SMELLY PARALLEL MCMC CHAINS

L. Martino[⋆]*, V. Elvira*[†]*, D. Luengo*[‡]*, A. Artés-Rodríguez*[†]*, J. 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 useful tools for Bayesian inference and stochastic optimization that have been widely applied in signal processing and machine learning. A well-known class of MC methods are Markov Chain Monte Carlo (MCMC) algorithms. In this work, we introduce a novel parallel interacting MCMC scheme, where the parallel chains share information, thus yielding a faster exploration of the state space. The interaction is carried out generating a dynamic repulsion among the "smelly" parallel chains that takes into account the entire population of current states. The ergodicity of the scheme and its relationship with other sampling methods are discussed. Numerical results show the advantages of the proposed approach in terms of mean square error, robustness w.r.t. to initial values and parameter choice.

Index Terms— Markov Chain Monte Carlo; parallel and interacting chains; Bayesian inference

1. INTRODUCTION

Monte Carlo (MC) methods are widely used in signal processing and communications [1, 2, 3]. Markov Chain Monte Carlo (MCMC) techniques [4] are well-known Monte Carlo methodologies to draw random samples and compute efficiently integrals involving a complicated multidimensional target probability density function (pdf), $\pi(\mathbf{x})$ with $\mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^n$. MCMC schemes only need to be able to evaluate the target pdf, but the difficulty of diagnosing and speeding up the convergence has motivated an intense research activity. For instance, several adaptive MCMC methods have been developed in order to adequately fix the parameters of the proposal density used to draw candidate samples [5, 6]. Nevertheless, guaranteeing the theoretical convergence is still an issue in most of the cases. In order to explore the state space faster (and specially to deal with high-dimensional applications), several schemes with parallel chains have been recently proposed [6, 7, 8, 9], as well as multiple try and interacting schemes [10], but the problem is still far from being solved.

In this work, we present a novel parallel MCMC chains scheme, the "smelly" parallel MCMC technique (SP-MCMC), where N different chains cooperate together in order to obtain a better exploration of the state space. The interaction is achieved by creating a repulsion among the chains during a fixed number of iterations. The motivation of this approach is trying to avoid that several chains remain trapped inside the same region of the state space (e.g., around a single mode of the target distribution), thus eliminating the risk of having a very slow convergence of the chains. Indeed, the starting point for the SP-MCMC algorithm comes from the fact that we can obtain a better mixing if each chain explores a different portion of the state space during the "burn-in" period. Therefore, in SP-MCMC a repulsion among the chains is produced during this "burn-in" period. The goal is making sure that each chain can smell the "stink" of the other chains, thus forcing it to move away and explore other regions of the state space. The strength of this repulsion should depend both on the locations of the chains (several chains could be allowed to coexist around regions of high probability of the target pdf) and the effective support of the proposal associated to each chain (as larger supports imply bigger jumps for the chain).

In this paper, we propose carrying out the interaction by building different modified target densities for each chain and iteration. Hence, the SP-MCMC scheme presents certain connections and similar advantages as the tempering approach [11, 12]. The resulting algorithm exhibits both flexibility and robustness w.r.t. initial values and parameter choice, as shown by the numerical results. The paper is structured as follows. Section 2 outlines the problem addressed. Then, Section 3 shows the generic version of the SP-MCMC, whereas Section 4 provides an example of a modified target. Section 5 discusses the relationship of the SP-MCMC to other approaches in the literature and Section 6 displays the numerical results The paper ends up with the conclusions in Section 7.

2. PROBLEM STATEMENT

In many applications, we are interested in inferring a variable of interest given a set of observations or measurements. Let

This work has been supported by the Spanish government (projects TEC2012-38058-C03-01, TEC2012-38800-C03-01, TEC2012-38883-C02-01, TEC2013-41718-R and S2010/BMD-2422); by the BBVA Foundation (MG-FIAR project); by the ERC grant 239784 and AoF grant 251170.

