:k-1}), y_k, \{w_{k-1}^l, x_{0:k-1}^l\}_{l=1}^N \)
Output: \( \{w_k^i, x_{0:k}^i\}_{i=1}^N \)
for \( 1 \leq i \leq N \) do
    for \( 1 \leq j \leq N \) do
        \( S_{i,j} \sim q(x_k|x_{0:k-1}^j) \)
        \( w_k^{i,j} \propto w_{k-1}^{i,j} \frac{f_k(x_k|x_{0:k-1}^j) y_k}{q(x_k|x_{0:k-1}^j)} \), \( \sum_{j=1}^N w_k^{i,j} = 1 \)
    end for
    \( R_i \sim \text{Pr}(L \in \{j \mid \{x_{0:k-1}^j, \tilde{x}_{i,j}^j\}_{j=1}^N\}) = w_k^{i,j} \)
    Set \( x_{0:k}^i = (x_{0:k-1}^i, \tilde{x}_{i,j}^i) \), \( w_k^i = \frac{1}{N} \)
end for
```

We now propose a new estimate \( \hat{\Theta}_k^{1-\text{SIR}} \) of \( \Theta \) which is based on the set \( \{x_k^i\}_{i=1}^N \) produced by Algorithm 4:

\[
\hat{\Theta}_k^{1-\text{SIR}} = \frac{1}{N} \sum_{i=1}^N \varphi(x_k^i).
\]
Comparing (32) with (27), remember that the samples \( \{x^i_{0:k-1}\}^N_{i=1} \) share the same pdf \( \tilde{q}_{k}^{N} \), but that in (32) they are now independent given \( \{x^i_{0:k-1}\}^N_{i=1} \). Starting from a dataset \( \{x^i_{0:k-1}\}^N_{i=1} \), it is ensured that \( \tilde{\Theta}_{k}^{N} \) outperforms \( \Theta_{k}^{N} \) since

\[
E(\tilde{\Theta}_{k}^{1-\text{SIR},N} | \{x^i_{0:k-1}\}^N_{i=1}) = E(\Theta_{k}^{\text{SIR},N} | \{x^i_{0:k-1}\}^N_{i=1}),
\]

\[
\text{var}(\tilde{\Theta}_{k}^{\text{SIR},N} | \{x^i_{0:k-1}\}^N_{i=1}) = \text{var}(\Theta_{k}^{1-\text{SIR},N} | \{x^i_{0:k-1}\}^N_{i=1})
\]

\[
+ \frac{N-1}{N} \text{var}(\tilde{\Theta}_{k}^{\text{SIR},N} | \{x^i_{0:k-1}\}^N_{i=1}).
\]

Of course, computing \( \tilde{\Theta}_{k}^{1-\text{SIR},N} \) via the samples produced by Algorithm 4 requires an extra computational cost. This point will be discussed in detail in our Simulations section, but for the moment let us make two comments: first, this algorithm can be seen as an alternative SIR mechanism which ensures the diversity of the resampled support without changing the conditional distribution of the final samples; if resampling needs to be performed rarely, then the independent resampling procedure may be used only when necessary. On the other hand, we will see that \( \tilde{\Theta}_{k}^{1-\text{SIR},N} \) can also provide an interesting alternative to \( \Theta_{k}^{\text{SIR},N} \) but requires an extra computational cost; so if we want to perform the independent resampling procedure at each time step we will decrease the number \( N \) of particles associated with \( \tilde{\Theta}_{k}^{1-\text{SIR},N} \) in order to reach the same computational cost associated with \( \Theta_{k}^{\text{SIR},N} \).

**Remark 1.** Note that the idea of using extra MC samples has already been proposed in the context of Island PFs [29]. The idea behind this class of techniques is to exploit parallel architectures, and the rationale is as follows. Instead of considering a unique set of \( N \) particles, the method consists in dividing the population of \( N \) samples into \( N_1 \) sets of \( N_2 \) samples with \( N_1N_2 = N \). It is well known that such a configuration does not improve the classical PF with \( N \) samples, but it has the advantage of splitting the associated computational cost when parallel architectures are available. In other words, the objective of the Island PF is not to struggle against the support impoverishment.

**C. Interpretation of the independent resampling scheme in terms of APF**

At this point, we have seen that it was possible to obtain an estimate of \( \Theta_{k}^{N} \) based on conditionally i.i.d. samples from the conditional pdf \( \tilde{q}_k^N \). As in the static case, we now wonder whether the final weights \( 1/N \) used to compute \( \tilde{\Theta}_{k}^{1-\text{SIR},N} \) (see eq. (32)) are optimal when \( N \) is finite. To this end we would like to make use of the expression of \( \tilde{q}_k^N \) to propose an alternative weighting mechanism. At first glance, the computation of a weight which would rely on (30)-(31) seems compromised because \( \tilde{q}_k^N \) involves a sum of \( N \) terms which should be computed for each final sample \( x^i_k \). As we will see, the interpretation of the independent SIR algorithm as a particular first step of an APF algorithm will help circumvent this limitation. Let us first begin with a brief presentation of APF filters.

1) A brief presentation of APF: In model (1), the filtering density at time \( k \) can be written in terms of that at time \( k-1 \),

\[
p(x_k|y_{0:k}) \propto g_k(y_k|x_k) \int f_k(x_k|x_{k-1})p(x_{k-1}|y_{0:k-1})dx_{k-1}.
\]

Plugging an MC approximation \( \{w^i_{k-1},x^i_{k-1}\}^N_{i=1} \) of \( p(x_{k-1}|y_{0:k-1}) \) into (35) yields

\[
\tilde{p}(x_k|y_{0:k}) \propto g_k(y_k|x_k) \sum_{i=1}^{N} w^i_{k-1} f_k(x_k|x^i_{k-1})
\]

\[
\propto \sum_{i=1}^{N} w^i_{k-1} p(y_k|x_{k-1}) p(x_k|x^i_{k-1}, y_k)
\]

where \( p(y_k|x_{k-1}) = \int f_k(x_k|x_{k-1})g_k(y_k|x_k)dx_k \) and \( p(x_k|x_{k-1}, y_k) = f_k(x_k|x_{k-1})g_k(y_k|x_k) \). Sampling from \( \tilde{p}(x_k|y_{0:k}) \) in (36) leads to a particular SMC algorithm referred to as the fully adapted APF (FA-APF) [30]. However sampling directly from \( \tilde{p}(x_k|y_{0:k}) \) is often impossible because \( p(y_k|x^i_{k-1}) \) or \( p(x_k|x^i_{k-1}, y_k) \) are unavailable. Some approaches have been proposed to sample approximately from the mixture (36). For example, [31] has proposed to first approximate \( p(y_k|x^i_{k-1}) \) via a local MC method and to sample from \( p(x_k|x^i_{k-1}, y_k) \) via rejection sampling. However, the computational cost associated to this method is random and the rejection method can be particularly inefficient in informative or high dimensional models. An alternative approach called the APF has been proposed [30] to obtain samples from an instrumental mixture pdf

\[
\tilde{q}(x_k) = \sum_{i=1}^{N} \mu(x^i_{0:k-1}) \tau(x_k|x^i_{0:k-1})
\]

and to use IS in augmented dimension; finally APF aims at targeting the mixture pdf \( \tilde{p}(x_k|y_{0:k}) \) in (36) which itself, targets the filtering distribution \( p(x_k|y_{0:k}) \). The resulting algorithm is displayed below.

