A GRADIENT ADAPTIVE POPULATION IMPORTANCE SAMPLER

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

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

ABSTRACT

Monte Carlo (MC) methods are widely used in signal processing and machine learning. A well-known class of MC methods is composed of importance sampling and its adaptive extensions (e.g., population Monte Carlo). In this paper, we introduce an adaptive importance sampler using a population of proposal densities. The novel algorithm dynamically optimizes the cloud of proposals, adapting them using information about the gradient and Hessian matrix of the target distribution. Moreover, a new kind of interaction in the adaptation of the proposal densities is introduced, establishing a trade-off between attaining a good performance in terms of mean square error and robustness to initialization.

Index Terms— Monte Carlo methods, adaptive importance sampling, population Monte Carlo (PMC), Hamiltonian Monte Carlo

1. INTRODUCTION

Monte Carlo methods are widely used in signal processing and communications [1, 2, 3]. The importance sampling (IS) technique [4, 5] is a well-known Monte Carlo (MC) methodology to efficiently compute integrals involving a complicated multidimensional target probability density function (pdf), $\pi(\mathbf{x})$ with $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^n$. The IS technique draws samples from a simple proposal pdf, $q(\mathbf{x})$, assigning weights to them according to the ratio between the target and the proposal, i.e., $w(\mathbf{x}) = \frac{\pi(\mathbf{x})}{q(\mathbf{x})}$. However, although the validity of this approach is guaranteed under mild assumptions, the variance of the estimator depends critically on the discrepancy between the shape of the proposal and the target. In order to solve this issue, several works are devoted to the design of adaptive IS (AIS) schemes [5], where the proposal density is updated by learning from all the previously generated samples [6, 7, 8]. In all of them, the adaptation is based on sampling and estimation arguments, so they only require to be able to evaluate the target function. AIS schemes have been applied successfully in different fields and problems [9, 10, 11, 12, 13]. However, several issues still remain: the robustness w.r.t. the initial conditions, the stability, or the ability to capture all the relevant features (e.g., the modes) for complicated distributions. These three aspects are strictly connected and the challenge grows with the dimension of the state space $\mathbb{R}^n$. Whenever information about the derivatives of the target is available, it is possible to increase the robustness of AIS schemes. This idea has been applied in the MCMC framework, where several Hamiltonian MCMC approaches have been proposed [14]. In all of these methods, the increased computational cost of evaluating the derivatives is balanced by a great enhancement in the performance. However, this idea has not been exploited yet in the importance sampling framework as far as we know.

In this work, we introduce the *gradient adaptive population importance sampling* (GAPIS) algorithm. At each iteration, GAPIS uses the information of the derivatives of the target to update the location (mean) and the scale (covariance) parameters of each proposal pdf. The mean adaptation is performed as a combination of two processes: an independent gradient ascent of each proposal (trying to reach the modes of the target), and a repulsion term that depends on all the proposals. This repulsion increases the exploratory ability of GAPIS w.r.t. other competing algorithms (like PMC [6, 15], AMIS [7], or APIS [8]), as it allows one proposal to describe a specific portion of the state space, while the remaining proposal pdfs explore other regions. The covariance adaptation is achieved by computing the Hessian of the logarithm of the target in the way described in Appendix A of [16]. Finally, GAPIS builds a multiple importance sampling estimator, weighting the samples according to the so-called *deterministic mixture* approach [17, 18]. Numerical results show that GAPIS exhibits an excellent performance, regardless of the initial conditions and parameters.

The paper is structured as follows. Section 2 provides the problem statement. Section 3 describes the GAPIS algorithm, including remarks about its implementation. The performance of GAPIS is then compared to other methods (PMC, APIS and static IS) through numerical simulations in Section 4. The paper ends with the conclusions in Section 5.

