Gauss-Hermite Quadrature for non-Gaussian Inference via an Importance Sampling Interpretation

Víctor Elvira
Communication Systems Dept.
IMT Lille Douai & CRIStAL laboratory
Villeneuve d'Ascq, France
victor.elvira@imt-lille-douai.fr

Pau Closas
Electrical & Computer Engineering Dept.
Northeastern University
Boston, USA
closas@northeastern.edu

Luca Martino
Signal Theory Dept.
Universidad Rey Juan Carlos
Madrid, Spain
luca.martino@urjc.es

***Abstract*—Intractable integrals appear in a plethora of problems in science and engineering. Very often, such integrals involve also a targeted distribution which is not even available in a closed form. In both cases, approximations of the integrals must be performed. Monte Carlo (MC) methods are a usual way of tackling the problem by approximating the integral with random samples. Quadrature methods are another alternative, where the integral is approximated with deterministic points and weights. However, the choice of these points and weights is only possible in a selected number of families of distributions. In this paper, we propose a deterministic method inspired in MC for approximating generic integrals. Our method is derived via an importance sampling (IS) interpretation, a MC methodology where the samples are simulated from the so-called proposal density, and weighted properly. We use Gauss-Hermite quadrature rules for Gaussian distributions, transforming them for approximating integrals with respect to generic distributions, even in the case where its normalizing constant is unknown. The novel method allows the use of several proposal distributions, allowing for the incorporation of recent advances in the multiple IS (MIS) literature. We discuss the convergence of the method, and we illustrate its performance with two numerical examples.**

***Index Terms*—Gauss-Hermite quadrature, importance sampling, Monte Carlo, Bayesian inference**

I. Introduction

Most inference problems require the approximation of intractable integrals. Monte Carlo (MC) methods are a family of statistical algorithms that allow for the approximation of such integrals, typically involving a targeted distribution [1]. Importance sampling (IS) is a very flexible MC methodology that approximates this kind of integrals by sampling from a so-called *proposal* distribution, circumventing the simulation from the often intractable targeted distribution which usually can be evaluated only up to a normalizing constant [2]. Another common approach for approximating these integrals is the use of quadrature methods (a.k.a. Gaussian quadrature rules) [3]. Quadrature methods approximate a targeted integral by selecting deterministically a set of N pairs of points (or nodes) and associated weights. Quadrature methods have been successfully used in signal processing problems mostly in the context of stochastic filtering in state-space models [4], e.g., the Quadrature Kalman filter (QKF) [3], [5], [6] and its adaptation for high-dimensional setups [7]–[9]. One of the reasons that hinders a wider applicability in batch inference is that the choice of points and weights is only clear for few families of integrands, i.e., integrals that involve specific families of probability density functions. For instance, Gauss-Hermite methods are designed to integrate functions with respect to (w.r.t.) Gaussian distributions [10]. Another limitation of quadrature methods is their need to evaluate the integrand. This is not always possible, e.g., in Bayesian inference the targeted distribution is available only up to a normalizing constant. In summary, the scope of quadrature methods is the same as in IS methods, providing better performance in very specific scenarios, but being restricted to a subset of problems compared to IS.

In this paper, we extend the applicability and improve the performance of quadrature methods by exploiting an IS perspective. In particular, we transform intricate problems that involve intractable integrals w.r.t. non-Gaussian distributions to a problem where Gauss-Hermite quadrature can be applied. This transformation is possible through the introduction of an auxiliary Gaussian *proposal* distribution (both multiplying and dividing in the integrand), similarly to the usual re-arrangement in IS. Due to this similarity, we call our method *importance Gauss-Hermite* (IGH). Then, we discuss the choice of those parameters analyzing the similarities and differences with the problem of choosing a proposal distribution in IS. We show that the use of a unique proposal, like in IS, is usually too constraining, and we extend the basic IGH framework to the case where several Gaussian *proposals* are introduced. Similarly to the case of multiple IS (MIS), the use of several proposals opens the door for many possible schemes. We propose and discuss two possibilities, inspired by the MIS literature [11].