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

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

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

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

such that $\widetilde{\pi}(\mathbf{x}) \propto \pi(\mathbf{x})$. Our goal is computing efficiently some moment of $\mathbf{x}$, i.e., an integral measure w.r.t. the target,

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

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. GENERIC SMELLY PARALLEL MCMC

3.1. Outline of the Smelly Parallel MCMC Algorithm

Consider N parallel chains,

$$\{\mathbf{x}_{i,t}\}_{t=0}^{\infty} \quad \text{with} \quad i = 1, ..., N,$$

generated by several MCMC schemes using different proposal pdfs, $q_i(\mathbf{x}|\mathbf{x}_{i,t})$. In the standard parallel scheme the chains are independent [13], typically considering the same target pdf $\widetilde{\pi}(\mathbf{x})$ and converging to it as $t \to \infty$. In this work, we incorporate an *interaction* among the chains in such a way that the chains cooperate to perform a better exploration of the state space. More specifically, this interaction is carried out in terms of a *repulsion* among the chains.

From the sampling point of view, we introduce this repulsion by considering a modified target density for each chain and iteration. These modified target densities should be constructed taking into account both the true target pdf and the proposals for all the chains. Furthermore, when one chain is far away from the remaining chains (i.e., almost no repulsion occurs) the modified target should be almost equal to the true target. For the sake of simplicity, let us assume that the modified target for the i-th chain at the t-th iteration is constructed taking into account only the current states of the remaining chains, i.e.,[1]

$$\widetilde{\varphi}_{i,t}(\mathbf{x}) \propto \varphi_{i,t}(\mathbf{x}) = \phi_i(\mathbf{x}|\mathbf{P}_{\neg i,t}), \tag{4}$$

where

$$\mathbf{P}_{\neg i,t} = [\mathbf{x}_{1,t}, \ldots, \mathbf{x}_{i-1,t}, \mathbf{x}_{i+1,t} \ldots, \mathbf{x}_{N,t}]. \tag{5}$$

[1] Let us remark that part or even the whole set of past states of the chains could be used to build $\widetilde{\varphi}_{i,t}(\mathbf{x}) \propto \varphi_{i,t}(\mathbf{x}) = \phi_i(\mathbf{x}|\mathbf{P}_{\neg i,0:t})$.

Table 1. Generic Smelly Parallel (SP-MCMC) method.

Initialization:

1. Choose the N initial points,

$$\{\mathbf{x}_{1,0}, \mathbf{x}_{2,0}, \ldots, \mathbf{x}_{N,0}\},$$

and the total number of iterations T.

For $t = 0, \ldots, T-1$:
For $i = 1, \ldots, N$ (in parallel):

2. Draw $\mathbf{z} \sim q_i(\mathbf{x}|\mathbf{x}_{i,t})$.

3. Set $\mathbf{x}_{i,t+1} = \mathbf{z}$ with probability

$$\alpha = \min\left[1, \frac{\varphi_{i,t}(\mathbf{z})q_i(\mathbf{x}_{i,t}|\mathbf{z})}{\varphi_{i,t}(\mathbf{x}_{i,t})q_i(\mathbf{z}|\mathbf{x}_{i,t})}\right].$$

Otherwise, set $\mathbf{x}_{i,t+1} = \mathbf{x}_{i,t}$.

4. Decrease the interaction among the chains, so that each function, $\varphi_{i,t+1}(\mathbf{x})$ for $i \in \{1, \ldots, N\}$, becomes closer to $\pi(\mathbf{x})$. Namely, defining

$$D_t = \int_{\mathbb{R}^n} |\varphi_{i,t}(\mathbf{x}) - \pi(\mathbf{x})|d\mathbf{x}, \tag{6}$$

update $\varphi_{i,t}$ in such a way that $D_{t+1} \leq D_t$, with $D_t \to 0$ as $t \to \infty$.

