IEEE Open Journal of
Signal Processing

Received 23 July 2020; revised 28 October 2020; accepted 2 November 2020. Date of publication 16 November 2020; date of current version 11 December 2020. The review of this article was arranged by Associate Editor Prof. Usman A Khan.

Digital Object Identifier 10.1109/OJSP.2020.3038048

Flexible Effective Sample Size Based on the Message Importance Measure

ZHEFAN LI [1,2], PINGYI FAN [1,2] (Senior Member, IEEE), AND YUNQUAN DONG [3] (Member, IEEE)

[1]Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
[2]Beijing National Research Center for Information Science and Technology(BNRist), Beijing 100084, China
[3]School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China

CORRESPONDING AUTHORS: YUNQUAN DONG AND PINGYI FAN (e-mail: yunquandong@nuist.edu.cn; fpy@tsinghua.edu.cn).

The work of Pingyi Fan was supported in part by the China Major State Basic Research Development Program (973 Program) No. 2012CB316100(2) and in part by Beijing Natural Science Foundation under Grant 4202030. The work of Yunquan Dong was supported in part by the National Natural Science Foundation of China (NSFC) under Grants 61701247 and 62071237.

ABSTRACT In Monte Carlo based importance sampling estimations, Effective Sample Size (ESS) is an important index of simulation efficiency, since ESS can measure the divergence between the target distribution and the proposal distribution effectively, and thus is widely used to decide whether resampling is needed or not. Among several well-known variants of ESS, the Shannon entropy based perplexity has been widely used. In this paper, however, we propose a new ESS function (E-MIM) by using the message importance measure (MIM) instead of Shannon entropy. We show that E-MIM satisfies all of the five conditions for ESS generalizations. We also propose an MIM based divergence and investigate its approximation to E-MIM. Moreover, we present a resampling threshold selection method for the ratio between E-MIM and the corresponding actual sample size. Finally, we investigate the performance of E-MIM and other ESS functions through numerical simulations. By a particle filter experiment, we show that E-MIM outperforms other ESS functions in terms of mean-squared error.

INDEX TERMS Effective sample size, message importance measure, particle filter, sequential Monte Carlo.

I. INTRODUCTION

Sequential Monte Carlo (SMC), also called bootstrap filter or particle filter, is an important tool for Bayesian inference [1], which is widely applied in statistics [2]–[4], signal processing [5]–[7] and economics [8]–[10]. One of the key factors for the success of the SMC method is the use of resampling, without which the SMC method will suffer from serious weight degeneracy [11]–[13]. However, the resampling method also introduces some side effect, such as loss of diversity in particles and extra computational cost [12]. Therefore, resampling should be performed only when it is necessary, and it is important to schedule the sampling process and determine whether resampling is need or not [11]. In particular, existing schedules mainly fall into two categories, deterministic ones and adaptive ones [11]. In a deterministic schedule, resampling is used at some fixed time epochs (usually with equal intervals) [11]; in an adaptive schedule, resampling is performed only if the effectiveness of current particles is below a certian threshold [11]. Due to its flexibility, most adaptive schedules perform better than deterministic schedules [14].

Effective Sample Size (ESS) is a widely used criteria to measure the efficiency of different Monte Carlo methods [13], [15]–[19], such as importance sampling (IS) and Markov Chain Monte Carlo (MCMC). ESS is also referred to as the measurement of particle degeneracy (MPD) in particle filtering [20]. ESS is theoretically defined as the equivalent number of samples independently drawn from the target distribution, performing the same efficiency with the estimation obtained by IS or MCMC methods [12]. One explicit definition of ESS in mathematics is the ratio between the variance of the ideal samples directly drawn from the target distribution, and the variance of IS or MCMC estimators [18].

In general, it is difficult to obtain ESS explicitly [11], [12]. In importance sampling, a well-known approximation is $\widehat{ESS} = (\sum_{n=1}^{n} \bar{w}_n^2)^{-1}$ [11]–[13], [21], in which $\bar{w}_n$ is the normalized importance weights, for $n = 1, 2, \ldots, N$. Although widely used in practice, the derivation of this approximation involves several steps and assumptions, which prohibits accurate estimations of ESS, except in some specific cases [12]. It was discussed in [22] what undesirable consequences will be caused using the expression and the authors also showed even in toy models, assumptions involved in the deviation can be unrealistic, leading to poor approximation results .For this reason, several other expressions of ESS have been proposed. For instance, perplexity was proposed by using the discrete entropy [23] of the normalized importance weights [3], [14], [24]. Moreover, three approximation formulas were proposed in [25], based on the minimum variance rule and the minimum bias rule, respectively. In [26], the authors defined a family of ESS functions called p-ESS controlled by a parameter $p > 0$, and proved that the standard approximation $\widehat{ESS} = (\sum_{n=1}^{n} \bar{w}_n^2)^{-1}$ is a special case of p-ESS with $p = 2$.

Based on aforementioned considerations on ESS, the generalized effective sample size (G-ESS) was proposed, and five necessary conditions for G-ESS functions were presented [12]. To be specific, each ESS function should fully characterize the divergence between the proposal distribution and the target distribution of IS methods [12]. It has been shown that in certain scenarios, there exist differences among the performances of different ESS functions [12].

Considering the application of discrete entropy in the assessment of ESS [12], [14] and adaptive importance sampling [24], [27], it is promising to use information theory tools in SMC. Message importance measure was introduced to deal with minority subsets in big data [28]. Different from Shannon entropy [29] and Renyi entropy [30], MIM adopted an exponential form of probability to emphasize the influence of the small probability events, and used a positive importance coefficient $\bar{\omega}$ to enable MIM with more flexibility [28]. Several recent results confirmed the applications of MIM in various scenarios and showed that MIM could focus on different probability events through certain parameter selection [31]. By generalizing the domain of the importance coefficient, it has been shown that the parameter $\bar{\omega}$ acts as a switch of user's interests. That is, when $\bar{\omega} > 0$, MIM focuses on small-probability events and when $\bar{\omega} < 0$, MIM focuses on big-probability events [32]. Since sampling can also be considered as the process of collecting the information of the target distribution, MIM would also be a promising method in resampling. In this paper, therefore, we propose an effective sample size measure based on the message importance measure, which is referred as E-MIM. E-MIM is a new family of ESS functions with an adjustble parameter α. We prove that E-MIM satisfies all of the five conditions for the generalized effective sample size, and explore its relationship with former ESS functions. Furthermore, inspired by the K-L divergence based on the Shannon entropy, we propose a MIM divergence and show its approximation to E-MIM.

The rest of this paper is organized as follows. First, we present a brief introduction of importance sampling and some widely used ESS functions in Section II. Section III present the definition of E-MIM. In Section IV, the properties of E-MIM are discussed. Then, we explore the relationship between E-MIM and former ESS functions in Section V. In Section VI, we present a threshold selection method of E-MIM to determine the time for resampling. Some numerical results are given in Section VII. Finally, we complete the paper with conclusions in Section VIII.

II. IMPRTANCE SAMPLING AND EFFECTIVE SAMPLE SIZE

In this section, we present a brief introduction of importance sampling and effective sample size.

Let us denote the target distribution as $\bar{\pi}(x) \propto \pi(x)$ for $x \in \mathcal{X}$, which is known up to a normalizing constant. We denote $h(x)$ as a square-integrable function on $\mathcal{X}$. We define

$$I = \int_{\mathcal{X}} h(x)\bar{\pi}(x)dx \tag{1}$$

and shall approximate the integral through the Monte Carlo methods. We draw N samples $x_1, x_2, \ldots, x_N$ from $\bar{\pi}$ independently and approximate I by

$$\hat{I} = \frac{1}{N}\sum_{n=1}^{N} h(x_n). \tag{2}$$

However, it is usually difficult to sample from the target distribution $\bar{\pi}(x)$ directly. Thus, we shall draw N samples $x_1, x_2, \ldots, x_N$ from a simpler proposal distribution $q(x)$ instead. According to the importance sampling, we assign each sample with the weight $w_n = \pi(x_n)/q(x_n)$, $n = 1, \ldots, N$. By using the following normalized importance weight

$$\bar{w}_n = \frac{w_n}{\sum_{n=1}^{N} w_n} n = 1, \ldots, N, \tag{3}$$

the approximation of the IS estimator can be expressed as

$$\tilde{I} = \sum_{n=1}^{N} \bar{w}_n h(x_n) \approx I. \tag{4}$$

Generally speaking, the performance of $\tilde{I}$ will be not so good as $\hat{I}$, since the samples are not drawn directly from the target distribution $\bar{\pi}(x)$. In practical implementations [7], [21], [27], therefore, it is necessary to measure the loss in efficiency when IS methods are used [12]. To this end, the Effective Sample Size (ESS) was defined as the ratio between the variances of the two approximation results [33]. That is,

$$ESS = N\frac{\mathrm{Var}_{\pi}(\hat{I})}{\mathrm{Var}_{q}(\tilde{I})}. \tag{5}$$

ESS measures the equivalent number of samples needed to be drawn directly from $\bar{\pi}(x)$ to achieve the same efficiency as that of the IS estimator, which draws N samples from the proposal distribution $q(x)$. Since the definition is related with both the target distribution and proposal distribution, as well as the function $h(x)$, it is generally difficult to obtain the theoretical value of ESS [11]. Thus, approximation based methods are used as alternative solutions in practice.

The most popular approximation for ESS is [11], [12], [34]

$$\begin{aligned}\widehat{ESS} &= \frac{1}{\sum_{n=1}^{N} \bar{w}_n^2} \\ &\triangleq P_N^{(2)}(\bar{\mathbf{w}}).\end{aligned} \tag{6}$$

Another approximation method referred to as the *perplexity*

$$\widehat{ESS} = 2^{H(\bar{\mathbf{w}})} \triangleq Per_N(\bar{\mathbf{w}}), \tag{7}$$

used the discrete entropy of the following normalized weights

$$H(\bar{\mathbf{w}}) = -\sum_{n=1}^{N} \bar{w}_n \log_2 \bar{w}_n. \tag{8}$$

More recently, a maximum normalized weight based ESS function was proposed as [12], [14],

$$\widehat{ESS} = \frac{1}{\max[\bar{w}_1, \ldots, \bar{w}_N]} \triangleq D_N^{(\infty)}(\bar{\mathbf{w}}). \tag{9}$$

As is shown in [12], this method outperforms the $\widehat{ESS} = (\sum_{n=1}^{n} \bar{w}_n^2)^{-1}$ measure in most cases.

The symbols, $P_N^{(2)}(\bar{\mathbf{w}})$, $Per_N(\bar{\mathbf{w}})$ and $D_N^{(\infty)}(\bar{\mathbf{w}})$ above follow the statement in previous literatures [12], to distinguish different approximation functions of ESS.

III. THE DEFINITION OF E-MIM

In this Section, we present the proposed E-MIM method and illustrate its relationship with MIM.