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

In many applications, the aim is inferring a variable of interest given a set of observations. Let us consider the variable of interest, $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^n$, and let $\mathbf{y} \in \mathcal{Y} \subseteq \mathbb{R}^d$ be the observed data. The posterior pdf is then

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

where $\ell(\mathbf{y}|\mathbf{x})$ is the likelihood function, $g(\mathbf{x})$ is the prior pdf, and $Z(\mathbf{y})$ is the model evidence or partition function (useful in model selection). In general, $Z(\mathbf{y})$ is unknown, so we consider the corresponding (generally unnormalized) target pdf,

$$\pi(\mathbf{x}) = \ell(\mathbf{y}|\mathbf{x})g(\mathbf{x}). \tag{2}$$

Our goal is computing efficiently some moment of $\mathbf{x}$, i.e., an integral measure w.r.t. the target pdf,

$$I = \frac{1}{Z}\int_{\mathcal{X}} f(\mathbf{x})\pi(\mathbf{x})d\mathbf{x}, \tag{3}$$

where f is typically assumed to be a smooth function of $\mathbf{x}$ and $Z = \int_{\mathcal{X}} \pi(\mathbf{x})d\mathbf{x}$.

3. THE GAPIS ALGORITHM

3.1. Algorithm description

The GAPIS technique allows the estimation of both integrals I and Z by drawing samples from a population of adaptive proposal pdfs. GAPIS iteratively refines the estimation, adjusting the location and scale parameters of the proposal pdfs making use of the first and second order statistics of the target (i.e., the mean and covariance). The adaptation is performed computing the gradient and the Hessian matrix of the log-target at the current location parameter $\boldsymbol{\mu}_i$ of the i-th proposal. The steps of the algorithm are described below.

1. **Initialization:** Set $t = 1$, $\hat{I}_0 = 0$ and $L_0 = 0$. Choose N normalized proposal pdfs,

   $$q_i^{(0)}(\mathbf{x}; \boldsymbol{\mu}_i^{(0)}, \mathbf{C}_i^{(0)}), \qquad i = 1, \dots, N,$$

   where $\boldsymbol{\mu}_i^{(0)}$ is the mean vector, and $\mathbf{C}_i^{(0)}$ is the covariance matrix. Let K be the number of samples drawn from each proposal per iteration, and T the total number of iterations. Calculate the gradient of the log-target distribution, $\nabla \log\, \pi(\mathbf{x})$.

2. **Mean adaptation**:

   (a) Update the repulsion term $G_t \geq 0$ using a suitable deterministic rule $G_t = h(t)$ (see Section 3.2 for further details).

   (b) The mean of the i-th proposal is adapted as

   $$\boldsymbol{\mu}_i^{(t)} = \boldsymbol{\mu}_i^{(t-1)} + \lambda \nabla \log\left(\pi(\boldsymbol{\mu}_i^{(t-1)})\right) + \sum_{j=1}^{N} \mathbf{r}_{i,j}^{(t-1)}, \tag{4}$$

   where $\lambda > 0$ is a constant,

   $$\mathbf{r}_{i,j}^{(t-1)} = G_t \frac{m_i m_j}{\|\mathbf{d}_{i,j}^{(t-1)}\|^3} \mathbf{d}_{i,j}^{(t-1)}, \tag{5}$$

   where $\|\cdot\|$ denotes the Euclidean norm operator, $\mathbf{d}_{i,j}^{(t-1)} = \boldsymbol{\mu}_i^{(t-1)} - \boldsymbol{\mu}_j^{(t-1)}$, and $m_i, m_j > 0$ are two positive terms that depend on the i-th and j-th proposals respectively.[1]

3. **Covariance adaptation**: The covariance matrix of the i-th proposal is adapted as

   $$\mathbf{C}_i^{(t)} = \left(\mathbf{H}_{\boldsymbol{\mu}_i^{(t)}}\right)^{-1},$$

   where $\mathbf{H}_{\boldsymbol{\mu}_i^{(t)}}$ is the Hessian matrix of $-\log(\pi(\mathbf{x}))$ evaluated at $\boldsymbol{\mu}_i^{(t)}$ (the new positions returned by Step 2).