The generic SP-MCMC algorithm is described in Table 1 using Metropolis-Hastings (MH) kernels, although any other valid MCMC kernel could be easily used as the basic building block. Observe that, given the cloud $\{\mathbf{x}_{i,t}\}$ at the t-th iteration, the N chains can be moved forward in parallel to the iteration $t+1$. Note also that we could consider adaptive proposal pdfs $q_{i,t}$, as in adaptive MCMC approaches [5, 14, 15]. In this case, $q_{i,t}$ could be updated using all the past samples instead of just the current state $\mathbf{x}_{i,t}$.

3.2. Ergodicity

The convergence of the chains to the true target, $\widetilde{\pi}(\mathbf{x}) \propto \pi(\mathbf{x})$, can only be ensured if

$$D_t = 0, \qquad \forall t \geq \tau, \tag{7}$$

for some where iteration $\tau < T < \infty$.[2] In this case, for $t \geq \tau$ we have N independent parallel MCMC chains converging to the true target. Thus, all the previous samples for $t < \tau$

[2] Note that this is a sufficient condition to guarantee the ergodicity. However, by making $D_t \to 0$ fast enough as $t \to \infty$ the ergodicity could also be ensured. Both necessary and sufficient conditions for attaining ergodicity will be considered in future works.

should be discarded for the final estimation. In this work, the underlying idea is to take advantage of the transient evolution of a chain (the so-called "burn-in" period) and apply the interaction in these first steps of the algorithm (note again that the samples in the "burn-in" period have to be discarded anyway). Unfortunately, the length of the "burn-in" period is unknown, so the instant τ should be set by the user based on a-priori information. Alternatively, considering the approach with parallel chains, the length "burn-in" period could be estimated as suggested in [13]. The same procedure has been applied in [6] for adapting the proposal's parameters.

3.3. Alternative scheme

So far, fixing an index i, we have considered that the function $\varphi_{i,t}(\mathbf{x}) = \phi_i(\mathbf{x}|\mathbf{P}_{\neg i,t})$ depends on the population

$$\mathbf{P}_{\neg i,t} = [\mathbf{x}_{1,t}, \ldots, \mathbf{x}_{i-1,t}, \mathbf{x}_{i+1,t} \ldots, \mathbf{x}_{N,t}],$$

so that all the chains can be moved forward simultaneously to the iteration $t+1$. However, in a Gibbs sampling fashion, it is possible to consider a sequential update of the population in $\mathbf{P}_{\neg i,t}$. Namely, we could consider

$$\mathbf{P}_{\neg i,t} = [\mathbf{x}_{1,t+1}, \ldots, \mathbf{x}_{i-1,t+1}, \mathbf{x}_{i+1,t} \ldots, \mathbf{x}_{N,t}], \quad (8)$$

where $\mathbf{x}_{1,t+1}, \ldots, \mathbf{x}_{i-1,t+1}$ have already been updated for the previous $i-1$ chains. This scheme could perform better than the previous one, but it loses the advantage of the parallel computation due to the sequential update of the chains.[3]

4. MODIFIED TARGET PDFS

In this section, we provide an example of possible functions $\varphi_{i,t}$ taking into account both the true target π and the current position of the states $\{\mathbf{x}_{i,t}\}_{t=0}^{\infty}$. For instance, we consider

$$\widetilde{\varphi}_{i,t}(\mathbf{x}) \propto \varphi_{i,t}(\mathbf{x}) = \frac{\pi(\mathbf{x})}{\left(\frac{1}{N-1}\sum\limits_{\substack{j=1\\ j\neq i}}^{N} q_j(\mathbf{x}|\mathbf{x}_{j,t})\right)^{\gamma_t}}. \quad (9)$$

where $\gamma_t \geq 0$ determines the strength of the repulsion. Denoting $V(\mathbf{x}) = \log[\pi(\mathbf{x})]$ and

$$W_t(\mathbf{x}) = \log\left[\frac{1}{N-1}\sum_{\substack{j=1\\ j\neq i}}^{N} q_i(\mathbf{x}|\mathbf{x}_{j,t})\right],$$

then we can simply write $\varphi_{i,t}(\mathbf{x})$ as

$$\widetilde{\varphi}_{i,t}(\mathbf{x}) \propto \varphi_{i,t}(\mathbf{x}) = \exp\left(V(\mathbf{x}) - \gamma_t W_t(\mathbf{x})\right). \quad (10)$$

[3]Note that the ergodicity is also ensured for this alternative scheme as long as (7) is fulfilled.

