ANTI-TEMPERED LAYERED ADAPTIVE IMPORTANCE SAMPLING

Luca Martino[⋆]*, Víctor Elvira*[†]*, David Luengo*[‡]

[⋆] Image Processing Laboratory, Universitat de Valencia (Spain).

[†] IMT Lille Douai & CRIStAL (UMR CNRS 9189), Villeneuve d'Ascq (France).

[‡] Dep. of Signal Theory and Communic., Universidad Politécnica de Madrid, 28031 Madrid (Spain).

ABSTRACT

Monte Carlo (MC) methods are widely used for Bayesian inference in signal processing, machine learning and statistics. In this work, we introduce an adaptive importance sampler which mixes together the benefits of the Importance Sampling (IS) and Markov Chain Monte Carlo (MCMC) approaches. Different parallel MCMC chains provide the location parameters of the proposal probability density functions (pdfs) used in an IS method. The MCMC algorithms consider a tempered version of the posterior distribution as invariant density. We also provide an exhaustive theoretical support explaining why, in the presented technique, even an anti-tempering strategy (reducing the scaling of the posterior) can be beneficial. Numerical results confirm the advantages of the proposed scheme.

Index Terms— Adaptive importance sampling, MCMC methods, parallel chains, Bayesian inference.

1. INTRODUCTION

Importance sampling (IS) and Markov Chain Monte Carlo (MCMC) are two well-known Monte Carlo (MC) techniques widely used in signal processing for efficiently computing *a-posteriori* estimators [1–5]. This requires the approximation of integrals involving a complicated multidimensional target probability density function (pdf), $\pi(\mathbf{x})$ with $\mathbf{x} \in \mathbb{R}^{D_x}$ (the posterior pdf in Bayesian framework) [1, 5–7]. Both approaches use a simpler proposal pdf, $q(\mathbf{x})$, to draw random candidates which are weighted or filtered in different ways. In both cases, the quality of the performance is directly related to the discrepancy between the shape of the proposal and the target. For this reason, several adaptive schemes have been proposed [8–14].

In this work, we extend the approach proposed in [15]. More specifically, we describe an adaptive importance sampler which mixes together the IS and MCMC approaches, while preserving the advantages of both (see also [16, 17]). Indeed, MCMC outputs (upper layer) provide the location parameters for the proposal pdfs used in importance sampler (lower layer), where deterministic mixture (DM) weighting schemes are employed [18–23]. Unlike in [15], the MCMC chains address a tempered target pdf. In the first iterations, a standard tempering approach is used increasing the scale of the target, in order to foster the exploration behavior of the Markov chains. Then, an *anti-tempered* target is employed reducing the scale of the target (addressed by the Markov chains; the posterior considered in the IS layer remains always unchanged). The underlying idea is based on an equivalent proposal formulation that we describe later on. Therefore, a theoretical support of the proposed approach is given.

We also provide a specific adaptive IS algorithm belonging to the family described above, called *Anti Tempered Layered Adaptive Importance Sampler* (ANTE-LAIS). The cloud of N different proposal pdfs is updated considering N independent parallel Metropolis-Hastings (MH) chains [5, 24]. The outputs of the MH methods are then used as location parameters for N proposal densities. In a lower level, these proposal pdfs interact for providing a unique global IS estimator, using the DM weighting procedure [15]. Unlike other techniques in the literature (e.g., [9, 25–28]), the novel method does not require resampling steps, thus avoiding the loss of diversity in the population. The ANTE-LAIS is also an excellent algorithm in order to estimate the normalizing constant (a.k.a., marginal likelihood) of the target density. The new algorithm improves the performance (in terms of mean squared error) as shown in the numerical simulations.

