Available online at www.sciencedirect.com

SciVerse ScienceDirect

Stochastic Processes and their Applications 122 (2012) 758–786

stochastic processes and their applications

www.elsevier.com/locate/spa

Scaling analysis of multiple-try MCMC methods

Mylène Bédard[a,*], Randal Douc[b], Eric Moulines[c]

[a] *Département de mathématiques et de statistique, Université de Montréal, C.P. 6128, succ. Centre-ville, Montréal, H3C 3J7, Canada*

[b] *SAMOVAR, CNRS UMR 5157 - Institut Télécom/Télécom SudParis, 9 rue Charles Fourier, 91000 Evry, France*

[c] *LTCI, CNRS UMR 8151 - Institut Télécom /Télécom ParisTech, 46, rue Barrault, 75634 Paris Cedex 13, France*

Received 9 December 2010; received in revised form 14 November 2011; accepted 15 November 2011
Available online 25 November 2011

Abstract

Multiple-try methods are extensions of the Metropolis algorithm in which the next state of the Markov chain is selected among a pool of proposals. These techniques have witnessed a recent surge of interest because they lend themselves easily to parallel implementations. We consider extended versions of these methods in which some dependence structure is introduced in the proposal set, extending earlier work by Craiu and Lemieux (2007).

We show that the speed of the algorithm increases with the number of candidates in the proposal pool and that the increase in speed is favored by the introduction of dependence among the proposals. A novel version of the hit-and-run algorithm with multiple proposals appears to be very successful.

MSC: 60F05; 65C40

Keywords: Random walk Metropolis; Weak convergence; Diffusion; Correlated proposals; Auxiliary random variables

Markov chain Monte Carlo (MCMC) methods allow generation of samples from virtually any target distribution π. In this paper, we focus on the multiple-try Metropolis (MTM) algorithm. Starting from the state x, the algorithm first generates K trial values. A candidate is then selected

* Corresponding author. Tel.: +1 514 343 6111; fax: +1 514 343 5700.
E-mail addresses: bedard@dms.umontreal.ca (M. Bédard), randal.douc@it-sudparis.eu (R. Douc), eric.moulines@telecom-paristech.fr (E. Moulines).

doi:10.1016/j.spa.2011.11.004

from the trial set with probability proportional to some positive weights. This candidate is finally accepted or rejected.

The idea of proposing multiple trials has been introduced in Monte Carlo simulations in molecular dynamics (see [7, Chapter 13], [8, Section 6.7]) and later in computational statistics by [10] under the name of MTM algorithm; see [9] and the references therein for a recent survey.

In the original version of the MTM algorithm, the proposals in the pool are independent and identically distributed. In their paper, [4] have shown how to adapt the MTM algorithm to dependent proposals, leading to the *multiple correlated-try Metropolis* (MCTM) algorithm. The simulations presented in [4] suggest that it is beneficial to design the joint proposal distribution so as to maximize the average squared distance between any pair of proposals in the pool (this algorithm is referred to as the extreme antithetic MTM).

In this paper, we consider an even more extreme form of dependence in which all the proposals are drawn using a common random variable (this algorithm is referred to as the MTM-C). We show that the acceptance ratio of the MTM-C can be computed without drawing additional auxiliary variables to guarantee reversibility. An instantiation of this algorithm is a novel version of the hit-and-run algorithm, in which the proposals in the pool are all obtained along the same search direction, with different (deterministic) step sizes.

As is usually the case with refined algorithms, extra steps and computational effort are required in order to implement the multiple-try Metropolis. It is of course of interest to understand under which scenarios these implementations are preferable to the plain Metropolis algorithm. To allow for a fair comparison, we need to be careful in the selection of the proposal distributions and the tuning of the algorithms.

Comparing different algorithms under general settings is a difficult task: when working with low-dimensional target distributions, there might exist various measures of efficiency which might lead to different tunings and conclusions. This issue however disappears in high-dimensional settings, as all efficiency measures reduce to a common criterion, the speed of the diffusion obtained as a weak limit of an appropriately scaled – in space and time – version of the algorithms (see [11–13,1,2]).

To be able to carry out a comparison of the methods mentioned, we thus work under the particular framework of high-dimensional target distributions formed of independent and identically distributed (i.i.d.) components. We develop in Appendix a general theory which extends the results obtained in [11] for the Metropolis algorithm to Multiple-Try Metropolis methods (as well as other MCMC algorithms; see [3]). In particular, we first establish in Theorem 5 a convergence in the Skorokhod topology of some rescaled process (ultimately our preferred algorithm) to a Markov process under general assumptions. In Theorem 7, we then specialize this result to a general class of multiple-try Metropolis algorithms. We finally prove the convergence of the MCTM algorithm (see Theorem 2) and MTM-C algorithm (see Theorem 3) to Langevin diffusions with explicit expressions for their respective speed.

In this framework, we show that the extreme antithetic proposals improve upon the MTM with independent proposals (see Section 1.1, Algorithm 1 for a description of the algorithms and Theorem 2 for a precise statement of the results). Since the introduction of correlation makes the computation of the acceptance ratio more complex, this increase in complexity might make the extreme antithetic proposals less efficient than the MTM with independent proposals in practical implementations (we refer the reader to the discussion in Section 4).

Our preferred choice is the MTM hit-and-run algorithm (see the description of the algorithm in Section 1.2, the statement of the results in Theorem 3 and the discussion in Section 4).

In particular, it is shown that the use of MTM-C with two deterministic antithetic proposals has a speed which is twice that of the Metropolis algorithm at the price of a marginal increase of the computational cost in many scenarios.

The paper is arranged as follows. The multiple-try Metropolis algorithm and its variants are introduced in Section 1. Results for the multiple correlated-try Metropolis (MCTM) algorithms are exposed in Section 2. The results for the multiple-try Metropolis with common random variables (MTM-C) are presented in Section 3. Section 4 compares the performance of the different implementations of the MTM algorithm and discusses the impact of the dependence in the proposal pool. The proof of the various theorems are presented in Appendix.

1. Multiple-try algorithms

Denote by X the state space of the Markov chain, assumed to be equipped with a countably generated σ-field $\mathcal{X}$. We denote by μ a σ-finite measure, and assume that the target distribution π has a density (also denoted by π) with respect to μ.

1.1. Multiple correlated-try algorithms (MCTM)

Following [4], it is assumed that the proposal pool is generated jointly. For K a positive integer, let $q(x;\cdot)$ denote the conditional density of the proposal $(Y^1,\ldots,Y^K)$ given the current state $X=x$. For $j\in\{1,\ldots,K\}$, denote by $q_j(x;\cdot)$ the marginal conditional distribution of Y^j given $X=x$:

$$q_j(x;y^j)=\int\cdots\int q(x;y^1,\ldots,y^K)\prod_{i\neq j}\mu(\mathrm{d}y^i). \tag{1}$$

It is assumed in the sequel that, for any $j\in\{1,\ldots,K\}$, $q_j(x;y)>0$ if and only if $q_j(y;x)>0$, μ-a.e. For $j\in\{1,\ldots,K\}$, let $\bar{q}_j(x,y^j;\cdot)$ denote the Markov transition density from X^2 to X^{K-1} given by

$$\bar{q}_j(x,y^j;(y^i)_{i\neq j})=\frac{q(x;y^1,\ldots,y^K)}{q_j(x;y^j)}. \tag{2}$$

By construction, $\bar{q}_j(x,y^j;\cdot)$ is the conditional density of the random vector $(Y^i)_{i\neq j}$ given $X=x$ and $Y^j=y^j$.

Let $\{w_j(x,y)\}_{j=1}^K$ be positive functions on $\mathsf{X}\times\mathsf{X}$. Denoting by $X=x$ the current state of the chain, the multiple correlated-try Metropolis algorithm (MCTM) is defined as follows.

Algorithm 1 (*MCTM*)**.**

(a) Draw K trials $(Y^1,\ldots,Y^K)$ jointly from the transition density $q(x;\cdot)$.
(b) Draw an index $J\in\{1,\ldots,K\}$, with probability proportional to

$$[w_1(Y^1,x),\ldots,w_K(Y^K,x)].$$

(c) Draw $K-1$ auxiliary variables $\{\tilde{Y}^{J,i}\}_{i\neq J}$ from the auxiliary kernel $\bar{q}_J(Y^J,x;\cdot)$, where $(\bar{q}_j,1\leq j\leq K)$ are defined in (2).
(d) Accept the proposal Y^J with probability $\alpha^J(x,(Y^i)_{i=1}^K,(\tilde{Y}^{J,i})_{i\neq J})$, where, for $j\in\{1,\ldots,K\}$,

$$\alpha^j(x,(y^i)_{i=1}^K,(\tilde{y}^{j,i})_{i\neq j}) = 1 \wedge \frac{\pi(y^j)w_j(x,y^j)q_j(y^j;x)\left(\sum_{i\neq j} w_i(y^i,x)+w_j(y^j,x)\right)}{\pi(x)w_j(y^j,x)q_j(x;y^j)\left(\sum_{i\neq j} w_i(\tilde{y}^{j,i},y^j)+w_j(x,y^j)\right)} \tag{3}$$

and reject otherwise.

Note that given $J=j$, the auxiliary variables $\{\tilde{Y}^{j,i}\}_{i\neq j}$ in step (c) are simulated according to the conditional distribution $\bar{q}_j(Y^j,x;\cdot)$ while the direct sample $(Y^i)_{i\neq j}$ is simulated according to $\bar{q}_j(x,Y^j;\cdot)$. Following [10, Theorem 1] and [4, Proposition 2.1], it is easily seen that this Markov chain satisfies the detailed balance condition for π. If $w_j(x,y)=\pi(x)$ and the marginal proposal transition kernels are symmetric ($q_j(x;y)=q_j(y;x)$), then the acceptance ratio becomes

$$\alpha^j(x,(y^i)_{i=1}^K,(\tilde{y}^{j,i})_{i\neq j}) = 1 \wedge \frac{\sum_{i\neq j}\pi(y^i)+\pi(y^j)}{\sum_{i\neq j}\pi(\tilde{y}^{j,i})+\pi(x)}. \tag{4}$$

This simple version of the Multiple-Try algorithm corresponds to the method of *orientational biased Monte Carlo* for molecular simulations introduced in [7, Chapter 13].

In the original multiple-try Metropolis of [10], the global transition kernel is equal to the product of the marginal kernels: $q(x;y^1,\dots,y^K)=\prod_{j=1}^K q_j(x;y^j)$ and $\bar{q}_j(x,y^j;(y^i)_{i\neq j}) = \prod_{i\neq j} q_i(x;y^i)$. In [4], the authors investigate the use of dependent exchangeable proposals with the same marginal distributions, i.e. $q_j=q_1$, for $j=1,\dots,K$. They put a particular emphasis on the situation where the proposals are multivariate normals with covariance Σ. In such a case, the auxiliary transition kernels ($\bar{q}_j, 1\le j\le K$) are easy to compute and to sample from (provided that Σ is positive definite). Several possible choices for the covariance of the proposal pool are discussed in [4]; among these choices, the so-called extreme antithetic proposal, which maximizes the expected value of the sum of the pairwise Euclidean distances among the members of the proposal pool, is the most appealing.

1.2. The multiple-try Metropolis algorithm with common random variables

We may alternatively generate all the proposals in the pool using the same random vector. Such a solution has been considered in [4], with the underlying idea of coupling the Metropolis algorithm with the quasi-Monte Carlo methods. However, common random numbers can be worthwhile in other settings, as will be shown below. The algorithm presented in [4] differs from the one proposed here in the way the acceptance ratio is computed. Another possibility consists in selecting a common search direction for all the proposals in the pool and proposing candidates along this search direction with different step sizes. This yields a version of the hit-and-run algorithm where the step sizes are chosen deterministically.

We assume in the sequel that the random variables $\{Y^j\}_{j=1}^K$ are distributed marginally according to transition kernels $\{q_j(x;\cdot)\}_{j=1}^K$, which may or may not be different. It is assumed that

(MTM-Ca) For all $i\in\{1,\dots,K\}$, $Y^i=\Psi^i(x,V)$ where V is a uniform random vector in $[0,1)^r$ and $(\Psi^i)_{i=1}^K$ are measurable functions, $\Psi^i:\mathsf{X}\times[0,1)^r\to\mathsf{X}$.

(MTM-Cb) For any $(i, j) \in \{1, \ldots, K\}^2$, there exists a measurable function $\Psi^{j,i} : \mathsf{X} \times \mathsf{X} \to \mathsf{X}$ such that

$$Y^i = \Psi^{j,i}(x, Y^j). \tag{5}$$

In words, all the proposals are sampled from the same random vector V using different transformations. The only constraint is that, given x and any member Y^j in the proposal pool, it is possible to reconstruct any other member Y^i. This is, in practice, a mild restriction. Supposing that the current state of the chain is $X = x$, one iteration of the multiple-try Metropolis algorithm with common random numbers (MTM-C) is defined as follows.

Algorithm 2 (*MTM-C*)**.**

(a) Draw a uniform random vector V in $[0, 1)^r$ and set $Y^i = \Psi^i(x, V)$ for $i = 1, \ldots, K$.
(b) Draw an index $J \in \{1, \ldots, K\}$, with probability proportional to

$$[w_1(Y^1, x), \ldots, w_K(Y^K, x)].$$

(c) Accept $Y = Y^J$ with probability $\bar{\alpha}^J(x, Y)$, where, for $j \in \{1, \ldots, K\}$,

$$\bar{\alpha}^j(x, y^j) = \alpha^j\left(x, \{\Psi^{j,i}(x, y^j)\}_{i=1}^K, \{\Psi^{j,i}(y^j, x)\}_{i\neq j}\right), \tag{6}$$

with α^j given in (3) and reject it otherwise.

