Transdimensional transformation based Markov chain Monte Carlo

Moumita Das and Sourabh Bhattacharya

Indian Statistical Institute

Abstract. Variable dimensional problems, where not only the parameters, but also the number of parameters are random variables, pose serious challenge to Bayesians. Although in principle the Reversible Jump Markov Chain Monte Carlo (RJMCMC) methodology is a response to such challenges, the dimension-hopping strategies need not be always convenient for practical implementation, particularly because efficient "move-types" having reasonable acceptance rates are often difficult to devise.

In this article, we propose and develop a novel and general dimension-hopping MCMC methodology that can update all the parameters as well as the number of parameters simultaneously using simple deterministic transformations of some low-dimensional (often one-dimensional) random variable. This methodology, which has been inspired by Transformation based MCMC (TMCMC) of (*Stat. Mehodol.* (2014) **16** 100–116), facilitates great speed in terms of computation time and provides reasonable acceptance rates and mixing properties. Quite importantly, our approach provides a natural way to automate the move-types in variable dimensional problems. We refer to this methodology as Transdimensional Transformation based Markov Chain Monte Carlo (TTMCMC). Comparisons with RJMCMC in gamma and normal mixture examples demonstrate far superior performance of TTMCMC in terms of mixing, acceptance rate, computational speed and automation. Furthermore, we demonstrate good performance of TTMCMC in multivariate normal mixtures, even for dimension as large as 20. To our knowledge, there exists no application of RJMCMC for such high-dimensional mixtures.

As by-products of our effort on the development of TTMCMC, we propose a novel methodology to summarize the posterior distributions of the mixture densities, providing a way to obtain the mode of the posterior distribution of the densities and the associated highest posterior density credible regions. Based on our method, we also propose a criterion to assess convergence of variable-dimensional algorithms. These methods of summarization and convergence assessment are applicable to general problems, not just to mixtures.

1 Introduction

Markov chain Monte Carlo (MCMC) is known to have revolutionized Bayesian computation. In modern times, it is often required to analyze high-dimensional,

complex data, and the Bayesian paradigm, with the MCMC machinery, provides an ideal package to the statistical scientist for the purpose. As is to be anticipated, to simulate from complex Bayesian posteriors, development of quite sophisticated MCMC methods were necessary, and various approaches based on component-wise and joint updating of the parameters, such as the adaptive direction sampling (Gilks, Roberts and George (1994)), the multiple-try Metropolis method (Liu, Liang and Wong (2000)), the auxiliary variable approach (Storvik (2011)), parallel MCMC methods (Martino et al. (2016)), have emerged in response to the needs of the modern Bayesian.

However, the above methods are appropriate when the number of parameters is known in advance. When one of the unknown parameters is the number of parameters itself, then none of the traditional MCMC methods are applicable, irrespective of how sophisticated they are. Indeed, simultaneous inference on both model and parameter space is an issue that is fundamental to modern statistical practice (Sisson (2005)). Examples of such problems arise in mixture analysis where the parameters associated with the mixture components as well as the number of mixture components are unknown (see, for example, Richardson and Green (1997)); in change point analysis where the locations and the number of change points are unknown (see, for example, Green (1995)); in variable selection problems where the number of covariates and the associated coefficients are unknown (Dellaportas, Forster and Ntzoufras (2002), Dellaportas and Forster (1999)); in spline smoothing where the location and the number of knots are unknown (see Denison, Mallick and Smith (1998) for instance); in continuous wavelet representation of unknown functions with a finite, but unknown number of wavelet basis functions and the corresponding parameters (Chu, Clyde and Liang (2009)); in autoregressive time series models where the order of the autoregression and the associated parameters are unknown (Vermaak et al. (2004)); in factor analysis where the dimension of the latent factor loading matrix and the associated parameters are unknown (Lopes and West (2004)); in spatial point processes where the locations and the number of points are random (see Møller and Waagepetersen (2004)); to name only a few.

A general MCMC strategy which can explore variable dimensional spaces by jumping between different dimensions has been proposed by Green (1995), and is well known as Reversible Jump MCMC (RJMCMC). The versatility of the methodology is well-reflected in the large varieties of variable-dimensional problems to which it has been applied; indeed, all the aforementioned examples make use of RJMCMC. However, one difficulty is frequently encountered when designing reversible jump algorithms is the construction of efficient proposals. Typically, dimension jumping moves in reversible jump samplers exhibit much lower acceptance rate than in fixed-dimensional moves. Al-Awadhi and Jennison (2004) observed that models with multimodal distributions yield particularly low acceptance rates. There have been many attempts of creating automatic RJMCMC samplers which also maintain high acceptance rates; see, for example, Brooks, Giudici and Roberts (2003), Robert (2003), Green (2003), Godsill (2003), Robert and Casella

(2004), Sisson (2005), Fan and Sisson (2011) and the references therein. However, in spite of the commendable attempts, these ideas are perhaps relevant in quite specific models with several restrictive assumptions; see Robert and Casella (2004), Sisson (2005), Fan and Sisson (2011).

The issues discussed above point towards the need to develop general and natural move types that can change dimensions as well as update the other (within model) parameters simultaneously, while maintaining reasonable acceptance rates and mixing properties. In this regard, the transformation based MCMC (TMCMC) approach of Dutta and Bhattacharya (2014) in the fixed dimensional set-up provides the necessary motivation. The key concept of TMCMC is to propose a move-type from a set of available move-types, simulate a single, one-dimensional random variable from some arbitrary distribution and propose simple deterministic transformations to all the parameters using the one-dimensional random variable, within the proposed move-type. In this article, we show that the same concept of deterministic transformations of a single random variable can be exploited to construct, for any general variable dimensional problem, a generic and effective dimension-hopping sampler which can change dimensions and update all the parameters of the proposed model in a single block while maintaining reasonable acceptance rates and mixing properties. We refer to this general variable dimensional MCMC sampler as Transdimensional Transformation based Markov Chain Monte Carlo (TTMCMC).

1.1 Overview of contributions and organisation of this paper

Before a formal introduction of TTMCMC, it is necessary to provide a brief overview of the basic concept of TMCMC. We do this in Section 2.

We introduce TTMCMC in Section 3, and in Section 4 we extend our proposed methodology to more general situations where one wishes to jump more than one dimension at a time. That TTMCMC thus developed closely qualifies as an automatic variable dimensional sampler, is argued in Section 5.

Although our proposed sampler is quite general and readily applicable to all transdimensional sampling frameworks, for the purpose of illustration and comparison with RJMCMC we restrict ourselves to gamma and normal mixture problems with unknown number of components. In this regard, in Section 6 we first conduct four simulation experiments with gamma mixtures with true number of components being 1, 2, 3 and 4, respectively. In Section 7, we provide details regarding applications of our methods to analyse three well-studied real data sets, namely, the enzyme, acidity and the galaxy data (see Richardson and Green (1997), for instance). In Section 8, we demonstrate the application of TTMCMC in mixtures of multivariate normal densities. In particular, we consider three simulation studies for dimensions 3, 10 and 20.

We show that the simplest possible TTMCMC algorithm, which is based on additive transformations, puts up excellent performance in all the examples, even in

all the multivariate scenarios, providing ample support to our claim of automation. Also interestingly, the TTMCMC applications are able to capture very precise information regarding the number of mixture components, for both simulated and real data sets. None of the previous methods (see Richardson and Green (1997) and the references therein) were able to capture so precise information as TTM-CMC. Moreover, there possibly does not exist any RJMCMC algorithm that works for multivariate mixtures with dimension as high as 20. Hence, from the high-dimensional perspective, TTMCMC is clearly far ahead of RJMCMC.

For the gamma mixtures and the normal mixtures associated with the real data applications we compare additive TTMCMC with the closest RJMCMC analogue of additive TTMCMC, based on random walk proposals. This RJMCMC algorithm seems to be the more natural, intuitive and computationally far simpler alternative to the random walk-motivated "automatic generic transdimensional RJMCMC sampler" proposed in Green (2003). Indeed, the approach of Green (2003) is appropriate only when a small set of models is considered in the variable-dimensional problem, and as such not a viable option for our normal mixtures with maximum of 30 components; see Section 6.6 for details.

Unfortunately, the random walk RJMCMC algorithm analogue of additive TTMCMC fails to produce satisfactory results in a way that even convergence is not assured in any of the examples. In particular, with the same scales of additive TTMCMC, random walk RJMCMC yields extremely poor acceptance rate in general. Moreover, the RJMCMC-based posterior of the number of components tends to assign higher posterior probabilities to implausibly large values, clearly indicating lack of convergence. We argue that the same issue persists with general RJMCMC algorithms. This suggests that complex and difficult-to-implement algorithms with extremely large convergence time are required for RJMCMC to yield sensible results, and that there is no default choice of such algorithms. On the other hand, the potentiality of additive TTMCMC in conjunction with the results of our experiments demonstrate that additive TTMCMC is close to qualifying as the default variable-dimensional algorithm, even for large dimensions.

We summarize our work and make concluding remarks in Section 9. Additional details are provided in the supplement (Das and Bhattacharya (2017)), whose sections have the prefix "S-" when referred to in this paper.

2 A brief overview of the key idea of TMCMC

In order to obtain a valid algorithm based on transformations, Dutta and Bhattacharya (2014) design appropriate move types so that detailed balance and irreducibility hold. We first illustrate the basic idea of transformation based moves with a simple example. Given that we are in the current state x, we may propose the "forward move" $x' = x + \varepsilon$, where $\varepsilon > 0$ is a simulation from some arbitrary density $\varrho(\cdot)$ which is supported on the positive part of the real line. To move back

to x from x', we need to apply the "backward transformation" $x' - \varepsilon$. In general, given ε and the current state x, we shall denote the forward transformation by $T(x, \varepsilon)$, and the backward transformation by $T^b(x, \varepsilon)$. For fixed ε the forward and backward transformations must be one-to-one and onto, and must satisfy $T^b(T(x, \varepsilon), \varepsilon) = x = T(T^b(x, \varepsilon), \varepsilon)$; see Dutta and Bhattacharya (2014) for a detailed discussion regarding these.

The simple idea discussed above has been generalized to the multi-dimensional situation by Dutta and Bhattacharya (2014). Remarkably, for any dimension, the moves can be constructed by simple deterministic transformations of the one-dimensional random variable ε, which is simulated from any arbitrary distribution on some relevant support. We provide some examples of such moves in the next section after introducing some necessary notation borrowed from Dutta and Bhattacharya (2014).

2.1 Notation

Suppose that $\mathcal{X}$ is a k-dimensional space of the form $\mathcal{X} = \prod_{i=1}^{k} \mathcal{X}_i$ so that $T = (T_1, \ldots, T_k)$ where each $T_i : \mathcal{X}_i \times \mathcal{D} \to \mathcal{X}_i$, for some set $\mathcal{D}$, are the component-wise transformations. Let $z = (z_1, \ldots, z_k)$ be a vector of indicator variables, where, for $i = 1, \ldots, k$, $z_i = 1$ and $z_i = -1$ indicate, respectively, application of forward transformation and backward transformation to x_i, and let $z_i = 0$ denote no change to x_i. This "no change" step is sufficient to ensure irreducibility of TMCMC in non-additive transformations; see Dutta and Bhattacharya (2014). Given any such indicator vector z, let us define $T_z = (g_{1,z_1}, g_{2,z_2}, \ldots, g_{k,z_k})$ where

$$g_{i,z_i} = \begin{cases} T_i^b & \text{if } z_i = -1, \\ x_i & \text{if } z_i = 0, \\ T_i & \text{if } z_i = 1. \end{cases}$$

Corresponding to any given z, we also define the following 'conjugate' vector $z^c = (z_1^c, z_2^c, \ldots, z_k^c)$, where

$$z_i^c = -z_i.$$

With this definition of z^c, T_{z^c} can be interpreted as the conjugate of T_z.

Since 3^k values of z are possible, it is clear that T, via z, induces 3^k many types of 'moves' of the forms $\{T_{z_i}; i = 1, \ldots, 3^k\}$ on the state-space. Suppose now that there is a subset $\mathcal{Y}$ of $\mathcal{D}$ such that the sets $T_{z_i}(\mathbf{x}, \mathcal{Y})$ and $T_{z_j}(\mathbf{x}, \mathcal{Y})$ are disjoint for every $z_i \neq z_j$. In fact, $\mathcal{Y}$ denotes the support of the distribution $\varrho(\cdot)$ from which ε is simulated. This mutual exclusiveness is required to satisfy the detailed balance property; see Dutta and Bhattacharya (2014) for the details. Thus, although $\mathcal{D}$ denotes the actual range of values that ε can assume in principle, for implementation of TMCMC we must restrict the support of ε to $\mathcal{Y}$.

2.2 Examples of transformations on two-dimensional state-space using single ε

Although for the sake of illustration, we provide below examples pertaining to two-dimensional cases it is important to remark at the outset that these examples can be easily generalized to any dimension; see Dutta and Bhattacharya (2014).

1. *Additive transformation*: Suppose $\mathcal{X} = \mathcal{D} = \mathbb{R}^2$. With two positive scale parameters a_1 and a_2, we can then consider the following additive transformation: $T_{(1,1)}(\mathbf{x}, \varepsilon) = (x_1 + a_1\varepsilon, x_2 + a_2\varepsilon)$, $T_{(-1,1)}(\mathbf{x}, \varepsilon) = (x_1 - a_1\varepsilon, x_2 + a_2\varepsilon)$, $T_{(1,-1)}(\mathbf{x}, \varepsilon) = (x_1 + a_1\varepsilon, x_2 - a_2\varepsilon)$ and $T_{(-1,-1)}(\mathbf{x}, \varepsilon) = (x_1 - a_1\varepsilon, x_2 - a_2\varepsilon)$. We set $\mathcal{Y} = (0, \infty)$.

2. *Multiplicative transformation*: Suppose $\mathcal{X} = \mathcal{D} = \mathbb{R}^2$. Then we may consider the following multiplicative transformation: $T_{(1,1)}(\mathbf{x}, \varepsilon) = (x_1\varepsilon, x_2\varepsilon)$, $T_{(-1,1)}(\mathbf{x}, \varepsilon) = (x_1/\varepsilon, x_2\varepsilon)$, $T_{(1,-1)}(\mathbf{x}, \varepsilon) = (x_1\varepsilon, x_2/\varepsilon)$, $T_{(-1,-1)}(\mathbf{x}, \varepsilon) = (x_1/\varepsilon, x_2/\varepsilon)$, $T_{(1,0)}(\mathbf{x}, \varepsilon) = (x_1\varepsilon, x_2)$, $T_{(1,0)}(\mathbf{x}, \varepsilon) = (x_1\varepsilon, x_2)$, $T_{(-1,0)}(\mathbf{x}, \varepsilon) = (x_1/\varepsilon, x_2)$, $T_{(0,1)}(\mathbf{x}, \varepsilon) = (x_1, x_2\varepsilon)$, $T_{(0,-1)}(\mathbf{x}, \varepsilon) = (x_1, x_2/\varepsilon)$, $T_{(0,0)}(\mathbf{x}, \varepsilon) = (x_1, x_2)$. We choose $\mathcal{Y} = \{(-1, 1) - \{0\}\}$.

3. *Additive-multiplicative transformation*: It is possible to combine additive and multiplicative transformations, but here we need at least two ε's, one for the additive, and another for the multiplicative transformation. For instance, if $\mathcal{X} = \mathcal{D} = \mathbb{R}^2$, then we may consider the following moves: $T_{(1,1)}(\mathbf{x}, \varepsilon_1, \varepsilon_2) = (x_1 + \varepsilon_1, x_2\varepsilon_2)$, $T_{(-1,1)}(\mathbf{x}, \varepsilon_1, \varepsilon_2) = (x_1 - \varepsilon_1, x_2\varepsilon_2)$, $T_{(1,-1)}(\mathbf{x}, \varepsilon_1, \varepsilon_2) = (x_1 + \varepsilon_1, x_2/\varepsilon_2)$, $T_{(-1,-1)}(\mathbf{x}, \varepsilon_1, \varepsilon_2) = (x_1 - \varepsilon_1, x_2/\varepsilon_2)$, $T_{(1,0)}(\mathbf{x}, \varepsilon_1, \varepsilon_2) = (x_1 + \varepsilon_1, x_2)$, $T_{(-1,0)}(\mathbf{x}, \varepsilon_1, \varepsilon_2) = (x_1 - \varepsilon_1, x_2)$, $T_{(0,1)}(\mathbf{x}, \varepsilon_1, \varepsilon_2) = (x_1, x_2\varepsilon_2)$, $T_{(0,-1)}(\mathbf{x}, \varepsilon_1, \varepsilon_2) = (x_1, x_2/\varepsilon_2)$, $T_{(0,0)}(\mathbf{x}, \varepsilon_1, \varepsilon_2) = (x_1, x_2)$. We let $\mathcal{Y} = (0, \infty) \times \{(-1, 1) - \{0\}\}$. Although this example uses two ε's for two dimensions, it is important to note that for any dimension higher than two, at most two ε's will be required for validity of additive-multiplicative TMCMC, one for the additive part and another for the multiplicative part, irrespective of the dimensionality. Thus, the minimum effective dimensionality of additive TMCMC and multiplicative TMCMC is 1, while that of additive-multiplicative TMCMC in this setting is 2, for any dimensionality greater than one.

The key observation underlying the above examples is that it is always possible to construct valid transformations in high-dimensional spaces using combinations of appropriate transformations on one-dimensional spaces. These transformations and the underlying principle remain valid even in TTMCMC.

2.3 The general form of the TMCMC algorithm

For a $k(\geq 1)$-dimensional target distribution, with current state $\mathbf{x} = (x_1, \ldots, x_k)$, Dutta and Bhattacharya (2014) apply forward and backward transformations to x_i with probabilities p_i and q_i, respectively and keep x_i unchanged with probability $1 - p_i - q_i$, for $i = 1, \ldots, k$. Thus, z can now be interpreted as a random vector

such that for $i = 1, \ldots, k$, $z_i \in \{-1, 0, 1\}$ with probabilities $q_i, 1 - p_i - q_i, p_i$, respectively. Thus, we simulate $z_i \sim \text{Multinomial}(1; p_i, q_i, 1 - p_i - q_i)$ independently for $i = 1, \ldots, k$, draw $\varepsilon \sim \varrho(\cdot)$, and form the proposed move $\mathbf{x} \mapsto \mathbf{x}' = T_z(\mathbf{x}, \varepsilon)$, which is accepted with probability

$$\alpha(\mathbf{x}, \varepsilon) = \min\left(1, \frac{P(z^c)}{P(z)} \frac{\pi(\mathbf{x}')}{\pi(\mathbf{x})} \left| \frac{\partial(T_z(\mathbf{x}, \varepsilon), \varepsilon)}{\partial(\mathbf{x}, \varepsilon)} \right| \right), \tag{2.1}$$

where

$$\frac{P(z^c)}{P(z)} = \prod_{\{i_1 : z_{i_1} = -1\}} \frac{p_{i_1}}{q_{i_1}} \prod_{\{i_2 : z_{i_2} = 1\}} \frac{q_{i_2}}{p_{i_2}}.$$

Note that the acceptance ratio is *always* independent of the proposal density ϱ.

The redundant move-type $\mathbf{x} \mapsto \mathbf{x}$ has positive probability of occurrence, and hence Dutta and Bhattacharya (2014) suggest rejection of this move whenever it appears. That is, sampling of z is to be continued until at least one $z_i \neq 0$. This rejection sampling of z is very efficient since the rejection region is a singleton and has very small probability, particularly in high dimensions. The normalizing constant that arises because of this truncation cancels in the acceptance ratio of TMCMC, as shown in Dutta and Bhattacharya (2014).

3 TTMCMC for updating the dimension and the parameters in a single block using deterministic transformations of a single random variable

First, we illustrate the main idea of TTMCMC informally using the additive transformation.

