Hamiltonian Adaptive Importance Sampling

Hamiltonian Adaptive Importance Sampling

Ali Mousavi, Reza Monsefi, and Víctor Elvira, *Senior Member, IEEE*,

Abstract—Importance sampling (IS) is a powerful Monte Carlo (MC) methodology for approximating integrals, for instance in the context of Bayesian inference. In IS, the samples are simulated from the so-called proposal distribution, and the choice of this proposal is key for achieving a high performance. In adaptive IS (AIS) methods, a set of proposals is iteratively improved. AIS is a relevant and timely methodology although many limitations remain yet to be overcome, e.g., the curse of dimensionality in high-dimensional and multi-modal problems. Moreover, the Hamiltonian Monte Carlo (HMC) algorithm has become increasingly popular in machine learning and statistics. HMC has several appealing features such as its exploratory behavior, especially in high-dimensional targets, when other methods suffer. In this paper, we introduce the novel Hamiltonian adaptive importance sampling (HAIS) method. HAIS implements a two-step adaptive process with parallel HMC chains that cooperate at each iteration. The proposed HAIS efficiently adapts a population of proposals, extracting the advantages of HMC. HAIS can be understood as a particular instance of the generic layered AIS family with an additional resampling step. HAIS achieves a significant performance improvement in high-dimensional problems w.r.t. state-of-the-art algorithms. We discuss the statistical properties of HAIS and show its high performance in two challenging examples.

Index Terms—Adaptive importance sampling, Hamiltonian Monte Carlo

I. INTRODUCTION

IN statistical signal processing, many tasks require the computation of expectations with respect to a probability density function (pdf). Unfortunately, obtaining closed-form solutions to these expectations is infeasible in many real-world challenging problems. There are various approximation techniques to solve this problem, the most popular of which is the Monte Carlo (MC) methodology, based on the generation of random samples [1]. Arguably, the two main subfamilies of MC methods are importance sampling (IS) [2] and Markov chain Monte Carlo (MCMC) [1, Chapter 6]. In this paper, we focus on the IS methods where the samples are obtained by simulating from the so-called proposal distribution. The key of IS is an appropriate choice of the proposal distribution, which is a hard and very relevant problem. Since choosing a good proposal in advance is in general unfeasible, adaptive IS (AIS) methods adapt a mixture of proposals, iteratively improving the quality of the estimators by better fitting the proposals [3]. There are several families of AIS methods, such as the population Monte Carlo (PMC) [4]–[6], the AMIS algorithm [7], [8], or gradient-based techniques [9]–[11]. The research in AIS continues being very active and

A. Mousavi and R. Monsefi are with Computer Department, Engineering Faculty, Ferdowsi University of Mashhad (FUM), Mashhad, Iran (e-mail: mousavi@iau-neyshabur.ac.ir; monsefi@um.ac.ir).

V. Elvira is with the School of Mathematics, University of Edinburgh, United Kingdom (e-mail: victor.elvira@ed.ac.uk).

many crucial challenges remain open (see [12] for a recent survey). For instance, high-dimensional and multi-modal targets are particularly challenging to be explored and most AIS (and adaptive MCMC) methods fail to efficiently discover regions with relevant probability mass. Many efforts have been devoted to adapt the AIS proposals through an optimization process [5], [13]–[16]. Other works have addressed directly the reduction of the variability of the importance weight, ultimate responsible of the poor performance of the IS estimators [17]–[22]. Some recent methods aim at adapting the proposals through an independent process from the generated samples. We refer the interested reader to this class of AIS methods in [12, Fig. 4(c)]. Particular instances of this class are the algorithms in the LAIS framework [23] or the techniques in [9], [24]. Unlike our proposed algorithm, the three explicit methods presented in the LAIS framework, which implement an adaptation of the upper layer based on the Metropolis-Hastings (MH) algorithm (see [23] for more details). One limitation of these algorithms is the well-known random-walk behavior of the MH that makes the convergence of the Markov chain inefficient, especially in high-dimensional multi-modal distributions. In addition, the performance of LAIS algorithms is highly dependent to the scale parameter of the proposals in the upper layer. Hamiltonian (or hybrid) Monte Carlo (HMC) [25], [26] is a state-of-the-art family of MCMC algorithms. HMC implements Hamiltonian dynamics allowing the samples to reach more distant points with higher probability of acceptance, which provides a greater improvement on exploratory capabilities compared with the other state-of-the-arts algorithms. Despite the appealing properties of HMC, the method can be notoriously difficult to tune (see [27]–[29]). We note that HMC have been also incorporated to the SMC samplers framework [30].

