Robust Covariance Adaptation in Adaptive Importance Sampling

Yousef El-Laham,* Víctor Elvira,† and Mónica F. Bugallo*
* Dept. of Electrical and Computer Engineering, Stony Brook University, New York (USA)
† IMT Lille Douai & CRIStAL (UMR CNRS 9189), Villeneuve d'Ascq, France.

Abstract

Importance sampling (IS) is a Monte Carlo methodology that allows for approximation of a target distribution using weighted samples generated from another proposal distribution. Adaptive importance sampling (AIS) implements an iterative version of IS which adapts the parameters of the proposal distribution in order to improve estimation of the target. While the adaptation of the location (mean) of the proposals has been largely studied, an important challenge of AIS relates to the difficulty of adapting the scale parameter (covariance matrix). In the case of weight degeneracy, adapting the covariance matrix using the empirical covariance results in a singular matrix, which leads to poor performance in subsequent iterations of the algorithm. In this paper, we propose a novel scheme which exploits recent advances in the IS literature to prevent the so-called weight degeneracy. The method efficiently adapts the covariance matrix of a population of proposal distributions and achieves a significant performance improvement in high-dimensional scenarios. We validate the new method through computer simulations.

Index Terms

Monte Carlo, importance sampling, weight degeneracy, covariance adaptation, nonlinear weight transformation.

I. Introduction

Importance sampling (IS) is a powerful Monte Carlo tool that allows to learn properties of a distribution (called target) by utilizing weighted samples drawn from another distribution (called proposal) [1]. The choice of the proposal distribution in IS algorithms is critical for successful performance. In particular, one issue that can arise in IS when a poorly chosen proposal is used is *weight degeneracy*, which relates to the problem of only a few samples having significant weights that contribute to approximation of the target distribution. High mismatch between the proposal distribution and the target distribution yields a high variance IS estimator. Alternatively, one can use a method known as multiple importance sampling (MIS), in which samples are drawn from multiple proposal distributions [2], [3]. MIS methods have an advantage over static IS methods as they show larger sample diversity and alternative weighting schemes can be applied to reduce the variance of the IS estimator [4], [5], [6]. Adaptive importance sampling (AIS) methods were introduced in order to adapt the proposal distribution through an iterative learning algorithm [7]. While AIS methods have shown success in many practical applications [8], [9], major challenges of the method become apparent when considering multimodal and highly nonlinear target distributions. Moreover, AIS also suffers from the curse of dimensionality [10], whereby the issue of weight degeneracy intensifies and leads to reduced sample diversity in subsequent iterations of the algorithm.

Some advances have been made to AIS schemes to increase the diversity of the generated samples and limit the effects of weight degeneracy. The use of deterministic mixture weights [3] in the population Monte Carlo scheme (DM-PMC) [11] significantly reduces the variance of the IS estimator and maintains the diversity of the proposals through each iteration. Unfortunately, the DM-PMC algorithm does not adapt the covariance matrices of the proposal distributions. Alternatively, the nonlinear population Monte Carlo (N-PMC) scheme applies a nonlinear transformation of the importance weights, similar to the approaches of [12], [13], [14]. The weight transformation guarantees representation of more samples for proposal adaptation in the case of weight degeneracy. However, in high dimensional spaces, the effect of the weight transformation may lead to minuscule adaptation of the proposal. Other AIS methods such as mixture population Monte Carlo (M-PMC) [15], adaptive multiple importance sampling (AMIS) [16], and adaptive population importance sampling (APIS) [17], have also been proposed with some variations. Some of these AIS methods also adapt the covariance matrix, but the stability of the update is not guaranteed. In this paper, we propose a novel and robust method for adapting a set of proposal distributions, called *covariance adaptive importance sampling* (CAIS), which tackles the inherent problems posed by high-dimensional targets. Specifically, we condition the covariance adaptation of each proposal on a local measurement of the effective sample size in order to achieve robustness.

The rest of the paper is organized as follows. In Section II, we describe the problem. Section III describes the proposed novel approach for adapting the proposal distributions. We show simulation results in Section IV and conclude the paper in Section V.