4. **Sampling steps:**

   (a) Draw K independent samples from each proposal, i.e., $\mathbf{z}_{i,k}^{(t)} \sim q_i^{(t)}(\mathbf{x}; \boldsymbol{\mu}_i^{(t)}, \mathbf{C}_i^{(t)})$ for $k = 1 \dots, K$ and $i = 1, \dots, N$.

   (b) Compute the importance weights,

   $$w_{i,k}^{(t)} = \frac{\pi(\mathbf{z}_{i,k}^{(t)})}{\frac{1}{N}\sum_{j=1}^{N} q_j^{(t)}(\mathbf{z}_{i,k}^{(t)}; \boldsymbol{\mu}_i^{(t)}, \mathbf{C}_i^{(t)})}, \tag{6}$$

   for $\quad i = 1, \dots, N$, and $k = 1, \dots, K$.

   (c) Normalize the weights as

   $$\bar{w}_{i,k}^{(t)} = \frac{w_{i,k}^{(t)}}{S_t}, \tag{7}$$

   with $S_t = \sum_{i=1}^{N}\sum_{k=1}^{K} w_{i,k}^{(t)}$.

5. **Iterative IS estimation:** Obtain the "current" estimate of I as

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

   and the global estimate, using the recursive formula

   $$\hat{I}_t = \frac{1}{L_{t-1} + S_t}\left(L_{t-1}\hat{I}_{t-1} + S_t \hat{J}_t\right), \tag{9}$$

   where $L_t = L_{t-1} + S_t$. Note that the estimation of the normalizing constant is $\hat{Z}_t = \frac{1}{NKt} L_t$.

[1] The repulsion term in (4) mimicks Coulomb's law, with G_t playing the role of Coulomb's constant and the m_i playing the role of the electrical charges. Each proposal corresponds to an electron, and thus the repulsion between two of them is inversely proportional to the squared distance between their corresponding locations.

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

GAPIS returns the estimations $\hat{I}_T \approx I$ and $\hat{Z}_T \approx Z$ using KNT weighted samples.

3.2. Further observations and robust implementation

An important feature of GAPIS is that the adaptation procedure is completely decoupled from the sampling steps. This is one of the main differences w.r.t. similar techniques like PMC [6], AMIS [7], or APIS [8]. A second major difference is the use of information about the first and second derivatives of the target, which has been extensively exploited by MCMC algorithms (e.g., by Hamiltonian Monte Carlo methods), but not in the context of adaptive importance sampling.

Regarding the interaction among the proposal pdfs in Eq. (5), the simplest choice for $G_t = h(t)$ is a constant value $G_t = G$. In this case, the cloud of proposals tends to a static equilibrium around high probability regions of the target. The precise final static configuration depends on the initialization and the choice of G. In order to allow the location of the proposals to converge to the modes of the target pdf, this parameter should be reduced with time, vanishing to zero as $t \to \infty$. A possible simple choice in this case is

$$G_t = \exp(-\beta t), \tag{10}$$

for some $\beta > 0$. Another possible choice is $G_t = \frac{\beta}{\beta + t}$, frequently used in stochastic approximation algorithms to prevent the rate from decreasing too quickly [19, 20].

In the previous scenario, the proposal pdfs eventually converge to some modes of the target, thus completely ceasing their exploratory behaviour. However, some other modes of the target could be potentially missed, since there is no further exploratory activity. An alternative is to choose a periodic repulsion term $G_t \geq 0$, such as

$$G_t = \sin(2\pi\nu t) + \Delta, \qquad \nu \in \mathbb{R}^+, \Delta \geq 1. \tag{11}$$

This approach combines the advantages of placing the proposal pdfs around the modes for some iterations (when G_t becomes low and thus hardly any interaction among proposals occurs), with an increased range of the reachable sampling areas when the interaction term G_t increases (and thus high interaction levels are attained).

