

# Interacting Multiple Try Algorithms with Different Proposal Distributions

Roberto Casarin<sup>¶</sup> Radu V. Craiu<sup>‡</sup> Fabrizio Leisen<sup>§</sup>

<sup>¶</sup>Advanced School of Economics, Venice

<sup>†</sup>University of Brescia

<sup>§</sup>Universidad Carlos III de Madrid

<sup>‡</sup>University of Toronto

November 5, 2010

## Abstract

We propose a new class of interacting Markov chain Monte Carlo (MCMC) algorithms designed for increasing the efficiency of a modified multiple-try Metropolis (MTM) algorithm. The extension with respect to the existing MCMC literature is twofold. The sampler proposed extends the basic MTM algorithm by allowing different proposal distributions in the multiple-try generation step. We exploit the structure of the MTM algorithm with different proposal distributions to naturally introduce an interacting MTM mechanism (IMTM) that expands the class of population Monte Carlo methods. We show the validity of the algorithm and discuss the choice of the selection weights and of the different proposals. We provide numerical studies which show that the new algorithm can perform better than the basic MTM algorithm and that the interaction mechanism allows the IMTM to efficiently explore the state space.

## 1 Introduction

Markov chain Monte Carlo (MCMC) algorithms are now essential for the analysis of complex statistical models. In the MCMC universe, one of the most widely used class of algorithms is defined by the Metropolis-Hastings (MH) (Metropolis et al., 1953; Hastings, 1970) and its variants. An important generalization of the standard MH formulation is represented by the

multiple-try Metropolis (MTM) (Liu et al., 2000). While in the MH formulation one accepts or rejects a single proposed move, the MTM is designed so that the next state of the chain is selected among multiple proposals. The multiple-proposal setup can be used effectively to explore the sample space of the target distribution and subsequent developments made use of this added flexibility. For instance, Craiu and Lemieux (2007) propose to use antithetic and quasi-Monte Carlo samples to generate the proposals and to improve the efficiency of the algorithm while Pandolfi et al. (2010b) and Pandolfi et al. (2010a) apply the multiple-proposal idea to a trans-dimensional setup and combine Reversible Jump MCMC with MTM.

This work further generalizes the MTM algorithm presented in Liu et al. (2000) in two directions. First, we show that the original MTM transition kernel can be modified to allow for different proposal distributions in the multiple-try generation step while preserving the ergodicity of the chain. The use of different proposal distributions gives more freedom in designing MTM algorithms for target distributions that require different proposals across the sample space. An important challenge remains the choice of the distributions used to generate the proposals and we propose to address it by expanding upon methods used within the population Monte Carlo class of algorithms.

The class of population Monte Carlo procedures (Cappé et al., 2004; Del Moral and Miclo, 2000; Del Moral, 2004; Jasra et al., 2007) has been designed to address the inefficiency of classical MCMC samplers in complex applications involving multimodal and high dimensional target distributions (Pritchard et al., 2000; Heard et al., 2006). Its formulation relies on a number of MCMC processes that are run in parallel while learning from one another about the geography of the target distribution.

A second contribution of the paper is finding reliable generic methods for constructing the proposal distributions for the MTM algorithm. We propose an interacting MCMC sampling design for the MTM that preserves the Markovian property. More specifically, in the proposed interacting MTM (IMTM) algorithm, we allow the distinct proposal distributions to use information produced by a population of auxiliary chains. We infer that the resulting performance of the MTM is tightly connected to the performance of the chains' population. In order to maximize the latter, we propose a number of strategies that can be used to tune the auxiliary chains. We also adapt previous extensions of the MTM and link the use of stochastic overrelaxation, random-ray Monte Carlo method (see Liu et al., 2000) and simulated annealing to IMTM.

In the next section we discuss the IMTM algorithm, propose a number

of alternative implementations and prove their ergodicity. In Section 3 we focus on some special cases of the IMTM algorithm and in Section 4 the performance of the methods proposed is demonstrated with simulations and real examples. We end the paper with a discussion of future directions for research.

## 2 Interacting Monte Carlo Chains for MTM

We begin by describing the MTM and its extension for using different proposal distributions.

### 2.1 Multiple-Try Metropolis With Different Proposal Distributions

Suppose that of interest is sampling from a distribution  $\pi$  that has support in  $\mathcal{Y} \subset \mathbf{R}^d$  and is known up to a normalizing constant. Assuming that the current state of the chain is  $x$ , the update defined by the MTM algorithm of Liu et al. (2000) is described in Algorithm 1.

---

**Algorithm 1.** *Multiple-try Metropolis Algorithm (MTM)*

---

1. Draw  $M$  trial proposals  $y_1, \dots, y_M$  from the proposal distribution  $T(\cdot|x)$ . Compute  $w(y_j, x)$  for each  $j \in \{1, \dots, M\}$ , where  $w(y, x) = \pi(y)T(x|y)\lambda(y, x)$ , and  $\lambda(y, x)$  is a symmetric function of  $x, y$ .
2. Select  $y$  among the  $M$  proposals with probability proportional to  $w(y_j, x), j = 1, \dots, M$ .
3. Draw  $x_1^*, \dots, x_{M-1}^*$  variates from the distribution  $T(\cdot|y)$  and let  $x_M^* = x$ .
4. Accept  $y$  with generalized acceptance probability

$$\rho = \min \left\{ 1, \frac{w(y_1, x) + \dots + w(y_M, x)}{w(x_1^*, y) + \dots + w(x_M^*, y)} \right\}.$$

---

Note that while the MTM uses the same distribution to generate all the proposals, it is possible to extend this formulation to different proposal distributions without altering the ergodicity of the associated Markov chain.

Let  $T_j(x, \cdot)$ , with  $j = 1, \dots, M$ , be a set of proposal distributions for which  $T_j(x, y) > 0$  if and only if  $T_j(y, x) > 0$ . Define

$$w_j(x, y) = \pi(x)T_j(x, y)\lambda_j(x, y) \quad j = 1, \dots, M$$

where  $\lambda_j(x, y)$  is a nonnegative symmetric function in  $x$  and  $y$  that can be chosen by the user. The only requirement is that  $\lambda_j(x, y) > 0$  whenever  $T(x, y) > 0$ . Then the MTM algorithm with different proposal distributions is given in Algorithm 2.

---

**Algorithm 2.** *MTM with Different Proposal Distributions*

---

1. Draw independently  $M$  proposals  $y_1, \dots, y_M$  such that  $y_j \sim T_j(x, \cdot)$ . Compute  $w_j(y_j, x)$  for  $j = 1, \dots, M$ .
2. Select  $Y = y$  among the trial set  $\{y_1, \dots, y_M\}$  with probability proportional to  $w_j(y_j, x)$ ,  $j = 1, \dots, M$ . Let  $J$  be the index of the selected proposal. Then draw  $x_j^* \sim T_j(y, \cdot)$ ,  $j \neq J$ ,  $j = 1, \dots, M$  and let  $x_J^* = x$ .
3. Accept  $y$  with probability

$$\rho = \min \left\{ 1, \frac{w_1(y_1, x) + \dots + w_M(y_M, x)}{w_1(x_1^*, y) + \dots + w_M(x_M^*, y)} \right\}$$

and reject with probability  $1 - \rho$ .

---

It should be noted that Algorithm 2 is a special case of the interacting MTM presented in the next section and that the proof of ergodicity for the associated chain follows closely the proof given in Appendix A for the interacting MTM and is not given here.

In Section 4 we will show, through simulation experiments, that this algorithm is more efficient than a MTM algorithm with a single proposal distribution.

## 2.2 General Construction

Undoubtedly, Algorithm 2 offers additional flexibility in organizing the MTM sampler. This section introduces generic methods for using a population of MCMC chains to define the proposal distributions.

---

**Algorithm 3.** *Interacting Multiple Try Algorithm (IMTM)*

---

- For  $i = 1, \dots, N$

1. Let  $x = x_n^{(i)}$ , for  $j = 1, \dots, M_i$  draw  $y_j \sim T_j^{(i)}(\cdot | x_n^{(1:i-1)}, x, x_n^{(i+1:N)})$  independently and compute

$$w_j^{(i)}(y_j, x) = \pi(y_j) T_j^{(i)}(y_j | x_n^{(1:i-1)}, x, x_n^{(i+1:N)}) \lambda_j^{(i)}(y_j, x).$$

2. Select  $J \in \{1, \dots, M_i\}$  with probability proportional to  $w_j^{(i)}(y_j, x)$ ,  $j = 1, \dots, M_i$  and set  $y = y_J$ .
3. For  $j = 1, \dots, M_i$  and  $j \neq J$  draw  $x_j^* \sim T_j^{(i)}(\cdot | x_n^{(1:i-1)}, y, x_n^{(i+1:N)})$ , let  $x_J^* = x_n^{(i)}$  and compute

$$w_j^{(i)}(x_j^*, y) = \pi(x_j^*) T_j^{(i)}(x_j^* | x_n^{(1:i-1)}, y, x_n^{(i+1:N)}) \lambda_j^{(i)}(x_j^*, y).$$

4. Set  $x_{n+1}^{(i)} = y$  with probability

$$\rho_i = \min \left\{ 1, \frac{w_1^{(i)}(y_1, x) + \dots + w_{M_i}^{(i)}(y_{M_i}, x)}{w_1^{(i)}(x_1^*, y) + \dots + w_{M_i}^{(i)}(x_{M_i}^*, y)} \right\}$$

and  $x_{n+1}^{(i)} = x_n^{(i)}$  with probability  $1 - \rho_i$ .