M. F. B. thanks the support of the National Science Foundation under Award CCF-1617986. V. E. acknowledges support from the ANR under PISCES project (ANR-17-CE40-0031-01), and from the French-American Fulbright Commission under the Fulbright scholarship.

TABLE I: A basic framework for AIS methods.

1. Initialization: Select the initial proposal $q(\mathbf{x}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$
2. For $i = 1, ..., I$
a. Draw M samples from the proposal,

$$\mathbf{x}_i^{(m)} \sim q(\mathbf{x}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \qquad m = 1, ..., M.$$

b. Compute the standard importance weights,

$$w_i^{(m)} = \frac{\pi(\mathbf{x}_i^{(m)})}{q(\mathbf{x}_i^{(m)}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}, \quad m = 1, ..., M,$$

and normalize them, $\bar{w}_i^{(m)} = \frac{w_i^{(m)}}{\sum_{j=1}^M w_i^{(j)}}$, with $m = 1, ..., M$.
c. Adapt the proposal parameters $\boldsymbol{\mu}_{i+1}$ and $\boldsymbol{\Sigma}_{i+1}$.
3. Return the pairs $\{\mathbf{x}_i^{(m)}, w_i^{(m)}\}$ for $m = 1, ..., M$ and $i = 1, ..., I$.

II. Problem Formulation

We introduce the estimation problem in a Bayesian context, with the goal of computing the posterior distribution $p(\mathbf{x}|\mathbf{y})$, where $\mathbf{x} \in \mathbb{R}^{d_x}$ is a d_x-dimensional vector of unknown parameters related to a set of available observations $\mathbf{y} \in \mathbb{R}^{d_y}$. It follows from Bayes' rule that,

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \equiv \pi(\mathbf{x}), \tag{1}$$

where $p(\mathbf{y}|\mathbf{x})$ is the likelihood function of the data, while $p(\mathbf{x})$ is a prior distribution of the vector of unknown parameters. The goal is in computing the expectation of some function $g(\mathbf{X})$ w.r.t. $p(\mathbf{x}|\mathbf{y})$, such that

$$I = \mathbb{E}(g(\mathbf{X})) = \int_{-\infty}^{\infty} g(\mathbf{x})p(\mathbf{x}|\mathbf{y})d\mathbf{x}. \tag{2}$$

In particular, we may be interested in computing quantities such as the normalizing constant $Z = \int \pi(\mathbf{x})d\mathbf{x}$ or the mean of the posterior distribution. To that end, basic implementations of IS approximate $p(\mathbf{x}|\mathbf{y})$ using samples drawn from a proposal distribution $q(\mathbf{x})$ that are weighted properly.

III. Novel Covariance Adaptation Strategy

We assume a family of proposal distributions, $q(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$, where the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean and covariance matrix, respectively. Table I summarizes a basic AIS framework, where the proposal is adapted over I iterations.

A. General Formulation

For a multivariate normal distribution, the maximum likelihood (ML) estimate of the covariance matrix for M unweighted samples with mean $\hat{\boldsymbol{\mu}}_{ML}$ is given by,

$$\hat{\boldsymbol{\Sigma}}_{ML} = \frac{1}{M}\sum_{m=1}^{M}(\mathbf{x}^{(m)} - \hat{\boldsymbol{\mu}}_{ML})(\mathbf{x}^{(m)} - \hat{\boldsymbol{\mu}}_{ML})^{\mathsf{T}}. \tag{3}$$

It is well-known that if $\mathbf{x} \in \mathbb{R}^{d_x}$ and $M < d_x$, then rank($\hat{\boldsymbol{\Sigma}}_{ML}$) $< d_x$ and $\hat{\boldsymbol{\Sigma}}_{ML}$ cannot be inverted [18]. Analogous to the case of unweighted samples, suppose that we are given a set of M weighted samples $\{\mathbf{x}^{(m)}, \bar{w}^{(m)}\}_{m=1}^M$. If the weighted mean of the samples is $\hat{\boldsymbol{\mu}}_W$, then the weighted empirical covariance is given by,

$$\hat{\boldsymbol{\Sigma}}_W = \sum_{m=1}^{M}\bar{w}^{(m)}(\mathbf{x}^{(m)} - \hat{\boldsymbol{\mu}}_W)(\mathbf{x}^{(m)} - \hat{\boldsymbol{\mu}}_W)^{\mathsf{T}}. \tag{4}$$