2. PROBLEM STATEMENT AND BACKGROUND

In many applications, the goal is to make inference about a vector of unknowns as $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^{D_x}$ given the observed data as $\mathbf{y} \in \mathcal{Y} \subseteq \mathbb{R}^{D_y}$. The statistical information is summarized by the posterior pdf,

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

where $\ell(\mathbf{y}|\mathbf{x})$ is the likelihood function, $g(\mathbf{x})$ is the prior pdf, and $Z = Z(\mathbf{y})$ is the normalization factor. Generally, we are able to evaluate $\pi(\mathbf{x})$ where $\bar{\pi}(\mathbf{x}) = \frac{1}{Z}\pi(\mathbf{x})$. A generic

This work has been supported by the European Research Council (ERC) through the ERC Consolidator Grant SEDAL ERC-2014-CoG 647423.

moment related to $\bar{\pi}(\mathbf{x})$ is denoted as

$$I = \frac{1}{Z}\int_{\mathcal{X}} f(\mathbf{x})\pi(\mathbf{x})d\mathbf{x}, \tag{2}$$

where $f(\cdot)$ can be any integrable function of $\mathbf{x}$. For instance, with $f(\mathbf{x}) = \mathbf{x}$, the integral I represents the Minimum Mean Square Error (MMSE) estimator, $\hat{\mathbf{x}}_{MMSE} = \int_{\mathcal{X}} \mathbf{x}\bar{\pi}(\mathbf{x})d\mathbf{x}$.

2.1. Monte Carlo integration

In many practical scenarios, the integral I cannot be computed in a closed form, and numerical approximations are typically required. Many deterministic quadrature methods are available in the literature [29]. However, as the dimension D_x of the inference problem grows ($\mathbf{x} \in \mathbb{R}^{D_x}$), the deterministic quadrature schemes become less efficient.

Standard Monte Carlo. A common approach consists in approximating the integral I in Eq. (2) by using Monte Carlo (MC) quadrature [5, 24]. Namely, drawing K independent and identically distributed (i.i.d.) samples from the target pdf, i.e. $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(K)} \sim \bar{\pi}(\mathbf{x})$, we can build the estimator $\widehat{I}_K = \frac{1}{K}\sum_{k=1}^{K} f(\mathbf{x}^{(k)})$.

Sampling methods. However, generally, it is not possible to draw from $\bar{\pi}(\mathbf{x})$. Therefore, the MC algorithms employ a simpler proposal pdf $q(\mathbf{x})$ in order to generate random candidates which are filtered or weighted according to some suitable rule for building consistent estimators. For instance, the MCMC techniques yield a Markov chain where the invariant density is exactly the posterior, whereas in the IS schemes the samples drawn from $q(\mathbf{x})$ are properly weighted, for instance $w^{(k)} = \frac{\pi(\mathbf{x}^{(k)})}{q(\mathbf{x}^{(k)})}$, $k = 1, \ldots, K$. The resulting (correlated or weighted) samples are used for approximating I in Eq. (2). For instance, the IS (self-normalized) estimator is given by

$$\widetilde{I}_K = \sum_{k=1}^{K} \frac{w^{(k)}}{\sum_{i=1}^{K} w^{(i)}} f(\mathbf{x}^{(k)}). \tag{3}$$

Efficiency of sampling methods. The performance of a MC method depends on the discrepancy between the target $\bar{\pi}(\mathbf{x}) \propto \pi(\mathbf{x})$ and the proposal $q(\mathbf{x})$. Roughly speaking, the performance is improved if $q(\mathbf{x})$ is more similar to $\bar{\pi}(\mathbf{x})$. In the following, we describe a theoretical procedure for tuning the parameters of the proposal density and then present some practical implementations.

3. HIERARCHICAL PROCEDURE FOR PROPOSAL GENERATION

In general, tuning the parameters of the chosen proposal is a difficult task that requires statistical information of the target distribution. In this section, we deal with this important issue, focusing on the mean vector of the proposal pdf. More specifically, we consider a proposal defined by a mean vector $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$, which can be denoted as $q(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C}) \equiv q(\mathbf{x} - \boldsymbol{\mu}|\mathbf{C})$. We propose the following hierarchical procedure for generating a set of samples that will be employed afterwards within some Monte Carlo technique:

1. For $j = 1, \ldots, J$:
   (a) **[Upper layer]** Draw a mean vector $\boldsymbol{\mu}_j \sim h(\boldsymbol{\mu})$.
   (b) **[Lower layer]** Draw $\mathbf{x}_j^{(m)} \sim q(\mathbf{x}|\boldsymbol{\mu}_j, \mathbf{C})$ for $m = 1, \ldots, M$.
2. Use all the generated samples, $\mathbf{x}_j^{(m)}$ for $j = 1, \ldots, J$ and $m = 1, \ldots, M$, as candidates within some Monte Carlo method, for instance, an importance sampler.