In this paper, we present the novel *Hamiltonian adaptive importance sampling* (HAIS) algorithm which retains advantageous features from both HMC and AIS, achieving a high performance. HAIS proposes a novel two-layered HMC-based AIS, inheriting the structure of LAIS or GAPIS. The upper layer consists of a two-step adaptation procedure that runs cooperative parallel HMC blocks in order to adapt the multiple proposals in AIS, which benefits both the exploratory behavior (particularly useful in multi-modal and high-dimensional problems) and the parallelization of the implementation. Another important contribution of the work is the consistent cooperation step where information among the chains is exchanged in order to enhance the global exploration. In addition, HAIS requires little tuning, overcoming one of the well-known limitations of HMC samplers. Therefore, the validity of the estimators is ensured by IS arguments, unlike in HMC where the tuning can endanger the convergence of the Markov chain.

The rest of the paper is organized as follows. In Section II,

we describe the problem while Section III describes the novel HAIS method. Numerical examples are provided to compare the proposed HAIS with some other techniques in Section IV.

II. PROBLEM STATEMENT

A. Bayesian inference

Let us consider a random variable of interest $\mathbf{x} \in \mathbb{R}^{d_x}$ and $\mathbf{y} \in \mathbb{R}^{d_y}$ be a set of related measurements or observations. In the Bayesian framework, the variable of interest is characterized through the posterior probability function or the target pdf known as the

$$\widetilde{\pi}(\mathbf{x}|\mathbf{y}) = \frac{\ell(\mathbf{y}|\mathbf{x})p_0(\mathbf{x})}{Z(\mathbf{y})}, \tag{1}$$

where $\ell(\mathbf{y}|\mathbf{x})$ is the likelihood function, $p_0(\mathbf{x})$, is the prior pdf, and $Z(\mathbf{y})$ is the model evidence (from now on we drop $\mathbf{y}$ in the notation). In many applications the goal is to obtain a moment of $\mathbf{x}$ which can be expressed as the integral

$$I = \mathrm{E}_{\widetilde{\pi}}[f(\mathbf{x})] = \int_{\mathbb{R}^{d_x}} f(\mathbf{x})\widetilde{\pi}(\mathbf{x})d\mathbf{x}, \tag{2}$$

where $f : \mathbb{R} \to \mathbb{R}^{d_x}$ is some integrable function.

B. Importance sampling

Importance sampling (IS) is one of the main subfamilies of Monte Carlo methods. The basic idea of IS is to sample from a simpler pdf, the so-called proposal pdf $q(\mathbf{x})$, to approximate the integrals w.r.t. to the target distribution $\widetilde{\pi}(\mathbf{x})$ as

$$\hat{I}_{\text{IS}}(\mathbf{x}) = \frac{1}{MZ} \sum_{m=1}^{M} w_m f(\mathbf{x}_m), \tag{3}$$

where $\{\mathbf{x}_m\}_{m=1}^{M}$ are iid samples generated from $q(\mathbf{x})$, Z is the normalization constant, and $w_m = \frac{\pi(\mathbf{x}_m)}{q(\mathbf{x}_m)}$ is the importance weight associated to $\mathbf{x}_n$. Then, we can construct $\hat{I}_{\text{IS}}$ which is an unbiased and consistent estimator of I, and its variance is related to discrepancy between $|f(\mathbf{x})|\pi(\mathbf{x})$ and $q(\mathbf{x})$. When Z is unknown, the so-called *self-normalized importance sampling* (SNIS) estimator can be constructed by plugging in Eq. (3) the unbiased estimator $\hat{Z} = \frac{1}{M}\sum_{m=1}^{M} w_m$ instead of Z (see more details in [31]). Since finding a good $q(\mathbf{x})$ in advance is generally impossible, adaptive importance sampling (AIS) approaches are usually implemented, in order to iteratively improve the proposal. Relevant recent AIS methods are the PMC [4] and AMIS[7], and more recently the LAIS [23], DM-PMC [6], or [32] (see [12] for a review).