**Algorithm 5 The APF algorithm**

**Input:** \( \mu(x_{0:k-1}), \tau(x_k|x_{0:k-1}), y_k, \{w^i_{k-1},x^i_{0:k-1}\}^N_{i=1} \)

**Outut:** \( \{w^i_{k},x^i_{0:k}\}^N_{i=1} \)

for \( 1 \leq i \leq N \) do

R. \( P \sim \text{Pr}(\{x^i_{0:k-1}\}^N_{i=1}) = \mu(x^i_{0:k-1})^N_{i=1}; \)

S. \( x^i_k \sim \tau(x_k|x^i_{0:k-1}); \)

W. \( w^i_k \sim \frac{w^i_{k-1} f_k(x_k|x^i_{k-1}) g_k(y_k|x_k)}{\mu(x^i_{0:k-1}) \tau(x_k|x^i_{0:k-1})}, \sum_{i=1}^{N} w^i_k = 1; \)

Set \( x^i_{0:k} = (x^i_{0:k-1}, x^i_k) \)

end for

Let us comment the choice of the instrumental distribution \( \tilde{q}(x_k) \) in (37). Compared to the SIR algorithm of paragraph III-A we see that there is an additional degree of freedom, \( \mu(x_{0:k-1}) \), which is called the first stage weight; \( \tau(x_k|x_{0:k-1}) \) is some importance distribution. Generally, the objective of the first stage weights is to avoid the computational waste induced by the resampling step of the SIR algorithm by pre-selecting trajectories at time \( k-1 \) which are in accordance with the new observation \( y_k \). Designing this pdf \( \tilde{q}(x_k) \) is critical and classical approximations of the predictive likelihood such
as the likelihood taken at the mode of the transition pdf (i.e. \( \mu(x_{0,k-1}^i) \)). The mode of \( f_k(x_{0,k-1}^i) \) can actually damage the performance of the estimate. This is why it is often suggested in practice to build a first-stage weight as close as possible to \( w_{k-1} = p(y_k|x_{k-1}) \), although this problem is generally difficult due to the computation of \( p(y_k|x_{k-1}) \). It remains to choose \( \tau(x_{k}|x_{k-1}) \); as in the SIR algorithm, one generally tries to approximate the optimal importance distribution \( p(x_k|x_{k-1}, y_k) \). Finally, note that it has been proved in [34] that the estimate which results from the direct sampling from mixture (36) has not necessarily an optimal asymptotic variance.

Some improvements of the APF have been proposed. In particular, instead of using IS in augmented dimension, the marginal APF of [35] directly targets the mixture pdf \( \tilde{p}(x_k|y_{0:k}) \) in (36). In this case, the second stage weights become

\[
\begin{align*}
  w_k^i &= \frac{\sum_{j=1}^{N} w_{k-1}^j f_k(x_k|x_{k-1}^j) g_k(y_k|x_k^i)}{\sum_{j=1}^{N} \mu(x_{0,k-1}^j) \tau(x_k^i|x_{k-1}^j)}, \quad \sum_{i=1}^{N} w_k^i = 1. \quad (38)
\end{align*}
\]

As a result, the variance of weights \( w_k^i \) in (38) is reduced w.r.t. that of the second stage weights of Algorithm 5. However, this variance reduction is closely related to the choice of the first stage weights and of the importance distribution. In particular, in the case where \( \mu(x_{0,k-1}^i) \) and \( \tau(x_{k}|x_{k-1}^i) = f_k(x_k|x_{k-1}^i) \) (i.e. it is difficult to approximate \( p(y_k|x_{k-1}^i) \) and we can only sample according to \( f_k(x_k|x_{k-1}^i) \), there is no improvement since the M-APF reduces to the booststrap PF algorithm.

2) Independent resampling as the first step of a canonical APF algorithm: Let us now turn to the interpretation of our independent resampling procedure in terms of APF. Let us observe that \( q_k^N(x) \) in (30) can be rewritten as

\[
\begin{align*}
  q_k^N(x) &= \prod_{i=1}^{N} h_k^l(x_i) q_k^l(x_i) dx \times \frac{h_k^l(x_i) q_k^l(x_i)}{\int h_k^l(x_i) q_k^l(x_i) dx} \quad (39)
\end{align*}
\]

and thus can be seen as one particular mixture pdf \( q_k(x) \) in (37), in which the weights \( \mu_{x_{0,k-1}}(x_{0,k-1}) \) are given by \( \int h_k^l(x) q_k^l(x) dx \) and the components \( \tau_{x_{0,k-1}}(x_{0,k-1}) \) by \( h_k^l(x) q_k^l(x) \int h_k^l(x) q_k^l(x) dx \). We now check that the couple of samples \( \{l^i, x_k^i\} \) produced by the independent resampling algorithm (Algorithm 4) can indeed be seen as an augmented sample according to \( q_k^N(x) \) in (39):

- given \( \{x_{0,k-1}^j\}_{j=1}^{N} \) and \( \{x_{0,k-1}^{j'}\}_{j'=1}^{N} \), \( \Pr(l^i = l) = w_l^i \). Since \( x_k^{j'} \sim q_k^l(x) \), the distribution of \( l^i \) given \( \{x_{0,k-1}^{j'}\}_{j'=1}^{N} \) becomes \( \Pr(l^i = l) = E(w_l^i|\{x_{0,k-1}^{j'}\}_{j'=1}^{N}) = \int h_k^l(x) q_k^l(x) dx \); and
- given \( \{x_{0,k-1}^{j'}\}_{j'=1}^{N} \) and \( l^i, x_k^{j'} \), the dependency in \( \{x_{0,k-1}^{j'}\}_{j'=1}^{N} \) becomes \( \tau_{x_{0,k-1}}(x_{0,k-1}^i) \).

