WEIGHTING A RESAMPLED PARTICLES IN SEQUENTIAL MONTE CARLO (EXTENDED PREPRINT)

L. Martino[◇], *V. Elvira*[†], *F. Louzada*[◇]

[†] Dep. of Signal Theory and Communic., Universidad Carlos III de Madrid, Leganés (Spain).
[◇] Institute of Mathematical Sciences and Computing, Universidade de São Paulo, São Carlos (Brazil).

ABSTRACT

The Sequential Importance Resampling (SIR) method is the core of the Sequential Monte Carlo (SMC) algorithms (a.k.a., particle filters). In this work, we point out a suitable choice for weighting properly a resampled particle. This observation entails several theoretical and practical consequences, allowing also the design of novel sampling schemes. Specifically, we describe one theoretical result about the sequential estimation of the marginal likelihood. Moreover, we suggest a novel resampling procedure for SMC algorithms called partial resampling, involving only a subset of the current cloud of particles. Clearly, this scheme attenuates the additional variance in the Monte Carlo estimators generated by the use of the resampling.

Index Terms— Importance Sampling; Sequential Importance Resampling; Sequential Monte Carlo; Particle Filtering.

1. INTRODUCTION

Sequential Monte Carlo (SMC) methods have become essential tools for Bayesian analysis in statistical signal processing [2, 8, 9, 12, 18]. SMC algorithms (a.k.a., particle filters) are based on the importance sampling technique [5, 7, 6, 11, 16, 22] and its sequential version known as Sequential Importance Sampling (SIS) [10, 13]. Another essential piece of SMC is the application of resampling procedures [3, 10]. The combination of SIS and resampling is often referred as Sequential Importance Resampling (SIR).

Since the unnormalized importance weight of a resampled particle cannot be computed analytically using the standard IS weight definition, in the classical SIR formulation, the users consider only the estimators involving normalized weights. The concept of the unnormalized weight of a resampled particle is usually not considered, i.e., its computation is avoided and omitted [8, 9, 10].

In this work, we introduce a proper unnormalized importance weight for a particle resampled from a set of weighted samples, defined as the arithmetic mean of the importance weights of these samples. This weight choice is proper according to the Liu's definition [13, Section 2.5.4] since it provides unbiased IS estimators, as shown in this work. The introduction of this unnormalized proper weight for a resampled particle entails several interesting consequences from a practical and theoretical point of view. For instance, this weight definition has been already applied implicitly or heuristically in different works: in parallel particle filters [4, 19, 20] and parallel SMC schemes, e.g., *the island particle- double bootstrap* method [25, 26], or "unawarely" in certain classes of MCMC algorithms [1, 14] (where one particle is chosen among a set of candidates via resampling, before be testing as possible future state of the chain), as we can infer from the discussion in [17]. Similarly approaches have been implicitly used in the so-called α*-SMC* [27] and *Nested-SMC* methods [21].

Here, we also describe two additional consequences. First, we highlight that all the estimators derived in the SIS approach can also be employed in SIR using the weight definition of a resampled particle introduced here. We show it considering the estimation of the marginal likelihood (a.k.a., Bayesian evidence or partition function) [10, 18, 22]. In SIS, there are two possible estimators of the marginal likelihood which are completely equivalent [18]. Using the proper unnormalized weight for a resampled particle, we show that we can employ two equivalent estimators of the marginal likelihood also in SIR. They coincide with the estimators in SIS as special case, when no resampling is applied. Furthermore, we describe an alternative resampling procedure for particle filtering algorithms, called *partial resampling*, involving only a subset of the

This is an extension of the work [15]. This work has been supported by the ERC grant 239784 and AoF grant 251170, the Spanish government through the OTOSiS (TEC2013-41718-R), by the Grant 2014/23160-6 of the São Paulo Research Foundation (FAPESP) and by the Grant 305361/2013-3 of the National Council for Scientific and Technological Development (CNPq).

current population of particles. This scheme attenuates the loss of diversity in the population and the additional variance in the Monte Carlo estimators generated due to the application of the resampling steps.

2. IMPORTANCE SAMPLING

Let us denote the target probability density function (pdf) as $\bar{\pi}(\mathbf{x}) = \frac{1}{Z}\pi(\mathbf{x})$ (known up to a normalizing constant) with

$$\mathbf{x} = x_{1:D} = [x_1, x_2, \dots, x_D] \in \mathcal{X} \subseteq \mathbb{R}^{D\times\eta},$$

where $x_d \in \mathbb{R}^\eta$ for all $d = 1, \dots, D$. We consider the Monte Carlo approximation of complicated integrals involving the target $\bar{\pi}(\mathbf{x})$ and a square-integrable function $h(\mathbf{x})$, e.g.,

$$I = E_{\bar{\pi}}[h(\mathbf{X})] = \int_{\mathcal{X}} h(\mathbf{x})\bar{\pi}(\mathbf{x})d\mathbf{x}, \tag{1}$$