C. Hamiltonian Monte Carlo

Hamiltonian Monte Carlo (HMC) is a MCMC-based method that adopts Hamiltonian dynamics to explore the state space in order to propose future states in the Markov chain. More precisely, let us denote $U(\mathbf{x}) = -\log \pi(\mathbf{x})$, which is usually called potential energy function in the physics literature [25], [26]. We consider also the kinetic energy function, $V(\mathbf{p}) = \frac{1}{2}\mathbf{p}^T \boldsymbol{R}^{-1}\mathbf{p}$, with $\boldsymbol{R}$ as a positive definite mass matrix and $\mathbf{p} \in \mathbb{R}^{d_x}$ as momentum vector. The matrix $\boldsymbol{R}$ is typically

diagonal or isotropic. Algorithm 1 describes the basic HMC procedure which explores the joint probability density of $\mathbf{x}$ and $\mathbf{p}$, allowing for the simulation of samples. Starting at an initial state $[\mathbf{x}_0, \mathbf{p}_0]$, HMC simulates Hamiltonian dynamics for L steps using a discretization method. The common method is the leapfrog with the small step size parameter ϵ. Next, the state of the position and momentum variables at the end of the simulation is used as the proposed state variables. Finally, $(\mathbf{x}^*, \mathbf{p}^*)$ is accepted using an update rule analogous to the Metropolis acceptance criterion [26]. By means of controlling the leapfrog size (L) and ϵ, the acceptance rate of the HMC sampler can be adjusted [33]. The Hamiltonian dynamics benefit from several properties. Despite the significant benefits of HMC, especially in high dimension, HMC is known to be highly sensitive to the choice of parameters, particularly ϵ and L. Choosing a too large step size will result in a low acceptance rate for new proposed state. On the other hand, a too small step size will lead to slow exploration. Also, HMC encounters difficulties to sample from multi-modal distributions. We refer the interested reader to [26], [34].

Algorithm 1: Hamiltonian Monte Carlo

Input: step size ϵ, leapfrog size L, starting point $\mathbf{x}^1$, sample number M
for $m = 1, ..., M$ **do**
  Sample $\mathbf{p}_0 \sim \mathcal{N}(0, 1)$ and set $\mathbf{x}_0 = \mathbf{x}^m$
  Run leapfrog method starting at $(\mathbf{x}_0, \mathbf{p}_0)$ for L steps with step size ϵ to obtain proposed states $(\mathbf{x}^*, \mathbf{p}^*)$
  Generate u from $U[0, 1]$
  if $u \leq min[1, e^{U(\mathbf{x}^m)+V(\mathbf{p}^m)-U(\mathbf{x}^*)-V(\mathbf{p}^*)}]$
    $\mathbf{x}^{m+1} = \mathbf{x}^*$
  else
    $\mathbf{x}^{m+1} = \mathbf{x}^m$
end
Output: sample set $\{\mathbf{x}^m\}_{m=1}^{M}$

III. HAMILTONIAN ADAPTIVE IMPORTANCE SAMPLING

In this section, we present the novel *Hamiltonian adaptive importance sampling* (HAIS) method, which is summarized in Fig. 1 and precisely described in Alg. 2. HAIS is an iterative method that, at each iteration, performs three main operations: (a) sampling; (b) weighting; and (c) adaptation.