Note that $h(\boldsymbol{\mu})$ plays the role of a prior pdf over the mean vector of q. Hence, the equivalent density of all generated samples can be expressed as

$$\widetilde{q}(\mathbf{x}|\mathbf{C}) = \int_{\mathcal{X}} q(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C})h(\boldsymbol{\mu})d\boldsymbol{\mu}, \tag{4}$$

i.e. we can consider that $\mathbf{x}_j^{(m)} \sim \widetilde{q}(\mathbf{x}|\mathbf{C})$ for all possible values of j and m. The density $\widetilde{q}$ is an *equivalent* proposal density corresponding to the hierarchical generating procedure. Note that $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_J$, are not directly used in the Monte Carlo estimator, but only the samples $\mathbf{x}_j^{(m)}$, for $j = 1, \ldots, J$ and $m = 1, \ldots, M$. Hence, the computational cost per iteration of this hierarchical approach is higher than in a standard approach, but this leads to substantial computational savings in terms of improved convergence towards the target, and thus a reduced number of iterations required, as shown in [15]. Note that, unlike other schemes such as [9,26], the generation of the $\boldsymbol{\mu}_j$'s in the upper level is independent of the samples $\mathbf{x}_j^{(m)}$ drawn in the lower level, thus facilitating the theoretical analysis of the resulting algorithms.[1]

Table 1. **Generic Layered Adaptive Importance Sampler**

1. For $j = 1, \ldots, J$:
   (a) Draw a mean vector $\boldsymbol{\mu}_j \sim h(\boldsymbol{\mu})$.
   (b) Draw M samples, $\mathbf{x}_j^{(m)} \sim q(\mathbf{x}|\boldsymbol{\mu}_j, \mathbf{C})$, with $m = 1, \ldots, M$.
2. Weight all the samples, $\mathbf{x}_j^{(m)}$ with

$$w_j^{(m)} = \frac{\pi(\mathbf{x}_j^{(m)})}{\frac{1}{J}\sum_{j=1}^{J} q(\mathbf{x}_j^{(m)}|\boldsymbol{\mu}_j, \mathbf{C})}, \tag{5}$$

   for $m = 1, \ldots, M$ and $j = 1, \ldots, J$.
3. Return $\{\mathbf{x}_j^{(m)}, w_j^{(m)}\}$ for $m = 1, \ldots, M$ and $j = 1, \ldots, J$.

Deterministic Mixture (DM) weighting. Note that in Eq. (5) a DM approach is used [18–20], i.e., the denominator

[1] Note that, in the ideal case described here, each $\boldsymbol{\mu}_j$ is also independent of the other $\boldsymbol{\mu}$'s. However, in several parts of this work, we also consider cases where correlation among the mean vectors, $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_J$, is introduced (due to the application of an MCMC scheme).

of the weight is the mixture $\psi(\mathbf{x}) = \frac{1}{J}\sum_{j=1}^{J} q(\mathbf{x}|\boldsymbol{\mu}_j, \mathbf{C})$. Observe also that $\psi(\mathbf{x})$ is a Monte Carlo approximation of the equivalent proposal $\widetilde{q}(\mathbf{x}|\mathbf{C})$ (using J samples, $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_J$).