3.1 Illustration of the key idea of TTMCMC with a simple example

Assume that the current state is $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$. We first randomly select $u = (u_1, u_2, u_3) \sim \text{Multinomial}(1; w_b, w_d, w_{nc})$, where w_b, w_d, w_{nc} (> 0) such that $w_b + w_d + w_{nc} = 1$ are the probabilities of birth, death, and no-change moves, respectively. That is, if $u_1 = 1$, then we increase the dimensionality from 2 to 3; if $u_2 = 1$, then we decrease the dimensionality from 2 to 1, and if $u_3 = 1$, then we keep the dimensionality unchanged. In the latter case, when the dimensionality is unchanged, the acceptance probability remains the same as in TMCMC, given by (2.1).

If $u_1 = 1$, we can increase the dimensionality by first selecting one of x_1 and x_2 with probability $1/2$; for the sake of clarity, we assume that x_1 has been selected, Here, as in TMCMC, we draw $\varepsilon \sim \varrho(\cdot)$, where $\varrho(\cdot)$ is supported on the positive part of the real line, and draw z_2 where $z_2 = 1$ with probability p_2 and $z_2 = -1$ with probability $1 - p_2$. Also, as before, $z^c = (z_1^c, z_2^c)$ is the

conjugate of z, where $z_i^c = -z_i$. We then construct the move-type $T_{b,z}(\mathbf{x}, \varepsilon) = (x_1 + a_1\varepsilon, x_1 - a_1\varepsilon, x_2 + z_2 a_2\varepsilon) = (g_{1,z_1=1}(x_1, \varepsilon), g_{1,z_1^c=-1}(x_1, \varepsilon), g_{2,z_2}(x_2, \varepsilon))$, say. We relabel $\mathbf{x}' = T_{b,z}(\mathbf{x}, \varepsilon) = (x_1 + a_1\varepsilon, x_1 - a_1\varepsilon, x_2 + z_2 a_2\varepsilon)$ as (x_1', x_2', x_3'). Thus, $T_{b,z}(\mathbf{x}, \varepsilon)$ increases the dimension from 2 to 3.

We accept this birth move with probability

$$a_b(\mathbf{x}, \varepsilon) = \min\left\{1, \frac{1}{3} \times \frac{w_d}{w_b} \times \frac{p_2^{I_{\{1\}}(z_2^c)} q_2^{I_{\{-1\}}(z_2^c)}}{p_2^{I_{\{1\}}(z_2)} q_2^{I_{\{-1\}}(z_2)}} \right.$$
$$\left. \times \frac{\pi(x_1 + a_1\varepsilon, x_1 - a_1\varepsilon, x_2 + z_2 a_2\varepsilon)}{\pi(x_1, x_2)} \times \left| \frac{\partial(T_{b,z}(\mathbf{x}, \varepsilon))}{\partial(\mathbf{x}, \varepsilon)} \right| \right\}. \tag{3.1}$$

In (3.1),

$$\left| \frac{\partial(T_{b,z}(\mathbf{x}, \varepsilon))}{\partial(\mathbf{x}, \varepsilon)} \right| = \left| \frac{\partial(x_1 + a_1\varepsilon, x_1 - a_1\varepsilon, x_2 + z_2 a_2\varepsilon)}{\partial(x_1, x_2, \varepsilon)} \right|$$

$$= \left| \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \\ a_1 & -a_1 & z_2 a_2 \end{pmatrix} \right| = 2a_1. \tag{3.2}$$

Now let us illustrate the problem of returning to $=(x_1, x_2)$ ($\in \mathbb{R}^2$) from $T_{b,z}(\mathbf{x}, \varepsilon) = (x_1 + a_1\varepsilon, x_1 - a_1\varepsilon, x_2 + z_2 a_2\varepsilon)$ ($\in \mathbb{R}^3$). For our purpose, we can select $x_1 + a_1\varepsilon$ with probability $1/3$; then select $x_1 - a_1\varepsilon$ from the remaining two elements with probability $1/2$, and form the average $x_1^* = ((x_1 + a_1\varepsilon) + (x_1 - a_1\varepsilon))/2 = x_1$. For non-additive transformations, we can consider the averages of the backward moves of each of the selected elements. Even in this additive transformation example, after simulating ε as before we can consider the respective backward moves of $x_1 + a_1\varepsilon$ and $x_1 - a_1\varepsilon$, both yielding x_1, and then take the average denoted by x_1^*. For the remaining element $x_2 + z_2 a_2\varepsilon$, we need to simulate z_2^c and then consider the move $(x_2 + z_2 a_2\varepsilon) + z_2^c a_2\varepsilon = x_2$. Thus, we can return to (x_1, x_2) using this strategy.

Letting $\mathbf{x}' = (x_1', x_2', x_3')$, and denoting the average involving the first two elements by x_1^*, the death move is then given by $\mathbf{x}'' = T_{d,z}(\mathbf{x}', \varepsilon) = (x_1^*, x_3' + z_2^c a_2\varepsilon) = (\frac{x_1' + x_2'}{2}, x_3' + z_2^c a_2\varepsilon)$. Now observe that for returning to (x_1', x_2') from x_1^*, we must have $x_1^* + a_1\varepsilon^* = x_1'$ and $x_1^* - a_1\varepsilon^* = x_2'$, which yield $\varepsilon^* = (x_1' - x_2')/2a_1$. Hence, the Jacobian associated with the death move in this case is given by

$$\left| \frac{\partial(T_{d,z}(\mathbf{x}', \varepsilon), \varepsilon^*, \varepsilon)}{\partial(\mathbf{x}', \varepsilon)} \right| = \left| \frac{\partial(\frac{x_1' + x_2'}{2}, x_3' + z_2^c a_2\varepsilon, \frac{x_1' - x_2'}{2a_1}, \varepsilon)}{\partial(x_1', x_2', x_3', \varepsilon)} \right|$$

$$= \left| \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2a_1} & 0 \\ \frac{1}{2} & 0 & -\frac{1}{2a_1} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & z_2^c a_2 & 0 & 1 \end{pmatrix} \right| = \frac{1}{2a_1}. \tag{3.3}$$

We accept this death move with probability

$$a_d(\mathbf{x}'', \varepsilon, \varepsilon^*) = \min\left\{1, 3 \times \frac{w_b}{w_d} \times \frac{P(z^c)}{P(z)} \frac{\pi(\mathbf{x}'')}{\pi(\mathbf{x}')} \left| \frac{\partial(T_{d,z}(\mathbf{x}', \varepsilon), \varepsilon^*, \varepsilon)}{\partial(\mathbf{x}', \varepsilon)} \right| \right\}$$

$$= \min\left\{1, 3 \times \frac{w_b}{w_d} \times \frac{p_2^{I_{\{1\}}(z_2^c)} q_2^{I_{\{-1\}}(z_2^c)}}{p_2^{I_{\{1\}}(z_2)} q_2^{I_{\{-1\}}(z_2)}} \times \frac{\pi(\mathbf{x}'')}{\pi(\mathbf{x}')} \times \frac{1}{2a_1}\right\}.$$

(3.4)

In the general situation, we shall make the birth, death and no-change probabilities w_b, w_d, w_{nc} depend upon the current dimension k, and denote them by $w_{b,k}$, $w_{d,k}$ and $w_{nc,k}$, respectively, satisfying $w_{b,k} + w_{d,k} + w_{nc,k} = 1$ for every $k \geq 1$. Note that when the current dimension $k = 1$, then $w_{d,k} = 0$, as $k \geq 1$. Similarly, if in some cases there is reason to assume that the number of parameters can not exceed some finite quantity denoted by $k_{\max}$, then $w_{b,k_{\max}} = 0$.

Figure 1 illustrates the idea of TTMCMC schematically, and compares it with the RJMCMC principle, shown diagrammatically in Figure 2. As illustrated, for RJMCMC, the necessary " dimension matching" criterion is satisfied, but the criterion is not satisfied, indeed, not necessary, for TTMCMC.

3.2 General TTMCMC algorithm for jumping one dimension at a time

We now provide the TTMCMC algorithm in the general case, as follows.

Algorithm 3.1. *General TTMCMC algorithm based on a single ε.*

- Let the initial value be $\mathbf{x}^{(0)} \in \mathbb{R}^k$.
- For $t = 0, 1, 2, \ldots$

  1. Generate $u = (u_1, u_2, u_3) \sim \text{Multinomial}(1; w_{b,k}, w_{d,k}, w_{nc,k})$.
  2. If $u_1 = 1$ (increase dimension), then
     (a) Randomly select a co-ordinate from $\mathbf{x}^{(t)} = (x_1^{(t)}, \ldots, x_k^{(t)})$ assuming uniform probability $1/k$ for each co-ordinate. Let j denote the chosen co-ordinate.
     (b) Generate $\varepsilon \sim \varrho(\cdot)$ and for $i = 1, \ldots, k; i \neq j$ simulate

        $$z_i \sim \text{Multinomial}(1; p_i, q_i, 1 - p_i - q_i)$$

        independently.
     (c) Propose the following birth move:

        $$\mathbf{x}' = T_{b,z}(\mathbf{x}^{(t)}, \varepsilon)$$

        $$= \left(g_{1,z_1}(x_1^{(t)}, \varepsilon), \ldots, g_{j-1,z_{j-1}}(x_{j-1}^{(t)}, \varepsilon), g_{j,z_j=1}(x_j^{(t)}, \varepsilon),\right.$$

        $$\left. g_{j,z_j^c=-1}(x_j^{(t)}, \varepsilon), g_{j+1,z_{j+1}}(x_{j+1}^{(t)}, \varepsilon), \ldots, g_{k,z_k}(x_k^{(t)}, \varepsilon)\right).$$

        Re-label the elements of $\mathbf{x}'$ as $(x_1', x_2', \ldots, x_{k+1}')$.

(d) Calculate the acceptance probability of the birth move $\mathbf{x}'$:

$$a_b(\mathbf{x}^{(t)}, \varepsilon) = \min\left\{1, \frac{1}{k+1} \times \frac{w_{d,k+1}}{w_{b,k}}\right.$$

$$\left. \times \frac{P_{(j)}(z^c)}{P_{(j)}(z)} \frac{\pi(\mathbf{x}')}{\pi(\mathbf{x}^{(t)})} \left| \frac{\partial(T_{b,z}(\mathbf{x}^{(t)}, \varepsilon))}{\partial(\mathbf{x}^{(t)}, \varepsilon)} \right| \right\},$$

where

$$P_{(j)}(z) = \prod_{i \neq j=1}^{k} p_i^{I_{\{1\}}(z_i)} q_i^{I_{\{-1\}}(z_i)},$$

and

$$P_{(j)}(z^c) = \prod_{i \neq j=1}^{k} p_i^{I_{\{1\}}(z_i^c)} q_i^{I_{\{-1\}}(z_i^c)}.$$

(e) Set

$$\mathbf{x}^{(t+1)} = \begin{cases} \mathbf{x}' & \text{with probability } a_b(\mathbf{x}^{(t)}, \varepsilon), \\ \mathbf{x}^{(t)} & \text{with probability } 1 - a_b(\mathbf{x}^{(t)}, \varepsilon). \end{cases}$$

3. If $u_2 = 1$ (decrease dimension), then
   (a) Generate $\varepsilon \sim \varrho(\cdot)$.
   (b) Randomly select co-ordinate j with probability $1/k$, and randomly select co-ordinate j' from the remaining co-ordinates with probability $1/(k-1)$. Let $x_j^* = (g_{j,z_j^c=-1}(x_j, \varepsilon) + g_{j',z_{j'}=1}(x_{j'}, \varepsilon))/2$; re-place the co-ordinate x_j drawn first by the average x_j^*, and delete $x_{j'}$.
   (c) Simulate z by generating independently, for $i = 1, \ldots, k$, but $i \neq j, j'$, $z_i \sim \text{Multinomial}(1; p_i, q_i, 1 - p_i - q_i)$. For $i \neq j, j'$, apply the transformation $x_i' = g_{i,z_i}(x_i^{(t)}, \varepsilon)$.
   (d) Propose the following death move:

$$\mathbf{x}' = T_{d,z}(\mathbf{x}^{(t)}, \varepsilon)$$

$$= (g_{1,z_1}(x_1^{(t)}, \varepsilon), \ldots, g_{j-1,z_{j-1}}(x_{j-1}^{(t)}, \varepsilon), x_j^*, g_{j+1,z_{j+1}}(x_{j+1}^{(t)}, \varepsilon),$$

$$\ldots, g_{j'-1,z_{j'-1}}(x_{j'-1}^{(t)}, \varepsilon), g_{j'+1,z_{j'+1}}(x_{j'+1}^{(t)}, \varepsilon), \ldots, g_{k,z_k}(x_k^{(t)}, \varepsilon)).$$

Re-label the elements of $\mathbf{x}'$ as $(x_1', x_2', \ldots, x_{k-1}')$.

   (e) Solve for ε^* from the equations $g_{j,z_j=1}(x_j^*, \varepsilon^*) = x_j$ and $g_{j,z_j^c=-1}(x_j^*, \varepsilon^*) = x_{j'}$ and express ε^* in terms of x_j and $x_{j'}$.

(f) Calculate the acceptance probability of the death move:

$$a_d(\mathbf{x}^{(t)}, \varepsilon, \varepsilon^*) = \min\left\{1, k \times \frac{w_{b,k-1}}{w_{d,k}}\right.$$

$$\left. \times \frac{P_{(j,j')}(z^c)}{P_{(j,j')}(z)} \frac{\pi(\mathbf{x}')}{\pi(\mathbf{x}^{(t)})} \left| \frac{\partial (T_{d,z}(\mathbf{x}^{(t)}, \varepsilon), \varepsilon^*, \varepsilon)}{\partial (\mathbf{x}^{(t)}, \varepsilon)} \right| \right\},$$

where

$$P_{(j,j')}(z) = \prod_{i \neq j, j'=1}^{k} p_i^{I_{\{1\}}(z_i)} q_i^{I_{\{-1\}}(z_i)},$$

and

$$P_{(j,j)}(z^c) = \prod_{i \neq j, j'=1}^{k} p_i^{I_{\{1\}}(z_i^c)} q_i^{I_{\{-1\}}(z_i^c)}.$$

(g) Set

$$\mathbf{x}^{(t+1)} = \begin{cases} \mathbf{x}' & \text{with probability } a_d(\mathbf{x}^{(t)}, \varepsilon, \varepsilon^*), \\ \mathbf{x}^{(t)} & \text{with probability } 1 - a_d(\mathbf{x}^{(t)}, \varepsilon, \varepsilon^*). \end{cases}$$

4. If $u_3 = 1$ (dimension remains unchanged), then implement steps (1), (2), (3) of Algorithm 3.1 of Dutta and Bhattacharya (2014).

• End for

In Sections S-1 and S-2 of the supplement, we provide the proofs of detailed balance and ergodicity (irreducibility and aperiodicity) of the above TTMCMC method.

3.2.1 *Observations regarding Algorithm* 3.1.

• Note that the acceptance probabilities are independent of the proposal density $\varrho(\cdot)$ irrespective of its form, just as in TMCMC. The reason is that in TTMCMC we simulate $\varepsilon \sim \varrho$, for some appropriate density ϱ, for increasing, as well as for decreasing dimension (see the proof of detailed balance in Section S-1 for the precise details). In other words, the "dimension-matching" criterion of RJM-CMC is not required for TTMCMC. Indeed, recall that, to accomplish the birth step in RJMCMC one needs to simulate an ε, but in the death step two randomly chosen components are averaged to reduce the dimension, and no simulation of ε is done. As such, in RJMCMC the dimension-matching criterion is responsible for the presence of the proposal density in the acceptance ratio.

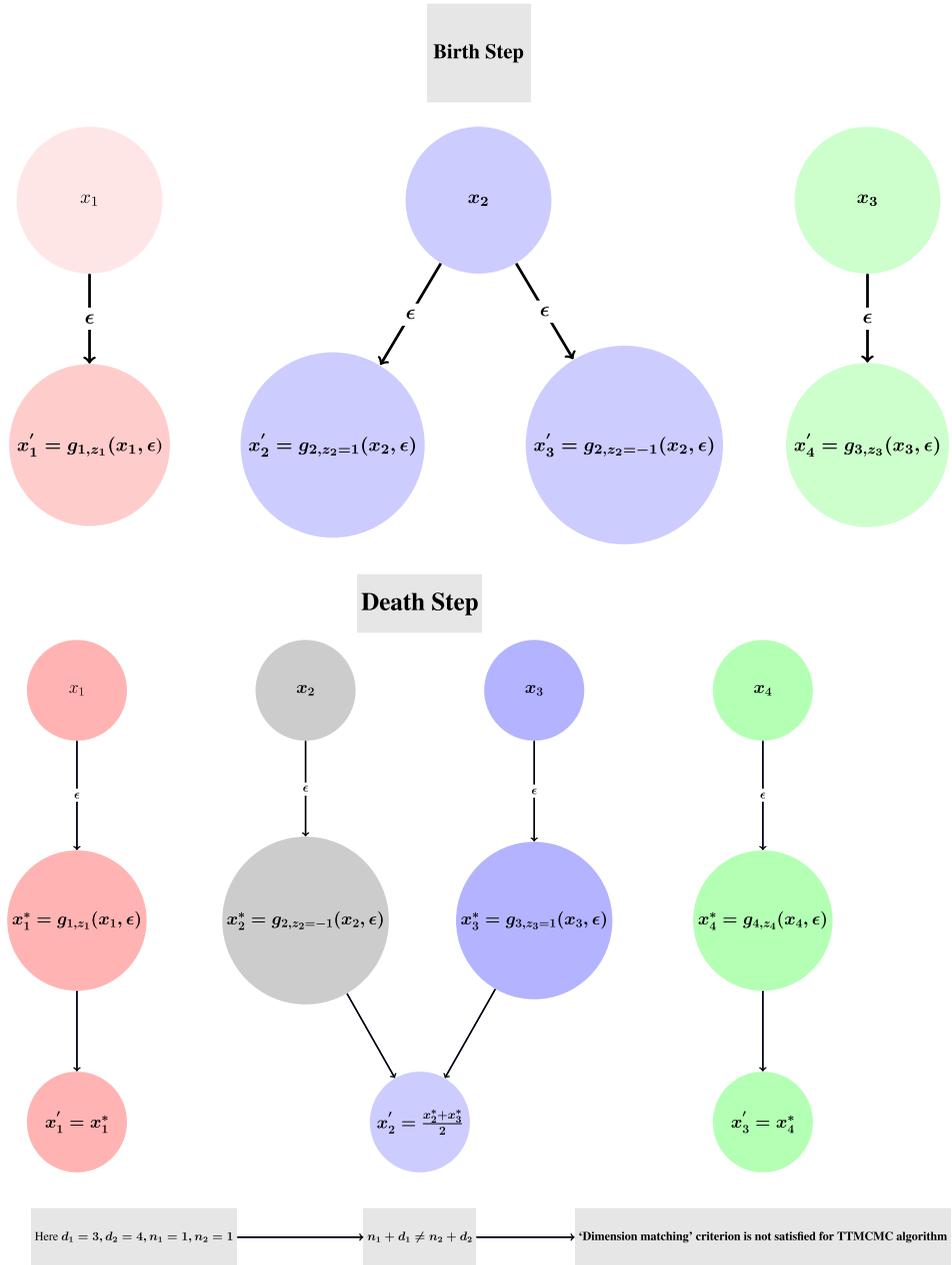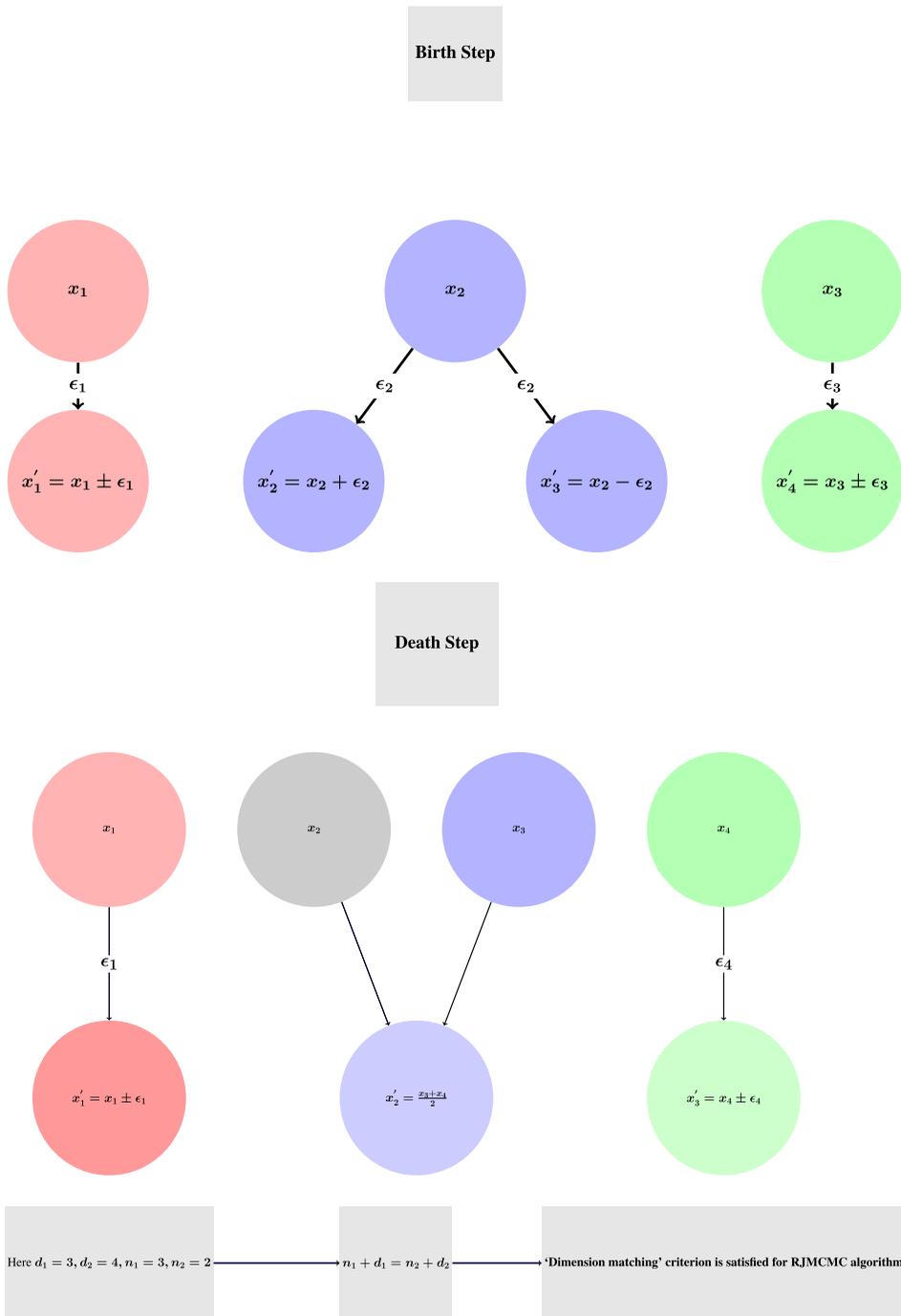**Figure 1** *Illustration of TTMCMC algorithm for jumping between dimension* 3 *and* 4.

