PARALLEL INTERACTING MARKOV ADAPTIVE IMPORTANCE SAMPLING

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

[⋆] Dep. of Mathematics and Statistics, University of Helsinki, 00014 Helsinki (Finland).
[†] Dep. of Signal Theory and Communic., Universidad Carlos III de Madrid, 28911 Leganés (Spain).
[‡] Dep. of Signal Theory and Communic., Universidad Politécnica de Madrid, 28031 Madrid (Spain).

ABSTRACT

Monte Carlo (MC) methods are widely used for statistical inference in signal processing applications. A well-known class of MC methods is importance sampling (IS) and its adaptive extensions. In this work, we introduce an iterated importance sampler using a population of proposal densities, which are adapted according to an MCMC technique over the population of location parameters. The novel algorithm provides a global estimation of the variables of interest iteratively, using all the samples weighted according to the deterministic mixture scheme. Numerical results, on a multi-modal example and a localization problem in wireless sensor networks, show the advantages of the proposed schemes.

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

1. INTRODUCTION

Monte Carlo methods are widely used in signal processing applications [1]. Importance sampling (IS) and Markov Chain Monte Carlo (MCMC) [2] are two well-known Monte Carlo (MC) techniques to efficiently compute integrals involving a complicated multidimensional target probability density function (pdf), $\pi(\mathbf{x})$ with $\mathbf{x} \in \mathbb{R}^{D_x}$. 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 variance of the corresponding estimators is directly related to the discrepancy between the shape of the proposal and the target. For this reason, several adaptive schemes have been proposed [3–7].

In this work, we mix together the IS and MCMC approaches, while preserving the advantages of both. We introduce the *Markov Adaptive Importance Sampling* (MAIS) method, where MCMC outputs provide the location parameters for the proposal pdf used to obtain the IS estimators. Therefore, after a burn-in period, the proposal pdf is essentially jumping around the modes of the target, extracting local features using the set of weighted samples. Moreover, the estimation of the normalizing constant of the target is straightforward in this case. A population approach is also introduced. We consider N parallel chains interacting through the *deterministic mixture* weighting scheme [8]. At each iteration, the *Parallel Interacting MAIS* (PI-MAIS) algorithm computes iteratively a *global* IS estimate, taking into account all the samples generated up to that iteration. The cloud of N different proposal pdfs is updated considering N parallel Metropolis-Hastings (MH) chains [2]. The outputs of the MH methods are then used as location parameters for N proposal densities, which are jointly applied in a multiple IS approach. Namely, in an upper level, independent parallel chains adapt the location parameters of different proposal densities. In a lower level, these proposal pdfs interact for providing a unique global IS estimator [12]. Unlike other techniques in the literature (e.g., [3]), the novel method does not require resampling steps, thus avoiding the loss of diversity in the population. The new algorithm improves the performance (in terms of mean squared error) and the robustness w.r.t. the choice of the proposal parameters.

This work has been supported by the Spanish government's projects ALCIT (TEC2012-38800-C03-01), AGES (S2010/BMD-2422), DISSECT (TEC2012-38058-C03-01), OTOSiS (TEC2013-41718-R), and COMPREHENSION (TEC2012-38883-C02-01); by the BBVA Foundation (MG-FIAR project); by the ERC grant 239784 and AoF grant 251170; and by the EU's 7th FP through the Marie Curie ITN MLPM2012 (Grant No. 316861).

2. PROBLEM STATEMENT AND BACKGROUND

In many applications, the goal is obtaining the posterior density function (pdf) of a set of unknown parameters given the observed data. Mathematically, denoting the vector of unknowns as $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^{D_x}$ and the observed data as $\mathbf{y} \in \mathcal{Y} \subseteq \mathbb{R}^{D_y}$, the posterior pdf is given by

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

where $\ell(\mathbf{y}|\mathbf{x})$ is the likelihood function, $g(\mathbf{x})$ is the prior pdf, and $Z(\mathbf{y})$ is the normalization factor.[1] The computation of a particular moment of $\mathbf{x}$ is obtained as

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

[1] For the sake of simplicity, in the sequel $\bar{\pi}(\mathbf{x}|\mathbf{y})$ and $\pi(\mathbf{x}|\mathbf{y})$ will be denoted as $\bar{\pi}(\mathbf{x})$ and $\pi(\mathbf{x})$ respectively.