In summary, our independent resampling procedure is nothing but the first step of one particular APF algorithm, because the pdf \( q_k^N(x) \) from which we draw i.i.d. samples (given \( \{w_{k-1}^i, x_{0,k-1}^i\}_{i=1}^{N} \)) coincides with the mixture pdf (39), which itself constitutes a class of instrumental distributions \( \tilde{q}(x_k) \) in (37) parametrized by \( q(x_k|x_{0,k-1}) \).

In order to appreciate the relevance of that particular solution let us comment on the choice of the first-stage weights \( \mu_{x_{0,k-1}}(x_{0,k-1}) \) and distributions \( \tau_{x_{0,k-1}}(x_{0,k-1}) \):

- at time \( k-1 \), trajectories \( \{x_{0,k-1}^i\}_{i=1}^{N} \) are first resampled according to first stage weights which coincide with the expectation of the importance weights \( w_k^i \) of the SIR algorithm defined in (25). In other words, these trajectories are preselected in such a way that the new importance weight \( w_k^i \) which will be assigned in the weighting step of the SIR algorithm will tend to be large;

- once a trajectory \( x_{0,k-1}^i \) has been selected, it is not ensured that its associated weight \( w_k^i \) will indeed be large. By sampling according to a pdf proportional to \( h_k^l(x) q_k^l(x) \), the objective is to produce a sample in a region where \( h_k^l(x) \) (the conditional expectation of the importance weight \( w_k^i \), given that \( x_{0,k-1}^i = x^i \)) and the distribution \( q_k^l(x) \) are both large.

Consequently, the mixture pdf \( q_k^N(x) \) appears as a natural instrumental candidate for the APF when the objective is to pre-select the trajectories and extend them in accordance with the given conditional importance distributions \( q_k^l(x) = q(x_k|x_{0,k-1}) \) used in the SIR algorithm. If the SIR algorithm IS densities \( q_k^l(x) \) coincide with the optimal importance distribution \( p(x_k|x_{k-1}, y_k) \), then one can easily see that our canonical APF instrumental pdf (39) reduces to the target mixture (36) (since \( h_k^l(x) \) in (31) is reduced to a term proportional to \( \int q_k^l(x) f_k(x_k|x_{k-1}) \)) and the independent SIR procedure to the FA-APF algorithm. In that case one can sample from \( q_k^N \) very efficiently (since (39) is a known mixture) and the resulting estimate outperforms the SIR estimate \( \Theta_k^{N \sim \text{SIR}} \) with optimal importance distribution [17] [18]. In case the FA-APF algorithm is not available, it remains possible to sample from the mixture pdf \( q_k^N(x) \) in (39) as soon as we can sample from the root pdf \( q_k^l(x) \), even when \( \mu_{x_{0,k-1}}(x_{0,k-1}) \) cannot be computed, or one cannot sample from \( \tau_{x_{0,k-1}}(x_{0,k-1}) \).

3) Reweighting the independent samples: We can finally use this APF interpretation in order to reweight our conditionally independent samples \( \{x_k^i\}_{i=1}^N \). Since \( q_k^N(x) \) can be seen as a mixture (37) with parameters \( \mu_{x_{0,k-1}}(x_{0,k-1}) \) and \( \tau_{x_{0,k-1}}(x_{0,k-1}) \), \( \mu_{x_{0,k-1}}(x_{0,k-1}) \times \tau_{x_{0,k-1}}(x_{0,k-1}) \) reduces to \( h_k^l(x) q_k^l(x) \). Finally when we target mixture (36), the second-stage weights associated with the independent samples \( x_{0,k-1}^i \) produced by Algorithm 4 read

\[
\begin{align*}
  w_k^i \propto \frac{w_{k-1}^i \int f_k(x_k|x_{0,k-1}^i) g_k(y_k|x_k^i)}{h_k^l(x_k^i) q_k^l(x_k^i)} \times \frac{1}{\sum_{i=1}^{N} w_k^i} = 1 \quad (40)
\end{align*}
\]

(note that we could compute these second stage weights in the context of the M-APF using (38) at the price of an extra-computational cost). We thus obtain a new estimate of \( \Theta_k \),

\[
\Theta_k^{1 \text{SIR} - \text{w}} = \sum_{i=1}^{N} w_k^i \varphi(x_k^i) \quad (41)
\]

where \( w_k^i \) are defined in (40). The practical computation of these final weights relies on that of \( h_k^l(x) \) in (31), which
can be approximated via the extra samples \( \tilde{x}^{i,j}_k \) generated in Algorithm 4,

\[
\hat{h}^M(x) = \frac{1}{N} \sum_{i=1}^{N} \frac{P_i(x)}{q_i(x)} + \sum_{j \neq i} \frac{P_j(\tilde{x}^{i,j}_k)}{q_j(\tilde{x}^{i,j}_k)}.
\]

Let us comment this approximation. First, it does not require an extra computational cost since samples \( \tilde{x}^{i,j}_k \) and \( \frac{P_j(\tilde{x}^{i,j}_k)}{q_j(\tilde{x}^{i,j}_k)} \) are recycled from Algorithm 4. Next, \( \hat{h}^M(x) \) can be seen as the conditional expectation of the \( h^M(x) \) is more complicated because the dimension of the integral in \( h^M(x) \) is directly related to the number of samples \( N \) used to approximate it. This point will be discussed in our Simulations section and we will see that in some cases it is possible to predict the behavior of \( \hat{h}^M(x) \).

Remark 2. Computing \( \hat{h}^{\text{SIR}} - \text{SIR} \) and \( \hat{h}^{\text{SIR}} - \text{SIR} \) is based on extra-intermediate samples as in [31], for example. However, contrary to such approach in which the extra samples are used to build a new target distribution, our extra samples build an important mixture which has an interpretation in terms of APF and enable us to obtain i.i.d. samples from this mixture and to reweight these samples.

D. Summary

Let us summarize the discussions of section III. When the objective is to compute \( \Theta_k \) in (2) we have several options:

1) using the classical SIR algorithm (see Algorithm 3) in which we compute \( \hat{\Theta}^{\text{SIR}}_k \) defined in (26). The resampling step which follows the computation of this estimate produces a conditionally dependent unweighted set of particles sampled from \( q_k^N \);

2) an alternative to avoid the local impoverishment induced by the traditional resampling step is to run Algorithm 4 and compute estimate \( \hat{\Theta}^{\text{SIR}}_k \). This estimate is still based on an unweighted set of particles marginally sampled from \( q_k^N \) but these samples have become conditionally independent;

3) finally, the samples produced by Algorithm 4 can also be seen as the result of a partial sampling procedure according to an APF instrumental mixture pdf (39). Further using the APF methodology with mixture \( q_k^N \), it is possible to target mixture (36) which itself is an approximation of \( p(x_k|y_{0:k}) \). This leads to estimate \( \hat{\Theta}^{\text{SIR}}_k \) in (41), in which the weights (40) are estimated by recycling the extra samples produced by Algorithm 4. This approximation of the weights introduces additional MC error but the interpretation of the previous point ensures that it is sufficient to propagate uniform weights.