where $\mathbf{X} \sim \bar{\pi}(\mathbf{x})$. In general, generating samples directly from the target $\bar{\pi}(\mathbf{x})$ is impossible. Thus, one usually considers a (simpler) proposal pdf, $q(\mathbf{x})$. The expression below

$$\begin{aligned} E_{\bar{\pi}}[h(\mathbf{X})] &= E_q[h(\mathbf{X})w(\mathbf{X})], \\ &= \frac{1}{Z}\int_{\mathcal{X}} h(\mathbf{x})\frac{\pi(\mathbf{x})}{q(\mathbf{x})}q(\mathbf{x})d\mathbf{x}, \end{aligned} \tag{2}$$

where $w(\mathbf{x}) = \frac{\pi(\mathbf{x})}{q(\mathbf{x})} : \mathcal{X} \to \mathbb{R}$, suggests an alternative procedure. Indeed, we can draw N samples (also called *particles*) $\mathbf{x}_1, \dots, \mathbf{x}_N$ from $q(\mathbf{x})$,[1] and then assign to each sample the following unnormalized weights

$$w(\mathbf{x}_n) = \frac{\pi(\mathbf{x}_n)}{q(\mathbf{x}_n)}, \quad n = 1, \dots, N. \tag{3}$$

If the target function $\pi(\mathbf{x})$ is normalized, i.e., $Z = 1$, $\bar{\pi}(\mathbf{x}) = \pi(\mathbf{x})$, a natural (unbiased) IS estimator [13, 22] is defined as

$$\widehat{I}_N = \frac{1}{N}\sum_{n=1}^{N} w(\mathbf{x}_n)h(\mathbf{x}_n), \quad \widehat{I}_N \xrightarrow[N\to\infty]{P} I, \tag{4}$$

where $\mathbf{x}_n \sim q(\mathbf{x})$, $n = 1, \dots, N$. If the normalizing constant Z is unknown, defining the normalized weights,

$$\bar{w}(\mathbf{x}_n) = \frac{w(\mathbf{x}_n)}{\sum_{i=1}^{N} w(\mathbf{x}_i)}, \quad n = 1, \dots, N, \tag{5}$$

an alternative self-normalized (biased) IS estimator [13, 22] is

$$\overline{I}_N = \sum_{n=1}^{N} \bar{w}(\mathbf{x}_n)h(\mathbf{x}_n), \quad \overline{I}_N \xrightarrow[N\to\infty]{P} I. \tag{6}$$

Moreover, an unbiased estimator of marginal likelihood, $Z = \int_{\mathcal{X}} \pi(\mathbf{x})d\mathbf{x}$, is given by

$$\widehat{Z} = \frac{1}{N}\sum_{i=1}^{N} w(\mathbf{x}_i), \quad \widehat{Z} \xrightarrow[N\to\infty]{P} Z, \tag{7}$$

where we have avoided the subindex N, in order to simplify the notation in the rest of the work.

[1] We assume that $q(\mathbf{x}) > 0$ for all $\mathbf{x}$ where $\bar{\pi}(\mathbf{x}) \neq 0$, and $q(\mathbf{x})$ has heavier tails than $\bar{\pi}(\mathbf{x})$.

2.1. Concept of proper weighted sample

Although the weights of Eq. (3) are broadly used in the literature, the concept of a *properly weighted sample*, suggested in [22, Section 14.2] and in [13, Section 2.5.4], can be used to construct more general weights. More specifically, following the definition in [13, Section 2.5.4], a set of weighted samples is considered *proper* with respect to the target π if, for any square integrable function h,

$$E_q[w(\mathbf{x}_n)h(\mathbf{x}_n)] = cE_{\bar{\pi}}[h(\mathbf{x}_n)], \quad \forall n = \{1, \ldots, N\}, \tag{8}$$

where c is a constant value, also independent from the index n, and the expectation of the left hand side is performed, in general, w.r.t. to the joint pdf of $w(\mathbf{x})$ and $\mathbf{x}$, i.e., $q(w, \mathbf{x})$. Namely, the weight $w(\mathbf{x})$, (for a given value of $\mathbf{x}$), could even be considered a random variable.

3. IMPORTANCE WEIGHT OF A RESAMPLED PARTICLE

Let us consider the following multinomial resampling procedure [3, 8, 9]:

1. Draw N particles $\mathbf{x}_n \sim q(\mathbf{x})$ and weight them with $w(\mathbf{x}_n) = \frac{\pi(\mathbf{x}_n)}{q(\mathbf{x}_n)}$, with $n = 1, \ldots, N$.
2. Draw one particle $\widetilde{\mathbf{x}}' \in \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ from the discrete probability mass

$$\widehat{\pi}(\mathbf{x}|\mathbf{x}_{1:N}) = \sum_{n=1}^{N} \bar{w}(\mathbf{x}_n)\delta(\mathbf{x} - \mathbf{x}_n), \tag{9}$$

   where $\bar{w}(\mathbf{x}_n) = \frac{w(\mathbf{x}_n)}{\sum_{i=1}^{N} w(\mathbf{x}_i)}$.

