Information Fusion via Importance Sampling

Augustin A. Saucan, *Member, IEEE*, Víctor Elvira, *Senior Member, IEEE*,
Pramod K. Varshney, *Life Fellow, IEEE*, and Moe Z. Win, *Fellow, IEEE*

***Abstract*—Information fusion is a procedure that merges information locally contained at the nodes of a network. Of high interest in the field of distributed estimation is the fusion of local probability distributions via a weighted geometrical average criterion. In numerous practical settings, the local distributions are only known through particle approximations, i.e., sets of samples with associated weights, such as obtained via importance sampling (IS) methods. Thus, prohibiting any closed-form solution to the aforementioned fusion problem. This article proposes a family of IS methods—called particle geometric–average fusion (PGAF)—that lead to consistent estimators for the geometrically-averaged density. The advantages of the proposed methods are threefold. First, the methods are agnostic of the mechanisms used to generate the local particle sets and, therefore, allow for the fusion of heterogeneous nodes. Second, consistency of estimators is guaranteed under generic conditions when the agents use IS-generated particles. Third, a low-communication overhead and agent privacy are achieved since local observations are not shared with the fusion center. Even more remarkably, for a sub-family of the proposed PGAF methods, the fusion center does not require the knowledge of the local priors used by the nodes. Implementation guidelines for the proposed methods are provided and theoretical results are numerically verified.**

***Index Terms*—Multiple importance sampling, information fusion, Monte Carlo methods, Kullback–Leibler divergence, kernel density estimate.**

I. Introduction

Information fusion generally encompasses various statistical methods employed for merging several sources of information according to some criterion. In engineering, information fusion is widely employed in multi-agent systems—such as sensor networks—for enhanced performance. In particular, complementary multi-modal sensor data is merged in order to enhance detection and tracking performance [1], [2].

In general, distributed inference [3], [4] assumes that each node (or agent) in an inference network constructs a local estimator of an unknown parameter by relying on local observations, and subsequently a fusion criterion is employed at a centralized node to merge these local estimators. Furthermore, in Bayesian settings [5], each node estimates a posterior distribution of the parameter by updating a local prior with local data. In numerous applications, due to non-linear an/or non-Gaussian sensing modalities, closed-form expressions for these local posteriors are unavailable and instead each agent evaluates its local posterior (up to a proportionality constant) at a set of discrete points called particles, leading to an empirical approximation of the true distribution. Sample-based approximations of distributions (i.e., empirical measures) are the basis of particle filters in sequential Monte Carlo [6], [7], and allow the representation of complex non-Gaussian distributions.

The barycenter distribution that minimizes a weighted-average Kullback–Leibler divergence (KLD) from the local distributions is often employed as a fusion criterion in engineering applications, such as, robotics and target tracking [8], [9]. More generally, information fusion is of relevance in many algorithms for networked sensing [10], [11], [12], [13], [14], [15], wireless communications [16], [17], [18], and Internet-of-Things [19] to name a few. Furthermore, the barycenter corresponds to the geometric average of the probability density functions of the agents, resulting in the geometric–average fusion (GAF) rule. The GAF rule leads to a closed-form result for Gaussian densities [8], in which case it is often named covaraince intersection (CI) [8], [9]. Moreover for the case of two agents, this rule is also called Chernoff information fusion in [20]. Decentralized CI was introduced in [21], while additional algorithms and results for distributed information fusion are found in [22].

Particle methods for GAF have seen an increased interest due to the prevalence of non-Gaussian distributions in a myriad of applications, e.g., sensing, finance, and biology to name a few. PGAF methods that resort to Gaussian fitting, as an intermediate step, followed by the CI rule have been proposed in [23], [24]. Gaussian-mixture fitting was employed in [11] to achieve a PGAF method. A PGAF method that does not resort to fitting was reported in [13], [25], [26]. More details about these state-of-the-art methods for PGAF are found in Section S-II of the supplementary material [27].

Manuscript received 15 January 2023; revised 26 May 2023; accepted 11 July 2023. Date of publication 7 August 2023; date of current version 9 April 2024. The fundamental research described in this paper was supported, in part, by the Agence Nationale de la Recherche of France under Grant PISCES ANR-17-CE40-0031-01, by the Leverhulme Research Fellowship under Grant RF-2021-593, by ARL/ARO under Grant W911NF-22-1-0235, by the National Science Foundation under Grant CNS-2148251, and by federal agency and industry partners in the RINGS program. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Gang Wang. *(Corresponding author: Moe Z. Win.)*

Augustin A. Saucan was with the Wireless Information and Network Sciences Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139 USA. He is now with the SAMOVAR Laboratory, Télécom SudParis, Institut Polytechnique de Paris, 91120 Evry, France (e-mail: asaucan@mit.edu).

Víctor Elvira is with the University of Edinburgh, EH8 9YL Edinburgh, U.K. (e-mail: victor.elvira@ed.ac.uk).

Pramod K. Varshney is with the Department of Electrical and Computer Engineering, Syracuse University, Syracuse, NY 13244 USA (e-mail: varshney@syr.edu).

Moe Z. Win is with the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: moewin@mit.edu).

This article has supplementary downloadable material available at https://doi.org/10.1109/TSIPN.2023.3299512, provided by the authors.

Digital Object Identifier 10.1109/TSIPN.2023.3299512

In this work, we propose several novel PGAF methods by relying on multiple importance sampling (MIS) [28] principles and kernel density estimators [29]. The agents send their IS-generated weighted particle sets to the fusion center, which subsequently constructs a new set of weighted particles (referred to as fused particles) that approximates the geometrically-fused density. The proposed PGAF methods have the following desirable properties. (1) *Consistency:* estimators based on the fused particle set are shown to converge in probability as the number of particles increases. The challenge in obtaining consistent PGAF estimators stems from correlations that arise due to the GAF target density being approximated via the local agent particles. (2) *Low-communication overhead:* the agents only need to transmit their weighted particle sets to the fusion center. (3) *Privacy:* the agents need not share their local observations and additionally, for a sub-family of the proposed PGAF methods, the knowledge of the local prior densities is not required by the fusion center. This is in contrast with other fusion methods, such as in [30]. While not a formal guarantee, this privacy feature allows the agents to withhold sensitive information such as observations and local prior densities while only sharing the posterior information required for fusion.

Several elements are necessary to construct the fused particle set: (i) the target distribution, (ii) the proposal (or importance) distributions, (iii) sampling mechanisms, and (iv) weighting schemes. In general, the target distribution is not available in closed form and, furthermore, for privacy-preservation reasons, the fusion center does not have access to the expression of the local distributions. More specifically, the fusion center has to rely on the particle sets supplied by the agents and no further evaluations of the local posteriors can be made. The IS samples and the IS weights transmitted by the agents are used to construct a kernel-based density estimate (or kernel-density estimate for short) of the GAF density which is employed as target density. We propose several sets of proposal densities, called families, from which the fusion center samples the fused particles (see Section V-B). For a given family of proposals, we propose and analyze several sampling and weighting schemes, that is, which proposals to use from the family and how to weigh the resulting samples.

Parallels between PGAF and other IS methods can be made. PGAF borrows from MIS the idea of sampling and weighting schemes, that is, the mechanisms to select the index of the proposal density to sample from and the choice of weighting the resulting particle. In contrast with MIS, in PGAF the target distribution is not fully known and a particle based approximation is used instead, thus motivating the need for novel consistency results in the case of PGAF. Additionally for PGAF methods, multiple sets of proposal densities can be identified, each leading to estimators with different statistical properties. Adaptive IS algorithms [31], [32], [33] gradually construct improved approximations of the optimal importance distribution, while PGAF approximates the target distribution.

The article is organized as follows. Section II introduces the notation, Section III provides the necessary background, and Section IV states the main challenges and the objective of this article. The proposed algorithms are given in Section V while the corresponding convergence results are presented in Section VI. Appendix A contains the main proofs for the convergence results. A detailed analysis of state-of-the-art methods, supplemental lemmas and their proofs are given in the supplementary material [27]. All section and lemma numbers from the supplementary material [27] are indexed with the prefix "S", i.e., Lemma 1 from [27] is referred to as Lemma S-1.

II. Notation

Let $(\Omega, \mathcal{F}, \mathbb{P})$ denote a complete probability space and $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ a Borel space on the real coordinate space of dimension d. A random variable x is defined as a measurable function $\mathsf{x} : \Omega \to \mathbb{R}$. A sequence of random variables $(\mathsf{x}_k)_k$ converges in probability to x, written $\mathsf{x}_k \xrightarrow{\mathbb{P}} \mathsf{x}$, if for all $\epsilon > 0$, $\mathbb{P}\{|\mathsf{x}_k - \mathsf{x}| > \epsilon\} \to 0$ as $k \to \infty$. Almost sure convergence, denoted with $\mathsf{x}_k \xrightarrow{a.s.} \mathsf{x}$, follows if $\mathbb{P}\{\lim_{k\to\infty} \mathsf{x}_k = \mathsf{x}\} = 1$. Random variables are displayed in sans serif, upright fonts; their realizations in serif, italic fonts. Vectors and matrices are denoted by bold lowercase and uppercase letters, respectively. For example, a random variable and its realization are denoted by x and x; a random vector and its realization are denoted by $\boldsymbol{\mathsf{x}}$ and $\boldsymbol{x}$; a random matrix and its realization are denoted by $\boldsymbol{\mathsf{X}}$ and $\boldsymbol{X}$, respectively. Random and deterministic sets are denoted by upright sans serif and calligraphic font, respectively. For example, a random set and its realization are denoted by X and $\mathcal{X}$, respectively. The expectation and variance operators are denoted via $\mathbb{E}\{\cdot\}$ and $\mathbb{V}\{\cdot\}$. For a set $\mathcal{X}$, set cardinality is denoted as $\#\mathcal{X}$ while $\mathbb{1}_{\mathcal{X}}(\cdot)$ denotes the indicator function of $\mathcal{X}$. The set of integers $\{i, i+1, \ldots, j-1, j\}$ is compactly denoted with $\mathbb{N}_i^j$. The set of bounded continuous functions on $\mathbb{R}^d$ is denoted via $\mathcal{C}_{\mathrm{b}}(\mathbb{R}^d)$, while the set of continuity points of a function f is denoted with $\mathcal{C}(f)$. Function supremum is denoted via $\sup(\cdot)$ while function support via $\mathrm{supp}(\cdot)$.

The Lebesgue measure on $\mathbb{R}^d$ is denoted as λ. Let $\mathcal{P}(\mathbb{R}^d)$ denote the set of probability measures on $\mathcal{B}(\mathbb{R}^d)$. For $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, we write $\mu \ll \nu$ if μ is absolutely continuous with respect to (w.r.t.) ν and denote with $d\mu/d\nu$ the corresponding Radon-Nikodym derivative. In this work, the agent probability distributions are assumed to be absolutely continuous w.r.t. λ. Furthermore, the KLD is defined as $\mathbb{D}_{\text{KL}}\{\mu, \nu\} \triangleq \int \log(\frac{d\mu}{d\nu}) d\mu$ when $\mu \ll \nu$ and $\mathbb{D}_{\text{KL}}\{\mu, \nu\} = \infty$ otherwise. For $p \in [1, \infty)$, $L^p(\lambda)$ represents the space of real-valued functions $f : \mathbb{R}^d \to \mathbb{R}$ such that $\int |f|^p d\lambda$ is finite. The L^2 vector norm on $\mathbb{R}^d$ is $\|\cdot\|$. The Dirac (singular) measure at point $\boldsymbol{x}$ is $\delta_{\boldsymbol{x}}(\cdot)$. The notation $f_{\mathcal{N}}(\boldsymbol{x}; \boldsymbol{m}, \boldsymbol{M})$ represents the d-dimensional standard normal probability density function of variable $\boldsymbol{x}$, mean vector $\boldsymbol{m}$, and covariance matrix $\boldsymbol{M}$. Finally, $f_{\mathcal{L}}(x; m, b) \triangleq (2b)^{-1} \exp(-|x - m|/b)$ denotes the Laplace (double exponential) probability density function of variable $x \in \mathbb{R}$, location $m \in \mathbb{R}$, and scale $b > 0$.

III. System Model

Consider a network of M agents ($M > 1$), indexed via $m \in \mathbb{N}_1^M$, where each agent is tasked with inferring an unknown quantity. Each agent constructs an estimate $\boldsymbol{\mathsf{x}}_m : \Omega \to \mathbb{R}^d$ of the unknown quantity, where the random variable x_m has a corresponding distribution $P_m \in \mathcal{P}(\mathbb{R}^d)$, i.e., the push-forward probability measure $P_m(\mathcal{A}) = \mathbb{P}\{\mathsf{x}_m^{-1}(\mathcal{A})\}$ for all $\mathcal{A} \in \mathcal{B}(\mathbb{R}^d)$. The distributions P_m are assumed to be absolutely continuous w.r.t. the Lebesgue measure λ and we denote their corresponding probability density functions as $p_m \triangleq dP_m/d\lambda$ for $m \in \mathbb{N}_1^M$. Typical applications that enjoy such a setup are Bayesian network localization and navigation and target tracking [1], [15], [17], [34], where a network of sensors collects measurements and concentrates all their local information into the posterior distributions P_m. Henceforth, the local distributions P_m will also be referred to as local posterior distributions or, simply, local posteriors.

In non-linear estimation, closed-form expressions of the distributions P_m for $m \in \mathbb{N}_1^M$ are unavailable and Monte Carlo, more specifically IS (see [35, Sec. 2.5] and [36], [37]), approximations are used instead. Thus a set of discrete points $\{\boldsymbol{x}_m^{(n)} \in \mathbb{R}^d : n = 1, 2, \ldots, N\}$, are used to describe the distributions P_m for $m \in \mathbb{N}_1^M$. These points are instances of the corresponding random variables $\{\mathsf{x}_m^{(n)}\}_{n=1}^N$ for $m \in \mathbb{N}_1^M$, henceforth referred to as particles. Since sampling directly from P_m is difficult in general, these particles are independent and identically distributed (i.i.d.) with a distribution Q_m, referred to as *importance* or *proposal* distribution. Assuming $Q_m \ll \lambda$, let q_m denote the respective probability density function for agent $m \in \mathbb{N}_1^M$. Assuming $\mathcal{D}_m \triangleq \operatorname{supp}(q_m) \supset \operatorname{supp}(p_m)$ for $m \in \mathbb{N}_1^M$, the mismatch between target P_m and importance Q_m distributions is corrected via the IS weights [6, Ch. 1.3.2]

$$\mathsf{w}_m^{(n)} = \frac{p_m\left(\mathsf{x}_m^{(n)}\right)}{q_m\left(\mathsf{x}_m^{(n)}\right)} \tag{1}$$

for $n \in \mathbb{N}_1^N$. On the side of each agent $m \in \mathbb{N}_1^M$, the samples $\{\mathsf{x}_m^{(n)}\}_{n=1}^N$ and weights $\{\mathsf{w}_m^{(n)}\}_{n=1}^N$ are locally constructed since both p_m and q_m are available for evaluation at any finite number of points. Moreover, p_m is only available through evaluation, while sampling directly from p_m is unachievable and a full analytic expression of p_m is often unavailable. On the side of the fusion center, the weighted particle sets $\{(\mathsf{w}_m^{(n)}, \mathsf{x}_m^{(n)})\}_{n=1}^N$ are received from each agent m, without having access to the target distributions $\{p_m\}_{m=1}^M$, neither for evaluation nor sampling. Based on the weighted particle set $\{(\mathsf{w}_m^{(n)}, \mathsf{x}_m^{(n)})\}_{n=1}^N$, two IS choices exist to approximate P_m: the standard IS empirical distribution

$$\widehat{\mathsf{P}}_m^N(\mathcal{A}) = \frac{1}{N}\sum_{n=1}^{N} \mathsf{w}_m^{(n)} \delta_{\mathsf{x}_m^{(n)}}(\mathcal{A}) \tag{2}$$

and, by defining $\bar{\mathsf{w}}_m^{(n)} = \mathsf{w}_m^{(n)} / \sum_{n=1}^N \mathsf{w}_m^{(n)}$, the self-normalized IS empirical distribution

$$\bar{\mathsf{P}}_m^N(\mathcal{A}) = \sum_{n=1}^{N} \bar{\mathsf{w}}_m^{(n)} \delta_{\mathsf{x}_m^{(n)}}(\mathcal{A}) \tag{3}$$

for any set $\mathcal{A} \in \mathcal{B}(\mathbb{R}^d)$. Note that although (2) is not a probability distribution (since in general $\widehat{\mathsf{P}}_m^N(\mathbb{R}^d) \neq 1$), it leads to unbiased IS estimates [38, Thm. 9.1]. Furthermore, (3) is a probability distribution that leads to consistent IS estimates, see [7, Ch. 3.3] and [39].