These three estimates are now going to be compared (in terms of performances and computational cost) in the next section.

IV. Simulations

We now validate our discussions through computer-generated experiments. In section IV-A we first illustrate the results of Section II and we compare the classical resampling mechanism to the independent one with both unweighted and weighted samples. We also discuss the computational cost associated with our independent resampling mechanism.

In section IV-B we next perform simulations in the ARCH model [36]. On the one hand, the FA-APF algorithm can be computed in this model [30]. On the other hand, remember that our weighted estimate (41) can be interpreted as the estimate deduced from a particular APF which uses the instrumental mixture pdf \( q_k^N \) in (39), from which it is always possible to sample from (with an extra computational cost). Thus the estimate deduced from the FA-APF algorithm is used as a benchmark and enables us to analyze the relevance of the instrumental pdf \( q_k^N \) in the APF algorithm.

Next in section IV-C we compute our independent estimates for a target tracking problem with range-bearing measurements. Our estimates are compared to those obtained from the classical SIR algorithm, for a given computational budget measured via the number of sampling operations; this means that we compare \( \hat{\Theta}^{\text{SIR}}_k \) and \( \hat{\Theta}^{\text{SIR}}_k \) (\( M \) is the number of particles after the independent resampling step) to \( \hat{\Theta}^{\text{SIR}}_k \) in which \( N = M^2 + M \). Thus all estimates are based on \( M^2 + M \) sampling operations (we do not distinguish if we sample according to a continuous or a discrete distribution).

The relative performances of the estimates are analyzed in function of the parameters of the state-space model.

Finally in section IV-D we compute our estimates in models where the dimension \( m \) of the hidden state is large and we analyze their performances w.r.t. classical PF estimates in function of the dimension \( m \) and with a fixed number of sampling operations. Finally throughout this section our simulations are averaged over \( P = 1000 \) MC runs, we set \( \varphi(x) = x \) in (2) and we use an averaged Root Mean Square Error (RMSE) criterion, defined as

\[
\text{RMSE}(\Theta_k) = \frac{1}{T} \sum_{k=1}^{T} \left( \frac{1}{P} \sum_{p=1}^{P} \frac{1}{T} \sum_{t=1}^{T} \sqrt{(\Theta_k^{(p)} - x_k^{(p)})^2} \right)^{1/2}
\]

where \( x_k^{(p)} \) is the true state at time \( k \) for the \( p \)-th realization, \( \Theta_k^{(p)} \) is an estimate of \( x_k^{(p)} \) and \( T \) is the time length of the scenario.

A. Comparison of static sampling procedures

Let us first consider the (static) Bayesian estimation problem in which we look for computing

\[
\Theta = E(x|y) = \int xp(x|y)dx
\]
via the techniques described in Section II. We assume that
\( p(x|y) \) is known up to a constant, \( p(x|y) \propto p(x)p(y|x) \) where
\( p(x) = \mathcal{N}(x; 0, \sigma_x^2) \) and \( p(y|x) = \mathcal{N}(y; x, \sigma_y^2) \) with \( \sigma_x^2 = 10 \)
and \( \sigma_y^2 = 3 \). We choose the IS distribution \( q(x) = p(x) \). For a
given number of final samples \( N \), we compute six estimates:
the estimate \( \hat{\Theta}^{\text{SIS},N} \) deduced from the IS mechanism with
importance distribution \( q(\cdot) \); the estimate \( \hat{\Theta}^{\text{SIR},N} \) deduced from
the SIR mechanism with \( N \) intermediate samples and
\( M_N = N \) final samples; our estimate \( \hat{\Theta}^{\text{SIR},N} \) based on \( N \)
unweighted independent samples drawn from \( q^N \) (see (15));
our estimate \( \hat{\Theta}^{\text{SIR}^*,N} \) based on \( N \) weighted independent
samples from \( q^N \) (see (21)). Remember that the computation
of the independent resampling mechanism is based on
the sampling of \( N \) intermediate particles and \( N \) resampling
steps and thus requires an extra computational cost w.r.t. the
dependent one. Consequently, we also compute \( \hat{\Theta}^{\text{SIR}^*,2-N} \)
based on the classical SIR procedure with \( N \) intermediate samples
and \( N \) (dependent) resampling steps; in other words
this estimate relies on \( N \) dependent samples obtained from
\( q^N \).

1) RMSE of estimates: In Fig. 1 we display the distance of
each estimate w.r.t. the true expectation \( E(x|y) \) in function
of the number of samples \( N \) while Table I describes
the performances of each estimate in terms of RMSE w.r.t. the
true \( x \) for small values of \( N \). As expected, the estimate
\( \hat{\Theta}^{\text{SIR},N} \) based on \( N \) independent samples drawn from \( q^N \)
outperforms the estimate \( \hat{\Theta}^{\text{SIS},N} \) which is computed from \( N \)
dependent samples drawn from \( q^N \). However, an interesting
result is that \( \hat{\Theta}^{\text{SIR},N} \) also outperforms \( \hat{\Theta}^{\text{SIR}^*,2,N} \). It means
that the distribution \( q^N \) produced by the SIR mechanism
is more adapted than the prior \( q(x) = p(x) \), which is not
surprising since \( q^N \) implicitly uses the observation \( y \)
through the resampling mechanism of intermediate samples. Of course,
the computation of \( \hat{\Theta}^{\text{SIR},N} \) requires an extra computational cost
but it is interesting to note that the size of the final
support is the same in the three cases. We then compare
the estimates based on the same computational cost. When
\( N \) increases, these estimates have the same asymptotical
behavior. It can be seen that the estimate \( \hat{\Theta}^{\text{SIR}^*,2-N} \) based on
\( N \) samples drawn from \( q^N \) outperforms \( \hat{\Theta}^{\text{SIR}^*,N} \) (and other
simulations as well as Table I show that the difference between
the two increases when \( N \) gets smaller, more specifically
\( N < 100 \)). However, when our i.i.d. samples are weighted
by a term proportional to \( p(x,y)/q^N(x) \) in an IS perspective,
our estimate \( \hat{\Theta}^{\text{SIR}^*,N} \) slightly outperforms \( \hat{\Theta}^{\text{SIR},N} \) for
small \( N \) and has the same performances when \( N \to \infty \), which
confirms the discussion in II-B2. However, remember that
the Independent SIR procedure can easily be parallelized, and that
\( \hat{\Theta}^{\text{SIR}^*,N} \) relies on the approximation (42) of (31) which
recycles the intermediate samples \( x_k^{(j)} \). To ensure the quality
of this approximation, we have estimated (31) with \( M \) additional
sets of \( N \) particles, \( M > N \), rather than only recycling the
\( N \) available sets of \( N \) particles; the performances obtained
for \( \hat{\Theta}^{\text{SIR}^*,N} \) using these additional sets of particles are identical,
which tends to show that recycling the particles \( x_k^{(j)} \) is indeed sufficient to approximate (31). It also been observed
that \( \hat{\Theta}^{\text{SIR}^*,N} \) outperforms other estimates when \( \sigma_y \to 0 \), i.e.
when the observations become informative. We do not display
results for such setting because this point will be discussed in
the dynamic case.