A. MIM-BASED EFFECTIVE SAMPLE SIZE

Definition 1: For a given set of normalized importance weights $\bar{\mathbf{w}} = [\bar{w}_1, \ldots, \bar{w}_N]$ and importance coefficient $\alpha \in \mathcal{R}$, the message importance measure based effective sample size (E-MIM), is defined as

$$E_N(\bar{\mathbf{w}}, \alpha) = -\frac{N\alpha}{\log \sum_{n=1}^{N} \bar{w}_n \exp(-N\alpha\bar{w}_n)}. \tag{10}$$

B. RELATIONSHIP WITH MIM

For a discrete probability distribution $\{p_1, p_2, \ldots, p_N\}$, and a chosen importance coefficient $\bar{\omega}$, the message importance measure, or MIM, is defined as [28]

$$L(p, \bar{\omega}) = \log \sum_{i=1}^{N} p_i \exp(\bar{\omega}(1 - p_i)). \tag{11}$$

Since MIM can also be expressed as

$$\begin{aligned} L(p, \bar{\omega}) &= \log \sum_{i=1}^{N} p_i \exp\left(\bar{\omega}(1 - p_i)\right) \\ &= \log \left(\exp(\bar{\omega}) \sum_{i=1}^{N} p_i \exp(-\bar{\omega} p_i) \right) \\ &= \log \left(\sum_{i=1}^{N} p_i \exp(-\bar{\omega} p_i) \right) + \bar{\omega}, \end{aligned} \tag{12}$$

E-MIM can be expressed in terms of MIM by

$$E_N(\bar{\mathbf{w}}, \alpha) = -\frac{N\alpha}{L(\bar{\mathbf{w}}, N\alpha) - N\alpha}. \tag{13}$$

IV. PROPERTIES OF E-MIM

In this section, we investigate the property of E-MIM.

A. SYMMETRY

For any permutation of the normalized weights $\bar{\mathbf{w}} = [\bar{w}_1, \ldots, \bar{w}_N]$, E-MIM does not change in value, i.e.

$$E_N(\bar{\mathbf{w}}, \alpha) = E_N(\bar{\mathbf{w}}', \alpha), \tag{14}$$

in which $\bar{\mathbf{w}}' = [w_{j_1}, \ldots, w_{j_N}]$, holds true for any possible set $\{j_1, j_2, \ldots, j_N\} = \{1, 2, \ldots, N\}$.

B. MAXIMUM CONDITION

E-MIM achieves its maximum value N if $\alpha < 1$ and $\bar{\mathbf{w}}^* = [1/N, \ldots, 1/N]$. That is

$$E_N(\bar{\mathbf{w}}^*, \alpha) = N \geq E_N(\bar{\mathbf{w}}, \alpha). \tag{15}$$

C. MINIMUM CONDITION

When $\bar{\mathbf{w}}^{(j)} = [\bar{w}_1 = 0, \ldots, \bar{w}_j = 1, \ldots, \bar{w}_N = 0]$, E-MIM reaches the mininum value 1, i.e.,

$$E_N(\bar{\mathbf{w}}^{(j)}, \alpha) = 1 \leq E_N(\bar{\mathbf{w}}, \alpha). \tag{16}$$

D. UNIQUENESS OF EXTREME VALUE

If $\alpha < 1$, E-MIM achieves its maximum value if and only if $\bar{\mathbf{w}} = \bar{\mathbf{w}}^*$. E-MIM achieves its minimum value 1 if and only if $\bar{\mathbf{w}} = \bar{\mathbf{w}}^{(j)}$, $j = 1, 2, \ldots, N$.

E. STABLITY– INVARIANCE OF THE RATE ESS/N

We consider a vector $\bar{\mathbf{w}} = [w_1, \ldots, w_N] \in R^N$ and a vector

$$\bar{\mathbf{v}} = [v_1, \ldots, v_{MN}] \in R^{MN}, M > 1, \tag{17}$$

which is a normalized repetition of $\bar{\mathbf{w}}$ by M times. That is

$$\bar{\mathbf{v}} = \frac{1}{M} \underbrace{[\bar{\mathbf{w}}, \ldots, \bar{\mathbf{w}}]}_{M\text{times}}. \tag{18}$$

Since we have $\sum_{n=1}^{MN} \bar{v}_n = \frac{1}{M}[M \sum_{n=1}^{N} \bar{w}_n] = 1$, the E-MIM would satisfy

$$\frac{E_N(\bar{\mathbf{w}}, \alpha)}{N} = \frac{E_{MN}(\bar{\mathbf{v}}, \alpha)}{MN} \tag{19}$$

$$E_N(\bar{\mathbf{w}}, \alpha) = \frac{1}{M} E_{MN}(\bar{\mathbf{v}}, \alpha). \tag{20}$$

Remark 1: The properties from A to E are five requirements that a G-ESS function should satisfy [12]. Further, properties B and C show an optimistic approximation of the theoretical value [12]. When the proposal distribution is the same as the target distribution, i.e., $\bar{\pi}(x) = q(x)$, it is clear that $\bar{\mathbf{w}} = \bar{\mathbf{w}}^*$ and $ESS = N$, yet the converse is not always true, i.e., even if $\bar{\mathbf{w}} = \bar{\mathbf{w}}^*$, the proposal distribution can be different from the target distribution. Therefore, $ESS \leq N$ holds ture in theory. So E-MIM gives an optimistic approximation in the maximum condition. The other extreme case is when $\bar{\mathbf{w}}^{(j)} = [\bar{w}_1 = 0, \ldots, \bar{w}_j = 1, \ldots, \bar{w}_N = 0]$, $j = 1, \ldots, N$, i.e., $\bar{w}_n = 0$, for $n \neq j$, and $\bar{w}_j = 1$. The best possible occasion is that the j-th sample is directly drawn from $\bar{\pi}(x)$, i.e., $\bar{\pi}(x_j) = q(x_j)$, thus $ESS \leq 1$. So E-MIM also gives an optimistic approximation in the minimum condition.

F. NON-DECREASING IN COEFFICIENT α

For a given normalized weights $\bar{\mathbf{w}} = [\bar{w}_1, \ldots, \bar{w}_N]$, E-MIM is monotonically non-decreasing with importance coefficient α.

Specially, despite when $\bar{\mathbf{w}}$ is uniformly distributed regardless of zero elements, E-MIM is monotonically increasing with importance coefficient α.

Remark 2: Property F suggests that the importance coefficient α characterizes the optimism level of approximation of ESS or the tolerance for the divergence between the proposal distribution and the target distribution. More specifically, α can be a good parameter to balance the approximation and divergence tolerance between $\bar{\pi}(x)$ and $q(x)$.

The proofs of properties A-F are shown in Appendix A.

V. RELATIONSHIP WITH OTHER ESS FUNCTIONS

In this section, we investigate the relationship between E-MIM and other measures, including $P_N^{(2)}(\bar{\mathbf{w}})$, $D_N^{(\infty)}(\bar{\mathbf{w}})$ and $Per_N(\bar{\mathbf{w}})$.

A. WITH COEFFICIENT α = 0

Theorem 1: If $\alpha = 0$, E-MIM reduces to $P_N^{(2)}(\bar{\mathbf{w}})$.

Proof:

$$\begin{aligned}\lim_{\alpha\to 0} E_N(\bar{\mathbf{w}}, \alpha) &= -\lim_{\alpha\to 0} \frac{N\alpha}{\log \sum_{n=1}^N \bar{w}_n \exp(-N\alpha\bar{w}_n)} \\ &= -\lim_{\alpha\to 0} \frac{N\sum_{n=1}^N \bar{w}_n \exp(-N\alpha\bar{w}_n)}{-N\sum_{n=1}^N \bar{w}_n^2 \exp(-N\alpha\bar{w}_n)} \\ &= \frac{1}{\sum_{n=1}^N \bar{w}_n^2}.\end{aligned} \tag{21}$$

■

B. WHEN COEFFICIENT α → −∞

Theorem 2: If $\alpha = -\infty$, E-MIM reduces to $D_N^{(\infty)}(\bar{\mathbf{w}})$.

Proof:

$$\begin{aligned}&\lim_{\alpha\to -\infty} E_N(\bar{\mathbf{w}}, \alpha) \\ &= -\lim_{\alpha\to -\infty} \frac{N\alpha}{\log \sum_{n=1}^N \bar{w}_n \exp(-N\alpha\bar{w}_n)} \\ &= \lim_{\alpha\to -\infty} \frac{\sum_{n=1}^N \bar{w}_n \exp(-N\alpha\bar{w}_n)}{\sum_{n=1}^N \bar{w}_n^2 \exp(-N\alpha\bar{w}_n)} \\ &= \lim_{\alpha\to -\infty} \frac{\sum_{n=1}^N \bar{w}_n \exp(-N\alpha(\bar{w}_n - max[\bar{w}_1, \ldots, \bar{w}_N]))}{\sum_{n=1}^N \bar{w}_n^2 \exp(-N\alpha(\bar{w}_n - max[\bar{w}_1, \ldots, \bar{w}_N]))} \\ &= \frac{1}{max[\bar{w}_1, \ldots, \bar{w}_N]}.\end{aligned} \tag{22}$$

■

Remark 3: Likewise, we have $\lim_{\alpha\to +\infty} E_N(\bar{\mathbf{w}}, \alpha) = \frac{1}{min[\bar{w}_1, \ldots, \bar{w}_N]}$, which confirms the conclusion that the importance coefficient performs like a switch controlling whether to focus on small-probability events or big-probability events. When $\alpha > 1$, however, E-MIM does not satisfy the requirement B that a G-ESS function should be no more than N, and thus is not considered in this paper.

Remark 4: Subsection A and B have shown that E-MIM can be a generalization of $P_N^{(2)}$ and $D_N^{(\infty)}$, while $P_N^{(2)}$ is the most widely used approximation and $D_N^{(\infty)}$ behaves as a lower bound of the theoretical value of ESS [12].