3.1. Optimal prior $h^*(\boldsymbol{\mu}|\mathbf{C})$

Let us first assume a given parametric form of $q(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C})$ with its covariance matrix $\mathbf{C}$ fixed. We consider the problem of finding the optimal prior $h^*(\boldsymbol{\mu}|\mathbf{C})$ over the mean vector $\boldsymbol{\mu}$. Note that, since $\boldsymbol{\mu}$ is a mean vector, we have $q(\mathbf{x}|\boldsymbol{\mu}, \mathbf{C}) = q(\mathbf{x} - \boldsymbol{\mu}|\mathbf{C})$ and we can write the equivalent proposal as

$$\widetilde{q}(\mathbf{x}|\mathbf{C}) = \int_{\mathcal{X}} q(\mathbf{x} - \boldsymbol{\mu}|\mathbf{C}) h(\boldsymbol{\mu}|\mathbf{C}) d\boldsymbol{\mu}. \tag{6}$$

Note that the equivalent proposal above can be also seen as the pdf of the sum of two independent random variables. A desirable scenario consists in having the equivalent proposal $\widetilde{q}(\mathbf{x}|\mathbf{C})$ coinciding exactly with the target $\bar{\pi}(\mathbf{x})$, i.e.,

$$\widetilde{q}(\mathbf{x}|\mathbf{C}) := \bar{\pi}(\mathbf{x}). \tag{7}$$

In IS, as discussed in Section 2.1, if only specific moment I of the target must be approximated, the best equivalent proposal is $\widetilde{q}(\mathbf{x}|\mathbf{C}) \equiv |f(\mathbf{x})|\bar{\pi}(\mathbf{x})$, which minimizes the variance of the estimator [5, 30]. In order to obtain the optimal prior $h^*(\boldsymbol{\mu}|\mathbf{C})$ and assuming $\widetilde{q}(\mathbf{x}|\mathbf{C}) = \bar{\pi}(\mathbf{x}$ and noting that Eq. (6) corresponds to a convolution, we can express it in the Fourier domain through the convolution theorem [31, Section 3.3] as

$$\bar{\Pi}(\boldsymbol{\nu}) = Q(\boldsymbol{\nu}|\mathbf{C}) H^*(\boldsymbol{\nu}|\mathbf{C}),$$

where we use $E_p[e^{i\boldsymbol{\nu}\mathbf{x}}]$ to denote the expectation of $e^{i\boldsymbol{\nu}\mathbf{x}}$ w.r.t. the pdf $p(\mathbf{x})$ (i.e., the characteristic function of $p(\mathbf{x})$), and thus $Q(\boldsymbol{\nu}|\mathbf{C}) = E_q[e^{i\boldsymbol{\nu}\mathbf{x}}]$, $H^*(\boldsymbol{\nu}|\mathbf{C}) = E_{h^*}[e^{i\boldsymbol{\nu}\mathbf{x}}]$ and $\bar{\Pi}(\boldsymbol{\nu}) = E_{\bar{\pi}}[e^{i\boldsymbol{\nu}\mathbf{x}}]$ are the characteristic functions of $q(\mathbf{x})$, $h^*(\boldsymbol{\mu}|\mathbf{C})$ and $\bar{\pi}(\mathbf{x})$ respectively. Hence, the optimal prior pdf has the following characteristic function[2]

$$H^*(\boldsymbol{\nu}|\mathbf{C}) = \frac{\bar{\Pi}(\boldsymbol{\nu})}{Q(\boldsymbol{\nu}|\mathbf{C})}. \tag{8}$$

Unfortunately, in general it is not possible to determine analytically the optimal prior pdf $h^*(\boldsymbol{\mu}|\mathbf{C})$ from $H^*(\boldsymbol{\nu}|\mathbf{C})$, i.e. a closed-form expression for the inverse Fourier transform cannot be obtained, and alternative choices are then required.[3]

4. TEMPERED PRIOR OVER THE PROPOSAL PARAMETERS

In [15], we discuss and justify the use of

$$h(\boldsymbol{\mu}) = \bar{\pi}(\boldsymbol{\mu}). \tag{9}$$

[2]Using Bochner's theorem [32, Theorem 1.8.9], it can be easily proved that $H^*(\boldsymbol{\nu}|\mathbf{C})$ is indeed a characteristic function.

[3]In some cases, the optimal prior pdf $h^*(\boldsymbol{\mu}|\mathbf{C})$ can even not exist.