2) Bias and variance of estimates: Next we study the
behaviour of the empirical variance and bias of these estimates.
Contrary to the previous simulation in this section we have made sure to fix the observation \( y \) for all MC
runs so as to display the empirical variance corresponding
to the same variance terms \( \text{var}(\hat{\Theta})^{\text{SIR},N} \) and \( \text{var}(\hat{\Theta})^{\text{SIS},N} \)
mentioned throughout this paper (such as in equation (17)),
which are conditional on the observations; The results are not
displayed for space reasons but the order of performance of
the estimators in terms of variance or bias remains the same
as that in terms of RMSE; the variances of both \( \hat{\Theta}^{\text{SIR},N} \) and
\( \hat{\Theta}^{\text{SIS},N} \) become much closer to each other as well as to that
of \( \hat{\Theta}^{\text{SIR}^*,N} \), compared to when we consider the RMSE
averaged over different observations.

3) Estimation of the predictive likelihood: We finally
consider the problem of estimating the predictive likelihood
\( p(y_k|y_{0:k-1}) \) which has a critical role for parameter estimation
or smoothing problems [37]. Let us consider two ways of approximating
\( p(y_k|y_{0:k-1}) \), based on the observation that it can be
deduced from the filtering pdf, since
\( p(y_k|y_{0:k-1}) = \int q(y_k|x_k)f(x_k|x_{k-1})p(x_{0:k-1}|y_{0:k-1})dx_{0:k} \). The first one
[31] is an unbiased estimate based on the \( N^2 \) intermediate
samples \( \tilde{x}_k^{(j)} \sim q(x_k|x_{k-1}) \):
\[
\hat{p}(y_k|y_{0:k-1}) = \frac{1}{N} \sum_{j=1}^{N} \frac{1}{N} \sum_{i=1}^{N} g_k(y_k|x_k^{(j)}) f_k(x_k^{(j)}|x_{k-1}^{(i)}) q(x_k^{(j)}|x_{k-1}^{(i)}).
\]
(45)
The second one uses the APF interpretation of our algorithm;
from our weighted samples, one can build the following
estimator of the predictive likelihood:
\[
\hat{p}(y_k|y_{0:k-1}) = \sum_{i=1}^{N} w_{k-1} g_k(y_k|x_k^{(i)}) f_k(x_k^{(i)}|x_{k-1}^{(i)}) h_k^{(i)} q(x_k^{(i)}|x_{k-1}^{(i)}),
\]
(46)
where the \( h_k^{(i)} \) have already been estimated by recycling the
\( N^2 \) extra particles.

Turning back to the static linear and Gaussian model, we
have estimated \( p(y) \) from (45) (which is also the estimate
deduced from IS with \( N^2 \) samples) and (46) as well as the
estimate deduced from IS with \( N \) samples (see [22]). It can
be observed from Fig. 2 that estimates (46) and (45) both
converge to the true value much quicker and more smoothly
than the IS one as \( N \) increases, and that (46) is slightly more
stable than (45), as long as the number of particles is not
extremely low (below 10 particles in this simulation).

<table>
<thead>
<tr>
<th>( N )</th>
<th>( \text{gIS}^{\text{SIS},N} )</th>
<th>( \text{gIS}^{\text{SIR}^*,N} )</th>
<th>( \text{gIS}^{\text{SIR},N} )</th>
<th>( \text{gIS}^{\text{SIR}^*,2-N} )</th>
<th>( \text{gIS}^{\text{SIR}^*,N} )</th>
<th>( \text{gIS}^{\text{SIR}^*,2-N} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>1.6844</td>
<td>1.6819</td>
<td>1.6524</td>
<td>1.5951</td>
<td>1.5618</td>
<td>1.5610</td>
</tr>
<tr>
<td>40</td>
<td>1.5925</td>
<td>1.5981</td>
<td>1.5763</td>
<td>1.5606</td>
<td>1.5446</td>
<td>1.5410</td>
</tr>
<tr>
<td>60</td>
<td>1.5792</td>
<td>1.5777</td>
<td>1.5637</td>
<td>1.5442</td>
<td>1.5385</td>
<td>1.5335</td>
</tr>
<tr>
<td>80</td>
<td>1.5623</td>
<td>1.5649</td>
<td>1.5530</td>
<td>1.5348</td>
<td>1.5309</td>
<td>1.5293</td>
</tr>
<tr>
<td>100</td>
<td>1.5519</td>
<td>1.5504</td>
<td>1.5410</td>
<td>1.5320</td>
<td>1.5290</td>
<td>1.5290</td>
</tr>
</tbody>
</table>

**TABLE I**

**STATIC LINEAR AND GAUSSIAN MODEL - RMSE VALUES OF EACH ESTIMATE.**
Other simulations as well as Table I show that $N$, meaning that $p$ for large $\tilde{q}$ consequently it is possible to obtain i.i.d. samples from the ARCH model which is a particular HMC model (1)

$E(\text{independent resampling Algorithm 4 in which the importance distribution is deduced from the APF with } \mu(x_{0:k-1}) \propto w_k^{-1} p(y_k|x_{k-1})$ and $\tau(x_k|x_{0:k-1}) = f_k(x_k|x_{k-1})$; with this configuration, the particles are pre-selected with the so-called optimal first stage weight and sampled from the transition pdf.