C. RELATIONSHIP WITH THE PERPLEXITY AND ENTROPY CRITERION

The perplexity $Per_N(\bar{\mathbf{w}})$ can be related to the entropy criterion [24] of the SMC methods. The entropy criterion considers the K-L divergence [35] or the relative entropy of the proposal distribution and the target distribution to measure the divergence between the two. It is clear that the smaller the relative entropy is, the closer the proposal distribution is to the target distribution, and the more efficient the IS estimator is. The K-L divergence or the relative entropy is defined as

$$D(\bar{\pi}(x) \,||\, q(x)) = \int \bar{\pi}(x) \log_2 \frac{\bar{\pi}(x)}{q(x)} dx. \tag{23}$$

Suppose the known distribution satisfies

$$\pi(x) = Z\bar{\pi}(x), \tag{24}$$

in which Z is the normalizing constant, we then have

$$D(\bar{\pi}(x) \,|\, q(x)) = \int \bar{\pi}(x) \log_2 \frac{\pi(x)}{Zq(x)} dx. \tag{25}$$

Based on the samples $x_1, x_2, \ldots, x_N$ from IS estimator, the approximation of the target distribution can then be expressed as

$$\hat{\bar{\pi}}(x) = \sum_{n=1}^N \bar{w}_n \delta(x - x_n). \tag{26}$$

The approximation of the constant Z is

$$\hat{Z} = \frac{1}{N} \sum_{n=1}^N w_n. \tag{27}$$

Thus, the approximation of the K-L divergence between $\bar{\pi}(x)$ and $q(x)$ is

$$\begin{aligned}\hat{D}(\bar{\pi}(x) \,||\, q(x)) &= \sum_{n=1}^N \bar{w}_n \log_2 \frac{w_n}{\frac{1}{N}\sum_{n=1}^N w_n} \\ &= \sum_{n=1}^N \bar{w}_n \log_2 \bar{w}_n + \log_2 N \\ &= \log_2 N - H(\bar{\mathbf{w}}).\end{aligned} \tag{28}$$

Then, we can obtain the relationship between $Per_N(\bar{\mathbf{w}})$ and the approximation of K-L divergence as

$$\begin{aligned}Per_N(\bar{\mathbf{w}}) &= 2^{H(\bar{\mathbf{w}})} \\ &= 2^{\log_2 N - \hat{D}(\bar{\pi}(x)||q(x))} \\ &= N * 2^{-\hat{D}(\bar{\pi}(x)||q(x))},\end{aligned} \tag{29}$$

i.e.,

$$\frac{Per_N(\bar{\mathbf{w}})}{N} = 2^{-\hat{D}(\bar{\pi}(x)||q(x))}. \tag{30}$$

Hence, the perplexity $Per_N(\bar{\mathbf{w}})$ shows the approximation of the K-L divergence between the target distribution and proposal distribution. When $\bar{\pi}(x) = q(x)$, the K-L divergence is zero, and $Per_N(\bar{\mathbf{w}})$ reaches the maximum value N.

Likewise, we can define MIM divergence to measure the divergence between two distributions.

Definition 2: The MIM divergence between two distributions, $\bar{\pi}(x)$ and $q(x)$, is

$$D_{MIM}(\bar{\pi}(x) \mid q(x), \alpha) = -\frac{1}{\alpha} \log \int \bar{\pi}(x) exp\left(\alpha\left(1 - \frac{\bar{\pi}(x)}{q(x)}\right)\right) dx, \quad (31)$$

in which, α is the importance coefficient.

Theorem 3: When $\alpha < 1$

$$D_{MIM}(\bar{\pi}(x) \mid q(x), \alpha) \geq 0. \quad (32)$$

The equality holds if and only if $\bar{\pi}(x) = q(x)$

Similarly with the approximation of K-L divergence, by substituting (26) and (27) in (31), we have

$$\begin{aligned} &\hat{D}_{MIM}(\bar{\pi}(x) \mid q(x), \alpha) \\ &= -\frac{1}{\alpha} \log \sum_{n=1}^{N} \bar{w}_n \exp\left(\alpha\left(1 - \frac{w_n}{\frac{1}{N}\sum_{n=1}^{N} w_n}\right)\right) \\ &= -\frac{1}{\alpha} \log\left(\sum_{n=1}^{N} \bar{w}_n \exp(-N\alpha\bar{w}_n)\right) - 1. \end{aligned} \quad (33)$$

Moreover, the relationship between E-MIM and MIM divergence is

$$\begin{aligned} E_N(\bar{\mathbf{w}}, \alpha) &= -\frac{N\alpha}{\log \sum_{n=1}^{N} \bar{w}_n \exp(-N\alpha\bar{w}_n)} \\ &= \frac{N}{\hat{D}_{MIM}(\bar{\pi}(x) \mid q(x), \alpha) + 1}. \end{aligned} \quad (34)$$

Hence, E-MIM shows the approximation of the MIM divergence between the target distribution and proposal distribution. When $\bar{\pi}(x) = q(x)$, the MIM divergence is zero, and E-MIM reaches the maximum value N.

Through the expression of the MIM divergence, we further have one understanding about the influence of the importance coefficient α. When $\alpha > 0$, the intervals with $\bar{\pi}(x)/q(x) < 1$ will have a greater influence on the MIM divergence, or the particles with their weights $\bar{w}_n < 1/N$ will contribute more to the approximation of ESS. When $\alpha < 0$, the intervals with $\bar{\pi}(x)/q(x) > 1$ will have a greater influence on the MIM divergence, or the particles with their weights $\bar{w}_n > 1/N$ will contribute more to the approximation of ESS. In extreme cases like in Subsection A and B, when $\alpha \to -\infty$, $E_N(\bar{\mathbf{w}}, \alpha) \to 1/\max[\bar{w}_1, \ldots, \bar{w}_1]$, E-MIM is dominated by the particle with maximum weight, when $\alpha \to +\infty$, $E_N(\bar{\mathbf{w}}, \alpha) \to 1/\min[\bar{w}_1, \ldots, \bar{w}_1]$, E-MIM is dominated by the particle with minimum weight. This property coordinates with the switch-like property of the importance coefficient in MIM, which can provide a guideline for the choice of α. When it needs to pay more attention to the particles with high weights, one can choose a negative value of α, while when it needs to pay more attention to the particles with low weights, one can choose a positive value of α. Numerical experiments in Section VII provide an example in which it needs to be more tolerant with particles with low weights, called bearing-only tracking. In this case, the low weight just represents that the particle has a deviation from the right direction currently, but the particle can have a high potential if its velocity is close to the true value, so it requires more tolerance to the particles with low weights. Simulation shows the selection of positive α improves the accuracy of tracking.

VI. SELECTION OF THE THRESHOLD

In most SMC iterations, we need to implement the resampling only when the ratio of the ESS function and the number of samples is below a given threshold ε, i.e., $ESS < \varepsilon N$ [12]. To be specific, resampling would be applied in every iteration if $\varepsilon = 1$ and would never occur if $\varepsilon = 0$. In the case $0 < \varepsilon < 1$, resampling will be applied adaptively, in which the frequency of resampling is generally increasing with the threshold ε. While large resampling rate leads to extra computational cost and the loss of the diversity of particles [11], [13], untimely and sparse resampling will result in severe degeneracy of weights [21]. Hence, it is important to select an appropriate threshold. In this section, we shall present a threshold selection scheme for E-MIM through exploring the distribution of E-MIM.

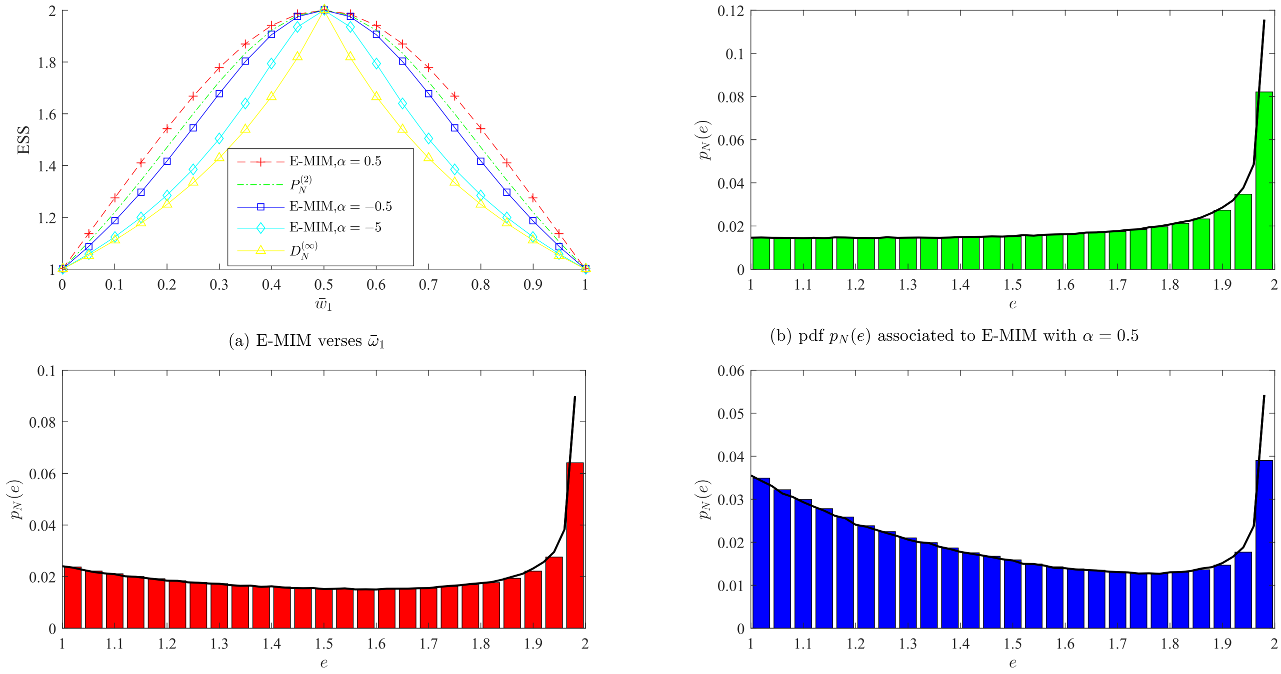
We consider the distribution of E-MIM over a unit simplex S_N, in which $\bar{\mathbf{w}}$ is a uniformly distributed random vector. That is, we consider the random vector $\bar{\mathbf{w}} \sim U(S_N)$ and the random variable $E = E_N(\bar{\mathbf{W}}, \alpha)$. The distribution function of E is denoted as $p_N(e)$, for which the support is $[1, N]$. The distribution of E can help us further understand the property of E-MIM.

Since it is generally difficult to express $p_N(e)$ analytically, we shall consider a simple case of $N = 2$ in this section and leave further discussions to Section VII. When $N = 2$, we have $\bar{\mathbf{w}} = [\bar{w}_1, \bar{w}_2] = [\bar{w}_1, 1 - \bar{w}_1]$. We plot how E changes with $\bar{w}_1$ in Fig. 1, in which the coefficient α take values from$\{0.5, -0.5, -5\}$. From Fig. 1(a), we observe that for the same distribution $\bar{\mathbf{w}}$ E-MIM increases with the coefficient α, which means it is more optimistic towards the efficiency of IS methods and more tolerant in terms of the degeneracy of weights of particles. Fig. 1 (b), 1(c), and 1(d) presents the distribution $p_N(e)$ of E-MIM, in which $\alpha = 0.5$, $\alpha = -0.5$, respectively. In the case $\alpha = 0.5$, $p_N(e)$ is not uniformly distributed and with larger value in the right-hand side. In the cases $\alpha = -0.5$ and $\alpha = -5$, although $p_N(e)$ is also unbalanced, we see that the distribution curve is convex with smaller e. All the three sub-figures indicate that E-MIM tends to move the values in the right-hand side to the left-hand side so that the efficiency of IS methods is reduced gradually and the tolerance to the degeneracy of weights becomes weaker.