Theorem 1. *Under assumptions* (MTM − Ca) *and* (MTM − Cb)*, the MTM-C algorithm described above satisfies the detailed balance condition and hence induces a reversible chain with stationary distribution* π.

Contrary to the MCTM algorithm, it is not required to draw a shadow sample: the acceptance ratio is therefore computationally simpler. Note that the MTM-C algorithm differs from the one proposed in [4], the latter using the same construction as the MCTM algorithm and therefore requiring to draw auxiliary variables. If $w_j(x, y) = \pi(x)$ and the transition densities are symmetric $q_j(x; y) = q_j(y; x)$, this expression boils down to

$$\bar{\alpha}^j(x, y) = 1 \wedge \frac{\sum_{i\neq j} \pi\left[\Psi^{j,i}(x, y)\right] + \pi(y)}{\sum_{i\neq j} \pi\left[\Psi^{j,i}(y, x)\right] + \pi(x)}. \tag{7}$$

We now give a specific instantiation of the MTM-C algorithm. In the sequel, it is assumed that $\mathsf{X} = \mathbb{R}^r$. We first draw a search direction $Z = \Phi^{-1}(V)$, where V has a uniform distribution over $(0, 1]^r$ and then move deterministically along this direction: $Y^j = x + \gamma^j Z$ for $j \in \{1, \ldots, K\}$, where $\{\gamma^j\}$ are deterministic step sizes chosen in the interval $[-\gamma, \gamma]$, with $\gamma > 0$ the size of the search interval. Compared to the hit-and-run algorithm outlined in [10, page 126], the step sizes are chosen deterministically in the interval $[-\gamma, \gamma]$ rather than being drawn at random (the randomized algorithm could also be analyzed in the present setting, but this extra randomization does not seem useful).

In this case, for $i \in \{1, \ldots, K\}$, the marginal kernels q_i are multivariate Gaussian with mean x and covariance $(\gamma^i)^2 \mathrm{I}_r$. The functions Ψ^i and $\Psi^{j,i}$ are given by $\Psi^i(x, v) = x + \gamma^i \Phi^{-1}(v)$ and $\Psi^{j,i}(x, y) = x + (\gamma^i/\gamma^j)(y - x)$.

2. Scaling analysis of the multiple correlated-try Metropolis algorithm

In this section, we consider the asymptotic behavior of MTM algorithms when the dimension, $T+1$, of the state space goes to infinity.

We focus on the case where $w(x, y) = \pi(x)$ and where the proposals are multivariate Gaussian. Consider the following assumptions:

(A1) The target density is an $(T+1)$-product density with respect to Lebesgue measure:

$$\pi_T(x_{0:T}) = \prod_{t=0}^{T} f(x_t), \quad \text{where } x_{0:T} \triangleq (x_0, \ldots, x_T). \tag{8}$$

The probability density function f is a positive twice continuously differentiable function, $[\ln f]''$ is Lipschitz bounded, and

$$\int f(x) \left|[\ln f]'(x)\right|^4 \mathrm{d}x < \infty.$$

We denote by $\left(\mathbf{X}_T[n] \triangleq \left(X_{T,t}[n]\right)_{t=0}^{T}, n \in \mathbb{N}\right)$ the sequence of Markov chains on $(\mathbb{R}^{T+1}, T \in \mathbb{N})$ defined by the MCTM algorithm (Algorithm 1) with target distribution π_T given in (8).

Define $\mathcal{F}_T = \left(\mathcal{F}_{T,n}, n \geq 0\right)$, the natural filtration of the Markov chain $\mathbf{X}_T$, i.e. for any $n \geq 0$,

$$\mathcal{F}_{T,n} \triangleq \sigma\left(\mathbf{X}_T[m], m = 0, \ldots, n\right). \tag{9}$$

For $j \in \{1, \ldots, K\}$ and Σ a positive definite matrix, consider the $(K-1) \times 1$ vector $\Sigma^{-j,j} = \left[\Sigma^{i,j}\right]_{i \neq j}$ obtained by extracting the j-th column of Σ and removing the j-th entry, and the $(K-1) \times (K-1)$ matrix $\Sigma^{-j,-j}$ obtained by deleting the j-th column and row of Σ. Denote

$$\mu^j(\Sigma) = \left[\mu^{j,i}(\Sigma)\right]_{i \neq j} = (\Sigma^{j,j})^{-1} \Sigma^{-j,j}, \tag{10}$$

$$\Delta^j(\Sigma) = \Sigma^{-j,-j} - \Sigma^{j,j} \mu^j(\Sigma) \left[\mu^j(\Sigma)\right]^T. \tag{11}$$

Provided that $(U^i)_{i=1}^K$ is a zero-mean Gaussian vector with covariance matrix Σ, which we denote $(U^i)_{i=1}^K \sim \mathcal{N}(0, \Sigma)$, the conditional distribution of the $(K-1) \times 1$ vector $(U^i)_{i \neq j}$ given the coordinate U^j, $j \in \{1, \ldots, K\}$, is Gaussian with mean $\mu^j(\Sigma) U^j$ and covariance matrix $\Delta^j(\Sigma)$.

Using these notations and assumptions, the n-th step of the MCTM algorithm can be formulated as follows:

Algorithm 3 (*MCTM with Product Target Density*).

1. Given the current state $\mathbf{X}_T[n]$ of the Markov chain at time n, a pool of proposals

$$\left(\mathbf{Y}_T^i[n+1]\right)_{i=1}^K \triangleq \left(Y_{T,t}^i[n+1], 0 \leq t \leq T\right)_{i=1}^K,$$

is generated according to

$$Y_{T,t}^i[n+1] = X_{T,t}[n] + T^{-1/2} U_t^i[n+1], \quad 0 \leq t \leq T,\ 1 \leq i \leq K, \tag{12}$$

where
(a) for any $t \in \{0, \dots, T\}$, $(U_t^i[n+1])_{i=1}^K \sim \mathcal{N}(0, \Sigma)$.
(b) The $T+1$ random vectors $\{(U_t^1[n+1], \dots, U_t^K[n+1])\}_{t=0}^T$ are independent conditionally to $\mathcal{F}_{T,n}$, where $\mathcal{F}_{T,n}$ is defined in (9).

2. An index $J_T[n+1]$ is drawn independently of $\mathcal{F}_{T,n}$ from a multinomial distribution with parameters proportional to

$$\left[\pi_T\left(\mathbf{Y}_T^1[n+1]\right), \dots, \pi_T\left(\mathbf{Y}_T^K[n+1]\right)\right].$$

3. Given the proposal pool $\left(\mathbf{Y}_T^i[n+1]\right)_{i=1}^K$, auxiliary variables

$$\left(\left(\tilde{\mathbf{Y}}_T^{j,i}[n+1]\right)_{i\neq j}\right)_{j=1}^K = \left(\left(\tilde{Y}_{T,t}^{j,i}[n+1], 0 \le t \le T\right)_{i\neq j}\right)_{j=1}^K,$$

are generated according to:

$$\tilde{Y}_{T,t}^{j,i}[n+1] = X_{T,t}[n] + T^{-1/2}\tilde{U}_t^{j,i}[n+1], \quad i \neq j, \tag{13}$$

$$\tilde{U}_t^{j,i}[n+1] = \left[1 - \mu^{j,i}(\Sigma)\right] U_t^j[n] + \tilde{W}_t^{j,i}[n+1], \tag{14}$$

where
(a) For any $1 \le j \le K$ and $0 \le t \le T$, $(\tilde{W}_t^{j,i}[n+1])_{i\neq j} \sim \mathcal{N}\left(0, \Delta^j(\Sigma)\right)$ where $\Delta^j(\Sigma)$ is defined in (11).
(b) The $T+1$ random vectors $\left\{(\tilde{W}_t^{j,i}[n+1])_{i\neq j}\right\}_{t=0}^T$ are independent conditionally to $\mathcal{F}_{T,n}$.
(c) $\left\{(U_t^i[n+1], 1 \le i \le K)\right\}_{t=0}^T$ and $\left\{(\tilde{W}_t^{j,i}[n+1])_{i\neq j}\right\}_{t=0}^T$ are independent.

4. Given $J_T[n+1] = j$, the proposal $\mathbf{Y}_T^{J_T[n+1]}[n+1]$ is then accepted with probability (see (4))

$$\alpha_T^j\left(\mathbf{X}_T[n], \left(\mathbf{Y}_T^i[n+1]\right)_{i=1}^K, \left(\tilde{\mathbf{Y}}_T^{j,i}[n+1]\right)_{i\neq j}\right),$$

where

$$\alpha_T^j\left(\mathbf{x}_T, \left(\mathbf{y}_T^i\right)_{i=1}^K, \left(\tilde{\mathbf{y}}_T^{j,i}\right)_{i\neq j}\right) = 1 \wedge \frac{\sum_{i=1}^K \pi_T\left[\mathbf{y}_T^i\right]}{\sum_{i=1}^{K-1} \pi_T\left[\tilde{\mathbf{y}}_T^{j,i}\right] + \pi_T(\mathbf{x}_T)}. \tag{15}$$

The conditional probability of selecting the j-th member of the proposal set and accepting it is equal to $\alpha_T^j(\mathbf{X}_T[n])$, where

$$\alpha_T^j(\mathbf{x}_T) = \mathbb{E}\left[A\left\{L_{0,T}(\mathbf{x}_T, \mathbf{U}_T^j), \left(L_{0,T}(\mathbf{x}_T, \mathbf{U}_T^i)\right)_{i\neq j}, \left(L_{0,T}(\mathbf{x}_T, \tilde{\mathbf{U}}_T^{j,i})\right)_{i\neq j}\right\}\right], \tag{16}$$

with

$$A\left(u, (v^i)_{i=1}^{K-1}, (w^i)_{i=1}^{K-1}\right) \triangleq \frac{\mathrm{e}^u}{\mathrm{e}^u + \sum_{i=1}^{K-1} \mathrm{e}^{v^i}} \wedge \frac{\mathrm{e}^u}{1 + \sum_{i=1}^{K-1} \mathrm{e}^{w^i}}, \tag{17}$$

$$L_{s,T}(x_{s:T}, u_{s:T}) = \sum_{t=s}^T \left\{\ln f(x_t + T^{-1/2}u_t) - \ln f(x_t)\right\}, \quad 0 \le s \le T. \tag{18}$$

Denote by ζ_T the projection on the first coordinate, that is $\zeta_T : \mathbb{R}^{T+1} \to \mathbb{R}$ such that $\zeta_T(x_{0:T}) = x_0$. Consider the progressive cadlag process $W_T \triangleq (W_T[s], s \in \mathbb{R}^+)$

$$s \mapsto W_T[s] = \zeta_T\left[\mathbf{X}_T\left[\lfloor Ts \rfloor\right]\right]. \tag{19}$$

Weak convergence in the Skorokhod topology is denoted by $\Rightarrow$ and the standard Brownian process is denoted by $\left(B[s], s \in \mathbb{R}^+\right)$.

For any integer L, define by $\mathcal{C}_L^+$ the set of positive symmetric $L \times L$ matrices. For Γ in $\mathcal{C}_{2K-1}^+$, let

$$\alpha(\Gamma) = \mathbb{E}\left[A\left\{\left(G^i - \mathbb{V}\mathrm{ar}[G^i]/2\right)_{i=1}^{2K-1}\right\}\right], \tag{20}$$

where A is defined in (17) and $(G^i)_{i=1}^{2K-1} \sim \mathcal{N}(0, \Gamma)$. Let $\mathcal{I}$ be the Fisher information quantity

$$\mathcal{I} = \int f(x)\left\{[\ln f]'(x)\right\}^2 \mathrm{d}x. \tag{21}$$

For $(\Gamma^1, \ldots, \Gamma^K)$ in $\mathcal{C}_{2K-1}^+ \times \cdots \times \mathcal{C}_{2K-1}^+$, denote

$$\lambda\left(\mathcal{I}, (\Gamma^j)_{j=1}^K\right) \triangleq \sum_{j=1}^K \Gamma_{1,1}^j \cdot \alpha\left[\mathcal{I}\Gamma^j\right], \tag{22}$$

where $\Gamma^j = (\Gamma_{k,\ell}^j)_{0 \le k,\ell \le 2K-1}$.

Theorem 2. *Assume* (A1) *and consider Algorithm 3. Suppose that $\mathbf{X}_T[0]$ is distributed according to the target density π_T. Then, the process W_T defined in (19) weakly converges in the Skorokhod topology to the stationary solution $(W[s], s \in \mathbb{R}^+)$ of the Langevin SDE*

$$\mathrm{d}W[s] = \lambda^{1/2}\mathrm{d}B[s] + \frac{1}{2}\lambda\left[\ln f\right]'(W[s])\mathrm{d}s,$$

with $\lambda \triangleq \lambda\left(\mathcal{I}, (\Gamma^j(\Sigma))_{j=1}^K\right)$, where $\Gamma^j(\Sigma)$, $1 \le j \le K$ denotes the covariance matrix of the random vector $(U_0^j, (U_0^i)_{i \ne j}, (\tilde{U_0}^{j,i})_{i \ne j})$ defined in (12) and (14).