The RMSE of each estimate is displayed in Fig. 3(a) as a function of the number of samples $N$. Interestingly enough, our weighted independent resampling algorithm which produces $\tilde{q}_k^{\text{i-SIR-wN}}$ has the same performance as the FA-APF algorithm when $N \geq 20$, without using the predictive likelihood $p(y_k|x_{k-1})$ nor the optimal importance distribution $p(x_k|x_{k-1}, y_k)$. It means that the mixture pdf $q_k^{\text{i-SIR}}$ which has been implicitly built by our algorithm differs from the target mixture (36) (because there is a significant difference between using uniform and non uniform weights) but is as relevant as this target mixture (36); indeed, one advantage of the mixture pdf $q_k^{\text{i-SIR}}$ deduced from the resampling mechanism is that its interpretation does not depend on the importance distribution which has been chosen and that it is possible to sample from it in general HMC models (1). We also observe that re-weighting the final samples is beneficial w.r.t. attributing uniform weights. In order to analyze the behavior of the weights associated to our estimate $\tilde{q}_k^{\text{i-SIR-wN}}$, we compute the normalized effective sample size defined as $N_{\text{norm,eff}} = N \sum_{x \in \omega} w_x^2$. It can be observed that $N_{\text{norm,eff}}$ quickly converges from 0.9 to nearly 1 as $N$ increases from 5 to 100 (reaching $N_{\text{norm,eff}} = 0.99$ for $N = 30$), meaning that these weights tend to become uniform, so estimates $\tilde{q}_k^{\text{i-SIR-wN}}$ and $\tilde{q}_k^{\text{i-SIR-wN}}$ become close when $N$ is sufficiently large. Additionally, we show in Fig. 3(b) the RMSE of the various estimators w.r.t. time. It can be observed that $\tilde{q}_k^{\text{i-SIR-wN}}$ (which is deduced from samples drawn from $f_k(x_k|x_{k-1})$) and $\tilde{q}_k^{\text{FA-wN}}$ have the same performances whatever time $k$.

B. Comparison with APF algorithms

We now focus on the comparison of our independent resampling algorithm in terms of APF. We study the ARCH model which is a particular HMC model (1) in which $f_k(x_k|x_{k-1}) = N(x_k; 0; \beta_0 x_{k-1}^2 + \beta_1)$ and $g_k(y_k|x_k) = N(y_k; x_k; R)$. We set $R = 1$, $\beta_0 = 3$ and $\beta_1 = 0.75$. In this model one can compute $p(y_k|x_{k-1}) = N(y_k; 0; \beta_0 R x_{k-1}^2 + \beta_1 x_{k-1}^2)$ and $p(x_k|x_{k-1}, y_k) = N(x_k; \beta_0 R x_{k-1}^2 + \beta_1 x_{k-1}^2, \beta_0 R x_{k-1}^2 + \beta_1 x_{k-1}^2)$; consequently, it is possible to obtain i.i.d. samples from the target mixture (36) and thus compute the estimate $\tilde{q}_k^{\text{i-SIR-wN}}$ based on the FA-APF algorithm. In such a setting, the M-APF described in III-C1 coincides with the FA-APF. Remember that the FA-APF can also be seen as a particular case of our independent resampling Algorithm 4 in which the importance distribution $q(x_k|x_{0:k-1})$ coincides with $p(x_k|x_{k-1}, y_k)$ (see section III-C2). However this setting can be implemented in specific models only, while Algorithm 4 can be used with any importance distribution $q(x_k|x_{0:k-1})$, while keeping the same interpretation as the FA-APF (see our discussion in section III-C2). So we also compute our estimates $\tilde{q}_k^{\text{i-SIR-wN}}$ and $\tilde{q}_k^{\text{i-SIR-wN}}$, which can be seen as an estimate deduced from the APF in which the importance mixture (37) coincides with $q_k^{\text{i-SIR}}$. We finally compute the estimate $\tilde{q}_k^{\text{i-SIR-wN}}$ which is deduced from the APF with $\mu(x_{0:k-1}) \propto w_k^{-1} p(y_k|x_{k-1})$ and $\tau(x_k|x_{0:k-1}) = f_k(x_k|x_{k-1})$; with this configuration, the particles are pre-selected with the so-called optimal first stage weight and sampled from the transition pdf.

C. Tracking from range-bearing measurements

We now study the performance of our algorithms in a tracking scenario with range-bearing measurements. We look for estimating the state vector $x_k = [p_{x,k}, \dot{p}_{x,k}, p_{y,k}, \dot{p}_{y,k}]^T$ (position and velocity in Cartesian coordinates) of a target from noisy range-bearing measurements $y_k$. The pdfs in model (1) associated with this tracking problem are $f_k(x_k|x_{k-1}) = \mathcal{N}(x_k; F x_{k-1}; Q)$ and $g_k(y_k|x_k) = \mathcal{N}(y_k; \sqrt{\sigma_p^2 + \sigma_q^2 \frac{1}{\arctan\left(\frac{\sigma_q}{\sigma_p}\right)}}; R)$. where $R = \begin{pmatrix} \sigma_x^2 & 0 \\ 0 & \sigma_y^2 \end{pmatrix}$, $F = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$, $Q = \sigma^2 \begin{pmatrix} 0 & 1/2 \\ 1/2 & 0 \\ 0 & 1/2 \\ 1/2 & 0 \end{pmatrix}$ (47)
case where \( \sigma_Q = \sqrt{10} \), \( \sigma_\rho = 0.05 \) and \( \sigma_\theta = \frac{\pi}{720} \). For the first configuration, we observe that \( \widehat{\Theta}^{1-\text{SIR}-\text{w},M}_k \) outperforms the other estimates and improves \( \Theta^{1-\text{SIR},M}_k \), which does not rely on weighted samples, for any \( M \). Compared to the classical SIS estimate, \( \Theta^{1-\text{SIR},M}_k \) also gives better performance as long as the number of samples \( M \) is weak (\( M < 30 \), so \( N < 465 \)) but is next outperformed. As shown in Fig. 4(b), when the observations become informative, \( \Theta^{1-\text{SIR},M}_k \) gives the best performances. Contrary to \( \Theta^{1-\text{SIR},N}_k \), \( \Theta^{1-\text{PF},N}_k \) and \( \Theta^{1-\text{SIR},N}_k \), our estimate does not suffer from the degeneration of the importance weights. Indeed when the measurements are informative (and so the likelihood is sharp), few importance weights have a non null value and even more advanced resampling schemes do not achieve to create a diversified final support. By contrast, the independent resampling procedure ensures the diversity of the final samples when we use uniform weights. Finally remember that \( \Theta^{1-\text{SIR}-\text{w},M}_k \) relies on the MC approximation (42). If we study the approximation \( \hat{H}_k \) in (42), any term \( \hat{q}_k^e(x_k|\tilde{x}_{k-1}) \) with \( q_k(x) = f_k(x_k|x_{k-1}) \) can be considered negligible when compared to any other such term because the model is very informative. Consequently, the final unnormalized approximated second stage weights tend to become equal to the intermediary unnormalized weights \( p_k^e(\cdot) \) of the corresponding particles and the estimate \( \Theta^{1-\text{SIR}-\text{w},M}_k \) is affected by the lack of diversity.

For reference, the averaged computational times for a single time iteration of the algorithms (without parallelization) when \( N = 50 \) are as follows (this experiment used Matlab software and a laptop featuring a 2.80 GHz Intel(R) Core(TM) i7-4810MQ CPU): \( t_{\text{SIR}} \approx 0.0010s, t_{\text{PF}} \approx 0.0014s, t_{1-\text{SIR}} \approx 0.0011s, t_{1-\text{SIR}-\text{w}} \approx 0.0020s \).