IV. Objective Statement

The task of the centralized node is to construct a new probability distribution $F \in \mathcal{P}(\mathbb{R}^d)$ that merges the information contained in all of the agents in a consistent manner. The *first challenge* stems from not having access to the analytical form of the local distributions and instead, the fusion has to be carried out only having access to the local particle sets $\{(\mathsf{w}_m^{(n)}, \mathsf{x}_m^{(n)}) : n \in \mathbb{N}_1^N, m \in \mathbb{N}_1^M\}$.

The *second challenge* stems from the agents not having the same prior distribution. This prohibits the usage of likelihood consensus [40] and a fusion rule is employed instead that searches for the barycenter distribution $F \in \mathcal{P}(\mathbb{R}^d)$ that minimizes a suitable dissimilarity metric (e.g. the KLD) w.r.t. the local distributions P_m. Let $(c_m)_{m=1}^M$ be a collection of weights such that $c_m \in (0, 1)\ \forall m$ and $\sum_{m=1}^M c_m = 1$. The weights $(c_m)_{m=1}^M$ can be interpreted as credibility coefficients given to each agent during the fusion process.

One such fusion rule, namely the GAF, results from the following Kullback–Leibler divergence average (KLDA) minimization program

$$F_{\mathrm{G}} \triangleq \underset{F \in \mathcal{P}(\mathbb{R}^d)}{\operatorname{argmin}} \sum_{m=1}^{M} c_m \, \mathbb{D}_{\mathrm{KL}}\{F, P_m\}\,. \tag{4}$$

Restricting ourselves to the class of distributions that are absolutely continuous w.r.t. the Lebesgue measure (i.e., $F \ll \lambda$) and assuming that the support domains of the local probability density functions verify $\lambda(\cap_{m=1}^M \operatorname{supp}(p_m)) \neq 0$, the GAF distribution is easily shown (see Section S-III of the Supplemental Material [27]) to have the probability density

$$f_{\mathrm{G}} \triangleq \frac{dF_{\mathrm{G}}}{d\lambda} = Z^{-1} \prod_{m=1}^{M} p_m^{c_m} \tag{5}$$

where $Z \triangleq \int \prod_{m=1}^M p_m^{c_m}(\boldsymbol{x}) d\lambda(\boldsymbol{x}) \in (0, 1]$ is a normalization constant. The *third challenge* stems from not having in general a closed-form expression for the GAF density of (5), due to the normalization constant Z, or not having access to the expressions of the local posteriors $\{p_m\}_{m=1}^M$ at the fusion center.

In order to address these challenges, the objective of this work is to provide IS solutions with convergence guarantees for the GAF rule when the local posterior densities $(p_m)_{m=1}^M$ are only approximately known through their particle sets $\{(\mathsf{w}_m^{(n)}, \mathsf{x}_m^{(n)})\}_{n=1}^N$ for $m \in \mathbb{N}_1^M$. More specifically, we assume that the posteriors $\{p_m\}_{m=1}^M$ can only be evaluated at a finite collection of points and cannot be directly sampled from. Thus, the fused density (5) is only approximately known, further hampering direct sampling. We aim to formulate mechanisms to generate weighted particle sets $\{(\mathsf{w}_{\mathrm{G}}^{(n)}, \mathsf{x}_{\mathrm{G}}^{(n)})\}_{n=1}^N$, that provide consistent estimators for the fused density f_{G}, i.e.,

$$\sum_{n=1}^{N} \varphi\left(\mathsf{x}_{\mathrm{G}}^{(n)}\right) \frac{\mathsf{w}_{\mathrm{G}}^{(n)}}{\sum_{j=1}^{N} \mathsf{w}_{\mathrm{G}}^{(j)}} \xrightarrow{\mathbb{P}} \int \varphi(\boldsymbol{x}) f_{\mathrm{G}}(\boldsymbol{x}) d\lambda(\boldsymbol{x}) \tag{6}$$

as $N \to \infty$ and for any function $\varphi \in \mathcal{C}_{\mathrm{b}}(\mathbb{R}^d)$.

Remark 1: The usage of the inclusive form of the KLD (i.e., $\mathbb{D}_{\text{KL}}\{P_m, F\}$) in place of the exclusive form $\mathbb{D}_{\text{KL}}\{F, P_m\}$ of (4) is also possible and leads to the arithmetic–average fusion (AAF) rule with resulting density $f_{\text{A}} \triangleq \sum_{m=1}^{M} c_m p_m$, as seen in Section S-III of the Supplemental Material [27]. For completeness, an IS approximation for the AAF density is easily obtained via the concatenation $\{(c_m \bar{\mathsf{w}}_m^{(n)}, \mathsf{x}_m^{(n)}) : n \in \mathbb{N}_1^N, m \in \mathbb{N}_1^M\}$. The exclusive and inclusive KLDs, called I-projection and M-projection in [41, Ch. 8.5], have different characteristics. More specifically, the exclusive KLD tends to allocate the entire mass of the approximating density on the most probable mode of the target density whereas the inclusive KLD attempts to cover most of the target density, that is, to assign reasonable probability to the entire support of the target density at the cost of higher variance. For this reason, the exclusive KLD is referred to as mode-seeking or zero-forcing and the inclusive KLD is referred to as zero avoiding in [42]. A comprehensive study of various fusion rules is given in [43]. □

Remark 2: In the special case of Gaussian distributions [36], closed-form expressions are available for GAF (see Section S-IV of the Supplemental Material [27]), while closed-form expressions for AAF are trivially obtained for Gaussian mixture distributions. When the cross-correlations between the local estimators are unknown, the CI rule leads to estimators with a variance that is equal to or larger than the variance of the optimal fusion estimator that has knowledge of the aforementioned cross-correlations [44]. Such properties confer robustness to the GAF rule and are preferred in applications such as robotics.□

V. Particle GAF Elements and Methods

The framework of all IS-based methods [6], [7], including MIS [28], is comprised of four elements: (i) the target distribution, (ii) the proposal distributions, (iii) sampling mechanisms, and (iv) weighting schemes. The standard IS setting (see [7, Ch. 3.3] and [6, Ch. 1.3.2]) is obtained when a single proposal distribution Q is utilized. More complex sampling and weighting mechanisms emerge whenever a set of proposal functions are available (see [28, Section 5]).

Particle fusion methods attempt to merge several local particle sets, and thus different families of proposal distributions naturally arise from the local empirical distributions. In contrast to MIS, however, PGAF methods differ on the following two points: (i) here the target distributions are also particle-based approximations leading to distinct estimators from [28] and (ii) multiple families of proposal distributions can be constructed in addition to $\{q_m\}_{m=1}^{M}$. In the following, we describe the construction of the necessary elements for IS with the GAF rule.

A. The GAF Target Distribution

In order to apply the KLDA metric of (4), or equivalently the GAF rule of (5), a regularization step is first performed on the local empirical distributions of (2). This is achieved via the convolution of the local empirical distributions with a kernel, defined as a Borel-measurable function $K : \mathbb{R}^d \to [0, \infty)$. Furthermore, for a positive sequence of real terms $(h_i)_{i=1}^{\infty}$, let $K_{h_i}(\boldsymbol{x}) \triangleq h_i^{-d} K(\boldsymbol{x}/h_i)\ \forall i$. Note that $h_i \in \mathbb{R}_{>0}$ are referred to as bandwidth terms $\forall i$. For each agent m, a regularized particle approximation of the density p_m is obtained as

$$\hat{\mathsf{p}}_m^N(\boldsymbol{x}) \triangleq \frac{1}{N} \sum_{n=1}^{N} \mathsf{w}_m^{(n)} K_{h_N}\left(\boldsymbol{x} - \mathsf{x}_m^{(n)}\right). \tag{7}$$

Since a closed-form expression for the GAF density is generally not known and only the local particle sets are available, we employ the regularized densities $\{\hat{\mathsf{p}}_m^N\}_{m=1}^{M}$ to construct the following target density

$$\mathsf{f}_{\text{G}}^N(\boldsymbol{x}) \triangleq \prod_{m=1}^{M} \left[\hat{\mathsf{p}}_m^N(\boldsymbol{x})\right]^{c_m}. \tag{8}$$

Note that the density of (8) is not normalized, i.e., $\int \mathsf{f}_{\text{G}}^N(\boldsymbol{x}) d\lambda(\boldsymbol{x}) \neq 1$ in general, and it depends on the number of particles N, the kernel K, and bandwidth h_N.

B. Proposal Families

The fusion center constructs a set of weighted GAF particles $\{(\mathsf{w}_{\text{G}}^{(n)}, \mathsf{x}_{\text{G}}^{(n)})\}_{n=1}^{N_{\text{G}}}$ by relying on the local particle sets $\{(\mathsf{w}_m^{(n)}, \mathsf{x}_m^{(n)})\}_{n=1}^{N}$ for $m \in \mathbb{N}_1^M$. Without loss of generality[1], we will consider $N_{\text{G}} = N$. Given the available information from the M agents, the fusion center can construct multiple sets $\mathcal{Q} \triangleq \{\mathsf{q}_m\}_{m=1}^{M_{\text{G}}}$ of proposal densities q_m. Here, we refer to the set $\mathcal{Q}$ as a proposal family. Moreover, we will consider $M_{\text{G}} = M$, where a single proposal density q_m is constructed for each agent m. Applications where several proposal densities are associated with a single agent can be addressed in a similar fashion to the case treated here. The design and choice of an optimal proposal family is related to adaptive importance sampling methods (see [45] for more details). The GAF particles can be sampled from a proposal family $\mathcal{Q} \triangleq \{\mathsf{q}_m\}_{m=1}^{M}$, where the generic proposals q_m take on specific forms for each family, as shown next.

- P1 *Local distributions*. The family $\mathcal{Q}$ is comprised of the local proposals that were used to sample the local particle sets, i.e., $\mathcal{Q} \triangleq \{q_m : m \in \mathbb{N}_1^M\}$.
- P2 *Mollified distributions*. The family $\mathcal{Q}$ is comprised of a mollified version of the local proposals, i.e., $\mathcal{Q} \triangleq \{\tilde{q}_m^N : m \in \mathbb{N}_1^M\}$, where each importance distribution is obtained by convolution with a mollifier, i.e., $\tilde{q}_m^N \triangleq q_m * K_{h_N}$. For the special case of a Gaussian kernel K, sampling from a proposal $\tilde{q}_m^N$ is easily obtained by sampling a particle from q_m and adding an independent Gaussian noise with distribution K_{h_N}.
- P3 *Regularized distributions*. The family $\mathcal{Q}$ is comprised of regularized particle approximations of the local posteriors, i.e., $\mathcal{Q} \triangleq \{\bar{\mathsf{q}}_m^N : m \in \mathbb{N}_1^M\}$, where each $\bar{\mathsf{q}}_m^N$ is defined as

$$\bar{\mathsf{q}}_m^N(\boldsymbol{x}) \triangleq \frac{1}{\mathsf{c}} \sum_{n=1}^{N} \left[\mathsf{w}_m^{(n)}\right]^{c_m} \left[K_{h_N}\left(\boldsymbol{x} - \mathsf{x}_m^{(n)}\right)\right]^{c_m} \tag{9}$$

[1]The consistency results of Section VI also hold when the local number N and the fused number N_{G} of particles jointly increase.

where $\mathsf{c} \triangleq (h_N^d)^{1-c_m} \big[\sum_{j=1}^N \big(\mathsf{w}_m^{(j)}\big)^{c_m} \big] \int K^{c_m} d\lambda < \infty \; \forall N$ and $c_m \in (0,1)$, as per assumption A.4 of Section VI. In particular for Gaussian kernels K and given the local particle sets, sampling from a proposal $\bar{\mathsf{q}}_m^N$ is equivalent to sampling from a Gaussian mixture and has a low computational overhead. Sampling from `P3` has the added privacy advantage: the fusion center requires neither local observations nor local proposal distributions $\{q_m\}_{m=1}^M$. Instead, the fusion center relies solely on the weighted particle sets $\big\{ \big(\mathsf{w}_m^{(n)}, \mathbf{x}_m^{(n)}\big) \big\}_{n=1}^N$ for $m \in \mathbb{N}_1^M$.

C. Sampling Methods

Borrowing from MIS concepts [28] and regardless of which proposal family is employed, several sampling methods are available to select the index $\mathsf{i}^{(n)}$ of the proposal used to sample the n-th particle $\mathbf{x}_{\mathrm{G}}^{(n)}$. Let $\mathcal{Q} = \big\{\mathsf{q}_m : m \in \mathbb{N}_1^M\big\}$ be the family of proposals and define an arbitrary collection $\boldsymbol{r} \triangleq (r_m)_{m=1}^M$, such that $\sum_{m=1}^M r_m = 1$ and $r_m > 0$ for $m \in \mathbb{N}_1^M$. In this work, the weights $(r_m)_{m=1}^M$ are assumed fixed. The choice of optimal wights is a subject of ongoing research in the MIS community (see [46], [47] and references within). In the following, we present several index-selection mechanisms for $\big\{\mathsf{i}^{(n)}\big\}_{n=1}^N$, with $\mathsf{i}^{(n)} \in \mathbb{N}_1^M$ and $n \in \mathbb{N}_1^N$, that indicate which proposals to sample from.

- `S1` *Categorical sampling (with replacement)*. The indices $\big\{\mathsf{i}^{(n)}\big\}_{n=1}^N$ of IS densities from $\mathcal{Q}$ are i.i.d. according to a categorical distribution with weights $\boldsymbol{r}$, i.e., $\mathsf{i}^{(n)} \sim \mathrm{Cat}(\boldsymbol{r}) \; \forall n$, where $\mathrm{Cat}(\boldsymbol{r})$ denotes the categorical distribution with event probabilities $(r_m)_{m=1}^M$. Given a sequence of realizations $\big\{i^{(n)}\big\}_n$, the particles are conditionally sampled as $\mathbf{x}_{\mathrm{G}}^{(n)} | i^{(n)} \sim \mathsf{q}_{i^{(n)}} \; \forall n$.
- `S2` *Recycling selection* (only in combination with the proposal family `P1`). The particles $\mathbf{x}_{\mathrm{G}}^{(n)}$ are deterministically mapped to the (already available) particles $\big\{\mathbf{x}_m^{(n)}\big\}_{n=1}^N$ for $m \in \mathbb{N}_1^M$. Let $\mathrm{Det}_M^N(\cdot)$ be a function $\mathbb{N}_1^N \ni n \mapsto \mathrm{Det}_M^N(n) \in \mathbb{N}_1^M \times \mathbb{N}_1^N$, i.e., that maps each GAF particle index n to a proposal index $i^{(n)}$ and a local particle index $j^{(n)}$. Thus, the `S2`-sampled particles are defined as $\mathbf{x}_{\mathrm{G}}^{(n)} = \mathbf{x}_{i^{(n)}}^{(j^{(n)})}$ where $\big(i^{(n)}, j^{(n)}\big) = \mathrm{Det}_M^N(n)$ for all $n \in \mathbb{N}_1^N$. Let κ_m^N denote the number of times the deterministic function $\mathrm{Det}_M^N(\cdot)$ maps to the m-th proposal in N consecutive calls, i.e, $\kappa_m^N \triangleq \#\big\{\big(i^{(n)}, j^{(n)}\big) : \big(i^{(n)}, j^{(n)}\big) = \mathrm{Det}_M^N(n),\, i^{(n)} = m,\, n \in \mathbb{N}_1^N\big\}$. While multiple definitions of $\mathrm{Det}_M^N(\cdot)$ are possible[2], here, we only require that $\mathrm{Det}_M^N(\cdot)$ be constructed such that $\lim_{N\to\infty} \kappa_m^N / N = r_m$ $\forall m$ and that the tuples $\big(i^{(n)}, j^{(n)}\big)$ are distinct for all n.