We set the threshold as the expectation of the E-MIM within the unit simplex S_N [12], i.e.,

$$\varepsilon = \mathbb{E}[E/N]. \quad (35)$$

In the case $N = 2$, the threshold of E-MIM and the probability to resample, i.e., $P(E < \varepsilon N)$, can be presented as a function of coefficient α, as shown in Fig. 2. We observe that the threshold is increasing with α, which is consist with the property F of E-MIM. That is, E-MIM is monotonically



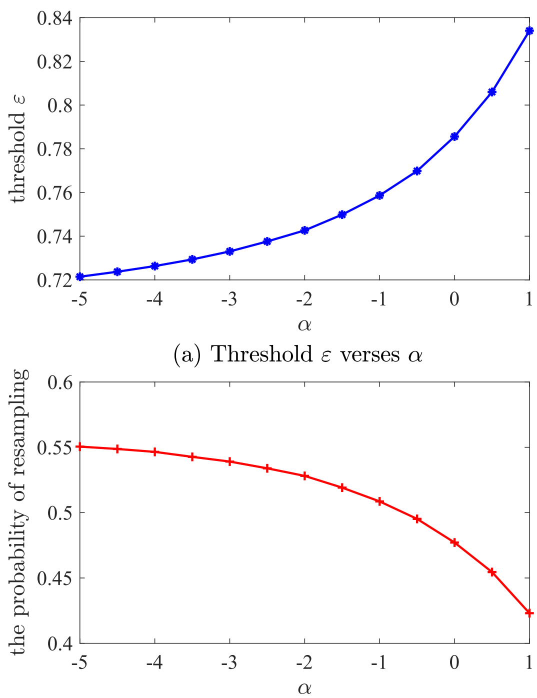
(a) E-MIM verses $\bar{\omega}_1$

(b) pdf $p_N(e)$ associated to E-MIM with $\alpha = 0.5$

(c) pdf $p_N(e)$ associated to E-MIM with $\alpha = -0.5$

(d) pdf $p_N(e)$ associated to E-MIM with $\alpha = -5$

FIGURE 1. Comparison of the distribution of E-MIM with different α.



(a) Threshold ε verses α

(b) The probability of resampling verses α

FIGURE 2. Threshold ε and the probability of resampling versus α.

non-decreasing with importance coefficient α. More importantly, the probability to resample is decreasing of α, which means with the increasing with α, one tends to give higher evaluation of the efficiency of the IS estimator and reduce the utilization of the resampling step, when using the above threshold choosing strategy.

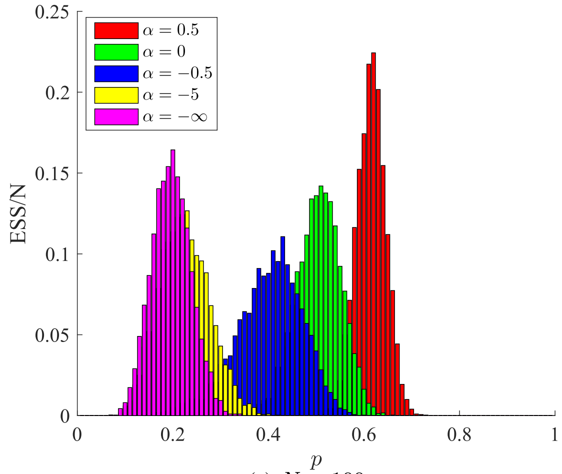
VII. SIMULATIONS

In this section, we provide some simulations to valid our theoretical results.

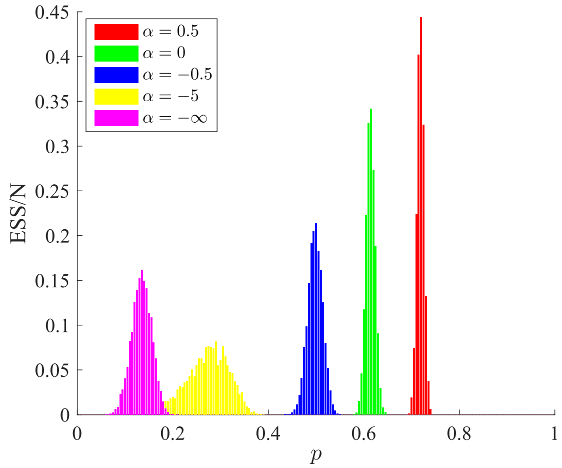
A. THE DISTRIBUTION OF E-MIM

To study the distribution of E-MIM with different coefficient α and number of samples, we draw the normalized weights $\bar{\mathbf{w}}$ uniformly from the unit simplex $S_N \in R^N$, and calculate the corresponding E-MIM. To be specific, we draw 3000 samples independently from S_N (the sampling methods refers to [36]) for each simulation. Fig. 2(a) and (b) show the histograms of distribution of E-MIM for $\alpha = 0.5$, $\alpha = 0, -0.5$, $\alpha = -5$, and $\alpha = -\infty$, with $N = 100$, and $N = 1000$, respectively. Fig. 3 displays how the expected value and variance of E-MIM change with the importance coefficient α. We observe that as α is increased, the expectation of E-MIM increases while the variance of E-MIM increases first and then decreases.

From Fig. 2 and 3, we observe that E-MIM functions with different α and N all concentrate around one mode, and different values of α and N will correspond to different locations. The variance is decreasing with the increase in sample size N. The distribution of E-MIM could help us decide the threshold for resampling. For example, $P_N^{(2)}(\bar{\mathbf{w}})$, or E-MIM with $\alpha = 0$, as suggested in Section VI, the threshold should be $\varepsilon = 0.5$, which shows the fact that with different sample size N, the mean of $P_N^{(2)}(\bar{\mathbf{w}}$ is always close to 0.5, as shown in Fig. 3.

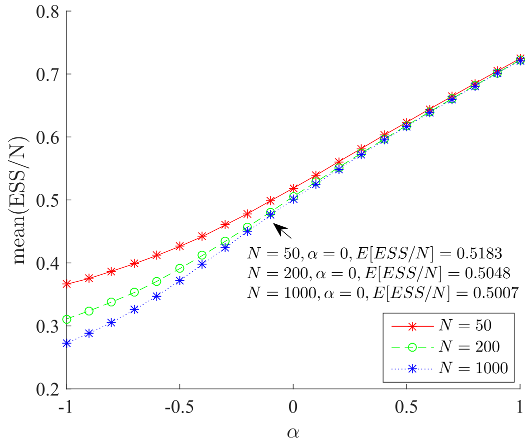


(a) $N = 100$.

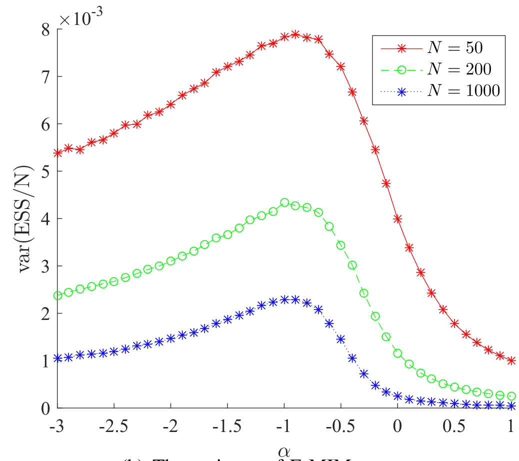


(b) $N = 1000$.

FIGURE 3. Distribution of E-MIM, for $\alpha = 0.5$, $\alpha = 0$, $\alpha = -0.5$, $\alpha = -5$, $\alpha = -\infty$, $N = 100$, and$N = 1000$, respectively.



(a) The expected value of E-MIM verses α.



(b) The variance of E-MIM verses α.

FIGURE 4. The mean and variance of E-MIM verses α.

B. APPROXIMATION OF THE THEOCRATICAL VALUE OF ESS

In this section, we compare the performance of different E-MIM functions in approximating the theoretical value of ESS.

$$ESS = N\frac{var_\pi(\hat{I}(h))}{var_\pi(\tilde{I}(h))}. \tag{36}$$

We consider the target distribution is a standard Gaussian distribution, i.e.,

$$\bar{\pi}(x) \sim \mathcal{N}(0, 1). \tag{37}$$

The proposal distribution is also a Gaussian distribution with mean μ and variance σ^2, i.e.,

$$q(x) \sim \mathcal{N}(\mu, \sigma^2). \tag{38}$$

As discussed in [12], we consider three different parameters settings to compare the estimation performance of E-MIM with different coefficient α, as well as $P_N^{(2)}(\bar{\mathbf{w}})$, $Per_N(\bar{\mathbf{w}})$ and $D_N^{(\infty)}(\bar{\mathbf{w}})$ for comparison, and set $h(x) = x$.

S1 $\sigma^2 = 1$ and $\mu \in [0, 2]$;

S2 $\mu = 1$ and $\sigma \in [0.3, 4]$;

S3 $\sigma^2 = 1$, $\mu \in \{0.5, 1, 1.5, 3\}$, and change N from 10 to 5000 to explore the influence of the sample size N.

In simulations with setting S1 and S2, we set $N = 10$ or $N = 1000$, repeat for $5 * 10^5$ times, and present the results in Fig. 5. In simulations with setting S3, we consider changing N from 10 to 5000, repeat for $5 * 10^5$ times and show the results in Fig. 6.

In the setting S1 with $N = 10$, we observe in Fig. 5(a) that compared to $P_N^{(2)}(\bar{\mathbf{w}})$, $D_N^{(\infty)}(\bar{\mathbf{w}})$, E-MIM with $\alpha = -0.5$ and E-MIM with $\alpha = -5$ perform more accurate approximations for the theoretical value of ESS. $Per_N(\bar{\mathbf{w}})$ and E-MIM with $\alpha = 0.5$ has relatively large deviation to the ESS in theory. In the setting S1 with $N = 1000$, we observe in Fig. 5(b) that E-MIM with $\alpha = -0.5$ and E-MIM with $\alpha = -5$ also have better approximations.

In the setting S2 with $N = 10$ shown in Fig. 5(c), E-MIM with $\alpha = -0.5$ and E-MIM with $\alpha = -5$ provide more accurate approximations when $\sigma < 1.5$; E-MIM with $\alpha = 0.5$ and $Per_N(\bar{\mathbf{w}})$ perform better than other schemes for $\sigma > 1.5$. In the setting S2 with $N = 1000$ shown in Fig. 5(d), we have similar observations.

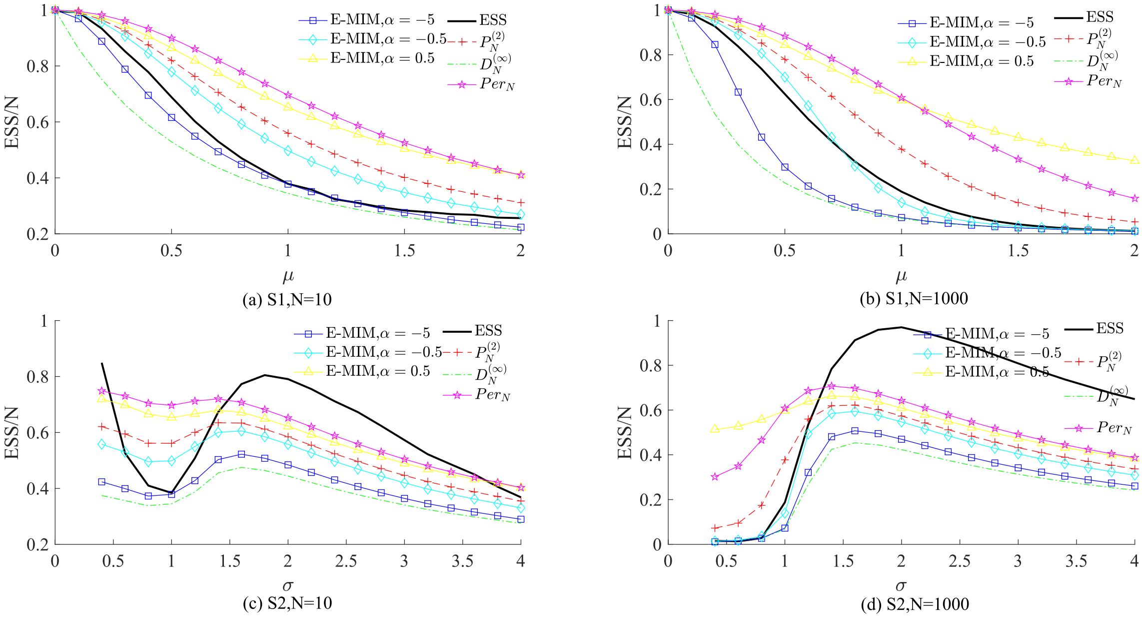


FIGURE 5. ESS/N versus standard variance σ under settings S1 and S2.