The rest of the paper is organized as follows. In Section II we present the problem and briefly review Gauss-Hermite

V.E. acknowledges support from the *Agence Nationale de la Recherche* of France under PISCES project (ANR-17-CE40-0031-01). P.C. was supported by the National Science Foundation under Awards CNS-1815349 and ECCS-1845833.

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

quadrature and IS. In Section III we present the basic IGH framework and the extension with several proposals. Finally, we show two numerical examples in Section IV, and some conclusions in Section IV.

II. Problem statement and background

This section provides insights on the basic problem we address, the Gauss-Hermite rules, and a brief summary of importance sampling. Building on those concepts, the next section describes the proposed methodology.

A. Addressed problem

We are interested in computing the (possibly intractable integral)

$$I = \int f(\mathbf{x})\tilde{\pi}(\mathbf{x})d\mathbf{x}, \tag{1}$$

where $\tilde{\pi}(\mathbf{x})$ is a probability density function (pdf), and f is an integrable function w.r.t. $\tilde{\pi}(\mathbf{x})$. Very often, the normalizing constant Z of $\tilde{\pi}(\mathbf{x})$ is unknown, and one has access only to the unnormalized non-negative function $\pi(\mathbf{x})$, in such a way that $\tilde{\pi}(\mathbf{x}) = \frac{\pi(\mathbf{x})}{Z}$. For instance, in the context of Bayesian inference, $\tilde{\pi}(\mathbf{x})$ is the posterior distribution of some unknown parameter $\mathbf{x} \in \mathbb{R}^{d_x}$ conditioned to the data (omitted in the notation), $\pi(\mathbf{x})$ is the product of likelihood and prior, and Z is the marginal likelihood.

B. Gauss-Hermite quadrature for Gaussian distributions

Let us consider the integral of the form

$$I = \int h(\mathbf{x})\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})d\mathbf{x}, \tag{2}$$

where $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a Gaussian pdf with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$, and h is an integrable function w.r.t. $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ Note that Eq. (2) is a particular case of Eq. (1). It is noteworthy that the integral in Eq. (2) can be efficiently approximated with certain deterministic rules usually referred to as *quadratures* for the one-dimensional case where $d_x = 1$ or *cubature*, for higher dimensions [12].[1]

Quadrature methods approximate the integral in Eq. (2) with a set of N weighted points $\mathcal{S} = \{\mathbf{x}_n, v_n\}_{n=1}^N$, where $\mathbf{x}_n$ is the n-th deterministically chosen point and v_n its associated quadrature weight. The integral is then approximated by

$$\hat{I}_{\text{Q}} = \sum_{n=1}^{N} v_n h(\mathbf{x}_n). \tag{3}$$

The way the points are selected goes beyond the scope of the paper, but intuitively one can think that the selected points are representative of the Gaussian distribution and the weights ensure both a convergence when N grows, and a low error (or even a perfect approximation) for a given N. In this paper, and without loss of generality, we focus on Gauss-Hermite quadrature for integrals in $\mathbb{R}^{d_x}$ [13] although other rules could be considered. In this scheme, α points per dimension are selected according to a deterministic rule. Hence, $N = \alpha^{d_x}$ points are chosen forming a lattice in a d_x-dimensional space. Note that the exponential growth of points with the dimension can be alleviated with sparse methods [14]. For a polynomial function h of order p in Eq. (2), the integral is computed without error if $p \leq 2\alpha - 1$. In the case where $p > 2\alpha - 1$, then a bounded approximation error can found [15].