In the case of the scheme `S2`, no additional sampling is carried out, and instead the fused particles $\big\{\mathbf{x}_{\mathrm{G}}^{(n)}\big\}_{n=1}^N$ are deterministically selected from the local (and already available) particles $\big\{\mathbf{x}_m^{(n)} : n \in \mathbb{N}_1^N,\, m \in \mathbb{N}_1^M\big\}$. This has the potential to significantly reduce computational complexity especially in applications involving high-dimensional spaces. Note that `S2` only works in conjunction with the proposal family `P1`, since the local particles obey the distributions $\big\{q_m : m \in \mathbb{N}_1^M\big\}$. In addition to the categorical sampling with replacement of `S1`, in [28] a sampling method without replacement is employed. However, this becomes infeasible in our formulation since $M \ll N$ in general. Moreover, for the convergence analysis the number of particles is taken to infinity while the number of agents M is fixed.

D. Weighting Schemes

An (un-normalized) weighting function is constructed as $\tilde{\mathsf{w}}_{\mathrm{G}}^N(\boldsymbol{x}) = \frac{\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{x})}{\psi_{\mathcal{P}_n}(\boldsymbol{x})}$, i.e., as the ratio of the target density $\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{x})$ and the importance function $\psi_{\mathcal{P}_n}(\boldsymbol{x})$. Note that $\psi_{\mathcal{P}_n}$ is a generic probability density function that obeys $\mathrm{supp}(\psi_{\mathcal{P}_n}) \supseteq \mathrm{supp}(\mathsf{f}_{\mathrm{G}}^N)$ and which is parameterized by a subset $\mathcal{P}_n \subseteq \big\{i^{(n)}\big\}_{n=1}^N$ of the indices used in the sampling of $\big\{\mathbf{x}_{\mathrm{G}}^{(n)}\big\}_{n=1}^N$. From [28], two choices for $\psi_{\mathcal{P}_n}$ are relevant to this work and lead to the weighting schemes presented next.

- `W1` *Direct weighting*. For any family of proposal densities and assuming that the sequence of indices $\big\{i^{(n)}\big\}_{n=1}^N$ was used to select the individual proposal densities to sample $\big\{\mathbf{x}_{\mathrm{G}}^{(n)}\big\}_{n=1}^N$, the importance function for $\mathbf{x}_{\mathrm{G}}^{(n)}$ is set to $\psi_{\mathcal{P}_n} = \mathsf{q}_{i^{(n)}}$. In other words, irrespective of the mechanism used to select the indices $\big\{i^{(n)}\big\}_{n=1}^N$, if $\mathbf{x}_{\mathrm{G}}^{(n)}$ was sampled from $\mathsf{q}_{i^{(n)}}$, then $\psi_{\mathcal{P}_n}(\boldsymbol{x}) = \mathsf{q}_{i^{(n)}}(\boldsymbol{x})$.
- `W2` *Mixture weighting*. For any family of proposal densities, the importance function is set to the mixture distribution $\psi_{\mathcal{P}_n}(\boldsymbol{x}) = \Psi(\boldsymbol{x}) \triangleq \sum_{m=1}^M r_m \mathsf{q}_m(\boldsymbol{x})$. In other words, irrespective of the mechanism used to select the indices $\big\{i^{(n)}\big\}_{n=1}^N$, each particle is interpreted as being drawn from the mixture of densities that comprise the proposal family $\mathcal{Q} = \{\mathsf{q}_m\}_{m=1}^M$. This alternative weighting scheme has received different names in the literature, such as balance heuristic or deterministic mixture weights. See [46], [48], [49] for more details.

E. Particle GAF Methods

PGAF methods arise as combinations of the previous elements: proposal family, sampling method, and weighting scheme. Table I presents the resulting PGAF methods. The GAF particle set is constructed as $\big\{\big(\mathsf{w}_{\mathrm{G}}^{(n)}, \mathbf{x}_{\mathrm{G}}^{(n)}\big)\big\}_{n=1}^N$, where the weights[3] are set to

$$\mathsf{w}_{\mathrm{G}}^{(n)} = \tilde{\mathsf{w}}_{\mathrm{G}}^N\big(\mathbf{x}_{\mathrm{G}}^{(n)}\big). \tag{10}$$

Note that regardless of the proposal family, sample method, or weighting scheme, the weights $\mathsf{w}_{\mathrm{G}}^{(n)}$ are correlated due to the use of the kernel density estimate $\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{x})$ as the target density in all samples $n \in \mathbb{N}_1^N$. From Table I, note again that the recycling sampling method `S2` is only employable with the proposal family `P1`, since here the fused particles are selected from the available local particles. Finally, the PGAF methods that employ

[2] For example, by partitioning the set $\mathbb{N}_1^N$ into M subsets with cardinalities of the type $\lfloor r_m N \rfloor$ for $m = 1, \ldots, M$, then one such function maps $n \mapsto (m, n)$ when n belongs to the m-th subset.

[3] In the following, the dependence of the importance weights on the number of particles N is dropped for notational compactness.

TABLE I
PARTICLE GAF METHODS: FOR A SPECIFIC PROPOSAL FAMILY $\mathcal{Q} = \{\mathsf{q}_m : m \in \mathbb{N}_1^M\}$ (E.G., P1, P2, OR P3), EACH COMBINATION OF SAMPLING AND WEIGHTING SCHEMES LEADS TO A VALID METHOD

	W1	W2
S1	$\mathsf{i}^{(n)} \sim \mathrm{Cat}(\boldsymbol{r})$ $\mathsf{x}_{\mathrm{G}}^{(n)} \vert i^{(n)} \sim \mathsf{q}_{i^{(n)}}$ $\tilde{\mathsf{w}}_{\mathrm{G}}^N(\boldsymbol{x}) = \frac{\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{x})}{\mathsf{q}_{i^{(n)}}(\boldsymbol{x})}$	$\mathsf{i}^{(n)} \sim \mathrm{Cat}(\boldsymbol{r})$ $\mathsf{x}_{\mathrm{G}}^{(n)} \vert i^{(n)} \sim \mathsf{q}_{i^{(n)}}$ $\tilde{\mathsf{w}}_{\mathrm{G}}^N(\boldsymbol{x}) = \frac{\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{x})}{\sum_{m=1}^M r_m \mathsf{q}_m(\boldsymbol{x})}$
S2 (only for P1)	$(i^{(n)}, j^{(n)}) = \mathrm{Det}_M^N(n)$ $\mathsf{x}_{\mathrm{G}}^{(n)} \leftarrow \mathsf{x}_{i^{(n)}}^{(j^{(n)})}$ $\tilde{\mathsf{w}}_{\mathrm{G}}^N(\boldsymbol{x}) = \frac{\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{x})}{q_{i^{(n)}}(\boldsymbol{x})}$	$(i^{(n)}, j^{(n)}) = \mathrm{Det}_M^N(n)$ $\mathsf{x}_{\mathrm{G}}^{(n)} \leftarrow \mathsf{x}_{i^{(n)}}^{(j^{(n)})}$ $\tilde{\mathsf{w}}_{\mathrm{G}}^N(\boldsymbol{x}) = \frac{\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{x})}{\sum_{m=1}^M r_m q_m(\boldsymbol{x})}$

the P3 proposal, do not require the knowledge of the priors $\{q_m\}_{m=1}^M$ used by the local nodes. This enhances the privacy of the nodes via-à-vis the fusion center when the latter employs P3-based PGAF methods.

For any bounded continuous test function $\varphi \in \mathcal{C}_{\mathrm{b}}(\mathbb{R}^d)$, the Z-normalized PGAF estimator is defined as

$$\upvarphi_{\mathrm{G}}^N \triangleq \frac{1}{ZN} \sum_{n=1}^N \varphi\big(\mathsf{x}_{\mathrm{G}}^{(n)}\big) \mathsf{w}_{\mathrm{G}}^{(n)} \tag{11}$$

with the weights $\mathsf{w}_{\mathrm{G}}^{(n)}$ defined in (10). In the case of unknown Z, the estimator $\hat{Z}^N = \frac{1}{N}\sum_{n=1}^N \mathsf{w}_{\mathrm{G}}^{(n)}$ is employed instead of Z in (11), leading to the self-normalized PGAF estimator

$$\bar{\upvarphi}_{\mathrm{G}}^N \triangleq \sum_{n=1}^N \varphi\big(\mathsf{x}_{\mathrm{G}}^{(n)}\big) \frac{\mathsf{w}_{\mathrm{G}}^{(n)}}{\sum_{j=1}^N \mathsf{w}_{\mathrm{G}}^{(j)}}\,. \tag{12}$$

VI. CONVERGENCE RESULTS

Throughout the convergence guarantees presented in this section, several of the following assumptions on the kernel function K, the bandwidth sequence $(h_n)_{n=1}^\infty$, and the densities $\{p_m\}_{m=1}^M$ and $\{q_m\}_{m=1}^M$ are assumed to hold:

A.1(a) $K(\boldsymbol{x}) = K(-\boldsymbol{x})$ and $K(\boldsymbol{x}) \geqslant 0\ \forall \boldsymbol{x}$,
A.1(b) $\int K d\lambda = 1$,
A.1(c) $\sup_{\boldsymbol{x}\in\mathbb{R}^d} K(\boldsymbol{x}) = M_{\mathrm{K}} < \infty$,
A.1(d) $\|\boldsymbol{x}\|^d K(\boldsymbol{x}) \to 0$ as $\|\boldsymbol{x}\| \to \infty$,
A.2(a) $\lim_{n\to\infty} h_n = 0$,
A.2(b) $\lim_{n\to\infty} n h_n^d = \infty$,
A.2(c) $\sum_{n=1}^\infty \exp(-\alpha n h_n^d) < \infty$ for any $\alpha > 0$,
A.3(a) $\sup_{\boldsymbol{x}\in\mathcal{D}_m} p_m(\boldsymbol{x})/q_m(\boldsymbol{x}) = M_m < \infty \quad \forall m \in \mathbb{N}_1^M$ (recall $\mathcal{D}_m \triangleq \mathrm{supp}(q_m)$),
A.3(b) p_m are λ-almost everywhere continuous $\forall m \in \mathbb{N}_1^M$,
A.3(c) $p_m/q_m \in L^1(\mathbb{R}^d)\ \forall m \in \mathbb{N}_1^M$,
A.4 $K^{c_m} \in L^1(\mathbb{R}^d)$ (i.e., $\int K^{c_m} d\lambda < \infty$) $\forall m \in \mathbb{N}_1^M$.

Remark 3: Assumptions A.1(a)-A.3(b) are standard in the kernel-density estimate and IS literature. For $d = 1$ in [50] and for $d \geq 1$ in [51], assumptions A.1(b), A.1(c), A.1(4), A.2(b), and A.2(a) are employed to ensure the asymptotic unbiasedness and weak consistency of the standard kernel density estimate, that is, with a direct sampling construction (sampling from p_m instead of q_m). Additionally by assuming A.2(c) and A.3(b), both strong consistency and convergence in $L^1(\mathbb{R}^d)$ are obtained in [52] for the same standard kernel density estimate. Assumption A.3(a) implies that the importance density q_m has heavier tails than p_m and is a typical assumption for IS methods, see [7, Ch.3.3.2] and [53]. Furthermore, this condition is strengthened in Assumption A.3(c), to ensure the convergence of the PGAF methods that rely on the recycling strategy S2. Assumption A.4 is necessary for the proper definition of the P3 proposal family in (9). □

In the following, several convergence results are presented for the PGAF estimators in (11) and (12), where the fused particle set $\{\mathsf{x}_{\mathrm{G}}^{(n)}\}_{n=1}^N$ and the weighting function $\tilde{\mathsf{w}}_{\mathrm{G}}^N$ are subject to the particular choice of PGAF method, as seen in Table I.

Theorem VI.1 (Asymptotic unbiasedness (I)–for PGAF methods with : S1 sampling) Under assumptions A.1(a)-A.3(b) (and A.4 for the methods involving P3) and when the normalization constant Z of (5) is known, the Z-normalized PGAF methods P1S1W1, P2S1W1, P1S1W2, P2S1W2, P3S1W1, and P3S1W2, from Table I are asymptotically unbiased, i.e., $\lim_{N\to\infty} \mathbb{E}\{\upvarphi_{\mathrm{G}}^N\} = \int \varphi(\boldsymbol{x}) f_{\mathrm{G}}(\boldsymbol{x}) d\lambda(\boldsymbol{x})$, for any test function $\varphi \in \mathcal{C}_{\mathrm{b}}(\mathbb{R}^d)$. □

Theorem VI.2 (Asymptotic unbiasedness (II)–for PGAF methods with : S2 sampling) Under assumptions A.1(a)-A.3(b), A.3(c), and when the normalization constant Z of (5) is known, the Z-normalized PGAF methods P1S2W1 and P1S2W2 from Table I are asymptotically unbiased, that is, $\lim_{N\to\infty} \mathbb{E}\{\upvarphi_{\mathrm{G}}^N\} = \int \varphi(\boldsymbol{x}) f_{\mathrm{G}}(\boldsymbol{x}) d\lambda(\boldsymbol{x})$, for any test function $\varphi \in \mathcal{C}_{\mathrm{b}}(\mathbb{R}^d)$. □

Remark 4: Note that the PGAF methods with the S2 sampling method require a stronger assumption on the tails of the importance distributions $\{q_m\}_{m=1}^M$ than the methods with S1—namely, A.3(c) in contrast to A.3(a). This is due to the recycling of samples that leads to terms of the form $\frac{p_m(\mathsf{x}_m)}{q_m^2(\mathsf{x}_m)}$ where $\mathsf{x}_m \sim q_m$ for $\forall m \in \mathbb{N}_1^M$. The condition of finite expectation of these aforementioned terms is ensured by assumption A.3(c). □

Theorem VI.3 (Consistency of PGAF methods): For all PGAF methods of Table I and under the specific assumptions listed below, both the Z-normalized (11) and the self-normalized (12) estimators are (weakly) consistent, i.e., $\upvarphi_{\mathrm{G}}^N \xrightarrow{\mathbb{P}} \int \varphi(\boldsymbol{x}) f_{\mathrm{G}}(\boldsymbol{x}) d\lambda(\boldsymbol{x})$ and $\bar{\upvarphi}_{\mathrm{G}}^N \xrightarrow{\mathbb{P}} \int \varphi(\boldsymbol{x}) f_{\mathrm{G}}(\boldsymbol{x}) d\lambda(\boldsymbol{x})$ as $N \to \infty$ for any test function $\varphi \in \mathcal{C}_{\mathrm{b}}(\mathbb{R}^d)$.

- Consistency for P1S1W1 results by assuming A.1(a)-A.3(b) and

$$(p_m * K_{h_N})/q_k \in L^1(\mathbb{R}^d) \tag{13}$$

for all $m, k \in \mathbb{N}_1^M$ and $\forall N$. Whereas for P1S1W2, the above only needs to hold for all $m = k$.

- Consistency for P1S2W1 follows by assuming A.1(a)-A.3(b),

$$\sup_{\boldsymbol{x}\in\mathcal{D}_m} p_m(\boldsymbol{x})/q_m^2(\boldsymbol{x}) < \infty \tag{14}$$

for all $m \in \mathbb{N}_1^M$, and (13) $\forall m, k \in \mathbb{N}_1^M$. Whereas for P1S2W2 (13) only needs to hold for all $m = k$.

- Consistency for `P2S1W1` follows by assuming A.1(a)-A.3(b), and

$$(p_m * K_{h_N})/(q_k * K_{h_N}) \in L^1(\mathbb{R}^d) \tag{15}$$

  for all $n, k \in \mathbb{N}_1^M$ and $\forall N$. Whereas for `P2S1W2` the above only needs to hold for all $n = k$.
- Consistency for `P3S1W1` and `P3S1W2` follow from A.1(a)-A.3(b) and A.4. □

Remark 5: The methods involving the scheme `S2` employ recycling of samples and, consequently, have the lowest computational and memory requirements at the cost of the more stringent conditions of (14) on the tails of $\{q_m\}_{m=1}^M$. The conditions in (15) for the methods involving `P2` are less severe than those of (13) and (14). Most noteworthy are the methods that rely on the `P3` proposal family, which in virtue of their construction—the distribution in (9)—do not require any additional constraints on the tails of $\{q_m\}_{m=1}^M$. □

VII. Implementation and Experimental Results

A. Implementation Aspects

This section presents a discussion on the assumptions of Section VI, specifically the kernel function and the bandwidth sequence. Furthermore, we also offer examples of kernels and bandwidth sequences that satisfy these assumptions. It can be shown that assumptions A.1(a)-A.1(d) are satisfied by all kernel functions (both Gaussian and non-Gaussian) given in Table I of [50] for the single dimensional case ($d = 1$). For $d > 1$, multidimensional kernels are straightforwardly obtained by product kernels [51, Sec. 4]. In this work, we will employ the multidimensional normal kernel $K(\boldsymbol{x}) = (2\pi)^{-d/2} \exp(-\boldsymbol{x}^\top \boldsymbol{x}/2)$ for $\boldsymbol{x} \in \mathbb{R}^d$.