Birth Step

Death Step

Here $d_1 = 3, d_2 = 4, n_1 = 3, n_2 = 2$ → $n_1 + d_1 = n_2 + d_2$ → **'Dimension matching' criterion is satisfied for RJMCMC algorithm**

Figure 2 *Illustration of RJMCMC algorithm for jumping between dimension* 3 *and* 4.

- Consequently, it is not possible to interpret TTMCMC as a special case of RJM-CMC. Also, neither is RJMCMC a special case of TTMCMC, even though in fixed-dimensional problems, TMCMC with additive transformations contains the random walk Metropolis algorithm as a special case when as many ε's as the number of variables to be updated are used for TMCMC.

- Independence of the acceptance ratio of the proposal density ϱ has pleasing consequences for TTMCMC in the sense that for any finite TTMCMC sample (which is always the case in practice), the possible bias in the acceptance probabilities of birth and death moves due to involvement of ϱ is absent. Since for RJMCMC this is not the case, the performance may be seriously affected. For instance, if ϱ is strictly bounded above by 1, then the birth move will have significantly greater acceptance probability than the death move. The advantage of TTMCMC and disadvantage of RJMCMC in this regard are clearly reflected in all our experiments that we report in this article.

- In the acceptance probabilities, $\frac{P_{(j)}(z^c)}{P_{(j)}(z)} = 1$ and $\frac{P_{(j,j')}(z^c)}{P_{(j,j')}(z)} = 1$ if $p_i = q_i$ for each i. This results in simplification of the acceptance ratio computation. The birth, death and the no-change probabilities given by $w_{b,k}$, $w_{d,k}$ and $w_{nc,k}$ can also be chosen to be equal for every $k > 1$, which will result in further simplification of the computation of the acceptance ratio.

- In our algorithm, the new variables created from one variable are never "necessarily adjacent". Even in the case of adjacency, our method does absolutely fine; indeed, for the death step, we only need to have appropriate positive probability of selecting the two variables for combining them into one (or deleting one) such that the detailed balance holds. Specifically, suppose that we create adjacent variables in the birth move. Then, in the corresponding death move we will choose adjacent pairs with appropriate probability and combine them into one. Alternatively, one may select two variables, but should reject the entire death move if the selected variables are not adjacent. In fact, the issue of adjacency is nothing specific to TTMCMC, and can be handled by RJMCMC as well as by TTMCMC.

3.3 Structured dependence within the moves

In Algorithm 3.1 we have assumed that for $i \in \{1, \ldots, k\} \setminus \{j\}$ and for $i \in \{1, \ldots, k\} \setminus \{j, j'\}$ (accordingly as the move-type is birth move or death move), z_i are independently simulated in every iteration. Although the co-ordinate-wise moves are dependent since the same ε is used for updating them, more flexible and structured dependence can be induced within the moves in the TTMCMC context. Such structured dependence allows for selecting the coordinate-wise forward or backward transformations in ways that take account of the posterior correlation between the parameters, thus facilitating more efficient moves.

Briefly, at each iteration, for $i = 1, \ldots, k$, we can reparameterize p_i and q_i as

$$p_i = \frac{\exp(\psi_{1i})}{\sum_{j=1}^{3} \exp(\psi_{ji})};$$

$$q_i = \frac{\exp(\psi_{2i})}{\sum_{j=1}^{3} \exp(\psi_{ji})}; \qquad (3.5)$$

$$1 - p_i - q_i = \frac{\exp(\psi_{3i})}{\sum_{j=1}^{3} \exp(\psi_{ji})},$$

where, for $j = 1, 2, 3$,

$$(\psi_{j1}, \psi_{j2}, \ldots, \psi_{jk}) \sim N_k(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \qquad (3.6)$$

independently, where $(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$; $j = 1, 2, 3$ may be estimated from a pilot run of TMCMC with the dimensionality fixed at $k = k_{\max}$. Specifically, from a pilot run of TMCMC with $p_i = q_i$, for each variable x_i, $i = 1, \ldots, k_{\max}$, we may consider the three empirical means of x_i associated with $z_i = 1, -1$ and 0, as good candidates for the ith components of $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ and $\boldsymbol{\mu}_3$, respectively. For the covariance matrices $\boldsymbol{\Sigma}_j$, the empirical estimates of the covariances between x_i and x_j associated with $(z_i = 1, z_j = 1)$, $(z_i = -1, z_j = -1)$, and $(z_i = 0, z_j = 0)$ may be considered as the (i, j)th elements of $\boldsymbol{\Sigma}_1$, $\boldsymbol{\Sigma}_2$ and $\boldsymbol{\Sigma}_3$, respectively. The above strategy yields three $k_{\max}$-dimensional vectors $\tilde{\boldsymbol{\mu}}_j$; $j = 1, 2, 3$, and three $k_{\max} \times k_{\max}$-dimensional covariance matrices $\tilde{\boldsymbol{\Sigma}}_j$; $j = 1, 2, 3$. The required k-dimensional $\boldsymbol{\mu}_j$ and $k \times k$-dimensional $\boldsymbol{\Sigma}_j$ are then simply relevant sub-vectors and sub-matrices of $\tilde{\boldsymbol{\mu}}_j$ and $\tilde{\boldsymbol{\Sigma}}_j$, respectively.

At each iteration of TTMCMC, we then first simulate $(\psi_{j1}, \psi_{j2}, \ldots, \psi_{jk})$; $j = 1, 2, 3$ using (3.6), obtain $\{p_i, q_i, 1 - p_i - q_i; i = 1, \ldots, k\}$ using (3.5); then given $\{p_i, q_i, 1 - p_i - q_i; i = 1, \ldots, k\}$ we simulate $z_i \sim \text{Multinomial}(1; p_i, q_i, 1 - p_i - q_i)$ independently as before, where $i \in \{1, \ldots, k\} \setminus \{j\}$ or $i \in \{1, \ldots, k\} \setminus \{j, j'\}$.

As in the case of TMCMC, it can be easily verified that our modified TTMCMC algorithm with this hierarchical dependence structure for the distribution of z satisfies detailed balance.

4 Jumping more than one dimensions at a time

We now consider the situations where instead of jumping one dimension, one wishes to jump several dimensions at a time. That is, we now consider the more general framework where $\mathbf{x} = (x_1, \ldots, x_k) \in \mathbb{R}^k$ and that we wish to increase the dimension to $k + m$, or to decrease the dimension from $k + m$ to k, where $1 \leq m \leq k$. It follows that TTMCMC can jump from k to $2k$ dimensions and from $2k$ to k dimensions at the maximum. RJMCMC does not have such restriction, but

jumping many dimensions at a time will only add to the general inefficiency of RJMCMC.

For an illustrative TTMCMC example where jumping more than one dimension is desired, assume that $k = 3$ and $m = 2$, so that it is required to jump from $\mathbb{R}^3$ to $\mathbb{R}^5$. For simplicity, we illustrate with the additive transformation. One may anticipate that this can be accomplished by simulating a single positive $\varepsilon \sim \varrho(\cdot)$, selecting, say, x_1 and x_2 at random without replacement from $\mathbf{x} = (x_1, x_2, x_3)$, simulating z_3, and then constructing the birth move $\mathbf{x}' = T_{b,z_2}(\mathbf{x}, \varepsilon) = (x_1 + a_1\varepsilon, x_1 - a_1\varepsilon, x_2 + a_2\varepsilon, x_2 - a_2\varepsilon, x_3 + z_3a_3\varepsilon) = (x_1', x_2', x_3', x_4', x_5')$. However, for this move, the dimension of $(\mathbf{x}, \varepsilon) = (x_1, x_2, x_3, \varepsilon)$ is 4, while that of $\mathbf{x}' = (x_1', x_2', x_3', x_4', x_5')$ is 5. In other words, the Jacobian $|\frac{\partial(T_{b,z}(\mathbf{x}, \varepsilon))}{\partial(\mathbf{x}, \varepsilon)}|$ is not well-defined.

To get past the above difficulty with dimensions, we need to simulate two ε's from $\varrho(\cdot)$: ε_1 for splitting x_1 into $x_1 + a_1\varepsilon_1$ and $x_1 - a_1\varepsilon_1$, and ε_2 for splitting x_2 into $x_2 + a_2\varepsilon_2$ and $x_2 - a_2\varepsilon_2$, and also to update x_3 to $x_3 + z_3a_3\varepsilon_2$ (ε_1 can also be used to update x_3). Hence, the birth move takes the form $\mathbf{x}' = T_{b,z_3}(\mathbf{x}, \varepsilon_1, \varepsilon_2) = (x_1 + a_1\varepsilon_1, x_1 - a_1\varepsilon_1, x_2 + a_2\varepsilon_2, x_2 - a_2\varepsilon_2, x_3 + z_3a_3\varepsilon_2) = (x_1', x_2', x_3', x_4', x_5')$. Now the dimensions of both $\mathbf{x}' = (x_1', x_2', x_3', x_4', x_5')$ and $(\mathbf{x}, \varepsilon_1, \varepsilon_2) = (x_1, x_2, x_3, \varepsilon_1, \varepsilon_2)$ are the same and equals 5; hence the Jacobian

$$
\left| \frac{\partial(T_{b,z_3}(\mathbf{x}, \varepsilon_1, \varepsilon_2))}{\partial(\mathbf{x}, \varepsilon_1, \varepsilon_2)} \right|
$$

$$
= \left| \frac{\partial(x_1 + a_1\varepsilon_1, x_1 - a_1\varepsilon_1, x_2 + a_2\varepsilon_2, x_2 - a_2\varepsilon_2, x_3 + z_3a_3\varepsilon_2)}{\partial(x_1, x_2, x_3, \varepsilon_1, \varepsilon_2)} \right|
$$

$$
= 4a_1a_2,
$$

is well-defined. The acceptance probability of the birth move in this example is given by

$$
a_b(\mathbf{x}, \varepsilon_1, \varepsilon_2) = \min\left\{ 1, \frac{1}{(3+2)(3+1)} \times \frac{w_{d,5}}{w_{b,3}} \times \frac{p_3^{I_{\{1\}}(z_3^c)} q_3^{I_{\{-1\}}(z_3^c)}}{p_3^{I_{\{1\}}(z_3)} q_3^{I_{\{-1\}}(z_3)}} \right.
$$

$$
\left. \times \frac{\pi(\mathbf{x}')}{\pi(\mathbf{x})} \times \left| \frac{\partial(T_{b,z_3}(\mathbf{x}, \varepsilon_1, \varepsilon_2))}{\partial(\mathbf{x}, \varepsilon_1, \varepsilon_2)} \right| \right\}
\tag{4.1}
$$

$$
= \min\left\{ 1, \frac{1}{20} \times \frac{w_{d,5}}{w_{b,3}} \times \frac{p_3^{I_{\{1\}}(z_3^c)} q_3^{I_{\{-1\}}(z_3^c)}}{p_3^{I_{\{1\}}(z_3)} q_3^{I_{\{-1\}}(z_3)}} \frac{\pi(\mathbf{x}')}{\pi(\mathbf{x})} \times 4a_1a_2 \right\}.
$$

For the corresponding death move, that is, for moving from $\mathbf{x}' = (x_1', x_2', x_3', x_4', x_5')$ to $\mathbf{x}'' = T_{d,z}(\mathbf{x}', \varepsilon_1) = (\frac{x_1'+x_2'}{2}, \frac{x_3'+x_4'}{2}, x_5' + z_3^c a_3\varepsilon_1) = (x_1'', x_2'', x_3'')$, we must have, for the reverse of this death move, $x_1'' + a_1\varepsilon_1^* = x_1'$, $x_1'' - a_1\varepsilon_1^* = x_2'$, $x_2'' + a_2\varepsilon_2^* = x_3'$, $x_2'' - a_2\varepsilon_2^* = x_4'$. The first two equations yield $\varepsilon_1^* = \frac{x_1'-x_2'}{2a_1}$ and the last

two equations yield $\varepsilon_2^* = \frac{x_3' - x_4'}{2a_2}$. The Jacobian is given by

$$
\left| \frac{\partial(T_{d,z_3}(\mathbf{x}', \varepsilon_1); \varepsilon_1^*, \varepsilon_2^*, \varepsilon_1)}{\partial(\mathbf{x}', \varepsilon_1)} \right|
$$

$$
= \left| \frac{\partial(\frac{x_1' + x_2'}{2}, \frac{x_3' + x_4'}{2}, x_5' + z_3^c a_3 \varepsilon_1, \frac{x_1' - x_2'}{2a_1}, \frac{x_3' - x_4'}{2a_2}, \varepsilon_1)}{\partial(x_1', x_2', x_3', x_4', x_5', \varepsilon_1)} \right| \quad (4.2)
$$

$$
= \frac{1}{4a_1 a_2}.
$$

We accept this death move with probability

$$
a_d(\mathbf{x}'', \varepsilon_1, \varepsilon_1^*, \varepsilon_2^*)
$$

$$
= \min \left\{ 1, 5 \times 4 \times \frac{w_{b,3}}{w_{d,5}} \right.
$$

$$
\left. \times \frac{P(\mathbf{z}^c)}{P(\mathbf{z})} \frac{\pi(\mathbf{x}'')}{\pi(\mathbf{x}')} \left| \frac{\partial(T_{d,z_3}(\mathbf{x}', \varepsilon_1); \varepsilon_1^*, \varepsilon_2^*, \varepsilon_1)}{\partial(\mathbf{x}', \varepsilon_1)} \right| \right\} \quad (4.3)
$$

$$
= \min \left\{ 1, 20 \times \frac{w_{b,3}}{w_{d,5}} \times \frac{p_3^{I_{\{1\}}(z_3)} q_3^{I_{\{-1\}}(z_3)}}{p_3^{I_{\{1\}}(z_3^c)} q_3^{I_{\{-1\}}(z_3^c)}} \times \frac{\pi(\mathbf{x}'')}{\pi(\mathbf{x}')} \times \frac{1}{4a_1 a_2} \right\}.
$$

We illustrate the idea of this algorithm in Figure 3 diagrammatically for the ease of understanding.

Thus, in general, for moving from dimension k to dimension $k + m$, we need to simulate $\varepsilon_1, \ldots \varepsilon_m$ for updating $\mathbf{x} = (x_1, \ldots, x_k)$ to $\mathbf{x}' = (x_1', x_2', \ldots, x_k', x_{k+1}', \ldots, x_{k+m}')$. The associated general TTMCMC algorithm for jumping m dimensions is provided as Algorithm S-3.1 of Section S-3, and the proof of its detailed balance is provided in Section S-4.

In variable dimensional problems such as mixtures, changing the dimension of one set of parameters necessitates changing the dimensions of the other sets of parameters. Thus, more than one dimension must be changed at a time, while the parameters are inter-related. We provide the details and the relevant algorithm (Algorithm S-5.1) in Section S-5 of the supplement. Indeed, for our mixture applications of TTMCMC, we implement Algorithm S-5.1, choosing the additive transformation.

Note that exactly as discussed in Section 3.3, we can incorporate a hierarchical dependence structure on the distribution of z in Algorithms S-3.1 and S-5.1, which does not hamper the detailed balance condition.