Question 1. What is the distribution of the resampled particle $\widetilde{\mathbf{x}}'$ (not conditioned to $\mathbf{x}_{1:N}$)? We can easily write its corresponding density as

$$\widetilde{q}(\mathbf{x}) = \int_{\mathcal{X}^N} \left[\prod_{i=1}^{N} q(\mathbf{x}_i)\right] \widehat{\pi}(\mathbf{x}|\mathbf{x}_{1:N})d\mathbf{x}_{1:N}. \tag{10}$$

where $\widehat{\pi}$ is given in Eq. (9). However, the integral above cannot be computed analytically.

Question 2. Can we obtain a proper importance weight associated to the resampled particle $\widetilde{\mathbf{x}}$? As a consequence of the previous observations, we are not able to evaluate the corresponding standard importance weight, $w(\widetilde{\mathbf{x}}) = \frac{\pi(\widetilde{\mathbf{x}})}{\widetilde{q}(\widetilde{\mathbf{x}})}$.

For solving this issue, let us consider N resampled particles $\widetilde{\mathbf{x}}_1, \ldots, \widetilde{\mathbf{x}}_N$ independently obtained by the resampling procedure above. In SMC and adaptive IS applications, [5, 7, 8, 9], the *unnormalized* importance weights of $\widetilde{\mathbf{x}}_1, \ldots, \widetilde{\mathbf{x}}_N$ are not usually needed, but only the *normalized* ones. Thus, a well known proper strategy [8, 9, 10] in this case is to consider

$$w(\widetilde{\mathbf{x}}_1) = w(\widetilde{\mathbf{x}}_2) = \ldots = w(\widetilde{\mathbf{x}}_N), \tag{11}$$

and, as a consequence, the normalized weights are

$$\bar{w}(\widetilde{\mathbf{x}}_1) = \bar{w}(\widetilde{\mathbf{x}}_2) = \ldots = \bar{w}(\widetilde{\mathbf{x}}_N) = \frac{1}{N}. \tag{12}$$

The reason why this approach is suitable lies on the Liu's definition of proper importance weights in Section 2.1. Indeed, considering the random variable $\widetilde{\mathbf{X}} \sim \widehat{\pi}(\mathbf{x}|\mathbf{x}_{1:N})$, we have

$$\begin{aligned} E_{\widehat{\pi}}[h(\widetilde{\mathbf{X}})|\mathbf{x}_{1:N}] &= \int h(\mathbf{x})\widehat{\pi}(\mathbf{x}|\mathbf{x}_{1:N})d\mathbf{x}, \\ &= \sum_{n=1}^{N} \bar{w}(\mathbf{x}_n)h(\mathbf{x}_n) = \overline{I}_N. \end{aligned} \tag{13}$$

where $\mathbf{x}_n \sim q(\mathbf{x})$ for $n = 1, \ldots, N$, are considered fixed in the expectation $E_{\widehat{\pi}}[h(\widetilde{\mathbf{X}})|\mathbf{x}_{1:N}]$. Now, let us resample M times. The self-normalized IS estimator using the M resampled particles is

$$\widetilde{I}_M = \frac{1}{M}\sum_{m=1}^{M} h(\widetilde{\mathbf{x}}_m) \xrightarrow[M\to\infty]{P} E_{\widehat{\pi}}[h(\widetilde{\mathbf{X}})|\mathbf{x}_{1:N}] = \overline{I}_N. \tag{14}$$

Hence, we have

$$\widetilde{I}_M \xrightarrow[M\to\infty]{P} \overline{I}_N \xrightarrow[N\to\infty]{P} I, \tag{15}$$

due to Eqs. (13)-(14). This proves that the choice $\bar{w}(\widetilde{\mathbf{x}}_m) = \frac{1}{M}$, for all $m = 1, \ldots, M$, is proper by Liu's definition. However, for several theoretical and practical reasons (some of them discussed below), it is interesting to define also a proper *unnormalized* importance weight of a resampled particle. Let us consider the following definition.

Definition 1. *A proper choice for an unnormalized importance weight value (following Section 2.1) of a resampled particle* $\widetilde{\mathbf{x}} \in \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ *is*

$$\rho(\widetilde{\mathbf{x}}) = \rho(\widetilde{\mathbf{x}}|\mathbf{x}_{1:N}) = \widehat{Z} = \frac{1}{N}\sum_{i=1}^{N} w(\mathbf{x}_i). \tag{16}$$

Indeed, in this case, we have

$$E_{\widetilde{Q}(\mathbf{x},\mathbf{x}_{1:N})}[\rho(\mathbf{x}|\mathbf{x}_{1:N})h(\mathbf{x})] = cE_{\bar{\pi}}[h(\mathbf{x})], \tag{17}$$

where $\widetilde{Q}(\mathbf{x},\mathbf{x}_{1:N}) = \widehat{\pi}(\mathbf{x}|\mathbf{x}_{1:N})\left[\prod_{i=1}^{N} q(\mathbf{x}_i)\right]$.