---

Consider a population of  $N$  chains,  $X^{(i)} = \{x_n^{(i)}\}_{n \in \mathbb{N}}$  and  $i = 1, \dots, N$ . For full generality we assume that the  $i$ th chain has MTM transition kernel with  $M_i$  different proposals  $\{T_j^{(i)}\}_{1 \leq j \leq M_i}$  (if we set  $M_i = 1$  we imply that the chain has a MH transition kernel). The interacting mechanism allows each proposal distribution to possibly depend on the values of the chains at the previous step. Formally, if  $\Xi_n = \{x_n^{(i)}\}_{i=1}^N$  is the vector of values taken at iteration  $n \in \mathbb{N}$  by the population of chains, then we allow each proposal distribution used in updating the population at iteration  $n+1$  to depend on  $\Xi_n$ . The mathematical formalization is used in the description of Algorithm 3. One expects that the chains in the population are spread throughout the sample space and thus the proposals generated are a good representation of the sample space  $\mathcal{Y}$  ultimately resulting in better mixing for the chain of

interest.

In order to give a representation of the IMTM transition density let us introduce the following notation. Let  $T_j^{(i)}(x, y) = T_j^{(i)}(y|x_n^{(1:i-1)}, x, x_n^{(i+1:N)})$ ,  $T^{(i)}(x, y_{1:M_i}) = \prod_{k=1}^{M_i} T_k^{(i)}(x, y_k)$  and  $T_{-j}^{(i)}(x, y_{1:M_i}) = \prod_{k \neq j}^{M_i} T_k^{(i)}(x, y_k)$  and define  $dy_{1:M_i} = \prod_{k=1}^{M_i} dy_k$  and  $dy_{-j} = \prod_{k \neq j}^{M_i} dy_k$ .

The transition density associated to the population of chains is then

$$K(\Xi_n, \Xi_{n+1}) = \prod_{i=1}^N K_i(x_n^{(i)}, x_{n+1}^{(i)}) \quad (1)$$

where

$$K_i(x, y) = \sum_{j=1}^{M_i} A_j^{(i)}(x, y) T_j^{(i)}(x, y) + \left( 1 - \sum_{j=1}^{M_i} B_j^{(i)}(x) \right) \delta_x(y) \quad (2)$$

is the transition kernel associated to the  $i$ -th chain of algorithm with

$$A_j^{(i)}(x, y) = \int_{\mathcal{Y}^{2(M_i-1)}} \tilde{w}_j^{(i)}(y, x) \rho_j^{(i)}(x, y) T_{-j}^{(i)}(y, x_{1:M_i}^*) T_{-j}^{(i)}(x, y_{1:M_i}) dx_{-j}^* dy_{-j}$$

and

$$B_j^{(i)}(x) = \int_{\mathcal{Y}^{2(M_i-1)+1}} \rho_j^{(i)}(x, y) T_{-j}^{(i)}(y, dx_{1:M_i}^*) T^{(i)}(x, dy_{1:M_i}) dx_{-j}^* dy_{1:M_i}.$$

In the above equations  $\tilde{w}_j^{(i)}(y_j, x) = w_j^{(i)}(y_j, x) / (w_j^{(i)}(y_j, x) + \bar{w}_{-k}^{(i)}(y_{1:M_i}|x))$ , with  $j = 1, \dots, M_i$  and  $\bar{w}_{-j}^{(i)}(y_{1:M_i}|x) = \sum_{k \neq j}^{M_i} w_k^{(i)}(y_k, x)$ , are the normalized weights used in the selection step of the IMTM algorithm and

$$\rho_j^{(i)}(x, y) = \min \left\{ 1, \frac{w_j^{(i)}(y, x) + \bar{w}_{-j}^{(i)}(y_{1:M_i}|x)}{w_j^{(i)}(x, y) + \bar{w}_{-j}^{(i)}(x_{1:M_i}^*|y)} \right\}$$

is the generalized MH ratio associated to a MTM algorithm.

The validity of Algorithm 3 relies upon the detailed balance condition.

**Theorem 1.** *The transition density  $K_i(x_n^{(i)}, x_{n+1}^{(i)})$  associated to the  $i$ -th chain of the IMTM algorithm satisfies the conditional detailed balanced condition.*

**Proof** See Appendix A.

Since the transition  $K_i(x_n^{(i)}, x_{n+1}^{(i)})$ ,  $i = 1, \dots, N$  has  $\pi(x)$  as stationary distribution and satisfies the conditional detailed balance condition then the joint transition  $K(\Xi_n, \Xi_{n+1})$  has  $\pi(x)^N$  as a stationary distribution.

An important issue directly connected to the practical implementation of the IMTM is the choice of proposal distributions and the choice of  $\lambda_j^{(i)}(x, y)$ . First it should be noted that at each iteration of the interacting chains the computational complexity of the algorithm is  $\mathcal{O}(N \sum_{i=1}^N M_i)$ . When considering the number of chains and the number of proposals, there are two possible strategies in designing the interaction mechanism.

The first strategy is to use a small number of chains, say  $2 \leq N \leq 5$ , in order to improve the mixing of each chain and to allow for large jumps between different regions of the state space. When applying this strategy to our IMTM algorithms it is possible to set the number of proposals to be equal to the number of chains, i.e.  $M_i = N$ , for all  $1 \leq i \leq N$ . In this way all the chains can interact at each iteration of the algorithm and many search directions can be included among the proposals.

A second strategy is to use a higher number of chains, e.g.  $N = 100$ , in order to possibly have, at each iteration, a good approximation of the target or a much higher number of search directions for a good exploration of the sample space. This algorithm design strategy is common in Population Monte Carlo or Interacting MCMC methods. Clearly when a high number of chains is used within IMTM, it is necessary to set  $M_i < N$ . In the next section we discuss a few strategies to build the  $M_i$  proposals.

### 2.3 Parsing the Population of Auxiliary Chains

One of the strategies that revealed to be successful in our applications consists in the random selection of a certain number of chains of the population in order to build the proposals. More specifically, we let  $M_i = M$ , for all  $i$ , and when updating the  $i$ -th chain of the population we sample  $I_1, \dots, I_M$  random indexes from the uniform distribution  $\mathcal{U}_{\{1, \dots, N\}}$ , with  $N > M$ , and then set the proposals:  $T_j^{(i)}(y, x) = T_j^{(i)}(\cdot | x, x_n^{(I_1)}, \dots, x_n^{(I_M)})$ , for all  $j = 1, \dots, M$ . On the basis of our simulation experiments we found that the following choice  $T_j^{(i)}(\cdot | x_n^{(I_1)}, \dots, x_n^{(I_M)}) = T_j(\cdot | x_n^{(I_j)})$  works well in improving the mixing of the chains.

Previously suggested forms for the function  $\lambda_j^{(i)}(x, y)$  (Liu et al., 2000) are:

- a)  $\lambda_j^{(i)}(x, y) = 1$
- b)  $\lambda_j^{(i)}(x, y) = 2\{T_j^{(i)}(x, y) + T_j^{(i)}(y, x)\}^{-1}$
- c)  $\lambda_j^{(i)}(x, y) = \{T_j^{(i)}(x, y)T_j^{(i)}(y, x)\}^{-\alpha}$ ,  $\alpha > 0$ .

Here we propose to include in the choice of  $\lambda$  the information provided by the population of chains. Therefore we suggest to modify the above functions as follows

- a')  $\lambda_j^{(i)}(x, y) = \nu_j$
- b')  $\lambda_j^{(i)}(x, y) = 2\nu_j \{T_j^{(i)}(x, y) + T_j^{(i)}(y, x)\}^{-1}$
- c')  $\lambda_j^{(i)}(x, y) = \nu_j \{T_j^{(i)}(x, y)T_j^{(i)}(y, x)\}^{-\alpha}$ ,  $\alpha > 0$

where the factor  $\nu_j$  captures the behaviour of the auxiliary chains at the previous iteration

$$\nu_j = \frac{1}{N} \sum_{i=1}^N \mathbb{I}_{\{j\}}(J_{n-1}^{(i)}), \quad j = 1, \dots, M$$

where  $J_{n-1}^{(i)}$  is the random index of the selection step at the iteration  $n-1$  for the  $i$ -th chain. The modifications proposed for  $\lambda(\cdot, \cdot)$  would increase the use of those proposal distributions favoured by the population of chains at previous iteration. Since  $\nu_j$  depends only on samples generated at the previous step by the population of chains, the ergodicity of the IMTM chain is preserved. An alternative strategy is to sample the random indexes  $I_1, \dots, I_M$  with probabilities proportional to  $\nu_j$ .