5 TTMCMC: Towards automation

Algorithms 3.1, S-3.1 and S-5.1 provide concrete ways to implement our TTMCMC procedure, in general variable dimensional problems. Below, we de-
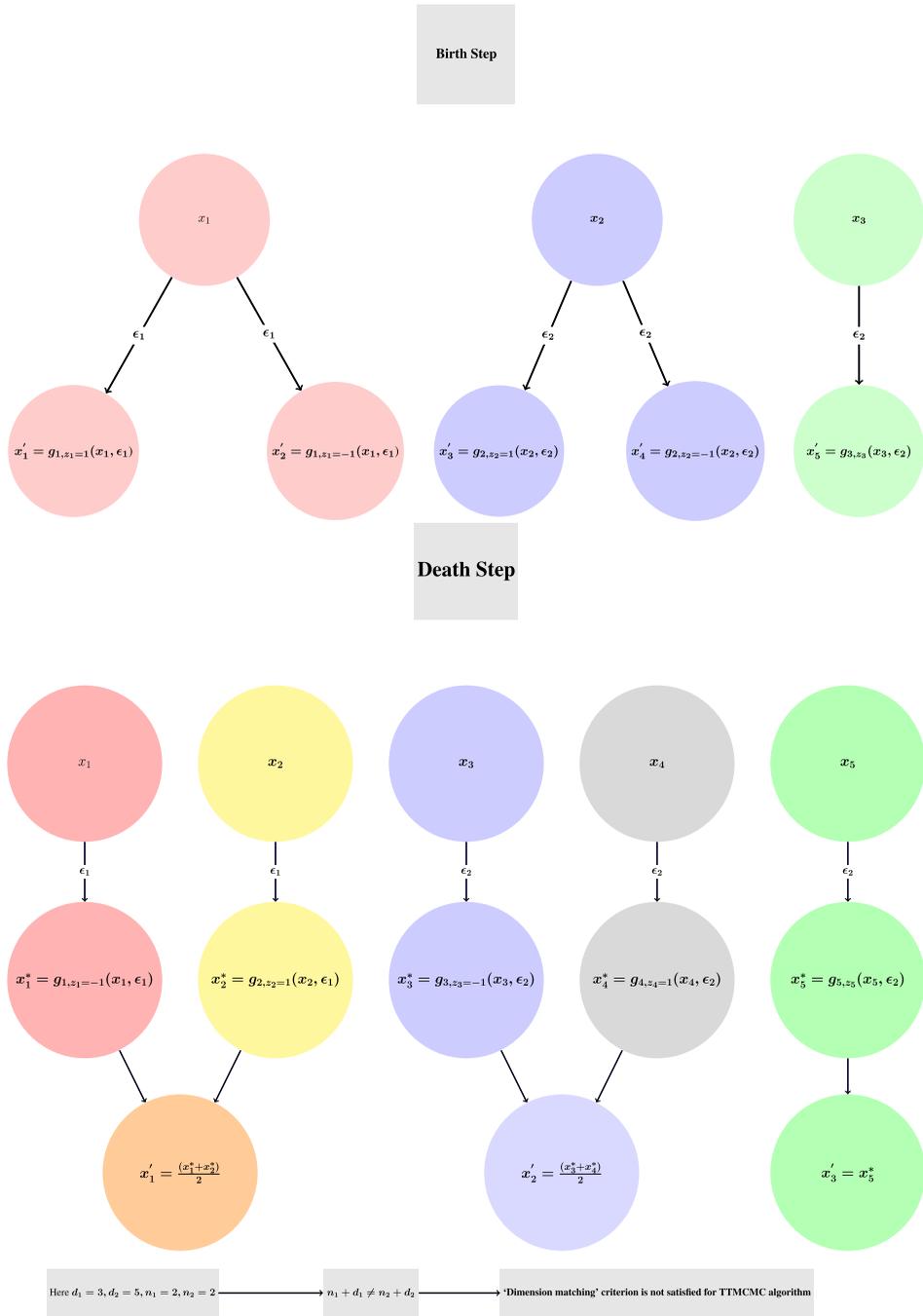
Figure 3 *Illustration of TTMCMC algorithm for jumping more than one dimension.*

tail the manyfold advantages of TTMCMC, which point towards the fact that TTMCMC is close to qualifying as an automatic sampler in variable dimensional problems.

5.1 Reasonably high acceptance rate

The additive and the multiplicative transformations, and combinations of them can be effectively utilized, in conjunction with just a few, fixed number of ε's, to accomplish transdimensional movement. The methodology reduces the variable dimensional problem to effectively fixed dimensional, indexed by a fixed and small number of ε's. The fixed and low-dimensional nature of ε (or the set $\{\varepsilon_1, \ldots, \varepsilon_m\}$) ensures reasonably high acceptance rate. Indeed, for high-dimensional proposals, with high probability at least one component would be ill-proposed, which would render the acceptance probability extremely small, even in fixed-dimensional cases. In the context of TMCMC, theoretical and empirical results are provided in Dutta and Bhattacharya (2014), Dey and Bhattacharya (2017a, 2017b). Our experiments in this paper provide ample support to our claim of adequate acceptance rate of TTMCMC.

5.2 Good mixing properties in high-dimensional and multimodal cases

Dutta and Bhattacharya (2014) discussed that in one-dimensional situations, TMCMC reduces to a Metropolis–Hastings algorithm with a specialized mixture proposal density, and hence, is expected to explore multimodal target densities quite efficiently (see Guan and Krone (2007), for example). In higher dimensions, due to singularity, the proposal does not admit a Lebesgue-measure-dominated mixture density form directly, but since the method employs similar principles, good convergence properties of TMCMC are to be expected for high-dimensional multimodal targets as well. Since TTMCMC samplers are also based on the same principles of deterministic transformations and construction of move types within each of the birth, death and no-change move types, good convergence properties are expected when the target density is multimodal for each dimension. In the context of TMCMC, Dey and Bhattacharya (2017a, 2017b, 2016) demonstrate far superior mixing of TMCMC compared to random walk Metropolis–Hastings. The results of our TTMCMC applications reported in this paper provide ample support to this discussion.

5.3 Applicability to all variable dimensional problems

The construction of TTMCMC sampler does not require any assumptions regarding the model, such as existence of moments or unimodality. Note that in the attempts made so far for constructing generic RJMCMC samplers, these assumptions are quite crucial; see Sisson (2005), Fan and Sisson (2011) for comprehensive discussions regarding these assumptions. So, for the construction of TTMCMC

sampler for switching between two models, namely, from $\mathcal{M}_k$ to $\mathcal{M}_{k'}$, we only need to determine if some sets of parameters are related and decide on the number of parameters to be added or deleted, in a single step. Accordingly we will choose one of the above mentioned algorithms and update all the parameters in a single block. Hence, our proposed sampler is very much applicable to any variable dimensional problem.

5.4 Default TTMCMC algorithm and its tuning

In order to design efficient MCMC algorithms, it has become standard practice to tune the proposals. For the default, random walk proposals, this is synonymous with choosing the scales optimally. Dutta and Bhattacharya (2014) recommended additive TMCMC as the default TMCMC proposal since this transformation requires much smaller number of move-types and the corresponding acceptance probability has a simple form in that it is free of the Jacobian of transformations. Already Dey and Bhattacharya (2017a, 2017b) have developed some theory on optimal scaling in the context of additive TMCMC. In keeping with Dutta and Bhattacharya (2014), we advocate additive TTMCMC as the default TTMCMC sampler, which again requires specification of the scaling constants. In this regard, in Section S-8.2 of the supplement, we propose a convergence diagnostic that is generally applicable. Guided by our proposed convergence diagnostic it is possible to find the appropriate value of scaling constants. Instances of the idea are illustrated in Sections 6 and 7. The results of our experiments demonstrate great ease of implementation and excellent performance of the default additive TTMCMC sampler in all the examples. Further experiments with additive TTMCMC, conducted by these authors and their colleagues in challenging, high-dimensional spatio-temporal problems (see, for example, Das and Bhattacharya (2016)), variable-selection problems, (high-dimensional) curve-fitting problems also yielded excellent results. Thus, it seems that additive TTMCMC is close towards the kind of automation that we desire.

6 Simulation studies with mixtures of gamma distributions with unknown number of components

Wiper, Insua and Ruggeri (2001) implement RJMCMC in mixtures of gamma distributions of the form $\mathcal{G}(v, \frac{v}{\mu})$, where by $\mathcal{G}(a, b)$ we mean a gamma distribution with mean a/b and variance a/b^2. In other words, Wiper, Insua and Ruggeri (2001) consider the following mixture density for $y > 0$:

$$f(y|\boldsymbol{v}_k, \boldsymbol{\mu}_k, \boldsymbol{\pi}_k, k) = \sum_{j=1}^{k} \pi_j \frac{(v_j/\mu_j)^{v_j}}{\Gamma(v_j)} y^{v_j-1} \exp\left(-\frac{v_j}{\mu_j} y\right), \qquad (6.1)$$

where $\boldsymbol{v}_k = (v_1, \ldots, v_k)$, $\boldsymbol{\mu}_k = (\mu_1, \ldots, \mu_k)$, and $\boldsymbol{\pi}_k = (\pi_1, \ldots, \pi_k)$. Given $k > 0$, for each j, $v_j > 0$, $\mu_j > 0$, $0 < \pi_j < 1$ such that $\sum_{j=1}^{k} \pi_j = 1$. We assume k to be unknown, so that the dimension of the model (that is, the number of the component parameters) is unknown and considered random.

6.1 Prior structure

Wiper, Insua and Ruggeri (2001) assumed the following prior structure given k:

$$\boldsymbol{\pi}_k \sim \mathcal{D}(1, \ldots, 1); \tag{6.2}$$

$$v_j \overset{\text{i.i.d.}}{\sim} \mathcal{E}(100); \qquad j = 1, \ldots, k; \tag{6.3}$$

$$\mu_j^{-1} \overset{\text{i.i.d.}}{\sim} \mathcal{G}(1, 1); \qquad j = 1, \ldots, k, \tag{6.4}$$

such that $\mu_1 < \cdots < \mu_k$. In (6.2), $\mathcal{D}(1, \ldots, 1)$ denotes the Dirichlet distribution with all the parameters equal to 1, and in (6.3), $\mathcal{E}(100)$ stands for the exponential distribution with mean 100. As regards k, Wiper, Insua and Ruggeri (2001) consider the discrete uniform distribution on $\{1, \ldots, 10\}$.

For the implementation purpose, we reparameterize v_j and μ_j as $\exp(v_j^*)$ and $\exp(\mu_j^*)$, where $v_j^* \sim \log(\text{Exponential}(100))$ and $(\mu_j^*)^{-1} \sim \log(\mathcal{G}(1, 1))$. Since $-\infty < v_j^* < \infty$ and $-\infty < \mu_j^* < \infty$, this reparameterization frees the parameter space from any restrictions, allowing TTMCMC to move freely, while keeping the original prior distributions intact. We denote $(v_1^*, \ldots, v_k^*)$ by $\boldsymbol{v}_k^*$ and $(\mu_1^*, \ldots, \mu_k^*)$ by $\boldsymbol{\mu}_k^*$.

For $\boldsymbol{\pi}$, we propose the following prior based on reparameterization: for $j = 1, \ldots, k$,

$$\pi_j = \frac{\exp(\omega_j)}{\sum_{\ell=1}^{k} \exp(\omega_j)}; \qquad \omega_1, \ldots, \omega_k \overset{\text{i.i.d.}}{\sim} N\big(\mu_\omega, \sigma_\omega^2\big), \tag{6.5}$$

where $\omega_j \overset{\text{i.i.d.}}{\sim} \log(\mathcal{G}(1, 1))$, so that the prior (6.2) remains intact. Thus, we need to update $\boldsymbol{\omega}_k = (\omega_1, \ldots, \omega_k)$, instead of $\boldsymbol{\pi}$, using TTMCMC.

6.2 Label switching

A brief account of the so-called "label-switching problem" associated with identifiability of mixtures is provided in Section S-6 of the supplement. In this article, our goal is to demonstrate TTMCMC with inference regarding posterior distributions of densities. Since inference on densities is not affected by label switching, the problem of label switching is not of much importance in our context. Moreover, we argue in Section S-6 that identifiability in the mixture context is not generally desirable. However, since Wiper, Insua and Ruggeri (2001) enforced the restriction $\mu_1 < \cdots < \mu_k$ in an attempt to mitigate identifiability problems, for fair comparison we also impose the same restriction.

6.3 Posterior summary

An important aspect to any Bayesian analysis is summarization of the posterior in the sense of obtaining a measure of central tendency and appropriate credible regions. Here we are interested in the posterior distribution of the entire mixture density, induced by the posterior of the unknown number of parameters. Thus, we need a measure of central tendency for the set of mixture densities supported by the posterior, and appropriately constructed credible regions. Indeed, in Section S-7 of the supplement, we develop a methodology for obtaining the modal mixture density associated with the posterior, along with the desired credible regions and highest posterior density (HPD) credible regions. In the context of our experiments we shall display the modal mixture densities and several other mixture densities falling within the 95% HPD regions.

6.4 Convergence diagnostics

Convergence assessment even in fixed-dimensional set-ups is a difficult proposition; in variable-dimensional problems, the challenges increase manyfold. We provide a briefing on these in Section S-8.1 of the supplement. As an attempt to make some progress on convergence assessment in variable-dimensional problems we propose a convergence diagnostic in Section S-8.2 of the supplement, which is based on the methodology for summarizing the posterior. In a nutshell, we obtain 95% (or any other desired) credible regions from the first and second halves of a complete run of TTMCMC, and then obtain the minimum increments of the radii required for the credible regions to contain one another; small values of the increments indicate convergence of TTMCMC. Not only do we assess convergence of TTMCMC with this method, we exploit this idea to select the scales of the additive transformation that we employ for the illustrations.

6.5 General TTMCMC strategy for our experiments

We conduct four simulation studies, with data generated from the same 1-component, 2-component, 3-component and 4-component gamma mixtures as considered by Wiper, Insua and Ruggeri (2001) and apply TTMCMC and compare our results with those obtained by the RJMCMC algorithm of Wiper, Insua and Ruggeri (2001). In particular, we apply Algorithm S-3.1, updating $(k, \boldsymbol{v}^*, \boldsymbol{\mu}^*, \boldsymbol{\omega})$ simultaneously in a single block using the additive transformation; we choose the proposal density to be $\varrho(\varepsilon) \equiv N(\varepsilon : 0, 1)\mathbb{I}_{(0,\infty)}(\varepsilon)$, where $N(\varepsilon : 0, 1)$ denotes the normal density with mean 0, variance 1, and evaluated at ε; $\mathbb{I}_{(0,\infty)}(\cdot)$ denotes the indicator function for the set $(0, \infty)$. For every iteration of TTMCMC we choose equal move-type probabilities of birth, death and no-change strategies. Also, for the underlying additive transformation, we choose equal probabilities of forward and backward transformations. The forms of the Jacobian for the birth and the death moves are given by $8a_{v_j^*}a_{\mu_j^*}a_{\omega_j}$ and $(8a_{v_j^*}a_{\mu_j^*}a_{\omega_j})^{-1}$ respectively, where

$a_{\nu_j^*}$, $a_{\mu_j^*}$ and a_{ω_j} are the scales for additive TTMCMC updating of ν_j^*, μ_j^* and ω_j respectively. We base the choices of these scales on the convergence diagnostic proposed in Section S-8.2 of the supplement. The experimental details are provided in the context-specific applications. All our codes are written in C and implemented on a 32 bit, dual core (2.53 GHz $\times$ 2) laptop with 2.8 GiB memory. However, for high-dimensional multivariate experiments we implemented our C codes on a VMWare.

6.6 An RJMCMC algorithm based on random walk proposals

Since, in this paper, we apply additive TTMCMC to our examples, it makes sense to compare our TTMCMC results with those obtained by the RJMCMC algorithm based on random walk, which is the closest to additive TTMCMC among all RJMCMC algorithms. Recall that random walk involves additive transformations of the same form as additive TTMCMC, but with independent jump sizes for every variable, unlike TTMCMC. Also, unlike TTMCMC, the acceptance ratios for the birth and death moves involves products of the densities $\varrho(u_i) \equiv N(u_i : 0, 1)\mathbb{I}_{(0,\infty)}(u_i)$; $i = 1, 2, 3$, corresponding to the birth proposals for $(\nu_1^*, \tau_1^*, \omega_1)$. Since the proposals of additive TTMCMC and random walk have the same additive form, the variabilities of the jump sizes of the competing proposals are not expected to be different. This is confirmed by the optimal scaling theory of TMCMC developed by Dey and Bhattacharya (2017b), where it is shown that the optimal scales of additive TMCMC and random walk are the same. Hence, in this work, we choose the same scales of random walk RJMCMC as additive TTMCMC.

The main difference between our random walk RJMCMC and the proposal of Green (2003) is that the latter is deterministic unless movement to a higher dimension is attempted; the moves also involve dimension-specific mean vectors and covariance matrices, which are to be estimated from the dimension-specific posteriors. Even for moderate number of models this is a difficult and computationally burdensome proposition; see, for example, Fan and Sisson (2011). Indeed, as stressed in Green (2003), the approach is unlikely to be useful for more than a small set of models.

However, for all our examples related to the gamma mixture, our random walk RJMCMC had very small overall acceptance rate, and completely failed to change the dimension in any such example. Hence, we do not provide further details regarding the performance of the random walk RJMCMC in gamma mixtures. In the normal mixture context, random walk RJMCMC performed somewhat better, although still not at all satisfactorily. Since this algorithm fails even in univariate contexts, we do not pursue this for the multivariate situations.

6.7 First simulation study with data generated from a one-component gamma mixture

Following Wiper, Insua and Ruggeri (2001) we generate 400 realizations from $\mathcal{G}amma(3, 3)$, and model the realized data with the gamma mixture of the form

(6.1). Assuming the same prior structure described in Section 6.1, we then simulate from the resulting variable-dimensional posterior using TTMCMC.

For implementing TTMCMC it is necessary to select the scales $a_{v_j^*}$, $a_{\tau_j^*}$, a_{ω_j} appropriately for each $j = 1, \ldots, k$. Rather than selecting the scales in order to optimize the acceptance rate (see Dey and Bhattacharya (2017b) for optimal scaling theory in the context of additive TMCMC), here we choose the scales by directly quantifying convergence of the TTMCMC chain using the convergence diagnostic procedure proposed in Section S-8.2 of the supplement. We experimented by setting, for every $j = 1, \ldots, k$, the scale values $a_{v_j^*} = a_{v^*}$; $a_{\mu_j^*} = a_{\mu^*}$, and $a_{\omega_j} = a_\omega$, with a_{v^*}, a_{τ^*}, a_ω being one of the trial values $0.05, 0.1, 0.12, 0.15, 0.20, 0.25, 0.50$. With every trial value, we ran our TTMCMC algorithm for a burn-in of 7.5×10^5 iterations, and a further 15×10^5 iterations, storing one in 150 iterations, thus obtaining a total of 10,000 realizations from the posterior distribution. For each trial run, we assessed convergence of our TTMCMC chain using the method proposed in Section S-8.2. . We divided our TTMCMC samples into two parts, one part consisting of the first 5000 realizations and the other part containing the next 5000 realizations. Constructing the approximate 95% credible regions as prescribed, we then obtained the minimum increment, η_1, of the radius of the first credible region such that the increased first credible region wholly contains the second credible region. Similarly, we obtained η_2, the radius increment associated with the second credible region. Small values of η_1 and η_2 indicate convergence of the algorithm. We selected that set of trial values of the scales which yielded the smallest η_1 and η_2 among the trial runs. Indeed, the smallest η_1 and η_2 turned out to be $\eta_1 = 0.041460$ and $\eta_2 = 0.027130$, which corresponded to $a_v = a_\mu = 0.5$ and $a_\omega = 1.5$. Hence, we report our results with respect to these trial values. Moreover, since both these quantities are small, we conclude that convergence has taken place appropriately. We remark here that the rather long burn-in that we had considered was unnecessary, as further experiments showed that the chain converged in far less number of iterations. But we feel it is a good practice to allow large enough burn-in when it is feasible computationally. The overall acceptance rate, evaluated empirically, turned out to be 0.036596. The birth, death, and no-change rates are 0.004206, 0.053131 and 0.067679, respectively. Our TTMCMC implementation with the scales selected as above took 10 minutes and 57 seconds.