For a bandwidth sequence $(h_n)_{n=1}^\infty$ with general term $h_n = cn^{-p}$, for some constant $c > 0$ and $0 < p < d^{-1}$, assumptions A.2(a) and A.2(b) are readily verified while A.2(c) follows from the integral test for series. More specifically, the convergence of the series $\sum_{n=1}^\infty \exp(-\alpha n h_n^d) < \infty$ follows from the facts: $\exp(-n^{\Theta(1)}) < n^{-\beta}$ for all $\beta > 1$ and n sufficiently large; and $\sum_{n \geq 1} n^{-\beta} < \infty$ for every $\beta > 1$. Note that this class of bandwidth sequences includes the optimal sequence $h_n = O(n^{-1/(d+4)})$, i.e., the sequence that minimizes the mean-squared error for a fixed n [51, p. 186].

B. Numerical Experiments

We consider two fusion scenarios. The first scenario considers the fusion of Gaussian densities, where closed-form solutions exist for the resulting fused density. Furthermore, we consider the test function $\varphi(x) = 1\ \forall x \in \mathbb{R}$, leading to estimators of the normalization constant Z. The parameters in this case are: $d = 1$, $M = 2$, and the bandwidth sequence has the general term $h_n = 0.1 \times n^{-1/5}$. The local densities are Gaussian: $p_1(x) = f_{\mathcal{N}}(x; 1, 1)$ and $p_2(x) = f_{\mathcal{N}}(x; -1, 2)$; the proposal densities are also Gaussian: $q_1(x) = f_{\mathcal{N}}(x; 1.3, 4.2)$ and $q_2(x) = f_{\mathcal{N}}(x; -0.5, 5)$; while $c_1 = 0.2$, $c_2 = 0.8$, and $r_1 = r_2 = 0.5$. Note that in addition to all the assumptions of Section VI, the strong assumption $q_m \in L^4(\mathbb{R}^d)$, $\forall m \in \{1, 2\}$, is also met in this single-Gaussian case. This last assumption is only required to ensure the convergence of the uncorrected particle geometric–average fusion (U-PGAF) method (see Section VI of the Supplemental Material [27]). Furthermore, the exact GAF density is $f_{\mathrm{G}}(x) = f_{\mathcal{N}}(x; -1/3, 5/3)$ while the normalization constant of (5) is $Z \approx 0.75$. For the proposed PGAF methods, an estimator for Z is constructed via $\hat{\mathsf{Z}}^N = \frac{1}{N} \sum_{n=1}^N \tilde{\mathsf{w}}_{\mathrm{G}}^N\left(\mathsf{x}_{\mathrm{G}}^{(n)}\right)$, where the particles and the weights are obtained according to the various PGAF methods resulting from Table I.

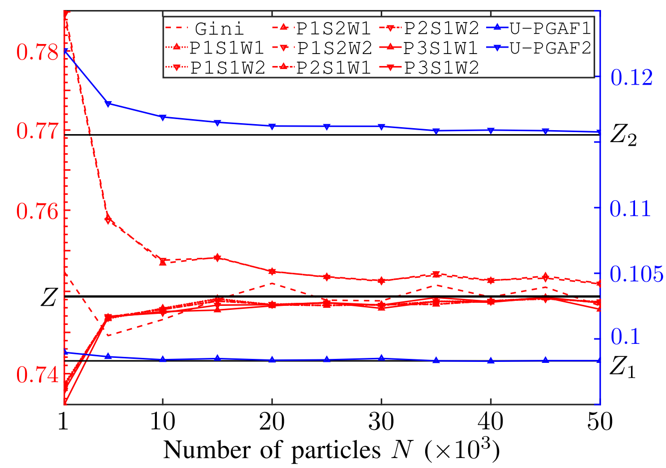

Fig. 1. Bias analysis of the proposed and state-of-the-art PGAF methods for the estimation of the normalization constant in the Gaussian case (i.e., Gaussian density fusion). The proposed methods and a variant of the Gaussian mixture particle geometric–average fusion (GM-PGAF) method with perfect knowledge of the local posteriors (referred to as `Gini`) are represented on the left axis in red. The two resulting estimators of the U-PGAF method are shown on the right axis in blue. The exact values of Z, Z_1, and Z_2 are also indicated on the respective axes.

We compare the proposed PGAF methods with two variants based on the state-of-the-art methods presented in Section S-II. First, we construct a variant of the Gaussian mixture particle geometric–average fusion method (referred to as `Gini`) that has perfect knowledge of the local posteriors, that is, the Gaussian-mixture distributions of (S-29) are replaced with the exact local posteriors: $\tilde{\mathsf{p}}_m \leftarrow p_m\ \forall m \in \mathbb{N}_1^M$. Thus, the `Gini` method employs the exact target density whereas all other methods use kernel density estimates instead. Second, the U-PGAF method of Section S-II leads to two normalization constant estimators since $M = 2$. These estimators are given by $\hat{\mathsf{Z}}_m^N = \frac{1}{N} \sum_{n=1}^N \mathsf{f}_{\mathrm{G}}^N\left(\mathsf{x}_m^{(n)}\right)$, for $m \in \{1, 2\}$, with the U-PGAF weights given in (S-48). Note that according to the consistency analysis of the U-PGAF method conducted in Section S-VI, the estimators converge to $Z_m \triangleq Z \int f_{\mathrm{G}}(x) q_m(x) d\lambda(x)$ for $m \in \{1, 2\}$. Numerically in this case, $Z_1 \approx 0.098$ and $Z_2 \approx 0.115$.

Fig. 1 shows the mean value of the resulting Z estimators computed over 100 independent simulations using the proposed methods and the state-of-the-art variants. Note that the proposed PGAF methods and the Gini method asymptotically approach Z whereas the U-PGAF method approaches Z_1 and Z_2, corresponding to the two nodes respectively. This result confirms the asymptotic unbiasedness of the proposed PGAF methods while also numerically showing the asymptotic bias of the U-PGAF estimators $\hat{\mathsf{Z}}_1^N$ and $\hat{\mathsf{Z}}_2^N$ as estimators of Z.

Fig. 2 presents the estimated variance of the Z estimators for the `Gini` and the proposed PGAF methods. Two remarks

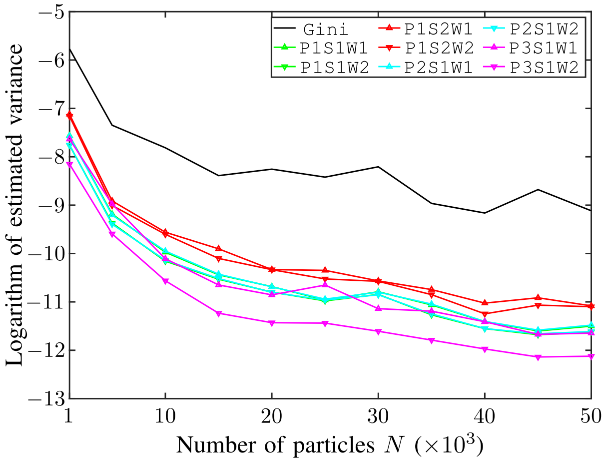

Fig. 2. Variance analysis of the normalization constant estimator for the proposed and state-of-the-art PGAF methods in the Gaussian case (i.e., Gaussian density fusion).

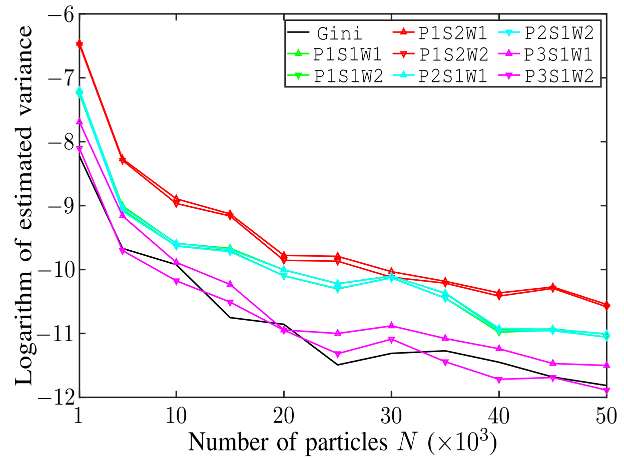

Fig. 4. Variance analysis of the normalization constant estimator for the proposed and state-of-the-art PGAF methods in the non-Gaussian case (i.e., Laplace density fusion).

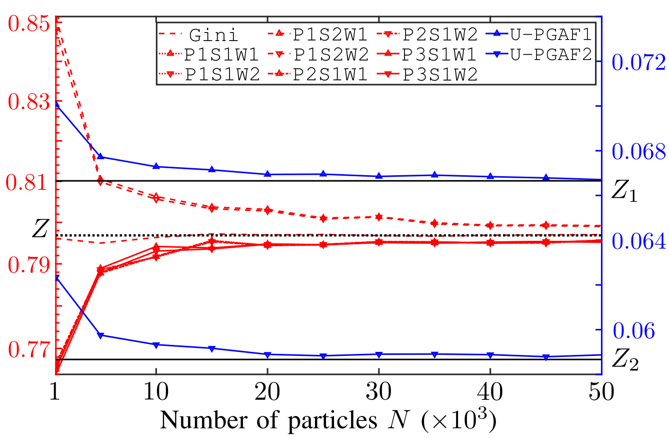

Fig. 3. Bias analysis of the proposed and state-of-the-art PGAF methods for the estimation of the normalization constant in the non-Gaussian case (i.e., Laplace density fusion). The proposed methods and a variant of the GM-PGAF method with perfect knowledge of the local posteriors (referred to as Gini) are represented on the left axis in red. The two resulting estimators of the U-PGAF method are shown on the right axis in blue. The exact values of Z, Z_1, and Z_2 are also indicated on the respective axes.

TABLE II

ESTIMATED QUARTILES q_1, q_2 (MEDIAN), AND q_3 OF THE ABSOLUTE ERROR $|Z - \hat{Z}^N|$ FOR THE Gini AND SEVERAL PGAF METHODS, AS A FUNCTION OF THE NUMBER OF PARTICLES N. THE ESTIMATED QUARTILE VALUES REPORTED IN THE TABLE WERE MULTIPLIED WITH 10^3

		Number of particles N $(\times 10^3)$					
		1	10	20	30	40	50
Gini	q_1	9.8	3.9	3.0	2.5	2.3	2
	q_2	20	8.2	6.3	5.3	4.8	4.4
	q_3	33.2	13.8	10.7	9.1	8.0	7.4
P1S1W2	q_1	7.4	1.8	1.4	1.3	1	0.8
	q_2	18	4.1	3.2	2.8	2.3	1.9
	q_3	27.9	7.2	5.1	5.8	3.8	3.4
P1S2W2	q_1	19.5	2.9	2.7	2.3	1.2	1.1
	q_2	32.8	6.5	4.7	3.9	2.9	2.8
	q_3	56	10.9	7	6.7	5	5.1
P2S1W2	q_1	8.2	1.8	1.4	1.2	0.9	0.8
	q_2	17	4.3	3	2.8	2.2	2
	q_3	28.2	7.4	5.1	5.9	3.8	3.5
P3S1W2	q_1	5.5	1.8	1.1	1	0.8	0.7
	q_2	12.7	3.7	2.1	1.9	1.9	1.7
	q_3	22.1	6.8	3.9	3.7	2.7	2.8

are in order about these results. First, the estimated variance of all proposed PGAF methods is decreasing with N. Second, the P3S1W2 method has the lowest estimated variance among all the methods. These conclusions are also supported by the results in Table II. The table presents estimated quartile values for the absolute error $|Z - \hat{Z}^N|$ in the case of the Gini and the proposed PGAF methods with W2 weighting. Similar to MIS [28], PGAF methods with the W2 weighting scheme are shown to exhibit a lower estimated variance than those with W1. Furthermore, the P3 family defined by (9) can be seen as providing a closer fit to the set of local posterior densities than the families P1 and P2, thus justifying its lower estimated variance.

A second scenario is considered where the densities to be fused are non-Gaussian. More specifically, the two Laplace distributions $p_1(x) = f_{\mathcal{L}}(x; 1, 1)$ and $p_2(x) = f_{\mathcal{L}}(x; -1, 2)$ are to be fused with equal weights $c_1 = c_2 = 0.5$. The proposal densities are also Laplace: $q_1(x) = f_{\mathcal{L}}(x; 1.3, 4.2)$ and $q_2(x) = f_{\mathcal{L}}(x; -0.5, 5)$; while other parameters are kept identical to the previous scenario. As before, the proposed PGAF methods are compared with the U-PGAF method and a Gini method that has access to the exact target density. Fig. 3 showcases the estimators for the normalization constant in this non-Gaussian case. The proposed PGAF methods and the Gini method are numerically converging to the true normalization constant $Z \approx 0.7968$, whereas the two U-PGAF-based estimators converge to $Z_1 \approx 0.0667$ and $Z_2 \approx 0.0587$. The logarithm of the variance of these estimators is numerically shown to vanish with increasing number of particles in Fig. 4. Similar to the Gaussian scenario, the P3S1W2 method has the lowest variance among the proposed PGAF methods.

VIII. Conclusion

The article proposed several IS methods for approximating GAF densities in multi-agent systems. The aim of these particle GAF methods is to sample particles that approximate the weighted geometric-average of local densities, where the local and the resulting GAF densities are difficult to sample from and/or do not have closed-form expressions. The local particle approximations serve to establish a kernel-density estimate of the GAF density which is subsequently employed as target density in a MIS framework. We also constructed several sets of importance densities and propose sampling and weighting schemes in order to construct the fused particle set, i.e., the particles that approximate the GAF density. The proposed PGAF methods are privacy-preserving, that is, the agents need not share their measurements, while for a sub-family of PGAF methods, the expressions of their prior distributions are not shared either. Moreover, the agents only need to transmit a weighted set of particles to the fusion center, thus ensuring a low communication overhead. Finally, consistency results were established for estimators based on the fused particles. The proposed PGAF methods are aimed at a broad spectrum of applications, such as, robotics, sensing, and multi-agent learning.

Appendix
Proofs of Theorems

Without loss of generality, let $\varphi \geqslant 0$ throughout this appendix.