Consider the set of indices $K = \{k \mid \bar{w}^{(k)} < \epsilon\}$. As $\epsilon \to 0$, we have that $\hat{\boldsymbol{\Sigma}}_W \approx \sum_{k \notin K}\bar{w}^{(k)}(\mathbf{x}^{(k)} - \hat{\boldsymbol{\mu}}_W)(\mathbf{x}^{(k)} - \hat{\boldsymbol{\mu}}_W)^{\mathsf{T}}$ and rank($\hat{\boldsymbol{\Sigma}}_W$) $\leq |K^C|$. If $|K^C| < d_x$, then (4) yields a singular matrix. For this reason, adaptation of the covariance matrix in AIS is challenging due to the chance of weight degeneracy of the samples at each iteration.

We can characterize the degeneracy of the importance weights in AIS at any iteration through the effective sample size (ESS) [19], which can approximated as,

$$\hat{\eta}_i = \frac{1}{\sum_{m=1}^M (\bar{w}_i^{(m)})^2}. \tag{5}$$

It is important to note that the approximation in (5) is only true under certain conditions [20]. Moreover, other approximations for ESS can be used, such as $\tilde{\eta}_i = \max(\bar{w}_i^{(m)})^{-1}$, though for our experiments we use eq. (5). Under the assumption that $\text{rank}(\mathbf{\Sigma}_{ML}) = M$ at each AIS iteration (i.e. large enough variability in the generated samples), our criterion for robust covariance adaptation rests upon guaranteeing that the ESS at each iteration of the algorithm satisfies some lower bound condition.

B. Algorithm Description

In this section, we formalize a novel algorithm, *covariance adaptive importance sampling* (CAIS), which guarantees robust covariance adaptation for a population of proposal distributions.

1) **Initialization:** Initialize the proposal parameters $\boldsymbol{\mu}_{1,d}, \mathbf{\Sigma}_{1,d}$ for $d = 1, ..., D$ and set $i = 1$.
2) **Generate samples:** Draw N samples from each mixand,
$$\mathbf{x}_{i,d}^{(n)} \sim q(\mathbf{x}; \boldsymbol{\mu}_{i,d}, \mathbf{\Sigma}_{i,d}), \quad n = 1, ..., N, \quad d = 1, ..., D.$$
3) **Compute weights:** Evaluate the standard importance weights of each sample,
$$w_{i,d}^{(n)} = \frac{\pi(\mathbf{x}_{i,d}^{(n)})}{q(\mathbf{x}_{i,d}^{(n)}; \boldsymbol{\mu}_{i,d}, \mathbf{\Sigma}_{i,d})}, \quad n = 1, ..., N, \quad d = 1, ..., D,$$
and normalize locally,
$$\bar{w}_{i,d}^{(n)} = \frac{w_{i,d}^{(n)}}{\sum_{j=1}^{N} w_{i,d}^{(j)}}, \quad n = 1, ..., N, \quad d = 1, ..., D. \tag{6}$$
We note that alternative weighting schemes can be used for target approximation, such as the deterministic mixture (DM) weights presented in [3], [2], [4].
4) **Compute local ESS:** For each set of samples and normalized weights, $\{\mathbf{x}_{i,d}^{(n)}, \bar{w}_{i,d}^{(n)}\}_{n=1}^{N}$, compute the local ESS, $\hat{\eta}_{i,d}$, by applying eq. (5).
5) **Update mean:** Update the mean of each mixand by taking the weighted sample mean,
$$\boldsymbol{\mu}_{i+1,d} = \sum_{n=1}^{N} \mathbf{x}_{i,d}^{(n)} \bar{w}_{i,d}^{(n)}, \quad d = 1, ..., D. \tag{7}$$
6) **Update covariance matrix:** For $d = 1, ..., D$,
   a) If $\hat{\eta}_{i,d} \geq N_T$, update the covariance matrix as,
   $$\mathbf{\Sigma}_{i+1,d} = \sum_{n=1}^{N} \bar{w}_{i,d}^{(n)} (\mathbf{x}_{i,d}^{(n)} - \boldsymbol{\mu}_{i,d})(\mathbf{x}_{i,d}^{(n)} - \boldsymbol{\mu}_{i,d})^{\intercal}.$$
   b) If $\hat{\eta}_{i,d} < N_T$, transform the importance weights as $w_{i,d}^{(n)*} = \psi(w_{i,d}^{(n)})$ for $n = 1, ..., N$. Normalize the transformed weights by applying eq. (6). Compute the tempered mean $\boldsymbol{\mu}_{i,d}^{*}$ by applying eq. (7) given the normalized transformed weights. Update the covariance matrix as,
   $$\mathbf{\Sigma}_{i+1,d} = \sum_{n=1}^{N} \bar{w}_{i,d}^{(n)*} (\mathbf{x}_{i,d}^{(n)} - \boldsymbol{\mu}_{i,d}^{*})(\mathbf{x}_{i,d}^{(n)} - \boldsymbol{\mu}_{i,d}^{*})^{\intercal}.$$
7) **Check stopping condition:** If $i = I$, then return $\{\mathbf{x}_{i,d}^{(n)}, w_{i,d}^{(n)}\}$ for $d = 1, ..., D$ and $n = 1, ..., N$. Otherwise, set $i = i+1$ and go to Step 2.