The trace plots of k, v_1^*, μ_1^* and ω_1, provided in Figure 4, exhibits quite adequate mixing properties consistent with our more formal test of convergence. Also very encouragingly, the posterior distribution of k gives probabilities 0.9344, 0.0649 and 0.0007 to $k = 1, 2, 3$ respectively, heavily supporting the true, single-component gamma mixture. Since the data size is rather large, such high support to the truth is expected. Indeed, with further simulation studies we demonstrate in Section S-9.1 of the supplement, that as the data size increases, the posterior distribution of k concentrates around the truth, namely, $k = 1$.
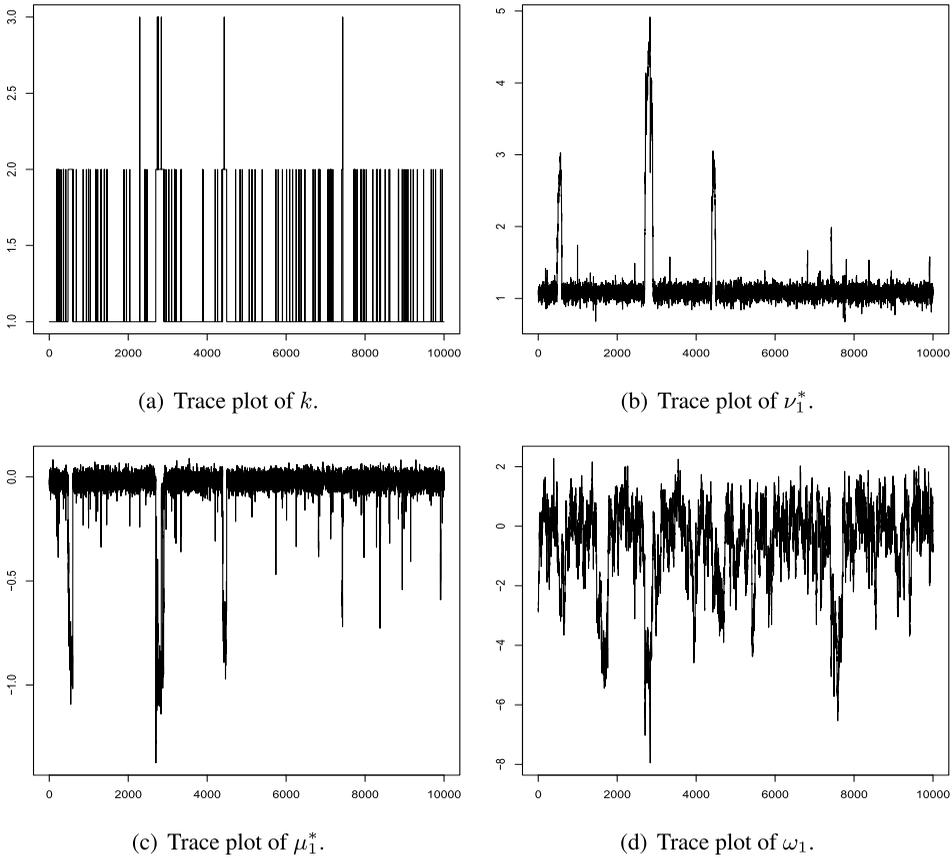
(a) Trace plot of k.

(b) Trace plot of ν_1^*.

(c) Trace plot of μ_1^*.

(d) Trace plot of ω_1.

Figure 4 *TTMCMC for 1-component gamma mixture: Trace plots of k, ν_1^*, μ_1^* and ω_1.*

Figure 5 shows the modal density (thick, black curve), along with some other densities within the 95% HPD region overlapped on the histogram of the simulated data. Excellent fit of the posterior distribution of the densities to the data is indicated by the diagram.

6.7.1 *Comparison with the results obtained by Whiper, Insua and Ruggeri* (2001). In sharp contrast with our TTMCMC results, Wiper, Insua and Ruggeri (2001), using an RJMCMC algorithm that is very similar to that proposed by Richardson and Green (1997) for normal mixtures, obtained a posterior distribution that supports all possible values of $k \in \{1, \ldots, 10\}$. In particular, their posterior probabilities of $k = 1, 2, 3, 4, 5$ turned out to be 0.41, 0.24, 0.12, 0.08 and 0.05, respectively, with other values of k having posterior probabilities less than 0.03. In other words, driven by RJMCMC, the true value $k = 1$ received lower posterior support, in
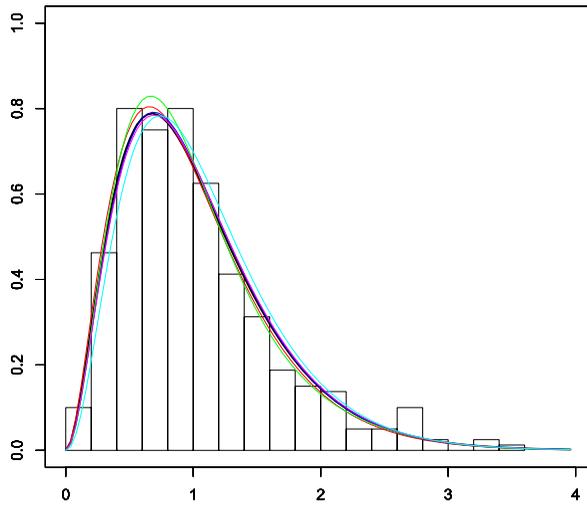
Figure 5 *TTMCMC for* 1-*component gamma mixture*: *Goodness of fit of the posterior distribution of densities* (*coloured curves*) *to the simulated data* (*histogram*). *The thick black curve is the modal density and the other coloured curves are some densities contained in the* 95% *HPD.*

comparison with our TTMCMC based posterior. This performance can possibly be attributed to the $\mathcal{G}(5, 5)$ proposal density they used for their dimension-changing move. Since this density is uniformly less than one and features in the acceptance ratio, heavy bias towards large values of k is to be expected as per our discussion in the third point following Algorithm 3.1. Thus, there seems to be good reasons to suspect the convergence of the RJMCMC algorithm in this case. In fact, as we shall show, the same issue hinders convergence of the RJMCMC algorithms for the remaining experiments as well.

It is important to remark in this context that the actual mixture density can be approximated well in spite of poor mixing, provided that k takes on large values with significant posterior probabilities. Therefore fitting the actual density alone can be very misleading as a criterion of assessment of variable-dimensional algorithms, particularly for RJMCMC algorithms, because of their inherent bias towards large values of k in any practical implementation. In all the four simulation examples considered by Wiper, Insua and Ruggeri (2001), the actual densities are well-approximated by RJMCMC, but in all the cases, large values of k seemed to play vital important roles in this regard. Such an issue is clearly of more concern in real data cases where the truth is unknown. As we demonstrate with TTMCMC in the supplement with the real galaxy data example of Richardson and Green (1997), their prior structure perhaps actually supports unimodal density, while the histogram is highly multimodal. However, because of large values of k supported by RJMCMC, the approximated density seems to appear as a good fit.

6.8 Second simulation study with data generated from a two-component gamma mixture

Following Wiper, Insua and Ruggeri (2001), we now generate 400 realizations from the two-component mixture $0.1 \times \mathcal{G}(9, 27) + 0.9 \times \mathcal{G}(90, 270)$.

In this case, for TTMCMC implementation we obtained $a_{\nu*} = 0.05$; $a_{\mu*} = 0.005$, and $a_\omega = 0.05$ using our convergence diagnostic procedure. We set a considerably large burn-in time of 3×10^6 iterations as convergence seemed to be somewhat slow compared to the one-component example. We stored one in 150 iterations of a further run of 15×10^5 iterations, so that, as before we stored a total of 10,000 realizations from the posterior distribution. This took 31 minutes 6 seconds and yielded an overall acceptance rate 0.229365. Also, the birth, death and no-change rates are 0.000017, 0.000022 and 0.688025, respectively. In this case, we obtained $\eta_1 = 0.28333$ and $\eta_2 = 0.30828$, which are reasonably small, providing reasonably strong evidence in support of convergence of our TTMCMC chain. This is further supported strongly by the visual information carried by the trace plots of k, ν_1^*, μ_1^* and ω_1, shown in Figure 6.

Interestingly, after burn-in, TTMCMC gives full mass to 2 components, thus completely supporting the truth. However, as demonstrated in Section S-9 of the supplement with simulation studies for different data sizes (see Section S-9.2 for simulations with this 2-component mixture), it is possible that the actual posterior distribution of k gives "almost" point mass to $k = 2$, such that with probability close to zero some other components may also occur, but might have been missed by us in this case due to the finite run length of our algorithm.

As before, Figure 7 shows excellent fit of the posterior distribution of the densities to the simulated data.

6.8.1 *Comparison with the results obtained by Whiper, Insua and Ruggeri* (2001). As to be anticipated, bias towards large values of k continued in this example. Indeed, although Wiper, Insua and Ruggeri (2001) obtained $k = 2$ as the mode of their RJMCMC based posterior of k, they also found that their RJMCMC algorithm yielded the posterior probability about 0.01 for $k = 1$, and supported other larger values of k. Thus, compared to TTMCMC, which identifies the truth very precisely, RJMCMC manages to facilitate only weak inference because of its lack of convergence.

6.9 Third simulation study with data generated from a three-component gamma mixture

Here we generate 400 realizations from the three-component mixture $0.2 \times \mathcal{G}(40, 20) + 0.6 \times \mathcal{G}(6, 1) + 0.2 \times \mathcal{G}(200, 20)$, following Wiper, Insua and Ruggeri (2001).

Again, we obtained $a_{\nu*} = 0.05$; $a_{\mu*} = 0.005$, and $a_\omega = 0.05$ using our convergence diagnostic procedure. Here a burn-in of 15×10^5 iterations turned out to be
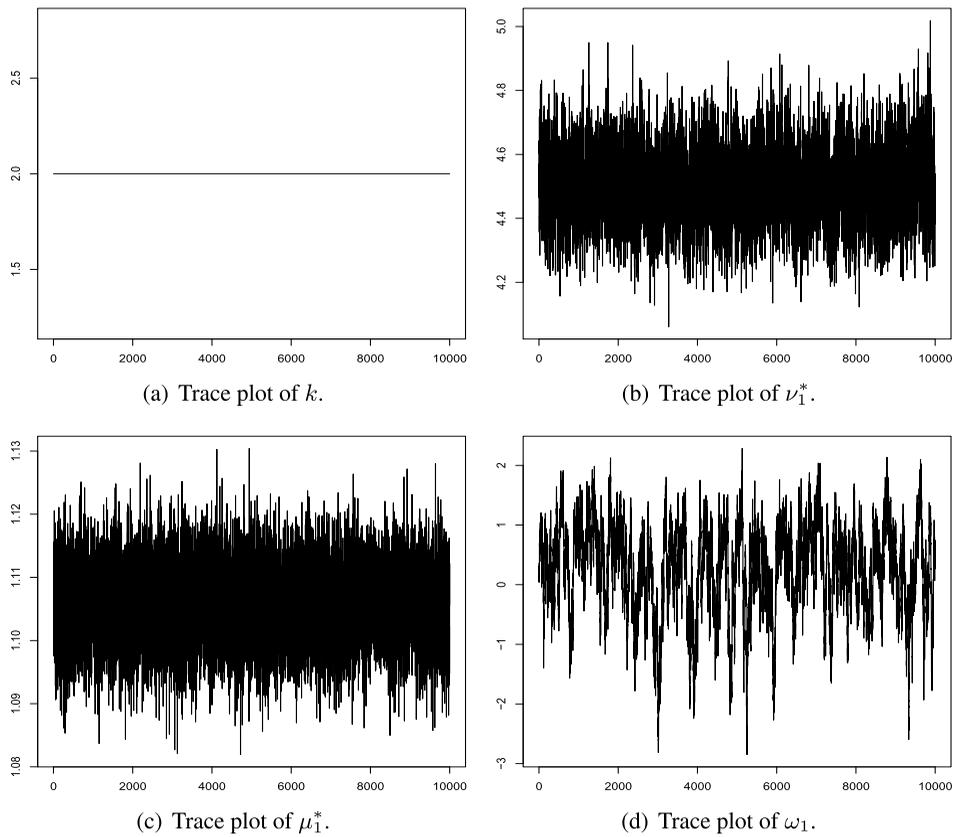
(a) Trace plot of k.

(b) Trace plot of ν_1^*.

(c) Trace plot of μ_1^*.

(d) Trace plot of ω_1.

Figure 6   *TTMCMC for 2-component gamma mixture: Trace plots of k, ν_1^*, μ_1^* and ω_1.*

more than sufficient. As before we stored 10,000 realizations from the posterior distribution out of a further 15×10^5 iterations after the burn-in with a thinning of size 150. The overall acceptance rate was 0.240443 and the time taken was 36 minutes and 5 seconds. The birth, death and no-change rates are 0.00001, 0.000017 and 0.720547, respectively. As regards the convergence diagnostic, $\eta_1 = 0.01602$ and $\eta_2 = 0.01757$, which are both small enough to let us conclude that the TTM-CMC chain has converged very well. The trace plots displayed in Figure 8 completely support our conclusion regarding convergence.

Again, the posterior distribution of k completely supports the truth, giving full mass to 3, which, in this example, is the correct number of components. The simulation study in Section S-9.3 of the supplement demonstrates that it is possible that here TTMCMC has missed $k = 4$, which might have occurred with extremely small probability.

As to be expected, Figure 9 confirms excellent fit of the posterior distribution of the densities to the simulated data.
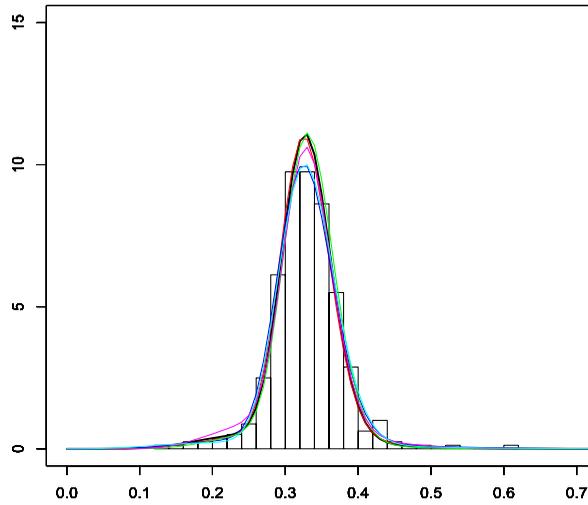
Figure 7 *TTMCMC for* 2-*component gamma mixture*: *Goodness of fit of the posterior distribution of densities* (*coloured curves*) *to the simulated data* (*histogram*). *The thick black curve is the modal density and the other coloured curves are some densities contained in the* 95% *HPD.*

6.9.1 *Comparison with the results obtained by Whiper, Insua and Ruggeri* (2001). Specific RJMCMC based results pertaining to the three component mixture are not provided in Wiper, Insua and Ruggeri (2001), but larger values of k compared to the truth, are certain to occur with significant probabilities.

6.10 Fourth simulation study with data generated from a four-component gamma mixture

For the final simulation study with gamma mixtures, following Wiper, Insua and Ruggeri (2001) we generate 400 realizations from the four-component mixture $0.25 \times \mathcal{G}(200, 100) + 0.25 \times \mathcal{G}(400, 100) + 0.25 \times \mathcal{G}(600, 100) + 0.25 \times \mathcal{G}(800, 100)$.

Here we obtained $a_{\nu^*} = 0.05$; $a_{\mu^*} = 0.005$, and $a_\omega = 0.12$, with a burn-in of 15×10^5 iterations and with respect to 10,000 realizations from the posterior distribution stored as before after burn-in with a thinning of size 150. The time to implement TTMCMC was 40 minutes and 35 seconds and we obtained an overall acceptance rate 0.117432. The birth, death and no-change rates are 0.000417, 0.000401 and 0.345363, respectively. That the chain converged reasonably well can be inferred since $\eta_1 = 0.08154$ and $\eta_2 = 0.11839$ are both reasonably small. As before, the trace plots displayed in Figure 10 confirm our conclusion regarding convergence.

Here the posterior distribution of k gives almost full mass to the truth $k = 4$, and seems to be consistent with the further simulation study conducted in Section S-9.4 of the supplement, considering a data of size 1000.
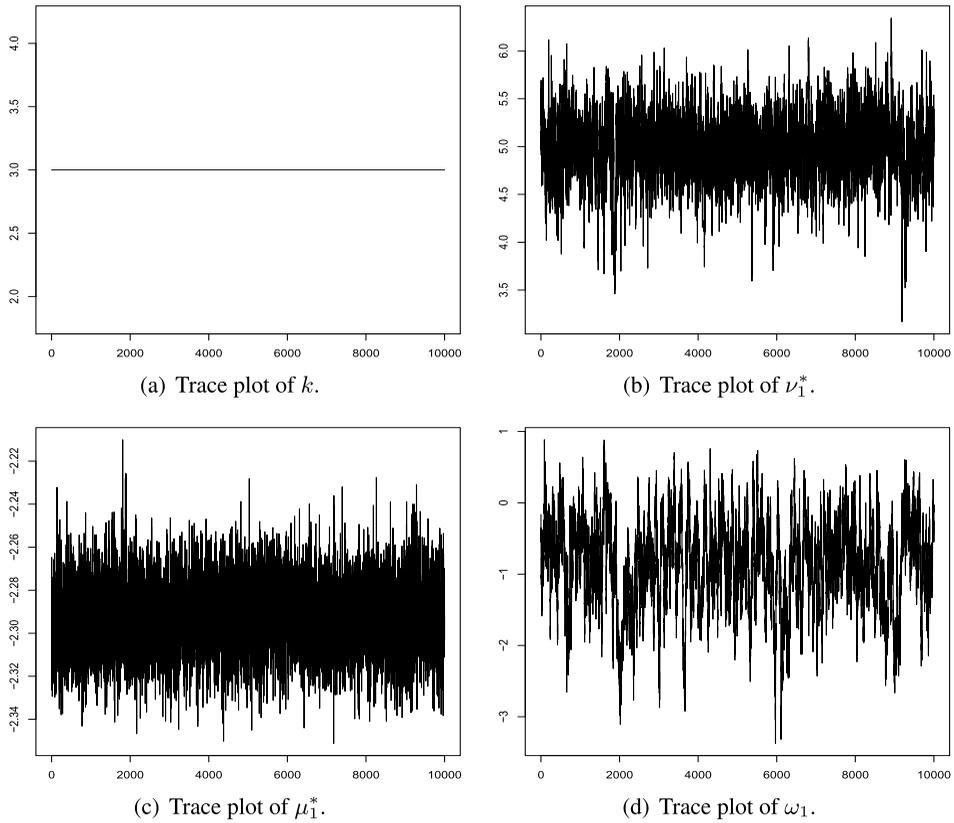
(a) Trace plot of k.

(b) Trace plot of ν_1^*.

(c) Trace plot of μ_1^*.

(d) Trace plot of ω_1.

Figure 8   *TTMCMC for 3-component gamma mixture: Trace plots of k, ν_1^*, μ_1^* and ω_1.*

As before, Figure 11 shows that excellent fit of the posterior distribution of the densities to the simulated data has been achieved.

6.10.1 *Comparison with the results obtained by Whiper, Insua and Ruggeri* (2001).   Even for this 4-component example specific RJMCMC based results are not provided in Wiper, Insua and Ruggeri (2001), but as in the other RJMCMC based examples, larger values of k compared to the truth, are certain to occur with significant probabilities.

7 Comparison of TTMCMC and RJMCMC in the normal mixture set up with unknown number of components

We now illustrate TTMCMC on normal mixture models with unknown number of components with application to the well-studied enzyme, acidity and the galaxy data sets. Richardson and Green (1997) modeled these data sets using parametric
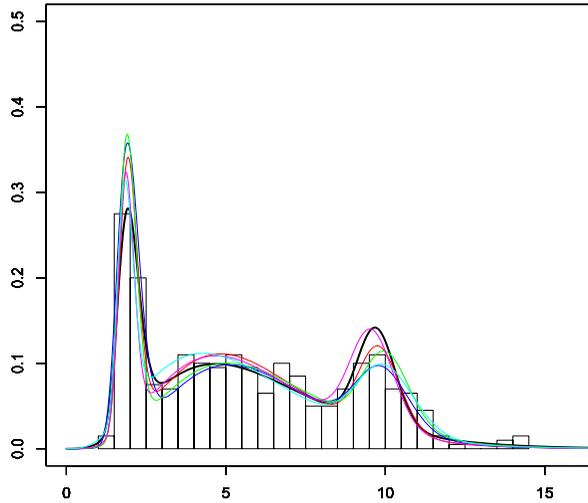
Figure 9 *TTMCMC for* 3-*component gamma mixture*: *Goodness of fit of the posterior distribution of densities* (*coloured curves*) *to the simulated data* (*histogram*). *The thick black curve is the modal density and the other coloured curves are some densities contained in the* 95% *HPD.*

normal mixtures and applied RJMCMC for Bayesian inference. On the other hand, Bhattacharya (2008) (see also Escobar and West (1995)) proposed a semi parametric normal mixture model based on Dirichlet process and used Gibbs sampler for Bayesian inference.