In addition, $\alpha\left[\mathcal{I}\Gamma^j\right]$ is the limit as $T \to \infty$ of the expected acceptance rate of the j-th component in stationarity

$$\alpha\left[\mathcal{I}\Gamma^j\right] = \lim_{T \to \infty} \int \cdots \int \alpha_T^j(\mathbf{x}_T)\pi_T(\mathrm{d}\mathbf{x}_T), \tag{23}$$

where α_T^j is the mean acceptance rate of the j-th component given in (16).

Proof. The proof of this result is postponed to Appendix. □

3. Scaling analysis of the multiple-try Metropolis algorithm with common random numbers

We now turn to the scaling analysis of the MTM-C algorithm. Consider the following assumption

(A2) For any $i \in \{1, \ldots, K\}$, $\varphi^i : [0, 1] \to \mathbb{R}$ is the quantile function of a symmetric distribution on $\mathbb{R}$. In addition, φ^i is invertible and $\int_0^1 |\varphi^i(v)|^3 \mathrm{d}v < \infty$.

The MTM-C algorithm defines a sequence of Markov chains

$$\left(\mathbf{X}_T[n] = \left(X_{T,t}[n]\right)_{t=0}^{T}, n \in \mathbb{N}\right)$$

on the sequence of state spaces $(\mathbb{R}^{T+1}, T \in \mathbb{N})$ as follows. Define

$$\psi_T^{j,i}(x, y) = x + T^{-1/2}\varphi^i \circ (\varphi^j)^{-1}\left[T^{1/2}(y - x)\right].$$

Algorithm 4 (*MTM-C with Product Target Density*)**.**

1. Given the current state $\mathbf{X}_T[n]$ of the Markov chain at time n, a family of proposals
$$\left(\mathbf{Y}_T^i[n+1]\right)_{i=1}^{K} \triangleq \left(Y_{T,t}^i[n+1], 0 \le t \le T\right)_{i=1}^{K},$$
is generated according to
$$Y_{T,t}^i[n+1] = x_t + T^{-1/2}U_t^i[n+1], \quad U_t^i[n+1] = \varphi^i(V_t[n+1]), \tag{24}$$
where $(V_t[n+1])_{t=0}^{T}$ is a vector of uniform random variables on $[0,1]^{T+1}$ independent of $\mathcal{F}_{T,n} \triangleq \sigma\left(\mathbf{X}_T[m], m = 0, \ldots, n\right)$.
2. An index $J_T[n+1]$ is drawn independently of $\mathcal{F}_{T,n}$ from a multinomial distribution with parameters proportional to
$$\left[\pi_T\left(\mathbf{Y}_T^1[n+1]\right), \ldots, \pi_T\left(\mathbf{Y}_T^K[n+1]\right)\right].$$
3. Given the proposal pool $(Y_{T,t}^i[n+1])_{i=1}^{K}$, auxiliary variables
$$\left(\left(\tilde{\mathbf{Y}}_T^{j,i}[n+1]\right)_{i\neq j}\right)_{j=1}^{K} \triangleq \left(\left(\tilde{Y}_{T,t}^{j,i}, 0 \le t \le T\right)_{i\neq j}\right)_{j=1}^{K}$$
are constructed according to
$$\tilde{Y}_{T,t}^{j,i}[n+1] = \psi_T^{j,i}(Y_{T,t}^j[n+1], X_{T,t}[n]) = X_{T,t}[n] + T^{-1/2}\tilde{U}_t^{j,i}[n+1],$$
where $\left(\tilde{U}_t^{j,i}[n+1], 0 \le t \le T\right)_{i\neq j}$ is defined as:
$$\tilde{U}_t^{j,i}[n+1] = U_t^j[n+1] + \varphi^i \circ (\varphi^j)^{-1}\left[-U_t^j[n+1]\right]. \tag{25}$$
4. Given $J_T[n+1] = j$, the proposal $\mathbf{Y}_T^j[n+1]$ is then accepted with probability (see (4))
$$\alpha_T^j\left(\mathbf{X}_T[n], \left(\mathbf{Y}_T^i[n+1]\right)_{i=1}^{K}, \left(\tilde{\mathbf{Y}}_T^{j,i}[n+1]\right)_{i\neq j}\right)$$
where α_T^j is defined in (15).

The conditional probability of selecting and accepting the j-th member of the proposal set is equal to $\bar{\alpha}_T^j(\mathbf{X}_T[n])$, where

$$\bar{\alpha}_T^j(\mathbf{x}_T) = \mathbb{E}\left[A\left\{L_{0,T}(\mathbf{x}_T, \mathbf{U}_T^j), \left(L_{0,T}(\mathbf{x}_T, \mathbf{U}_T^i)\right)_{i\neq j}, (L_{0,T}(\mathbf{x}_T, \tilde{\mathbf{U}}_T^{j,i}))_{i\neq j}\right\}\right],$$

with A and $L_{0,T}$ as defined in (A.27) and (18), and $\mathbf{U}_T^i$ and $\tilde{\mathbf{U}}_T^{j,i}$ as defined in (24) and (25).

Theorem 3. *Assume* (A1) *and* (A2). *Consider the MTM-C algorithm given in Algorithm 4. Suppose that* $\mathbf{X}_T[0]$ *is distributed according to the target density* π_T. *Then, the process* W_T *defined in (19) weakly converges in the Skorokhod topology to the stationary solution* $(W[s], s \in \mathbb{R}^+)$ *of the Langevin SDE*

$$\mathrm{d}W[s] = \lambda^{1/2}\mathrm{d}B[s] + \frac{1}{2}\lambda \left[\ln f\right]' (W[s])\mathrm{d}s,$$

where $\lambda = \lambda\left(\mathcal{I}, (\Gamma^j)_{j=1}^K\right)$ *and* α *are defined in (22) and (20) respectively, and* Γ^j *is the covariance matrix of the random vector* $(U_0^j, (U_0^i)_{i\neq j}, (\tilde{U_0}^{j,i})_{i\neq j})$ *defined in (24) and (25).*

In addition, $\alpha\left(\mathcal{I}\Gamma^j\right)$ *is the limit as* $T \to \infty$ *of the expected acceptance rate of the* j*-th component in stationarity*

$$\alpha\left(\mathcal{I}\Gamma^j\right) = \lim_{T\to\infty} \int \cdots \int \bar{\alpha}_T^j(\mathbf{x}_T)\pi_T(\mathrm{d}\mathbf{x}_T).$$

Proof. This proof follows the same lines as that of Theorem 2 and is thus omitted for the sake of brevity. □

We now consider a special case of the MTM-C algorithm, namely the hit-and-run algorithm.

The hit-and-run algorithm: Denoting by $\{\gamma^i\}_{i=1}^K$ a sequence of numbers in the interval $[-\gamma, \gamma]$, where $\gamma > 0$, we define, for $i \in \{1, \ldots, K\}$:

$$\varphi^i : v \mapsto \varphi^i(v) = \gamma^i \Phi^{-1}(v), \quad v \in [0, 1].$$

In this case, $Y_{T,t}^i = x_t + T^{-1/2}\gamma^i \Phi^{-1}(V_t)$ are Gaussian with mean x_t and variance $(\gamma^i)^2$. For $y \in \mathbb{R}$, the inverse of φ^i is given by

$$(\varphi^i)^{-1} : y \mapsto (\varphi^i)^{-1}(y) = \Phi(y/\gamma^i). \tag{26}$$

4. Discussion

The limiting processes in Theorems 2 and 3 may be expressed as a time-scaled version $(V[\lambda s], s \geq 0)$ of the stationary solution of a Langevin diffusion $(V[s], s \geq 0)$

$$\mathrm{d}V[s] = \mathrm{d}B[s] + \frac{1}{2}\left[\ln f\right]' (V[s])\mathrm{d}s.$$

Since the speed λ is the only quantity that depends on the proposal construction, all possible efficiency criteria become asymptotically equivalent as T goes to infinity under (A1); see [11, 13]. Therefore, a natural criterion to optimize is the speed.

4.1. The MCTM algorithm

We optimize the speed $\lambda \triangleq \lambda(\mathcal{I}, (\Gamma^j(\Sigma))_{j=1}^K)$ over a subset $\mathcal{G}$ of $\mathcal{C}_K^+$. The choice of $\mathcal{G}$ has a direct impact on the complexity of the resulting algorithm. The following choices are considered:

- $\mathcal{G} = \{\Sigma = \ell^2 \mathrm{I}_K, \ell \in \mathbb{R}\}$: only the global scale of the proposal is adjusted but the proposals are made independently, which is the default option for the MTM algorithm;
- $\mathcal{G} = \left\{\Sigma = \mathrm{diag}(\ell_1^2, \ldots, \ell_K^2), (\ell_1, \ldots, \ell_K) \in \mathbb{R}^K\right\}$: the proposals have different scales but are independent.

Table 1
Optimal scaling constants, value of the speed (when $\mathcal{I} = 1$), and mean acceptance rate for independent proposals.

K	1	2	3	4	5
$\ell^\star$	2.38	2.64	2.82	2.99	3.12
$\lambda^\star$	1.32	2.24	2.94	3.51	4.00
$a^\star$	0.23	0.32	0.37	0.39	0.41

Table 2
Optimal scaling constants, value of the speed (when $\mathcal{I} = 1$), and mean acceptance rate for extreme antithetic proposals.

K	1	2	3	4	5
$\ell^\star$	2.38	2.37	2.64	2.83	2.99
$\lambda^\star$	1.32	2.64	3.66	4.37	4.91
$a^\star$	0.23	0.46	0.52	0.54	0.55

- $\mathcal{G} = \left\{\Sigma = \ell^2 \Sigma_a, \ell^2 \in \mathbb{R}\right\}$, where Σ_a is the extreme antithetic covariance matrix:
$$\Sigma_a \triangleq \frac{K}{K-1}\mathrm{I}_K - \frac{1}{K-1}\mathbb{1}_K \mathbb{1}_K^T$$
with $\mathbb{1}_K = (1, \ldots, 1)^T$.

We also consider the case where $\mathcal{G} = \mathcal{C}_K^+$.

Consider first the case where $\mathcal{G}$ is chosen to be the subset of diagonal positive matrices. The first interesting result is that the optimum is obtained in the class of covariance matrices which are not only diagonal, but proportional to the identity matrices: it is optimal to let the variables in the proposal pool be exchangeable.

Proposition 4. *For any $K \geq 2$,*
$$\max_{(\ell_1^2,\ldots,\ell_K^2)>0} \lambda\left(\mathcal{I}, \left(\Gamma^j\left(\mathrm{diag}(\ell_1^2, \ldots, \ell_K^2)\right)\right)_{j=1}^K\right) = \max_{\ell^2>0} \lambda\left(\mathcal{I}, \left(\Gamma^j\left(\ell^2 \mathrm{I}_K\right)\right)_{j=1}^K\right).$$

Proof. The proof is omitted for brevity. □

The optimal values $\ell^\star$ of the scale ℓ for different values of K, the associated values of the speed (when $\mathcal{I} = 1$), and the average acceptance probabilities are summarized in Table 1. Not surprisingly, the optimal value of ℓ and the average acceptance rate increase with K. As the pool of proposals grows, optimal efficiency is thus attained by finding a compromise between more aggressive candidates and candidates that are accepted on average more frequently than with smaller values of K.

We then consider the extreme antithetic proposals of [4]. The optimal values $\ell^\star$ for different values of K, the associated values of the speed (when $\mathcal{I} = 1$), and the average acceptance probabilities are summarized in Table 2. The improvement is very significant, especially when going from $K = 1$ to $K = 2$. In this case, the speed of the algorithm is almost multiplied by a factor of 2. It is interesting to note that the optimal scales for the extreme antithetic proposals are smaller than the optimal scales for the independent proposals, but the mean acceptance rates and the resulting speeds are higher. Under this scheme, it is more rewarding to accept a large proportion of candidates than to favor aggressive candidates. We could thus say that the correlation structure present in the proposal pool yields high-quality candidates; the scale ℓ only needs minor adjustments as K grows.

Table 3
Optimal scaling constants, value of the speed (when $\mathcal{I} = 1$), and mean acceptance rate for the optimal covariance.

K	1	2	3	4	5
$\ell^\star$	2.38	2.37	2.66	2.83	2.98
$\lambda^\star$	1.32	2.64	3.70	4.40	4.93
$a^\star$	0.23	0.46	0.52	0.55	0.56

Table 4
Optimal scaling constants, value of the speed (when $\mathcal{I} = 1$), and mean acceptance rate for the hit-and-run algorithm.

K	1	2	4	6	8
$\ell^\star$	2.38	2.37	7.11	11.85	16.75
$\lambda^\star$	1.32	2.64	2.65	2.65	2.65
$a^\star$	0.23	0.46	0.46	0.46	0.46

We finally consider the unconstrained optimization of the covariance matrix. The speed is a highly non-linear function of the covariance of the proposals and the results in Table 3 have been obtained by numerical optimization. There is no significant improvement in the speed of convergence with respect to the extreme antithetic sampling which is therefore our preferred solution.