978-0-9928626-3-3/15/$31.00 ©2015 IEEE

where $f(\cdot)$ can be any integrable function of $\mathbf{x}$. In many practical scenarios, obtaining a closed-form expression for (2) is impossible, and Monte Carlo sampling techniques are used to obtain an approximate solution instead.

2.1. Importance Sampling (IS)

Let us consider K samples $(\mathbf{x}_1, \ldots, \mathbf{x}_K)$ drawn from a proposal pdf, $q(\mathbf{x})$, with heavier tails than the target, $\pi(\mathbf{x})$. The classical importance weights associated to the samples are given by

$$w_k = \frac{\pi(\mathbf{x}_k)}{q(\mathbf{x}_k)}, \quad k = 1, \ldots, K. \tag{3}$$

Using the samples and weights, (2) can be approximated as

$$\hat{I} = \frac{1}{K\hat{Z}} \sum_{k=1}^{K} w_k f(\mathbf{x}_k), \tag{4}$$

where $\hat{Z} = \frac{1}{K}\sum_{k=1}^{K} w_k$ is an unbiased estimator of $Z = \int_{\mathcal{D}} \pi(\mathbf{x}) d\mathbf{x}$ [2]. Eq. (4) provides a consistent estimator of I by means of a particle approximation of the target distribution, $\{\mathbf{x}_k, \bar{\rho}_k\}_{k=1}^{K}$ with

$$\bar{\rho}_k = \frac{w_k}{\sum_{k=1}^{K} w_k}, \quad k = 1, \ldots, K.$$

For an arbitrary $f(\mathbf{x})$, the variance of the *importance sampling* (IS) estimators depends on the discrepancy between $\pi(\mathbf{x})$ and $q(\mathbf{x})$ [2]. Therefore, having a proposal tailored to the target is essential to obtain good IS estimators.

2.2. Multiple Importance Sampling (MIS)

The IS approach can be easily extended using different proposal densities, $q_i(\mathbf{x})$ for $i = 1, \ldots, Q = \frac{K}{M}$. In this case, we draw M samples from each proposal, $\mathbf{x}_k \sim q_i(\mathbf{x})$ for $k = 1, \ldots, K = MQ$ and $i = \lceil \frac{k}{M} \rceil$, obtaining the *multiple importance sampling* (MIS) scheme.[2] For the sake of simplicity, let us consider a population of proposal densities, $q_i(\mathbf{x}_i|\boldsymbol{\mu}_i, \mathbf{C}_i)$, completely defined by their location ($\boldsymbol{\mu}_i$) and scale ($\mathbf{C}_i$) parameters. Then, the effective proposal pdf for the MIS approach is given by the following mixture of densities,

$$\psi(\mathbf{x}) = \frac{1}{Q} \sum_{i=1}^{Q} q_i(\mathbf{x}|\boldsymbol{\mu}_i, \mathbf{C}_i). \tag{5}$$

Now, at least two approaches can be taken to obtain a particle approximation of the target: using the classical IS weights given by (3) or computing the deterministic mixture (DM) weights [8]: $\tilde{w}_k = \pi(\mathbf{x}_k)/\psi(\mathbf{x}_k)$. The DM-MIS approach entails a larger computational cost, but provides estimators with a reduced variance w.r.t. the standard MIS [7, 9].

[2] Note that $\hat{I}$ is still given by (4), although the samples now come from different proposals, and thus $w_k = \pi(\mathbf{x}_k)/q_i(\mathbf{x}_k)$.

3. MARKOV ADAPTIVE IMPORTANCE SAMPLING

The underlying idea of this paper is trying to provide a good configuration for the location parameters ($\boldsymbol{\mu}_i$ for $i = 1, \ldots, Q$) iteratively. Note that the mixture in Eq. (5) can be seen as a kernel density approximation of the target pdf, $\bar{\pi}(\mathbf{x}) \propto \pi(\mathbf{x})$, where the proposals q_i play the role of the kernels [10, Chapter 6]. Hence, following kernel density estimation arguments, the ideal configuration for the location parameters is $\boldsymbol{\mu}_i \sim \bar{\pi}(\boldsymbol{\mu})$. In order to obtain $\boldsymbol{\mu}_i \sim \bar{\pi}(\boldsymbol{\mu})$, we consider an MCMC technique driving the IS algorithm: at each iteration, the MCMC outputs provide new location parameters, $\boldsymbol{\mu}_i$, for the proposal pdfs used in the IS estimation.