### D. High dimensional problems

We finally study the impact of the dimension of the hidden state \( x_k \). We consider a state vector of dimension \( m = 4 \times l \), \( x_k = [p_{x,k}(1), p_{x,k}(1), p_{y,k}(1), \ldots, p_{x,l}(1), p_{y,l}(1), p_{y,l}(1)]^T \). Each component \( x_{k,l} = [p_{x,l}(1), p_{x,l}(1), p_{y,l}(1), p_{y,l}(1)]^T \) evolves independently from all the other components, according to \( f_k(x_k(l)|x_{k-1}(l)) = N(x_{k,l}; Fx_{k-1,l}; Q) \) where \( F \) and \( Q \) given by (47), and is observed independently via \( g_k(y_k|x_{k,l}) = N(y_k; Hx_k(l); R) \) where

\[
H = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix},
R = \begin{pmatrix}
\sigma^2_\rho & 0 & 0 & 0 \\
0 & \sigma^2_\rho & 0 & 0 \\
0 & 0 & \sigma^2_\rho & 0 \\
0 & 0 & 0 & \sigma^2_\rho
\end{pmatrix}.
\]

Again, we compute the estimate based on classical PF \( \Theta^{\text{SIS},N}_k \) (see (26)). It is well known that the PF tends to degenerate when the dimension of the hidden state increases. We also compute \( \Theta^{1-\text{SIR},M}_k \) (see (32)), \( \Theta^{1-\text{SIR}-\text{w},M}_k \) (see (41)) with \( N = \frac{M^2}{2} \) for \( M = 100 \) and \( M = 1000 \) as a function of dimension \( m \) to see how the dimension impacts our estimate and the classical PF estimate. Note that we use the basic importance distribution \( f_k(x_k|x_{k-1}) \) in order to measure the impact of our method but the improvements developed for high dimensional models (such as e.g. [25]) can also be included with the independent SIR algorithm.
The results are displayed in Fig. 5. It can be seen that the estimates $\hat{\Theta}_k^{1-\text{SIR},\tilde{M}}$ and $\hat{\Theta}_k^{1-\text{SIR},w,\tilde{M}}$ outperform $\hat{\Theta}_k^{\text{SIS},M}$ more and more significantly as the dimension increases, due to the local impoverishment phenomenon. First, $\hat{\Theta}_k^{1-\text{SIR},w,\tilde{M}}$ outperforms $\hat{\Theta}_k^{1-\text{SIR},M}$ as long as the dimension of the hidden state is low ($m = 4$ and $m = 8$); when $m$ increases, the estimate based on weighted samples from $q^N$ limits the degeneration phenomenon w.r.t. that based on weighted samples from $q$ but using unweighted samples when the dimension is large ensures the diversity and gives better performance. Note that the dependent and independent SIR algorithms approximately give the same performance when $m$ is low but the gap between the dependent and the independent SIR estimates increases with the dimension. More precisely, this gap appears and is significantly amplified when $m$ increases from 8 to 20; for higher values of $m$, it then increases linearly with $m$ which is solely due to the RMSE adding up an increasing number of factors which always present the same gap of performance per-dimension as when $m = 20$ (which means that around $m = 20$, the dependent SIR is already completely degenerated, with only one or very few non-null weights). The same performance gap can be observed between the unweighted and weighted independent SIR estimates, except that this time for $m < 12$ the weighted estimate actually performs slightly better than the unweighted one, and the subsequent gap for $m > 20$ (which also increases linearly in $m$) is smaller.

**V. CONCLUSION**

SMC algorithms in Hidden Markov models are based on the sequential application of the IS principle. However the direct sequential application of the IS principle leads to the degeneration of the weights, against which multinomial resampling has been proposed. This rejuvenation scheme, which is now routinely used in SIR algorithms, enables to discard particles (or trajectories) with low weights, but particles with large weights will be resampled several times, which leads to dependency and support degeneracy. In this paper we thus revisited the resampling step used in classical SIR algorithms. We first addressed the static case, showed that the particles sampled by Rubin’s SIR mechanism are dependent samples drawn from some pdf $q^N$, and proposed an alternative sampling mechanism which produces independent particles drawn from that same marginal pdf $q^N$. This set of independent samples enables us to build a moment estimator which outperforms the classical SIR-based one, both from a theoretical point of view and from a practical point of view.
We next adapted this methodology to the dynamic case, in order to estimate a moment of interest in an HMC model. The computation of the post-resampling weights is more challenging than in the static case, but reinterpreting our independent resampling scheme as the first step of a particular APF algorithm enables us to make full use of the APF methodology and thus reweight the final samples via the second-stage APF weights. Finally we validated our discussions through computer-generated experiments and carefully took into account the computational budget. Simulations in a model where the FA-APF algorithm is computable show that the independent resampling with reweighting algorithm gives a performance close to the FA-APF algorithm. Consequently, it confirms the relevance of the instrumental mixture pdf implicitly used by the independent resampling PF which can be used in any HMC model since it does not require to compute the predictable likelihood nor the optimal importance distribution. Finally independent PF gives very satisfying results when applied in highly informative models which are challenging for classical PF and limits the degeneration phenomenon in high dimensional models.

APPENDIX

PROOF OF Proposition 1
Let $A$ be any Borel set. Let $I_A(x) = 1$ if $x \in A$ and 0 otherwise. Then for any $l, 1 \leq l \leq M_N$, 

$$
Pr(x^i \in A) = \int_{\mathbb{R}^N} [\int_{A} w^i(\tilde{x}^i, \ldots, \tilde{x}^N)] I_A(\tilde{x}^i) \prod_{j=1}^{N} q(\tilde{x}^j) d\tilde{x}^{1:N}
$$

$$
= \sum_{i=1}^{N} \int_{\mathbb{R}^N} w^i(\tilde{x}^i, \ldots, \tilde{x}^N) I_A(\tilde{x}^i) \prod_{j=1}^{N} q(\tilde{x}^j) d\tilde{x}^{1:N}
$$

$$
= \sum_{i=1}^{N} \int_{A} \int_{\mathbb{R}^N} \prod_{j=1}^{N} q(\tilde{x}^j) d\tilde{x}^{1:N} \cdot \sum_{i=1}^{N} q(\tilde{x}^i) d\tilde{x}^i
$$

so $x^i$ has pdf $\tilde{q}_N$ w.r.t. Lebesgue measure.