## 2.4 Annealed IMTM

Our belief in IMTM's improved performance is underpinned by the assumption that the population of Monte Carlo chains is spread throughout the sample space. This can be partly achieved by initializing the chains using draws from a distribution overdispersed with respect to  $\pi$  (see also Jennison, 1993; Gelman and Rubin, 1992) and partly by modifying the stationary distribution for some of the chains in the population. Specifically, we consider the sequence of annealed distributions  $\pi_t = \pi^t$  with  $t \in \{\xi_1, \xi_2, \dots, \xi_N\}$ , where  $1 = \xi_1 > \xi_2 > \dots > \xi_N$ , for instance  $\xi_t = 1/t$ . When  $t, s$  are close temperatures,  $\pi_t$  is similar to  $\pi_s$ , but  $\pi = \pi_1$  may be much harder



to sample from than  $\pi_{\xi_N}$  as has been long recognized in the simulated annealing and simulated tempering literature (see Marinari and Parisi, 1992; Geyer and Thompson, 1994; Neal, 1994). Therefore, it is likely that some of the chains designed to sample from  $\pi_1, \dots, \pi_N$  have good mixing properties, making them good candidates for the population of MCMC samplers needed for the IMTM.

We thus consider the Monte Carlo population made of the  $N - 1$  chains having  $\{\pi_2, \dots, \pi_N\}$  as stationary distributions. An example of annealed interacting MTM is given in Algorithm 4. Note that we let the  $i$ -th chain to interact only with the  $N - i + 1$  chains at higher temperature by sampling  $I_1, \dots, I_M$  from  $\mathcal{U}_{\{1, \dots, N-i+1\}}$ .

An astute reader may have noticed that the use of MTM for *each* auxiliary chain may be redundant since for smaller  $\xi_i$ 's the distribution  $\pi_i$  is easy to sample from. In Algorithm 5 we present an alternative implementation of the annealed IMTM in which each auxiliary chain is MH with target  $\pi_i$ ,  $1 \leq i \leq N - 1$ .

---

**Algorithm 4.** *Annealed IMTM Algorithm (AIMTM1)*

---

- For  $i = 1, \dots, N$

1. Let  $x = x_n^{(i)}$  and sample  $I_1, \dots, I_M$  from  $\mathcal{U}_{\{1, \dots, N-i+1\}}$ .
2. For  $j = 1, \dots, M$  draw  $y_j \sim T_j^{(i)}(\cdot | x_n^{(I_j)})$  independently and compute

$$w_j^{(i)}(y_j, x) = \pi(y_j) T_j^{(i)}(y_j | x_n^{(I_j)}) \lambda_j^{(i)}(y_j, x).$$

3. Select  $J \in \{1, \dots, M\}$  with probability proportional to  $w_j^{(i)}(y_j, x)$ ,  $j = 1, \dots, M$  and set  $y = y_J$ .
4. For  $j = 1, \dots, M$  and  $j \neq J$  draw  $x_j^* \sim T_j^{(i)}(\cdot | x_n^{(1:i-1)}, y, x_n^{(i+1:N)})$ , let  $x_j^* = x_n^{(i)}$  and compute

$$w_j^{(i)}(x_j^*, y) = \pi(x_j^*) T_j^{(i)}(x_j^* | y) \lambda_j^{(i)}(x_j^*, y).$$

5. Set  $x_{n+1}^{(i)} = y$  with probability  $\rho_i$ , where  $\rho_i$  is the generalized M.H. ratio of the IMT algorithm and  $x_{n+1}^{(i)} = x_n^{(i)}$  with probability  $1 - \rho_i$ .
-

---

**Algorithm 5.** *Annealed IMTM Algorithm (AIMTM2)*

---

- For  $i = 1$

1. Let  $x = x_n^{(i)}$  and sample  $I_1, \dots, I_M$  from  $\mathcal{U}_{\{1, \dots, N\}}$ .
2. For  $j = 1, \dots, M$  draw  $y_j \sim T_j^{(i)}(\cdot | x_n^{(I_j)})$  independently and compute

$$w_j^{(i)}(y_j, x) = \pi(y_j) T_j^{(i)}(y_j | x_n^{(I_j)}) \lambda_j^{(i)}(y_j, x).$$

3. Select  $J \in \{1, \dots, M\}$  with probability proportional to  $w_j^{(i)}(y_j, x)$ ,  $j = 1, \dots, M$  and set  $y = y_J$ .
4. For  $j = 1, \dots, M$  and  $j \neq J$  draw  $x_j^* \sim T_j^{(i)}(\cdot | x_n^{(1:i-1)}, y, x_n^{(i+1:N)})$ , let  $x_j^* = x_n^{(i)}$  and compute

$$w_j^{(i)}(x_j^*, y) = \pi(x_j^*) T_j^{(i)}(x_j^* | y) \lambda_j^{(i)}(x_j^*, y).$$

5. Set  $x_{n+1}^{(i)} = y$  with probability  $\rho_i$ , where  $\rho_i$  is the generalized M.H. ratio of the IMT algorithm and  $x_{n+1}^{(i)} = x_n^{(i)}$  with probability  $1 - \rho_i$ .

- For  $i = 2, \dots, N$

1. Let  $x = x_n^{(i)}$  and update the proposal function  $T^{(i)}(\cdot | x)$ .
2. Draw  $y \sim T^{(i)}(\cdot | x)$  and compute

$$\rho_i = \min \left\{ 1, \frac{\pi(y)^{\xi_i} T^{(i)}(x | y)}{\pi(x)^{\xi_i} T^{(i)}(y | x)} \right\}.$$

3. Set  $x_{n+1}^{(i)} = y$  with probability  $\rho_i$  and  $x_{n+1}^{(i)} = x_n^{(i)}$  with probability  $1 - \rho_i$ .
- 

The chain of interest, corresponding to  $\xi = 1$ , has an MTM transition kernel with  $M = N - 1$  proposal distributions. At time  $n$  the  $j$ th proposal distribution used for the chain ergodic to  $\pi$  is  $T_j^{(1)} = T^{(j)}$ , is the same as the proposal used by the  $j$ th auxiliary chain, for all  $2 \leq j \leq N$ .

An additional gain could be obtained if the auxiliary chains' transition kernels are modified using adaptive MCMC strategies (see also Chauveau and Vandekerckhove, 2002, for another example of adaption for interacting chains). However, it should be noted that letting the auxiliary chains adapt indefinitely results in complex theoretical justifications for the IMTM which go beyond the scope of this paper and will be presented elsewhere. Our current recommendation is to use finite adaptation for the auxiliary chains prior to the start of the IMTM. One could take advantage of multi-processor computing units and use parallel programming to increase the computational efficiency of this approach.

The adaptation of  $\lambda_j^{(i)}(\cdot, \cdot)$ , through the weights  $\nu_j$  defined in the previous section, should be used cautiously in this case. The aim of the annealing procedure is to allow the higher temperatures chains to explore widely the sample space and to improve the mixing of the MTM chain. Using  $\nu_j$  the context of annealed IMTM could arbitrarily penalize some of the higher temperature proposals and reduce the effectiveness of the annealing strategy.

It is possible to have a Monte Carlo approximation of a quantity of interest by using the output produced by all the chains in the population. For example let

$$\mathcal{I} = \int_{\mathcal{Y}} h(x)\pi(x)dx$$

be the quantity of interest where  $h$  is a test function. It is possible to approximate this quantity as follows

$$\mathcal{I}_{NT} = \frac{1}{T} \sum_{n=1}^T \frac{1}{\bar{\zeta}} \sum_{j=1}^N h(x_n^{(j)})\zeta_j(x_n^{(j)})$$

where  $x_n^{(i)}$ , with  $n = 1, \dots, T$  and  $i = 1, \dots, N$  is the output of a IMTM chains with target  $\pi^{\xi_i}$  and  $\zeta_j(x) = \pi(x)/\pi^{\xi_j}(x)$  is a set of importance weights with normalizing constant  $\bar{\zeta} = \sum_{j=1}^N \zeta_j(x_n^{(j)})$ .

## 2.5 Gibbs within IMTM update

It should be noticed that in the proposed algorithm at the  $n$ -th iteration the  $N$  chains are updated simultaneously. In the interacting MCMC literature a sequential updating scheme (Gibbs-like updating) has been proposed for example in Mengersen and Robert (2003) and Campillo et al. (2009). In the following we show that the Gibbs-like updating also apply to our IMTM context. In the Gibbs-like interacting MTM (GIMTM) algorithm given in

Algorithm 6 the different proposals functions of the  $i$ -th chain, with  $i = 1, \dots, N$ , may depend on the current values of the updated chains  $x_{n+1}^{(j)}$ , with  $j = 1, \dots, (i-1)$  and on the last values  $x_n^{(j)}$ , with  $j = (i+1), \dots, N$ , of the chains which have not yet been updated.