This choice allows very good performance when the posterior is multimodal and/or is heavy-tailed [15]. In this work, we propose to employ a sequence of tempered functions, i.e.,

$$h_t(\boldsymbol{\mu}) = [\bar{\pi}(\boldsymbol{\mu})]^{\beta_t}, \tag{10}$$

where $\beta_t > 0$, $t \in \mathbb{N}$ and

$$\beta_0 \leq \beta_1 \leq \ldots \leq \beta_t \ldots \leq \beta_\infty = \infty. \tag{11}$$

Therefore, we allow $\beta_t > 1$ (*anti-tempering*), so that the scale of the posterior is even reduced. More specifically, we have $\beta_t = \phi_\beta(\beta_{t-1})$, where ϕ_β is a suitable non-decreasing positive function.

Theoretical justification. Eq. (6) represents a convolution between the functions q and h. The convolution works as a smoother of $h(\boldsymbol{\mu})$ by using the proposal q, i.e., it is low-pass filter. Since $\det[\mathbf{C}] > 0$, in order to obtain $\widetilde{q}(\mathbf{x}|\mathbf{C}) = \bar{\pi}(\mathbf{x})$, the prior of the means $h(\boldsymbol{\mu}|\mathbf{C})$ must be sharper than the target $\bar{\pi}$ (given a fixed covariance matrix $\mathbf{C}$). This is the reason why the anti-tempering is advantageous in the proposed scheme. Due to the law of total variance,

$$\mathrm{Cov}_{\widetilde{q}} = \mathrm{Cov}_q + \mathrm{Cov}_h = \mathrm{Cov}_q + \mathrm{Cov}_{\bar{\pi}^\beta}. \tag{12}$$

Since ideally, $\widetilde{q} := \bar{\pi}$, it would be desirable that $\mathrm{Cov}_{\widetilde{q}} = \mathrm{Cov}_{\bar{\pi}}$. We can control $\mathrm{Cov}_{\widetilde{q}}$ by jointly tuning β and $\mathbf{C}$, where $\beta > 1$. See the example in Section 6.

Let us consider a covariance matrix $\mathbf{C}_t$ varying with the iteration t and, for simplicity, let us assume $\mathbf{C}_t = \sigma_t^2 \mathbf{I}$. Eq. (12) suggests us to use

$$\sigma_0 \leq \sigma_1 \leq \ldots \leq \sigma_t, \tag{13}$$

i.e., $\sigma_t = \phi_\sigma(\sigma_{t-1})$, where ϕ_σ is another suitable non-decreasing function. Since sampling directly from $\bar{\pi}(\boldsymbol{\mu})$ or $[\bar{\pi}(\boldsymbol{\mu})]^{\beta_t}$ is in general unfeasible, we apply an approximate procedure, i.e., MCMC steps are performed with invariant density $h_t(\boldsymbol{\mu}) = [\bar{\pi}(\boldsymbol{\mu})]^{\beta_t}$.

5. ANTI-TEMPERED LAIS

The previous ideas can be applied to a population of N proposal pdfs which also evolves with iteration index $t = 1, \ldots, T$. The Anti Tempered LAIS (ANTE-LAIS) algorithm adapts the location parameters of the N proposals with MCMC steps with invariant pdf $h_t(\boldsymbol{\mu}) = [\bar{\pi}(\boldsymbol{\mu})]^{\beta_t}$. At each iteration, the temperated parameter β_t is increased. For fostering the convergence of the chains is convenient to start with $\beta_0 < 1$. ANTE-LAIS is outlined in Table 2.