C. Algorithm Summary

The novel technique adapts both the mean and covariance matrix for a population of proposal distributions. We emphasize that the untransformed weights are used in order to adapt the mean of each proposal, while the covariance adaptation is conditioned on the local ESS of the samples drawn by each proposal. If the local ESS is above a threshold N_T, the untransformed weights are used for the covariance matrix adaptation. Otherwise, the importance weights are transformed using a weight transformation function $\psi(\cdot)$ and the covariance is adapted using the transformed weights, yielding a wider and hence more stable proposal.

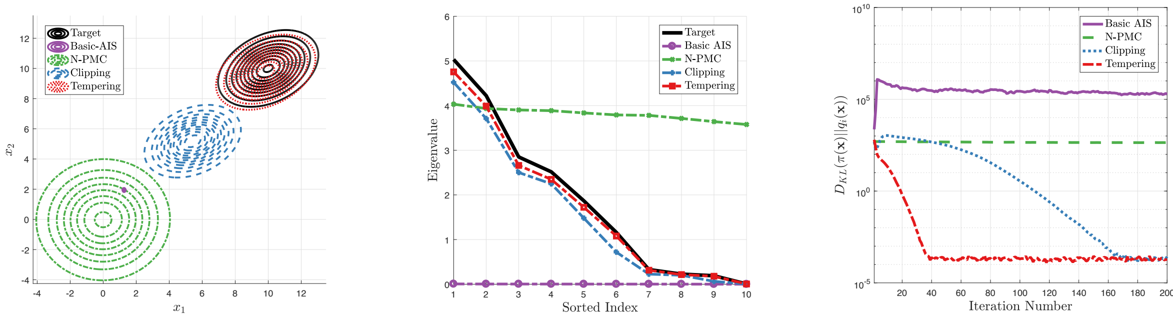

Fig. 1: (Example 1) The left figure shows the slice $[x_1, x_2]^\top$ of the adapted proposal distribution after 25 iterations. Note that the basic AIS provides a poor covariance adaptation with the proposal degenerating to a delta in a part of space which is far away from the target distribution. The middle figure compares the sorted eigenvalues of the target distribution's covariance matrix with that of the proposal distribution after 25 iterations. The right figure shows the evolution of the Kullback-Liebler divergence between the target distribution and the proposal distribution over 200 iterations.