In this section, we introduce the Markov Adaptive Importance Sampling (MAIS), which is the single chain version of the Parallel Interacting Markov Adaptive Importance Sampling (PI-MAIS) algorithm described in the following section. Essentially, an MCMC method (in this case, a simple MH) is used to provide a suitable location parameter $\boldsymbol{\mu}_t$ and then a standard IS is performed:

1. Set $t = 1$ and $H_0 = 0$, and choose the values of M, T, $\boldsymbol{\mu}_1$ and $\mathbf{C}$.
2. Draw $\boldsymbol{\mu}'$ from a proposal pdf $\varphi(\boldsymbol{\mu}|\boldsymbol{\mu}_t, \mathbf{C})$, where $\boldsymbol{\mu}_t$ and $\mathbf{C}$ are the location and scale parameters, respectively.
3. Set $\boldsymbol{\mu}_{t+1} = \boldsymbol{\mu}'$ with probability

$$\alpha = \min\left[1, \frac{\pi(\boldsymbol{\mu}')\varphi(\boldsymbol{\mu}_t|\boldsymbol{\mu}', \mathbf{C})}{\pi(\boldsymbol{\mu}_t)\varphi(\boldsymbol{\mu}'|\boldsymbol{\mu}_{t-1}, \mathbf{C})}\right].$$

   Otherwise, set $\boldsymbol{\mu}_{t+1} = \boldsymbol{\mu}_t$ (with probability $1 - \alpha$).
4. Draw $\mathbf{x}_t^{(j)} \sim q(\mathbf{x}|\boldsymbol{\mu}_{t+1}, \mathbf{C})$ for $j = 1, \ldots, M$.
5. Weight the samples as

$$w_t^{(j)} = \frac{\pi(\mathbf{x}_t^{(j)})}{q(\mathbf{x}_t^{(j)}|\boldsymbol{\mu}_{t+1}, \mathbf{C})}.$$

6. Normalize the weights

$$\bar{\rho}_t^{(j)} = \frac{w_t^{(j)}}{\sum_{\tau=0}^{t}\sum_{j=1}^{M} w_\tau^{(j)}} = \bar{\rho}_{t-1}^{(j)} \frac{S_t}{H_t},$$

   where $S_t = \sum_{j=1}^{M} w_t^{(j)}$ and $H_t = H_{t-1} + S_t$.
7. If $t < T$, then set $t = t + 1$ and repeat from Step 2. Otherwise, stop.
8. Output $\{\mathbf{x}_t^{(j)}, \bar{\rho}_t^{(j)}\}$ for $j = 1, \ldots, M$ and $t = 1, \ldots, T$, and $\hat{Z}_T = \frac{1}{MT} H_t$.

Note that we have two different proposal densities in this Monte Carlo algorithm: one proposal for the MH adaptation (φ) and another one for the IS estimation (q). The MH technique is applied to obtain good location parameters for the underlying IS scheme. The global estimation is then obtained iteratively using $K = MT$ samples.

4. PARALLEL INTERACTING MAIS

The idea of the MAIS algorithm can be easily extended to a population of proposal pdfs. The Parallel Interacting MAIS (PI-MAIS) algorithm provides a particle approximation of the target by drawing samples from a population of proposal pdfs whose location parameters are updated following an MCMC scheme. The *deterministic mixture* approach [8] is applied to build the importance weights, providing the interaction among the N proposal pdfs. MAIS is a special case of PI-MAIS with $N = 1$. PI-MAIS is described below:

1. **Initialization:** Set $t = 1$, $\hat{I}_0 = 0$ and $H_0 = 0$. Choose the initial population,

$$\mathcal{P}_0 = \{\boldsymbol{\mu}_{1,0}, ..., \boldsymbol{\mu}_{N,0}\},$$

and N covariance matrices $\mathbf{C}_i$ $(i = 1, \ldots, N)$. Choose the parametric form of the N *normalized* proposal pdfs, $q_{i,t}$, with parameters $\boldsymbol{\mu}_{i,t}$ and $\mathbf{C}_i$ (e.g., Gaussians or t-Student). Let T be the total number of iterations.