---

**Algorithm 6.** *Gibbs-like IMTM Algorithm (GIMTM)*

---

• For  $i = 1, \dots, N$

1. For  $j = 1, \dots, M_i$  draw  $y_j \sim T_j^{(i)}(\cdot | x_{n+1}^{(1:i-1)}, x_n^{(i)}, x_n^{(i+1:N)})$  independently and compute

$$w_j^{(i)}(y_j, x) = \pi(y_j) T_j^{(i)}(y_j | x_{n+1}^{(1:i-1)}, x_n^{(i)}, x_n^{(i+1:N)}) \lambda_j^{(i)}(y_j, x_n^{(i)}).$$

2. Select  $J \in \{1, \dots, M_i\}$  with probability proportional to  $w_j^{(i)}(y_j, x)$ ,  $j = 1, \dots, M_i$  and set  $y = y_J$ .

3. For  $j = 1, \dots, M_i$  and  $j \neq J$  draw  $x_j^* \sim T_j^{(i)}(\cdot | x_{n+1}^{(1:i-1)}, y, x_n^{(i+1:N)})$ , let  $x_j^* = x_n^{(i)}$  and compute

$$w_j^{(i)}(x_j^*, y) = \pi(x_j^*) T_j^{(i)}(x_j^* | x_{n+1}^{(1:i-1)}, y, x_n^{(i+1:N)}) \lambda_j^{(i)}(x_j^*, y).$$

4. Set  $x_{n+1}^{(i)} = y$  with probability

$$\rho_i = \min \left\{ 1, \frac{w_1^{(i)}(y_1, x) + \dots + w_{M_i}^{(i)}(y_{M_i}, x)}{w_1^{(i)}(x_1^*, y) + \dots + w_{M_i}^{(i)}(x_{M_i}^*, y)} \right\}$$

and  $x_{n+1}^{(i)} = x_n^{(i)}$  with probability  $1 - \rho_i$ .

---

In the GIMTM algorithm the iteration mechanism between the chains is not the same as in the IMTM algorithm. The chains are no longer independent since the proposals may depend on the current values for some of the chains in the population. The transition kernel for the whole population is

$$K(\Xi_n, \Xi_{n+1}) = \prod_{i=1}^N K_i(x_n^{(i)}, x_{n+1}^{(i)} | x_{n+1}^{(1:i-1)}, x_n^{(i+1:N)})$$

and in this case the validity of the algorithm still relies upon the conditional detail balance condition given for the IMTM algorithm.

Finally we remark that the GIMTM algorithm allows us to introduce further possible choices for the  $\lambda_j^{(i)}(x, y)$  functions. In particular a repulsive factor (see Mengersen and Robert (2003)) can be introduced in the selection weights in order to induce negative dependence between the chains. We let the study of the GIMTM algorithm and the use of repulsive factors for future research and focus instead on the properties of the IMTM algorithm.

### 3 Some generalizations

In the following we will discuss some possible generalization of the IMTM algorithm. First we show how to use the stochastic overrelaxation method to possibly have a further gain in the efficiency. Secondly we suggest two possible strategies to built the different proposal functions of the IMTM. The first strategy consists in proposing values along different search directions and represents an extension of the random-ray Monte Carlo algorithm presented in Liu et al. (2000). The second strategy relies upon a suitable combination of target tempering and adaptive MCMC chains.

#### 3.1 Stochastic Overrelaxation

Stochastic overrelaxation (SOR) is a Markov chain Monte Carlo technique developed by Adler (1981) for normal densities and subsequently extended by Green and Han (1992) for non-normal targets. The idea behind this approach is to induce negative correlation between consecutive draws of a single MCMC process.

Within the MTM algorithm we can implement SOR by inducing negative correlation between the proposals and between the proposals and the current state of the chain,  $x$ . A natural and easy to implement procedure may be based on the assumption that  $(y_1, \dots, y_{M-1}, x)^T \sim N_{d \times M}(\mathbf{0}, V)$  where  $V$ 's structure is dictated by the desired negative dependence between the proposals  $y_1, \dots, y_n$ 's and  $x$ , specifically

$$V = \begin{pmatrix} \Sigma_1 & \Psi_{12} & \dots & \Psi_{1M} \\ \Psi_{12} & \Sigma_2 & \dots & \Psi_{2M} \\ \dots & \dots & \dots & \dots \\ \Psi_{1M} & \Psi_{2M} & \dots & \Sigma_M \end{pmatrix}.$$

For instance, we can set  $\Psi_{ij} = 0$  whenever  $i, j \neq M$  and  $\Psi_{iM} = \Sigma_i^{1/2} R_{iM} \Sigma_M^{1/2}$  where  $R_{iM}$  is a correlation matrix which corresponds to extreme negative correlation (see Craiu and Meng, 2005, for a discussion of

extreme dependence) between any two components (with same index) of  $y_i$  and  $x$ , for any  $1 \leq i \leq M - 1$ . The general construction falls within the context of dependent proposals as discussed by Craiu and Lemieux (2007) with the additional bonus of "pulling" the proposals away from the current state due to the imposed negative correlation. This essentially ensures that no proposals are exceedingly close to the current location of the chain. Also note that the construction is general enough and can be applied for Algorithms 2 and 3 as long as the proposal distributions are Gaussian.

### 3.2 Multiple Random-ray Monte Carlo

We show here that the use of different proposals for the MTM algorithm allows also to extend the random-ray Monte Carlo method given in Liu et al. (2000). In particular the proposed algorithm allows to deal with multiple search directions at each iteration of the chains. At the  $n$ -th iteration of the chain, in order to update the set of chains  $\Xi_n$ , the algorithm performs for each chain  $x_n^{(r)} \in \Xi_n$ , with  $r = 1, \dots, N$ , the following steps:

1. Evaluate the gradient  $\log \pi(x)$  at  $x_n^{(r)}$  and find the mode  $a_n$  along  $x_n^{(r)} + ru_n$  where  $u_n = x_n^{(r)} - x_{n-1}^{(r)}$ .
2. Sample  $I_1, \dots, I_M$  from the uniform  $\mathcal{U}_{\{1, \dots, r-1, r+1, \dots, N\}}$ .
3. Let  $e_{n,j} = (a_n - x_n^{(I_j)}) / \|a_n - x_n^{(I_j)}\|$  and sample  $r_j$  from  $\mathcal{N}(0, \sigma^2)$ .

and then use the set of proposals  $T_j$  which depends on  $e_{n,j}$  to perform a MTM transition with different proposals as in the IMTM algorithm (see Alg. 1).

## 4 Simulation Results

### 4.1 Single-chain results

In this section we carry out, through some examples, a comparison between the single-chain multiple try algorithms MTM-DP with different proposals and the algorithm in Liu et al. (2000). In the MTM-DP algorithm we consider four Gaussian random-walk proposals  $\mathbf{y}_j \sim \mathcal{N}_n(\mathbf{x}, \Lambda_j)$  with  $\Lambda_1 = 0.1Id_n$ ,  $\Lambda_2 = 5Id_n$ ,  $\Lambda_3 = 50Id_n$  and  $\Lambda_4 = 100Id_n$ , where  $Id_n$  denotes the  $n$ -order identity matrix. In the MTM selection weights we set  $\lambda_j(\mathbf{x}, \mathbf{y}) = 2\alpha_j / (T_j(\mathbf{x}, \mathbf{y}) + T_j(\mathbf{y}, \mathbf{x}))$ , where  $\alpha_j = 0.25$ , for  $j = 1, \dots, 4$ .

In order to compare the MTM algorithm with different proposals and the Multiple Try algorithm of Liu et al. (2000) we consider 20,000 iterations and use four trials  $\mathbf{y}_j$  generated by the following proposal

$$T_j(\mathbf{x}, \mathbf{y}_j) = \sum_{j=1}^4 \alpha_j \mathcal{N}_n(\mathbf{x}, \Lambda_j) \quad (3)$$

where  $\Lambda_j$ ,  $j = 1, \dots, 4$ , have been defined above. In the weights of the selection step we set  $\lambda(\mathbf{x}, \mathbf{y}) = 2/(T_j(\mathbf{x}, \mathbf{y}) + T_j(\mathbf{y}, \mathbf{x}))$ .

#### 4.1.1 Bivariate Mixture with two components

We consider the following bivariate mixture of two normals

$$\frac{1}{3} \mathcal{N}_2(\boldsymbol{\mu}_1, \Sigma_1) + \frac{2}{3} \mathcal{N}_2(\boldsymbol{\mu}_2, \Sigma_2) \quad (4)$$

with  $\boldsymbol{\mu}_1 = (0, 0)'$ ,  $\boldsymbol{\mu}_2 = (10, 10)'$ ,  $\Sigma_1 = \text{diag}(0.1, 0.5)$  and  $\Sigma_2 = \text{diag}(0.5, 0.1)$ .

In Fig. 1 the ACF with 30 lags for the first component of the bivariate MH chain. The autocorrelation is lower for the MTM algorithm with different proposals.