A. Algorithm description

Algorithm 2 describes the HAIS method. We consider N (parametric) proposal distributions, $q_n(\mathbf{x}; \boldsymbol{\mu}_n, \boldsymbol{\nu}_n)$ where $\boldsymbol{\mu}_n$ is the location parameter (e.g., the mean in a Gaussian pdf) and $\boldsymbol{\nu}_n$ contains the other parameters (e.g., $\boldsymbol{\nu}_n = \boldsymbol{\Sigma}_n$ is the covariance matrix in a Gaussian distribution). The algorithm starts with the initialization of the proposal parameters. At each iteration, KN samples are generated from the N proposals (exactly K samples per proposal). Then, the importance weights are computed following deterministic mixture (DM) scheme. It has been recently shown that the unnormalized IS estimator of Eq. (3) with DM weights outperforms the same estimator if only the proposal q_n is in the denominator (instead of the whole mixture) [31]. Finally, adaptation procedure is performed. Next section is devoted to describe this adaptation.
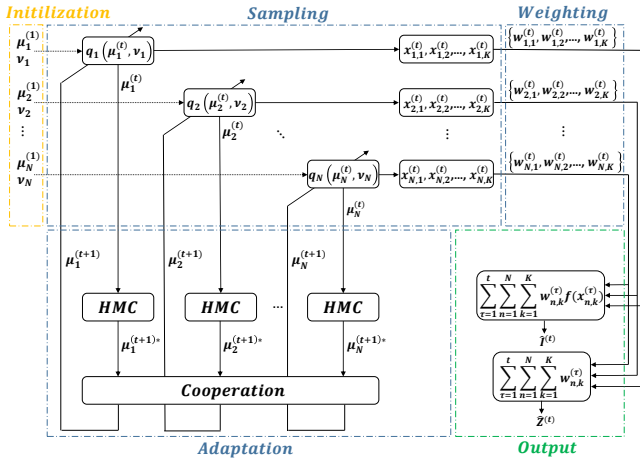
Fig. 1: The flowchart of the proposed HAIS method showing the three main operations at iteration t which results in the estimated values.

B. Adaptation process

Fig. 1 shows a two-step procedure in the adaptation of location parameters of the proposals (parallel HMC and cooperation). First, each HMC block explores the state space independently which is in a more efficient manner to discover local relevant features of the target, compared to other mechanisms such as MH-based methods or even naive gradient-based methods as in [9]. The parallel structure amplifies the exploratory behavior, improving the local exploration capability of HAIS in high-dimensional multi-modal targets especially when the modes are distant. Second, the information among the output of N parallel HMC is exchanged in order to improve the global exploration (see a discussion on local-global exploration in [6]). It is also important to remark that the cooperation step presents a theoretically sound and consistent procedure which does not require any free parameter to be tuned. We note that the adaptation of the proposal locations is completely independent from the samples for target estimation which helps the parallelization of the method (see a classification of AIS algorithm according to the adaptive mechanism in [12]). Note that, unlike HMC, the performance of HAIS does not critically depend on a precise tuning of the HMC parameters, since we use HMC for adapting the proposals not the final samples (that are properly weighted via IS). Assuming that all the HMC blocks obtain chains converging to the target pdf, the convergence of the overall HAIS is guaranteed. In the following we present a description of the novel approach.

1) Parallel HMC step: We propose to run N independent HMC method in a parallel way, each of which is shown as a HMC block in Fig. 1. Let us consider N parallel chains $\{\boldsymbol{\mu}_n\}_{n=1}^{N}$ generated by those HMC blocks with $\boldsymbol{\mu}_n^{(1)}$ as the initial d-dimensional starting point for the n-th chain. We apply one iteration of N parallel chains, one for each $\boldsymbol{\mu}_n^{(t)}$, returning $\boldsymbol{\mu}_n^{(t+1)*}$ for $n = 1, ..., N$. Next, we compute the normalized DM weight of each $\boldsymbol{\mu}_n^{(t+1)*}$,

$$\bar{w}_{\boldsymbol{\mu}_n^{(t+1)*}} \propto \frac{\pi(\boldsymbol{\mu}_n^{(t+1)*})}{\sum_{i=1}^{N} q_i(\boldsymbol{\mu}_i^{(t+1)*}; \boldsymbol{\mu}_i^{(t)}, \boldsymbol{\nu_i})}. \tag{4}$$

Now we consider $\widetilde{\pi}^{(t+1)*}(\boldsymbol{\mu})$ as a random measure that approximates the target distribution, i.e.,

$$\widetilde{\pi}^{(t+1)*}(\boldsymbol{\mu}) = \sum_{n=1}^{N} \bar{w}_{\boldsymbol{\mu}_n^{(t+1)*}} \delta(\boldsymbol{\mu} - \boldsymbol{\mu}_n^{(t+1)*}), \tag{5}$$

The theoretical motivation is that, after the burn-in periods, the N parallel HMC chains have converged to the target, so $\boldsymbol{\mu}_n^{(t+1)*} \sim \pi$. As a result, the random measure based on weighted mean vectors in Eq. (5) approximates the target distribution.