2. **Upper Level - Update of the location parameters:** Perform one transition of an MCMC technique over the current population of location parameters,

$$\mathcal{P}_{t-1} = \{\boldsymbol{\mu}_{1,t-1}, ..., \boldsymbol{\mu}_{N,t-1}\},$$

to obtain a new population,

$$\mathcal{P}_t = \{\boldsymbol{\mu}_{1,t}, ..., \boldsymbol{\mu}_{N,t}\}.$$

3. **Lower Level - Interacting IS steps:**
   (a) Draw $\mathbf{z}_{i,t}^{(j)} \sim q_{i,t}(\mathbf{x}|\boldsymbol{\mu}_{i,t}, \mathbf{C}_i)$ for $j = 1, \ldots, M$ and $i = 1, \ldots, N$.
   (b) Compute the importance weights,

$$w_{i,t}^{(j)} = \frac{\pi(\mathbf{z}_{i,t}^{(j)})}{\frac{1}{N}\sum_{k=1}^{N} q_{k,t}(\mathbf{z}_{i,t}^{(j)}|\boldsymbol{\mu}_{k,t}, \mathbf{C}_k)}, \quad (6)$$

   with $i = 1, \ldots, N$, $j = 1, \ldots, M$; normalize them,

$$\bar{w}_{i,t}^{(j)} = \frac{w_{i,t}^{(j)}}{\sum_{i=1}^{N}\sum_{j=1}^{M} w_{i,t}^{(j)}}, \quad (7)$$

   and set $S_t = \sum_{i=1}^{N}\sum_{j=1}^{M} w_{i,t}^{(j)}$.
   (c) Calculate the current estimate of I,

$$\hat{J}_t = \sum_{i=1}^{N}\sum_{j=1}^{M} \bar{w}_{i,t}^{(j)} f(\mathbf{z}_{i,t}^{(j)}), \quad (8)$$

   update the *global estimate* using the following recursive formula,

$$\begin{cases} \hat{I}_t = \dfrac{1}{H_{t-1} + S_t}\left(H_{t-1}\hat{I}_{t-1} + S_t\hat{J}_t\right), \\ \hat{Z}_t = H_{t-1} + S_t, \end{cases} \quad (9)$$

   and set $H_t = H_{t-1} + S_t$.

4. **Stopping rule:** If $t < T$, then set $t = t + 1$ and repeat from Step 2. Otherwise, end.

5. **Outputs:** Return all the pairs $\{\mathbf{z}_{i,t}^{(j)}, \bar{\rho}_{i,t}^{(j)}\}$, where

$$\bar{\rho}_{i,t}^{(j)} = \frac{w_{i,t}^{(j)}}{\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{M} w_{i,t}^{(j)}} = \frac{1}{NMT}\frac{w_{i,t}^{(j)}}{\hat{Z}_T}. \quad (10)$$

   The estimate of the desired integral is then obtained as

$$\hat{I}_T \approx I = \frac{1}{Z}\int_{\mathcal{X}} f(\mathbf{x})\pi(\mathbf{x})d\mathbf{x}, \quad (11)$$

   and the normalizing constant of the target pdf is given by

$$\hat{Z}_T \approx Z = \int_{\mathcal{X}} \pi(\mathbf{x})d\mathbf{x}. \quad (12)$$

   The final locations of the means, $\boldsymbol{\mu}_{i,T}$ for $i = 1, \ldots, N$, can be used to estimate the locations of the modes of $\pi(\mathbf{x})$ (e.g., to perform maximum a posteriori estimation).

4.1. Upper Level: MCMC adaptation

The simplest possibility is to apply one iteration of N parallel MH techniques (one for each $\boldsymbol{\mu}_{i,t}$) in order to return $\boldsymbol{\mu}_{i,t+1}$:

For $i = 1, \ldots, N$:

1. Draw $\boldsymbol{\mu}'$ from a proposal pdf $\varphi_i(\boldsymbol{\mu}|\boldsymbol{\mu}_{i,t})$.
2. Set $\boldsymbol{\mu}_{i,t+1} = \boldsymbol{\mu}'$ with probability

$$\alpha = \min\left[1, \frac{\pi(\boldsymbol{\mu}')\varphi_i(\boldsymbol{\mu}_{i,t}|\boldsymbol{\mu}')}{\pi(\boldsymbol{\mu}_{i,t})\varphi_i(\boldsymbol{\mu}'|\boldsymbol{\mu}_{i,t})}\right].$$

   Otherwise, set $\boldsymbol{\mu}_{i,t+1} = \boldsymbol{\mu}_{i,t}$ (with probability $1 - \alpha$).