C. Importance Sampling (IS)

Importance sampling is another alternative for approximating Eq. (1) by first rewriting it as

$$\begin{aligned} I &= \int f(\mathbf{x})\tilde{\pi}(\mathbf{x})d\mathbf{x} \\ &= \int \frac{f(\mathbf{x})\tilde{\pi}(\mathbf{x})}{q(\mathbf{x})} q(\mathbf{x})d\mathbf{x}, \end{aligned} \tag{4}$$

where q is a pdf with non-zero value for all $\mathbf{x}$ where $f(\mathbf{x})\tilde{\pi}(\mathbf{x})$ is not zero. Due to Monte Carlo arguments, the integral in Eq. (4) can be approximated by a sampling-weighting procedure. First, a set of N samples $\{\mathbf{x}_n\}_{n=1}^N$ is randomly simulated from the so-called proposal pdf, q. Second, an importance weight is assigned to each sample as

$$w_n = \frac{\pi(\mathbf{x}_n)}{q(\mathbf{x}_n)}, \quad n = 1, \ldots, N. \tag{5}$$

Finally, the *unnormalized* IS (UIS) estimator is given by

$$\hat{I}_{\text{IS}} = \frac{1}{NZ} \sum_{n=1}^{N} w_n f(\mathbf{x}_n). \tag{6}$$

We remind that Z is usually unknown, which precludes the use of the UIS estimator in general. The alternative self-normalized IS (SNIS) estimator approximates Eq. (1) by

$$\tilde{I}_{\text{IS}} = \sum_{n=1}^{N} \bar{w}_n f(\mathbf{x}_n), \tag{7}$$

where $\bar{w}_n = \frac{w_n}{\sum_{j=1}^N w_j}$ are the normalized weights. The choice of the proposal is a hard but key problem since the variance of UIS and SNIS estimators increase with the discrepancy between $\pi(\mathbf{x})|f(\mathbf{x})|$ and $q(\mathbf{x})$ [11], [16], [17]. Hence, adaptive schemes are usually implemented in order to iteratively increase the efficiency of the method by improving the proposal pdf [18].

III. Gauss-Hermite quadrature for non-Gaussian distributions

A. Basic importance Gaussian-Hermite (IGH) algorithm

Let us recall that in the targeted integral of Eq. (1), $\tilde{\pi}$ is in general a non-Gaussian distribution, and hence, Gauss-Hermite quadrature cannot be applied directly. However, we propose to make use of the IS trick of Eq. (4) and re-write I as

$$I \equiv \int h(\mathbf{x})q(\mathbf{x})d\mathbf{x}, \tag{8}$$

where we set $q(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$ parameters, and $h(\mathbf{x}) \equiv \frac{f(\mathbf{x})\tilde{\pi}(\mathbf{x})}{q(\mathbf{x})}$. Note that the Gaussian

[1] For the sake of brevity, in this paper we will use the term *quadrature* indistinctly regardless the dimension.

pdf q plays a similar role as the proposal distribution in IS. The problem now can be addressed with Gauss-Hermite quadrature rules as in Eq. (2), since we have an integral of the form of a nonlinear function times a Gaussian distribution. Algorithm1 summarizes the algorithm that we have named *importance Gauss-Hermite* method. In Step 1, we select the set of points according to Gauss-Hermite quadrature rules. In Step 2, we compute the importance weights similarly to those used in IS. In Step 3, we combine the importance and the quadrature weights to construct the weights w'_n. The multiplication factor N becomes apparent in Step 4, where the unnormalized IGH estimator $\hat{I}_{\text{IGH}}$ has a clear parallelism with the estimator $\hat{I}_{\text{IS}}$ of the UIS of Eq. (6). Like in IS, the unnormalized estimator can be used only when the normalizing constant Z is known.

B. Discussion

Note that Algorithm 1 could be re-written by simply evaluating h in all points, instead of separating the evaluation of $\frac{\tilde{\pi}(\mathbf{x})}{q(\mathbf{x})}$ and $f(\mathbf{x})$. However, this formulation (inspired in IS) allows us to go beyond the simple re-arrangement of Eq. (8). For instance, now we are able to build the self-normalized IGH estimator

$$\tilde{I}_{\text{IGH}} = \sum_{n=1}^{N} \bar{w}'_n f(\mathbf{x}_n) \tag{9}$$

where $\bar{w}'_n = \frac{w'_n}{\sum_{j=1}^{N} w'_j}$. The estimator $\tilde{I}_{\text{IGH}}$ can be used even if Z is unknown. In this case, the normalizing constant Z can be approximated with the estimator