Regarding the terms m_i and m_j in Eq. (5), they can be simply set as $m_i = m_j = 1$. A more sophisticated choice is making them depend on the scale of the pdf, e.g., $m_{i,t} = \text{tr}(\mathbf{C}_i^{(t)})$. This choice is motivated by sampling reasons: the trace of the covariance matrix measures the range of the sampling area of the corresponding proposal. The idea is to avoid that two proposals cover the same portion of the state space.

In other possible variant of the algorithm, additive white Gaussian noise could be added to the update equation of $\boldsymbol{\mu}_i$ in order to facilitate the exploration of the state space and reduce the dependence on the initial conditions. Then, Eq. (5) becomes

$$\boldsymbol{\mu}_i^{(t)} = \boldsymbol{\mu}_i^{(t-1)} + \lambda \nabla \log(\pi(\mathbf{x})) + \sum_{i=1}^{N} \mathbf{r}_{i,j}^{(t-1)} + \eta_i,$$

with $\eta_i \sim \mathcal{N}(0, \epsilon)$ for $i = 1, \ldots, N$. Given the space constraints, here we focus on using a periodic repulsion term, with $m_i = 1$ for $1 \leq i \leq N$ and no noise added to the update equation. All of the variants mentioned in this section will be explored in a longer paper in the future.

As a final remark, note that the final locations of the proposal pdfs (i.e., their means, $\boldsymbol{\mu}_i^{(t)}$) can be used to estimate the positions of the modes of $\pi(\mathbf{x})$, and the final covariance matrices $\mathbf{C}_i^{(t)}$ could be used to estimate their dispersion.

4. EXPERIMENTS

In order to evaluate the performance of GAPIS, we have performed an experiment where the target is a bivariate multimodal target pdf formed by a mixture of 5 Gaussians, i.e.,

$$\pi(\mathbf{x}) = \frac{1}{5} \sum_{i=1}^{5} \mathcal{N}(\mathbf{x}; \nu_i, \boldsymbol{\Sigma}_i), \quad \mathbf{x} \in \mathbb{R}^2, \tag{12}$$

with means $\nu_1 = [-10, -10]^\top$, $\nu_2 = [0, 16]^\top$, $\nu_3 = [13, 8]^\top$, $\nu_4 = [-9, 7]^\top$, and $\nu_5 = [14, -14]^\top$, 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 $\boldsymbol{\Sigma}_5 = [2,\ -0.1; -0.1,\ 2]$. In this example, the moments of the target in (12) can be easily computed analytically, thus allowing us to check the performance of the different techniques. For the sake of simplicity, we also consider Gaussian proposals.

We have run the GAPIS algorithm with $N = 100$ Gaussian proposals, estimating the mean (with known value $\bar{\mathbf{x}} = [1.6, 1.4]^\top$) and the normalizing constant ($Z = 1$) of the target. The means of the proposals have been randomly positioned inside a "bad" area (free from any modes of the target) in order to test the robustness of the algorithm w.r.t. a bad initialization. In this way, the initial means of the proposals are selected uniformly within a square $\boldsymbol{\mu}_i^{(0)} \sim \mathcal{U}([-4, 4] \times [-4, 4])$. Initially, the same isotropic covariance matrix, $\mathbf{C}_i^{(0)} = \sigma^2 \mathbf{I}_2$, is used for every proposal. We test different values of $\sigma \in \{1, 5, 10\}$. Then, different non-isotropic diagonal covariance matrices, $\mathbf{C}_i^{(0)} = \text{diag}(\sigma_{i,1}^2, \sigma_{i,2}^2)$ with $\sigma_{i,j} \sim \mathcal{U}([1, 10])$ for $i = 1, \ldots N$ and $j \in \{1, 2\}$, are also tested.