Note that the proposal pdfs for the MCMC stage, φ_i, could be the same as those used for the IS estimation, $q_{i,t}$. Furthermore, different schemes for parallel chains proposed in literature [11] can also be applied. Finally, the proposal pdf, φ_i, could also incorporate gradient information, as in the Metropolis adjusted Langevin algorithm (MALA).

4.2. Important Remarks

PI-MAIS is an iterated importance sampler where the cloud of location parameters is moved according to a transition of MCMC techniques. It is important to remark that all the proposal pdfs must be normalized to ensure a correct IS estimation. At each iteration t, M i.i.d. samples are drawn from each proposal. Then, PI-MAIS computes the current estimate of the desired integral, $\hat{J}_t$, and updates recursively the global estimates of the desired integral and the normalizing constant, $\hat{I}_t$ and $\hat{Z}_t$ respectively. The importance weights and, hence, the current estimate, $\hat{J}_t$, are obtained using the *deterministic mixture* approach proposed in [8] for a fixed (i.e., non-adaptive) setting, as this leads to more robust and efficient IS estimators [7, 9].

The final estimators, $\hat{I}_T$ and $\hat{Z}_T$, are iteratively obtained through an importance sampling approach using $K = NMT$ total samples drawn from $Q = NT$ proposal pdfs: N initial proposals chosen by the user, and $N(T-1)$ proposals updated by the algorithm. Indeed, denoting as $\mathbf{z}_{i,t}^{(j)}$ the i-th sample from the j-th proposal pdf at the t-th iteration and the corresponding weights as $w_{i,t}^{(j)}$ in Eq. (6), the final global estimator $\hat{I}_T$ can be expressed as

$$\hat{I}_T = \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{M} \bar{\rho}_{i,t}^{(j)} f(\mathbf{z}_{i,t}^{(j)}), \tag{13}$$

with $\bar{\rho}_{i,t}^{(j)}$ defined as in Eq. (10), and

$$\hat{Z}_T = \frac{1}{NMT}\sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{j=1}^{M} w_{i,t}^{(j)}. \tag{14}$$

The consistency of $\hat{I}_T$ and $\hat{Z}_T$ as the number of samples per iteration ($M \times N$) and/or the number of iterations (T) go to infinity is guaranteed by standard IS arguments [2, Chapter 14]. See [12] for the theoretical analysis and further comments about PI-MAIS and other hierarchical MCMC schemes.

5. NUMERICAL SIMULATIONS

5.1. Multimodal target distribution

In order to compare the performance of the proposed scheme with other algorithms, we first consider a bivariate multimodal target pdf, which is a mixture of 5 Gaussians:

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

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 $\boldsymbol{\Sigma}_1 = [2,\ 0.6; 0.6,\ 1]$, $\boldsymbol{\Sigma}_2 = [2,\ -0.4; -0.4,\ 2]$, $\boldsymbol{\Sigma}_3 = [2,\ 0.8; 0.8,\ 2]$, $\boldsymbol{\Sigma}_4 = [3,\ 0; 0,\ 0.5]$ and $\boldsymbol{\Sigma}_5 = [2,\ -0.1; -0.1,\ 2]$. We address the problem of computing the mean of the target, $E[\mathbf{X}] = [1.6, 1.4]^\top$ with $\mathbf{X}\sim\pi(\mathbf{x})$ (i.e., $Z = 1$, since the target is normalized). We compare the performance, in terms of Mean Squared Error (MSE), in the estimation using different sampling methodologies: **(1)** standard, non-adaptive, Multiple IS (MIS) approach using the classical IS weights; **(2)** the AMIS technique [4]; **(3)** the original PMC scheme in [3]; **(4)** the proposed PI-MAIS scheme. For all the previous techniques, we choose deliberately a "bad" initialization of the means (they are placed far away from the modes of the target). Thus, we can test the robustness of the algorithms and their adaptation ability. More specifically, the initial locations are selected uniformly within a square which does not contain any of the modes of the target: $\boldsymbol{\mu}_{i,0}\sim\mathcal{U}([-4,4]\times[-4,4])$ for $i = 1,\dots,N$.