6. NUMERICAL SIMULATIONS

In this section we provide a numerical example where we evidence the effect of the tempering in the equivalent proposal

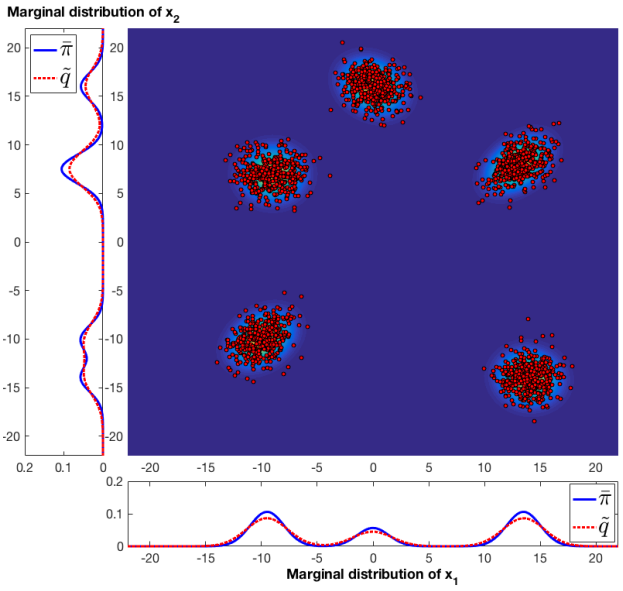

Fig. 1. Target distribution with standard LAIS ($\beta = 1$) with the means of the proposals represented with red solid circles. The marginals of $\bar{\pi}$ and $\widetilde{q}$ are also displayed.

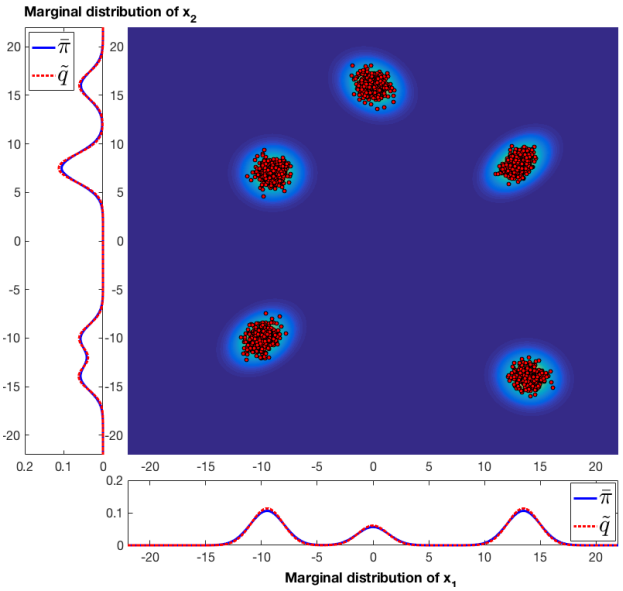

Fig. 2. Target distribution with standard ANTE-LAIS ($\beta = 4$) with the means of the proposals represented with red solid circles. The marginals of $\bar{\pi}$ and $\widetilde{q}$ are also displayed.

Table 2. Anti Tempered LAIS (ANTE-LAIS)

- For $t = 1, \ldots, T$:

1. Set $\beta_t = \phi_\beta(\beta_{t-1})$, $\sigma_t = \phi_\sigma(\sigma_{t-1})$ and $\mathbf{C}_{n,t} = \sigma_t^2 \mathbf{I}$, for all n.
2. Perform one transition of an MCMC technique with invariant pdf $h_t(\boldsymbol{\mu}) = [\bar{\pi}(\boldsymbol{\mu})]^{\beta_t}$ over the current population of means,
$$\mathcal{P}_{t-1} = \{\boldsymbol{\mu}_{1,t-1}, ..., \boldsymbol{\mu}_{N,t-1}\},$$
to obtain a new population of means,
$$\mathcal{P}_t = \{\boldsymbol{\mu}_{1,t}, ..., \boldsymbol{\mu}_{N,t}\}.$$
3. Draw $\mathbf{x}_{n,t}^{(m)} \sim q_{n,t}(\mathbf{x}|\boldsymbol{\mu}_{n,t}, \mathbf{C}_{n,t})$ for $m = 1, \ldots, M$ and $n = 1, \ldots, N$.
4. Compute the importance weights,
$$w_{n,t}^{(m)} = \frac{\pi(\mathbf{x}_{n,t}^{(m)})}{\frac{1}{N}\sum_{k=1}^{N} q_{k,t}(\mathbf{x}_{n,t}^{(m)}|\boldsymbol{\mu}_{k,t}, \mathbf{C}_{k,t})}, \qquad (14)$$
with $n = 1, \ldots, N$, and $m = 1, \ldots, M$.