7.1 Normal mixture

Let the data points $y_1, \ldots, y_n$ be independently and identically distributed (i.i.d.) as the normal mixture of the following form: for $i = 1, \ldots, n$

$$f(y_i | \boldsymbol{\nu}_k, \boldsymbol{\tau}_k, \boldsymbol{\pi}_k, k) = \sum_{j=1}^{k} \pi_j \sqrt{\frac{\tau_j}{2\pi}} \exp\left\{-\frac{\tau_j}{2}(y_i - \nu_j)^2\right\}, \qquad (7.1)$$

where $\boldsymbol{\nu}_k = (\nu_1, \ldots, \nu_k)$, $\boldsymbol{\tau}_k = (\tau_1, \ldots, \tau_k)$, and $\boldsymbol{\pi}_k = (\pi_1, \ldots, \pi_k)$. Given $k > 0$, for each j, $-\infty < \nu_j < \infty$, $\tau_j > 0$, $0 < \pi_j < 1$ such that $\sum_{j=1}^{k} \pi_j = 1$. As before, we assume that k is unknown.

7.2 Prior structure

Note that the semi parametric mixture model of Bhattacharya (2008) can be viewed as a parametric model when the scale parameter associated with the base distribution of the Dirichlet process prior tends to infinity. Hence, from that perspective, the base distributions of ν_j and τ_j may be regarded as the respective priors for our current parametric mixture context. Thus, motivated by Bhattacharya (2008), we
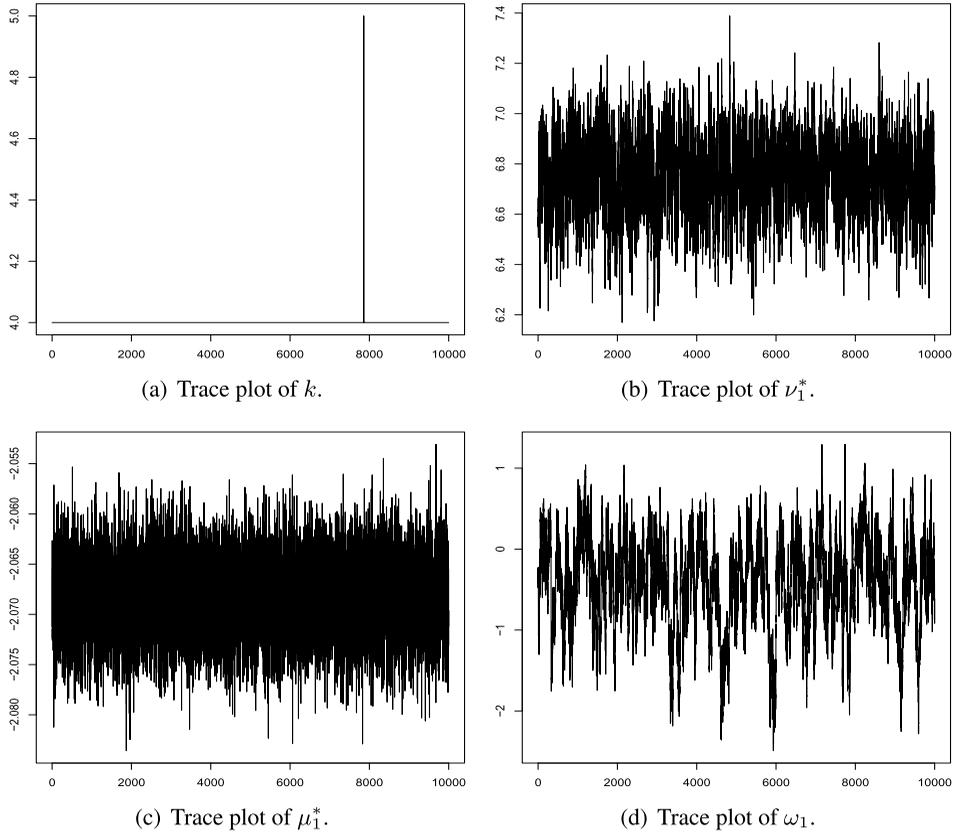
(a) Trace plot of k.

(b) Trace plot of ν_1^*.

(c) Trace plot of μ_1^*.

(d) Trace plot of ω_1.

Figure 10   *TTMCMC for 4-component gamma mixture*: *Trace plots of k, ν_1^*, μ_1^* and ω_1.*

consider the following prior for ν and τ:

$$[\tau_j] \sim \mathcal{G}\left(\frac{s}{2}, \frac{S}{2}\right); \tag{7.2}$$

$$[\nu_j | \tau_j] \sim N\left(\nu_0, \frac{\psi}{\tau_j}\right). \tag{7.3}$$

In the above, $N(\mu, \sigma^2)$ denotes the normal distribution with mean μ and variance σ^2. Specifications of the values of the hyperparameters s, S, ν_0, ψ are discussed in the context of the applications.

Analogous to the gamma mixture context here we reparameterize τ_j as $\exp(\tau_j^*)$, where $\tau_j^* \sim \log(\mathcal{G}(s/2, S/2))$. We denote $(\tau_1^*, \ldots, \tau_k^*)$ by $\boldsymbol{\tau}_k^*$.

For $\boldsymbol{\pi}$ we propose the same reparameterization (6.5). In this case, we consider two kinds of priors on ω. One is $\omega_j \sim N(\mu_\omega, \sigma_\omega^2)$, and the other is $\omega_j \sim \log(\mathcal{G}(\alpha_j, 1))$ independently, for $j = 1, \ldots, k$, where $\alpha_j > 0$; $j = 1, \ldots, k$. Note
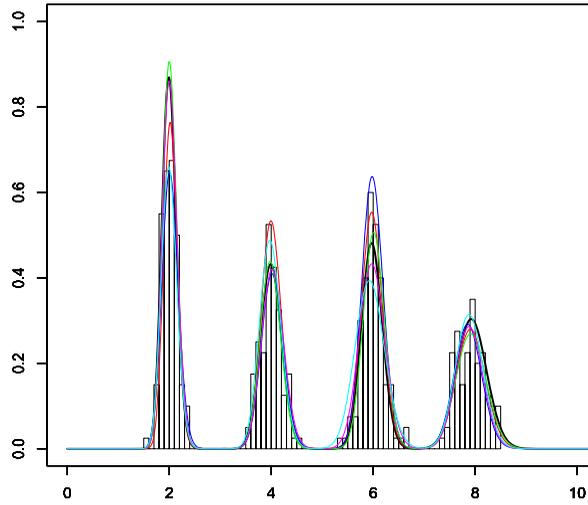
Figure 11 *TTMCMC for 4-component gamma mixture*: *Goodness of fit of the posterior distribution of densities* (*coloured curves*) *to the simulated data* (*histogram*). *The thick black curve is the modal density and the other coloured curves are some densities contained in the* 95% *HPD*.

that, for the normal prior on ω_j, the induced prior on $\boldsymbol{\pi}$ is not the traditional Dirichlet distribution, while the second prior implies that $\boldsymbol{\pi} \sim \mathcal{D}(\alpha_1, \ldots, \alpha_k)$.

As regards the prior on k, we consider the uniform distribution on $\{1, 2, \ldots, 30\}$, the truncated Poisson distribution on $\{1, 2, \ldots, 30\}$ and the discretized normal with mean μ_k and variance σ_k^2 on $\{1, 2, \ldots, 30\}$ (that is, the normal density with mean μ_k and variance σ_k^2 evaluated and re-normalized on $\{1, 2, \ldots, 30\}$ to render it a discrete probability mass function).

We fit normal mixture models to each of the three data sets—enzyme, acidity, and galaxy, using the general TTMCMC strategy provided in Section 6.5. The details are provided in the context-specific applications.

We compare the performance of additive TTMCMC with random walk RJM-CMC, which is analogous to additive TTMCMC but with independent jump-sizes for every co-ordinate and with the proposal density associated with the birth move incorporated within the acceptance ratio, unlike TTMCMC; see Section 6.6.

Our main aim is to demonstrate that the simplest version of TTMCMC, namely, TTMCMC with the additive transformation, is efficient enough for adequately exploring the complicated mixture-based posteriors in all the three applications, while the corresponding RJMCMC version, composed of random walk based moves, fails miserably.

Specific details of inference and implementation of our methodologies follow.

7.3 Enzyme data

Following Bhattacharya (2008) we set $s = 4.0$; $S = 2 \times (0.2/1.22) = 0.3278689$; $\nu_0 = 1.45$; $\psi = 33.3$. Rather than assuming $\omega_j \sim \log(\mathcal{G}(\alpha_j, 1))$ which induce the

traditional Dirichlet distribution for $\boldsymbol{\pi}$, here we assume that $\omega_j \sim N(\mu_{\omega_j}, \sigma^2_{\omega_j})$, with $\mu_\omega = 0$ and $\sigma^2_\omega = 0.25$. We chose somewhat small variance to reflect our belief that ω_j's are relatively close to constant, so that *a priori* the mixing probabilities $\boldsymbol{\pi}$ are approximately the same. We specify the uniform distribution on $\{1, \ldots, 30\}$ as the prior on k.

As in the gamma mixture set-up we experimented by setting, for every $j = 1, \ldots, k$, the scale values $a_{v^*_j} = a_{v^*}$; $a_{\tau^*_j} = a_{\tau^*}$, and $a_{\omega_j} = a_\omega$, with $a_{v^*}, a_{\tau^*}, a_\omega$ being one of the trial values 0.05, 0.1, 0.12, 0.15, 0.20, 0.25, 0.50. We considered a burn-in of 3.75×10^5 iterations and a further 15×10^5 iterations, storing as before one in 150 iterations to obtain 10,000 realizations from the posterior. Here η_1 and η_2 turned out to be $\eta_1 = 0.07291$ and $\eta_2 = 0.039230$, which corresponded to $a_v = a_\tau = a_\omega = 0.05$. The results we report are with respect to these trial values. Since both η_1 and η_2 are small, we conclude that convergence has taken place appropriately. The overall acceptance rate, evaluated empirically, turned out to be 0.05284032, and the birth, death, no-change rates are 0.000306, 0.000304 and 0.157810, respectively. Our TTMCMC implementation with the scales selected as above took 2 minutes and 56 seconds.

We also verified convergence of our TTMCMC chain with informal trace plots. Figure 12 displays the trace plots of k, v^*_1, τ^*_1 and ω_1. As seen in panel (a) of Figure 12 the posterior distribution of k placed highest mass on 2 components (posterior probability 0.986), followed by 3 components (posterior probability 0.0137), and then by 4 components (probability 0.0003). In other words, our Bayesian analysis strongly supports bimodality. Indeed, the information regarding bimodality is particularly strong thanks to the small range on which the data are supported and the large size of the data (the data set contains 245 observations on an effective support $(0, 3)$). Panels (b), (c) and (d) of Figure 12 show adequate mixing properties of the chain. Thus, the mixing information provided by these trace plots supports the conclusion obtained by our proposed credible region based convergence assessment method.

Figure 13 shows excellent fit of the posterior distribution of the densities to the data.

7.4 Acidity data

Again following Bhattacharya (2008) we set $s = 4.0$; $S = 2 \times (0.2/0.573) = 0.6980803$; $v_0 = 5.02$; $\psi = 33.3$. Here also we assume that $\omega_j \sim N(\mu_{\omega_j}, \sigma^2_{\omega_j})$, with $\mu_\omega = 0$ and $\sigma^2_\omega = 0.25$. As before, we put the uniform prior distribution on $\{1, \ldots, 30\}$ on k.

Following the convergence diagnostic method detailed above for choosing appropriate scales here, we obtain $a_{v^*_j} = a_{\tau^*_j} = a_{\omega^*_j} = 0.05$ for $j = 1, \ldots, k$. For these scales, we obtained $\eta_1 = 0.0049$ and $\eta_2 = 0.0080$, which are very small, indicating very good convergence.
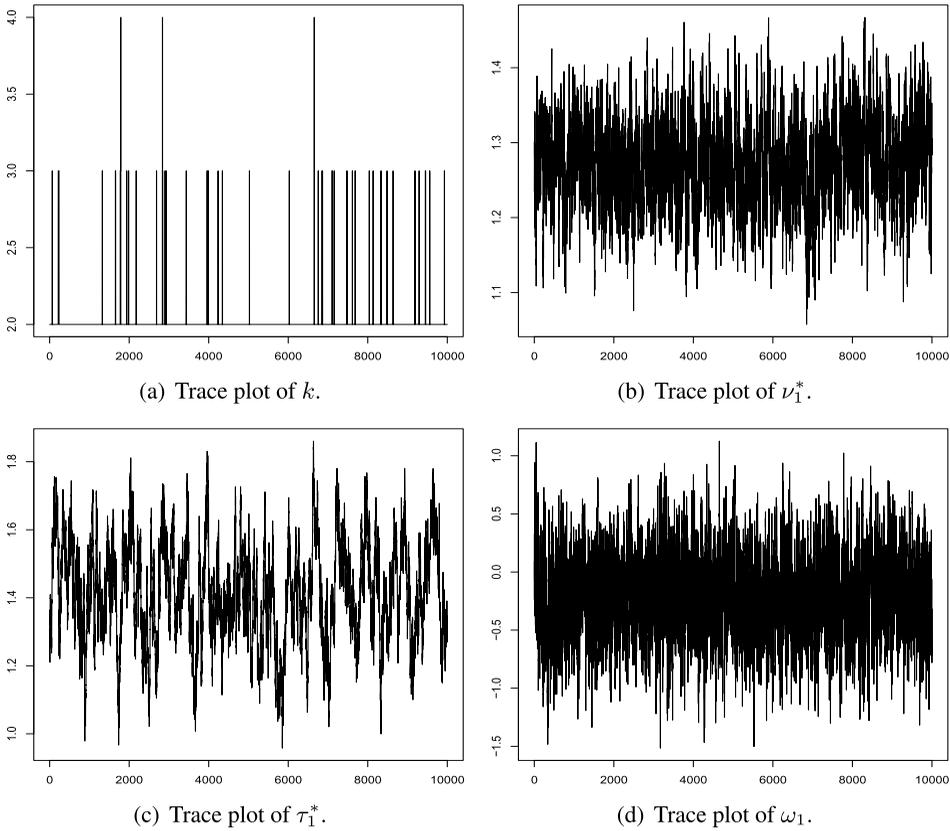
(a) Trace plot of k.

(b) Trace plot of ν_1^*.

(c) Trace plot of τ_1^*.

(d) Trace plot of ω_1.

Figure 12 *TTMCMC for the enzyme data: Trace plots of k, ν_1^*, τ_1^* and ω_1.*

With the chosen scales our implementation took 1 minute and 43 seconds to yield 10,000 realizations following a burn-in of 3×10^5 iterations, after storing one in 150 iterations out of further 15×10^5 iterations after the burn-in period. The overall acceptance rate turned out to be 0.198572, and the birth, death, no-change rates turned out to be 0.000795, 0.000842 and 0.593601, respectively.

The trace plots of k, ν_1^*, τ_1^* and ω_1, shown in Figure 14, again indicate quite good mixing properties and are consistent with the conclusions of our proposed credible region based convergence assessment criterion.

With our prior structure here the posterior distribution of k again strongly favoured 2 and 3 components, with $k = 2$ receiving significantly larger posterior mass 0.9941 compared to the posterior probability of $k = 3$. The reason for the strong support for bimodality can be attributed to the large size of the data contained in the relatively small interval $(2, 8)$.

The modal density and sample densities falling in the 95% HPD region, overlapped on the histogram of the observed data are shown in Figure 15. Once again, good fit to the data is indicated.
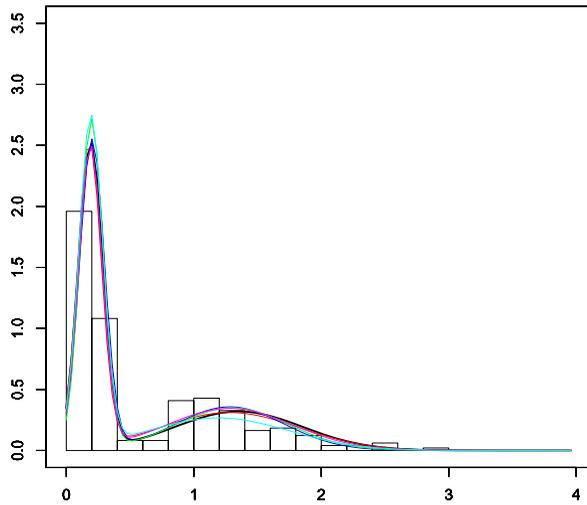
Figure 13 *TTMCMC for the enzyme data*: *Goodness of fit of the posterior distribution of densities* (*coloured curves*) *to the observed data* (*histogram*). *The thick black curve is the modal density and the other coloured curves are some densities contained in the* 95% *HPD*.

7.5 Galaxy data

In contrast with the previous two cases of the enzyme and the acidity data, the galaxy data, which is much more sparse and seems to exhibit far greater number of modes, seems to be much more challenging to analyze. Thus, we consider a somewhat different prior structure to reflect our beliefs regarding the Bayesian mixture analysis.

Here, following Bhattacharya (2008) we set $s = 4.0$; $S = 2$; $v_0 = 20$; $\psi = 33.3$. However, unlike the previous two cases here we assume that $\omega_j \sim \log(\mathcal{G}(5, 1))$, so that π follows the Dirichlet distribution with all the parameters equal to 5. The prior mean and mode of π_j associated with this Dirichlet distribution are $1/k$ and the variance is $(k - 1)/6k^2$. Note that the mean and the variance of the uniform Dirichlet distribution, which corresponds to taking all the parameters equal to 1, are $1/k$ and $(k - 1)/\{k(k + 1)\}$, respectively. Hence, for large k, the variance of our prior distribution is about $1/6$ times that of the uniform Dirichlet. This lesser variability ensures that the minor local modes receive non-negligible prior weights, and hence makes sense in this galaxy data scenario. As regards the prior on k, here we choose a discretized normal distribution on $\{1, \ldots, 30\}$ with mean 15 and variance 50. This reflects our belief that although all the values in $\{1, \ldots, 30\}$ receive significant prior masses, relatively large number of components is preferable in this application where many local modes are exhibited by the data.