We run the GAPIS algorithm with different number of iterations $T \in \{100, 400\}$. In order to perform a fair comparison, the total number of samples drawn in all the simulations is the same, i.e., we set $L = KNT = 2 \cdot 10^5$ samples and choose $K = L/(NT) \in \{20, 5\}$. Regarding the gradient step of (4) we set $\lambda = 1$, whereas the "masses" in (5) are set to

Algorithm	Spec. parameters		$\sigma = 1$	$\sigma = 5$	$\sigma = 10$	$\sigma_{i,j} \sim \mathcal{U}([1, 10])$
MIS	$K = 2000$	$T = 1$	41.95	2.17	0.0147	4.55
PIS	$K = 2000$	$T = 1$	47.74	0.2424	0.0124	0.0651
GAPIS	$K = 5$	$T = 400$	0.0024	**0.0022**	**0.0008**	**0.0012**
	$K = 20$	$T = 100$	0.0031	0.0068	0.0021	0.0050
GAPIS (fixed $\mathbf{C}_i$)	$K = 5$	$T = 400$	**0.0023**	0.0104	0.0178	0.0040
	$K = 20$	$T = 100$	0.0440	0.0103	0.0143	0.0048
APIS	Best [21]		2.45	0.2424	0.0185	0.0045
PMC	Best [21]		107.58	0.6731	0.0744	0.0732

Table 1. MSE in the estimation of the mean of the target (first component), for several values of the initial σ and different techniques, keeping fixed the total number of evaluations of the target $L = KNT = 2 \cdot 10^5$ (with $N = 100$ proposal pdfs). The best results for each value of σ are highlighted in bold-face.

$m_i = 1 \; \forall i$ and the interaction term G_t is a periodic function, as discussed in Subsection 3.2, with $\nu = 0.05$ and $\Delta = 1.04$.

We compare the performance of GAPIS with two static schemes, where no adaptation is performed (i.e., $T = 1$): a standard multiple importance sampler (MIS) and another static approach where the deterministic mixture is applied, so called population importance sampler (PIS) [21]. We also compare GAPIS with other state-of-the-art adaptive methods, such as APIS and PMC. The results displayed for APIS and PMC correspond to the choice of parameters that lead to the best performance, as shown in [21]. Finally, we also consider a GAPIS scheme where the location parameters of the proposal pdfs are adapted, but their scale parameters are fixed.

Table 1 shows the mean square error (MSE) in the estimation of the first component of the mean of the target. All the results are averaged over 1000 independent experiments. GAPIS largely outperforms the non-adaptive standard IS methods. The best results are obtained when the mean and the covariance matrices are updated with $K = 5$ and $T = 400$. Note that, as discussed in Section 3.2, the extra computational cost depends on the number of iterations T, since the means and the covariance matrices of the proposals are adapted at the beginning of each iteration. Table 1 also shows that GAPIS is robust w.r.t. the initial choice of the covariance of the proposals. Indeed, since both the mean vectors and the covariance matrices are adapted by GAPIS, after few iterations the proposals are always suitably placed and scaled regardless of the initialization.

Figure 1 shows the MSE evolution in the estimation of the first component of the target described above for GAPIS, PIS, and APIS (with the parameters that result in the best performance, as found in [21]). GAPIS has been run with $T = 100$ ($K = 20$) and $T = 400$ ($K = 5$), and the number of proposals is $N = 100$ in both cases. All the methods have been simulated with the same number of samples ($L = KNT = 2 \cdot 10^5$) and the MSE curves have been averaged over 1000 independent experiments. The results show that all the adaptive schemes outperform the static PIS. GAPIS works better in terms of MSE when more adaptations are performed ($T = 400$), outperforming APIS in this case, although at the expense of an increased computational cost per iteration. Finally, note that GAPIS also seems to exhibit a better asymptotic behavior than APIS or PIS.