We use the same isotropic covariance matrix, $\mathbf{C}_i = \sigma^2\mathbf{I}_2$, for all the proposals, testing different values of $\sigma \in \{1,2,5,10\}$. For the MCMC adaptation, we use Gaussian proposals φ_i with covariance matrices $\boldsymbol{\Psi} = \lambda^2\mathbf{I}_2$ and $\lambda \in \{5,10,70\}$. We configure all the algorithms to yield always the same number of evaluations of the target: $L = 2\cdot 10^5$. In PI-MAIS the number of evaluations is $L = TN(M+1)$ and we have used $N = 100$, $T \in \{20,100,1000\}$ and $M = L/(TN) - 1 \in \{99,19,1\}$. In PMC we have set $N = 100$ and $T = 2000$, whereas in AMIS we have tested several sets of parameters and we only show the best and the worst results as in [7]. Table 1 shows the Mean Squared Error (MSE) in the estimation of the mean of the first component. All the results have been averaged over 2000 independent experiments. PI-MAIS outperforms the other algorithms in all cases except for $\sigma = 10$, where it outperforms PMC and the loss w.r.t. AMIS is negligible. Table 1 also shows that PI-MAIS is more robust w.r.t. the choice of the parameters than AMIS or PMC.

Alg. \ Std			$\sigma = 1$	$\sigma = 2$	$\sigma = 5$	$\sigma = 10$
MIS			41.95	64.51	2.17	0.015
PI-MAIS	$\lambda = 5$	$T = 20$	0.522	0.593	0.021	0.014
		$T = 100$	0.121	0.042	0.009	0.014
		$T = 1000$	**0.002**	0.016	0.010	0.027
	$\lambda = 10$	$T = 20$	0.155	0.288	0.018	0.013
		$T = 100$	0.012	0.053	**0.009**	0.014
		$T = 1000$	0.002	0.002	0.010	0.025
	$\lambda = 70$	$T = 20$	5.579	1.493	0.038	0.013
		$T = 100$	0.918	0.131	0.015	0.014
		$T = 1000$	0.104	**0.014**	0.012	0.027
AMIS		(best)	121.21	100.23	0.864	**0.012**
		(worst)	123.38	114.82	16.92	0.013
PMC			114.11	47.97	2.34	0.056

Table 1. MSE in the estimation of the mean of the target (first component). We set $N = 100$ and $L = 2\cdot 10^5$ total number of target evaluations for MIS, PI-MAIS, AMIS and PMC. For AMIS, we only show the best and worst results.

5.2. Localization Problem in a Wireless Sensor Network

We consider the problem of positioning a target in a bidimensional space using range measurements, which is a problem that appears frequently in localization applications in wireless sensor networks [13]. Namely, we consider a random vector $\mathbf{X} = [X_1, X_2]^\top$ to denote the target position in the plane $\mathbb{R}^2$. The position of the target is then a specific realization $\mathbf{X} = \mathbf{x}$. The range measurements are obtained from 3 sensors located at $\mathbf{h}_1 = [-10,2]^\top$, $\mathbf{h}_2 = [8,8]^\top$ and $\mathbf{h}_3 = [-20,-18]^\top$. The observation equations are given by

$$Y_j = a\log\left(\frac{||\mathbf{x}-\mathbf{h}_j||}{0.3}\right) + \Theta_j, \quad j = 1,\dots,3, \tag{16}$$

where Θ_j are i.i.d. Gaussian random variables with pdfs $\mathcal{N}(\vartheta_j; 0, \omega^2)$ for $j \in \{1, 2, 3\}$. We consider a prior density over ω, $p(\omega) = \mathcal{N}(\omega; 0, 25)I(\omega > 0)$, where $I(\omega > 0)$ is 1 if $\omega > 0$ and 0 otherwise. The parameter $A = a$ is also unknown and we consider again a half-Gaussian prior, $A \sim p(a) = \mathcal{N}(a; 0, 25)I(a > 0)$. Moreover, we also apply Gaussian priors over $\mathbf{X}$, $p(x_i) = \mathcal{N}(x_i; 0, 25)$ with $i \in \{1, 2\}$. Thus, the target pdf is

$$\pi(x_1, x_2, a, \omega|\mathbf{y}) \propto \ell(\mathbf{y}|x_1, x_2, a, \omega)p(x_1)p(x_2)p(a)p(\omega),$$

where $\mathbf{y} \in \mathbb{R}^{D_y}$ is the measurement vector.