We remark again that we have a different function $\varphi_{i,t}$ for each chain ($1 \leq i \leq N$) and iteration $t \in \mathbb{N}^+$. In this case, the modified target associated to each chain is directly proportional to the true target and inversely proportional to the probability mass generated by the remaining proposal pdfs at the t-th iteration. This effect yields a *repulsion* among the chains according to the "intensity of the smell" of the remaining chains. The underlying idea is that there is no reason for the rest of the chains to explore the portion of the state space easily reachable by the i-th chain during the "burn-in" period. The effective support (i.e., the range) of the i-th chain is then taken to be proportional to the probability mass of the i-th proposal pdf. In practice, this implies that larger variances or heavier tails correspond to greater ranges, as expected. Finally, let us remark that in order to obtain a suitable and feasible approach, we need to fulfill two conditions:

- The parameter γ_t has to decrease as t increases:

  $$\gamma_1 \geq \ldots \geq \gamma_t \geq \gamma_{t+1} \geq \ldots \geq \gamma_\tau = 0,$$

  with $\gamma_{\tau'} = 0$ for $\tau' > \tau$, so that $\varphi_{i,t}(\mathbf{x}) = \pi(\mathbf{x})$ for all $i \in \{1, ..., N\}$ and $t \geq \tau$.

- The proposal pdfs, q_i, must have heavier tails than the target, π, so that $\int_{\mathcal{X}} \varphi_{i,t}(\mathbf{x})d\mathbf{x} < +\infty$.

Figure 1 shows an example of a modified target pdf for a univariate and unimodal case.

4.1. Dynamics of the modified target pdfs

On the one hand, note that the variation of the i-th modified target $\varphi_{i,t}$ depends on the evolution of the states of the other chains. On the other hand, the evolution of the chains depends on the variation of the targets $\varphi_{i,t}$. Thus, it is not straightforward to analyze the dynamics of this system of moving states $\{\mathbf{x}_{i,t}\}$ for $t < \tau$, as well as to study the stationary pdf for each chain if $\tau \to \infty$. However, let us remark that an infinite divergence among the chains $\{\mathbf{x}_{i,t}\}$ is not possible. Indeed, when the states are far from each other the intensity of the repulsion

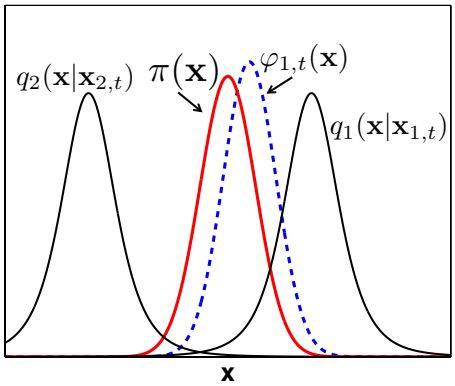

Fig. 1. Example of a modified target for a univariate case.