$$\hat{Z}_{\text{IGH}} = \frac{1}{N}\sum_{n=1}^{N} w'_n \tag{10}$$

at no extra cost, since it only requires the weights w'_n of Step 3. Note that $\tilde{I}_{\text{IGH}}$ can be derived by substituting $\hat{Z}_{\text{IGH}}$ in $\hat{I}_{\text{IGH}}$, when Z is unknown, which ensures its convergence with N. In the re-arrangement of Eq. (8) that leads to Algorithm 1, we have introduced the auxiliary Gaussian pdf q, which requires selecting the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Similarly to what happens in IS, a low error in the IGH estimator depends on an appropriate choice of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. The optimal IS proposal is $q(\mathbf{x}) \propto \pi(\mathbf{x})|f(\mathbf{x})|$, which yields zero variance estimators [16]. In general, the use of this proposal is not possible, and the strategy is usually in selecting a proposal that minimizes the mismatch w.r.t. $\pi(\mathbf{x})|f(\mathbf{x})|$. In Gauss-Hermite quadrature, the integral is perfectly computed if the function $h(\mathbf{x})$ is polynomial with order $p \leq 2\alpha - 1$. Then, an appropriate criterion is selecting $q(\mathbf{x})$ in such a way $h(\mathbf{x}) = \frac{f(\mathbf{x})\tilde{\pi}(\mathbf{x})}{q(\mathbf{x})}$ can be approximated with a low order polynomial. Note that, if we apply the optimal (and unfeasible) IS proposal $q(\mathbf{x}) \propto \pi(\mathbf{x})|f(\mathbf{x})|$ to Algorithm 1, then $h(\mathbf{x}) = I$, i.e., it is a constant with the true value, yielding both $\hat{I}_{\text{IGH}}$ and $\tilde{I}_{\text{IGH}}$ estimators with the exact value. As a general note, this novel IS interpretation for applying Gauss-Hermite rules in a non-Gaussian problem allows us to use recent advances in the IS literature for improving the performance of the basic IGH method and enlarging its range of applicability beyond the restrictive case of Gaussian distributions.

Algorithm 1 Basic importance Gaussian-Hermite (IGH)

Input: Number of points $N = \alpha^{d_x}$, and parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$

1: Select $\mathcal{S} = \{\mathbf{x}_n, v_n\}_{n=1}^{N}$ according to Gauss-Hermite rules.

2: Compute the importance weights

$$w_n = \frac{\pi(\mathbf{x}_n)}{q(\mathbf{x}_n)}, \qquad n = 1, \ldots, N \tag{11}$$

3: Compute the quadrature important weights

$$w'_n = w_n v_n N \tag{12}$$

4: The unnormalized estimator is built as

$$\hat{I}_{\text{IGH}} = \frac{1}{ZN}\sum_{n=1}^{N} w'_n f(\mathbf{x}_n) \tag{13}$$

when Z is known.

Output: $\{\mathbf{x}_n, w'_n\}_{n=1}^{N}$

C. Standard multiple IGH (SM-IGH)

The IS interpretation opens the door for extending the basic IGH algorithm of previous section to more complicated setups, making use of recent advances in multiple IS (MIS) [11]. In previous section, we discuss that when q approximates well the integrand, then the IGH estimators improve their performance. However, $\tilde{\pi}$ can be skewed, multimodal, or with different tails than a Gaussian. While the Gaussian restriction is limiting, it is widely accepted that under mild assumptions, a non-negative function can be approximated by a mixture of Gaussians [19], [20]. Let us then consider a set of M Gaussian pdfs $\{q_m(\mathbf{x})\}_{m=1}^{M}$, with $q_m(\mathbf{x}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)$. In this case, a re-arrangement similar to Eq. (8) is possible such that

$$I = \frac{1}{M}\sum_{m=1}^{M}\int \frac{f(\mathbf{x})\tilde{\pi}(\mathbf{x})}{q_m(\mathbf{x})} q_m(\mathbf{x}) d\mathbf{x} \tag{14}$$

$$= \frac{1}{M}\sum_{m=1}^{M}\int h_m(\mathbf{x}) q_m(\mathbf{x}) d\mathbf{x}, \tag{15}$$