Proof. *We show that Eq.* (17) *holds. Note that*

$$\begin{aligned} E_{\widetilde{Q}(\mathbf{x},\mathbf{x}_{1:N})}[\rho(\mathbf{x}|\mathbf{x}_{1:N})h(\mathbf{x})] &= \int_{\mathcal{X}}\int_{\mathcal{X}^N} h(\mathbf{x})\rho(\mathbf{x}|\mathbf{x}_{1:N})\widetilde{Q}(\mathbf{x},\mathbf{x}_{1:N})d\mathbf{x}d\mathbf{x}_{1:N}, &(18)\\ &= \int_{\mathcal{X}}\int_{\mathcal{X}^N} h(\mathbf{x})\rho(\mathbf{x}|\mathbf{x}_{1:N})\widehat{\pi}(\mathbf{x}|\mathbf{x}_{1:N})\left[\prod_{i=1}^{N} q(\mathbf{x}_i)\right] d\mathbf{x}d\mathbf{x}_{1:N}. &(19)\end{aligned}$$

Recalling that

$$\begin{aligned} \widehat{\pi}(\mathbf{x}|\mathbf{x}_{1:N}) &= \sum_{n=1}^{N} \bar{w}(\mathbf{x}_n)\delta(\mathbf{x}-\mathbf{x}_n) = \frac{1}{\sum_{n=1}^{N} w(\mathbf{x}_n)}\sum_{n=1}^{N} w(\mathbf{x}_n)\delta(\mathbf{x}-\mathbf{x}_n), &(20)\\ &= \frac{1}{N\widehat{Z}}\sum_{n=1}^{N} w(\mathbf{x}_n)\delta(\mathbf{x}-\mathbf{x}_n), &(21)\end{aligned}$$

where $\widehat{Z} = \widehat{Z}(\mathbf{x}_{1:N}) = \frac{1}{N}\sum_{n=1}^{N} w(\mathbf{x}_n)$. *Recalling also* $w(\mathbf{x}_n) = \frac{\pi(\mathbf{x}_n)}{q(\mathbf{x}_n)}$, *we can rearrange the expectation above as*

$$\begin{aligned} E_{\widetilde{Q}(\mathbf{x},\mathbf{x}_{1:N})}[\rho(\mathbf{x}|\mathbf{x}_{1:N})h(\mathbf{x})] &= \int_{\mathcal{X}} h(\mathbf{x})\left[\sum_{j=1}^{N}\left(\int_{\mathcal{X}^{N-1}}\rho(\mathbf{x}|\mathbf{x}_{1:N})\frac{w(\mathbf{x})}{N\widehat{Z}}\left[\prod_{\substack{i=1\\ i\neq j}}^{N} q(\mathbf{x}_i)\right]d\mathbf{x}_{\neg j}\right)\right]d\mathbf{x}, &(22)\\ &= \int_{\mathcal{X}} h(\mathbf{x})\left[\pi(\mathbf{x})\sum_{j=1}^{N}\left(\int_{\mathcal{X}^{N-1}}\rho(\mathbf{x}|\mathbf{x}_{1:N})\frac{1}{N\widehat{Z}}\left[\prod_{\substack{i=1\\ i\neq j}}^{N} q(\mathbf{x}_i)\right]d\mathbf{x}_{\neg j}\right)\right]d\mathbf{x}, &(23)\end{aligned}$$

where $\mathbf{x}_{\neg j} = [\mathbf{x}_1, \ldots, \mathbf{x}_{j-1}, \mathbf{x}_{j+1}, \ldots, \mathbf{x}_N]$. *The we have,*

$$E_{\widetilde{Q}(\mathbf{x},\mathbf{x}_{1:N})}[\rho(\mathbf{x}|\mathbf{x}_{1:N})h(\mathbf{x})] = \int_{\mathcal{X}} h(\mathbf{x})\pi(\mathbf{x})\left[\sum_{j=1}^{N}\left(\int_{\mathcal{X}^{N-1}}\rho(\mathbf{x}|\mathbf{x}_{1:N})\frac{1}{N\widehat{Z}}\left[\prod_{\substack{i=1\\ i\neq j}}^{N} q(\mathbf{x}_i)\right]d\mathbf{x}_{\neg j}\right)\right]d\mathbf{x}, \tag{24}$$

If we choose $\rho(\mathbf{x}|\mathbf{x}_{1:N}) = \widehat{Z}$ and replace in the expression above, we obtain

$$E_{\widetilde{Q}(\mathbf{x},\mathbf{x}_{1:N})}[\rho(\mathbf{x}|\mathbf{x}_{1:N})h(\mathbf{x})] = \int_{\mathcal{X}} h(\mathbf{x})\pi(\mathbf{x}) \left[\sum_{j=1}^{N} \left(\int_{\mathcal{X}^{N-1}} \widehat{Z}\frac{1}{N\widehat{Z}} \left[\prod_{\substack{i=1\\i\neq j}}^{N} q(\mathbf{x}_i)\right] d\mathbf{x}_{\neg j}\right)\right] d\mathbf{x}, \quad (25)$$