In this application, following the previous convergence diagnostic method, we found the appropriate scales to be $a_{v_j^*} = a_{\tau_j^*} = a_{\omega_j} = 1$ for $j = 1, \ldots, k$. These
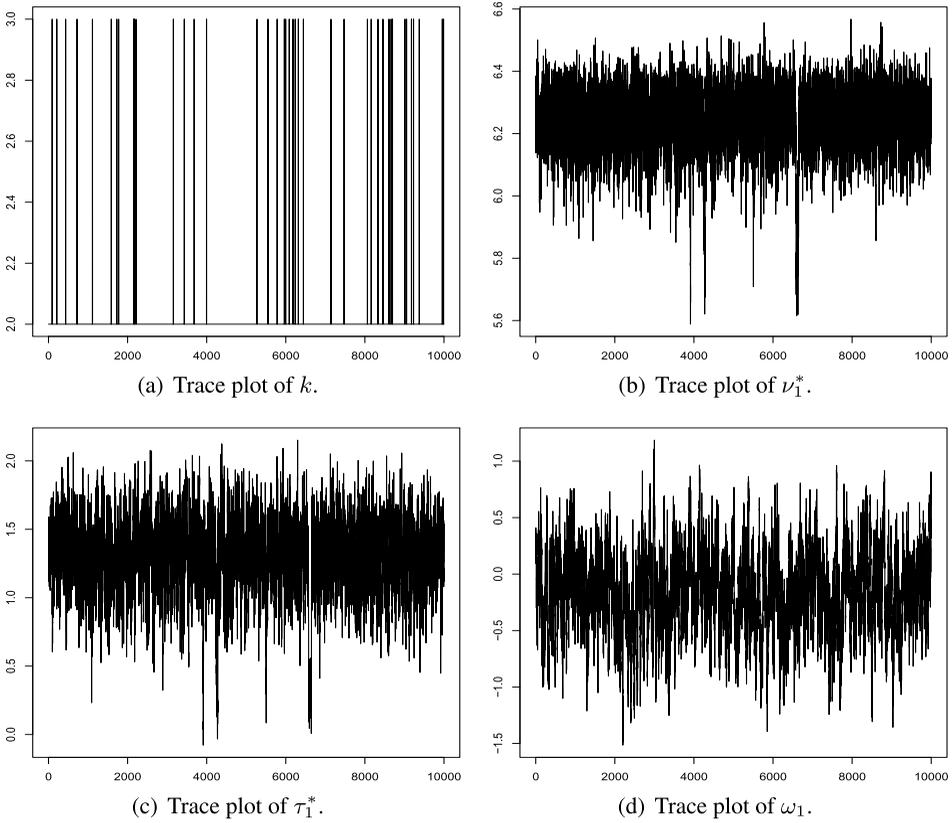
(a) Trace plot of k.

(b) Trace plot of ν_1^*.

(c) Trace plot of τ_1^*.

(d) Trace plot of ω_1.

Figure 14 *TTMCMC for the acidity data*: *Trace plots of k, ν_1^*, τ_1^* and ω_1.*

scales correspond to $\eta_1 = 0.01657$ and $\eta_2 = 0.01039$, which indicate good convergence. Here the overall acceptance rate, computed over 18×10^5 iterations, turned out to be 0.036388, while the birth, death and no-change rates are 0.007517, 0.007559 and 0.094195, respectively.

The implementation of TTMCMC in this application took 6 minutes and 33 seconds to yield 10,000 realizations after discarding a burn-in of 3×10^5 iterations, and then storing one iteration in every 150 iterations out of further 15×10^5 iterations following the burn-in period.

Note that, even in this challenging galaxy data application, the trace plots turned out to be quite reasonable, as shown in Figure 16. Thus, reasonable overall mixing behavior of the TTMCMC chain is indicated by the trace plots, consistent with the results of our credible region based convergence assessment criterion.

In this problem, the posterior distribution of k turned out to be much more variable than in the previous two cases. Here $k \in \{7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25\}$ with respective probabilities $\{0.0002, 0.0005, 0.0059, 0.0191, 0.0455, 0.0784, 0.1044, 0.1371, 0.1596, 0.1457, 0.1115, 0.0869,$
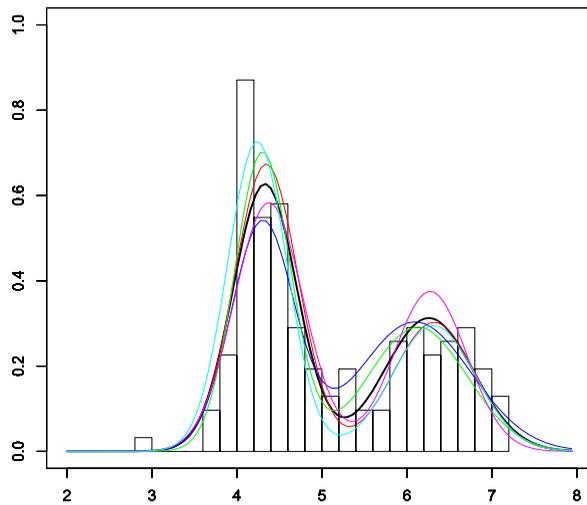
Figure 15 *TTMCMC for the acidity data*: *Goodness of fit of the posterior distribution of densities* (*coloured curves*) *to the observed data* (*histogram*). *The thick black curve is the modal density and the other coloured curves are some densities contained in the* 95% *HPD*.

0.0513, 0.0277, 0.0128, 0.0097, 0.0018, 0.0016, 0.0003}. Thus most of the possible values of k received positive posterior masses. It is also difficult to single out any particular value of k that is very strongly favoured by the posterior, unlike the previous two applications.

Figure 17 depicts the modal density and sample densities falling in the 95% HPD region, overlapped on the histogram of the observed data. The fit to the data seems to be quite encouraging with the sample densities capturing even the minor modes located at the extreme ends of the support of the data.

7.6 Comparison of TTMCMC with random walk RJMCMC with respect to the three real data sets

To save space, we have provided the details of the comparisons in Section S-10 of the supplement. Briefly, in all the three examples, random walk RJMCMC places much higher posterior mass to large number of components that are very implausible. The reason for this can be attributed to the product of the left truncated standard normal densities that features in the denominator of the acceptance ratio of the birth move of RJMCMC; since the aforementioned densities are bounded above by 1, this makes the acceptance rate for the birth move exceeding large, which, in effect, seriously slows down convergence. In addition, for the somewhat challenging galaxy data set, the random walk RJMCMC chain has extremely poor acceptance rate, and the chain hardly moved. Recall that this was the case for all
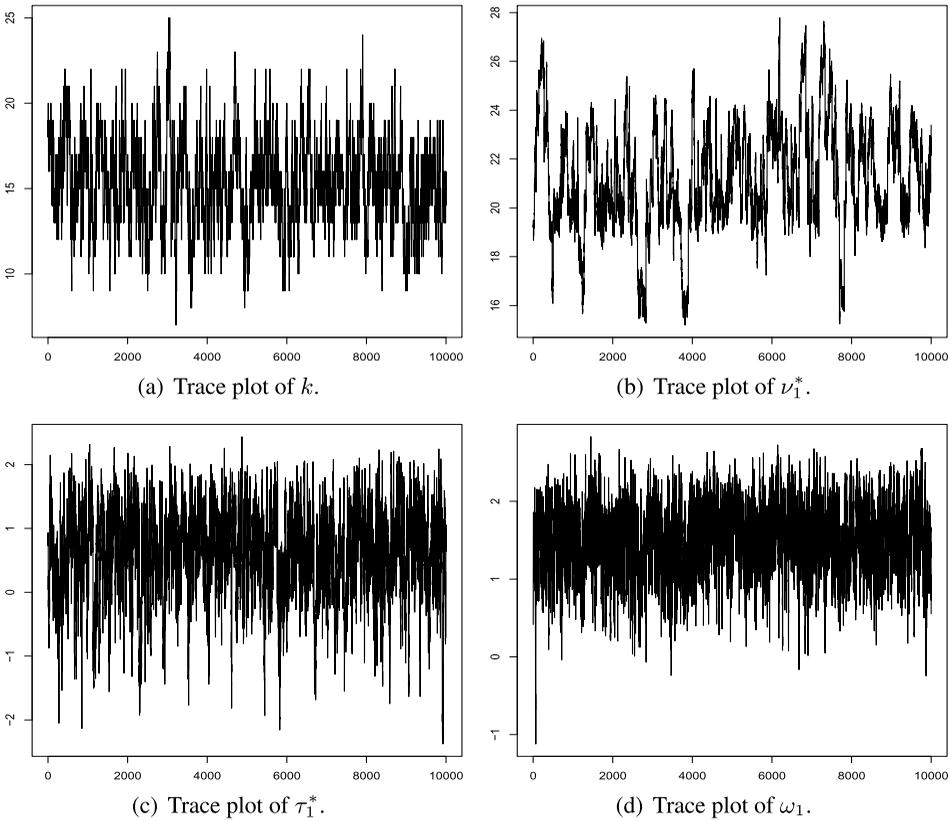
(a) Trace plot of k.

(b) Trace plot of ν_1^*.

(c) Trace plot of τ_1^*.

(d) Trace plot of ω_1.

Figure 16 *TTMCMC for the galaxy data*: *Trace plots of k, ν_1^*, τ_1^* and ω_1. Good mixing behavior of the TTMCMC chain is exhibited by the above panels.*

the four gamma mixture examples as well. Thus, random walk RJMCMC completely fails to act as the default RJMCMC algorithm.

7.7 Relevance of autocorrelation plots for convergence diagnosis in variable dimensions

Convergence assessment with the help of autocorrelations is not always appropriate in variable dimensional MCMC algorithms. Since there is no fixed Euclidean structure, parameters may not retain the same meaning throughout the iterations. To proceed with autocorrelation plots, it is necessary to focus attention on those parameters which retain constant interpretation across all models. In the mixture case the number of components may be considered. In this regard, the autocorrelation plots presented in Figure S-4 of the supplement reveal far superior mixing of the k-chain obtained by our TTMCMC sampler compared to random walk RJMCMC for all the three real data sets. In particular, for the galaxy data set, the RJMCMC based autocorrelations are simply hopeless!
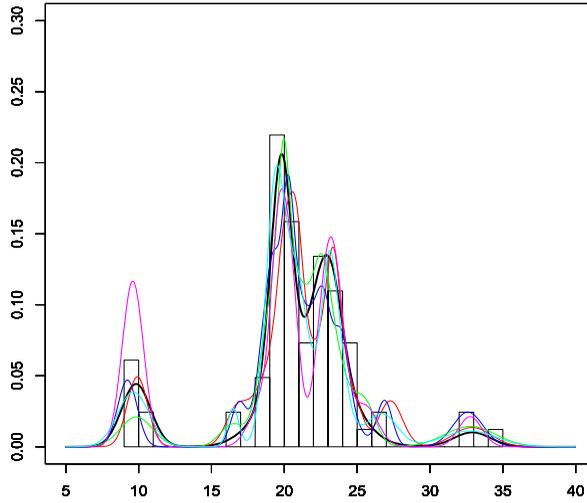
Figure 17 *TTMCMC for the galaxy data: Goodness of fit of the posterior distribution of densities (coloured curves) to the observed data (histogram). The thick black curve is the modal density and the other coloured curves are some densities contained in the 95% HPD.*

7.8 Comparison between TTMCMC and RJMCMC when the prior of Richardson and Green (1997) is considered

Further comparisons between TTMCMC and RJMCMC with respect to the prior structure and the algorithm of Richardson and Green (1997), are provided in Section S-10 of the supplement, in the context of the challenging galaxy data. We argue that actually their prior structure, where τ are made dependent in a way that they are approximately of the same size, is not expected to provide good fit to the observed histogram, but the large number of components supported by their algorithm as a result of its inherent bias as discussed, create the appearance of good fit. We further argue that the prior structure of Cappé, Robert and Rydén (2003), which is essentially the prior of Richardson and Green (1997) but τ are independent *a priori*, is a more appropriate prior for capturing the varieties of modes in the galaxy data.

8 TTMCMC for multivariate normal mixtures

We now consider i.i.d. *p*-variate data $\{\mathbf{y}_i = (y_{i1}, \ldots, y_{ip})^T; i = 1, \ldots, n\}$ arising from the *p*-variate normal mixture having the following density when the number of components is k: for $i = 1, \ldots, n$,

$$f(\mathbf{y}_i | \mathbf{\Theta}_k) = \sum_{j=1}^{k} \pi_j \frac{1}{(2\pi)^{p/2} |\mathbf{\Sigma}_j|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\mathbf{y}_i - \boldsymbol{\mu}_j)^T \mathbf{\Sigma}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_j)\right\}, \quad (8.1)$$

where $\mathbf{\Theta}_k = \{\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k, \mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_k, \pi_1, \ldots, \pi_k\}$.

Letting $\bar{\boldsymbol{y}}$ denote the p-dimensional sample mean vector and $\boldsymbol{S} = \text{diag}\{s_1^2, \ldots, s_p^2\}$ denote the diagonal matrix with the sample variances in the diagonal, we transform the data y_i, following Dellaportas and Papageorgiou (2006), to $\boldsymbol{S}^{-1/2}(\boldsymbol{y}_i - \bar{\boldsymbol{y}})$, once the data are generated.

8.1 Prior structure

Following Dellaportas and Papageorgiou (2006), we assume that *a priori*

$$[\boldsymbol{\mu}_j | \boldsymbol{\Sigma}_j] \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_j), \tag{8.2}$$

a p-variate normal with mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{\Sigma}_j$. We also assume following Dellaportas and Papageorgiou (2006) that

$$[\boldsymbol{\Sigma}_j] \sim W^{-1}(p+1, \boldsymbol{\Omega}), \tag{8.3}$$

an inverse-Wishart distributon with $(p+1)$ degrees of freedom and diagonal matrix $\boldsymbol{\Omega}$. However, instead of considering the gamma prior on the diagonal elements of $\boldsymbol{\Omega}$ as in Dellaportas and Papageorgiou (2006), we set all the diagonal elements equal to 1. This we do to avoid oversmoothness induced by the dependence structure between the $\boldsymbol{\Sigma}_j$; $j = 1, \ldots, k$, and to facilitate adaptive learning from the data. Recall that (see Section 7.8) a similar issue of oversmoothness seems to render the prior of Richardson and Green (1997) less appropriate for capturing the varieties of modes as compared to the prior of Cappé, Robert and Rydén (2003), in the univariate normal mixture case.

As before, we consider a discrete uniform prior for k on $\{1, 2, \ldots, 30\}$. Here we remark that although Dellaportas and Papageorgiou (2006) also report a discrete uniform prior on k, they did not specify the range.

8.2 TTMCMC strategy for multivariate situations

As before we reparameterize π_j as $\exp(w_j)/\sum_{i=1}^{k} \exp(w_i)$. As for $\boldsymbol{\Sigma}_j$, we consider the Cholesky decomposition $\boldsymbol{\Sigma}_j = \boldsymbol{L}\boldsymbol{L}^T$, where $\boldsymbol{L} = (L_{rs})_{r,s=1,\ldots,p}$ is the appropriate lower triangular matrix. Thus, there are $1 + p + p(p+1)/2$ number of parameters to be split in any given birth move given that the jth mixture component is chosen; w_j, the p components of $\boldsymbol{\mu}_j = (\mu_{j1}, \ldots, \mu_{jp})^T$ and $p(p+1)/2$ non-zero elements of $\boldsymbol{L}_j$. Thus, we need $1 + p + p(p+1)/2$ ε's to define our additive TTMCMC move types. The Jacobian of the birth move is given $2^{1+p+p(p+1)/2} \times a_{w_j} \times \prod_{r=1}^{p} a_{\mu_{jr}} \prod_{r \geq s=1}^{p} a_{L_{jrs}}$, where $a_{\mu_{jr}}$ is the scale for the additive transformation of the rth component of $\boldsymbol{\mu}_j$ and $a_{L_{jrs}}$ is the same for the (r, s)th element of $\boldsymbol{L}_{jrs}$, where $r \geq s$. The Jacobian for the death move is the inverse of that of the birth move with the relevant scale values. We reject the entire move if any of the diagonal elements of $\boldsymbol{L}$ becomes negative.

8.3 Simulation experiment with $p = 3$

Following Dellaportas and Papageorgiou (2006), we set generate 80, 100 and 100 data points from 3-variate normal distributions with means $\boldsymbol{\mu}_1 = (6, 4, 2)^T$, $\boldsymbol{\mu}_2 = (-11, -4, -1)^T$, $\boldsymbol{\mu}_3 = (-7, -11, -5)^T$ and covariance matrices

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 5 & 0 \\ 1 & 0 & 4 \end{pmatrix}, \qquad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 2 & -1.5 & 1 \\ -1.5 & 5 & 2 \\ 1 & 2 & 3 \end{pmatrix},$$

$$\boldsymbol{\Sigma}_3 = \begin{pmatrix} 5 & -1 & 1 \\ -1 & 4 & -2 \\ 1 & -2 & 3 \end{pmatrix},$$

respectively. Thus our data set consists of 280 data points and we fit our 3-variate mixture model to the data assuming unknown number of components.

Considering a burn-in of 3×10^6 iterations, we ran the TTMCMC algorithm for a further 3×10^6 iterations, storing one in 300 iterations, to obtain 10,000 realizations from the posterior. The implementation took 49 minutes and 14 seconds on our laptop. The overall acceptance rate turned out to be 0.038231 when the scales of the additive transformations are set to be 0.05 and 0.05 for the means and the elements of the Cholesky factors, and 0.5 for the weights. The birth, death and the no-change rates are 0.000002, 0.000015 and 0.114539, respectively. The trace plots shown in Figure 18 confirm excellent convergence properties of our algorithm, even in the multivariate case. Importantly, we obtained point mass at the true number of mixture components (as before, we do not rule out the possibility of missing some component other than 3 in our finite TTMCMC run). In contrast, Dellaportas and Papageorgiou (2006) report 6 models associated with $k = 1, \ldots, 6$, with 3 components receiving 0.9493 posterior probability.

Figure 19 depicts the modal density and sample densities falling in the 95% HPD region, overlapped on the histogram of the first component $\{y_{i1}; i = 1, \ldots, n\}$ (here $n = 280$) of the observed data. Excellent fit to the data is clearly indicated.

8.4 Simulation experiment with $p = 10$

We now consider application of TTMCMC to mixtures of $p = 10$ dimensional multivariate normals. Specifically, we first generate two mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ from two 10-dimensional, normal distributions $N_{10}(4\mathbf{1}_{10}, \boldsymbol{I}_{10})$ and $N_{10}(-5\mathbf{1}_{10}, \boldsymbol{I}_{10})$, where, for any integer $p \geq 1$, $\mathbf{1}_p$ is a d-component vector with each component 1, and $\boldsymbol{I}_p$ is the identity matrix of order p. Corresponding to the mean vectors $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, we specify covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ of the following form: the off-diagonal elements are given by $\sigma_j^2 \rho$ and the diagonal elements
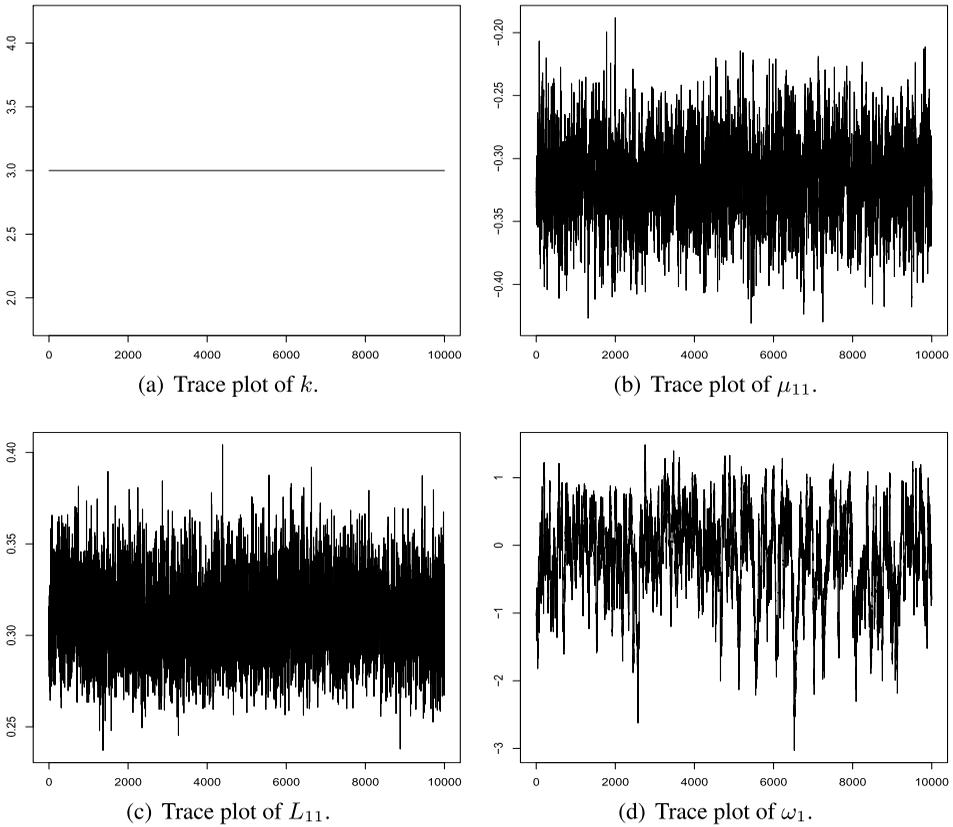
(a) Trace plot of k.