Proof of Theorem VI.1:

- `(P1` or `P2)S1W1` Considering either the proposal family `P1` or `P2` (generically denoted via $\mathcal{Q}=\{\mathsf{q}_m\}_{m=1}^M$) coupled with the sampling method `S1` leads to i.i.d. particles $\{\mathbf{x}_{\mathrm{G}}^{(n)}\}$ that obey $\mathbf{x}_{\mathrm{G}}^{(n)}|\{\mathsf{i}^{(n)}=i^{(n)}\}\sim \mathsf{q}_{i^{(n)}}$ and to i.i.d. indices with laws $\mathsf{i}^{(n)}\sim \mathrm{Cat}(\boldsymbol{r})\ \forall n$. By the law of total expectation and Fubini's theorem,

$$\begin{aligned}&\mathbb{E}\{\upvarphi_{\mathrm{G}}^N\}\\&=\frac{1}{ZN}\sum_{n=1}^N\mathbb{E}\left\{\mathbb{E}\left\{\varphi\left(\mathbf{x}_{\mathrm{G}}^{(n)}\right)\frac{\mathsf{f}_{\mathrm{G}}^N\left(\mathbf{x}_{\mathrm{G}}^{(n)}\right)}{\mathsf{q}_{\mathsf{i}^{(n)}}\left(\mathbf{x}_{\mathrm{G}}^{(n)}\right)}\middle|\mathsf{i}^{(n)}\right\}\right\}\\&=\frac{1}{ZN}\sum_{n=1}^N\sum_{m=1}^M r_m\,\mathbb{E}\left\{\varphi\left(\mathbf{x}_{\mathrm{G}}^{(n)}\right)\frac{\mathsf{f}_{\mathrm{G}}^N\left(\mathbf{x}_{\mathrm{G}}^{(n)}\right)}{\mathsf{q}_m\left(\mathbf{x}_{\mathrm{G}}^{(n)}\right)}\middle|\mathsf{i}^{(n)}=m\right\}\\&=\frac{1}{Z}\int\varphi\left(\boldsymbol{y}\right)\mathbb{E}\{\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{y})\}\,d\lambda(\boldsymbol{y})\end{aligned}$$

Lemma (S-V.4) (i) leads to $\varphi(\boldsymbol{y})\,Z^{-1}\mathbb{E}\{\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{y})\}\to\varphi(\boldsymbol{y})\,f_{\mathrm{G}}(\boldsymbol{y})$ λ-a.e., similarly the sequence with general term $g_N(\boldsymbol{y})\triangleq\sup(\varphi)Z^{-1}\mathbb{E}\{\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{y})\}$ also converges $g_N(\boldsymbol{y})\to g(\boldsymbol{y})$ λ-a.e. as $N\to\infty$, with $g(\boldsymbol{y})\triangleq\sup(\varphi)f_{\mathrm{G}}(\boldsymbol{y})$. Moreover, $\int g_N d\lambda<\infty$ since $\int\mathbb{E}\{\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{y})\}d\lambda\leqslant\sum_m c_m\int p_m * K_{h_N}d\lambda=1$ $\forall N$, while from Lemma (S-V.5) (ii) it follows that $\lim_{N\to\infty}\int g_N d\lambda=\int g d\lambda=\sup(\varphi)<\infty$. Thus, the claim follows from a generalized form of the dominated convergence theorem (DCT).

- `(P1` or `P2)S1W2` We consider either the proposal family `P1` or `P2` (generically denoted via $\mathcal{Q}=\{\mathsf{q}_m\}_{m=1}^M$), the sampling method `S1`, and the weighting scheme `W2`. By marginalizing over the sampling index $\mathsf{i}^{(n)}$, the i.i.d. fused particles $\{\mathbf{x}_{\mathrm{G}}^{(n)}\}_{n=1}^N$ can be seen as drawn from the mixture density $\mathbf{x}_{\mathrm{G}}^{(n)}\sim\Psi$, where $\Psi(\boldsymbol{y})=\sum_{m=1}^M r_m\mathsf{q}_m(\boldsymbol{y})$. Similarly, by the law of total expectation

$$\begin{aligned}\mathbb{E}\{\upvarphi_{\mathrm{G}}^N\}&=\frac{1}{ZN}\sum_{n=1}^N\mathbb{E}\left\{\varphi\left(\mathbf{x}_{\mathrm{G}}^{(n)}\right)\frac{\mathsf{f}_{\mathrm{G}}^N\left(\mathbf{x}_{\mathrm{G}}^{(n)}\right)}{\sum_{m=1}^M r_m\,\mathsf{q}_m\left(\mathbf{x}_{\mathrm{G}}^{(n)}\right)}\right\}\\&=\frac{1}{Z}\int\varphi(\boldsymbol{y})\frac{\mathbb{E}\{\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{y})\}}{\sum_{m=1}^M r_m\,\mathsf{q}_m(\boldsymbol{y})}\,\Psi(\boldsymbol{y})d\lambda(\boldsymbol{y})\,.\end{aligned}$$

The claim follows from a similar reasoning as in the previous case.

- `(P3S1W1)` We fist denote the set of all local particles via $\mathsf{X}_M^N\triangleq\{\mathbf{x}_m^{(n)}:m\in\mathbb{N}_1^M,n\in\mathbb{N}_1^N\}$ and note that, conditionally on $\mathsf{X}_M^N=\mathcal{X}_M^N$, the random density functions $\bar{\mathsf{q}}_m^N$ and $\mathsf{f}_{\mathrm{G}}^N$ become deterministic, i.e., $\bar{\mathsf{q}}_m^N|\mathcal{X}_M^N\triangleq\bar{q}_m^N$ and $\mathsf{f}_{\mathrm{G}}^N|\mathcal{X}_M^N\triangleq f_{\mathrm{G}}^N$. The expectation $\mathbb{E}\{\upvarphi_{\mathrm{G}}^N\}$ can be written as

$$\begin{aligned}\mathbb{E}\{\upvarphi_{\mathrm{G}}^N\}&=\frac{1}{ZN}\sum_{n=1}^N\mathbb{E}\left\{\mathbb{E}\left\{\varphi\left(\mathbf{x}_{\mathrm{G}}^{(n)}\right)\frac{\mathsf{f}_{\mathrm{G}}^N\left(\mathbf{x}_{\mathrm{G}}^{(n)}\right)}{\bar{\mathsf{q}}_{\mathsf{i}^{(n)}}^N\left(\mathbf{x}_{\mathrm{G}}^{(n)}\right)}\middle|\mathsf{i}^{(n)},\mathsf{X}_M^N\right\}\right\}\\&=\frac{1}{ZN}\sum_{n=1}^N\mathbb{E}\left\{\int\varphi(\boldsymbol{y})\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{y})d\lambda(\boldsymbol{y})\right\}\\&=\frac{1}{Z}\int\varphi(\boldsymbol{y})\,\mathbb{E}\{\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{y})\}\,d\lambda(\boldsymbol{y})\end{aligned}$$

where Fubini's theorem was employed in the last line. Subsequently, a combination of Lemma (S.V.4) (i), Lemma (S.V.5) (ii), and DCT lead to the desired claim.

- `(P3S1W2)` The claim follows in a similar manner to `P1S1W2` and from the law of total expectation by conditioning on X_M^N.

$\boxtimes$

Proof of Theorem VI.2:

- `P1S2W1` The combination `P1S2` implies that the GAF particles are mapped to the local particles, i.e., $\mathbf{x}_{\mathrm{G}}^{(n)}=\mathbf{x}_{i^{(n)}}^{(j^{(n)})}$ where $(i^{(n)},j^{(n)})=\mathrm{Det}_M^N(n)$ is a deterministic mapping (see Section V-C). Let $\mathcal{J}_m^N\triangleq\{j^{(n)}:(i^{(n)},j^{(n)})=\mathrm{Det}_M^N(n),i^{(n)}=m,\ n\in\mathbb{N}_1^N\}$ and $\kappa_m^N=\#\mathcal{J}_m^N$ which denotes the number of times the deterministic function $\mathrm{Det}_M^N(\cdot)$ maps to the m-th proposal in N consecutive calls. The expectation of the Z-normalized estimator becomes

$$\begin{aligned}\mathbb{E}\{\upvarphi_{\mathrm{G}}^N\}&=\frac{1}{ZN}\sum_{n=1}^N\mathbb{E}\left\{\varphi\left(\mathbf{x}_{i^{(n)}}^{j^{(n)}}\right)\frac{\mathsf{f}_{\mathrm{G}}^N\left(\mathbf{x}_{i^{(n)}}^{(j^{(n)})}\right)}{q_{i^{(n)}}\left(\mathbf{x}_{i^{(n)}}^{(j^{(n)})}\right)}\right\}\\&=\frac{1}{ZN}\sum_{m=1}^M\sum_{j\in\mathcal{J}_m^N}\mathbb{E}\left\{\varphi\left(\mathbf{x}_m^{(j)}\right)\frac{\mathsf{f}_{\mathrm{G}}^N\left(\mathbf{x}_m^{(j)}\right)}{q_m\left(\mathbf{x}_m^{(j)}\right)}\right\}\\&=\frac{1}{Z}\sum_{m=1}^M\frac{\kappa_m^N}{N}\,\mathbb{E}\left\{\varphi\left(\mathbf{x}_m^{(1)}\right)\frac{\mathsf{f}_{\mathrm{G}}^N\left(\mathbf{x}_m^{(1)}\right)}{q_m\left(\mathbf{x}_m^{(1)}\right)}\right\}\end{aligned}\tag{16}$$

where the last equality results by noting that the expectation is identical for all $j \in \mathcal{J}_m^N$. Lemma (S-V.6) and the continuous mapping theorem lead to the convergence

$$\varphi\left(\mathsf{x}_m^{(1)}\right)\frac{\mathsf{f}_{\mathrm{G}}^N\left(\mathsf{x}_m^{(1)}\right)}{q_m\left(\mathsf{x}_m^{(1)}\right)} \xrightarrow{a.s.} \varphi\left(\mathsf{x}_m^{(1)}\right)\frac{Z f_{\mathrm{G}}\left(\mathsf{x}_m^{(1)}\right)}{q_m\left(\mathsf{x}_m^{(1)}\right)} \tag{17}$$

as $N \to \infty$ and for all $m \in \mathbb{N}_1^M$. Furthermore employing the weighted arithmetic mean geometric mean (WAMGM) inequality, a dominating sequence can be constructed as

$$\varphi\left(\mathsf{x}_m^{(1)}\right)\frac{\mathsf{f}_{\mathrm{G}}^N\left(\mathsf{x}_m^{(1)}\right)}{q_m\left(\mathsf{x}_m^{(1)}\right)} \leqslant \sup(\varphi)\sum_{i=1}^M c_i \frac{\hat{\mathsf{p}}_i^N\left(\mathsf{x}_m^{(1)}\right)}{q_m\left(\mathsf{x}_m^{(1)}\right)}. \tag{18}$$

The sequence with general term given in the right-hand-side above, i.e., $\mathsf{y}_N \triangleq \sup(\varphi)\sum_{i=1}^M c_i \frac{\hat{\mathsf{p}}_i^N(\mathsf{x}_m^{(1)})}{q_m(\mathsf{x}_m^{(1)})}$, converges almost surely to $\mathsf{y} \triangleq \sup(\varphi)\sum_{i=1}^M c_i \frac{p_i(\mathsf{x}_m^{(1)})}{q_m(\mathsf{x}_m^{(1)})}$ as $N \to \infty$ via Lemma (S-V.6) and the continuous mapping theorem. Furthermore, the mean $\mathbb{E}\{\mathsf{y}_N\}$ also converges to $\mathbb{E}\{\mathsf{y}\}$ as $N \to \infty$. This follows by first noting that

$$\mathbb{E}\left\{\frac{\hat{\mathsf{p}}_i^N\left(\mathsf{x}_m^{(1)}\right)}{q_m\left(\mathsf{x}_m^{(1)}\right)}\right\} = \int \mathbb{E}\left\{\hat{\mathsf{p}}_i^N(\boldsymbol{x})\right\} d\lambda(\boldsymbol{x}) = 1$$

for all $i \neq m$ and $\forall N$. Subsequently, the decomposition hods

$$\mathbb{E}\left\{\frac{\hat{\mathsf{p}}_m^N\left(\mathsf{x}_m^{(1)}\right)}{q_m\left(\mathsf{x}_m^{(1)}\right)}\right\} = \mathbb{E}\left\{\frac{\tilde{\mathsf{p}}_m^{N,1}\left(\mathsf{x}_m^{(1)}\right)}{q_m\left(\mathsf{x}_m^{(1)}\right)}\right\} + \frac{K(0)}{Nh_N^d}\mathbb{E}\left\{\frac{p_m\left(\mathsf{x}_m^{(1)}\right)}{q_m^2\left(\mathsf{x}_m^{(1)}\right)}\right\} - \frac{1}{N}\mathbb{E}\left\{\frac{p_m\left(\tilde{\mathsf{x}}_m\right)K_{h_N}\left(\tilde{\mathsf{x}}_m - \mathsf{x}_m^{(1)}\right)}{q_m\left(\tilde{\mathsf{x}}_m\right)q_m\left(\mathsf{x}_m^{(1)}\right)}\right\} \tag{19}$$

where $\tilde{\mathsf{x}}_m \sim q_m$ is independent of $\left\{\mathsf{x}_m^{(n)} : \forall n\right\}$ and $\tilde{\mathsf{p}}_m^{N,1}$ is defined in (S-41). Due to the construction of $\tilde{\mathsf{p}}_m^{N,1}$ (see proof of Lemma (S.V.6) from [27]), the first expectation on the right-hand-side above converges to one. The other two terms converge to zero by invoking $p_m/q_m \in L^1(\mathbb{R})$ from assumption A.3(c). It follows that $\{\mathsf{y}_N : N \in \mathbb{N}\}$ is uniformly integrable [54, Thm. 5.5.2], as well as $\left\{\varphi\left(\mathsf{x}_m^{(1)}\right)\frac{\mathsf{f}_{\mathrm{G}}^N(\mathsf{x}_m^{(1)})}{q_m(\mathsf{x}_m^{(1)})} : N \in \mathbb{N}\right\}$ due to (18). The latter coupled with (17) and assuming $\lim_{N\to\infty}\kappa_m^N/N = r_m$ leads to the claim.

- `P1S2W2` Recalling that $\Psi(\boldsymbol{x}) = \sum_{m=1}^M r_m q_m(\boldsymbol{x})$, the expectation of the estimator becomes

$$\mathbb{E}\left\{\varphi_{\mathrm{G}}^N\right\} = \frac{1}{Z}\sum_{m=1}^M \frac{\kappa_m^N}{N}\,\mathbb{E}\left\{\varphi\left(\mathsf{x}_m^{(1)}\right)\frac{\mathsf{f}_{\mathrm{G}}^N\left(\mathsf{x}_m^{(1)}\right)}{\Psi\left(\mathsf{x}_m^{(1)}\right)}\right\}$$

Similarly to the previous case, Lemma (S-V.6) and the continuous mapping theorem lead to the convergence

$$\varphi\left(\mathsf{x}_m^{(1)}\right)\frac{\mathsf{f}_{\mathrm{G}}^N\left(\mathsf{x}_m^{(1)}\right)}{\Psi\left(\mathsf{x}_m^{(1)}\right)} \xrightarrow{a.s.} \varphi\left(\mathsf{x}_m^{(1)}\right)\frac{Z f_{\mathrm{G}}\left(\mathsf{x}_m^{(1)}\right)}{\Psi\left(\mathsf{x}_m^{(1)}\right)} \tag{20}$$

as $N \to \infty$ and for all $m \in \mathbb{N}_1^M$. A dominating sequence is easily constructed from $\varphi\left(\mathsf{x}_m^{(1)}\right)\frac{\mathsf{f}_{\mathrm{G}}^N(\mathsf{x}_m^{(1)})}{\Psi(\mathsf{x}_m^{(1)})} \leqslant r_m^{-1}\sup(\varphi)\frac{\mathsf{f}_{\mathrm{G}}^N(\mathsf{x}_m^{(1)})}{q_m(\mathsf{x}_m^{(1)})}$, where the sequence on the right-hand side forms a uniformly integrable family (see the case `P1S2W1`). Thus, (20) and $\lim_{N\to\infty}\kappa_m^N/N = r_m$ lead to the claim through Vitali's convergence theorem.

⊠

Proof of Theorem VI.3: For all PGAF methods from Table I, Theorems VI.1 and VI.2 assert the convergence $\lim_{N\to\infty}\mathbb{E}\left\{Z\varphi_{\mathrm{G}}^N\right\} = Z\int\varphi(\boldsymbol{x})f_{\mathrm{G}}(\boldsymbol{x})d\lambda(\boldsymbol{x})$. Subsequently, if the variance vanishes $\mathbb{V}\left\{Z\varphi_{\mathrm{G}}^N\right\} \to 0$ as $N \to \infty$, then the consistency of the Z-normalized estimator is guaranteed from [54, Thm. 2.2.4], i.e., $\varphi_{\mathrm{G}}^N \xrightarrow{\mathbb{P}} \int\varphi(\boldsymbol{x})f_{\mathrm{G}}(\boldsymbol{x})d\lambda(\boldsymbol{x})$ as $N \to \infty$. The special case for $\varphi(\boldsymbol{x}) = 1\ \forall\boldsymbol{x}$ also follows: $\frac{1}{N}\sum_{n=1}^N \tilde{\mathsf{w}}_{\mathrm{G}}^N\left(\mathsf{x}_{\mathrm{G}}^{(n)}\right) \xrightarrow{\mathbb{P}} Z$ as $N \to \infty$. Furthermore, the consistency of the self-normalized estimator follows via the continuous mapping theorem and by noticing that $\bar{\varphi}_{\mathrm{G}}^N = \frac{Z\varphi_{\mathrm{G}}^N}{\frac{1}{N}\sum_{n=1}^N \tilde{\mathsf{w}}_{\mathrm{G}}^N(\mathsf{x}_{\mathrm{G}}^{(n)})}$.