We simulate $D_y = 30$ observations from the model ($D_y/3 = 10$ from each of the three sensors), fixing $x_1 = 3$, $x_2 = 3$, $a = -20$ and $\omega = 5$. Our goal is computing the expected value of $(X_1, X_2, A, \Omega) \sim \pi(x_1, x_2, a, \omega|\mathbf{y})$ via Monte Carlo, in order to estimate the position of the target, the parameter a and the standard deviation ω of the noise in the system. We apply the PI-MAIS and PMC schemes using N Gaussian proposals as in the previous example. For both algorithms, we initialize the cloud of particles to be spread throughout the space of the variables of interest, i.e.,

$$\boldsymbol{\mu}_{i,0} \sim \mathcal{N}(\boldsymbol{\mu}; \mathbf{0}, 30^2\mathbf{I}_4), \quad i = 1, ..., N,$$

and fix the scale parameters as $\mathbf{C}_i = \text{diag}(\sigma_{i,1}^2, \ldots, \sigma_{i,4}^2)\mathbf{I}_4$ with $i = 1, \ldots, N$. The values of the standard deviations, $\sigma_{i,j}$, are chosen randomly for each Gaussian pdf: $\sigma_{i,j} \sim \mathcal{U}([1, R])$ with $R \in \{5, 10, 30\}$. For the MCMC adaptation, all proposals φ_i are again Gaussians with $\boldsymbol{\Psi} = \lambda^2\mathbf{I}_4$ and $\lambda \in \{5, 10, 70\}$. The MSE of the estimators (averaged over 3000 independent runs) are provided in Table 2 for $N = 100$ and $T \in \{20, 100, 1000\}$ in PI-MAIS. PMC has been simulated using $N = 100$ and $T = 2000$. PI-MAIS always outperforms PMC, showing both its robustness and flexibility.

Alg. \ $\sigma_{i,j} \sim$			$\mathcal{U}([1,5])$	$\mathcal{U}([1,10])$	$\mathcal{U}([1,30])$
PI-MAIS	$\lambda = 5$	$T = 20$	0.382	0.351	0.363
		$T = 100$	0.073	0.074	0.071
		$T = 1000$	**0.017**	**0.016**	**0.017**
	$\lambda = 10$	$T = 20$	0.570	0.594	0.561
		$T = 100$	0.139	0.143	0.143
		$T = 1000$	0.040	0.041	0.039
	$\lambda = 70$	$T = 20$	4.365	4.001	4.190
		$T = 100$	16.23	16.78	16.40
		$T = 1000$	35.52	215.19	64.21
PMC			0.064	0.435	0.153

Table 2. MSE in the estimation of $E[(X_1, X_2, A, \Omega)]$, using PI-MAIS and PMC, for different random choices of the scale parameters, drawing $L = 2 \cdot 10^5$ samples in all cases.

6. CONCLUSIONS

In this paper, we have proposed a novel hierarchical Monte Carlo technique that combines the Markov chain Monte Carlo (MCMC) and importance sampling (IS) approaches: Parallel Interacting Markov adaptive importance sampling (PI-MAIS). Essentially, 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 proposed approach outperforms other adaptive IS methods (such as PMC or AMIS) in terms of efficiency and robustness w.r.t. the initialization parameters.

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

[3] 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.

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

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

[6] L. Martino, J. Read, and D. Luengo, "Independent doubly adaptive rejection Metropolis sampling," *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2014.

[7] L. Martino, V. Elvira, D. Luengo, and J. Corander, "An adaptive population importance sampler: Learning from the uncertainty," *IEEE Transactions on Signal Processing (DOI:10.1109/TSP.2015.2440215)*, 2015.

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

[9] V. Elvira *et al.*, "Efficient multiple importance sampling estimators," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1757–1761, Oct. 2015.

[10] D. W. Scott, *Multivariate density estimation: theory, practice, and visualization*, John Wiley & Sons, 2009.

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

[12] L. Martino *et al.*, "Layered adaptive importance sampling," *arXiv:1505.04732*, 2015.

[13] L. Martino and J. Míguez, "Generalized rejection sampling schemes and applications in signal processing," *Signal Processing*, vol. 90, no. 11, pp. 2981–2995, Nov. 2010.