D. Choice of the Weight Transformation Function

1) Weight Clipping: One choice for $\psi(\cdot)$ is the clipping function utilized in N-PMC. N-PMC transforms the weights by building a sorted permutation of the unnormalized weights $\{w^{(n)}\}_{n=1}^N$ as

$$w^{(k_1)} \geq w^{(k_2)} \geq ... \geq w^{(k_N)}, \tag{8}$$

and then transforming the subset of the greatest $1 < N_T < N$ weights as

$$\psi(w^{(n)}) = \min\big(w^{(n)}, w^{(k_{N_T})}\big). \tag{9}$$

This choice of $\psi(\cdot)$ guarantees that the ESS determined by the transformed weights is bounded below by N_T. Intuitively, in the case of extreme degeneracy, this can be interpreted as adapting the covariance matrix using the ML estimate of the covariance considering only the N_T highest weighted samples.

2) Weight Tempering: While (9) guarantees that the minimum ESS condition is satisfied, it also forces the N_T largest weights to take identical value, eliminating any information about which of these weights are most representative of the target distribution. An alternative choice, which offers more flexibility, is the weight tempering function,

$$\psi(w_i^{(n)}) = \big(w_i^{(n)}\big)^{\frac{1}{\gamma}}, \quad \gamma \geq 1, \tag{10}$$

where γ controls the transformation of the importance weights.

Note that all strategies transform the importance weights but preserve the position of the samples. More advanced mechanisms could be devised by implementing covariance estimators where the samples are also transformed, following the ideas of [21].

E. Choice of Parameters

The choice of the parameter N_T depends on the dimension of the target, d_x. In order to guarantee stability of the covariance update, we must have that $N_T > d_x$. However, we note that under poor initialization of the proposal distribution, a choice of large N_T (relative to N) will force a strong nonlinear transformation of the importance weights, regardless of the $\psi(\cdot)$ chosen. Consequently, the adapted covariance matrix will not differ from the one from the previous iteration, resulting in slower convergence to the true target.

The weight tempering function should be such that the ESS of the transformed weights is $\hat{\eta}^* \approx N_T$. In order to find the optimal γ, it is sufficient to employ a combination of loose grid search and fine grid search, similar to the approach used in [22]. For γ that is ϵ-suboptimal (i.e. $|N_T - \hat{\eta}^*| < \epsilon$), the computational complexity of a standard one-dimensional grid-search is $\mathcal{O}(\frac{1}{\epsilon})$. For a reasonable choice of ϵ (not too small), this is overshadowed by the computational complexity of each covariance matrix update, which is $\mathcal{O}(Nd_x^2)$.

IV. Numerical Examples

A. Example 1: Unimodal Gaussian Target

We consider a toy example to demonstrate fundamental differences between existing AIS schemes which adapt the covariance matrix and the novel strategy. In particular, the target of interest is a 10-dimensional unimodal Gaussian distribution with $\nu_i = 10$ for $i = 1, ..., 10$ and a non-diagonal covariance matrix randomly generated from a Wishart distribution. We use a single proposal

Method	$\sigma = 0.5$	$\sigma = 1$	$\sigma = 2$	$\sigma = 3$	$\sigma = 5$	$\sigma = 7$	$\sigma = 10$
N-PMC [best]	8.8445	8.0085	5.7241	2.3912	1.8289	5.6683	10.1791
N-PMC [worst]	16.3521	17.1335	14.8301	12.7081	9.6968	11.3779	16.3428
APIS [best]	2.6422	0.5863	0.3556	3.2842	12.9770	17.9266	22.6413
APIS [worst]	3.5907	1.1710	0.6811	5.0860	14.3068	18.5037	23.7821
DM-PMC [best]	3.0323	0.9643	1.2454	5.7601	14.1088	17.6142	21.8480
DM-PMC [worst]	15.2424	16.7463	18.1355	15.1963	18.6797	23.4776	34.9426
CAIS [C1]	4.7118	2.8164	1.2644	0.3865	0.4045	1.5499	5.9985
CAIS [C2]	6.7016	2.9347	0.4929	0.1780	**0.3214**	**0.7211**	**4.6918**
CAIS [T1]	1.7740	0.3834	0.2185	0.2378	0.5632	2.1462	8.8714
CAIS [T2]	**0.5793**	**0.1931**	**0.0995**	**0.1362**	0.4035	0.9717	7.9077