2) Cooperation step: The final mean vector of the proposals for the next iteration are obtained by sampling from $\widetilde{\pi}^{(t+1)*}(\boldsymbol{\mu})$, (via resampling), i.e., $\boldsymbol{\mu}_n^{(t+1)} \sim \widetilde{\pi}^{(t+1)*}(\boldsymbol{\mu})$. After resampling, a modified and unweighted random measure is produced as

$$\widetilde{\pi}^{(t+1)}(\boldsymbol{\mu}) = \sum_{n=1}^{N} \delta(\boldsymbol{\mu} - \boldsymbol{\mu}_n^{(t+1)}). \tag{6}$$

In the following, we provide theoretical justification for using the cooperation step. The convergence of this adaptation process is given by Theorem 1.

Theorem 1 *The moments of the random measure $\widetilde{\pi}^{(t+1)}(\boldsymbol{\mu})$ obtained from outputs of the cooperation step (i.e., the means $\boldsymbol{\mu}_n^{(t+1)}$) converge almost surely to those of the true distribution $\widetilde{\pi}(\mathbf{x})$ when $N \to \infty$.*

Proof. See the appendix.

Algorithm 2: Hamiltonian adaptive importance sampling

Input: initial location, $\{\boldsymbol{\mu}_n^{(1)}\}_{n=1}^{N}$, and scale, $\{\boldsymbol{\nu}_n\}_{n=1}^{N}$, of proposals, N proposals, K sample per proposal, T iterations
for $t = 1, ..., T$ **do**
    (a) **sampling**: Draw K samples per individual proposal or mixand,

$$\mathbf{x}_{n,k}^{(t)} \sim q_n(\mathbf{x}; \boldsymbol{\mu}_n^{(t)}, \boldsymbol{\nu}_n), n = 1, ..., N, k = 1, ..., K.$$

    (b) **weighting**: Compute the deterministic mixture (DM) weight of each sample,

$$w_{n,k}^{(t)} = \frac{\pi(\mathbf{x}_{n,k}^{(t)})}{\sum_{i=1}^{N} q_i(\mathbf{x}_{i,k}^{(t)}; \boldsymbol{\mu}_i^{(t)}, \boldsymbol{\nu}_i)}.$$

    (c) **adaptation**: Update location of proposals for the next iteration through the two-step adaptation procedure.
    1. parallel HMC step: Generate new population of proposals, $\{\boldsymbol{\mu}_1^{(t+1)*}, \boldsymbol{\mu}_2^{(t+1)*}, ..., \boldsymbol{\mu}_N^{(t+1)*}\}$, by applying one iteration of N parallel HMC blocks.
    2. cooperation step: Generate final population of proposals, $\{\boldsymbol{\mu}_1^{(t+1)}, \boldsymbol{\mu}_2^{(t+1)}, ..., \boldsymbol{\mu}_N^{(t+1)}\}$, for the next iteration by sampling from equation (5).
end
Output: a set of KNT samples with associated weights, $\{\mathbf{x}_{n,k}^{(t)}, w_{n,k}^{(t)}\}$ to obtain the estimation.

IV. NUMERICAL EXAMPLES

A. Bimodal target distribution in $d_x = 20$

In this section we consider a high-dimensional bi-modal target pdf in order to compare HAIS with some alternative methods. The target is a mixture of two Gaussians, i.e., $\pi(\mathbf{x}) = \sum_{i=1}^{2} \frac{1}{2} \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$. Here $\mathbf{x} \in \mathbb{R}^{20}$, $\boldsymbol{\mu}_i = [\mu_{i,1}, \mu_{i,2}, ..., \mu_{i,20}]^T$, and $\boldsymbol{\Sigma}_i = c\mathbf{I}_{20}$, for $i \in \{1, 2\}$ where $\mathbf{I}_{20}$ is an identity matrix of dimension 20. We set $\mu_{1,d} = 8$ and $\mu_{2,d} = -8$ for all dimensions and also we set $c = 5$.