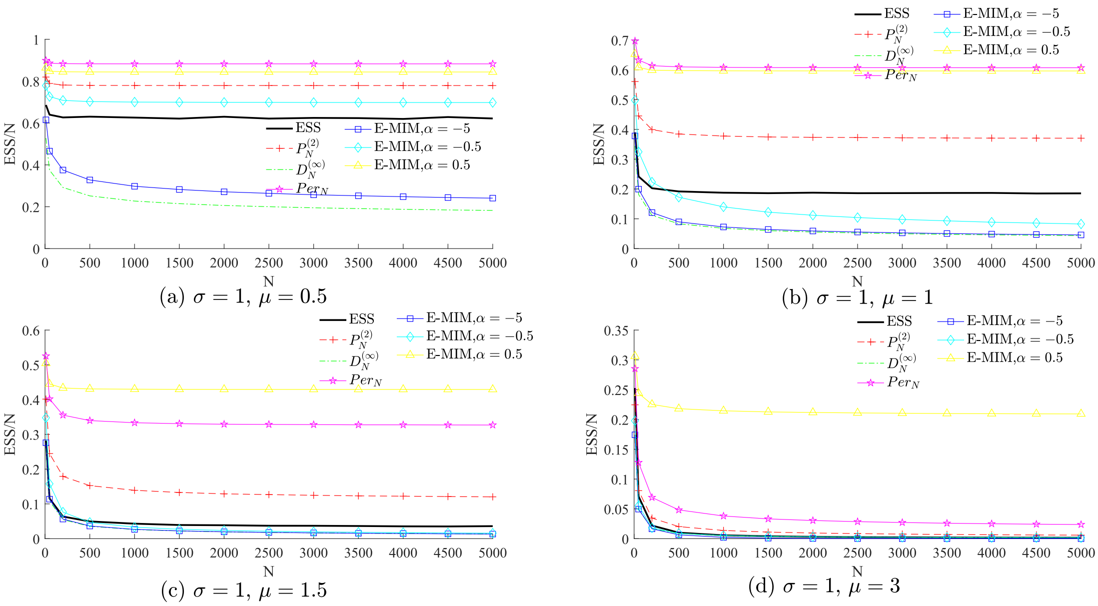


FIGURE 6. ESS/N versus standard variance σ under setting S3.

For setting S1 and S2, therefore, it concludes that E-MIM would outperform previous ESS functions by adjusting its coefficient. To be specific, E-MIM with a negative coefficient performs better when the divergence between the target distribution and the proposal distribution is small; E-MIM with a positive coefficient performs better if the divergence is large.

For setting S3, we compare the variance tendency of ESS/N with N in different situations. In the four cases with different mean value μ, it is observed that ESS/N approach some constants for all ESS estimation method when $N > 500$, which is consistent with the theoretical value of ESS. This result confirms the stability of E-MIM. That is, the performance of E-MIM is irrelevant of sample size N. Moreover, in all of the four cases, E-MIM with $\alpha = -0.5$ approximates the real ESS most accurately. When $\mu = 1.5$ or $\mu = 3$, all ESS estimation functions perform well except $Per_N(\bar{\mathbf{w}})$.

C. APPLICATION IN PARTICLE SAMPLING

In this section, we apply E-MIM to the particle filter, for which the widely used Bearing-Only Tracking is considered [1], [37]. In this model, the target moves within x-y plane

TABLE 1 **Average MSE Under Different ESS Functions**

Criterion	E-MIM	$P_N^{(2)}(\bar{\mathbf{w}})$	$D_N^{(\infty)}(\bar{\mathbf{w}})$
RMSE	0.3817	0.3990	0.4505

according to the following rules

$$x_t = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix} x_{t-1} + \begin{bmatrix} T^2/2 & 0 \\ T & 0 \\ 0 & T^2/2 \\ 0 & T \end{bmatrix} \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix}. \quad (39)$$

The status of the target is described by $x_t = [x_{1,t}, x_{2,t}, x_{3,t}, x_{4,t}]$, in which $x_{1,t}$ and $x_{3,t}$ represent the positions while $x_{2,t}$ and $x_{4,t}$ represent the velocities, in the x direction and the y direction. $v_{1,t}$ and $v_{2,t}$ are independent Gaussian noises, i.e., $v_{1,t} \sim N(0, \sigma_{v1}^2)$, $v_{2,t} \sim N(0, \sigma_{v2}^2)$. T is the sampling interval. The radar is located at the origin and can only observe the target bearing

$$\theta_t = \arctan(x_{3,t}/x_{1,t}) + w_t, \quad (40)$$

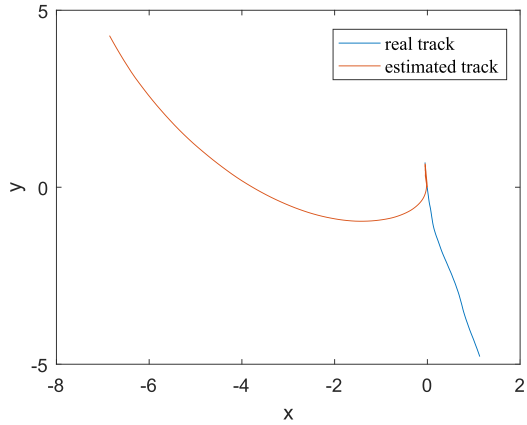
in which, w_t is also independent Gaussian noise, i.e., $w_t \sim N(0, \sigma_w^2)$.

The parameter is set as follows, $T = 1$, $\sigma_{v1}^2 = \sigma_{v2}^2 = 0.001^2$, $\sigma_w^2 = 0.005^2$, the initial state is $x_1 = (-0.05, 0.001, 0.7, -0.055)^T$. The prior distribution for particle filtering is Gaussian distribution, with mean $\hat{x} = (0, 0, 0.4, -0.05)^T$ and covariance matrix $M = \mathrm{diag}(0.5^2, 0.005^2, 0.3^2, 0.01^2)$. For simplicity, we adopted a standard particle filter. The propagation function is chosen as $q(x_t \mid x_{t-1,i}, \theta_t) = p(x_t \mid x_{t-1,i})$, i.e., $x_{t,i} \sim p(x_t|x_{t-1,i})$. According to the law of large numbers, any resampling schedule will provide good performance if the sample size N is big enough. For a relatively small sample size, it is necessary to consider how to do the resampling.

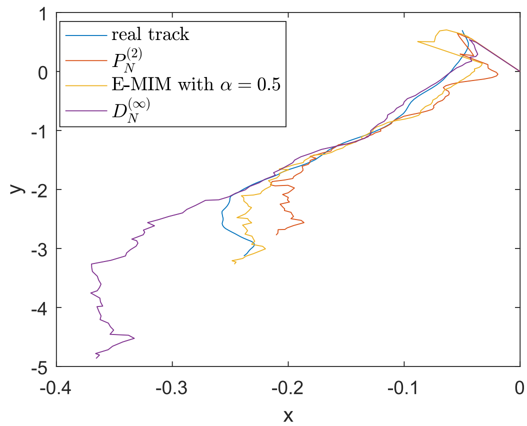
Since σ_w^2 is relatively small, the particle deviates from the real direction a litte will be assigned quite low weights, though it could have relatively accurate velocity. So, we should pay more attention to the low-weighted particles and choose a positive α as 0.5. We set $N = 2000$ and implement E-MIM with $\alpha = 0.5$, as well as $P_N^{(2)}(\bar{\mathbf{w}})$, i.e., E-MIM with $\alpha = 0$, and $D_N^{(\infty)}(\bar{\mathbf{w}})$, i.e., E-MIM with $\alpha = -\infty$, respectively for the resampling criterion. We use the same threshold as in section VI, i.e., $\varepsilon = 0.62, 0.5, 0.125$. We set the time for tracking as $T_t = 100$, and repeat the experiments independently for 2000 times to compare the mean square error of tracking results. The average MSE of the estimation is calculated as follows

$$MSE = \frac{1}{T_t} \sum_{n=1}^{T_t} \left[(\hat{x}_{1,t} - x_{1,t})^2 + (\hat{x}_{3,t} - x_{3,t})^2 \right]. \quad (41)$$

The MSE of E-MIM, $P_N^{(2)}(\bar{\mathbf{w}})$ and $D_N^{(\infty)}(\bar{\mathbf{w}})$ is shown in Table 1, from which we can observe that E-MIM provides the lowest average MSE. Fig. 7(a) shows when the filter loses target lock, the estimated track is far away from the real one, and such extreme case should be abandoned. Fig. 7(b) shows E-MIM manifests better tracking ability than $P_N^{(2)}(\bar{\mathbf{w}})$ and $D_N^{(\infty)}(\bar{\mathbf{w}})$. We also compare how MSE changes with time. In



(a) When losing the track



(b) When tracking correctly

FIGURE 7. **The cases of losing the target lock and tracking successfully.**

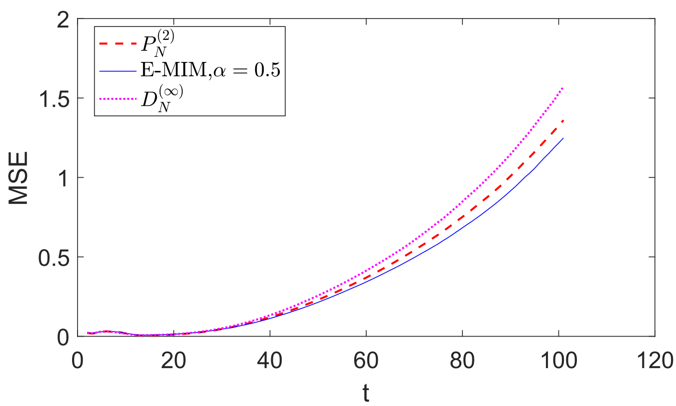


FIGURE 8. **MSE versus simulation time.**

Fig. 8, E-MIM with appropriate α tends to help decrease the divergence in tracking.