- Return $\{\mathbf{x}_{n,t}^{(m)}, w_{n,t}^{(m)}\}$ for all m, n and t.

$\widetilde{q}$. Let us consider the bivariate target pdf, which is a mixture of 5 Gaussians,

$$\bar{\pi}(\mathbf{x}) = \frac{1}{5}\sum_{i=1}^{5} \mathcal{N}(\mathbf{x}; \boldsymbol{\nu}_i, \boldsymbol{\Sigma}_i), \quad \mathbf{x} \in \mathbb{R}^2, \qquad (15)$$

with means $\boldsymbol{\nu}_1 = [-10, -10]^\top$, $\boldsymbol{\nu}_2 = [0, 16]^\top$, $\boldsymbol{\nu}_3 = [13, 8]^\top$, $\boldsymbol{\nu}_4 = [-9, 7]^\top$, $\boldsymbol{\nu}_5 = [14, -14]^\top$, and covariance matrices $\boldsymbol{\Sigma}_1 = [2,\ 0.6; 0.6,\ 2]$, $\boldsymbol{\Sigma}_2 = [2,\ -0.4; -0.4,\ 2]$, $\boldsymbol{\Sigma}_3 = [2,\ 0.8; 0.8,\ 2]$, $\boldsymbol{\Sigma}_4 = [2,\ 0; 0,\ 2]$ and $\boldsymbol{\Sigma}_5 = [2,\ -0.1; -0.1,\ 2]$. We first run the traditional LAIS, i.e. with $\beta = 1$ and hence with no tempering. We set the parameters $T = 200$, $N = 50$, and the IS proposals in the lower level are Gaussians with covariance $\mathbf{C} = \sigma^2\mathbf{I}$, where $\sigma^2 = 1.5$. Figure 1 shows the multimodal bivariate target with the means $\boldsymbol{\mu}_{n,t}$ represented with red solid circles. We represent only the last half of iterations in order to remove the burn-in period of the MCMC in the upper layer. The two marginals of the target are displayed below and to the left of the main figure, both in solid blue line. We also show the approximation of the two marginals of the equivalent proposal $\widetilde{q}$, both in red dashed line. In Fig. 2 we run ANTE-LAIS in the same setup with $\beta = 4$. First, note that in ANTE-LAIS the means are more concentrated in the middle of the modes, since $\bar{\pi}^\beta$ is sharper than $\bar{\pi}$. Secondly, we see that the marginals of $\bar{\pi}$ are much better approximated by $\widetilde{q}$ with ANTE-LAIS ($\beta = 4$) than with standard LAIS ($\beta = 1$). In fact, ANTE-LAIS seems to replicate almost perfectly the target with the equivalent proposal $\widetilde{q}$. In standard LAIS, the equivalent proposal $\widetilde{q}$ seems to be a smoothed version of the target, due to the low-pass effect implicit in the convolution of Eq. (6) and explained in the theoretical justification of Section 4.

7. CONCLUSIONS

We have proposed a novel scheme that combines the benefits of the Markov chain Monte Carlo (MCMC) and importance sampling (IS) approaches. MCMC methods are run on top of a multiple IS scheme in order to find good location parameters for the population of proposal pdfs, which interact for providing a unique global IS estimator. The MCMC algorithms address tempered versions of the posterior pdf. First, a standard tempering is used facilitating the convergence of the Markov chains. In a second stage, an anti-tempering strategy is employed in order to obtain an *equivalent* proposal pdf more similar to the target, and therefore improving the efficiency of the IS estimators.

REFERENCES

[1] W. J. Fitzgerald, "Markov chain Monte Carlo methods with applications to signal processing," *Signal Processing*, vol. 81, no. 1, pp. 3–18, January 2001.