Hence it is sufficient to prove that the variance $\mathbb{V}\{Z\varphi_{\mathrm{G}}^N\}$ vanishes as $N \to \infty$ for $\forall\,\varphi \in \mathcal{C}_{\mathrm{b}}(\mathbb{R}^d)$ and each PGAF method. The variance $\mathbb{V}\left\{Z\varphi_{\mathrm{G}}^N\right\}$ becomes

$$\begin{aligned}&\mathbb{V}\left\{\frac{1}{N}\sum_{n=1}^N \varphi\left(\mathsf{x}_{\mathrm{G}}^{(n)}\right)\tilde{\mathsf{w}}_{\mathrm{G}}^N\left(\mathsf{x}_{\mathrm{G}}^{(n)}\right)\right\}\\ &= \frac{1}{N^2}\sum_{n=1}^N \mathbb{E}\left\{\left(\varphi\left(\mathsf{x}_{\mathrm{G}}^{(n)}\right)\tilde{\mathsf{w}}_{\mathrm{G}}^N\left(\mathsf{x}_{\mathrm{G}}^{(n)}\right)\right)^2\right\} - \left[Z\,\mathbb{E}\{\varphi_{\mathrm{G}}^N\}\right]^2\\ &\quad + \frac{1}{N^2}\sum_{n=1}^N\sum_{\ell=1,\ell\neq n}^N \mathbb{E}\left\{\varphi\left(\mathsf{x}_{\mathrm{G}}^{(n)}\right)\tilde{\mathsf{w}}_{\mathrm{G}}^N\left(\mathsf{x}_{\mathrm{G}}^{(n)}\right)\varphi\left(\mathsf{x}_{\mathrm{G}}^{(\ell)}\right)\tilde{\mathsf{w}}_{\mathrm{G}}^N\left(\mathsf{x}_{\mathrm{G}}^{(\ell)}\right)\right\}.\end{aligned} \tag{21}$$

In the following, the variance of (21) is addressed separately for each PGAF method.

- *Consistency of* $\bar{\varphi}_{\mathrm{G}}^N$ *for the case* `P1S1W1`. The first term in (21) is written as

$$\begin{aligned}&\frac{1}{N^2}\sum_{n=1}^N \mathbb{E}\left\{\left(\varphi\left(\mathsf{x}_{\mathrm{G}}^{(n)}\right)\tilde{\mathsf{w}}_{\mathrm{G}}^N\left(\mathsf{x}_{\mathrm{G}}^{(n)}\right)\right)^2\right\}\\ &= \frac{1}{N^2}\sum_{n=1}^N\sum_{m=1}^M r_m\int\varphi^2(\boldsymbol{y})\frac{\mathbb{E}\left\{\left[\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{y})\right]^2\right\}}{q_m(\boldsymbol{y})}d\lambda(\boldsymbol{y})\\ &= \frac{1}{N}\sum_{m=1}^M r_m\int\frac{\varphi^2(\boldsymbol{y})}{q_m(\boldsymbol{y})}\mathbb{E}\left\{\left[\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{y})\right]^2\right\}d\lambda(\boldsymbol{y})\\ &\leqslant \frac{\sup(\varphi^2)}{N}\sum_{m=1}^M r_m\int\frac{1}{q_m(\boldsymbol{y})}\sum_{\ell=1}^M c_\ell\,\mathbb{E}\left\{\left[\hat{\mathsf{p}}_\ell^N(\boldsymbol{y})\right]^2\right\}d\lambda(\boldsymbol{y})\\ &\leqslant \frac{\sup(\varphi^2)M_{\mathrm{K}}}{Nh_N^d}\sum_{m,\ell=1}^M\left(\frac{M_\ell}{N}+1\right)\int\frac{p_\ell * K_{h_N}(\boldsymbol{y})}{q_m(\boldsymbol{y})}d\lambda(\boldsymbol{y})\end{aligned}$$

which is finite for all N and converges to zero as $N\to\infty$ when (13) holds. The last term in (21) can be written as

$$\frac{1}{N^2}\sum_{n=1}^{N}\sum_{\ell=1,\ell\neq n}^{N}\mathbb{E}\big\{\varphi(\mathsf{x}_{\mathrm{G}}^{(n)})\,\tilde{\mathsf{w}}_{\mathrm{G}}^{N}(\mathsf{x}_{\mathrm{G}}^{(n)})\,\varphi(\mathsf{x}_{\mathrm{G}}^{(\ell)})\,\tilde{\mathsf{w}}_{\mathrm{G}}^{N}(\mathsf{x}_{\mathrm{G}}^{(\ell)})\big\}$$
$$=\frac{N-1}{N}\iint\mathbb{E}\big\{\varphi(\boldsymbol{x})\,\mathsf{f}_{\mathrm{G}}^{N}(\boldsymbol{x})\,\varphi(\boldsymbol{y})\,\mathsf{f}_{\mathrm{G}}^{N}(\boldsymbol{y})\big\}d\lambda(\boldsymbol{x})\,d\lambda(\boldsymbol{y})\,. \tag{22}$$

Lemma (S.V.4) (ii) leads to the convergence

$$\mathbb{E}\big\{\varphi(\boldsymbol{x})\,\mathsf{f}_{\mathrm{G}}^{N}(\boldsymbol{x})\,\varphi(\boldsymbol{y})\,\mathsf{f}_{\mathrm{G}}^{N}(\boldsymbol{y})\big\}\to Z^2\,\varphi(\boldsymbol{x})\,f_{\mathrm{G}}(\boldsymbol{x})\,\varphi(\boldsymbol{y})\,f_{\mathrm{G}}(\boldsymbol{y})$$

for $(\lambda\times\lambda)$-a.e. $\boldsymbol{x}$ and $\boldsymbol{y}$ as $N\to\infty$. Furthermore, the integrand $\mathbb{E}\big\{\varphi(\boldsymbol{x})\,\mathsf{f}_{\mathrm{G}}^{N}(\boldsymbol{x})\,\varphi(\boldsymbol{y})\,\mathsf{f}_{\mathrm{G}}^{N}(\boldsymbol{y})\big\}$ on the right-hand side of (22) is bounded by

$$(\sup(\varphi))^2\sum_{m=1}^{M}c_m\,\mathbb{E}\big\{\hat{\mathsf{p}}_m^N(\boldsymbol{x})\,\hat{\mathsf{p}}_m^N(\boldsymbol{y})\big\}$$
$$\leqslant(\sup(\varphi))^2\sum_{m=1}^{M}c_m\left[a_m^N(\boldsymbol{x},\boldsymbol{y})+b_m^N(\boldsymbol{x},\boldsymbol{y})\right]$$

where

$$a_m^N(\boldsymbol{x},\boldsymbol{y})=\frac{M_m}{N}\int p_m(\boldsymbol{z})\,K_{h_N}(\boldsymbol{x}-\boldsymbol{z})\,K_{h_N}(\boldsymbol{y}-\boldsymbol{z})d\lambda(\boldsymbol{z})$$
$$b_m^N(\boldsymbol{x},\boldsymbol{y})=\frac{N-1}{N}\,(p_m*K_{h_N}(\boldsymbol{x}))\,(p_m*K_{h_N}(\boldsymbol{y}))\,.$$

For any m, the functions a_m^N and b_m^N are integrable w.r.t. the product measure $(\lambda\times\lambda)$ for all N and we denote

$$A_m^N\triangleq\iint a_m^N(\boldsymbol{x},\boldsymbol{y})d\lambda(\boldsymbol{x})d\lambda(\boldsymbol{y})=\frac{M_m}{N} \tag{23}$$
$$B_m^N\triangleq\iint b_m^N(\boldsymbol{x},\boldsymbol{y})d\lambda(\boldsymbol{x})d\lambda(\boldsymbol{y})=\frac{N-1}{N}\,. \tag{24}$$

Since $a_m^N(\boldsymbol{x},\boldsymbol{y})\to 0$ and $b_m^N(\boldsymbol{x},\boldsymbol{y})\to p_m(\boldsymbol{x})p_m(\boldsymbol{y})$ for $(\lambda\times\lambda)$-a.e. as well as $\lim_{N\to\infty}A_m^N=0$ and $\lim_{N\to\infty}B_m^N=1$, the application of the generalized DCT w.r.t. the product measure $(\lambda\times\lambda)$ leads to the convergence of the quantity in (22) to $[Z\int\varphi(\boldsymbol{x})f_{\mathrm{G}}(\boldsymbol{x})d\lambda(\boldsymbol{x})]^2$ and subsequently the variance $\mathbb{V}\{Z\bar{\upvarphi}_{\mathrm{G}}^N\}\to 0$ as $N\to\infty$.

- *Consistency of* $\bar{\upvarphi}_{\mathrm{G}}^N$ *for the case* `P1S1W2`. The first term of the variance in (21) is written as

$$\frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}\Big\{\Big(\varphi(\mathsf{x}_{\mathrm{G}}^{(n)})\,\tilde{\mathsf{w}}_{\mathrm{G}}^{N}(\mathsf{x}_{\mathrm{G}}^{(n)})\Big)^2\Big\}$$
$$=\frac{1}{N^2}\sum_{n=1}^{N}\sum_{i^{(n)}=1}^{M}r_{i^{(n)}}\int\frac{\varphi^2(\boldsymbol{y})\mathbb{E}\Big\{\big[\mathsf{f}_{\mathrm{G}}^{N}(\boldsymbol{y})\big]^2\Big\}}{\big[\sum_{m=1}^{M}r_m\,q_m(\boldsymbol{y})\big]^2}q_{i^{(n)}}(\boldsymbol{y})\,d\lambda(\boldsymbol{y})$$
$$\leqslant\frac{\sup(\varphi^2)}{N}\sum_{\ell=1}^{M}\int c_\ell\frac{\mathbb{E}\Big\{\big[\hat{\mathsf{p}}_\ell^N(\boldsymbol{y})\big]^2\Big\}}{\sum_{m=1}^{M}r_m\,q_m(\boldsymbol{y})}d\lambda(\boldsymbol{y})$$
$$\leqslant\frac{\sup(\varphi^2)M_{\mathrm{K}}}{Nh_N^d}\sum_{\ell=1}^{M}\frac{c_\ell}{r_\ell}\left(\frac{M_\ell}{N}+1\right)\int\frac{p_\ell*K_{h_N}(\boldsymbol{y})}{q_\ell(\boldsymbol{y})}d\lambda(\boldsymbol{y})$$

which is finite for all N and converges to zero as $N\to\infty$ when (13) holds. Analogous to the case of `P1S1W1`, the quantity in (22) converges to $[Z\int\varphi(\boldsymbol{x})f_{\mathrm{G}}(\boldsymbol{x})d\lambda(\boldsymbol{x})]^2$ and subsequently the variance $\mathbb{V}\{Z\bar{\upvarphi}_{\mathrm{G}}^N\}\to 0$ as $N\to\infty$ leading to the consistency of the self-normalized estimator for `P1S1W2`.

- *Consistency of* $\bar{\upvarphi}_{\mathrm{G}}^N$ *for the case* `P1S2W1`. Under the proposal and sampling choice `P1S2`, the GAF particles are mapped to local particles, i.e., $\mathsf{x}_{\mathrm{G}}^{(n)}=\mathsf{x}_{i^{(n)}}^{(j^{(n)})}$ where $\big(i^{(n)},j^{(n)}\big)=\mathrm{Det}_M^N(n)$ is a deterministic mapping (see Section V-C). Employing the WAMGM inequality and $|x+y|^p\leqslant 2^p(|x|^p+|y|^p)$, $\forall x,y\in\mathbb{R}$ and $p>0$, the first term in the variance of (21) is bounded by

$$\frac{1}{N^2}\sum_{n=1}^{N}\mathbb{E}\Big\{\Big(\varphi(\mathsf{x}_{i^{(n)}}^{j^{(n)}})\tilde{\mathsf{w}}_{\mathrm{G}}^{N}(\mathsf{x}_{i^{(n)}}^{(j^{(n)})})\Big)^2\Big\}$$
$$\leqslant\frac{\sup(\varphi^2)}{N^2}\sum_{\ell=1}^{M}\kappa_\ell^N\sum_{m=1}^{M}c_m\,\mathbb{E}\left\{\frac{\Big[\hat{\mathsf{p}}_m^N(\mathsf{x}_\ell^{(1)})\Big]^2}{q_\ell^2(\mathsf{x}_\ell^{(1)})}\right\}$$
$$\leqslant\frac{\sup(\varphi^2)}{N^2}\sum_{\ell=1}^{M}\kappa_\ell^N\Bigg\{\frac{4c_\ell K^2(0)}{N^2h_N^{2d}}\left[\sup\left(\frac{p_\ell}{q_\ell^2}\right)\right]^2$$
$$+\frac{4c_\ell M_{\mathrm{K}}}{h_N^d}\left(\frac{M_\ell}{N}+1\right)\int\frac{p_\ell*K_{h_N}(\boldsymbol{y})}{q_\ell(\boldsymbol{y})}d\lambda(\boldsymbol{y})$$
$$+\sum_{\substack{m=1\\m\neq\ell}}^{M}c_m\frac{M_{\mathrm{K}}}{h_N^d}\left(\frac{M_m}{N}+1\right)\int\frac{p_m*K_{h_N}(\boldsymbol{y})}{q_\ell(\boldsymbol{y})}d\lambda(\boldsymbol{y})\Bigg\}$$

which is finite for all N and converges to zero, as $N\to\infty$, when both (13) and (14) hold. Let $\mathcal{J}_m^N\triangleq\big\{j^{(n)}:\big(i^{(n)},j^{(n)}\big)=\mathrm{Det}_M^N(n),i^{(n)}=m,\,n\in\mathbb{N}_1^N\big\}$ and note that $\kappa_m^N=\#\mathcal{J}_m^N$. The last term of (21) becomes

$$\frac{1}{N^2}\sum_{\substack{n,k=1\\n\neq k}}^{N}\mathbb{E}\Bigg\{\varphi(\mathsf{x}_{i^{(n)}}^{(j^{(n)})})\varphi(\mathsf{x}_{i^{(k)}}^{(j^{(k)})})\frac{\mathsf{f}_{\mathrm{G}}^N(\mathsf{x}_{i^{(n)}}^{(j^{(n)})})}{q_{i^{(n)}}(\mathsf{x}_{i^{(n)}}^{(j^{(n)})})}\frac{\mathsf{f}_{\mathrm{G}}^N(\mathsf{x}_{i^{(k)}}^{(j^{(k)})})}{q_{i^{(k)}}(\mathsf{x}_{i^{(k)}}^{(j^{(k)})})}\Bigg\}$$
$$=\frac{1}{N^2}\sum_{m=1}^{M}\sum_{i\in\mathcal{J}_m^N}\sum_{\substack{j\in\mathcal{J}_m^N\\j\neq i}}\mathbb{E}\Bigg\{\varphi(\mathsf{x}_m^{(i)})\varphi(\mathsf{x}_m^{(j)})\frac{\mathsf{f}_{\mathrm{G}}^N(\mathsf{x}_m^{(i)})}{q_m(\mathsf{x}_m^{(i)})}\frac{\mathsf{f}_{\mathrm{G}}^N(\mathsf{x}_m^{(j)})}{q_m(\mathsf{x}_m^{(j)})}\Bigg\}$$
$$+\frac{1}{N^2}\sum_{\substack{m,\ell=1\\m\neq\ell}}^{M}\sum_{i\in\mathcal{J}_m^N}\sum_{j\in\mathcal{J}_\ell^N}\mathbb{E}\Bigg\{\varphi(\mathsf{x}_m^{(i)})\varphi(\mathsf{x}_\ell^{(j)})\frac{\mathsf{f}_{\mathrm{G}}^N(\mathsf{x}_m^{(i)})}{q_m(\mathsf{x}_m^{(i)})}\frac{\mathsf{f}_{\mathrm{G}}^N(\mathsf{x}_\ell^{(j)})}{q_\ell(\mathsf{x}_\ell^{(j)})}\Bigg\}$$
$$=\sum_{m=1}^{M}\frac{\kappa_m^N(\kappa_m^N-1)}{N^2}\mathbb{E}\Bigg\{\varphi(\mathsf{x}_m^{(1)})\varphi(\mathsf{x}_m^{(2)})\frac{\mathsf{f}_{\mathrm{G}}^N(\mathsf{x}_m^{(1)})}{q_m(\mathsf{x}_m^{(1)})}\frac{\mathsf{f}_{\mathrm{G}}^N(\mathsf{x}_m^{(2)})}{q_m(\mathsf{x}_m^{(2)})}\Bigg\}$$
$$+\sum_{m=1}^{M}\sum_{\substack{\ell=1\\\ell\neq m}}^{M}\frac{\kappa_m^N\kappa_\ell^N}{N^2}\mathbb{E}\Bigg\{\varphi(\mathsf{x}_m^{(1)})\varphi(\mathsf{x}_\ell^{(2)})\frac{\mathsf{f}_{\mathrm{G}}^N(\mathsf{x}_m^{(1)})}{q_m(\mathsf{x}_m^{(1)})}\frac{\mathsf{f}_{\mathrm{G}}^N(\mathsf{x}_\ell^{(2)})}{q_\ell(\mathsf{x}_\ell^{(2)})}\Bigg\} \tag{25}$$