where $h_m(\mathbf{x}) = \frac{f(\mathbf{x})\tilde{\pi}(\mathbf{x})}{q_m(\mathbf{x})}$. We can solve the M integrals in Eq. (8) by performing M IGH algorithms in parallel as described in previous section, each with N quadrature points. We denote this quadrature algorithm as *standard multiple IGH* (SM-IGH), where $\hat{I}_{\text{SM-IGH}}$ is built as an equally weighted combination of the M IGH algorithms, and the self-normalized $\tilde{I}_{\text{SM-IGH}}$ normalizes the set of MN weights before combining all points. Note that this re-arrangement is similar to the MIS interpretation of known algorithms, such as the PMC algorithm [21]. However, note that if $\tilde{\pi}$ is complicated (i.e., it does not have a standard form), regardless the choice of parameters, all Gaussians $q_m(\mathbf{x})$ will be unable to mimic the target, yielding $h_m(\mathbf{x})$ very different from a low-order polynomial. This can be a problem regardless the number M of employed Gaussians. Interestingly, the re-arrangement of (14) is inspired in the standard multiple MIS scheme (SM-MIS), a.k.a. `N1` scheme in [11], which is known to show a poor performance in terms of variance w.r.t. other advanced MIS schemes (see also the discussion in [22, Section 4.1.1.]).

D. Deterministic mixture IGH (DM-IGH)

Let us define the mixture $\psi(\mathbf{x}) \equiv \frac{1}{M}\sum_{m=1}^{M} q_m(\mathbf{x})$ of all previously defined Gaussian proposals. An alternative re-arrangement is

$$I = \int \frac{f(\mathbf{x})\tilde{\pi}(\mathbf{x})}{\psi(\mathbf{x})}\psi(\mathbf{x})d\mathbf{x}$$

$$= \int \frac{f(\mathbf{x})\tilde{\pi}(\mathbf{x})}{\psi(\mathbf{x})}\frac{1}{M}\sum_{m=1}^{M} q_m(\mathbf{x})d\mathbf{x} \quad (16)$$

$$= \frac{1}{M}\sum_{m=1}^{M}\int h(\mathbf{x})q_m(\mathbf{x})d\mathbf{x}, \quad (17)$$

where $h(\mathbf{x}) = \frac{f(\mathbf{x})\tilde{\pi}(\mathbf{x})}{\psi(\mathbf{x})}$ does not depend on m. First, note Eq. (16) is inspired in the MIS scheme known as deterministic mixture MIS (DM-MIS), a.k.a. `N3` scheme in [11], that has been recently shown theoretically to provide the best performance in terms of variance of the UIS estimator [11]. Second, the M integrands share the same function $h(\mathbf{x})$, where all Gaussians appear at its denominator. This facilitates the search of a set of M Gaussians in such a way $h(\mathbf{x})$ can be approximated by a low-order polynomial. The search of a good mixture approximation of a given distribution can be performed with algorithms such as EM [23]. Ideally, if the mixture of Gaussians $\psi(\mathbf{x}) \propto |f(\mathbf{x})|\tilde{\pi}(\mathbf{x})$, then the integration is perfect, unlike in SM-IGH. Algorithm 2 describes the DM-IGH method with the so-called DM weights of Eq. (18), where all the Gaussian proposals appear in the denominator of all weights (unlike in SM-IGH).

IV. Numerical example

A. Example 1: Unimodal unidimensional example

Let us consider a unidimensional generalized Gaussian distribution (GGD) as a target pdf, i.e., $\tilde{\pi}(x) = \mathcal{GG}(x;\nu,c,\beta) \equiv \kappa \exp\left(-\left(\frac{|x-\nu|}{c}\right)^{\beta}\right)$, where $\kappa = \frac{\beta}{2c\Gamma\left(\frac{1}{\beta}\right)}$, and $\Gamma(\cdot)$ is the gamma function. In this example, we set $\nu = 0$, $c = 1$, and $\beta = 4$. We test the single proposal IS method of Section II-C, and the basic IGH for approximating the mean of the target, with Gaussian proposal $q(x) = \mathcal{N}(x;\mu,\sigma)$ of parameters $\mu = 1$, and $\sigma = 1.3$. Figure 1 shows the (mean) absolute error of the unnormalized and self-normalized estimators when we increase the number of samples N. The error decreases with N in all estimators, and that IGH obtains a better convergence rate in this example. Note that in this setup, there is a target-proposal mismatch in the location, the scale, and the tails.