#### 4.1.2 Multivariate Normal Mixture

We compare the algorithms for high-dimensional targets. We consider the following multivariate mixture of two normals with a sparse variance-covariance structure

$$\frac{1}{3} \mathcal{N}_{20}(\boldsymbol{\mu}_1, \Sigma_1) + \frac{2}{3} \mathcal{N}_{20}(\boldsymbol{\mu}_2, \Sigma_2) \quad (5)$$

with  $\boldsymbol{\mu}_1 = (3, \dots, 3)'$ ,  $\boldsymbol{\mu}_2 = (10, \dots, 10)'$  and  $\Sigma_j$ , with  $j = 1, 2$ , generated independently from a Wishart distribution  $\Sigma_j \sim \mathcal{W}_{20}(\nu, Id_{20})$  where  $\nu$  is the degrees of freedom parameter of the Wishart. In the experiments we set  $\nu = 21$ .

The autocorrelation function of the chain for one of the experiment is given in Fig. 2. The ACF has been evaluated for each components of the 20-dimensional chain. The values of the ACF of the MTM-DP (black lines in Fig. 2) are less than those of the original MTM (gray lines of the same figure) in all the directions of the support space. We conclude that the MTM algorithm with different proposals (Algorithm 3) outperforms the Liu et al. (2000) MTM algorithm.

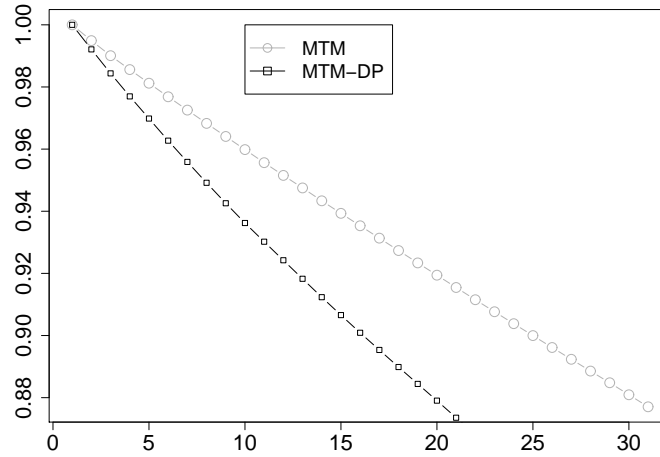


Figure 1: Autocorrelation function of the Liu's MTM (gray line) and MTM with different proposals (black line).

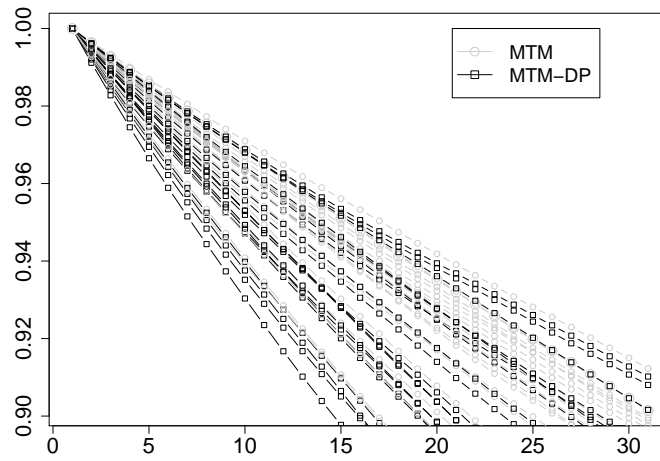


Figure 2: Autocorrelation function (ACF) of the 20 components of the multivariate MH chain for the Liu's MTM (gray lines) and MTM with different proposals (black lines).



## 4.2 Multiple-chains results

In this section we show real and simulated data results of the general interacting multiple try algorithm in Alg. 1.

### 4.2.1 Bivariate Mixture with two components

The target distribution is the following bivariate mixture of two normals

$$\frac{1}{3}\mathcal{N}_2(\boldsymbol{\mu}_1, \Sigma_1) + \frac{2}{3}\mathcal{N}_2(\boldsymbol{\mu}_2, \Sigma_2) \quad (6)$$

with  $\boldsymbol{\mu}_1 = (0, 0)'$ ,  $\boldsymbol{\mu}_2 = (10, 10)'$ ,  $\Sigma_1 = \text{diag}(0.1, 0.5)$  and  $\Sigma_2 = \text{diag}(0.5, 0.1)$ .

We consider a population of  $N = 50$  chains with  $M = 50$  proposals and 1,000 iterations of the IMTM algorithm. For each chain we consider the case  $T_j^{(i)}(\mathbf{y}|\mathbf{x}_{n+1}^{(1:i-1)}, \mathbf{x}_n^{(i)}, \mathbf{x}_n^{(i+1:N)}) = T^{(j)}(\mathbf{y}, \mathbf{x}_n^{(j)})$  and draw

$$\mathbf{y}_j \sim \mathcal{N}_2(\mathbf{x}_n^{(j)}, \Lambda_j) \quad (7)$$

where  $\Lambda_j = (0.1 + 5j)Id_2$ . In this specification of the IMTM algorithm each chain has  $M$  independent proposals with conditional mean given by the previous values of the chains in the population.

In this experiment we consider two kind of weights. First we set  $\lambda_j^{(i)}(x, y) = (T_j^{(i)}(x, y)T_j^{(i)}(y, x))^{-1}$ , that corresponds to use importance sampling selection weights, secondly we consider  $\lambda_j^{(i)}(x, y) = 2(T_j^{(i)}(x, y) + T_j^{(i)}(y, x))^{-1}$ , which implies a symmetric MTM algorithm. We denote with IMTM-IS and IMTM-TA respectively the resulting algorithms.

Fig. 3 show the results of the IMTM-IS and IMTM-TA algorithms at the last iteration of the population of chains (black dots). In both of the algorithms the population is visiting the two modes of the distribution in the right proportion. Moreover each chain is able to jump from one mode to the other. (the light-gray line represents the sample path of one of the chain).

### 4.2.2 Beta-Binomial Model

We consider here the problem of the genetic instability of esophageal cancers. During a neoplastic progression the cancer cells undergo a number of genetic changes and possibly lose entire chromosome sections. The loss of a chromosome section containing one allele by abnormal cells is called *Loss of Heterozygosity* (LOH). The LOH can be detected using laboratory assays

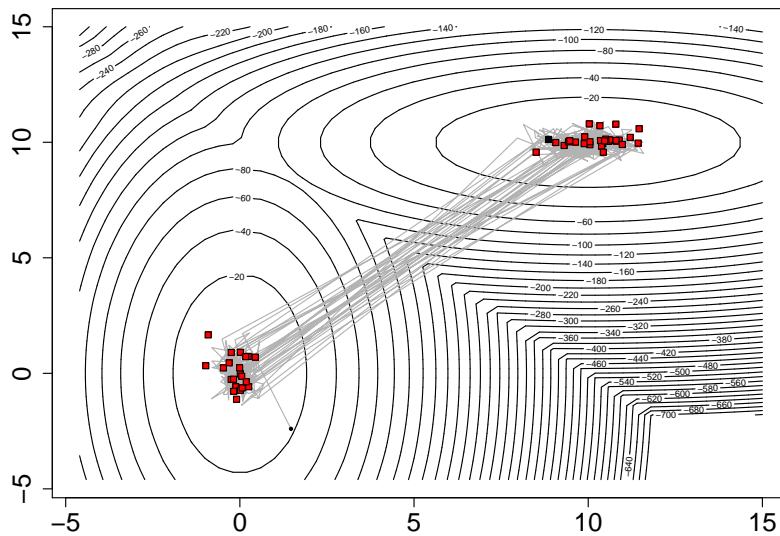
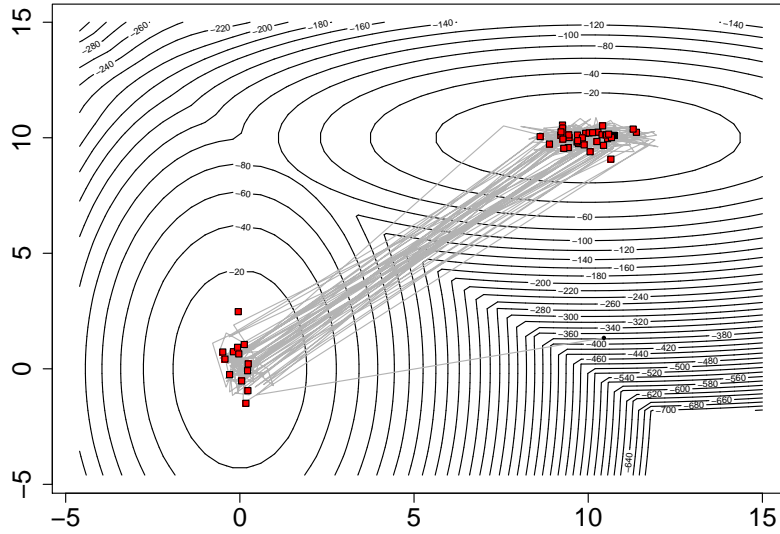


Figure 3: Log-target level sets (solid lines), the MH chains (red dots) at the last iteration of the IMTM algorithms and the path of one of the chain of the set (gray line). Up: output of the IMTM-IS with importance sampling selection weights. Bottom: output of the IMTM-TA with symmetric lambda as in (c).

on patients with two different alleles for a particular gene. Chromosome regions containing genes which regulate cell behavior, are hypothesized to have a high rates of LOH. Consequently the loss of these chromosome sections disables important cellular controls.