$$= \int_{\mathcal{X}} h(\mathbf{x})\pi(\mathbf{x}) N\frac{1}{N} d\mathbf{x}, \quad (26)$$

$$= \int_{\mathcal{X}} h(\mathbf{x})\pi(\mathbf{x}) d\mathbf{x} \quad (27)$$

$$= cE_{\bar{\pi}}[h(\mathbf{x})], \quad (28)$$

where $c = Z$. □

Considering N independent resampled particles, Note that with this definition we again have

$$\rho(\widetilde{\mathbf{x}}_1) = \rho(\widetilde{\mathbf{x}}_2) = \ldots = \rho(\widetilde{\mathbf{x}}_N),$$

so that also $\bar{\rho}(\widetilde{\mathbf{x}}_n) = \frac{1}{N}$, for all $n = 1, \ldots, N$, denoting with $\bar{\rho}(\widetilde{\mathbf{x}}_n)$ the corresponding normalized weights.

Remark 1. *The previous definition allows is to estimate Z using the resampled particles as well. Indeed,*

$$\widetilde{Z} = \frac{1}{N}\sum_{i=1}^{N} \rho(\widetilde{\mathbf{x}}_i) = \frac{1}{N}\left(N\widehat{Z}\right) = \widehat{Z}, \quad (29)$$

is an unbiased estimator of Z (equivalent to $\widehat{Z}$).

4. APPLICATION IN SIR

Let recall $\mathbf{x} = x_{1:D} = [x_1, x_2, \ldots, x_D] \in \mathcal{X} \subseteq \mathbb{R}^{D\times\eta}$ where $x_d \in \mathbb{R}^{\eta}$ for all $d = 1, \ldots, D$ and let us consider a target pdf $\bar{\pi}(\mathbf{x})$ factorized as

$$\bar{\pi}(\mathbf{x}) \propto \pi(\mathbf{x}) = \gamma_1(x_1)\prod_{d=2}^{D} \gamma_d(x_d|x_{1:d-1}), \quad (30)$$

where $\gamma_1(x_1)$ is a marginal pdf and $\gamma_d(x_d|x_{1:d-1})$ are conditional pdfs. We also denote the joint probability of $[x_1, \ldots, x_d]$,

$$\bar{\pi}_d(x_{1:d}) = \frac{1}{Z_d}\pi_d(x_{1:d}), \quad (31)$$

where

$$\pi_d(x_{1:d}) = \gamma_1(x_1)\prod_{j=2}^{d} \gamma_j(x_j|x_{1:j-1}).$$

Clearly, $\bar{\pi}(\mathbf{x}) \equiv \bar{\pi}_D(x_{1:D})$. We can also consider a proposal pdf decomposed in the same fashion,

$$q(\mathbf{x}) = q_1(x_1)q_2(x_2|x_1)\cdots q_D(x_D|x_{1:D-1}).$$

In a batch IS scheme, given the n-th sample $\mathbf{x}_n = x_{1:D}^{(n)} \sim q(\mathbf{x})$, we assign the importance weight

$$w(\mathbf{x}_n) = \frac{\pi(\mathbf{x}_n)}{q(\mathbf{x}_n)}$$

$$= \frac{\gamma_1(x_1^{(n)})\gamma_2(x_2^{(n)}|x_1^{(n)})\cdots\gamma_D(x_D^{(n)}|x_{1:D-1}^{(n)})}{q_1(x_1^{(n)})q_2(x_2^{(n)}|x_1^{(n)})\cdots q_D(x_D^{(n)}|x_{1:D-1}^{(n)})}.$$

The previous expression suggests a recursive procedure for computing the importance weights. Indeed, in a sequential Importance sampling (SIS) approach [8, 9], we can write

$$w_d^{(n)} = w_{d-1}^{(n)}\beta_d^{(n)} = \prod_{j=1}^{d}\beta_j^{(n)}, \qquad n = 1,\dots,N, \tag{32}$$

where we have set

$$w_1^{(n)} = \beta_1^{(n)} = \frac{\pi(x_1^{(n)})}{q(x_1^{(n)})} \quad \text{and} \quad \beta_d^{(n)} = \frac{\gamma_d(x_d^{(n)}|x_{1:d-1}^{(n)})}{q_d(x_d^{(n)}|x_{1:d-1}^{(n)})}, \tag{33}$$

for $d = 2,\dots,D$. Clearly, $w(\mathbf{x}_n) \equiv w_D^{(n)}$. The estimator of the normalizing constant $Z_d = \int_{\mathbb{R}^{d\times\eta}} \pi_d(x_{1:d})dx_{1:d}$ at the d-th iteration is

$$\begin{aligned}\widehat{Z}_d &= \frac{1}{N}\sum_{n=1}^{N} w_d^{(n)} = \frac{1}{N}\sum_{n=1}^{N} w_{d-1}^{(n)}\beta_d^{(n)}, &(34)\\ &= \frac{1}{N}\sum_{n=1}^{N}\left[\prod_{j=1}^{d}\beta_j^{(n)}\right]. &(35)\end{aligned}$$