4.2. MTM-C

We consider the multiple-try hit-and-run algorithm (MTM-HR) with regularly spaced step sizes $(\gamma^i)_{i=1}^K$ in $[-\ell, \ell]$. The improvement from $K = 1$ to 2 is the same as in the extreme antithetical case and there is no further improvement for larger K. Nevertheless, the case $K = 2$ is very interesting since the implementation of the MTM-HR algorithm requires simulating only one random variable per iteration, and the overhead introduced by the evaluation of the $2K - 1$ likelihood functions is generally modest in this setting. The results in Table 4 justify a deeper analysis of this algorithm as they might seem intriguing at first sight.

When going from $K = 1$ to $K = 2$, the improvement in the MTM-HR is more important than in the MTM with independent proposals. In a trial set of size 2, there is a higher chance of getting at least one good candidate when moving in opposite directions. If, for instance, the MTM-HR proposes an unlikely candidate (e.g. in the tails of the target), a second candidate in the opposite direction will likely move towards higher values for the target density and will result in a more easily accepted candidate for the chain. When the second candidate is chosen independently from the first one (MTM with independent proposals), the direction in which it is proposed might be as bad as the first one. The MTM-HR with $K = 2$ thus reduces the risk of generating two bad candidates in the same proposal set.

Applying the MTM-HR with $K > 2$ is not worthwhile under the framework considered. For $K = 2$, the optimal scale is $\ell^* = 2.37$ and the optimal acceptance rate is $a^* = 0.46$. When $K = 4$, the trial values are still generated along a common search direction (according to deterministic step sizes). If we let $\ell = 2.37$ as before, candidates closer to the current value of the chain are included in the proposal set, which automatically increases the acceptance rate while reducing the speed (otherwise, this would contradict the fact that $\ell = 2.37$ is optimal for $K = 2$). In order for the MTM-HR to remain as efficient as for $K = 2$, it is thus necessary to preserve the optimal acceptance rate of 0.46 by letting ℓ^* increase with K.

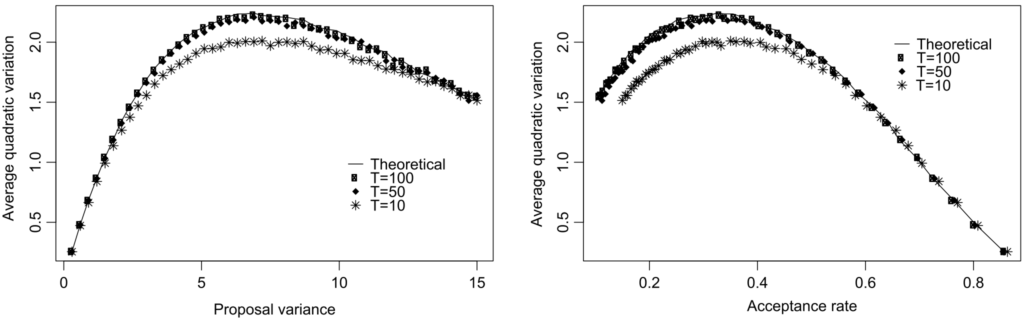

Fig. 1. AQV as a function of the proposal variance (left) and the acceptance rate (right) for the MTM algorithm with a normal target and $K = 2$ independent proposals. The plotted symbols are the results of simulation studies in different dimensions, while the solid lines represent the theoretical curves.

4.3. Multiple-try with importance weights

In this paper, we focused on the analysis of multiple-try algorithms with $w_j(x, y) = \pi(x)$, $j \in \{1, \ldots, K\}$: candidates are selected from the pool of proposals with probability proportional to the target density. Alternative choices for the weights are possible; an appealing option is given by the importance weights, $w_j(x, y) = \pi(x)/q_j(y_j; x)$, $j \in \{1, \ldots, K\}$ (see [10]). This weight function places higher probability on selecting proposals that are further away from the current state of the chain.

Based on some simulation studies, it would seem that this weight function does not perform as well as the former for target densities satisfying Assumption 1. In the case of the MTM algorithm with independent proposals, we found that the speed (when $\mathcal{I} = 1$) increases from 1.32 ($K = 1$) to approximately 2 ($K = 5$). The optimal scale ℓ^* also increases with K, but the optimal acceptance rate a^* decreases (from 0.23 to about 0.15 when $K = 5$). When the size of the proposal set grows, candidates become considerably more aggressive, resulting in decreasing optimal acceptance rates.

Although the importance weights might yield good performances in specific situations, the numerical results obtained were more convincing when using the weights that are proportional to the target density.

4.4. Numerical examples

We validate the conclusions stated above with a simulation study. Let f be the standard normal density; π_T is thus a multivariate normal distribution with a diagonal covariance matrix. This toy example is popular for validating optimal scaling results; in this setting, $\mathcal{I} = 1$.

We first consider the MTM algorithm with $K = 2$ independent proposals. The graphs in Fig. 1 display the average quadratic variation (AQV) as a function of the proposal variance and of the average acceptance rate. The AQV is a convenient measure as it is a function of the sample obtained only, *i.e.* it is independent of the specific estimates that we might be interested in obtaining from the Markov chain. It is computed as $\sum_{i=1}^{T} \sum_{j=1}^{N} \left(X_{T,i}[j] - X_{T,i}[j-1] \right)^2 / NT$, where N is the number of iterations performed (see [12]). The solid lines in each graph represent the speed of the limiting diffusion as a function of the proposal variance and the global average acceptance probability respectively, while the dotted lines represent the corresponding AQV curves obtained by running the MTM algorithm in various dimensions. Each dotted curve is

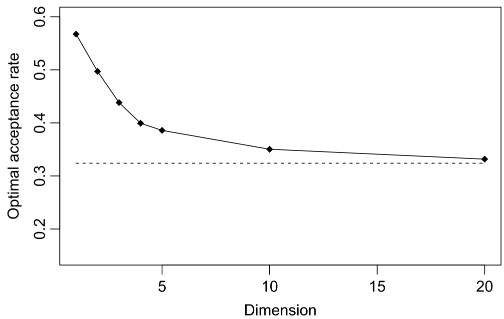

Fig. 2. Optimal acceptance rate of the MTM algorithm with $K = 2$ as a function of the dimension T.

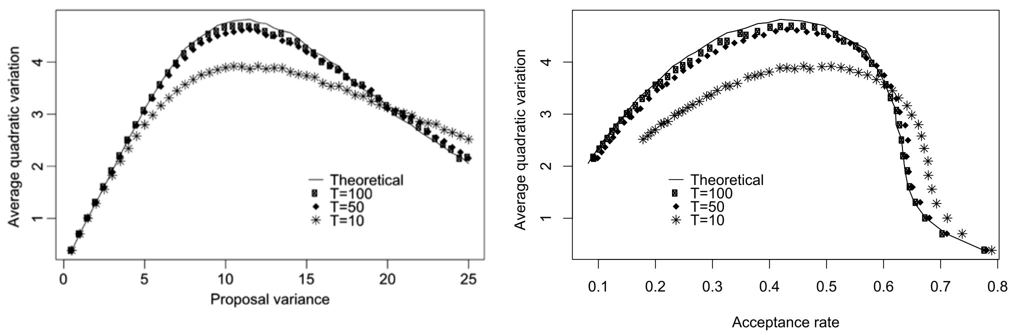

Fig. 3. AQV as a function of the proposal variance (left) and the acceptance rate (right) for the MTM algorithm with a normal target and $K = 7$ independent proposals. The plotted symbols are the results of simulation studies in different dimensions, while the solid lines represent the theoretical curves.

produced by running 50 replications of the MTM algorithm in a given dimension T; the points in a given curve are the results of 10^5-iteration runs with different proposal variances. We also estimated the acceptance rate by the proportion of accepted moves in the algorithm.

Both graphs highlight the fact that low-dimensional algorithms behave similarly to higher dimensional ones. Fig. 2 corroborates this conclusion by showing the relationship between optimal acceptance rates and the dimension of the MTM algorithm with independent proposals when $K = 2$. Convergence towards the asymptotically optimal acceptance rate of 0.32 happens rapidly. We have repeated the same experiment for the MTM algorithm with $K = 7$ independent proposals (see Fig. 3). The curve of the 10-dimensional algorithm is not as close to the asymptotic curve as it was for $K = 2$. For larger values of K, we must then be cautious about the dimensionality of the problem considered when using the available optimal scaling results. Similar graphs may be obtained for the MCTM and MTM-C algorithms, and were thus omitted for brevity.

4.5. Conclusion of the discussion

It is difficult to draw general conclusions without taking into account a precise expression for the target density and the cost of simulations. Nevertheless, the above results provide us with some useful guidelines.

In the asymptotic theory considered here, it appears that the extreme antithetic proposals improve upon the MTM with independent proposals. Although the introduction of correlation makes the computation of the acceptance ratio more complex, this algorithm might be more efficient than the MTM with independent proposals in some practical implementations, but a definitive answer to this question is not available from the theory derived in this paper.

The advantage of the MTM-C algorithms stems from the fact that only one simulation is required for obtaining the pool of proposals and auxiliary variables. In many statistical models, the evaluation of the likelihood at $2K - 1$ points is much simpler for the MTM-HR algorithm because the proposals are along the same direction. In particular, the case $K = 2$ induces a speed which is twice that of the Metropolis algorithm whereas the computational cost is almost the same in many scenarios.

Acknowledgments

The first author's work is supported by the National Sciences and Engineering Research Council of Canada. The second and third authors' work is supported by the Agence Nationale de la Recherche (ANR, 212, rue de Bercy 75012 Paris) through the 2009–2012 project Big MC.

Appendix. On scaling approximations

Scaling approximations have been introduced in the MCMC literature by [11] (see [1] and the references therein). In this section, we extend these results for scaling analysis of a general class of random-walk type MCMC algorithms involving auxiliary random variables and covering both the MCTM and the MTM-C algorithms (and, presumably, most of the MCMC algorithms using auxiliary random variables in a symmetric random walk framework).

A.1. Convergence to a continuous-time Markov process

In this first section, we derive some general results on the convergence to a one-dimensional Markov process. In what follows, the set $\mathbb{R}^{T+1}$ is equipped with its Borel σ-field; $x_{0:T}$ denotes $x_{0:T} \triangleq (x_0, \ldots, x_T) \in \mathbb{R}^{T+1}$. Denote by $\zeta_T : \mathbb{R}^{T+1} \to \mathbb{R}$ the projection on the first coordinate, that is $\zeta_T(x_{0:T}) = x_0$. For any function $h : \mathbb{R} \to \mathbb{R}$, define the function $P_T h : \mathbb{R}^{T+1} \to \mathbb{R}$ by

$$P_T h : x_{0:T} \mapsto P_T h(x_{0:T}) = h(x_0) = h \circ \zeta_T(x_{0:T}). \tag{A.1}$$

Let $(\mathbf{X}_T[n],\ n \in \mathbb{N})_{T \geq 1}$ be a sequence of homogeneous Markov chains taking values in $\mathbb{R}^{T+1}$ with transition kernel Q_T. For all $s \geq 0$, denote the continuous-time process

$$s \mapsto W_T[s] = \zeta_T\left(\mathbf{X}_T\left[\lfloor Ts \rfloor\right]\right), \tag{A.2}$$

obtained by speeding up by a factor T the first coordinate of the Markov chain $\mathbf{X}_T$. Whereas $\left(W_T[s],\ s \in \mathbb{R}^+\right)$ is not itself a Markov process, it is a progressive $\mathbb{R}$-valued process and the aim of this section is to establish that $\left(W_T[s],\ s \in \mathbb{R}^+\right)$ converges in the Skorokhod topology to some Markov process under some general assumptions that will be stated below. In subsequent sections, we specialize these general results to multiple-try Metropolis algorithms.

Define $G_T = T(Q_T - I)$ and denote by C_c^∞ the set of compactly supported indefinitely continuously differentiable functions defined on $\mathbb{R}$. Let $\{\mathsf{F}_T\}_{T \geq 0}$ be a sequence of Borel subsets of $\mathbb{R}^T$ and consider the following assumptions:

(B1) For all $T \in \mathbb{N}$, the transition kernel Q_T has a unique stationary distribution denoted by π_T. Moreover, for any Borel non negative function h on $\mathbb{R}$,

$$\pi_T(P_T h) = \int h(x_0) f(x_0) \mathrm{d}x_0, \tag{A.3}$$

where $f : \mathbb{R} \to \mathbb{R}$ is a probability density function.

(B2) $\lim_{T\to\infty} \pi_T\left(\mathbb{R}\times \mathsf{F}_T\right) = 1$.
(B3) There exists $p > 1$ such that for any $h \in C_c^\infty$,

$$\sup_{T\geq 0}\int \sup_{x_{1:T}\in\mathbb{R}^T} \left|G_T\left[P_T h\right](x_{0:T})\right|^p f(x_0)\mathrm{d}x_0 < \infty.$$

(B4) There exists a Markov process $(W[s], s \in \mathbb{R}^+)$ with cadlag sample paths and (infinitesimal) generator G such that C_c^∞ is a core for G and for any $h \in C_c^\infty$,

$$\lim_{T\to\infty}\int \sup_{x_{1:T}\in\mathsf{F}_T} \left|G_T\left[P_T h\right](x_{0:T}) - Gh(x_0)\right| f(x_0)\mathrm{d}x_0 = 0.$$

Theorem 5. *Assume* (B1)–(B4). *Then,* $W_T \Longrightarrow W$ *in the Skorokhod topology where* $W[0]$ *is distributed according to* f.