where the last equality results by noting that for all $m,\ell\in\mathbb{N}_1^M$ with $m\neq\ell$, the expectation is identical in the cases of $i\neq j$ and $i=j$. Lemma (S-V.6) and the continuous mapping theorem lead to the convergence

$$\frac{\mathsf{f}_{\mathrm{G}}^N(\mathsf{x}_m^{(1)})}{q_m(\mathsf{x}_m^{(1)})}\frac{\mathsf{f}_{\mathrm{G}}^N(\mathsf{x}_\ell^{(2)})}{q_\ell(\mathsf{x}_\ell^{(2)})}\xrightarrow{a.s.}Z^2\frac{f_{\mathrm{G}}(\mathsf{x}_m^{(1)})}{q_m(\mathsf{x}_m^{(1)})}\frac{f_{\mathrm{G}}(\mathsf{x}_\ell^{(2)})}{q_\ell(\mathsf{x}_\ell^{(2)})} \tag{26}$$

as $N \to \infty$ and for all $m, \ell \in \mathbb{N}_1^M$. Furthermore, the WAMGM inequality leads to

$$\varphi\big(\mathsf{x}_m^{(1)}\big)\,\varphi\big(\mathsf{x}_\ell^{(2)}\big)\frac{\mathsf{f}_{\mathrm{G}}^N\big(\mathsf{x}_m^{(1)}\big)}{q_m\big(\mathsf{x}_m^{(1)}\big)}\frac{\mathsf{f}_{\mathrm{G}}^N\big(\mathsf{x}_\ell^{(2)}\big)}{q_\ell\big(\mathsf{x}_\ell^{(2)}\big)} \leqslant \sup(\varphi)^2\sum_{i=1}^M c_i\left[\frac{\hat{\mathsf{p}}_i^N\big(\mathsf{x}_m^{(1)}\big)}{q_m\big(\mathsf{x}_m^{(1)}\big)}\frac{\hat{\mathsf{p}}_i^N\big(\mathsf{x}_\ell^{(2)}\big)}{q_\ell\big(\mathsf{x}_\ell^{(2)}\big)}\right] \quad (27)$$

and to the construction of the dominating sequence $(\mathsf{y}_{i,m,\ell}^N)_N$ with general term $\mathsf{y}_{i,m,\ell}^N \triangleq \frac{\hat{\mathsf{p}}_i^N(\mathsf{x}_m^{(1)})}{q_m(\mathsf{x}_m^{(1)})}\frac{\hat{\mathsf{p}}_i^N(\mathsf{x}_\ell^{(2)})}{q_\ell(\mathsf{x}_\ell^{(2)})}$. Again, Lemma (S-V.6) and the continuous mapping theorem lead to the convergence $\mathsf{y}_{i,m,\ell}^N \xrightarrow{a.s.} \mathsf{y}_{i,m,\ell}$ as $N \to \infty$, $\forall i, m, \ell \in \mathbb{N}_1^M$, and where $\mathsf{y}_{i,m,\ell} \triangleq \frac{p_i(\mathsf{x}_m^{(1)})}{q_m(\mathsf{x}_m^{(1)})}\frac{p_i(\mathsf{x}_\ell^{(2)})}{q_\ell(\mathsf{x}_\ell^{(2)})}$. In the following, we will show that $\mathbb{E}\{\mathsf{y}_{i,m,\ell}^N\} \to \mathbb{E}\{\mathsf{y}_{i,m,\ell}\} = 1$ as $N \to \infty$ and subsequently the family $\{\mathsf{y}_{i,m,\ell}^N\}_N$ is uniformly integrable $\forall i, m, \ell$. For the case $i \neq m \neq \ell$, we obtain

$$\mathbb{E}\{\mathsf{y}_{i,m,\ell}^N\} = \int \mathbb{E}\{\hat{\mathsf{p}}_i^N(\boldsymbol{x})\,\hat{\mathsf{p}}_i^N(\boldsymbol{y})\}d\lambda(\boldsymbol{x})d\lambda(\boldsymbol{y}) = \frac{1}{N}\int\frac{p_i^2(\boldsymbol{z})}{q_i(\boldsymbol{z})}\,d\lambda(\boldsymbol{z}) + \frac{N-1}{N}$$

which converges to one (in virtue of assumption A.3(a)). For the case $i = m \neq \ell$, we obtain

$$\mathbb{E}\{\mathsf{y}_{i,i,\ell}^N\} = \frac{K_{h_N}(0)}{N^2}\int\left[\frac{p_i^2(\boldsymbol{x})}{q_i^2(\boldsymbol{x})} + (N-1)\frac{p_i(\boldsymbol{x})}{q_i(\boldsymbol{x})}\right]d\lambda(\boldsymbol{x}) + \frac{N-1}{N^2}\int\left[\frac{p_i(\boldsymbol{x})}{q_i(\boldsymbol{x})}(p_i * K_{h_N})(\boldsymbol{x}) + \frac{p_i^2(\boldsymbol{x})}{q_i(\boldsymbol{x})}\right]d\lambda(\boldsymbol{x}) + \frac{(N-1)(N-2)}{N^2}$$

where all terms converge to zero (assumption A.3(a) and $p_m/q_m \in L^1(\mathbb{R}^d)$ $\forall m$ due to (14)), except the last term which converges to one. Similarly, for the case $i \neq m = \ell$, it can be easily shown that $\mathbb{E}\{\mathsf{y}_{i,m,m}^N\} \to 1$ as $N \to \infty$. Lastly, for $i = m = \ell$, we obtain

$$\begin{aligned}\mathbb{E}\{\mathsf{y}_{i,i,i}^N\} &= \frac{K_{h_N}^2(0)}{N^2}\left[\int\frac{p_i(\boldsymbol{x})}{q_i(\boldsymbol{x})}\,d\lambda(\boldsymbol{x})\right]^2 + \frac{1}{N^2}\iint\frac{p_i(\boldsymbol{x})}{q_i(\boldsymbol{x})}\frac{p_i(\boldsymbol{y})}{q_i(\boldsymbol{y})}\,K_{h_N}^2(\boldsymbol{x}-\boldsymbol{y})\,d\lambda(\boldsymbol{y})d\lambda(\boldsymbol{x}) \\ &+ \frac{2(N-2)}{N^2}\int\frac{p_i(\boldsymbol{x})}{q_i(\boldsymbol{x})}(p_i * K_{h_N})(\boldsymbol{x})\,d\lambda(\boldsymbol{x}) + \frac{2K_{h_N}(0)}{N^2}\left[\int\frac{p_i^2(\boldsymbol{x})}{q_i^2(\boldsymbol{x})}\,d\lambda(\boldsymbol{x}) + (N-2)\int\frac{p_i(\boldsymbol{x})}{q_i(\boldsymbol{x})}\,d\lambda(\boldsymbol{x})\right] \\ &+ \frac{N-2}{N^2}\int\frac{p_i^2(\boldsymbol{x})}{q_i(\boldsymbol{x})}\,d\lambda(\boldsymbol{x}) + \frac{(N-2)(N-3)}{N^2}\end{aligned}$$

which also converges to one in virtue of assumption A.3(a) and since (14) holds $\forall\ m$. The uniform integrability of the family $\{\mathsf{y}_{i,m,\ell}^N\}_N$ and the inequality in (27) lead to the uniform integrability of the family $\left\{\varphi\big(\mathsf{x}_m^{(1)}\big)\,\varphi\big(\mathsf{x}_\ell^{(2)}\big)\frac{\mathsf{f}_{\mathrm{G}}^N(\mathsf{x}_m^{(1)})}{q_m(\mathsf{x}_m^{(1)})}\frac{\mathsf{f}_{\mathrm{G}}^N(\mathsf{x}_\ell^{(2)})}{q_\ell(\mathsf{x}_\ell^{(2)})} : N \in \mathbb{N}\right\}$.

Coupled with (26), this results in the convergence

$$\lim_{N\to\infty}\mathbb{E}\left\{\varphi\big(\mathsf{x}_m^{(1)}\big)\,\varphi\big(\mathsf{x}_\ell^{(2)}\big)\frac{\mathsf{f}_{\mathrm{G}}^N\big(\mathsf{x}_m^{(1)}\big)}{q_m\big(\mathsf{x}_m^{(1)}\big)}\frac{\mathsf{f}_{\mathrm{G}}^N\big(\mathsf{x}_\ell^{(2)}\big)}{q_\ell\big(\mathsf{x}_\ell^{(2)}\big)}\right\} = \left(Z\int\varphi(\boldsymbol{x})\,f_{\mathrm{G}}(\boldsymbol{x})\,d\lambda(\boldsymbol{x})\right)^2$$

for all $m, \ell \in \mathbb{N}_1^M$. Finally, $\lim_{N\to\infty}\kappa_m^N/N = r_m$ and $\sum_{m=1}^M r_m = 1$ lead to the convergence of (25) to $[Z\int\varphi(\boldsymbol{x})f_{\mathrm{G}}(\boldsymbol{x})d\lambda(\boldsymbol{x})]^2$. Subsequently the variance $\mathbb{V}\{Z\upvarphi_{\mathrm{G}}^N\} \to 0$ as $N \to \infty$ leading to the consistency of the self-normalized estimator with the `P1S2W1` construction.

- *Consistency of $\bar{\upvarphi}_{\mathrm{G}}^N$ for the case* `P1S2W2` follows along similar arguments to the case of `P1S2W1` with the exception being that (13) only needs to hold for all $m, k \in \mathbb{N}_1^M$ with $m = k$.
- *Consistency of $\bar{\upvarphi}_{\mathrm{G}}^N$ for the case* `P2S1W1` follows along similar lines to the case `P1S1W1` by assuming (15) to hold for all $m, k \in \mathbb{N}_1^M$.
- *Consistency of $\bar{\upvarphi}_{\mathrm{G}}^N$ for the case* `P2S1W2` follows along similar lines to the case `P1S1W2` by assuming (15) to hold for all $m, k \in \mathbb{N}_1^M$ with $m = k$.
- *Consistency of $\bar{\upvarphi}_{\mathrm{G}}^N$ for the case* `P3S1W1`. Assuming A.4, note the bound $\frac{\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{x})}{\bar{\mathsf{q}}_m^N(\boldsymbol{x})} \leqslant N^{1-c_m}C$ for some finite constant $C > 0$ (see Section VIII of the Supplemental Material [27]). By the law of total expectation, the first term in (21) is written as

$$\begin{aligned}&\frac{1}{N^2}\sum_{n=1}^N\mathbb{E}\left\{\mathbb{E}\left\{\left(\varphi\big(\mathsf{x}_{\mathrm{G}}^{(n)}\big)\,\tilde{\mathsf{w}}_{\mathrm{G}}^N\big(\mathsf{x}_{\mathrm{G}}^{(n)}\big)\right)^2\,\middle|\,\mathsf{i}^{(n)}, \mathsf{X}_M^N\right\}\right\} \\ &= \frac{1}{N^2}\sum_{n=1}^N\sum_{i^{(n)}=1}^M r_{i^{(n)}}\int\varphi^2(\boldsymbol{y})\,\mathbb{E}\left\{\frac{\big(\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{y})\big)^2}{\bar{\mathsf{q}}_{i^{(n)}}^N(\boldsymbol{y})}\right\}d\lambda(\boldsymbol{y}) \\ &\leqslant C\sum_{m=1}^M r_m\,N^{-c_m}\int\varphi^2(\boldsymbol{y})\,\mathbb{E}\{\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{y})\}\,d\lambda(\boldsymbol{y}) \\ &\leqslant \sup(\varphi^2)C\sum_{m=1}^M r_m\,N^{-c_m}\int\sum_{\ell=1}^M c_\ell\,\mathbb{E}\{\hat{\mathsf{p}}_\ell^N(\boldsymbol{y})\}\,d\lambda(\boldsymbol{y}) \\ &= \sup(\varphi^2)C\sum_{m=1}^M r_m\,N^{-c_m} \quad (28)\end{aligned}$$

where, in the last line, we employed the result $\mathbb{E}\{\hat{\mathsf{p}}_\ell^N(\boldsymbol{x})\} = p_\ell * K_{h_N}(\boldsymbol{x})$. Furthermore, the bound (28) vanishes as $N \to \infty$. The last term in (21) becomes

$$\begin{aligned}&\frac{1}{N^2}\sum_{\substack{n,\ell=1\\ n\neq\ell}}^N\mathbb{E}\left\{\mathbb{E}\left\{\varphi\big(\mathsf{x}_{\mathrm{G}}^{(n)}\big)\frac{\mathsf{f}_{\mathrm{G}}^N\big(\mathsf{x}_{\mathrm{G}}^{(n)}\big)}{\bar{\mathsf{q}}_{\mathsf{i}^{(n)}}^N\big(\mathsf{x}_{\mathrm{G}}^{(n)}\big)} \times \varphi\big(\mathsf{x}_{\mathrm{G}}^{(\ell)}\big)\frac{\mathsf{f}_{\mathrm{G}}^N\big(\mathsf{x}_{\mathrm{G}}^{(\ell)}\big)}{\bar{\mathsf{q}}_{\mathsf{i}^{(\ell)}}^N\big(\mathsf{x}_{\mathrm{G}}^{(\ell)}\big)}\,\middle|\,\mathsf{i}^{(n)}, \mathsf{i}^{(\ell)}, \mathsf{X}_M^N\right\}\right\} \\ &= \frac{N-1}{N}\iint\mathbb{E}\{\varphi(\boldsymbol{x})\,\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{x})\,\varphi(\boldsymbol{y})\,\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{y})\}\,d\lambda(\boldsymbol{x})\,d\lambda(\boldsymbol{y})\end{aligned}$$

with the double integral converging to $[Z \int \varphi(\boldsymbol{x}) f_{\mathrm{G}}(\boldsymbol{x}) d\lambda(\boldsymbol{x})]^2$ according to the same arguments given for the case P1S1W1. This implies that the variance $\mathbb{V}\{Z \upvarphi_{\mathrm{G}}^N\} \to 0$, as $N \to \infty$, for the P3S1W1 PGAF method and thus assuring its consistency.

- The consistency of $\bar{\upvarphi}_{\mathrm{G}}^N$ for the case P3S1W2 by additionally assuming A.4. The first term in (21) becomes

$$\frac{1}{N^2} \sum_{n=1}^{N} \mathbb{E}\left\{ \int \frac{\varphi^2(\boldsymbol{y}) \left[\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{y})\right]^2}{\sum_{j=1}^{M} r_j \, \bar{\mathsf{q}}_j^N(\boldsymbol{y})} d\lambda(\boldsymbol{y}) \right\}$$
$$\leqslant \frac{1}{N r_1} \mathbb{E}\left\{ \int \varphi^2(\boldsymbol{y}) \mathsf{f}_{\mathrm{G}}^N(\boldsymbol{y}) \left(\frac{\mathsf{f}_{\mathrm{G}}^N(\boldsymbol{y})}{\bar{\mathsf{q}}_1^N(\boldsymbol{y})} \right) d\lambda(\boldsymbol{y}) \right\}$$
$$\leqslant C \frac{\sup(\varphi^2)}{r_1 \, N^{c_1}} \int \mathbb{E}\left\{ \mathsf{f}_{\mathrm{G}}^N(\boldsymbol{y}) \right\} d\lambda(\boldsymbol{y})$$

which is finite for all N and vanishes as $N \to \infty$. In a similar manner to the case P3S1W2, the last term in (21) converges to $[Z \int \varphi(\boldsymbol{x}) f_{\mathrm{G}}(\boldsymbol{x}) d\lambda(\boldsymbol{x})]^2$, leading to $\mathbb{V}\{Z \upvarphi_{\mathrm{G}}^N\} \to 0$, as $N \to \infty$. Thus, proving the consistency for the P3S1W2 PGAF method.