TABLE I: MSE in the approximation of the mean and normalizing constant of multi-modal target distribution.

| Method	$\sigma = 1$		$\sigma = 2$		$\sigma = 5$	
	$E[\mathbf{x}]$	Z	$E[\mathbf{x}]$	Z	$E[\mathbf{x}]$	Z
GR-PMC	64.92	0.9418	64.14	0.3126	65.70	0.3412
LR-PMC	**21.57**	0.9974	26.73	0.0998	65.24	0.9997
PI-MAIS ($\lambda = 5$)	46.10	1	38.99	0.4531	36.63	0.5421
PI-MAIS ($\lambda = 10$)	60.51	1	60.30	1	53.28	0.9878
HAIS ($\epsilon = 5$)	41.09	**0.8649**	17.57	0.0201	13.20	0.0031
HAIS ($\epsilon = 10$)	42.76	0.8828	**17.32**	**0.0162**	**12.87**	**0.0016**

Multimodal settings are challenging, and in this example the two modes are distant, which over-complicates the exploration process at high dimension. Moreover, the proposal densities are Gaussian pdfs with uniformly selected initial means, i.e., $\boldsymbol{\mu}_n \sim U([-4 \times 4]^{d_x})$ for $n = 1, ..., N$, where none of the modes of the target fall within this area. We use the same isotropic covariance for all of proposals, $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_{20}$ with $\sigma \in \{1, 2, 5\}$. We test for a constant number of samples, $K = 5$, and number of proposals, $N = 100$. For each algorithm the number of iterations, T, is set so they have the same total number of target evaluations, $E = 2 \times 10^5$. We focus in estimation the mean and the normalizing constant of the target, which are obviously known in this example ($E[\mathbf{x}] = 0$ and $Z = 1$). We compared the proposed HAIS with recent high-performance AIS methods: GR-PMC and LR-PMC [6], and LAIS [23]. For the upper layer of the LAIS, we also consider Gaussian pdfs $\varphi_n(\mathbf{x}|\boldsymbol{\mu}_n, \Lambda_n)$ where covariance matrices $\Lambda_n = \lambda^2 I_2$ with $\lambda \in \{5, 10\}$. In our HAIS we consider all the HMC blocks to have the same step size parameter choosing from $\epsilon \in \{5, 10\}$ and a fixed value of $L = 50$. The results are averaged over 200 independent runs. The simulation results are summarized in Table I in terms of mean squared error (MSE) in the estimators. First, note that many settings/algorithms obtain very large MSE values. Those situations often correspond to the case where one or both modes are failed to be discovered. In the situation of missing both modes, the estimation of normalizing constant of the target is $\hat{Z} \approx 0$ which corresponds to a MSE≈ 1 as it happens in several settings. Second, we see that the HAIS algorithm generally outperforms the other methods. We note that the optimum value of σ which yields the smallest MSE, depends on the scale parameter of the target (here it is $c = 5$).

B. High-dimensional banana-shaped target distribution

We now consider a benchmark multidimensional banana-shaped target distribution [35], which is a challenging example because of its nonlinear nature. The target pdf is given by

$$\bar{\pi}(x_1, ..., x_{d_x})$$
$$\propto \exp\left(-\frac{x_1^2}{2\sigma^2} - \frac{(x_2 + b(x_1^2 - \sigma^2))^2}{2\sigma^2} - \sum_{i=3}^{d_x} \frac{x_i^2}{2\sigma^2}\right). \quad (7)$$

We set $b = 3$ and $\sigma = 1$, then the true value for $E_{\bar{\pi}}[\mathbf{x}] = 0$. Here Gaussian densities are considered as proposals with initial locations similar to the previous experiment. We use the isotropic covariance matrix for all of proposal pdfs, $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_{d_x}$ with $\sigma^2 = 1$. The selected values of K, N, and T is the same as in previous experiment. We compute the
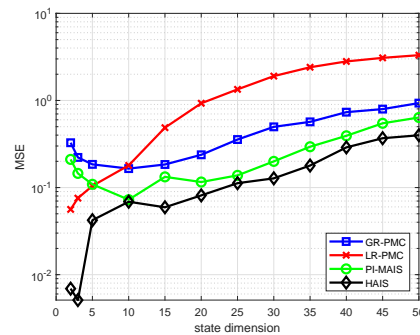
Fig. 2: log-MSE of $E[\mathbf{x}]$, using $N = 100$ proposals with $\sigma = 1$, as the dimension of the target pdf, d_x, increases.