Again, $Z \equiv Z_D$ and $\widehat{Z} \equiv \widehat{Z}_D$. However, an alternative formulation is often used [9, 10]

$$\begin{aligned}\overline{Z}_d &= \prod_{j=1}^{d}\left[\sum_{n=1}^{N}\bar{w}_{j-1}^{(n)}\beta_j^{(n)}\right] = \prod_{j=1}^{d}\left[\frac{\sum_{n=1}^{N} w_j^{(n)}}{\sum_{n=1}^{N} w_{j-1}^{(n)}}\right],\\ &= \widehat{Z}_1\prod_{j=2}^{d}\left[\frac{\widehat{Z}_j}{\widehat{Z}_{j-1}}\right] = \widehat{Z}_1\frac{\widehat{Z}_2}{\widehat{Z}_1}\cdots\frac{\widehat{Z}_d}{\widehat{Z}_{d-1}} = \widehat{Z}_d. &(36)\end{aligned}$$

Therefore, in SIS, $\widehat{Z}_d$ in Eq. (34) and $\overline{Z}_d$ in Eq. (36) are equivalent formulations of the same estimator of Z_d [18]. Furthermore, note that $\widehat{Z}_d$ can be written in a recursive form as

$$\overline{Z}_d = \overline{Z}_{d-1}\left[\sum_{n=1}^{N}\bar{w}_{d-1}^{(n)}\beta_d^{(n)}\right]. \tag{37}$$

4.1. Estimators of the marginal likelihood in SIR

Sequential Importance Resampling (SIR) [13, 22, 23, 24] combines the SIS approach with the application of the resampling procedure described in Section 3. Considering the proper importance weight of a resampled particle given in Definition 1 and recalling that $w_d^{(n)}$ the weight at the d-th iteration, we obtain the following recursion, $w_1^{(n)} = \beta_1^{(n)}$ and

$$w_d^{(n)} = w_{d-1}^{(n)}\beta_d^{(n)}, \tag{38}$$

for $d = 2,\dots,D$, where

$$w_{d-1}^{(n)} = \begin{cases} w_{d-1}^{(n)}, & \text{without resampling at } (d-1)\text{-th it.,}\\ \widehat{Z}_{d-1}, & \text{with resampling at } (d-1)\text{-th it.,}\end{cases} \tag{39}$$

i.e., if a resampling is applied at $(d-1)$-th iteration then $\xi_{d-1}^{(n)} = \widehat{Z}_{d-1}, \forall n = 1,\dots,N$.

Remark 2. *Using the Definition 1 and the recursive definition of the weights $w_d^{(n)}$ in Eqs.* (38)-(39)*, $\widehat{Z}_d$ and $\overline{Z}_d$ are both consistent and equivalent estimators of the marginal likelihood, also in SIR. Namely, the two estimators are*

$$\widehat{Z}_d = \frac{1}{N}\sum_{n=1}^{N} w_d^{(n)}, \quad \overline{Z}_d = \prod_{j=1}^{d}\left[\sum_{n=1}^{N}\bar{w}_{j-1}^{(n)}\beta_j^{(n)}\right] \tag{40}$$

are equivalent, $\widehat{Z}_d \equiv \overline{Z}_d$*. For instance, if the resampling is applied at each iteration, they become*

$$\overline{Z}_d = \prod_{j=1}^{d} \left[\frac{1}{N} \sum_{n=1}^{N} \beta_j^{(n)} \right], \tag{41}$$

$$\widehat{Z}_d = \widehat{Z}_{d-1} \left[\frac{1}{N} \sum_{n=1}^{N} \beta_d^{(n)} \right] = \prod_{j=1}^{d} \left[\frac{1}{N} \sum_{n=1}^{N} \beta_j^{(n)} \right], \tag{42}$$

and clearly coincide.

Figure 1 summarizes this scenario. In SIS, the estimators $\overline{Z}_d$ and $\widehat{Z}_d$ coincide with the formula $\overline{Z}_d \equiv \widehat{Z}_d = \frac{1}{N}\sum_{n=1}^{N}\left[\prod_{j=1}^{d}\beta_j^{(n)}\right]$ in Eq. (35). With a proper weighting of a resampled particles, in an adaptive resampling context, we have again $\overline{Z}_d \equiv \widehat{Z}_d$. If the resampling is applied at each iteration, they coincides with the expression $\overline{Z}_d \equiv \widehat{Z}_d = \prod_{j=1}^{d}\left[\frac{1}{N}\sum_{n=1}^{N}\beta_j^{(n)}\right]$.[2]