⊠

References

[1] A. A. Saucan and M. Z. Win, "Information-seeking sensor selection for ocean-of-things," *IEEE Internet Things J.*, vol. 7, no. 10, pp. 10072–10088, Oct. 2020.

[2] Y. Bar-Shalom, P. K. Willett, and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*. Storrs, CT, USA: Yaakov Bar-Shalom, 2011.

[3] P. K. Varshney, *Distributed Detection and Data Fusion*. Berlin, Germany: Springer, 1997.

[4] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proc. IEEE*, vol. 98, no. 11, pp. 1847–1864, Nov. 2010.

[5] C. E. Rasmussen and C. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.

[6] A. de Doucet, N. Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*. Berlin, Germany: Springer, 2001.

[7] C. P. Robert and G. Casella, *Monte Carlo Statistical Methods*. Berlin, Germany: Springer, 2005.

[8] S. J. Julier and J. K. Uhlmann, "A non-divergent estimation algorithm in the presence of unknown correlations," in *Proc. IEEE Amer. Control Conf.*, 1997, vol. 4, pp. 2369–2373.

[9] J. K. Uhlmann, S. J. Julier, and M. Csorba, "Nondivergent simultaneous map-building and localization using covariance intersection," *Proc. SPIE*, vol. 3087, pp. 2–11, 1997.

[10] Y. Shen and M. Z. Win, "Fundamental limits of wideband localization–Part I: A general framework," *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 4956–4980, Oct. 2010.

[11] J. Li and A. Nehorai, "Distributed particle filtering via optimal fusion of Gaussian mixtures," *IEEE Trans. Signal Inf. Process. Over Netw.*, vol. 4, no. 2, pp. 280–292, Jun. 2018.

[12] M. Z. Win, Y. Shen, and W. Dai, "A theoretical foundation of network localization and navigation," *Proc. IEEE*, vol. 106, no. 7, pp. 1136–1165, Jul. 2018.

[13] O. Tsil, O. Aharon, and A. Carmi, "Distributed estimation using particles intersection," in *Proc. Int. Conf. Inf. Fusion*, 2018, pp. 1653–1660.

[14] M. Z. Win, W. Dai, Y. Shen, G. Chrisikos, and H. V. Poor, "Network operation strategies for efficient localization and navigation," *Proc. IEEE*, vol. 106, no. 7, pp. 1224–1254, Jul. 2018.

[15] A. A. Saucan, M. J. Coates, and M. Rabbat, "A multisensor multi-Bernoulli filter," *IEEE Trans. Signal Process.*, vol. 65, no. 20, pp. 5495–5509, Oct. 2017.

[16] A. Conti et al., "Location awareness in beyond 5G networks," *IEEE Commun. Mag.*, vol. 59, no. 11, pp. 22–27, Nov. 2021.

[17] P. Sharma, A.-A. Saucan, D. J. Bucci, and P. K. Varshney, "Decentralized Gaussian filters for cooperative self-localization and multi-target tracking," *IEEE Trans. Signal Process.*, vol. 67, no. 22, pp. 5896–5911, Nov. 2019.

[18] W.-P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Data fusion trees for detection: Does architecture matter?," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 4155–4168, Sep. 2008.

[19] A. Conti, S. Mazuelas, S. Bartoletti, W. C. Lindsey, and M. Z. Win, "Soft information for localization-of-things," *Proc. IEEE*, vol. 107, no. 11, pp. 2240–2264, Nov. 2019.

[20] F. Nielsen, "An information-geometric characterization of Chernoff information," *IEEE Signal Process. Lett.*, vol. 20, no. 3, pp. 269–272, Mar. 2013.

[21] S. Julier and J. K. Uhlmann, "General decentralized data fusion with covariance intersection," in *Handbook of Multisensor Data Fusion Theory and Practice*, 2nd ed., M. E. Liggins, D. Hall, and J. Llinas, Eds. Boca Raton, FL, USA: CRC Press, 2009, ch. 14, pp. 319–343.

[22] K. Chang, C. Chong, and S. Mori, "Analytical and computational evaluation of scalable distributed fusion algorithms," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 4, pp. 2022–2034, Oct. 2010.

[23] G. Hao, Y. Li, M. Zhao, H. Li, and Y. Dou, "Covariance intersection fusion particle filter for nonlinear systems," in *Proc. IEEE Chin. Control Decis. Conf.*, 2017, pp. 5501–5504.

[24] G. Soysal and M. Efe, "Data fusion in a multistatic radar network using covariance intersection and particle filtering," in *Proc. IEEE Int. Conf. Inf. Fusion*, 2011, pp. 1–7.

[25] O. Tsil and A. Carmi, "Information fusion using particles intersection," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, 2018, pp. 4269–4273.

[26] O. Tslil, N. Lehrer, and A. Carmi, "Approaches to Chernoff fusion with applications to distributed estimation," *Digit. Signal Process.*, vol. 107, 2020, Art. no. 102877.

[27] A. A. Saucan, V. Elvira, P. K. Varshney, and M. Z. Win, "Information fusion via importance sampling: Supplementary material," *IEEE Trans. Signal Inf. Process. Netw.*, doi: 10.1109/TSIPN.2023.3299512.

[28] V. Elvira, L. Martino, D. Luengo, and M. Bugallo, "Generalized multiple importance sampling," *Stat. Sci.*, vol. 34, no. 1, pp. 129–155, 2019.

[29] L. Devroye and L. Györfi, *Nonparametric Density Estimation : The L1 View* (Wiley Interscience Series in Discrete Mathematics). Hoboken, NJ, USA:Wiley, 1985.

[30] P. Wu, T. Imbiriba, V. Elvira, and P. Closas, "Bayesian data fusion with shared priors," 2022, *arXiv:2212.07311*.

[31] O. Cappé, P. Pudlo, and M. Sedki, "Adaptive importance sampling in general mixture classes," *Statist. Comput.*, vol. 18, no. 4, pp. 447–459, 2008.

[32] R. Douc, A. Guillin, J.-M. Marin, and C. P. Robert, "Convergence of adaptive mixtures of importance sampling schemes," *Ann. Statist.*, vol. 35, no. 1, pp. 420–448, 2007.

[33] J.-M. Marin, P. Pudlo, and M. Sedki, "Consistency of adaptive importance sampling and recycling schemes," *Bernoulli*, vol. 25, no. 3, pp. 1977–1998, 2019.

[34] A. A. Saucan, T. Chonavel, C. Sintes, and J.-M. Le Caillec, "CPHD-DOA tracking of multiple extended sonar targets in impulsive environments," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1147–1160, Mar. 2016.

[35] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*. Berlin, Germany: Springer, 2004.

[36] D. Luengo, L. Martino, V. Elvira, and M. Bugallo, "Efficient linear fusion of partial estimators," *Digit. Signal Process.*, vol. 78, pp. 265–283, 2018.

[37] A. Owen and Y. Zhou, "Safe and effective importance sampling," *J. Amer. Statist. Assoc.*, vol. 95, no. 449, pp. 135–143, 2000.

[38] A. B. Owen, "Monte Carlo theory, methods and examples," 2013. Accessed: May 05, 2023. [Online],Aavailable: https://statweb.stanford.edu/owen/mc/

[39] V. Elvira and L. Martino, "Advances in importance sampling," *Wiley StatsRef: Statist. Reference Online*, pp. 1–22, 2021.

[40] O. Hlinka, O. Slučiak, F. Hlawatsch, P. M. Djurić, and M. Rupp, "Likelihood consensus and its application to distributed particle filtering," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4334–4349, Aug. 2012.

[41] D. Koller and N. Friedman, *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA, USA: MIT Press, 2009.

[42] T. Minka, "Divergence measures and message passing," Microsoft Research, Ltd., Cambridge, U.K., Tech. Rep. MSR-TR-2005-173, 2005.

[43] G. Koliander, Y. El-Laham, P. M. Djurić, and F. Hlawatsch, "Fusion of probability density functions," *Proc. IEEE*, vol. 110, no. 4, pp. 404–453, Apr. 2022.

[44] S. J. Julier and J. K. Uhlmann, "A new extension of the Kalman filter to nonlinear systems," *Proc. SPIE*, vol. 3068, pp. 182–193, 1997.

[45] M. F. Bugallo, V. Elvira, L. Martino, D. Luengo, J. Míguez, and P. M. Djurić, "Adaptive importance sampling: The past, the present, and the future," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 60–79, Jul. 2017.

[46] M. Sbert and V. Elvira, "Generalizing the balance heuristic estimator in multiple importance sampling," *Entropy*, vol. 24, no. 2, 2022, Art. no. 191.

[47] H. Y. He and A. B. Owen, "Optimal mixture weights in multiple importance sampling," 2014, *arXiv:1411.3954*.

[48] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Efficient multiple importance sampling estimators," *IEEE Signal Process. Let.*, vol. 22, no. 10, pp. 1757–1761, Oct. 2015.

[49] E. Veach and L. Guibas, "Optimally combining sampling techniques for Monte Carlo rendering," in *Proc. 22nd Int. ACM Conf. Comput. Graph. Interactive Techn.*, 1995, pp. 419–428.

[50] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Stat.*, vol. 33, no. 3, pp. 1065–1076, Sep. 1962.

[51] T. Cacoullos, "Estimation of a multivariate density," *Ann. Inst. Stat. Math.*, vol. 18, no. 1, pp. 179–189, Dec. 1966.

[52] L. P. Devroye and T. J. Wagner, "The L_1 convergence of kernel density estimates," *Ann. Statist.*, vol. 7, no. 5, pp. 1136–1139, 1979.

[53] J. F. Geweke, "Bayesian inference in economic models using Monte Carlo integration," *Econometrica*, vol. 45, no. 7, pp. 788–799, 1989.

[54] R. Durrett, *Probability: Theory and Examples*, 4th ed. New York, NY, USA: Cambridge Univ. Press, 2013.

Augustin A. Saucan (Member, IEEE) received the M.Sc. and Ph.D. degrees in electrical engineering from Institut Mines-Télécom Atlantique (formerly Télécom Bretagne), Brest, France, in 2012 and 2016, respectively.

From 2016 to 2017, he held several Postdoctoral appointments with McGill University, Montreal, QC, Canada; also with Syracuse University, Syracuse, NY, USA, from 2017 to 2019; and also the Wireless Information and Network Sciences Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA, from 2019 to 2022. In 2022, he joined Télécom SudParis, Institut Polytechnique de Paris (IP Paris), Palaiseau, France, as an Assistant Professor. His research interests include statistical signal processing, Monte Carlo methods, localization and tracking, and reinforcement learning.

Dr. Saucan was a reviewer of several journals, including IEEE TRANSACTION ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS, and IEEE TRANSACTIONS ON SIGNAL AND INFORMATION PROCESSING OVER NETWORKS.

Víctor Elvira (Senior Member, IEEE) received the Ph.D. degree in statistical signal processing from the University of Cantabria, Santander, Spain, in 2011.

From 2013 to 2016, he was an Assistant Professor with the University Carlos III of Madrid, Getafe, Spain. From 2016 to 2019, he was an Associate Professor with the Engineering School IMT Lille Douai, France. He is currently a Reader (Associate Professor) of statistics and data science with the School of Mathematics, and the Director of the Centre for Statistics, The University of Edinburgh, Edinburgh, U.K. He was also a Visiting Researcher with several institutions, such as Stony Brook University of New York, Stony Brook, NY, USA, and Paris-Dauphine University, Paris, France.

He has coauthored more than 120 journal and refereed conference papers. His research interests include computational statistics, statistical signal processing, and machine learning, in particular in Bayesian inference and Monte Carlo methods with different applications including sensor networks, wireless communications, target tracking, ecology, and biomedicine. He is also a Fulbright Fellow, Marie Curie Fellow, Leverhulme Fellow, and Alan Turing Fellow.

Pramod K. Varshney (Life Fellow, IEEE) was born in Allahabad, India, in 1952. He received the B.S. degree (Highest Hons.) in electrical engineering and computer science and the M.S. and Ph.D. degrees in electrical engineering from the University of Illinois at Urbana–Champaign, Champaign, IL, USA, in 1972, 1974, and 1976, respectively.

From 1972 to 1976, he held teaching and research assistantships with the University of Illinois. Since 1976, he has been with Syracuse University, Syracuse, NY, USA, where he is currently a Distinguished Professor of electrical engineering and computer science and the Director of the Center for Advanced Systems and Engineering. As the Director of CASE, he is responsible for technology transition of university expertise to make economic impact in the high-tech economy of New York. He was an Associate Chair of the Department from 1993 to 1996. He is also an Adjunct Professor of radiology with Upstate Medical University, Syracuse, NY, USA. He has published extensively. He is the author of *Distributed Detection and Data Fusion* (Springer, 1997). His research interests include distributed sensor networks and data fusion, detection and estimation theory, wireless communications, image processing, radar signal processing, physical layer security, and machine learning.

Dr. Varshney was a James Scholar, a Bronze Tablet Senior, and a Fellow while with the University of Illinois. He is a Member of Tau Beta Pi. He was the recipient of the 1981 ASEE Dow Outstanding Young Faculty Award. In 1997, he was elected to the grade of fellow of the IEEE for his contributions in the area of distributed detection and data fusion. In 2000, he was the recipient of the Third Millennium Medal from the IEEE and Chancellor's Citation for Exceptional Academic Achievement at Syracuse University. He was also the recipient of the IEEE 2012 Judith A. Resnik Award, Doctor of Engineering Honoris Causa from Drexel University in 2014, ECE Distinguished Alumni Award from the University of Illinois in 2015, ISIFs Yaakov Bar-Shalom Award for a Lifetime of Excellence in Information Fusion in 2018, Claude Shannon-Harry Nyquist Technical Achievement Award from the IEEE Signal Processing Society, Pioneer Award from the IEEE Aerospace and Electronic Society in 2021, and Syracuse University Chancellors Lifetime Achievement Award in 2023. He was the Guest Editor of the special issue on data fusion of the IEEE PROCEEDINGS in 1997. He has served on the Editorial Boards of the *Journal on Advances in Information Fusion*, IEEE TRANSACTIONS OF SIGNAL PROCESSING, and *IEEE Signal Processing Magazine*. He was the President of International Society of Information Fusion in 2001.

Moe Z. Win (Fellow, IEEE) is a Professor at the Massachusetts Institute of Technology and the founding director of the Wireless Information and Network Sciences Laboratory. Prior to joining MIT, he was with AT&T Research Laboratories and with NASA Jet Propulsion Laboratory.

His research encompasses fundamental theories, algorithm design, and network experimentation for a broad range of real-world problems. His current research topics include ultra-wideband systems, network localization and navigation, network interference exploitation, and quantum information science. He has served the IEEE Communications Society as an elected Member-at-Large on the Board of Governors, as elected Chair of the Radio Communications Committee, and as an IEEE Distinguished Lecturer. Over the last two decades, he held various editorial positions for IEEE journals and organized numerous international conferences. He has served on the SIAM Diversity Advisory Committee.

Dr. Win is an elected Fellow of the AAAS, EURASIP, IEEE, and IET. He was honored with two IEEE Technical Field Awards: the IEEE Kiyo Tomiyasu Award (2011) and the IEEE Eric E. Sumner Award (2006, jointly with R. A. Scholtz). His publications, coauthored with students and colleagues, have received several awards. Other recognitions include the MIT Everett Moore Baker Award (2022), IEEE Vehicular Technology Society James Evans Avant Garde Award (2022), IEEE Communications Society Edwin H. Armstrong Achievement Award (2016), Cristoforo Colombo International Prize for Communications (2013), Copernicus Fellowship (2011) and *Laurea Honoris Causa* (2008) from the Università degli Studi di Ferrara, and U.S. Presidential Early Career Award for Scientists and Engineers (2004).