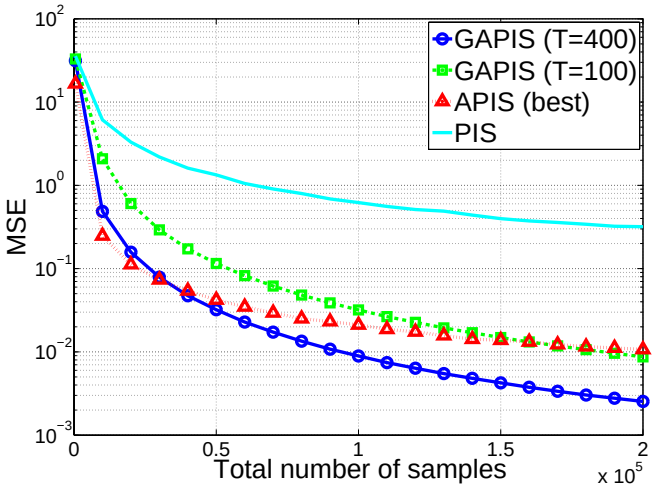

Fig. 1. MSE evolution in the estimation of the first component for GAPIS, APIS, and PIS methods.

5. CONCLUSIONS

In this work, we have presented a novel adaptive importance sampling method using jointly a population of proposal pdfs. The gradient adaptive population importance sampling (GAPIS) is an iterative importance sampler that dynamically optimizes the cloud of proposals by using the information of the gradient and Hessian matrix of the target distribution. Numerical results have shown the excellent performance of GAPIS compared to other adaptive importance samplers. GAPIS also exhibits an attractive robustness with respect to the choice of the algorithm's parameters, including the location and scale of the cloud of proposals.

6. 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] X. Wang, R. Chen, and J. S. Liu, "Monte Carlo Bayesian signal processing for wireless communications," *Journal of VLSI Signal Processing*, vol. 30, pp. 89–105, 2002.

[3] A. Doucet and X. Wang, "Monte Carlo methods for signal processing," *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 152–170, Nov. 2005.

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

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

[6] 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.

[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, December 2012.

[8] L. Martino, V. Elvira, D. Luengo, and J. Corander, "An adaptive population importance sampler," *Proc. of the 39th ICASSP, Florence, Italy*, May 4-9 2014.

[9] R. Srinivasan, *Importance sampling: Applications in communications and detection*, Springer, 2002.

[10] A. Wong, A. Mishra, P. Fieguth, and D. Clausi, "An adaptive Monte Carlo approach to nonlinear image denoising," in *Proc. 19th Int. Conf. on Pattern Recognition (ICPR)*, 2008, pp. 1–4.

[11] J.-C. Chen, "Partial transmit sequences for peak-to-average power ratio reduction of OFDM signals with the cross-entropy method," *IEEE Signal Processing Letters*, vol. 16, no. 6, pp. 545–548, June 2009.

[12] P. Minvielle, E. Tantar, A. Tantar, and P. Bérisset, "Sparse antenna array optimization with the cross-entropy method," *IEEE Transactions on Antennas and Propagation*, vol. 59, no. 8, pp. 2862–2871, Aug. 2011.

[13] J. Sirén, P. Marttinen, and J. Corander, "Reconstructing population histories from single nucleotide polymorphism data," *Molecular biology and evolution*, vol. 28, no. 1, pp. 673–683, 2011.

[14] R. Neal, *MCMC using Hamiltonian dynamics*, Appears as Chapter 5 of the Handbook of Markov Chain Monte Carlo Edited by Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng; Chapman and Hall CRC Press, 2011.

[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] K.-V. Yuen, *Bayesian methods for structural dynamics and civil engineering*, John Wiley & Sons, 2010.

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

[18] 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.

[19] H. J. Kushner and D. S. Clark, *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, Springer, 1978.

[20] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*, John Wiley & Sons, 2004.

[21] L. Martino, V. Elvira, D. Luengo, and J. Corander, "An adaptive population importance sampler: Learning from uncertainty," *viXra:1405.0280*, 2014.