B. Example 2: Multimodal bivariate example

Let us consider that the pdf we want to integrate from is a mixture of five bivariate Gaussians,

$$\tilde{\pi}(\mathbf{x}) = \frac{1}{5}\sum_{i=1}^{5}\mathcal{N}(\mathbf{x};\boldsymbol{\nu}_i,\mathbf{C}_i), \quad \mathbf{x}\in\mathbb{R}^2, \quad (22)$$

Algorithm 2 Deterministic mixture IGH (DM-IGH)

Input: Number of points $N = \alpha^{d_x}$, and $\{\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m\}_{m=1}^{M}$

1: Select $\mathcal{S}_m = \{\mathbf{x}_{m,n}, v_{m,n}\}_{n=1}^{N}$ for each $q_m(\mathbf{x}) = \mathcal{N}(\mathbf{x};\boldsymbol{\mu}_m,\boldsymbol{\Sigma}_m)$ according to Gauss-Hermite rules.

2: Compute the importance weights

$$w_{m,n} = \frac{\pi(\mathbf{x}_{m,n})}{\frac{1}{M}\sum_{j=1}^{M} q_j(\mathbf{x}_{m,n})}, \quad \begin{array}{l} n = 1,\cdots,N, \\ m = 1,\cdots,M \end{array} \quad (18)$$

3: Compute the quadrature important weights

$$w'_{m,n} = w_{m,n}v_{m,n}N, \quad \begin{array}{l} n = 1,\cdots,N, \\ m = 1,\cdots,M \end{array} \quad (19)$$

4: The unnormalized estimator is built as

$$\hat{I}_{\text{DM-IGH}} = \frac{1}{ZMN}\sum_{m=1}^{M}\sum_{n=1}^{N} w'_{m,n}f(\mathbf{x}_{m,n}), \quad (20)$$

when Z is known, and the self-normalize estimator as

$$\tilde{I}_{\text{DM-IGH}} = \sum_{m=1}^{M}\sum_{n=1}^{N} \bar{w}'_{m,n}f(\mathbf{x}_{m,n}), \quad (21)$$

with $\bar{w}'_{m,n} = \frac{w'_{m,n}}{\sum_{j=1}^{M}\sum_{i=1}^{N} w'_{j,i}}$.

Output: $\{\mathbf{x}_{m,n}, w'_{m,n}\}_{n=1,m=1}^{N,M}$.

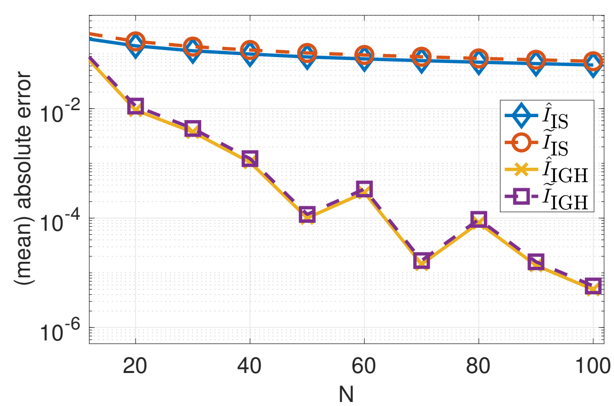

Fig. 1. **Ex. 1.** Mean absolute error in the approximation of the target distribution with IS and IGH estimators.