VIII. CONCLUSION

In this paper, we proposed a MIM based ESS function, which is referred to as the E-MIM. First, we proved that E-MIM satisfies the five requirement of G-ESS, and showed that E-MIM is monotonically non-decreasing with the importance coefficient α. We also considered the relationship between E-MIM and other widely used ESS functions, and showed that $P_N^{(2)}(\bar{\mathbf{w}})$ is a special case of E-MIM with $\alpha = 0$, $D_N^{(\infty)}(\bar{\mathbf{w}})$ is a special form of E-MIM with $\alpha = -\infty$. Inspired by the connection between the perplexity and the K-L divergence, we then proposed the MIM divergence to measure the divergence between two probability distributions. Based on these results, we further explored the influence of the coefficient α upon E-MIM. Specifically, when $\alpha > 0$, particles with low weights $\bar{w}_n < 1/N$ contributes more to the approximation of ESS while when $\alpha < 0$, particles with high weights $\bar{w}_n > 1/N$ contributes more to the approximation of ESS. This property provides useful guidance for the selection of α. That is, we can choose a negative value of α if we prefer to focusing on particles with high weights, while we select a positive value of α if low weight particles need more attention. Thus, we proposed a threshold selection strategy of E-MIM for resampling based on the distribution of E-MIM. Finally, by adjusting its coefficient, we confirmed that E-MIM provides better approximations for the theoretical ESS through numerical simulations.

APPENDIX A PROOF OF THE PROPERIES OF E-MIM

In this section, we give proofs of the property A-F of E-MIM.

A. PROOF OF PROPERTY A

Proof:

$$\begin{aligned} E_N(\bar{\mathbf{w}}', \alpha) &= -\frac{N\alpha}{\log \sum_{n=1}^N \bar{w}_{j_n} \exp(-N\alpha \bar{w}_{j_n})} \\ &= -\frac{N\alpha}{\log \sum_{n=1}^N \bar{w}_n \exp(-N\alpha \bar{w}_n)} \\ &= E_N(\bar{\mathbf{w}}, \alpha). \end{aligned} \tag{42}$$

■

B. PROOF OF PROPERTY B

Proof: First, we show that when $\alpha < 1$ and $\bar{\mathbf{w}}^* = [1/N, \ldots, 1/N]$, E-MIM reaches N.

$$\begin{aligned} E_N(\bar{\mathbf{w}}^*, \alpha) &= -\frac{N\alpha}{\log \sum_{n=1}^N \frac{1}{N} \exp(-N\alpha \frac{1}{N})} \\ &= -\frac{N\alpha}{-\alpha} \\ &= N. \end{aligned} \tag{43}$$

Second, we show that N is the maximum value.

1) WHEN $0 < \alpha < 1$

To prove

$$E_N(\bar{\mathbf{w}}, \alpha) = -\frac{N\alpha}{\log \sum_{n=1}^N \bar{w}_n \exp(-N\alpha \bar{w}_n)} \le N, \tag{44}$$

is equivalent to prove

$$\sum_{n=1}^N \bar{w}_n \exp(-N\alpha \bar{w}_n) \le \exp(-\alpha). \tag{45}$$

Consider the following function

$$f(x) = x \exp(-N\alpha x), \alpha > 0, x \in [0, 1]. \tag{46}$$

The derivative of $f(x)$ is

$$f'(x) = (1 - N\alpha x) \exp(-N\alpha x). \tag{47}$$

Hence

$$f'\left(\frac{1}{N}\right) = (1 - \alpha) \exp(-\alpha). \tag{48}$$

The tangent line at the point $(\frac{1}{N}, \frac{1}{N} \exp(-\alpha))$ is

$$g(x) = \left((1 - \alpha)x + \frac{\alpha}{N}\right) \exp(-\alpha). \tag{49}$$

The second derivative of $f(x)$ is

$$f''(x) = -N\alpha(2 - N\alpha x) \exp(-N\alpha). \tag{50}$$

Hence, when $x < \frac{2}{N\alpha}$, $f(x)$ is upper convex and when $x > \frac{2}{N\alpha}$, $f(x)$ is lower convex.

Since $\alpha < 1$, we have $\frac{2}{N\alpha} > \frac{1}{N}$, thus, we obtain $g(x) \ge f(x)$, when $x < \frac{2}{N\alpha}$. And for $x > \frac{2}{N\alpha}$ we only need to prove

$$g(1) > f(1). \tag{51}$$

which is equivalent to prove

$$\begin{aligned} \left(1 - \alpha + \frac{\alpha}{N}\right) \exp(-\alpha) &> \exp(-N\alpha) \\ 1 - \alpha + \frac{\alpha}{N} &> \exp(-(N-1)\alpha). \end{aligned} \tag{52}$$

Since $\exp(-(N-1)\alpha)$ is a lower convex function of α, we only need to check (52) holds when $\alpha = 0$ and $\alpha = 1$, which is ture.

In summary,

$$f(x) \le g(x), x \ge 0. \tag{53}$$

the equality holds if and only if $x = \frac{1}{N}$.

Hence,

$$\begin{aligned} \sum_{n=1}^N \bar{w}_n \exp(-N\alpha \bar{w}_n) &\le \sum_{n=1}^N \left((1-\alpha)\bar{w}_n + \frac{\alpha}{N}\right) \exp(-\alpha) \\ &= \exp(-\alpha). \end{aligned} \tag{54}$$

the equality holds if and only if $\bar{w}_n = \frac{1}{N}$ for $n = 1, 2, \ldots, N$.

Hence, when $0 < \alpha < 1$

$$E_N(\bar{\mathbf{w}}*, \alpha) = N \ge E_N(\bar{\mathbf{w}}, \alpha), \tag{55}$$

the equality holds if and only if $\bar{\mathbf{w}} = \bar{\mathbf{w}}^*$.

2) WHEN $\alpha = 0$

$$E_N(\bar{\mathbf{w}}, \alpha) = \frac{1}{\sum_{n=1}^N \bar{w}_n^2} \le \frac{N}{(\sum_{n=1}^N \bar{w}_n)^2} = N = E_N(\bar{\mathbf{w}}^*, \alpha), \tag{56}$$

The equality holds if and only if $\bar{\mathbf{w}} = \bar{\mathbf{w}}^*$.

3) WHEN $\alpha < 0$

To prove

$$E_N(\bar{\mathbf{w}}, \alpha) = -\frac{N\alpha}{\log \sum_{n=1}^{N} \bar{w}_n \exp(-N\alpha \bar{w}_n)} \le N, \tag{57}$$

is equivalent to prove

$$\sum_{n=1}^{N} \bar{w}_n \exp(-N\alpha \bar{w}_n) \ge \exp(-\alpha). \tag{58}$$

Consider the function

$$f(x) = x \exp(-N\alpha x), \alpha < 0, x \in [0, 1]. \tag{59}$$

The derivative of $f(x)$ is

$$f'(x) = (1 - N\alpha x) \exp(-N\alpha x). \tag{60}$$

Hence

$$f'\left(\frac{1}{N}\right) = (1 - \alpha) \exp(-\alpha). \tag{61}$$

The tangent line at the point $(\frac{1}{N}, \frac{1}{N} \exp(-\alpha))$ is

$$g(x) = \left((1 - \alpha)x + \frac{\alpha}{N}\right) \exp(-\alpha). \tag{62}$$

The second derivative of $f(x)$ is

$$f''(x) = -N\alpha(2 - N\alpha x) \exp(-N\alpha), \tag{63}$$

and $f''(x)$ is positive when $x > 0$, so $f(x)$ is lower convex. Thus, we obtain

$$f(x) \ge g(x), x \ge 0, \tag{64}$$

the equality holds if and only if $x = \frac{1}{N}$.

Hence,

$$\sum_{n=1}^{N} \bar{w}_n \exp(-N\alpha \bar{w}_n) \ge \sum_{n=1}^{N} \left((1 - \alpha)\bar{w}_n + \frac{\alpha}{N}\right) \exp(-\alpha) = \exp(-\alpha), \tag{65}$$

the equality holds if and only if $\bar{w}_n = \frac{1}{N}$ for $n = 1, 2, \ldots, N$.

Hence, when $\alpha < 0$

$$E_N(\bar{\mathbf{w}}^*, \alpha) = N \ge E_N(\bar{\mathbf{w}}, \alpha), \tag{66}$$

the equality holds if and only if $\bar{\mathbf{w}} = \bar{\mathbf{w}}^*$.

In summary, if $\alpha < 1$

$$E_N(\bar{\mathbf{w}}^*, \alpha) = N \ge E_N(\bar{\mathbf{w}}, \alpha), \tag{67}$$

the equality holds if and only if $\bar{\mathbf{w}} = \bar{\mathbf{w}}^*$. ■

C. PROOF OF PROPERTY C

Proof: The proof is divided in three parts.

1) WHEN $\alpha > 0$

$$\sum_{n=1}^{N} \bar{w}_n \exp(-N\alpha \bar{w}_n) \ge \sum_{n=1}^{N} \bar{w}_n \exp(-N\alpha) = \exp(-N\alpha). \tag{68}$$

Hence,

$$\log \sum_{n=1}^{N} \bar{w}_n \exp(-N\alpha \bar{w}_n) \ge -N\alpha. \tag{69}$$

Hence,

$$E_N(\bar{\mathbf{w}}, \alpha) = -\frac{N\alpha}{\log \sum_{n=1}^{N} \bar{w}_n \exp(-N\alpha \bar{w}_n)} \ge 1, \tag{70}$$

the equality holds if and only if $\bar{\mathbf{w}} = \bar{\mathbf{w}}^{(j)} = [\bar{w}_1 = 0, \ldots, \bar{w}_j = 1, \ldots, \bar{w}_n = 0]$, $j = 1, 2, \ldots, N$.

2) WHEN $\alpha = 0$

$$E_N(\bar{\mathbf{w}}, \alpha) = \frac{1}{\sum_{n=1}^{N} \bar{w}_n^2} \ge \frac{1}{\sum_{n=1}^{N} \bar{w}_n} = 1, \tag{71}$$

the equality holds if and only if $\bar{\mathbf{w}} = \bar{\mathbf{w}}^{(j)} = [\bar{w}_1 = 0, \ldots, \bar{w}_j = 1, \ldots, \bar{w}_n = 0]$, $j = 1, 2, \ldots, N$.

3) WHEN $\alpha < 0$

$$\sum_{n=1}^{N} \bar{w}_n \exp(-N\alpha \bar{w}_n) \le \sum_{n=1}^{N} \bar{w}_n \exp(-N\alpha) = \exp(-N\alpha). \tag{72}$$

Hence,

$$\log \sum_{n=1}^{N} \bar{w}_n \exp(-N\alpha \bar{w}_n) \le -N\alpha. \tag{73}$$

Hence,

$$E_N(\bar{\mathbf{w}}, \alpha) = -\frac{N\alpha}{\log \sum_{n=1}^{N} \bar{w}_n \exp(-N\alpha \bar{w}_n)} \ge 1, \tag{74}$$

the equality holds if and only if $\bar{\mathbf{w}} = \bar{\mathbf{w}}^{(j)} = [\bar{w}_1 = 0, \ldots, \bar{w}_j = 1, \ldots, \bar{w}_n = 0]$, $j = 1, 2, \ldots, N$.