The statement of this theorem tailored to the analysis of MCMC algorithms is to the best of our knowledge original. The proof is obtained by assembling several results presented in [6]. We preface the proof with the following lemma:

Lemma 6. *Assume* (B1)–(B4). *Then,*

$$\lim_{T\to\infty}\int\cdots\int \left|G_T\left[P_T h\right](x_{0:T}) - Gh(x_0)\right| \pi_T(\mathrm{d}x_{0:T}) = 0. \tag{A.4}$$

Proof. Consider the following decomposition

$$\begin{aligned}&\pi_T\left|G_T[P_T h] - P_T Gh\right| \\ &\quad = \pi_T\left|\mathbb{1}_{\mathbb{R}\times\mathsf{F}_T}\left(G_T[P_T h] - P_T Gh\right)\right| + \pi_T\left|\mathbb{1}_{(\mathbb{R}\times\mathsf{F}_T)^c}\left(G_T[P_T h] - P_T Gh\right)\right|.\end{aligned} \tag{A.5}$$

We will show that each term of the right-hand side converges to 0 as T tends to infinity. Note that by (B1),

$$\pi_T\left|\mathbb{1}_{\mathbb{R}\times\mathsf{F}_T}\left(G_T[P_T h] - P_T Gh\right)\right| \leq \int \Delta_T[\mathsf{F}_T](x_0) f(x_0)\mathrm{d}x_0,$$

where we set, for any $A \subset \mathbb{R}^T$,

$$\Delta_T[A](x_0) \triangleq \sup_{x_{1:T}\in A} \left|G_T[P_T h](x_{0:T}) - Gh(x_0)\right|.$$

(B4) then implies $\lim_{T\to\infty} \pi_T\left|\mathbb{1}_{\mathbb{R}\times\mathsf{F}_T}\left(G_T[P_T h] - P_T Gh\right)\right| = 0$. We now turn to the second term in the right-hand side of Eq. (A.5). For any $M > 0$ and $p > 1$,

$$\begin{aligned}&\pi_T\left|\mathbb{1}_{(\mathbb{R}\times\mathsf{F}_T)^c}\left(G_T[P_T h] - P_T Gh\right)\right| \\ &\quad \leq \int\cdots\int \mathbb{1}\{x_{1:T}\notin \mathsf{F}_T\}\Delta_T[\mathbb{R}^T](x_0)\pi_T(\mathrm{d}x_{0:T}) \\ &\quad \leq \int \Delta_T[\mathbb{R}^T](x_0)\mathbb{1}\{\Delta_T(\mathbb{R}^T)(x_0) \geq M\} f(x_0)\mathrm{d}x_0 + M\pi_T((\mathbb{R}\times\mathsf{F}_T)^c) \\ &\quad \leq \frac{\sup_T \int \left|\Delta_T[\mathbb{R}^T](x_0)\right|^p f(x_0)\mathrm{d}x_0}{M^{p-1}} + M\pi_T((\mathbb{R}\times\mathsf{F}_T)^c).\end{aligned} \tag{A.6}$$

Provided that

$$\sup_T \int \left|\Delta_T[\mathbb{R}^T](x_0)\right|^p f(x_0)\mathrm{d}x_0 < \infty, \tag{A.7}$$

the first term in the right-hand side of (A.6) can be taken arbitrarily small by choosing M sufficiently large. Then, using $\lim_{T\to\infty} \pi_T((\mathbb{R}\times \mathsf{F}_T)^c) = 0$, we finally obtain

$$\lim_{T\to\infty} \pi_T \left|\mathbb{1}_{(\mathbb{R}\times\mathsf{F}_T)^c} (G_T[P_T h] - P_T G h)\right| = 0,$$

which completes the proof. It now remains to check (A.7) where p is defined as in Assumption (B3). Actually, considering (B3), we only need to show that

$$\int |G[h](x_0)|^p f(x_0)\mathrm{d}x_0 < \infty. \tag{A.8}$$

Let $X_0 \sim f$. Assumption (B4) yields $\lim_T \mathbb{E}\left[\Delta_T[\mathsf{F}_T](X_0)\right] = 0$ so that the sequence of nonnegative random variables $(\Delta_T[\mathsf{F}_T](X_0))_{T\geq 1}$ converges in probability to 0. This implies the existence of a (deterministic) sequence of integers (T_k) tending to infinity such that

$$\lim_{k\to\infty} \Delta_T[\mathsf{F}_{T_k}](X_0) = 0, \quad a.s.$$

Then, considering the definition of $\Delta_T[\mathsf{F}_{T_k}]$, there exist triangular arrays of random vectors $(X_{T_k,1:T_k})_{k\geq 1}$ taking values in $\mathbb{R}^{T_k}$ such that

$$G[h](X_0) = \lim_{k\to\infty} G_{T_k}[P_{T_k} h]([X_0, X_{T_k,1:T_k}]), \quad a.s.$$

Finally, by Fatou's lemma and (B3),

$$\begin{aligned}
\int |G[h](x_0)|^p f(x_0)\mathrm{d}x_0 &= \mathbb{E}\left[|G[h](X_0)|^p\right] = \mathbb{E}\left[\liminf_{k\to\infty} |G_{T_k}[P_{T_k} h]([X_0, X_{T_k,1:T_k}])|^p\right] \\
&\leq \liminf_{k\to\infty} \mathbb{E}\left[|G_{T_k}[P_{T_k} h]([X_0, X_{T_k,1:T_k}])|^p\right] \\
&\leq \sup_{T\geq 0} \int \sup_{x_{1:T}\in\mathbb{R}^T} |G_T\left[P_T h\right](x_{0:T})|^p f(x_0)\mathrm{d}x_0 < \infty,
\end{aligned}$$

showing (A.8). The proof is complete. □

Proof of Theorem 5. We first show that the finite-dimensional distributions of W_T defined in (A.2) by

$$s \mapsto W_T[s] = \zeta_T(\mathbf{X}_T[\lfloor Ts\rfloor])$$

converge weakly to those of the solution W of a Langevin equation starting with $W[0] \sim f$. We apply [6, Corollary 8.5] with

$$\begin{aligned}
\xi_T(s) &= P_T h(\mathbf{X}_T[\lfloor Ts\rfloor]) = h\circ\zeta_T(\mathbf{X}_T[\lfloor Ts\rfloor]) = h(W_T[s]) \\
\psi_T(s) &= G_T[P_T h](\mathbf{X}_T[\lfloor Ts\rfloor]).
\end{aligned}$$

The conditions [6, Theorem 8.2, (8.8)–(8.9), p. 227] are satisfied by definitions of ξ_T and ψ_T and (B3). Now, using [6, Remark 8.3, p. 227], conditions [6, Eqs. (8.10)–(8.11), p. 227] may be checked by showing that, for any $s\geq 0$,

$$\begin{aligned}
\mathbb{E}\left[|\psi_T(s) - Gh(W_T[s])|\right] &= \mathbb{E}\left[|G_T[P_T h](\mathbf{X}_T[0]) - Gh(W_T[0])|\right] \\
&= \int\cdots\int |G_T\left[P_T h\right](x_{0:T}) - Gh(x_0)|\, \pi_T(\mathrm{d}x_{0:T}) \to_{T\to\infty} 0,
\end{aligned}$$

which follows from Lemma 6. Thus, by [6, Corollary 8.5 and Theorem 8.2], the finite-dimensional distributions of W_T converge weakly to those of W. According to [6, Corollary 8.6], the convergence in the Skorokhod topology may be obtained by checking [6, (8.33)–(8.34) p. 231]. First note that [6, (8.33) p. 231] is direct by the definition of ξ_T. Moreover,

$$\begin{aligned}
&\sup_{T\in\mathbb{N}} \mathbb{E}\left[\left(\int_0^t |G_T[P_T h](\mathbf{X}_T[\lfloor Ts\rfloor])|^p \mathrm{d}s\right)^{1/p}\right] \\
&\quad\le \sup_{T\in\mathbb{N}} \left(\mathbb{E}\left[\int_0^t |G_T[P_T h](\mathbf{X}_T[\lfloor Ts\rfloor])|^p \mathrm{d}s\right]\right)^{1/p} \\
&\quad\le \sup_{T\in\mathbb{N}} \left(\int_0^t \mathbb{E}\left[|G_T[P_T h](\mathbf{X}_T[\lfloor Ts\rfloor])|^p\right] \mathrm{d}s\right)^{1/p} \\
&\quad\le \sup_{T\in\mathbb{N}} \left(\int_0^t \mathbb{E}\left[|G_T[P_T h](\mathbf{X}_T[0])|^p\right] \mathrm{d}s\right)^{1/p} \\
&\quad\le \sup_{T\in\mathbb{N}} \left(t\int \sup_{x_{1:T}\in\mathbb{R}^T} |G_T[P_T h](x_{0:T})|^p f(x_0)\mathrm{d}x_0\right)^{1/p} < \infty
\end{aligned}$$

by (B3). Thus condition [6, (8.34) p. 231] is also satisfied, which concludes the proof of the theorem. □

A.2. Scaling analysis of multiple-try algorithms

In this section, we now specialize Theorem 5 to a general class of multiple-try algorithms which encompass both MCTM and MTM-C algorithms. To apply Theorem 5, we need to specify the sequence of transition kernels $(Q_T)_{T\ge 1}$. Consider a sequence of homogeneous Markov chains $(\mathbf{X}_T[n],\ n\in\mathbb{N})$ taking values in $\mathbb{R}^{T+1}$ with transition kernel Q_T satisfying, for any measurable bounded function h on $\mathbb{R}$

$$\begin{aligned}
&Q_T\,[P_T h]\,(x_{0:T}) - h(x_0) \\
&\quad= \mathbb{E}\,[\,h(\zeta_T(\mathbf{X}_T[1]))\,|\,\mathbf{X}_T[0]=x_{0:T}] - h(x_0) \\
&\quad= \sum_{j=1}^{K} \mathbb{E}\left[\left\{h(x_0+T^{-1/2}U^j)-h(x_0)\right\}\beta_T^j(x_{0:T},T^{-1/2},x_0+T^{-1/2}U^j)\right]
\end{aligned} \tag{A.9}$$

where $\{U^j\}_{1\le j\le K}$ are random variables and, for $j\in\{1,\dots,K\}$, $\beta_T^j:\mathbb{R}^{T+1}\times\mathbb{R}\times\mathbb{R}\to[0,1]$,

$$(x_{0:T},\eta,y)\mapsto \beta_T^j(x_{0:T},\eta,y), \tag{A.10}$$

are nonnegative measurable functions. When applied to the MCTM or the MTM-C algorithm, $\beta_T^j(x_{0:T},\eta,y)$ will be the average probability of accepting the j-th component in the pool when the Markov chain is in state $x_{0:T}$, but it is not required to specify this function further at this stage. When studying the limiting behavior in (A.9), it will sometimes be convenient to write the expectation slightly differently. For $j\in\{1,\dots,K\}$, $\eta\ge 0$, and $u\in\mathbb{R}$, denote

$$\tilde\beta_T^j(x_{0:T},\eta,u)=\beta_T^j(x_{0:T},\eta,x_0+\eta u). \tag{A.11}$$

Alternatively, Eq. (A.11) can be rewritten as follows: for any $x_{0:T}\in\mathbb{R}^{T+1}$, $\eta\in\mathbb{R}$ and $y\in\mathbb{R}$,

$$\beta_T^j(x_{0:T},\eta,y)=\tilde\beta_T^j(x_{0:T},\eta,(y-x_0)/\eta), \tag{A.12}$$

with the convention $0/0 = 0$. With these notations,

$$Q_T\,[P_T h]\,(x_{0:T}) - h(x_0) = \sum_{j=1}^{K} \mathbb{E}\left[\left\{h(x_0 + T^{-1/2}U^j) - h(x_0)\right\} \tilde{\beta}_T^j(x_{0:T}, T^{-1/2}, U^j)\right], \tag{A.13}$$

so that $G_T = T(Q_T - I)$ can be rewritten as:

$$G_T\,[P_T h]\,(x_{0:T}) = T\,\{Q_T\,[P_T h]\,(x_{0:T}) - P_T h(x_{0:T})\} = \sum_{j=1}^{K} \mathbb{E}\left[T\left\{h(x_0 + T^{-1/2}U^j) - h(x_0)\right\} \tilde{\beta}_T^j(x_{0:T}, T^{-1/2}, U^j)\right]. \tag{A.14}$$

We now replace the assumptions (B3)–(B4) that may be difficult to check in practice by some more practical assumptions:

(C1) There exist constants $\{a_j\}_{j=1}^K \in \mathbb{R}^K$ such that for all $j \in \{1, \ldots, K\}$,

$$\lim_{T\to\infty} \int \sup_{x_{1:T}\in \mathsf{F}_T} \left|\beta_T^j(x_{0:T}, 0, x_0) - a_j\right| f(x_0)\mathrm{d}x_0 = 0.$$

(C2) There exists a family $\{w^j\}_{j=1}^K$ of measurable functions $w^j : \mathbb{R} \to \mathbb{R}$ such that for all $j \in \{1, \ldots, K\}$,

$$\lim_{T\to\infty} \int \sup_{x_{1:T}\in \mathsf{F}_T} \left|\frac{\partial \beta_T^j}{\partial y}(x_{0:T}, 0, x_0) - w^j(x_0)\right| f(x_0)\mathrm{d}x_0 = 0. \tag{A.15}$$

(C3) There exists $p > 1$ such that for any $j \in \{1, \ldots, K\}$,

$$\sup_{T\geq 0} \int \sup_{x_{1:T}\in \mathbb{R}^T} \left|\frac{\partial \beta_T^j}{\partial y}(x_{0:T}, 0, x_0)\right|^p f(x_0)\mathrm{d}x_0 < \infty, \tag{A.16}$$

$$\sup_{T\geq 0} \int \sup_{x_{1:T}\in \mathbb{R}^T} \left(\mathbb{E}\left[(U^j)^2 \sup_{0\leq\eta\leq T^{-1/2}} \left|\frac{\partial \tilde{\beta}_T^j}{\partial \eta}\left(x_{0:T}, \eta, U^j\right)\right|\right]\right)^p f(x_0)\mathrm{d}x_0 < \infty, \tag{A.17}$$

$$\sup_{T\geq 0} \int \sup_{x_{1:T}\in \mathbb{R}^T} \left(\mathbb{E}\left[|U^j| \sup_{0\leq\eta\leq T^{-1/2}} \left|\frac{\partial^2 \tilde{\beta}_T^j}{\partial \eta^2}(x_{0:T}, \eta, U^j)\right|\right]\right)^p f(x_0)\mathrm{d}x_0 < \infty. \tag{A.18}$$

(C4) For any $j \in \{1, \ldots, K\}$, $\mathbb{E}\left[U^j\right] = 0$ and $\mathbb{E}\left[|U^j|^3\right] < \infty$.