Chromosome regions with high rates of LOH are hypothesized to contain *Tumor Suppressor Genes* (TSGs), whose deactivation contributes to the development of esophageal cancer. Moreover the neoplastic progression is thought to produce a high level of background LOH in all chromosome regions.

In order to discriminate between "background" and TSGs LOH, the Seattle Barrett's Esophagus research project (Barrett et al. (1996)) has collected LOH rates from esophageal cancers for 40 regions, each on a distinct chromosome arm. The labeling of the two groups is unknown so Desai (2000) suggest to consider a mixture model for the frequency of LOH in both the "background" and TSG groups.

We consider the hierarchical Beta-Binomial mixture model proposed in Warnes (2001)

$$f(x, n|\eta, \pi_1, \pi_2, \gamma) = \eta \binom{n}{x} \pi_1^x (1 - \pi_1)^{n-x} + (1 - \eta) \binom{n}{x} \frac{\Gamma(1/\omega_2)}{\Gamma(\pi_2/\omega_2)\Gamma((1 - \pi_2)/\omega_2)} \frac{\Gamma(x + \pi_2/\omega_2)\Gamma(n - x + (1 - \pi_2)/\omega_2)}{\Gamma(n + 1/\omega_2)} \quad (8)$$

with  $x$  number of LOH sections,  $n$  the number of examined sections,  $\omega_2 = \exp\{\gamma\}/(2(1 + \exp\{\gamma\}))$ . Let  $\mathbf{x} = (x_1, \dots, x_m)$  and  $\mathbf{n} = (n_1, \dots, n_m)$  be a set of observations from  $f(x, n|\eta, \pi_1, \pi_2, \gamma)$  and let us assume the following priors

$$\eta \sim \mathcal{U}_{[0,1]}, \quad \pi_1 \sim \mathcal{U}_{[0,1]}, \quad \pi_2 \sim \mathcal{U}_{[0,1]} \quad \text{and} \quad \gamma \sim \mathcal{U}_{[-30,30]} \quad (9)$$

with  $\mathcal{U}$  the uniform distribution on  $[a, b]$ . Then the posterior distribution is

$$\pi(\eta, \pi_1, \pi_2, \gamma|\mathbf{x}, \mathbf{n}) \propto \prod_{j=1}^m f(x_j, n_j|\eta, \pi_1, \pi_2, \gamma) \quad (10)$$

The parametric space is of dimension four:  $(\eta, \pi_1, \pi_2, \gamma) \in [0, 1]^3 \times [-30, 30]$  and the posterior distribution has two well-separated modes making it difficult to sample using generic methods.

We apply the IMTM algorithm with 100 iterations,  $M = 4$  proposal functions randomly selected between a population of  $N = 100$  chains. For each chain we consider importance sampling weights in the selection step,

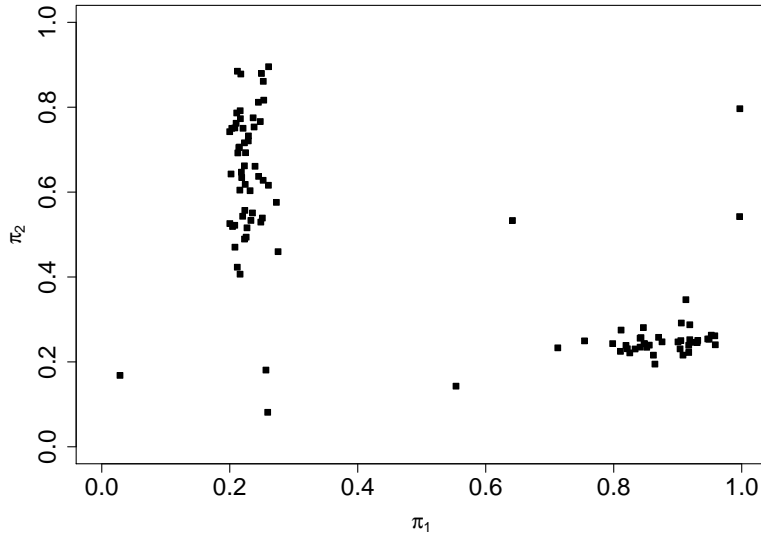


Figure 4: Values of the population of chains (dots) at the last iteration on the subspace  $(\pi_1, \pi_2)$ . The interaction is given by  $M = 4$  proposal functions randomly selected between the population of  $N = 100$  chains.

i.e. we set  $\lambda_j^{(i)}(x, y) = (T_j^{(i)}(x, y)T_j^{(i)}(y, x))^{-1}$  with  $j = 1, \dots, 4$  and  $i = 1, \dots, 100$ . The values of the population of chains (dots) at the last iteration on the subspace  $(\pi_1, \pi_2)$  is given in Fig. 4.2.2. The IMTM is able to visit both regions of the parameter space confirming the analysis of Craiu et al. (2009) and Warnes (2001).

### 4.2.3 Stochastic Volatility

The estimation of the stochastic volatility (SV) model due to Taylor (1994) still represents a challenging issue in both off-line (Celeux et al. (2006)) and sequential (Casarin and Marin (2009)) inference contexts. One of the main difficulties is due to the high dimension of the sampling space which hinders the use of the data-augmentation and prevents a reliable joint estimation of the parameters and the latent variables. As highlighted in Casarin et al. (2009) using multiple chains with a chain interaction mechanism could lead to a substantial improvement in the MCMC method for this kind of model.

We consider the SV model given in Celeux et al. (2006)

$$\begin{aligned} y_t|h_t &\sim \mathcal{N}\left(0, e^{h_t}\right) \\ h_t|h_{t-1}, \boldsymbol{\theta} &\sim \mathcal{N}\left(\alpha + \phi h_{t-1}, \sigma^2\right) \\ h_0|\boldsymbol{\theta} &\sim \mathcal{N}\left(0, \sigma^2/(1 - \phi^2)\right) \end{aligned}$$

with  $t = 1, \dots, T$  and  $\boldsymbol{\theta} = (\alpha, \phi, \sigma^2)$ . For the parameters we assume the noninformative prior (see Celeux et al., 2006)

$$\pi(\boldsymbol{\theta}) \propto 1/(\sigma\beta)\mathbb{I}_{(-1,1)}(\phi)$$

where  $\beta^2 = \exp(\alpha)$ . In order to simulate from the posterior we consider the full conditional distributions and apply a Gibbs algorithm. If we define  $\mathbf{y} = (y_1, \dots, y_T)$  and  $\mathbf{h} = (h_0, \dots, h_T)$  then the full conditionals for  $\beta$  and  $\phi$  are the inverse gamma distributions

$$\begin{aligned} \beta^2|\mathbf{h}, \mathbf{y} &\sim \mathcal{IG}\left(\sum_{t=1}^T y_t^2 \exp(-h_t)/2, (T-1)/2\right) \\ \sigma^2|\phi, \mathbf{h}, \mathbf{y} &\sim \mathcal{IG}\left(\sum_{t=2}^T (h_t - \phi h_{t-1})^2/2 + h_1^2(1 - \phi^2), (T-1)/2\right) \end{aligned}$$

and  $\phi$  and the latent variables have non-standard full conditionals

$$\begin{aligned} \pi(\phi|\sigma^2, \mathbf{h}, \mathbf{y}) &\propto (1 - \phi^2)^{1/2} \exp\left(-\frac{\phi^2}{2\sigma^2} \sum_{t=2}^{T-1} h_t^2 - \frac{\phi}{\sigma^2} \sum_{t=2}^T h_t h_{t-1}\right) \mathbb{I}_{(-1, +1)}(\phi) \\ \pi(h_t|\alpha, \phi, \sigma^2, \mathbf{h}, \mathbf{y}) &\propto \exp\left(-\frac{1}{2\sigma^2} \left((h_t - \alpha - \phi h_{t-1})^2 - \right. \right. \\ &\quad \left. \left. (h_{t+1} - \alpha - \phi h_t)^2\right) - \frac{1}{2} (h_t + y_t^2 \exp(-h_t))\right). \end{aligned}$$

In order to sample from the posterior we use an IMTM within Gibbs algorithm. A detailed description of the proposal distributions for  $\phi$  and  $h_t$  can be found in Celeux et al. (2006).

We consider the two parameter settings  $(\alpha, \phi, \sigma^2) = (0, 0.99, 0.01)$  and  $(\alpha, \phi, \sigma^2) = (0, 0.9, 0.1)$  which correspond, in a financial stock market context, to daily and weekly frequency data respectively. Note that as reported in Casarin and Marin (2009) inference in the daily example is more difficult. We compare the performance of MH within Gibbs and IMTM within Gibbs algorithms in terms of Mean Square Error (MSE) for the parameters and

of cumulative RMSE for the latent variables. We carry out the comparison in statistical terms and estimate the MSE and RMSE by running the algorithms on 20 independent simulated datasets of 200 observations. In the comparison we take into account the computational cost and for the IMTM within Gibbs we use  $N = 20$  interacting chains, 1,000 iterations and  $M = 5$  proposal functions. This setting corresponds to 100,000 random draws and is equivalent to the 100,000 iterations of the MH within Gibbs algorithm used in the comparison. Note that the proposal step of the IMTM selects at random the proposal functions between the other chains and the selection step uses  $\lambda_j^{(i)}(x, y) = (T_j^{(i)}(x, y)T_j^{(i)}(y, x))^{-1}$ , with  $h = 1, \dots, 5$  and  $i = 1, \dots, 20$ .