TABLE II: Average MSE in approximation of mean of multimodal target distribution. For the novel method (CAIS), we show four different parameter configurations. [C1] uses the clipping function and [T1] uses the tempering function with $D = 25$ and $N_T = 0.1 \times N$. [C2] uses the clipping function and [T2] uses the tempering function with $D = 50$ and $N_T = 0.3 \times N$.

distribution for each method and initialize the mean as $\boldsymbol{\mu}_1 = \mathbf{0}$ and the covariance matrix as $\boldsymbol{\Sigma}_1 = 4 \times \mathbb{I}_{10}$, where $\mathbb{I}_{10}$ is the 10×10 identity matrix. This is a poor initialization since it is far from the mean of the target distribution. We draw $N = 500$ samples over $I = 200$ iterations. For the novel covariance adaptation strategy, we fix $N_T = 50$, which is greater than the dimension $d_x = 10$. The results are averaged over 500 Monte Carlo runs.

Results for this example are shown in Figure 1. We tested four different schemes which continuously adapt the covariance matrix after each iteration of the algorithm, including the basic AIS algorithm from Table I, with proposal adaptation using eq. (4) and (7). In the left-hand plot of Figure 1, we can see that the basic AIS algorithm adapts to a peaky-proposal distribution, resulting in limited diversity in the generation of samples due to weight degeneracy. The N-PMC algorithm manages to keep the adaptation of the covariance matrix more robust. However, due to the effect of the transformed weights on the adaptation of the location parameter, the proposal distribution remains far from the target. The novel covariance adaptation strategies adapt the best, with the strategy utilizing the weight tempering function performing best.

The center plot shows the sorted eigenvalues of the adapted covariance matrix and compares to that of the target distribution. The basic AIS method results in a covariance matrix whose eigenvalues are practically 0. The N-PMC method manages to keep the eigenvalues nonzero, but fails to adapt to the covariance matrix of the target. The novel strategy works best with either choice of weight transformation function.

The right-hand plot shows the evolution of the Kullback-Liebler (KL) distance between the target and proposal as a function of iteration of the algorithm. This plot emphasizes the advantage of utilizing the weight tempering function over the weight clipping function. We can see that with the weight tempering function, after just 40 iterations, the KL distance reaches a minimum, whereas the weight clipping function does not reach this value until after 170 iterations.

B. Example 2: High-Dimensional Multimodal Target

We now consider the following target distribution

$$\pi(\mathbf{x}) = \frac{1}{3}\sum_{k=1}^{3}\mathcal{N}(\mathbf{x};\boldsymbol{\nu}_k,\boldsymbol{\Lambda}_k), \quad \mathbf{x} \in \mathbb{R}^{10}, \tag{11}$$

where $\nu_{1,j} = 6$, $\nu_{2,j} = -5$ for $j = 1, ..., 10$, and $\boldsymbol{\nu}_3 = [1, 2, 3, 4, 5, 5, 4, 3, 2, 1]^\top$. The covariances matrices, $\boldsymbol{\Lambda}_k$ are non-diagonal and were randomly generated from the Wishart distribution. Our objective is to estimate the mean of the target, which is given by, $E_\pi(\mathbf{X}) = [\frac{2}{3}, 1, \frac{4}{3}, \frac{5}{3}, 2, 2, \frac{5}{3}, \frac{4}{3}, 1, \frac{2}{3}]^\top$. The prior mean of each proposal distribution is drawn uniformly, such that $\boldsymbol{\mu}_{1,d} \sim \mathcal{U}([-10, 10]^{10})$. This is considered a good initialization since all modes of the target distribution are contained in this hypercube. We assume a prior covariance matrix for each proposal, $\boldsymbol{\Sigma}_{1,d} = \sigma^2\mathbb{I}_{10}$, where $\mathbb{I}_{10}$ is the 10×10 identity matrix. We test for a constant of $M = 10000$ samples, $D = \{20, 25, 40, 50, 80\}$ mixands for $I = \frac{4\times10^5}{M}$ iterations, where we draw $N = \frac{M}{D}$ samples from each proposal at each iteration. The results are averaged over 500 Monte Carlo simulations. The simulation results are summarized in Table II. The novel scheme (CAIS) is compared to three different state-of-the-art AIS algorithms: N-PMC, DM-PMC, and adaptive population importance sampling (APIS). The results indicate that the novel scheme outperforms all three schemes under their best parameter configurations for each value of σ.