		SP-MCMC						Independent parallel chains (IPCs)					
N	**In**	$\sigma = 1.1$	$\sigma = 2$	$\sigma = 5$	$\sigma = 10$	$\sigma = 20$	**Avg**	$\sigma = 1.1$	$\sigma = 2$	$\sigma = 5$	$\sigma = 10$	$\sigma = 20$	**Avg**
20	**In1**	5.92	5.59	5.27	5.44	5.34	**5.51**	8.65	8.48	7.66	7.50	7.96	**8.05**
100	**In1**	2.39	1.50	1.37	1.23	1.25	**1.54**	5.09	4.33	3.80	3.88	3.99	**4.21**
20	**In2**	7.18	7.13	5.78	7.39	6.17	**6.73**	6.93	7.27	6.55	6.11	6.19	**6.63**
100	**In2**	2.22	1.95	1.86	2.03	1.79	**1.97**	2.02	1.90	1.81	1.93	1.56	**1.84**

Table 2. Mean square error (MSE) in the estimation of the mean of the target (averaging the MSEs of the two components), for different values of the standard deviation of the proposal pdfs (σ).

vanishes to zero, and all the modified targets become the true target. Hence, all the chains would be attracted towards the modes of the true target, since the SP-MCMC approach becomes a standard parallel MH chains technique when the interaction is removed. For this reason, we set $\tau < \infty$ to ensure that the system converges to the true target after the initial stage, where repulsion is useful to enhance the exploratory behaviour of the parallel chains.

5. RELATIONSHIP WITH OTHER TECHNIQUES

Unlike the adaptive MCMC schemes, in the SP-MCMC algorithm the target is changing over the time, whereas the proposal pdfs are not adapted considering all the past samples. In this sense, the SP-MCMC scheme is more similar to the *tempering* approaches, where the scale of target is changed with the time index [11]. Initially, a large scale is considered using an auxiliary parameter. Then, the scale is progressively reduced in order to end up drawing samples from the true target density $\widetilde{\pi}(x)$. This idea is often used to help the sampling schemes in dealing with very sharp and narrow target pdfs (e.g., those typically arising when analyzing huge amounts of data). In the SP-MCMC approach, the complete shape of the target is changed, not just the scale parameter. However, considering Eq. (9), we notice that we also obtain a similar effect of spreading the probability mass of the target. This is due to the fact that the variances and the probability masses of the proposal pdfs are involved in building the modified targets.

6. NUMERICAL SIMULATIONS

In order to test the SP-MCMC scheme, we consider a bivariate multimodal target pdf, which is a mixture of 5 Gaussians:

$$\pi(\mathbf{x}) = \frac{1}{5}\sum_{i=1}^{5}\mathcal{N}(\mathbf{x};\boldsymbol{\nu}_i,\boldsymbol{\Sigma}_i), \tag{11}$$

with $\mathbf{x} \in \mathbb{R}^2$; means $\boldsymbol{\nu}_1 = [-10,-10]^\top$, $\boldsymbol{\nu}_2 = [0,16]^\top$, $\boldsymbol{\nu}_3 = [13,8]^\top$, $\boldsymbol{\nu}_4 = [-9,7]^\top$, and $\boldsymbol{\nu}_5 = [14,-14]^\top$; 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 apply the smelly parallel approach (SP-MCMC) and a standard scheme with independent parallel chains (IPCs) to estimate the mean of the target ($[1.6, 1.4]^\top$) using different values for the number of parallel chains, $N \in \{20, 100\}$. Furthermore, in order to test the robustness of the algorithm we choose deliberately a "bad" initialization, $\boldsymbol{x}_{i,0} \sim \mathcal{U}([-4,4]\times[-4,4])$ (denoted as **In1**), and a "better" initialization, $\boldsymbol{x}_{i,0} \sim \mathcal{U}([-20,20]\times[-20,20])$ (denoted as **In2**) for $i = 1, \ldots, N$. Moreover, defining $\mathbf{x} = [x^{(1)}, x^{(2)}]^\top$ and $\mathbf{x}_{i,t} = [x_{i,t}^{(1)}, x_{i,t}^{(2)}]^\top$, we consider t-Student proposal pdfs,

$$q_i(\mathbf{x}|\mathbf{x}_{i,t}) \propto \prod_{j=1}^{2}\left(1 - \frac{1}{\nu}\left(x^{(j)} - x_{i,t-1}^{(j)}\right)^2\right)^{\frac{\nu-1}{2}},$$