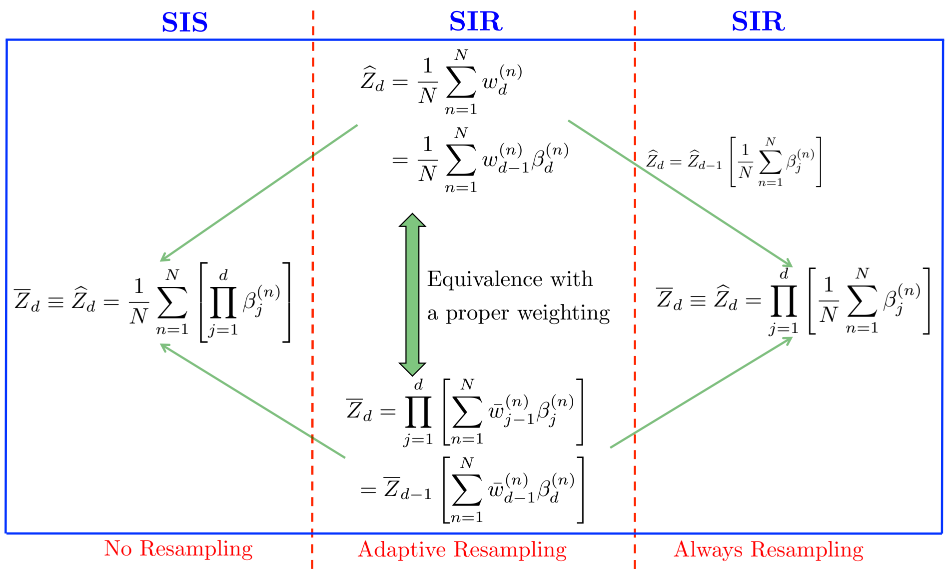

Fig. 1. Expressions of the marginal likelihood estimators $\overline{Z}_d$ and $\widehat{Z}_d$ in different SIR scenarios (without resampling, i.e., in SIS, adaptive resampling, and applying resampling at each iterations). If a proper weighting of a resampled particle is considered, they always coincide, $\overline{Z}_d \equiv \widehat{Z}_d$.

Remark 3. *Let us focus on the marginal likelihood estimators at the final iteration, i.e.,* $\widehat{Z} = \widehat{Z}_D$ *and* $\overline{Z} = \overline{Z}_D$*. Without using the Definition 1 and the recursive definition of the weights* $w_d^{(n)}$ *in Eqs.* (38)-(39)*, the only estimator of the marginal likelihood that can be properly computed in SIR is* $\overline{Z}$*, that involves only the computation of the normalized weights* $\bar{w}_d^{(n)}$ *(omitting the values of the corresponding unnormalized ones).*

5. PARTIAL RESAMPLING

The core of Sequential Monte Carlo methods is the SIR approach [8, 9, 23]. Namely, the weights are constructed recursively as in (32) and resampling steps, *involving all the particles*, are applied at some iterations. The combination of both, SIS and resampling schemes, is possible in the standard SIR approach *only* if the entire set of particles is employed in the resampling [8, 9], so that the assumption

$$\rho(\widetilde{\mathbf{x}}_1) = \rho(\widetilde{\mathbf{x}}_2) = \ldots = \rho(\widetilde{\mathbf{x}}_N), \tag{43}$$

[2] A related Matlab demo code is provided at `http://www.lucamartino.altervista.org/GIS.zip`.

is enough for computing $\overline{I}_N$ and $\overline{Z}$, since $\bar{w}(\widetilde{\mathbf{x}}_n) = \frac{1}{N}$ for all $n \in \{1, \dots, N\}$. If Definition 1 is used and then the recursive expression (38)-(39) is applied, we can define a resampling procedure *involving only a subset of particles*, as described in the following. We consider to apply a partial resampling scheme at the d-th iteration:

1. Choose randomly without replacement a subset of $M \leq N$ samples,

$$\mathcal{R} = \{x_d^{(j_1)}, \dots, x_d^{(j_M)}\},$$

   contained within the set of N particles $\{x_d^{(1)}, \dots, x_d^{(N)}\}$. Let us denote also the set

$$\mathcal{N} = \{x_d^{(1)}, \dots, x_d^{(N)}\} \backslash \mathcal{R},$$

   of the particles which do not take part in the resampling.

2. Give the set $\mathcal{R}$, resample with replacement M particles according to the normalized weights $\bar{w}_d^{(j_m)} = \frac{w_d^{(j_m)}}{\sum_{k=1}^M w_d^{(j_k)}}$, for $m = 1, \dots, M$, obtaining $\widetilde{\mathcal{R}} = \{\widetilde{x}_d^{(1)}, \dots, \widetilde{x}_d^{(M)}\}$. Clearly, $\widetilde{\mathcal{R}} \subseteq \mathcal{R}$.

3. For all the resampled particles in $\widetilde{\mathcal{R}}$ set

$$w_d^{(m)} = \frac{1}{M} \sum_{k=1}^{m} w_d^{(j_k)}, \tag{44}$$

   for $m = 1, \dots, M$, whereas the unnormalized importance weights of the particles in $\mathcal{N}$ remains invariant.

4. Go forward to the iteration $d + 1$ of the SIR method, using the recursive formula (32).

The procedure above is valid since yields proper weighted samples by Liu's definition. If $M = N$, it coincides with the traditional resampling procedure. This approach can reduce the loss of diversity due to the application of the resampling [8, 9, 13].