V. Conclusions

In this work, we addressed the problem of proposal parameter adaptation in adaptive importance sampling methods for high-dimensional scenarios. The new proposed methods adapt location (mean) parameters using the standard importance weights, while the adaptation of scale (covariance matrix) parameters are conditioned on the effective sample size and utilize transformed importance weights. This robust adaptation improves the performance of AIS and also allows us for better reconstruction of the target distribution as a mixture of kernels with different adapted covariances. Simulation results show excellent performance of the novel strategies compared to other state-of-the-art algorithms. Furthermore, the novel adaptation methods have the potential to extend to other adaptive importance sampling methods, improving their performance for higher-dimensional systems.

References

[1] C. P Robert and G. Casella, *Monte Carlo Statistical Methods*, vol. 95, 2004.
[2] E. Veach and L. Guibas, "Optimally combining sampling techniques for Monte Carlo rendering," *Proceedings of the 22nd Conf. on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pp. 419–428, 1995.
[3] A. Owen and Y. Zhou, "Safe and Effective Importance Sampling," *Journal of the American Statistical Association*, vol. 95, no. 449, pp. 135–143, 2000.
[4] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Generalized multiple importance sampling," *arXiv preprint arXiv:1511.03095*, 2015.
[5] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Efficient multiple importance sampling estimators," *Signal Processing Letters, IEEE*, vol. 22, no. 10, pp. 1757–1761, 2015.
[6] 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.
[7] M. 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.
[8] S. K. Au and J. L. Beck, "A new adaptive importance sampling scheme for reliability calculations," *Structural Safety*, vol. 21, no. 2, pp. 135–158, 1999.
[9] 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.
[10] T. Bengtsson, P. Bickel, and B. Li, "Curse-of-dimensionality revisited: Collapse of the particle filter in very large scale systems," *Probability and Statistics: Essays in Honor of David A. Freedman*, vol. 2, pp. 316–334, 2008.
[11] 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.
[12] E. L. Ionides, "Truncated importance sampling," *Journal of Computational and Graphical Statistics*, vol. 17, no. 2, pp. 295–311, 2008.
[13] 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, 2013.
[14] A. Vehtari and A. Gelman, "Pareto smoothed importance sampling," *arXiv preprint arXiv:1507.02646*, 2015.
[15] 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.
[16] 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.
[17] L. Martino, V. Elvira, D. Luengo, and J. Corander, "An Adaptive Population Importance Sampler: Learning from Uncertainty," *IEEE Transactions on Signal Processing*, vol. 63, no. 16, pp. 4422–4437, 2015.
[18] O. Ledoit and M. Wolf, "Honey, I Shrunk the Sample Covariance Matrix," *The Journal of Portfolio Management*, vol. 30, no. 4, pp. 110–119, 2004.
[19] A. Kong, "A note on importance sampling using standardized weights," *University of Chicago, Dept. of Statistics, Tech. Rep*, vol. 348, 1992.
[20] L. Martino, V. Elvira, and F. Louzada, "Effective sample size for importance sampling based on discrepancy measures," *Signal Processing*, vol. 131, pp. 386–401, 2017.
[21] L. Martino, V. Elvira, and G. Camps-Valls, "Group importance sampling for particle filtering and mcmc," *arXiv preprint arXiv:1704.02771*, 2017.
[22] C. Hsu, C. Chang, and C. Lin, "A Practical Guide to Support Vector Classification," *BJU international*, vol. 101, no. 1, pp. 1396–400, 2008.