[2] L. Martino and J. Míguez, "A novel rejection sampling scheme for posterior probability distributions," *Proc. of the 34th IEEE ICASSP*, April 2009.

[3] David Luengo and Luca Martino, "Almost rejectionless sampling from Nakagami-m distributions ($m \geq 1$)," *IET Electronics Letters*, vol. 48, no. 24, pp. 1559–1561, Nov. 2012.

[4] L. Martino, V. Elvira, D. Luengo, A. Artes, and J. Corander, "Orthogonal MCMC algorithms," *IEEE Statistical Signal Processing Workshop (SSP)*, 2014.

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

[6] M. Bugallo, L. Martino, and J. Corander, "Adaptive importance sampling in signal processing," *Digital Signal Processing*, vol. 47, pp. 36–49, 2015.

[7] F. Liang, C. Liu, and R. Caroll, *Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples*, Wiley Series in Computational Statistics, England, 2010.

[8] C. Andrieu and J. Thoms, "A tutorial on adaptive mcmc," *Statistics and Computing*, vol. 18, no. 4, pp. 343–373, 2008.

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

[10] J. M. Cornuet *et al.*, "Adaptive multiple importance sampling," *Scandinavian Journal of Statistics*, vol. 39, no. 4, pp. 798–812, Dec. 2012.

[11] D. Luengo and L. Martino, "Fully adaptive Gaussian mixture Metropolis-Hastings algorithm," *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013.

[12] L. Martino, V. Elvira, D. Luengo, and F. Louzada, "Adaptive population importance samplers: A general perspective," *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*, pp. 1–5, 2016.

[13] L. Martino, V. Elvira, D. Luengo, and J. Corander, "An adaptive population importance sampler," *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2014.

[14] L. Martino, V. Elvira, D. Luengo, and J. Corander, "An adaptive population importance sampler: Learning from the uncertainty," *IEEE Transactions on Signal Processing*, vol. 16, no. 63, pp. 4422–4437, 2015.

[15] L. Martino, V. Elvira, D. Luengo, and J. Corander, "Layered Adaptive Importance Sampling," *Statistics and Computing*, vol. 27, no. 3, pp. 599–623, 2017.

[16] F. Beaujean and Caldwell A, "Initializing adaptive importance sampling with Markov chains," *arXiv:1304.7808*, 2013.

[17] Z. I. Botev, P. LEcuyer, and B. Tuffin, "Markov chain importance sampling with applications to rare event probability estimation," *Statistics and Computing*, vol. 23, pp. 271–285, 2013.

[18] A. Owen and Y. Zhou, "Safe and effective importance sampling," *Journal of the American Statistical Association*, vol. 95, no. 449, pp. 135–143, 2000.

[19] V. Elvira, L. Martino, D. Luengo, and M. F. Bugallo, "Generalized Multiple Importance Sampling," *arXiv:1511.03095*, 2015.

[20] 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.

[21] 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.

[22] T. Hesterberg, "Weighted average importance sampling and defensive mixture distributions," *Technometrics*, vol. 37, no. 2, pp. 185–194, 1995.

[23] E. Veach and L. Guibas, "Optimally combining sampling techniques for Monte Carlo rendering," *In SIGGRAPH 1995 Proceedings*, pp. 419–428, 1995.

[24] L. Martino and V. Elvira, *Metropolis Sampling*, Wiley StatsRef: Statistics Reference Online, 2017.

[25] 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.

[26] 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.

[27] E. Koblents and J. Míguez, "A population Monte Carlo scheme with transformed weights and its application to stochastic kinetic models," *Statistics and Computing*, 2014.

[28] G. R. Douc, J. M. Marin, and C. P. Robert, "Convergence of adaptive mixtures of importance sampling schemes," *Annals of Statistics*, vol. 35, pp. 420–448, 2007.

[29] R. L. Burden and J. D. Faires, *Numerical Analysis*, Brooks Cole, 2000.

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

[31] E. Lukacs, *Characteristic functions*, Griffin, 1970.

[32] Nikolai G Ushakov, *Selected topics in characteristic functions*, Walter de Gruyter, 1999.