The main result of this section is the following theorem which establishes the weak convergence (in the Skorokhod topology) of $\left(W_T[s],\ s \in \mathbb{R}^+\right)$ defined in (A.2) to a Langevin diffusion.

Theorem 7. *Assume* (B1)–(B2) *and* (C1)–(C4). *Then,* $W_T \Longrightarrow W$ *in the Skorokhod topology where* $W[0]$ *is distributed according to* f *and* $(W[s],\ s \in \mathbb{R}^+)$ *satisfies the Langevin SDE*

$$\mathrm{d}W[t] = \sqrt{\lambda}\mathrm{d}B[t] + \frac{1}{2}\lambda\,[\ln f]'\,(W[t])\mathrm{d}t, \tag{A.19}$$

with

$$\lambda = \sum_{j=1}^{K} \mathbb{V}\mathrm{ar}[U^j] a_j. \tag{A.20}$$

In addition, for any $x \in \mathbb{R}$,

$$\sum_{j=1}^{K} \mathbb{V}\mathrm{ar}[U^j] w^j(x) = \frac{\lambda}{2} [\ln f]'(x). \tag{A.21}$$

Note that this theorem does not require to show that $\sum_{j=1}^{K} \mathbb{V}\mathrm{ar}[U^j] w^j(x) = \lambda [\ln f]'(x)/2$, and therefore simplifies the arguments presented in [11, Thm 1.1] and later developed in [1,2].

Denote by G the generator of the Langevin diffusion (A.19)

$$Gh(x) \triangleq \frac{\lambda}{2} \left(h'(x) [\ln f]'(x) + h''(x) \right), \tag{A.22}$$

where λ is defined in (A.20). Note that G_T defined in (A.14) is not itself the generator of $(W_T[s],\ s \in \mathbb{R}^+)$ since the latter is not a Markov process nor the first component of a Markov process (recall that $(W_T[s],\ s \in \mathbb{R}^+)$ is indeed constant over intervals of fixed length $1/T$ and not of exponential length). Of course, it is possible to obtain the same convergence results by considering instead the first component of the Markov process associated to the Markov chain $(X_{0:T}[s],\ s \in \mathbb{N})$ with jumps happening according to a Poisson process with rate T (see e.g. [11]). In this case, G_T can be seen as the generator of this Markov jump process. Nevertheless, this interpretation is not necessary and we decide here to consider directly $(W_T[s],\ s \in \mathbb{R}^+)$.

Proof of Theorem 7. It is a direct consequence of Theorem 5 and Lemma 8. □

Lemma 8. *Assume* (C1)–(C4). *Then* (B3)–(B4) *are satisfied with G defined in* (A.22).

Proof. Define, for any function $h \in C_c^\infty$,

$$\tilde{G}h(x) \triangleq h'(x) w(x) + \frac{1}{2} \lambda h''(x), \tag{A.23}$$

where λ is given by (A.20) and $w(x) = \sum_{j=1}^{K} \mathbb{V}\mathrm{ar}[U^j] w^j(x)$. We first show

$$\sup_{T \geq 0} \int \sup_{x_{1:T} \in \mathbb{R}^T} |G_T [P_T h](x_{0:T})|^p f(x_0) \mathrm{d}x_0 < \infty, \tag{A.24}$$

and

$$\lim_{T \to \infty} \int \sup_{x_{1:T} \in \mathsf{F}_T} \left| G_T [P_T h](x_{0:T}) - \tilde{G}h(x_0) \right| f(x_0) \mathrm{d}x_0 = 0. \tag{A.25}$$

Denote $\eta_T = T^{-1/2}$. Note that under (C4),

$$\mathbb{E}\left[U^j \tilde{\beta}_T^j (x_{0:T}, 0, U^j) \right] = \beta_T^j (x_{0:T}, 0, x_0) \mathbb{E}\left[U^j \right] = 0,$$

so that by a Taylor expansion of $\eta \mapsto h(x_0 + \eta U^j)$ in η in a neighborhood of $\eta = 0$ in (A.14), we obtain the following decomposition

$$G_T\left[P_T h\right](x_{0:T}) = \sum_{j=1}^{K} h'(x_0) A_T(x_{0:T}, h) + \sum_{j=1}^{K} \frac{1}{2} h''(x_0) B_T^j(x_{0:T}, h) + \sum_{j=1}^{K} \frac{\eta_T}{6} R_T^j(x_{0:T}, h) \tag{A.26}$$

where for $j \in \{1, \ldots, K\}$,

$$A_T(x_{0:T}, h) \triangleq \eta_T^{-1} \mathbb{E}\left[U^j \left\{\tilde{\beta}_T^j(x_{0:T}, \eta_T, U^j) - \tilde{\beta}_T^j(x_{0:T}, 0, U^j)\right\}\right], \tag{A.27}$$

$$B_T^j(x_{0:T}, h) \triangleq \mathbb{E}\left[(U^j)^2 \tilde{\beta}_T^j(x_{0:T}, \eta_T, U^j)\right], \tag{A.28}$$

$$R_T^j(x_{0:T}, h) \triangleq \mathbb{E}\left[(U^j)^3 h'''(x_0 + \bar{\eta}_T U^j) \tilde{\beta}_T^j(x_{0:T}, \eta_T, U^j)\right], \tag{A.29}$$

for some random variable $\bar{\eta}_T \in [0, \eta_T]$. First note that, for all $x_0 \in \mathbb{R}$ and $j \in \{1, \ldots, K\}$,

$$\sup_{x_{0:T} \in \mathbb{R}^{T+1}} \left|R_T^j(x_{0:T}, h)\right| \leq \left|h'''\right|_\infty \mathbb{E}\left[|U^j|^3\right]. \tag{A.30}$$

We now consider the terms $A_T(x_{0:T}, h)$ and $B_T^j(x_{0:T}, h)$, $j \in \{1, \ldots, K\}$. Noting that $\mathbb{E}\left[U^j\right] = 0$ and that for any $u \in \mathbb{R}$, using (A.11),

$$\frac{\partial \tilde{\beta}_T^j}{\partial \eta}(x_{0:T}, 0, u) = \frac{\partial \beta_T^j}{\partial \eta}(x_{0:T}, 0, x_0) + u \frac{\partial \beta_T^j}{\partial y}(x_{0:T}, 0, x_0),$$

a second-order expansion of $\eta \mapsto \tilde{\beta}_T^j(x_{0:T}, \eta, U^j)$ in η in a neighborhood of $\eta = 0$ in (A.27) yields

$$A_T(x_{0:T}, h) = \mathbb{V}\mathrm{ar}[U^j] \frac{\partial \beta_T^j}{\partial y}(x_{0:T}, 0, x_0) + \frac{\eta_T}{2} R_T^{A,j}(x_{0:T}, h)$$

where

$$R_T^{A,j}(x_{0:T}, h) \triangleq \mathbb{E}\left[U^j \frac{\partial^2 \tilde{\beta}_T^j}{\partial \eta^2}(x_{0:T}, \tilde{\eta}_T, U^j)\right],$$

for some random variable $\tilde{\eta}_T \in [0, \eta_T]$. Then,

$$\left|R_T^{A,j}(x_{0:T}, h)\right| \leq \mathbb{E}\left[|U^j| \sup_{0 \leq \eta \leq \eta_T} \left|\frac{\partial^2 \tilde{\beta}_T^j}{\partial \eta^2}(x_{0:T}, \eta, U^j)\right|\right], \tag{A.31}$$

and (A.16)–(A.18) therefore imply that

$$\sup_{T \geq 0} \int \sup_{x_{1:T} \in \mathbb{R}^T} |A_T(x_{0:T}, h)|^p f(x_0) \mathrm{d}x_0 < \infty. \tag{A.32}$$

We may further decompose $A_T(x_{0:T}, h)$ as follows

$$A_T(x_{0:T}, h) = \mathbb{V}\mathrm{ar}[U^j] \left\{w^j(x_0) + \left(\frac{\partial \beta_T^j}{\partial y}(x_{0:T}, 0, x_0) - w^j(x_0)\right)\right\} + \frac{\eta_T}{2} R_T^{A,j}(x_{0:T}, h).$$

Using (A.15), (A.18) and (A.31),

$$\lim_{T\to\infty}\int \sup_{x_{1:T}\in \mathsf{F}_T}\left|A_T(x_{0:T},h)-\mathbb{V}\mathrm{ar}[U^j]w^j(x_0)\right| f(x_0)\mathrm{d}x_0 = 0. \tag{A.33}$$

Consider finally $B_T^j(x_{0:T},h)$. Since, by definition, $\tilde{\beta}_T(x_{0:T},\eta,U^j)\le 1$, we have that

$$\sup_{T\ge 0}\int \sup_{x_{1:T}\in\mathbb{R}^T}\mathbb{E}\left[\left|B_T^j(x_{0:T},h)\right|\right]^p f(x_0)\mathrm{d}x_0 \le \mathbb{V}\mathrm{ar}[U^j]^p<\infty. \tag{A.34}$$

By a first order expansion of $\tilde{\beta}_T^j$ in η at 0 in (A.28), we obtain for $j\in\{1,\dots,K\}$,

$$B_T^j(x_{0:T},h)=\mathbb{V}\mathrm{ar}[U^j]\left\{a_j+\left[\beta_T^j(x_{0:T},0,x_0)-a_j\right]\right\}+\eta_T R_T^{B,j}(x_{0:T},h)$$

where

$$|R_T^{B,j}(x_{0:T},h)|\le \mathbb{E}\left[(U^j)^2\sup_{0\le\eta\le\eta_T}\left|\frac{\partial\tilde{\beta}_T^j(x_{0:T},\eta,U^j)}{\partial\eta}\right|\right]. \tag{A.35}$$

Thus, (C1) and (A.17) yield, for $j\in\{1,\dots,K\}$,

$$\lim_{T\to\infty}\int \sup_{x_{1:T}\in \mathsf{F}_T}\left|B_T^j(x_{0:T},h)-\mathbb{V}\mathrm{ar}[U^j]a_j\right| f(x_0)\mathrm{d}x_0 = 0. \tag{A.36}$$

The proof of (A.24) follows from (A.26) using (C3), (A.30), (A.32) and (A.34). The proof of (A.25) follows from (A.23) and (A.26) using (A.30), (A.33) and (A.36). To obtain (B3)–(B4), it thus remains to show that $\tilde{G}$ coincides with G almost everywhere. By construction, Q_T is a transition kernel with stationary distribution π_T; hence, by definition of G_T, for any $h\in C_c^\infty$, $\pi_T(G_T[P_T h])=0$. Now, since f is the stationary distribution of the Langevin diffusion (A.19), we get by (A.3) that $\pi_T(P_T Gh)=0$. Combining the two previous relations yields

$$\begin{aligned}&\pi_T\left(G_T[P_T h]-P_T Gh\right)\\&\quad=\pi_T\left(G_T[P_T h]-P_T\tilde{G}h\right)+\int f(x)\left[\tilde{G}h(x)-Gh(x)\right]\mathrm{d}x=0.\end{aligned} \tag{A.37}$$

By (A.4), for any $h\in C_c^\infty$,

$$\lim_{T\to\infty}\pi_T\left(G_T[P_T h]-P_T\tilde{G}h\right)=0, \tag{A.38}$$

from which we deduce that

$$\int f(x)\left[\tilde{G}h(x)-Gh(x)\right]\mathrm{d}x=\int f(x)h'(x)\left[w(x)-\frac{\lambda}{2}[\ln f]'(x)\right]\mathrm{d}x=0. \tag{A.39}$$

The latter relation being satisfied for any $h\in C_c^\infty$, this implies that, almost everywhere with respect to the Lebesgue measure,

$$w(x)=\frac{\lambda}{2}[\ln f]'(x). \tag{A.40}$$

Therefore, the two generators G and $\tilde{G}$ coincide. The proof is complete. □

A.3. Application to MCTM and MTM-C algorithms

The last step consists in applying Theorem 7 to the analysis of the MCTM and the MTM-C algorithms. Let η such that

$$0 < \eta < 1/4. \tag{A.41}$$

Then, define the sequence of sets $\{\mathsf{F}_T\}_{T=0}^{\infty}$ by

$$\mathsf{F}_T = \left\{x_{1:T} \in \mathbb{R}^T, |\mathcal{I}_T(x_{1:T}) - \mathcal{I}| \vee |\mathcal{J}_T(x_{1:T}) - \mathcal{I}| \vee \mathcal{S}_T(x_{1:T}) \leq T^{-\eta}\right\}, \tag{A.42}$$

where, for any $x_{1:T} \in \mathbb{R}^T$, we let

$$\mathcal{I}_T(x_{1:T}) = T^{-1} \sum_{t=1}^{T} \left\{[\ln f]'(x_t)\right\}^2, \tag{A.43}$$

$$\mathcal{J}_T(x_{1:T}) = -T^{-1} \sum_{t=1}^{T} [\ln f]''(x_t), \tag{A.44}$$

$$\mathcal{S}_T(x_{1:T}) = T^{-1/2} \mathcal{I}_T{}^{-1/2}(x_{1:T}) \sup_{t=1,\dots,T} \left|[\ln f]'(x_t)\right|, \tag{A.45}$$

and $\mathcal{I}$ is as in (21). We preface the proof of Theorem 2 by two technical lemmas: Lemma 9 will be used for checking Assumption (B2) and Lemma 10 for Assumptions (C1)–(C2).