6. CONCLUSIONS

In this work, we have introduced a proper choice of the unnormalized weight assigned to a resampled particle. This choice entails several theoretical and practical consequences. We have described two of them, regarding (1) the estimation of the marginal likelihood and, (2) the application of a partial resampling involving only a subset of the cloud of particles, within SIR techniques. Other novel algorithms (based on the partial resampling perspective) and theoretical consequences (also affecting well-known the MCMC techniques, such as the particle Metropolis-Hastings method [1, 17], and parallel SMC implementations) will be highlighted out in an extended version of this work.

References

[1] C. Andrieu, A. Doucet, and R. Holenstein. Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society B*, 72(3):269–342, 2010.

[2] M. S. Arulumpalam, S. Maskell, N. Gordon, and T. Klapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions Signal Processing*, 50(2):174–188, February 2002.

[3] M. Bolić, P. M. Djurić, and S. Hong. Resampling algorithms for particle filters: A computational complexity perspective. *EURASIP Journal on Advances in Signal Processing*, 2004(15):2267–2277, November 2004.

[4] M. Bolić, P. M. Djurić, and S. Hong. Resampling algorithms and architectures for distributed particle filters. *IEEE Transactions Signal Processing*, 53(7):2442–2450, July 2005.

[5] M. F. Bugallo, L. Martino, and J. Corander. Adaptive importance sampling in signal processing. *Digital Signal Processing*, (47):36–49, 2015.

[6] O. Cappe and T. Ryden E. Moulines. *Inference in Hidden Markov Models*. Springer, 2005.

[7] O. Cappé, A. Guillin, J. M. Marin, and C. P. Robert. Population Monte Carlo. *Journal of Computational and Graphical Statistics*, 13(4):907–929, 2004.

[8] P. M. Djurić, J. H. Kotecha, J. Zhang, Y. Huang, T. Ghirmai, M. F. Bugallo, and J. Míguez. Particle filtering. *IEEE Signal Processing Magazine*, 20(5):19–38, September 2003.

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

[10] A. Doucet and A. M. Johansen. A tutorial on particle filtering and smoothing: fifteen years later. *technical report*, 2008.

[11] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo. Generalized multiple importance sampling. *arXiv:1511.03095*, 2015.

[12] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.-J. Nordlund. Particle filters for positioning, navigation and tracking. *IEEE Transactions Signal Processing*, 50(2):425–437, February 2002.

[13] J. S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer, 2004.

[14] J. S. Liu, F. Liang, and W. H. Wong. The multiple-try method and local optimization in metropolis sampling. *Journal of the American Statistical Association*, 95(449):121–134, March 2000.

[15] L. Martino, V. Elvira, and F. Louzada. Weighting a resampled particle in Sequential Monte Carlo. *IEEE Statistical Signal Processing Workshop, (SSP)*, 122:1–5, 2016.

[16] L. Martino, V. Elvira, D. Luengo, and J. Corander. An adaptive population importance sampler: Learning from the uncertainty. *IEEE Transactions on Signal Processing*, 63(16):4422–4437, 2015.

[17] L. Martino, F. Leisen, and J. Corander. On multiple try schemes and the Particle Metropolis-Hastings algorithm. *viXra:1409.0051*, 2014.

[18] L. Martino, J. Read, V. Elvira, and F. Louzada. Cooperative parallel particle filters for on-line model selection and applications to urban mobility. *viXra:1512.0420*, 2015.

[19] J. Míguez. Analysis of parallelizable resampling algorithms for particle filtering. *Signal Processing*, 87(12):3155–3174, 2007.

[20] J. Miguez and M. A. Vazquez. A proof of uniform convergence over time for a distributed particle filter. *Signal Processing*, 122:152–163, 2016.

[21] C. A. Naesseth, F. Lindsten, and T. B. Schon. Nested Sequential Monte Carlo methods. *Proceedings of theInternational Conference on Machine Learning*, 37:1–10, 2015.

[22] C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, 2004.

[23] D. B. Rubin. A noniterative sampling/importance resampling alternative to the data augmentation algorithm for creating a few imputations when fractions of missing information are modest: the SIR algorithm. *Journal of the American Statistical Association*, 82:543–546, 1987.

[24] D. B. Rubin. Using the sir algorithm to simulate posterior distributions. *in Bayesian Statistics 3, ads Bernardo, Degroot, Lindley, and Smith. Oxford University Press, Oxford, 1988.*, 1988.

[25] C. Verg, C. Dubarry, P. Del Moral, and E. Moulines. On parallel implementation of sequential Monte Carlo methods: the island particle model. *Statistics and Computing*, 25(2):243–260, 2015.

[26] C. Verg, P. Del Moral, E. Moulines, and J. Olsson. Convergence properties of weighted particle islands with application to the double bootstrap algorithm. *arXiv:1410.4231*, pages 1–39, 2014.

[27] N. Whiteley, A. Lee, and K. Heine. On the role of interaction in sequential Monte Carlo algorithms. *Bernoulli*, 22(1):494–529, 2016.