with parameters $\boldsymbol{\nu}_1 = [-10,-10]^\top$, $\boldsymbol{\nu}_2 = [0,16]^\top$, $\boldsymbol{\nu}_3 = [13,8]^\top$, $\boldsymbol{\nu}_4 = [-9,7]^\top$, $\boldsymbol{\nu}_5 = [14,-14]^\top$, $\mathbf{C}_1 = [2,\ 0.6; 0.6,\ 1]$, $\mathbf{C}_2 = [2,\ -0.4; -0.4,\ 2]$, $\mathbf{C}_3 = [2,\ 0.8; 0.8,\ 2]$, $\mathbf{C}_4 = [3,\ 0; 0,\ 0.5]$, and $\mathbf{C}_5 = [2,\ -0.1; -0.1,\ 2]$. We first use the M-PMC algorithm [24], which implements an integrated stochastic EM algorithm for approximating the target, imposing a set of isotropic Gaussians $\{q_m(\mathbf{x}|\boldsymbol{\mu}_m, \sigma_m^2\mathbb{I}_2)\}_{m=1}^{M}$, where $\mathbb{I}_2$ is the bi-dimensional identity matrix. We run this M-PMC with 500 samples per iteration, $T = 5$ iterations, and $M \in \{1, 10\}$. Then, we apply standard IS and IGH with the proposal resulting from a M-PMC run (with $M = 1$), and the methods SM-MIS, DM-MIS, DM-IGH, and SM-IGH by using the proposals produced with

Algorithm	IS	SM-MIS	DM-MIS	IGH	SM-IGH	DM-IGH
$\hat{I}$ (M-PMC)	2.26	1.45	0.42	0.26	$\mathbf{1.94 \cdot 10^{-7}}$	$8.28 \cdot 10^{-7}$
$\tilde{I}$ (M-PMC)	3.71	0.21	0.18	1.74	$\mathbf{2.36 \cdot 10^{-7}}$	$9.34 \cdot 10^{-7}$
$\hat{I}$ (inflated)	5.86	2.07	0.05	6.46	0.2002	**0.0004**
$\tilde{I}$ (inflated)	6.12	0.07	0.06	11.74	0.1837	**0.0003**

TABLE I

Ex. 2. MSE/SE IN THE APPROXIMATION OF THE TARGET MEAN FOR THREE SAMPLING (LEFT) AND THREE QUADRATURE (RIGHT) METHODS. IN THIS FIRST TWO ROWS, ALL METHODS USE AS PROPOSALS THE RESULT OF AN STOCHASTIC EM ALGORITHM (M-PMC), WHILE IN THE LAST TWO ROWS, WE INFLATE THE STANDARD DEVIATION OF THE PROPOSALS BY A FACTOR OF 1.25.

a M-PMC run (with $M = 10$). All the six methods are run with a total number of 500 samples. Table I shows the mean squared error (MSE), averaged over both dimensions, in the unnormalized $\hat{I}$ and self-normalized $\tilde{I}$ estimators of the target mean. In the first two rows, the proposals are those obtained from the M-PMC run (as described above), while in the last two rows, the standard deviation of the proposals, σ_m, are inflated by a factor of 1.25 to increase the mismatch between target and proposal. The best performance is obtained by both SM-IGH and DM-IGH when we use the M proposals coming from M-PMC (that are the same we use for SM-MIS and DM-MIS). We remark that those proposals do not perfectly reconstruct the target (we use $M = 10$ proposals instead of the 5 modes of the target, and we impose isotropic structure). We see that inflating the proposals deteriorates the quadrature methods although they still show the best performance. IGH shows an intermediate performance in this multimodal example due to the use of a single proposal. It is interesting to see that the self-normalize estimator $\tilde{I}$ obtains always similar results, improving even the unnormalized estimator $\hat{I}$ in some cases, without needing to know the normalizing constant.

V. Conclusions

In this paper, we have presented a novel framework for applying Gauss-Hermite quadrature to non-Gaussians integration problems. In its basic form, the new IGH method reparametrizes the integral by introducing a Gaussian distribution inspired in the IS methodology. This new perspective allows for a new estimator, the self-normalized IGH estimator, and extensions to the case where more than one proposals are introduced. In particular, we have proposed and numerically evaluated the SM-IGH and the DM-IGH methods that allows for approximating problems with intricate distributions in a more flexible manner. The achieved performance in the reported example is promising, providing orders of magnitude improvement over state-of-the-art IS methods.