MSE in the estimation of $E[\mathbf{x}]$ and the results are averaged over 200 Monte Carlo simulations. In order to compare the performance of the proposed method with other approaches as the dimension of the state space increases, we vary the dimension of the state space testing different values of d_x (with $2 \leq d_x \leq 50$). Fig. 2 shows the log-MSE in the estimation of $E[\mathbf{x}]$ as a function of the dimension d_x of the state-space, regarding the same techniques as in the previous bi-dimensional example. The performance of all the methods degrades as the dimension becomes larger. The result indicates that the novel HAIS scheme outperforms all the other methods under fair computational complexity comparison.

V. CONCLUSION

In this paper we have proposed the HAIS algorithm, an adaptive importance sampler with particularly good behavior for high-dimensional and multi-modal problems. HAIS belongs to the family of layered AIS samplers, and implements an adaptive procedure by running parallel HMC steps followed by a cooperation step, instead of the MH algorithm deployed in the algorithms of the LAIS framework. The method is theoretically justified by HMC and resampling arguments. Simulation results have shown significant improvement compared with other state-of-the-art methods.

APPENDIX A
PROOF OF THEOREM 1

The proof of Theorem 1 is based on a generalization of the result in [6, Section 4.1]. Let $\mathbf{z} = [z_1, z_2, ..., z_{d_x}]$ be a new auxiliary variable the same size of the variable of interest, $\mathbf{x} = [x_1, x_2, ..., x_{d_x}]$. Now consider the desired square integrable function to be an indicator function, i.e., $f_z(\mathbf{x}) = \prod_{d=1}^{d_x} \mathbb{1}(x_d \leq z_d)$ where $\mathbb{1}$ denotes the indicator function. The integral $I_z = \int f_z(\mathbf{x})\pi(\mathbf{z})d\mathbf{z}$ becomes the multi-variate cumulative distribution function (cdf) of $\pi(\mathbf{x})$. In a similar way, we can obtain the cdf for $\widetilde{\pi}^{(t+1)*}(\boldsymbol{\mu})$, named $\widetilde{I}_z$. It can be shown that $\widetilde{I}_z \to I_z$ a.s. for any value of z as $N \to \infty$ [36]. As a result, as $N \to \infty$, the cdf associated to $\widetilde{\pi}^{(t+1)*}$ a.s. converges to the target cdf. Consequently, the outputs of the cooperation step are asymptotically distributed as the target $\widetilde{\pi}$, i.e., $\boldsymbol{\mu}_n^{(t+1)} \sim \pi(\mathbf{x})$ when $N \to \infty$.

REFERENCES

[1] C. Robert and G. Casella, "Monte carlo statistical methods. 2004."

[2] V. Elvira and L. Martino, "Advances in importance sampling," *Wiley StatsRef: Statistics Reference Online, arXiv:2102.05407*, 2021.

[3] M.-S. Oh and J. O. Berger, "Adaptive importance sampling in monte carlo integration," *Journal of Statistical Computation and Simulation*, vol. 41, no. 3-4, pp. 143–168, 1992.

[4] 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.

[5] 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, no. 4, pp. 447–459, 2008.

[6] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Improving population monte carlo: Alternative weighting and resampling schemes," *Signal Processing*, vol. 131, pp. 77–91, 2017.

[7] J.-M. Cornuet, J.-M. Marin, A. Mira, and C. P. Robert, "Adaptive multiple importance sampling," *Scandinavian Journal of Statistics*, vol. 39, no. 4, pp. 798–812, 2012.

[8] Y. El-Laham, L. Martino, V. Elvira, and M. F. Bugallo, "Efficient adaptive multiple importance sampling," in *2019 27th European Signal Processing Conference (EUSIPCO)*, pp. 1–5, IEEE, 2019.

[9] V. Elvira, L. Martino, D. Luengo, and J. Corander, "A gradient adaptive population importance sampler," in *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pp. 4075–4079, IEEE, 2015.

[10] I. Schuster, "Gradient importance sampling," tech. rep., 2015. https://arxiv.org/abs/1507.05781.

[11] M. Fasiolo, F. E. de Melo, and S. Maskell, "Langevin incremental mixture importance sampling," *Stat. Comput.*, vol. 28, no. 3, pp. 549–561, 2018.