PROOF OF Proposition 2
Let $x^i$ (for any $i, 1 \leq i \leq M_N$) be produced by the classical SIR mechanism. Then

$$
E(\phi(x^i)) = \hat{\Theta}^{IS,N}.
$$

(48)

So $E(\Theta^{IS,M_N}(x^i)|x^{1:N}) = \hat{\Theta}^{IS,N}$ and $E(\Theta^{IS,M_N}) = E(\hat{\Theta}^{IS,N})$. On the other hand $E(\Theta^{IS-1,SM,M_N}) = E(\hat{\Theta}^{IS,M_N})$, whence (16). Next

$$
\text{var}(\Theta^{SM,M_N}) = \frac{1}{M_N^2} \sum_{i=1}^{M_N} \text{var}(\phi(x^i)) + \frac{1}{M_N^2} \sum_{k,l=1}^{M_N} \text{Cov}(\phi(x^k), \phi(x^l)).
$$

(49)

in which $x^i \sim \tilde{q}^N$ for all $i$. The first term is equal to $\text{var}(\Theta^{IS-1,SM,M_N})$. Let us compute the second term. For all $k, l, 1 \leq k, l \leq M_N$ with $k \neq l$,

$$
E(\phi(x^k)\phi(x^l)|x^{1:N}) = (\hat{\Theta}^{IS,N})^2,
$$

so $E(\phi(x^k)\phi(x^l)) = E(E(\phi(x^k)\phi(x^l)|x^{1:N})) = E((\hat{\Theta}^{IS,N})^2)$. Using (48) again, we conclude that $Cov(\phi(x^k), \phi(x^l)) = \text{var}(\hat{\Theta}^{IS,N})$, whence (17).

PROOF OF THEOREM 1
We first introduce the following notations:

$$
\Theta(\phi) = \int \phi(x)p(x)dx,
$$

(50)

$$
\hat{\Theta}^{IS,N}(\phi) = \sum_{i=1}^{N} \frac{p(x^i)}{q(x^i)} \phi(x^i),
$$

(51)

$$
\hat{\Theta}^{IS-1,SM,M_N}(\phi) = \frac{1}{M_N} \sum_{i=1}^{M_N} \phi(x^i),
$$

(52)

Finally independent PF gives very satisfying results when applied in highly informative models which are challenging for classical PF and limits the degeneration phenomenon in high dimensional models.

Using $E(\hat{\Theta}^{IS-1,SM,M_N}(\phi)) = E(\hat{\Theta}^{IS,N}(\phi))$, we have

$$
\sqrt{M_N} (\hat{\Theta}^{IS-1,SM,M_N}(\phi) - \Theta(\phi)) \overset{D}{\to} N \left(0, E_\phi \left(\frac{p^2(x)}{q^2(x)} (\phi(x) - \Theta(\phi))^2\right)\right).
$$

(53)

According to [17, Theorem 9.11.0], $E((\sqrt{N}(\hat{\Theta}^{IS,N}(\phi) - \Theta(\phi)))^2)$ is bounded and so its upper bound is finite. According to [38, corollary of Theorem 25.12], it is ensured that

$$
\sqrt{N}E((\hat{\Theta}^{IS,N}(\phi) - \Theta(\phi))) \to 0;
$$

(54)

consequently

$$
\sqrt{M_N} \sqrt{\frac{1}{N}} E(\sqrt{N}(\hat{\Theta}^{IS,N}(\phi) - \Theta(\phi))) \to 0.
$$

(55)

Our objective is to show that $A_N$ converges to a centered Gaussian distribution with variance $\text{var}_\phi(\phi(x))$ and that $B_N$ converges to 0.

Convergence of $B_N$
We have recalled (see (10)) that under mild assumptions [5]

$$
\sqrt{N}(\hat{\Theta}^{IS,N}(\phi) - \Theta(\phi)) \overset{D}{\to} N \left(0, E_\phi \left(\frac{p^2(x)}{q^2(x)} (\phi(x) - \Theta(\phi))^2\right)\right).
$$

(56)

According to [17, Theorem 9.11.0], $E((\sqrt{N}(\hat{\Theta}^{IS,N}(\phi) - \Theta(\phi)))^2)$ is bounded and so its upper bound is finite. According to [38, corollary of Theorem 25.12], it is ensured that

$$
\sqrt{N}E((\hat{\Theta}^{IS,N}(\phi) - \Theta(\phi))) \to 0;
$$

(57)

Convergence of $A_N$

$A_N$ reads

$$
\sqrt{M_N} \left(\frac{1}{M_N} \sum_{i=1}^{M_N} \phi(x^i) - E(\phi(x^i))\right).
$$

To prove the convergence when $N \to \infty$, we need a CLT for triangular arrays and we use the version presented in [17, Theorem 9.15.13]. The required assumptions are:

1) $\{x^i\}_{i=1}^{M_N}$ are independent;

2) $\frac{1}{M_N} \sum_{i=1}^{M_N} E(\phi(x^i)^2) - (E(\phi(x^i)))^2 \to \text{var}_\phi(\phi(x));$

3) for any positive $C$, $\frac{1}{M_N} \sum_{i=1}^{M_N} E(\phi(x^i)^2 I_{|\phi(x^i)| \geq C}) \to \Theta(\phi^2 I_{|\phi| \geq C}).$
Assumption 1) is satisfied since \( (\varphi_i)_{i=1}^{M_N} \) are i.i.d. from \( \bar{q}_N \). Next, \( E(\varphi^2(x)) = E_{\bar{q}_N}(\varphi^2(x)) \) which coincides with \( E(\Theta_{N,i}^S,N(\varphi)) \). Using again [17, Theorem 9.1.10] and [38, Theorem 25.12], \( E(\Theta_{N,i}^S,N(\varphi)) \rightarrow \Theta(\varphi) \) when \( N \rightarrow \infty \).

Finally, \( E(\varphi^2(x)) = E_{\bar{q}_N}(\varphi^2(x)) \rightarrow \Theta(\varphi^2) = \varphi(x) \).

Consequently, assumption 2) is satisfied since

\[
\frac{1}{M_N} \sum_{i=1}^{M_N} E(\varphi^2(x)) - E(\varphi^2(x)) \rightarrow 0.
\]

Finally, \( E(\varphi^2(x)) \big| x \geq C = E(\Theta_{N,i}^S,N(\varphi^21_{x \geq C})) \) which converges to \( \Theta(\varphi^21_{x \geq C}) \) and assumption 3) is satisfied. Consequently,

\[
\sqrt{M_N} \left( \frac{1}{M_N} \sum_{i=1}^{M_N} \varphi(x) - E(\varphi^2(x)) \right) \xrightarrow{D} N(0, \varphi(x)).
\]  

Combining (56), (58) and (53) we obtain (18).

REFERENCES