References

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

[2] J. S. Liu, *Monte Carlo strategies in scientific computing*. Springer Science & Business Media, 2008.

[3] K. Ito and K. Xiong, "Gaussian filters for nonlinear filtering problems," *IEEE Trans. on Automatic Control*, vol. 45, no. 5, pp. 910–927, May 2000.

[4] P. Stano *et al.*, "Parametric Bayesian filters for nonlinear stochastic dynamical systems: a survey," *IEEE Trans. on Cybernetics*, vol. 43, no. 6, pp. 1607 – 1624, Dec. 2013.

[5] I. Arasaratnam, S. Haykin, and R. J. Elliot, "Discrete-time nonlinear filtering algorithms using Gauss-Hermite quadrature," *Proc. of the IEEE*, vol. 95, no. 5, pp. 953–977, 2007.

[6] I. Arasaratnam and S. Haykin, "Square-root quadrature Kalman filtering," *IEEE Trans. Signal Processing*, vol. 56, no. 6, pp. 2589–2593, June 2008.

[7] P. Closas, C. Fernández-Prades, and J. Vilà-Valls, "Multiple Quadrature Kalman filtering," *IEEE Trans. Signal Processing*, vol. 60, no. 12, pp. 6125–6137, Dec. 2012.

[8] P. Closas, J. Vilà-Valls, and C. Fernández-Prades, "Computational complexity reduction techniques for quadrature Kalman filters," in *Proc. of the CAMSAP'15*, Cancun, Mexico, Dec. 2015.

[9] J. Vilà-Valls, P. Closas, and A. García-Fernández, "Uncertainty exchange through multiple quadrature Kalman filtering," *IEEE signal processing letters*, vol. 23, no. 12, pp. 1825–1829, 2016.

[10] N. Steen, G. Byrne, and E. Gelbard, "Gaussian Quadratures for the Integrals $\int_0^\infty \exp(-x^2)f(x)dx$ and $\int_0^b \exp(-x^2)f(x)dx$," *Mathematics of Computation*, pp. 661–671, 1969.

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

[12] D. Ballreich, *Deterministic Numerical Integration*. Cham: Springer International Publishing, 2017, pp. 47–91.

[13] P. Jäckel, "A note on multivariate Gauss-Hermite quadrature," *London: ABN-Amro. Re*, 2005.

[14] B. Jia, M. Xin, and Y. Cheng, "Sparse-grid quadrature nonlinear filtering," *Automatica*, vol. 48, no. 2, pp. 327–341, 2012.

[15] W. Gautschi, "A survey of Gauss-Christoffel quadrature formulae," in *EB Christoffel*. Springer, 1981, pp. 72–147.

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

[17] A. Owen, *Monte Carlo theory, methods and examples*. http://statweb.stanford.edu/∼owen/mc/, 2013.

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

[19] W. Feller, *An Introduction to Probability and Its Applications*, ser. Wiley publication in mathematical statistics. Wiley India Pvt. Limited, 1966, vol. II.

[20] H. W. Sorenson and D. L. Alspach, "Recursive Bayesian estimation using Gaussian sums," *Automatica*, vol. 7, pp. 465–479, 1971.

[21] O. Cappé, A. Guillin, J. M. Marin, and C. P. Robert, "Population Monte Carlo," *Journal of Computational and Graphical Statistics*, vol. 13, no. 4, pp. 907–929, 2004.

[22] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Improving Population Monte Carlo: Alternative weighting and resampling schemes," *Signal Processing*, vol. 131, no. 12, pp. 77–91, 2017.

[23] G. J. McLachlan, S. X. Lee, and S. I. Rathnayake, "Finite mixture models," *Annual Review of Statistics and Its Application*, no. 0, 2000.

[24] O. Cappé, R. Douc, A. Guillin, J. M. Marin, and C. P. Robert, "Adaptive importance sampling in general mixture classes," *Statistics and Computing*, vol. 18, pp. 447–459, 2008.