[12] M. F. Bugallo, V. Elvira, L. Martino, D. Luengo, J. Miguez, 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.

[13] E. K. Ryu, *Convex optimization for Monte Carlo: Stochastic optimization for importance sampling*. PhD thesis, Stanford University, 2016.

[14] R. Douc, A. Guillin, J.-M. Marin, C. P. Robert, *et al.*, "Convergence of adaptive mixtures of importance sampling schemes," *The Annals of Statistics*, vol. 35, no. 1, pp. 420–448, 2007.

[15] R. Douc, A. Guillin, J.-M. Marin, and C. P. Robert, "Minimum variance importance sampling via population monte carlo," *ESAIM: Probability and Statistics*, vol. 11, pp. 427–447, 2007.

[16] Y. El-Laham and M. F. Bugallo, "Stochastic gradient population monte carlo," *IEEE Signal Processing Letters*, 2019.

[17] E. L. Ionides, "Truncated importance sampling," *Journal of Computational and Graphical Statistics*, vol. 17, no. 2, pp. 295–311, 2008.

[18] E. Koblents and J. Míguez, "A population monte carlo scheme with transformed weights and its application to stochastic kinetic models," *Statistics and Computing*, vol. 25, no. 2, pp. 407–425, 2015.

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

[20] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Heretical multiple importance sampling," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1474–1478, 2016.

[21] Y. El-Laham, V. Elvira, and M. F. Bugallo, "Robust covariance adaptation in adaptive importance sampling," *IEEE Signal Processing Letters*, vol. 25, no. 7, pp. 1049–1053, 2018.

[22] Y. El-Laham, V. Elvira, and M. Bugallo, "Recursive shrinkage covariance learning in adaptive importance sampling," in *2019 IEEE 8th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pp. 624–628, IEEE, 2019.

[23] L. Martino, V. Elvira, D. Luengo, and J. Corander, "Layered adaptive importance sampling," *Statistics and Computing*, vol. 27, no. 3, pp. 599–623, 2017.

[24] D. Rudolf, B. Sprungk, *et al.*, "On a metropolis–hastings importance sampling estimator," *Electronic Journal of Statistics*, vol. 14, no. 1, pp. 857–889, 2020.

[25] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, "Hybrid monte carlo," *Physics letters B*, vol. 195, no. 2, pp. 216–222, 1987.

[26] R. M. Neal *et al.*, "Mcmc using hamiltonian dynamics," *Handbook of markov chain monte carlo*, vol. 2, no. 11, p. 2, 2011.

[27] S. Mohamed, N. de Freitas, *et al.*, "Adaptive hamiltonian and riemann manifold monte carlo samplers," *arXiv preprint arXiv:1302.6182*, 2013.

[28] A. Beskos, N. Pillai, G. Roberts, J.-M. Sanz-Serna, A. Stuart, *et al.*, "Optimal tuning of the hybrid monte carlo algorithm," *Bernoulli*, vol. 19, no. 5A, pp. 1501–1534, 2013.

[29] O. Mangoubi and A. Smith, "Rapid mixing of hamiltonian monte carlo on strongly log-concave distributions," *arXiv preprint arXiv:1708.07114*, 2017.

[30] A. Buchholz, N. Chopin, and P. E. Jacob, "Adaptive tuning of hamiltonian monte carlo within sequential monte carlo," *Bayesian Analysis*, vol. 1, no. 1, pp. 1–27, 2021.

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

[32] V. Elvira and E. Chouzenoux, "Langevin-based strategy for efficient proposal adaptation in population monte carlo," in *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5077–5081, IEEE, 2019.

[33] B. Leimkuhler and S. Reich, *Simulating hamiltonian dynamics*, vol. 14. Cambridge university press, 2004.

[34] M. Betancourt, "A conceptual introduction to hamiltonian monte carlo," *arXiv preprint arXiv:1701.02434*, 2017.

[35] H. Haario, E. Saksman, J. Tamminen, *et al.*, "An adaptive metropolis algorithm," *Bernoulli*, vol. 7, no. 2, pp. 223–242, 2001.

[36] J. Geweke, "Bayesian inference in econometric models using monte carlo integration," *Econometrica: Journal of the Econometric Society*, pp. 1317–1339, 1989.