The results for the parameter estimation when applying IMTM are presented in Table 1 and show an effective improvement in the estimates, both for weekly and daily data, when compared to the results of a MH algorithm with an equivalent computational load.

		Daily Data				Weekly Data	
$\theta$	Value	MSE		$\theta$	Value	MSE	
		IMTM	MH			IMTM	MH
$\alpha$	0	0.04698 (0.00612)	0.09517 (0.00194)	$\alpha$	0	0.00146 (0.00139)	0.00849 (0.00105)
$\phi$	0.99	0.20109 (0.02414)	0.34825 (0.05187)	$\phi$	0.9	0.01328 (0.04014)	0.10746 (0.03629)
$\sigma^2$	0.01	0.00718 (0.00173)	0.02380 (0.00202)	$\sigma^2$	0.1	0.00136 (0.00141)	0.09175 (0.00358)

Table 1: Mean square error (MSE) and its standard deviation (in parenthesis) for the parameter estimation with IMTM and MH within Gibbs algorithms. Left panel: daily datasets. Right panel: weekly dataset.

A typical output of the IMTM for some chains of the population and for the latent variables  $\mathbf{h}$  is given in Fig. 4.2.3. Each chart shows for a given chain the estimated latent variables (dotted black line), the posterior quantiles (gray lines) and the true value of  $\mathbf{h}$  (solid black line).

Figures 4.2.3 and 4.2.3 exhibit the HPD region at the 90% (gray areas) and the mean (black lines) of the cumulative RMSE of each algorithm for the weekly and daily data, respectively. The statistics have been estimated from 20 independent experiments. The average RMSE shows that, in both parameter settings considered here, the IMTM (dashed black line) is more

efficient than the standard MH algorithm (solid black line).

## 5 Conclusions

In this paper we propose a new class of interacting Metropolis algorithm, with Multiple Try transition. These algorithms extend the existing literature in two directions. First we show a natural and not straightforward way to include the chains interaction in a multiple try transition. Secondly the multiple try transition has been extended in order to account for the use of different proposal functions. We give a proof of validity of the algorithm and show on real and simulated examples the effective improvement in the mixing property and exploration ability of the resulting interacting chains.

## Appendix A

### Proof

Without loss of generality, we drop the chain index  $i$  and the iteration index  $n$ , set  $M_i = N, \forall i$  and  $x_n^{(i)} = x$  and denote with  $T_j(y, x) = T_j(y|x_{n+1}^{(1:i-1)}, x_n^{(i)}, x_n^{(i+1:N)})$   $\lambda_j^{(i)}(y_j, x_n^{(i)})$  the  $j$  proposal of the  $i$ -th chain at the iteration  $n$  conditional on the past and current values,  $x_{n+1}^{(1:i-1)}$  and  $x_n^{(i+1:N)}$  respectively, of the other chains.

Let us define the following quantities

$$\bar{w}(y_{1:N}|x) = \sum_{j=1}^N w_j(y_j, x), \quad \bar{w}_{-k}(y_{1:N}|x) = \sum_{j \neq k}^N w_j(y_j, x)$$

and

$$S_N(J) = \frac{1}{\bar{w}(y_{1:N}|x)} \sum_{j=1}^N \delta_j(J) w_j(y_j, x)$$

with  $J \in \mathcal{J} = \{1, \dots, N\}$  the empirical measure generated by different proposals and by the normalized selection weights.

Let  $T(x, dy_{1:N}) = \bigotimes_{j=1}^N T_j(x, dy_j) : (\mathcal{Y} \times \mathcal{B}(\mathcal{Y}^N)) \mapsto [0, 1]$  the joint proposal for the multiple try and define  $T_{-k}(x, dy_{1:N}) = \bigotimes_{j \neq k}^N T_j(x, dy_j)$ . Let  $A(x, y)$  be the actual transition probability for moving from  $x$  to  $y$  in the IMT2 algorithm. Suppose that  $x \neq y$ , then the transition is a results two steps. The first step is a selection step which can be written as  $y = y_J$

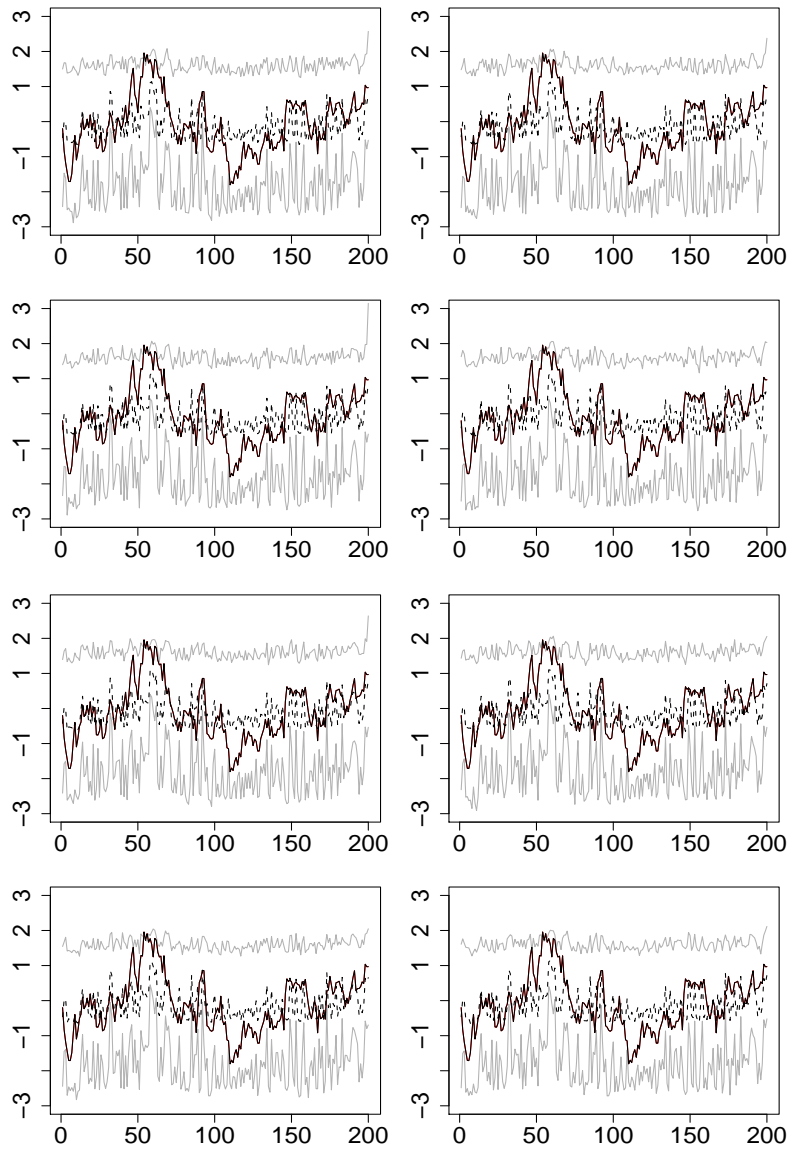


Figure 5: Typical output of the population of chains for a weekly dataset. For each chain the estimated latent variable (dotted black line), the 2.5% and 97.5% quantiles (gray lines) and the true value of  $\mathbf{h}$  (solid black line).