and test different degrees of freedom, ν, in order to obtain different standard deviations $\sigma = \sqrt{\frac{\nu}{\nu-2}} \in \{1.1, 2, 5, 10, 20\}$, to gauge the performance of SP-MCMC. We set $T = 1000$ and $\tau = 100$ for all the simulations and use all the generated samples without removing any "burn-in" period (in order to obtain a fair comparison with IPCs). This means that for SP-MCMC we also use the samples $\{\mathbf{x}_{i,t}\}$ for $t < \tau$ in the estimation. We set $\gamma_t = 400$ for $t < \tau$ and $\gamma_t = 0$ for $t \geq \tau$.

Table 2 shows the mean square error (MSE) in the estimation of the first component of the mean averaged over 500 independent runs. Note the better performance of SP-MCMC when compared to the IPCs approach, as well as its increased robustness w.r.t. the initialization. Indeed, for the "bad" initialization the SP-MCMC approach succeeds in decreasing the average MSE from 8.05 to 5.51 for $N = 20$ (roughly 68.4 % of the IPCs average MSE) and from 4.21 to 1.54 for $N = 100$ (around 36.5 % of the IPCs average MSE). In contrast, for the "better" initialization there is an almost negligible loss (around 1.5 % and 7.0 % for $N = 20$ and $N = 100$ respectively), caused by the repulsion.

7. CONCLUSIONS

We have introduced a novel family of parallel Markov chain Monte Carlo (MCMC) algorithms that incorporate interaction among the population of parallel MCMC chains. The smelly parallel MCMC (SP-MCMC) chains share information, thus yielding a faster exploration of the state space. Compared to the fully independent parallel chains approach, the novel technique shows a more robust behaviour w.r.t. the parameterization and the choice of the initialization.

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

[5] D. Luengo and L. Martino, "Fully adaptive Gaussian mixture Metropolis-Hastings algorithm," in *Proc. 38th IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Vancouver (Canada), 2013, pp. 6148–6152.

[6] R. Craiu, J. Rosenthal, and C. Yang, "Learn from thy neighbor: Parallel-chain and regional adaptive MCMC," *Journal of the American Statistical Association*, vol. 104, no. 448, pp. 1454–1466, 2009.

[7] J. Ye, A. Wallace, and J. Thompson, "Parallel Markov chain Monte Carlo computation for varying-dimension signal analysis," in *Proc. 17th European Signal Processing Conf. (EUSIPCO)*, Glasgow (Scotland), 24–28 Aug. 2009, pp. 2673–2677.

[8] L. Martino, V. Elvira, D. Luengo, A. Artes-Rodriguez, and J. Corander, "Orthogonal MCMC algorithms," *IEEE Workshop on Statistical Signal Processing (SSP)*, pp. 364–367, June 2014.

[9] J. Corander, M. Ekdahl, and T. Koski, "Parallel interacting MCMC for learning of topologies of graphical models," *Data Mining and Knowledge Discovery*, vol. 17, no. 3, pp. 431–456, December 2008.

[10] R. Casarin, R. V. Craiu, and F. Leisen, "Interacting multiple try algorithms with different proposal distributions," *Statistics and Computing*, vol. 23, no. 2, pp. 185–200, 2013.

[11] C. J. Geyer and E. A. Thompson, "Annealing Markov chain Monte Carlo with applications to ancestral inference," *Journal of the American Statistical Association*, vol. 90, no. 431, pp. 909–920, 1995.

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

[13] A. Gelman and D. B. Rubin, "Inference from iterative simulation using multiple sequences," *Statistical Science*, vol. 7, no. 4, pp. 457–511, February 1992.

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

[15] L. Martino, J. Read, and D. Luengo, "Independent doubly adaptive rejection Metropolis sampling," in *Proc. 39th IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Florence (Italy), 4–9 May 2014.