Lemma 9. *Assume* (A1). *Then* (B2) *is satisfied.*

Proof. First note that since $\eta < 1/2$, the law of iterated logarithm implies that

$$\lim_{T\to\infty} \pi_T\{x_{0:T} \in \mathbb{R}^{T+1}; |\mathcal{I}_T(x_{1:T}) - \mathcal{I}| \vee |\mathcal{J}_T(x_{1:T}) - \mathcal{I}| \leq T^{-\eta}\} = 1,$$

where $\mathcal{I}_T$ and $\mathcal{J}_T$ are defined in (A.43) and (A.44). To obtain (B2), it is thus sufficient to show that

$$\lim_{T\to\infty} \pi_T \left\{x_{0:T} \in \mathbb{R}^{T+1}; \mathcal{S}_T(x_{1:T}) \leq T^{-\eta}\right\} = 1, \tag{A.46}$$

where $\mathcal{S}_T$ is defined in (A.45). More precisely, we will show that for any constant c,

$$\lim_{T\to\infty} \pi_T \left\{x_{0:T} \in \mathbb{R}^{T+1}; T^{-1/2} \sup_{t=1,\dots,T} |[\ln f]'(x_t)| \leq cT^{-\eta}\right\} = 1,$$

which implies (A.46). Let $(X_t)_{t=0}^{T}$ be a sequence of i.i.d. random variables distributed under f; then

$$\begin{aligned}
&\pi_T \left\{x_{0:T} \in \mathbb{R}^{T+1}; T^{-1/2} \sup_{t=1,\dots,T} |[\ln f]'(x_t)| \leq cT^{-\eta}\right\} \\
&\quad = \mathbb{P}\left[\sup_{t=1,\dots,T} |[\ln f]'(X_t)| \leq cT^{1/2-\eta}\right] \\
&\quad = \exp\left\{T \ln\left(1 - \mathbb{P}\left[|[\ln f]'(X_1)| > cT^{1/2-\eta}\right]\right)\right\},
\end{aligned}$$

which tends to 1 as T grows to infinity if and only if

$$\mathbb{P}\left[|[\ln f]'(X_1)| > cT^{1/2-\eta}\right] = \mathrm{o}(T^{-1}).$$

Now, by Markov's inequality,

$$\mathbb{P}\left[|[\ln f]'(X_1)| > cT^{1/2-\eta}\right] \le \mathbb{E}\left[|[\ln f]'(X_1)|^4\right] / T^{4(1/2-\eta)},$$

which is $\mathrm{o}(T^{-1})$ since the right-hand side is finite by (A1) and $4(1/2-\eta) > 1$ (by (A.41)). This proves (B2). □

For any measurable bounded function $\zeta : \mathbb{R}^\ell \to \mathbb{R}^p$, and nonnegative $\ell \times \ell$ matrix Γ, define

$$a(\zeta, \Gamma) \triangleq \mathbb{E}\left[\zeta\left\{\left(G^i - \mathbb{V}\mathrm{ar}[G^i]/2\right)_{i=1}^{\ell}\right\}\right]. \tag{A.47}$$

Lemma 10. *Let $\zeta : \mathbb{R}^\ell \to \mathbb{R}$ be a Lipschitz and bounded function. Let Γ be a $(\ell \times \ell)$ nonnegative symmetric matrix and $\{V_t = (V_t^1, \ldots, V_t^\ell)\}_{t=1}^T$ be i.i.d. ℓ-dimensional random vectors with zero-mean and covariance matrix Γ. For $i = 1, \ldots, \ell$, let $H^i : \mathbb{R}^2 \to \mathbb{R}$ be functions such that for all $x \in \mathbb{R}$, $y \mapsto H^i(x, y)$ is differentiable at $y = x$ and $H^i(x, x) = 0$. Finally, for $x_{0:T} \in \mathbb{R}^{T+1}$ and $y \in \mathbb{R}$, let*

$$\Upsilon_T(x_{0:T}, y) \triangleq \mathbb{E}\left[\zeta\left\{\left(L_{1,T}(x_{1:T}, V_{1:T}^i) + H^i(x_0, y)\right)_{i=1}^{\ell}\right\}\right],$$

where $L_{1,T}$ is the log-likelihood ratio defined in (18). Then,

(i) $\lim_{T\to\infty} \sup_{\mathsf{F}_T} |\Upsilon_T(x_{0:T}, x_0) - a(\zeta, \mathcal{I}\Gamma)| = 0$, *where $\mathcal{I}$ is defined in (21).*
(ii) *If in addition ζ is differentiable and $\nabla\zeta$ is Lipschitz and bounded, then for all $x_{0:T} \in \mathbb{R}^{T+1}$, the function $y \mapsto \Upsilon_T(x_{0:T}, y)$ is differentiable at $y = x_0$ and*

$$\lim_{T\to\infty} \sup_{\mathsf{F}_T} \left|\frac{\partial \Upsilon_T}{\partial y}(x_{0:T}, x_0) - \left\langle \frac{\partial H}{\partial y}(x_0, x_0), a(\nabla\zeta, \mathcal{I}\Gamma)\right\rangle\right| = 0,$$

where $\frac{\partial H}{\partial y}(x_0, y) = \left(\frac{\partial H^i}{\partial y}(x_0, y)\right)_{i=1}^{K}$.

Proof. In this proof, denote by $\gamma_j^2 = \mathbb{V}\mathrm{ar}[V^j]$. We first consider statement (i). Define, for $v_{1:T} \in \mathbb{R}^T$,

$$W_T(x_{1:T}, v_{1:T}) \triangleq T^{-1/2}\mathcal{I}_T{}^{-1/2}(x_{1:T}) \sum_{t=1}^{T} [\ln f]'(x_t) v_t. \tag{A.48}$$

By a second order Taylor expansion of $L_{1,T}$, write the following decomposition for any $j \in \{1, \ldots, \ell\}$,

$$L_{1,T}(x_{1:T}, V_{1:T}^j) = \left[\mathcal{I}^{1/2} W_T(x_{1:T}, V_{1:T}^j) - \frac{1}{2}\mathcal{I}\gamma_j^2\right] + \sum_{u=1}^{4} R_T^{j,u}(x_{1:T}, V_{1:T}^j),$$

where

$$R_T^{j,1}(x_{1:T}, v_{1:T}) = \left\{\mathcal{I}_T{}^{1/2}(x_{1:T}) - \mathcal{I}^{1/2}\right\} W_T(x_{1:T}, v_{1:T}),$$

$$R_T^{j,2}(x_{1:T}, v_{1:T}) = -\frac{1}{2}\{\mathcal{J}_T(x_{1:T}) - \mathcal{I}\}\gamma_j^2,$$
$$R_T^{j,3}(x_{1:T}, v_{1:T}) = \frac{1}{2T}\sum_{t=1}^{T}[\ln f]''(x_t)(v_t^2 - \gamma_j^2),$$
$$R_T^{j,4}(x_{1:T}, v_{1:T}) = \frac{1}{2T}\sum_{t=1}^{T}\left\{[\ln f]''(x_t + \bar{\eta}_{T,t}v_t) - [\ln f]''(x_t)\right\}v_t^2,$$

and $\bar{\eta}_{T,t}$ belongs to $(0, T^{-1/2})$. Now, denote

$$\tilde{\zeta}\left((u_i)_{i=1}^{\ell}\right) = \zeta\left((\mathcal{I}^{1/2}u_i - \mathcal{I}\gamma_i^2/2)_{i=1}^{\ell}\right). \tag{A.49}$$

Of course, $\tilde{\zeta}$ implicitly depends on $\{\gamma_i\}_{i=1}^{\ell}$ but to simplify, we drop it from the notations. Since the function ζ is Lipschitz and $H^i(x_0, x_0) = 0$,

$$\sup_{\mathsf{F}_T}\left|\Upsilon_T(x_{0:T}, x_0) - \mathbb{E}\left[\tilde{\zeta}\left\{\left(W_T(x_{1:T}, V_{1:T}^i)\right)_{i=1}^{\ell}\right\}\right]\right|$$
$$\le [\zeta]_{\mathrm{lip}}\sum_{j=1}^{\ell}\sum_{u=1}^{4}\sup_{\mathsf{F}_T}\mathbb{E}\left[\left|R_T^{j,u}(x_{1:T}, V_{1:T}^j)\right|\right],$$

where $[\zeta]_{\mathrm{lip}}$ is the Lipschitz constant of ζ. We now show that the RHS converges to 0 as T tends to infinity. First, write

$$\sup_{\mathsf{F}_T}\mathbb{E}\left[|R_T^{j,1}(x_{1:T}, V_{1:T}^j)|\right] \le \sup_{\mathsf{F}_T}\left(\left|\mathcal{I}_T{}^{1/2}(x_{1:T}) - \mathcal{I}^{1/2}\right|\mathcal{S}_T(x_{1:T})\mathbb{E}\left[\sum_{i=1}^{\ell}|V_1^i|\right]\right),$$

which converges to 0 as $T \to \infty$ by definition of the set F_T. Using again the definition of the set F_T, we obtain immediately that

$$\lim_{T\to\infty}\sup_{\mathsf{F}_T}\mathbb{E}\left[|R_T^{j,2}(x_{1:T}, V_{1:T}^j)|\right] = 0.$$

Applying $\mathbb{E}[|R|] \le \mathbb{V}\mathrm{ar}[R]^{1/2}$ when $\mathbb{E}[R] = 0$, we obtain

$$\mathbb{E}\left[\left|R_T^{j,3}(x_{1:T}, V_{1:T}^j)\right|\right] \le \left(\frac{1}{4T^2}\sum_{t=1}^{T}\left([\ln f]''(x_t)\right)^2\mathbb{V}\mathrm{ar}[(V_1^j)^2]\right)^{1/2}$$
$$\le \frac{1}{2\sqrt{T}}\left|[\ln f]''\right|_{\infty}\left(\mathbb{V}\mathrm{ar}[(V_1^j)^2]\right)^{1/2}.$$

Therefore, $\lim_{T\to\infty}\sup_{x_{1:T}\in\mathsf{F}_T}\mathbb{E}\left[\left|R_T^{j,3}(x_{1:T}, V_{1:T}^j)\right|\right] = 0$. Finally, using that $(\ln f)''$ is Lipschitz,

$$\sup_{\mathsf{F}_T}\mathbb{E}\left[|R_T^{j,4}(x_{1:T}, V_{1:T}^j)|\right] \le \frac{1}{2\sqrt{T}}\left[(\ln f)''\right]_{\mathrm{lip}}\mathbb{E}\left[|V_1^j|^3\right],$$

which converges to 0 as T tends to infinity. Finally,

$$\lim_{T\to\infty}\sup_{\mathsf{F}_T}\left|\Upsilon_T(x_{0:T}, x_0) - \mathbb{E}\left[\tilde{\zeta}\left\{\left(W_T(x_{1:T}, V_{1:T}^i)\right)_{i=1}^{\ell}\right\}\right]\right| = 0.$$

To show (i), it thus remains to check that

$$\lim_{T\to\infty} \sup_{\mathsf{F}_T} \left| \mathbb{E}\left[\tilde{\zeta}\left\{\left(W_T(x_{1:T}, V^i_{1:T})\right)_{i=1}^{\ell}\right\}\right] - \mathbb{E}\left[\tilde{\zeta}\left\{\left(G^i\right)_{i=1}^{\ell}\right\}\right]\right| = 0,$$

where $(G^1, \ldots, G^\ell) \sim \mathcal{N}(0, \Gamma)$. Let $\{x_{T,1:T}\}_{T\geq 1} = \{(x_{T,s})_{s=1}^T\}_{T\geq 1}$ be a triangular array of (deterministic) real numbers satisfying for all $T \geq 0$,

$$\begin{aligned}\sup_{x_{1:T}\in \mathsf{F}_T} &\left| \mathbb{E}\left[\tilde{\zeta}\left\{\left(W_T(x_{1:T}, V^i_{1:T})\right)_{i=1}^{\ell}\right\}\right] - \mathbb{E}\left[\tilde{\zeta}\left\{\left(G^i\right)_{i=1}^{\ell}\right\}\right]\right| \\ &\leq \left| \mathbb{E}\left[\tilde{\zeta}\left\{\left(W_T(x_{T,1:T}, V^i_{1:T})\right)_{i=1}^{\ell}\right\}\right] - \mathbb{E}\left[\tilde{\zeta}\left\{\left(G^i\right)_{i=1}^{\ell}\right\}\right]\right| + 1/T.\end{aligned}$$