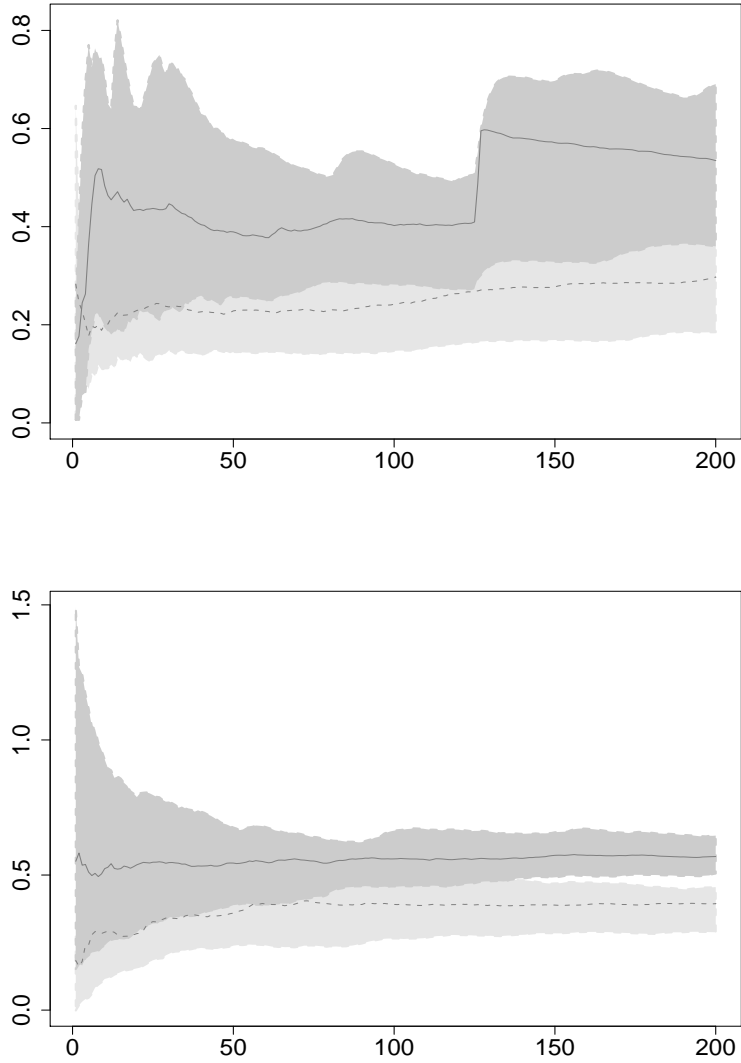


Figure 6: Cumulative RMSE for the IMTM (dashed line) and MH (solid line) and the 90% HPD regions for the IMTM (light gray) and MH (dark gray) estimated on 20 independent experiments for both the daily (up) and weekly (bottom) datasets.

and  $x_j^* = x$  with the random index  $J$  sampled from the empirical measure  $S_N(J)$ . The second step is a accept/reject step based on the generalized MH ratio which involves the generation of the auxiliary values  $x_j^*$  for  $j \neq J$ . Then

$$\begin{aligned}
\pi(x)A(x, y) &= \\
&= \pi(x) \int_{\mathcal{Y}^N} T(x, dy_{1:N}) \int_{\mathcal{J}} S_N(dJ) \int_{\mathcal{Y}^{N-1} \times \mathcal{Y}^2} T_{-J}(y, dx_{1:N}^*) \times \\
&\quad \times \delta_x(dx_J^*) \delta_{y_J}(dy) \min \left\{ 1, \frac{\bar{w}(y_{1:N}|x)}{\bar{w}(x_{1:N}^*|y)} \right\} \\
&= \pi(x) \sum_{j=1}^N \int_{\mathcal{Y}^{N-1}} T_{-j}(x, dy_{1:N}) T_j(x, y) \int_{\mathcal{Y}^{N-1}} T_{-j}(y, dx_{1:N}^*) \times \\
&\quad \times \frac{w_j(y, x)}{w_j(y, x) + \bar{w}_{-j}(y_{1:N}|x)} \min \left\{ 1, \frac{w_j(y, x) + \bar{w}_{-j}(y_{1:N}|x)}{w_j(x, y) + \bar{w}_{-j}(x_{1:N}^*|y)} \right\} \\
&= \sum_{j=1}^N \frac{w_j(x, y) w_j(y, x)}{\lambda_j(y, x)} \int_{\mathcal{Y}^{2(N-1)}} T_{-j}(x, dy_{1:N}) \times \\
&\quad \times T_{-j}(y, dx_{1:N}^*) \min \left\{ \frac{1}{w_j(y, x) + \bar{w}_{-j}(y_{1:N}|x)}, \frac{1}{w_j(x, y) + \bar{w}_{-j}(x_{1:N}^*|y)} \right\}
\end{aligned}$$

which is symmetric in  $x$  and  $y$ .

## References

- BARRETT, M., GALIPEAU, P., SANCHEZ, C., EMOND, M. and REID, B. (1996). Determination of the frequency of loss of heterozygosity in esophageal adeno-carcinoma nu cell sorting, whole genome amplification and microsatellite polymorphisms. *Oncogene* **12**.
- CAMPILLO, F., RAKOTOZAFY, R. and ROSSI, V. (2009). Parallel and interacting Markov chain Monte Carlo algorithm. *Mathematics and Computers in Simulation* **79** 3424–3433.
- CAPPÉ, O., GULLIN, A., MARIN, J. and ROBERT, C. P. (2004). Population Monte Carlo. *J. Comput. Graph. Statist.* **13** 907–927.
- CASARIN, R. and MARIN, J.-M. (2009). Online data processing: Comparison of Bayesian regularized particle filters. *Electronic Journal of Statistics* **3** 239–258.

- CASARIN, R., MARIN, J.-M. and ROBERT, C. (2009). A discussion on: Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations by Rue, H. Martino, S. and Chopin, N. *Journal of the Royal Statistical Society Ser. B* **71** 360–362.
- CELEUX, G., MARIN, J.-M. and ROBERT, C. (2006). Iterated importance sampling in missing data problems. *Computational Statistics and Data Analysis* **50** 3386–3404.
- CHAUVEAU, D. and VANDEKERKHOVE, P. (2002). Improving convergence of the hastings-metropolis algorithm with an adaptive proposal. *Scandinavian Journal of Statistics* **29** 13.
- CRAIU, R. V. and LEMIEUX, C. (2007). Acceleration of the multiple-try Metropolis algorithm using antithetic and stratified sampling. *Statistics and Computing* **17** 109–120.
- CRAIU, R. V. and MENG, X. L. (2005). Multi-process parallel antithetic coupling for forward and backward MCMC. *Ann. Statist.* **33** 661–697.
- CRAIU, R. V., ROSENTHAL, J. S. and YANG, C. (2009). Learn from thy neighbor: Parallel-chain adaptive and regional MCMC. *Journal of the American Statistical Association* **104** 1454–1466.
- DEL MORAL, P. (2004). *Feynman-Kac Formulae. Genealogical and Interacting Particle Systems with Applications*. Springer.
- DEL MORAL, P. and MICLO, L. (2000). Branching and interacting particle systems approximations of feynman-kac formulae with applications to non linear filtering. In *Séminaire de Probabilités XXXIV. Lecture Notes in Mathematics, No. 1729*. Springer, 1–145.
- DESAI, M. (2000). *Mixture Models for Genetic changes in cancer cells*. Ph.D. thesis, University of Washington.
- GELMAN, A. and RUBIN, D. B. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statist. Sci.* 457–511.
- GEYER, C. J. and THOMPSON, E. A. (1994). Annealing Markov chain Monte Carlo with applications to ancestral inference. Tech. Rep. 589, University of Minnesota.
- HASTINGS, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57** 97–109.

- HEARD, N. A., HOLMES, C. and STEPHENS, D. (2006). A quantitative study of gene regulation involved in the immune response of anopheline mosquitoes: an application of Bayesian hierarchical clustering of curves. *J. Amer. Statist. Assoc.* **101** 18–29.
- JASRA, A., STEPHENS, D. and HOLMES, C. (2007). On population-based simulation for static inference. *Statist. Comput.* **17** 263–279.
- JENNISON, C. (1993). Discussion of "Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods," by A.F.M. Smith and G.O. Roberts. *J. Roy. Statist. Soc. Ser. B* **55** 54–56.
- LIU, J., LIANG, F. and WONG, W. (2000). The multiple-try method and local optimization in Metropolis sampling. *Journal of the American Statistical Association* **95** 121–134.
- MARINARI, E. and PARISI, G. (1992). Simulated tempering: A new Monte Carlo scheme. *Europhysics Letters* **19** 451–458.
- MENGERSEN, K. and ROBERT, C. (2003). The pinball sampler. In *Bayesian Statistics 7* (J. Bernardo, A. Dawid, J. Berger and M. West, eds.). Springer-Verlag.
- METROPOLIS, N., ROSENBLUTH, A., ROSENBLUTH, M., TELLER, A. and TELLER, E. (1953). Equations of state calculations by fast computing machines. *J. Chem. Ph.* **21** 1087–1092.
- NEAL, R. M. (1994). Sampling from multimodal distributions using tempered transitions. Tech. Rep. 9421, University of Toronto.
- PANDOLFI, S., BARTOLUCCI, F. and FRIEL, N. (2010a). A generalization of the multiple-try metropolis algorithm for bayesian estimation and model selection. In *13th International Conference on Artificial Intelligence and Statistics (AIS-TATS), Chia Laguna Resort, Sardinia, Italy. ???*
- PANDOLFI, S., BARTOLUCCI, F. and FRIEL, N. (2010b). A generalized Multiple-try Metropolis version of the Reversible Jump algorithm. Tech. rep., <http://arxiv.org/pdf/1006.0621>.
- PRITCHARD, J. K., STEPHENS, M. and DONNELLY, P. (2000). Inference of population structure using multilocus genotype data. *Genetics* **155** 945–959.

- ROBERTS, G. O. and ROSENTHAL, J. S. (2007). Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *J. Appl. Probab.* **44** 458–475.
- TAYLOR, S. (1994). Modelling stochastic volatility. *Mathematical Finance* **4** 183–204.
- WARNES, G. (2001). The Normal kernel coupler: An adaptive Markov chain Monte Carlo method for efficiently sampling from multi-modal distributions. Technical report, George Washington University.