In summary, for any $\alpha \in R$

$$E_N(\bar{\mathbf{w}}, \alpha) \ge 1, \tag{75}$$

the equality holds if and only if $\bar{\mathbf{w}} = \bar{\mathbf{w}}^{(j)} = [\bar{w}_1 = 0, \ldots, \bar{w}_j = 1, \ldots, \bar{w}_n = 0]$, $j = 1, 2, \ldots, N$.

D. PROOF OF PROPERTY D

The proof is included in the proof of property B and C.

E. PROOF OF PROPERTY E

Proof: Since

$$\bar{\mathbf{v}} = \frac{1}{M} \underbrace{[\bar{\mathbf{w}}, \ldots, \bar{\mathbf{w}}]}_{M times}, \tag{76}$$

we have

$$\begin{aligned} E_{MN}(\bar{\mathbf{v}}, \alpha) &= -\frac{MN\alpha}{\log \sum_{n=1}^{MN} \bar{v}_n \exp(-MN\alpha \bar{v}_n)} \\ &= -\frac{MN\alpha}{\log M \sum_{n=1}^{N} \frac{1}{M} \bar{w}_n \exp(-MN\alpha \frac{1}{M} \bar{w}_n)} \\ &= -\frac{MN\alpha}{\log \sum_{n=1}^{N} \bar{w}_n \exp(-N\alpha \bar{w}_n)} \\ &= M E_N(\bar{\mathbf{w}}, \alpha), \end{aligned} \tag{77}$$

i.e.,

$$\frac{E_N(\bar{\mathbf{w}}, \alpha)}{N} = \frac{E_{MN}(\bar{\mathbf{v}}, \alpha)}{MN}. \tag{78}$$

F. PROOF OF PROPERTY F

Proof:

$$E_N(\bar{\mathbf{w}}, \alpha) = -\frac{N\alpha}{\log \sum_{n=1}^{N} \bar{w}_n \exp(-N\alpha\bar{w}_n)}. \quad (79)$$

We denote

$$\alpha' = N\alpha. \quad (80)$$

Then we need to prove that

$$E_N(\bar{\mathbf{w}}, \alpha') = -\frac{\alpha'}{\log \sum_{n=1}^{N} \bar{w}_n \exp(-\alpha'\bar{w}_n)}, \quad (81)$$

is monotonically nondecreasing with α'. Since $E_N(\bar{\mathbf{w}}, \alpha') \geq 1$, we need to prove

$$L(\bar{\mathbf{w}}, \alpha') = \frac{1}{E_N(\bar{\mathbf{w}}, \alpha')} = -\frac{\log \sum_{n=1}^{N} \bar{w}_n \exp(-\alpha'\bar{w}_n)}{\alpha'}, \quad (82)$$

is monotonically nonincreasing with α'.

The derivative of $L(\bar{\mathbf{w}}, \alpha')$ with respect to α' can be expressed as follows

$$\begin{aligned}
&\frac{dL(\bar{\mathbf{w}}, \alpha')}{d\alpha'} \\
&= \frac{1}{\alpha'^2}\left[\alpha' \frac{\sum_{n=1}^{N} \bar{w}_n^2 e^{-\alpha'\bar{w}_n}}{\sum_{n=1}^{N} \bar{w}_n e^{-\alpha'\bar{w}_n}} + \log \sum_{n=1}^{N} \bar{w}_n e^{-\alpha'\bar{w}_n}\right] \\
&= \frac{1}{\alpha'^2}\left[\alpha' \sum_{n=1}^{N} \frac{\bar{w}_n e^{-\alpha'\bar{w}_n}}{\sum_{n=1}^{N} \bar{w}_n e^{-\alpha'\bar{w}_n}} \bar{w}_n - \log \frac{\sum_{n=1}^{N} \bar{w}_n}{\sum_{n=1}^{N} \bar{w}_n e^{-\alpha'\bar{w}_n}}\right] \\
&= \frac{1}{\alpha'^2}\left[\sum_{n=1}^{N} \frac{\bar{w}_n e^{-\alpha'\bar{w}_n}}{\sum_{n=1}^{N} \bar{w}_n e^{-\alpha'\bar{w}_n}} \log(e^{\alpha'\bar{w}_n}) \right. \\
&\qquad \left. - \log \sum_{n=1}^{N} \frac{\bar{w}_n e^{-\alpha'\bar{w}_n}}{\sum_{n=1}^{N} \bar{w}_n e^{-\alpha'\bar{w}_n}} e^{\alpha'\bar{w}_n}\right]. \quad (83)
\end{aligned}$$

Since $\log(x)$ is an concave function, according to Jensen's inequality, we have

$$\begin{aligned}
&\sum_{n=1}^{N} \frac{\bar{w}_n e^{-\alpha'\bar{w}_n}}{\sum_{n=1}^{N} \bar{w}_n e^{-\alpha'\bar{w}_n}} \log(e^{\alpha'\bar{w}_n}) \\
&\qquad \leq \log \sum_{n=1}^{N} \frac{\bar{w}_n e^{-\alpha'\bar{w}_n}}{\sum_{n=1}^{N} \bar{w}_n e^{-\alpha'\bar{w}_n}} e^{\alpha'\bar{w}_n}, \quad (84)
\end{aligned}$$

the equality holds if and only if all the nonzero elements in $\bar{\mathbf{w}} = [\bar{w}_1, \ldots, \bar{w}_N]$ are equal.

Thus,

$$\frac{dL(\bar{\mathbf{w}}, \alpha')}{d\alpha'} \leq 0, \quad (85)$$

i.e.,

$$\frac{dE_N(\bar{\mathbf{w}}, \alpha)}{d\alpha} \geq 0, \quad (86)$$

the equality holds if and only if all the nonzero elements in $\bar{\mathbf{w}} = [\bar{w}_1, \ldots, \bar{w}_N]$ are equal. ■

The proof is completed.

APPENDIX B PROOF OF THE NON-NEGATIVITY OF MIM DIVERGENCE

Proof: The MIM divergence between distribution $\bar{\pi}(x)$ and $q(x)$ with the importance coefficient α is defined as,

$$\begin{aligned}
&D_{MIM}(\bar{\pi}(x) \mid q(x), \alpha) \\
&\quad = -\frac{1}{\alpha} \log \int \bar{\pi}(x) \exp\left(\alpha\left(1 - \frac{\bar{\pi}(x)}{q(x)}\right)\right) dx. \quad (87)
\end{aligned}$$

1) WHEN $\alpha > 0$

To prove

$$D_{MIM}(\bar{\pi}(x) \mid q(x), \alpha) \geq 0, \quad (88)$$

is equivalent to prove

$$\int \bar{\pi}(x) \exp\left(\alpha\left(1 - \frac{\bar{\pi}(x)}{q(x)}\right)\right) dx \leq 1. \quad (89)$$

Consider the function

$$f(x) = \exp\left(\alpha\left(1 - \frac{1}{x}\right)\right). \quad (90)$$

The derivative of $f(x)$ is

$$f'(x) = \frac{\alpha}{x^2} \exp\left(\alpha\left(1 - \frac{1}{x}\right)\right). \quad (91)$$

The tangent line at point (1,1) is

$$g(x) = \alpha x + (1 - \alpha). \quad (92)$$

The second derivative of $f(x)$ is

$$f''(x) = \frac{\alpha}{x^4}(\alpha - 2x) \exp\left(\alpha\left(1 - \frac{1}{x}\right)\right). \quad (93)$$

Thus, $f(x)$ is lower convex when $x < \alpha/2$, while $f(x)$ is upper convex when $x > \alpha/2$. So to make $f(x) \leq g(x)$ hold when $x > 0$, we only need

$$\begin{cases} \frac{\alpha}{2} < 1 \\ f(0) < g(0) \end{cases}. \quad (94)$$

So we need

$$\alpha < 1. \quad (95)$$

Then we have

$$f(x) \leq g(x), x \geq 0, \quad (96)$$

i.e.,

$$\exp\left(\alpha(1 - \frac{1}{x})\right) \leq \alpha x + (1 - \alpha), x \geq 0. \quad (97)$$

The equality holds if and only if $x = 1$.

Hence,

$$\begin{aligned}
&\int \bar{\pi}(x) \exp\left(\alpha\left(1 - \frac{\bar{\pi}(x)}{q(x)}\right)\right) dx \\
&\leq \int \bar{\pi}(x) \left(\alpha \frac{q(x)}{\bar{\pi}(x)} + (1 - \alpha)\right) dx \\
&= \alpha \int \bar{\pi}(x) dx + (1 - \alpha) \int q(x) dx \\
&= 1,
\end{aligned} \quad (98)$$

which means that the equality

$$D_{MIM}(\bar{\pi}(x) \mid q(x), \alpha) \geq 0, \quad (99)$$

holds if and only if $\frac{\bar{\pi}(x)}{q(x)} = 1$, i.e., $\bar{\pi}(x) = q(x)$.

2) WHEN $\alpha = 0$

$$\lim_{\alpha\to 0} -\frac{1}{\alpha}\log\int \bar{\pi}(x)\exp\left(\alpha\left(1-\frac{\bar{\pi}(x)}{q(x)}\right)\right)dx = \int \frac{\bar{\pi}^2(x)}{q(x)}dx - 1, \tag{100}$$

where the first term on the right side of the equation is the so-called second order Renyi divergence.

Since $1 - x$ is the tangent line of $1/x - 1$ at point (1,1), we obtain

$$\int \frac{\bar{\pi}^2(x)}{q(x)}dx - 1 = \int \bar{\pi}(x)\left(\frac{\bar{\pi}(x)}{q(x)} - 1\right)dx \geq \int \bar{\pi}(x)\left(1 - \frac{q(x)}{\bar{\pi}(x)}\right)dx = 0, \tag{101}$$

the equality holds if and only if $\frac{\bar{\pi}(x)}{q(x)} = 1$, i.e., $\bar{\pi}(x) = q(x)$.

3) WHEN $\alpha < 0$

To prove

$$D_{MIM}(\bar{\pi}(x) \mid q(x), \alpha) \geq 0, \tag{102}$$

is equivalent to prove

$$\int \bar{\pi}(x)\exp\left(\alpha\left(1-\frac{\bar{\pi}(x)}{q(x)}\right)\right)dx \geq 1. \tag{103}$$

Consider the function

$$f(x) = \exp\left(\alpha\left(1-\frac{1}{x}\right)\right). \tag{104}$$

The derivative of $f(x)$ is

$$f'(x) = \frac{\alpha}{x^2}\exp\left(\alpha\left(1-\frac{1}{x}\right)\right). \tag{105}$$

The tangent line at point (1,1) is

$$g(x) = \alpha x + (1-\alpha). \tag{106}$$

The second derivative of $f(x)$ is

$$f''(x) = \frac{\alpha}{x^4}(\alpha - 2x)\exp\left(\alpha\left(1-\frac{1}{x}\right)\right). \tag{107}$$

Thus,we have $f''(x) > 0$ when $x > 0$, i.e., $f(x)$ is lower convex when $x > 0$. Hence, when $x > 0$,

$$f(x) \geq g(x), \tag{108}$$

the equality holds if and only if $x = 1$.