Since the function $\tilde{\zeta}$ is continuous and bounded, the right-hand side converges to 0 as soon as we can show that the random vector $\left(W_T(x_{T,1:T}, V^i_{1:T})\right)_{i=1}^{\ell}$ converges weakly to $\left(G^i\right)_{i=1}^{\ell}$. Using the Cramer–Wold device, it is enough to show that for all scalars $(\alpha_i)_{i=1}^{\ell} \in \mathbb{R}^\ell$,

$$\sum_{i=1}^{\ell} \alpha_i W_T(x_{T,1:T}, V^i_{1:T}) \xrightarrow{\mathcal{D}}_{T\to\infty} \mathcal{N}\left(0, \sigma^2\right), \tag{A.50}$$

where $\sigma^2 = \mathbb{E}\left[\left(\sum_{i=1}^{\ell} \alpha_i G^i\right)^2\right] = \mathbb{E}\left[\left(\sum_{i=1}^{\ell} \alpha_i V^i_1\right)^2\right]$. Rewrite the left-hand side of (A.50) as:

$$\sum_{i=1}^{\ell} \alpha_i W_T(x_{T,1:T}, V^i_{1:T}) = \sum_{t=1}^{T} U_{T,t},$$

where $U_{T,t} \triangleq \left(T^{-1/2}\mathcal{I}_T{}^{-1/2}(x_{T,1:T}) [\ln f]'(x_{T,t})\right) \sum_{i=1}^{\ell} \alpha_i V^i_t$ and set $\mathcal{F}_{T,t} = \sigma(\{V^i_1, \ldots, V^i_t\}_{i=1}^{\ell})$. Since $U_{T,t}$ is centered and $\mathcal{F}_{T,t}$-measurable, we will show (A.50) by applying the CLT theorem for a triangular array of random variables (see [5]). We thus need to check that

$$\sum_{t=1}^{T} \mathbb{E}\left[U^2_{T,t} \,\middle|\, \mathcal{F}_{T,t-1}\right] - \left\{\mathbb{E}\left[U_{T,t} \,\middle|\, \mathcal{F}_{T,t-1}\right]\right\}^2 \xrightarrow{\mathrm{P}} \sigma^2, \tag{A.51}$$

$$\sum_{t=1}^{T} \mathbb{E}\left[U^2_{T,t}\mathbb{1}_{\{|U_{T,t}|\geq\epsilon\}} \,\middle|\, \mathcal{F}_{T,t-1}\right] \xrightarrow{\mathrm{P}} 0, \tag{A.52}$$

for any $\epsilon > 0$. (A.51) is immediate since by straightforward algebra,

$$\sum_{t=1}^{T} \mathbb{E}\left[U^2_{T,t} \,\middle|\, \mathcal{F}_{T,t-1}\right] - \left\{\mathbb{E}\left[U_{T,t} \,\middle|\, \mathcal{F}_{T,t-1}\right]\right\}^2 = \sigma^2.$$

Moreover, by the definition of $U_{T,t}$ and F_T, since $x_{T,1:T} \in \mathsf{F}_T$, we have that

$$U^2_{T,t}\mathbb{1}_{\{|U_{T,t}|\geq\epsilon\}} \leq T^{-1}\mathcal{I}_T{}^{-1}(x_{T,1:T}) \left\{[\ln f]'(x_{T,t})\right\}^2 \left(\sum_{i=1}^{\ell} \alpha_i V^i_t\right)^2 \mathbb{1}_{\{T^{-\eta}|\sum_{i=1}^{\ell} \alpha_i V^i_t|\geq\epsilon\}},$$

where η is the constant that appears in the definition of F_T (see (A.42)). Since the T random vectors $\{(V_t^1, \ldots, V_t^\ell)\}_{t=1}^T$ are i.i.d., this implies

$$\sum_{t=1}^T \mathbb{E}\left[U_{T,t}^2 \mathbb{1}_{\{|U_{T,t}|>\epsilon\}} \,\middle|\, \mathcal{F}_{T,t-1} \right] \le \mathbb{E}\left[\left(\sum_{i=1}^\ell \alpha_i V_0^i \right)^2 \mathbb{1}_{\{|\sum_{i=1}^\ell \alpha_i V_0^i| \ge \epsilon T^\eta\}} \right],$$

which converges to 0 as T tends to infinity. The proof of (i) follows. The statement (ii) is a direct consequence of (i). □

Proof of Theorem 2. The proof follows from Theorem 7 by checking successively (B1)–(B2) and (C1)–(C4). (B1) is derived from standard properties of MCMC algorithms and (B2) is direct from Lemma 9.

Now, rewrite $Q_T\,[P_T h]\,(x_{0:T}) - h(x_0)$ as in (A.9) where Q_T is the Markov kernel associated to the MCTM algorithm. Provided that at time 0 the state of the Markov chain is $\mathbf{x}_T = x_{0:T}$, the candidate $\mathbf{Y}_T^j[1] = Y_{T,0:T}^j$ is accepted with probability $\alpha_T^j\left(x_{0:T}, \left(Y_{T,0:T}^i[1] \right)_{i=1}^K, (\tilde{Y}_{T,0:T}^{j,i}[1])_{i\neq j} \right)$ given by Eq. (15), so that

$$\begin{aligned} &Q_T\,[P_T h]\,(x_{0:T}) - h(x_0) \\ &\quad = \sum_{j=1}^K \mathbb{E}\left[\alpha_T^j(x_{0:T}, (Y_{T,0:T}^i[1])_{i=1}^K, (\tilde{Y}_{T,0:T}^{j,i}[1])_{i\neq j}) \{h(Y_{T,0}^j[1]) - h(x_0)\} \right] \\ &\quad = \sum_{j=1}^K \mathbb{E}\left[\mathbb{E}\left[\alpha_T^j(x_{0:T}, (Y_{T,0:T}^i[1])_{i=1}^K, (\tilde{Y}_{T,0:T}^{j,i}[1])_{i\neq j}) \,\middle|\, Y_{T,0}^j \right] \{h(Y_{T,0}^j[1]) - h(x_0)\} \right]. \end{aligned}$$

By definition, $(Y_{T,1:T}^i[1])_{i=1}^K, (\tilde{Y}_{T,1:T}^{j,i}[1])_{i\neq j}$ are independent of $Y_{T,0}^j[1]$. Now, by (12)–(14), for any $j \in \{1, \ldots, K\}$ and $i \in \{1, \ldots, K\} \setminus \{j\}$,

$$\begin{aligned} Y_{T,0}^i[1] &= m^{j,i}(x_0, Y_0^j[1]) + T^{-1/2} W^{j,i}[1], \\ \tilde{Y}_{T,0}^{j,i}[1] &= m^{j,i}(Y_{T,0}^j[1], x_0) + T^{-1/2} \tilde{W}^{j,i}[1], \end{aligned}$$

where $m^{j,i}(x, y) = (1 - \mu^{j,i}(\Sigma))x + \mu^{j,i}(\Sigma)y$, $(W^{j,i})_{i\neq j} \sim \mathcal{N}\left(0, \Delta^j(\Sigma)\right)$ and $(\tilde{W}^{j,i})_{i\neq j} \sim \mathcal{N}\left(0, \Delta^j(\Sigma)\right)$ and the vectors $(W^{j,i})_{i\neq j}$, $(\tilde{W}^{j,i})_{i\neq j}$ and $Y_{T,0}^j$ are independent ($\mu^{j,i}$, Δ^j are defined in (10) and (11)). Therefore,

$$Q_T\,[P_T h]\,(x_{0:T}) - h(x_0) = \sum_{j=1}^K \mathbb{E}\left[\beta_T^j(x_{0:T}, T^{-1/2}, Y_{T,0}^j[1]) \{h(Y_{T,0}^j[1]) - h(x_0)\} \right]$$

where β_T^j is defined by

$$\begin{aligned} \beta_T^j(x_{0:T}, \eta, y) = \mathbb{E}\Bigg[A \Bigg\{ & L_{1,T}(x_{1:T}, U_{1:T}^j) + \ln f(y) - \ln f(x_0), \\ & \left(L_{1,T}(x_{1:T}, U_{1:T}^i) + \ln f[m^{j,i}(x_0, y) + \eta W^{j,i}] - \ln f(x_0) \right)_{i\neq j}, \\ & \left(L_{1,T}(x_{1:T}, \tilde{U}_{1:T}^i) + \ln f[m^{j,i}(y, x_0) + \eta \tilde{W}^{j,i}] - \ln f(x_0) \right)_{i\neq j} \Bigg\} \Bigg]. \end{aligned}$$

This expression allows us to define $\tilde{\beta}_T^j$ using the relation

$$\tilde{\beta}_T^j(x_{0:T}, \eta, u) = \beta_T^j(x_{0:T}, \eta, x_0 + \eta u).$$

Noting that the first and second order derivatives of A are all bounded and the fact that there exists a constant M such that for all $u \in \mathbb{R}$,

$$|[\ln f]'(u)| \le M|u|, \qquad |[\ln f]''(u)| \le M,$$

we obtain the existence of constants C and D (which do not depend on $x_{0:T}$ nor on η or u) such that for all $\eta \le T^{-1/2} \le 1$,

$$\sup_{x_{1:T} \in \mathbb{R}^T} \left| \frac{\partial \tilde{\beta}_T^j}{\partial \eta}(x_{0:T}, \eta, u) \right| \le C|u|(|x_0| + |u|) + D,$$
$$\sup_{x_{1:T} \in \mathbb{R}^T} \left| \frac{\partial^2 \tilde{\beta}_T^j}{\partial \eta^2}(x_{0:T}, \eta, u) \right| \le C|u|^2(|x_0| + |u|)^2 + D,$$

showing assumption (C3) for any $p > 1$. Finally, note that

$$\begin{aligned}\beta_T^j(x_{0:T}, 0, y) = \mathbb{E}\Bigg[A \Bigg\{ & L_{1,T}(x_{1:T}, U_{1:T}^j) + \ln f(y) - \ln f(x_0), \\ & \Big(L_{1,T}(x_{1:T}, U_{1:T}^i) + \ln f[m^{j,i}(x_0, y)] - \ln f(x_0) \Big)_{i \ne j}, \\ & \Big(L_{1,T}(x_{1:T}, \tilde{U}_{1:T}^i) + \ln f[m^{j,i}(y, x_0)] - \ln f(x_0) \Big)_{i \ne j} \Bigg\} \Bigg]\end{aligned}$$

and assumptions (C1) and (C2) are direct from Lemma 10 and the Dominated Convergence Theorem. (C4) is immediate. □

References

[1] M. Bédard, Weak convergence of Metropolis algorithms for non-i.i.d. target distributions, Ann. Appl. Probab. 17 (2007) 1222–1244.

[2] M. Bédard, Optimal acceptance rates for Metropolis algorithms: moving beyond 0.234, Stochastic Process. Appl. 118 (2008) 2198–2222.

[3] M. Bédard, R. Douc, E. Moulines, Scaling analysis of delayed rejection MCMC method (submitted for publication), Preprint, 2010.

[4] R.V. Craiu, C. Lemieux, Acceleration of the multiple-try metropolis algorithm using antithetic and stratified sampling, Stat. Comput. 17 (2007) 109–120.

[5] A. Dvoretzky, Asymptotic normality for sums of dependent random variables, in: Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Univ. California, Berkeley, Calif., 1970/1971, in: Probability Theory, vol. II, Univ. California Press, Berkeley, Calif., 1972, pp. 513–535.

[6] S.N. Ethier, T.G. Kurtz, Markov Processes: Characterization and Convergence, in: Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons Inc., New York, 1986.

[7] D. Frenkel, B. Smit, Understanding Molecular Simulations, Academic Press, 2002.

[8] D.P. Landau, K. Binder, A Guide to Monte Carlo Simulations in Statistical Physics, third ed., Cambridge University Press, Cambridge, 2009.

[9] F. Liang, C. Liu, R. Caroll, Advanced Markov Chain Monte Carlo Methods: Learning from Past Samples, Wiley, New York, 2010.

[10] J.S. Liu, F. Liang, W.H. Wong, The multiple-try method and local optimization in Metropolis sampling, J. Amer. Statist. Assoc. 95 (2000) 121–134.

[11] G.O. Roberts, A. Gelman, W.R. Gilks, Weak convergence and optimal scaling of random walk Metropolis algorithms, Ann. Appl. Probab. 7 (1997) 110–120.
[12] G.O. Roberts, J.S. Rosenthal, Markov chain Monte Carlo: some practical implications of theoretical results, Canad. J. Statist. 26 (1998) 5–32.
[13] G.O. Roberts, J.S. Rosenthal, Optimal scaling for various Metropolis–Hastings algorithms, Statist. Sci. 16 (2001) 351–367.