Hence,

$$\begin{aligned}&\int \bar{\pi}(x)\exp\left(\alpha\left(1-\frac{\bar{\pi}(x)}{q(x)}\right)\right)dx\\ &\geq \int \bar{\pi}(x)\left(\alpha\frac{q(x)}{\bar{\pi}(x)} + (1-\alpha)\right)dx\\ &= \alpha\int \bar{\pi}(x)dx + (1-\alpha)\int q(x)dx\\ &= 1,\end{aligned} \tag{109}$$

which means that the equality

$$D_{MIM}(\bar{\pi}(x) \mid q(x), \alpha) \geq 0, \tag{110}$$

holds if and only if $\frac{\bar{\pi}(x)}{q(x)} = 1$, i.e., $\bar{\pi}(x) = q(x)$.

The proof is completed. ∎

REFERENCES

[1] A. Doucet, N. Freitas, and N. Gordon, "An introduction to sequential Monte Carlo methods,," in *Sequential Monte Carlo Methods in Practice*, New York, NY, USA: Springer, 2001, ch. 1, pp. 3–14.

[2] P. Del Moral, A. Doucet, and A. Jasra, "Sequential Monte Carlo samplers," *J. Roy. Stat. Soc. Ser. B-Stat. Methodol.*, vol. 68, no. 3, pp. 411–436, 2006.

[3] C. P. Robert and G. Casella, *Introducing Monte Carlo Methods With R*. Berlin, Germany: Springer New York, 2010.

[4] G. Qian, L. Ning, and R. Huggins, "Using capture-recapture data and hybrid Monte Carlo sampling to estimate an animal population affected by an environmental catastrophe," *Comput. Statist. Data Anal.*, vol. 55, no. 1, pp. 655–666, 2011.

[5] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[6] Z. Yang and X. Wang, "A sequential Monte Carlo blind receiver for OFDM systems in frequency-selective fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 271–280, Feb. 2002.

[7] L. Zuo, R. Niu, and P. K. Varshney, "Conditional posterior Cramérao lower bounds for nonlinear sequential Bayesian estimation," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 1–14, Dec. 2011.

[8] J. R. Blevins, "Sequential Monte Carlo methods for estimating dynamic microeconomic models," *J. Appl. Econometrics*, vol. 31, no. 5, pp. 773–804, 2016.

[9] X. Kong, K. Cui, and H. Jia, "Capacity credit evaluation of wind power with sequential Monte Carlo method," in *Proc. Int. Conf. E-Product E-Serv. E-Entertainment*, 2010, pp. 1–4.

[10] R. S. Targino, G. W. Peters, and P. V. Shevchenko, "Sequential Monte Carlo samplers for capital allocation under copula-dependent risk models," *Insurance Math. Econ.*, vol. 61, pp. 206–226, Mar. 2015.

[11] T. Li, M. Bolic, and P. M. Djuric, "Resampling methods for particle filtering: Classification, implementation, and strategies," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 70–86, May 2015.

[12] L. Martino, V. Elvira, and F. Louzada, "Effective sample size for importance sampling based on discrepancy measures," *Signal Process.*, vol. 131, pp. 386–401, Feb. 2017.

[13] P. Djuric *et al.*, "Particle filtering," *IEEE Signal Process. Mag.*, vol. 20, no. 5, pp. 19–38, Sep. 2003.

[14] D. Crisan and A. Doucet, "Convergence of sequential Monte Carlo methods," Signal Process. Group, Dept. Eng. Univ. Cambridge, Tech. Rep. CUEDIF-INFENGrrR38., 2000.

[15] S. Chatterjee and P. Diaconis, "The sample size required in importance sampling," *Statistics*, vol. 28, pp. 1099–1135, 2018.

[16] J. S. Liu, "Metropolized independent sampling with comparisons to rejection sampling and importance sampling," *Statist. Comput.*, vol. 6, no. 2, pp. 113–119, 1996.

[17] J. S. Liu, and R. Chen, "Blind deconvolution via sequential imputations," *J. Amer. Statist. Assoc.*, vol. 90, no. 430, pp. 567–576, Jun. 1995.

[18] D. Gamerman and H. F. Lopes, *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Boca Raton, FL, USA: CRC Press, 2006.

[19] L. Martino, V. Elvira, D. Luengo, and J. Corander, "An adaptive population importance sampler: Learning from uncertainty," *IEEE Trans. Signal Process.*, vol. 63, no. 16, pp. 4422–4437, Aug. 2015.

[20] Z. Chen *et al.*, "Bayesian filtering: From Kalman filters to particle filters, and beyond," *Statistics*, vol. 182, no. 1, pp. 1–69, 2003.

[21] T. Li, M. Bolic, and P. M. Djuric, "Resampling methods for particle filtering: Classification, implementation, and strategies," *IEEE Signal Process. Mag.*, vol. 32, no. 3, pp. 70–86, May 2015.

[22] V. Elvira, L. Martino, and C. P. Robert, "Rethinking the effective sample size," Sep. 2018, *arXiv:1809.04129*.

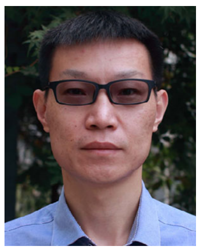
[23] T. Cover, *Elements of Information Theory.* Hoboken, NJ, USA: Wiley, 1999.

[24] O. Cappé, R. Douc, A. Guillin, J. M. Marin, and C. P. Robert, "Adaptive importance sampling in general mixture classes," *Statist. Comput.*, vol. 18, no. 4, pp. 447–459, 2008.
[25] N. Celik and Y.-J. Son, "State estimation of a shop floor using improved resampling rules for particle filtering," *Int. J. Prod. Econ.,* vol. 134, no. 1, pp. 224–237, 2011.
[26] J. H. Huggins and D. M. Roy, "Sequential Monte Carlo as approximate sampling: Bounds, adaptive resampling via ∞-ESS, and an application to Particle Gibbs," Mar. 2015, *arXiv:1503.00966*.
[27] P. D. Moral, A. Doucet, and A. Jasra, "On adaptive resampling procedures for sequential Monte Carlo methods," *Bernoulli*, vol. 18, no. 1, pp. 252–278, 2012.
[28] P. Fan, Y. Dong, J. Lu, and S. Liu, "Message importance measure and its application to minority subset detection in big data," in *Proc. IEEE Conf. Globecom Workshops*, Washington, DC, USA, Dec. 2016, p. 5.
[29] C. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.
[30] A. Rényi *et al.*, "On measures of entropy and information," in *Proc. 4th Berkeley Symp. Math. Statist. Probability*, *Contributions Theory Statist.*, 1961.
[31] S. Rui, S. Liu, Y. Dong, and P. Fan, "Focusing on a probability element: Parameter selection of message importance measure in big data," in *Proc. ICC IEEE Int. Conf. Commun.*, 2017, pp. 1–6.
[32] S. Liu, R. She, S. Wan, P. Fan, and Y. Dong, "A switch to the concern of user: Importance coefficient in utility distribution and message importance measure," in *Proc. 14th Int. Wireless Commun. Mobile Comput. Conf.*, 2018, pp. 1362–1367.
[33] A. Kong, J. S. Liu, and W. H. Wong, "Sequential imputations and Bayesian missing data problems," *J. Amer. Stat. Assoc.*, vol. 89, no. 425, pp. 278–288, Mar. 1994.
[34] S. Q. Hu and Z. L. Jing, "Overview of particle filter algorithm," *Control Decis.*, vol. 20, no. 4, p. 361, 2005.
[35] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 1951.
[36] N. A. Smith and R. W. Tromble, "Sampling uniformly from the unit simplex," Johns Hopkins Univ., Tech. Rep, vol. 29, pp. 1–6, 2004. [Online]. Available: http://www.cs.cmu.edu/~nasmith/papers/smith+tromble.tr04.pdf
[37] J. Carpenter, P. Clifford, and P. Fearnhead, "Improved particle filter for nonlinear problems," *Proc. IEE Radar Sonar Navigation*, vol. 146, no. 1, pp. 2–7, Feb. 1999.

ZHEFAN LI received the B.S. degree in 2020 from the Department of Electronic Engineering, Tsinghua University, Beijing, China, where he is currently working toward the Ph.D. degree. His current research interests include wireless communications in big data analysis, federated learning, and information theory.

PINGYI FAN (Senior Member, IEEE) received the B.S. degree from the Department of Mathematics, Hebei University, Baoding, China, in 1985, the M.S. degree from Nankai University, Tianjin, China, in 1990, and the Ph.D. degree from the Department of Electronic Engineering, Tsinghua University, Beijing, China, in 1994. He is currently a Professor with the Department of Electronic Engineering, Tsinghua University. From August 1997 to March 1998, he visited the Hong Kong University of Science and Technology as a Research Associate. From May 1998 to October 1999, he visited the University of Delaware, Newark, DE, USA, as a Research Fellow. In March 2005, he visited the NICT of Japan as a Visiting Professor. From June 2005 to May 2019, he visited the Hong Kong University of Science and Technology for many times and from July 2011 to September 2011, he was a Visiting Professor with the Institute of Network Coding, Chinese University of Hong Kong. His main research interests include B5G technology in wireless communications, machine learning and AI in wireless communications, information theory in big data analysis, network coding and network information theory, etc.

Dr. Fan is an overseas member of IEICE. He has attended to organize many international conferences including as TPC Co-Chair of Chinacom 2020, IEEE WCNIS 2010, and TPC member of IEEE ICC, Globecom, WCNC, VTC, Infocom, etc. He was an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, *Inderscience International Journal of Ad Hoc and Ubiquitous Computing* and *Wiley Journal of Wireless Communication and Mobile Computing*, *MDPI Electronics*, etc. He is also a Reviewer of more than 30 international journals including 24 IEEE Journals and eight EURASIP Journals. He has received some academic awards, including the IEEE ICC20, TAOS20, Globecom14, WCNC08 Best Paper Awards, ACM IWCMC10 Best Paper Award, and IEEE ComSoc Excellent Editor Award for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS in 2009.



YUNQUAN DONG (Member, IEEE) received the B.S. degree in electronic and information engineering from Qingdao University, Qingdao, China, in 2005, the M.S. degree in communication and information systems from the Beijing University of Posts and Telecommunications, Beijing, China, in 2008, and the Ph.D. degree in communication and information engineering from Tsinghua University, Beijing, China, in 2014.

From 2015 to 2016, he was a BK Assistant Professor with the Department of Electrical and Computer Engineering, Seoul National University, Seoul, South Korea. He is currently a Professor with the School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing, China. His research interests include the performance evaluations and performance optimizations of wireless networks, with recent focus on age of information, and ubiquitous sensing.

Dr. Dong was the recipient of the Best Paper Award of IEEE International Conference on Communication Technology in 2011, the National Scholarship for Postgraduates from China's Ministry of Education in 2013, and the Young Star of Information Theory Award from the China's Information Theory Society in 2014. He was selected as an Exemplary Reviewer of the IEEE WIRELESS COMMUNICATIONS LETTERS in 2017.