(b) Trace plot of μ_{11}.

(c) Trace plot of L_{11}.

(d) Trace plot of ω_1.

Figure 18 *TTMCMC for 3-dimensional case*: *Trace plots of k, μ_{11}, L_{11} and ω_1. Good mixing behavior of the TTMCMC chain is exhibited by the above panels.*

are all equal to σ_j^2, for $j = 1, 2$. For our illustration we consider $\sigma_1^2 = 4$, $\sigma_2^2 = 3$ and $\rho = 0.5$.

We then generate 300 realizations from $N_{10}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and 300 realizations from $N_{10}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, which constitute our data set $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_{600}\}$ of size 600.

We use the same TTMCMC algorithm as in the 3-dimensional experiment, but as to be anticipated for higher dimensions, the convergence was slower compared to the 3-dimensional example. To improve mixing, we employed the following strategy. At the end of each iteration $t \geq 1$, we simulated $r^{(t)} \sim N(0, 1)$ and proposed the further additive transformation $\boldsymbol{\Theta}^{(t)} \mapsto \boldsymbol{\Theta}^{(t)} + \boldsymbol{a}r^{(t)}$, where $\boldsymbol{\Theta}^{(t)}$ denotes the stage of the parameters at iteration t, and $\boldsymbol{a}$ denotes the vector of scaling constants for the additive transformation. We then calculated the acceptance probability of this proposal in the usual TMCMC set-up to either accept the new proposal $\boldsymbol{\Theta}^{(t)} + \boldsymbol{a}r^{(t)}$ or to remain at $\boldsymbol{\Theta}^{(t)}$. Such a strategy has also been employed by Mukhopadhyay and Bhattacharya (2013) to improve mixing in the con-
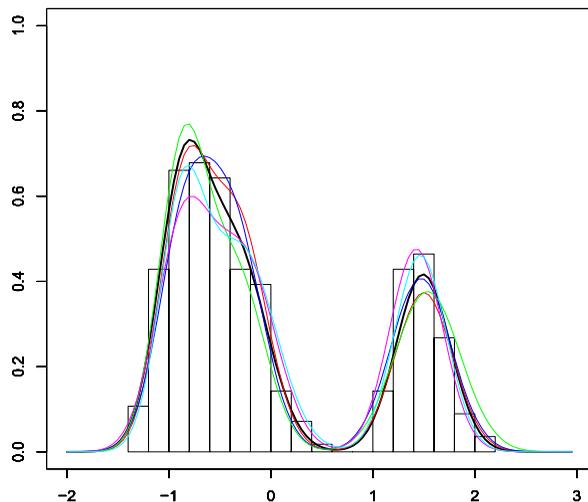
Figure 19 *TTMCMC for* 3-*dimensional case*: *Goodness of fit of the posterior distribution of densities* (*coloured curves*) *to the histogram of the first component of the observed data* $\{y_{i1}; i = 1, \ldots, 285\}$. *The thick black curve is the modal density and the other coloured curves are some densities contained in the* 95% *HPD.*

text of palaeoclimate modeling. The strategy is akin to the so-called generalized Gibbs/MH methods in fixed-dimensional set-ups have the potential of improving mixing (see, for example, Liu, Liang and Wong (2000), Liu (2001); see also Liu and Yu (1999)). Further details can be found in the supplement of Dutta and Bhattacharya (2014).

For our purpose, we chose the scales of the additive transformation associated with the original TTMCMC to be relatively large; 0.5 for the means, 0.05 for the Cholesky components and 1.5 for the weights, while for the mixing improvement step we chose the scales to be 1/10th of the above scales. This ensures relatively small acceptance rate but large moves for the original TTMCMC steps but much higher acceptance rate at the mixing improvement step.

However, in spite of the above strategy, the mixing improvement was not dramatic in our case, and still a considerably long run was necessary. As such, we discarded the first 3×10^7 iterations, and stored one in 300 iterations out of the next 12×10^7 iterations to store 4×10^5 iterations. We applied further thinning of size 40 to the stored samples, finally storing 10,000 iterations. The entire procedure took about 68 hours on our VMWare. The overall acceptance rate, birth rate, death rate and the no-change rates in this implementation are 0.008173, 0.00014, 0.00037 and 0.023949, respectively.

The trace plots and the goodness of fit (for the first co-ordinate of the 10-dimensional data) diagram shown in Figures 20 and 21 vindicate satisfactory per-
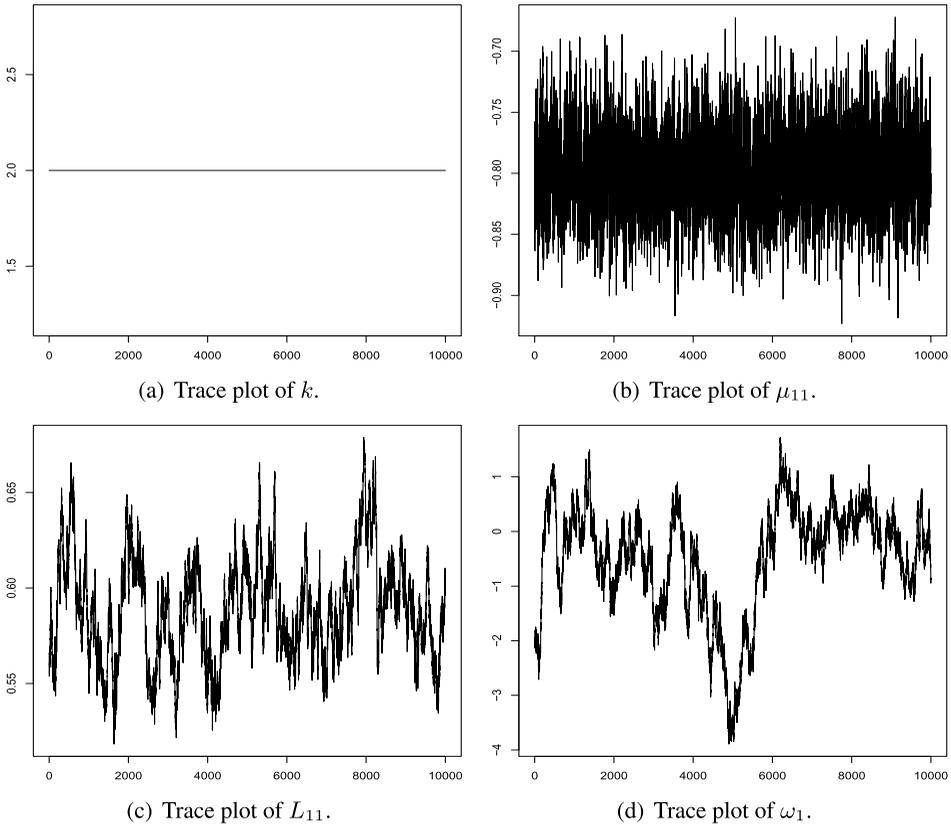
(a) Trace plot of k.

(b) Trace plot of μ_{11}.

(c) Trace plot of L_{11}.

(d) Trace plot of ω_1.

Figure 20 *TTMCMC for* 10-*dimensional case*: *Trace plots of* k, μ_{11}, L_{11} *and* ω_1. *Adequate mixing behavior of the TTMCMC chain is exhibited by the above panels*.

formance of our method, in spite of high dimensionality. Importantly, the correct number of components, namely, $k = 2$ has been identified correctly.

8.5 Simulation experiment with $p = 20$

We conduct a further experiment, now with dimension $p = 20$. Our data generation mechanism remains the same as in Section 8.4, only the dimension is increased from $p = 10$ to $p = 20$. Our TTMCMC algorithm also remains almost the same, with the same mixing improvement strategy. We again obtain 10,000 samples by thinning from a total of 15×10^7 iterations. In this case, the overall acceptance rate, birth rate, death rate and the no-change rate are 0.00741, 0.00019, 0.000421 and 0.02163, respectively. The time taken is 136 hours and 44 minutes. The trace plots and the goodness-of-fit diagram depicted in Figures 22 and 23 once again speak in favour of our ideas, in particular, the great automation of our method, irrespective of dimensions.
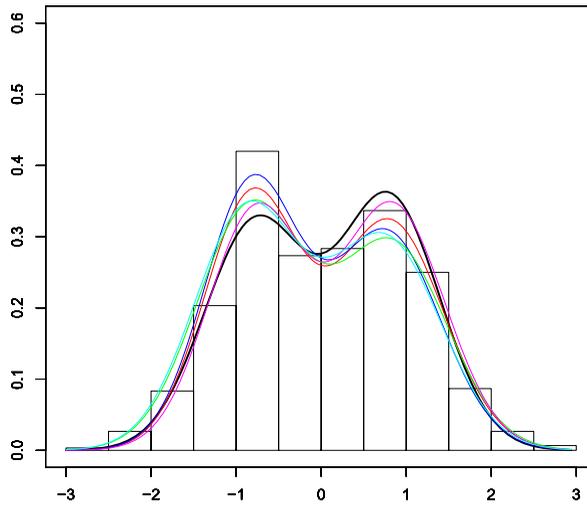
Figure 21 *TTMCMC for* 10-*dimensional case*: *Goodness of fit of the posterior distribution of densities* (*coloured curves*) *to the histogram of the first component of the observed data* $\{y_{i1}; i = 1, \ldots, 600\}$. *The thick black curve is the modal density and the other coloured curves are some densities contained in the* 95% *HPD*.

9 Conclusion

The transformation based concepts of TMCMC in the fixed-dimensional set-up has led to the interesting variable-dimensional counterpart TTMCMC just as the traditional Metropolis–Hastings methodology has led to RJMCMC. Consequently, the advantages of TMCMC over Metropolis–Hastings are expected to carry over to TTMCMC as compared to RJMCMC. Indeed, as we demonstrated in this paper, TTMCMC is simple to implement, can update all the (variable number of) parameters in a single block while maintaining reasonable acceptance rates thanks to drastic effective reduction of the dimensionality. In fact, TTMCMC effectively reduces the variable dimensional problem to a fixed dimensional problem involving a single ε or just a few, fixed number of ε's, given any move type within the birth, death or no-change moves. The block updating strategy of TTMCMC using ε or a few ε's also ensures huge computational savings. Furthermore, the mixture-type proposal distributions associated with TTMCMC ensures reasonable mixing properties.

There are three key features that manifested themselves in our comparative studies on TTMCMC and RJMCMC. First, TTMCMC yields reasonable acceptance rates, which are larger than those of RJMCMC for the same scales of the additive transformations. Importantly, in the gamma mixtures and the galaxy example, RJMCMC yields extremely poor acceptance rate, while that of TTMCMC is quite reasonable, for the same scales.
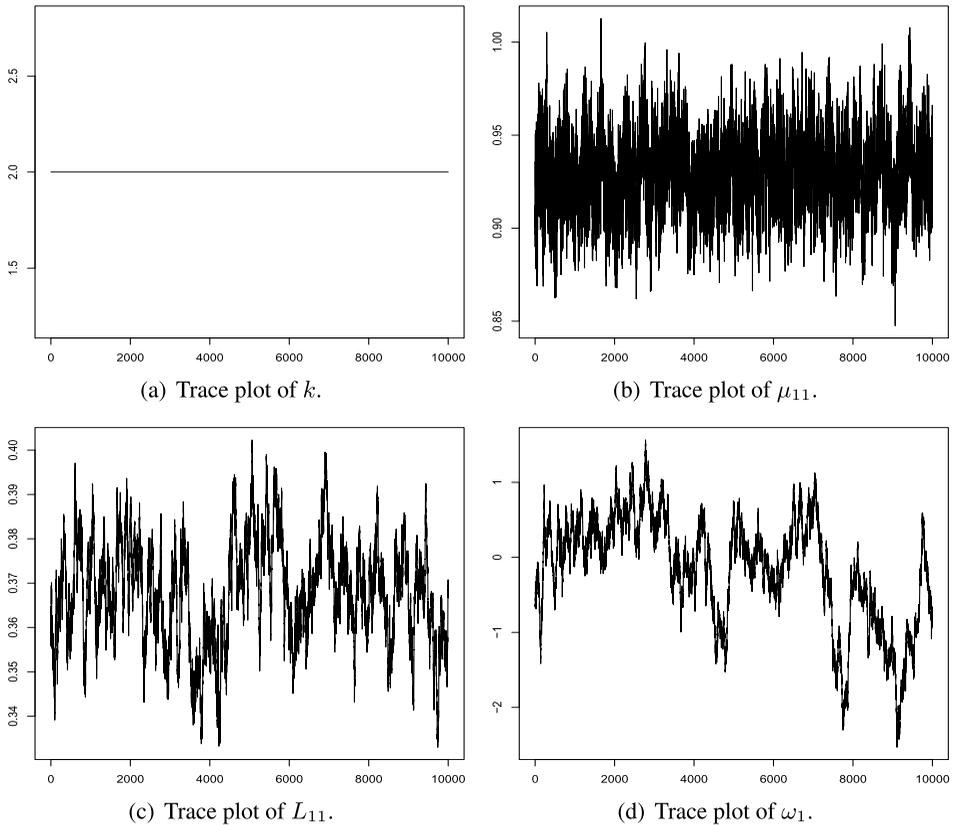
(a) Trace plot of k.

(b) Trace plot of μ_{11}.

(c) Trace plot of L_{11}.

(d) Trace plot of ω_1.

Figure 22 *TTMCMC for* 20*-dimensional case*: *Trace plots of* k, μ_{11}, L_{11} *and* ω_1. *Adequate mixing behavior of the TTMCMC chain is exhibited by the above panels.*

Second, ensuring reasonable mixing is a very challenging issue in variable dimensional problems. Here TTMCMC outperforms RJMCMC very significantly in all the cases, as vindicated by the autocorrelation plots shown in Figure S-4 of the supplement. In other words, even in univariate situations, the random walk RJMCMC completely fails to compete with TTMCMC.

Third, it seems to be infeasible to devise appropriate RJMCMC move types in high-dimensional contexts. Indeed, Dellaportas and Papageorgiou (2006) consider a maximum of only 5-dimensional example for RJMCMC application. On the other hand, we have demonstrated that our simple additive TTMCMC works even for dimensions as large as 20. In this regard it is useful to note that the split-merge proposals of Jain and Neal (2004) and Jain and Neal (2007) are perhaps better candidates compared to those of Richardson and Green (1997) and Dellaportas and Papageorgiou (2006) as they update all the allocation variables simultaneously, rather than Gibbs sampling. Since TTMCMC also generally updates all the vari-
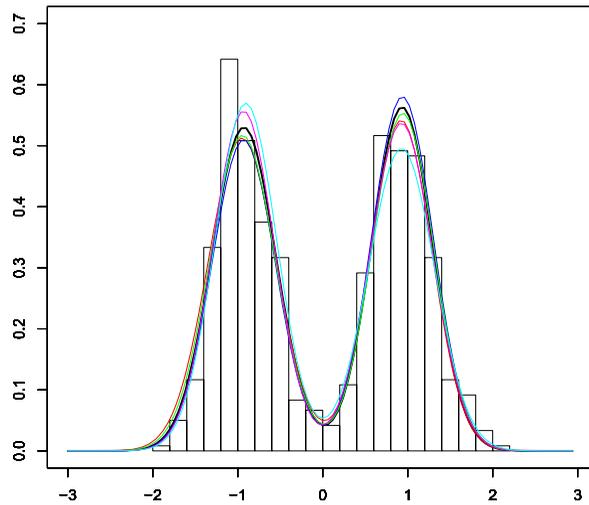
Figure 23 *TTMCMC for* 20-*dimensional case*: *Goodness of fit of the posterior distribution of densities* (*coloured curves*) *to the histogram of the first component of the observed data* $\{y_{i1}; i = 1, \ldots, 600\}$. *The thick black curve is the modal density and the other coloured curves are some densities contained in the* 95% *HPD*.

ables in a single block, the general principles of their algorithm and TTMCMC match. But a key difference is that we do not introduce allocation variables for mixture updation, and hence have much less number of variables to update, which is expected to lead to better acceptance rate in our case. It is also to be noted that the algorithms of Jain and Neal (2004) and Jain and Neal (2007) are devised for mixtures only, not for general variable-dimensional problems. In contrast, our default additive TTMCMC that we used for mixtures can be applied to all variable dimensional problems.

A further issue with RJMCMC is that it tends to support more components than are expected. The main issue responsible for this possible non-convergence is the requirement of dimension-matching for RJMCMC implementation. This condition forces the acceptance ratio for the dimension-changing moves to depend upon the proposal density either via the denominator (birth move) or through the numerator (death move). Thus, unlike fixed-dimensional Metropolis–Hastings, the acceptance ratio is not balanced by the presence of the proposal density in both numerator and denominator. As already remarked in the discussion following Algorithm 3.1, this unbalanced nature of the RJMCMC acceptance ratio causes large number of birth moves if the proposal density is uniformly bounded by 1, as in our examples. Since TTMCMC does not require dimension-matching, it has been possible to free the corresponding acceptance ratio of the proposal density, which, in turn, completely solves the problem of bias towards large number of models in finite number of iterations.

The wisdom that emerges from the investigations and the subsequent analyses is that even the simplest version of TTMCMC, namely, additive TTMCMC, is capable enough of exploring challenging variable-dimensional posteriors, providing ample support to our claim of automation inherent within TTMCMC. On the other hand, as our implementations show, the corresponding random walk RJMCMC do not measure up at all. In principle, there may exist RJMCMC algorithms which may perhaps perform reasonably in terms of convergence, but at the cost of being problem-specific, complicated, hard-to-implement, and computationally burdensome.

Also, very importantly, as we showed, our simple additive transformation exhibited very decent performance even in dimension as large as 20, thus providing a large boost to our claim of automation. To our knowledge, there exists no instance of RJMCMC that works in such high dimension.

Thus, as per our experiments and knowledge, TTMCMC is remarkably close to automation, while automation for RJMCMC is nowhere in sight.

Apart from developing TTMCMC, we have also proposed, in a separate supplementary material, a general methodology for summarizing the posterior distributions of densities. In particular, we have prescribed a procedure for obtaining the modes and desired HPD regions of the posterior distribution of density functions. Moreover, using these concepts as basis, we have proposed a convergence diagnostic criterion for the underlying TTMCMC algorithm, which is again very generally applicable. The convergence diagnostic method seems to be particularly useful in variable-dimensional contexts, where determining convergence is far more difficult than fixed-dimensional situations. Also, as we demonstrated with our applications, in the absence of optimal scaling theory in variable-dimensional situations, the criterion can provide guidance regarding choices of the scales of default additive TTMCMC.

Our results demonstrate that additive TTMCMC is promising enough to qualify as the default variable-dimensional algorithm. This is also vindicated by the excellent performances of TTMCMC in challenging spatio-temporal problems investigated by these authors and others. In this paper, we restricted ourselves to mixture models because of their high standing in statistics and challenging nature of the associated variable-dimensional problem. However, in a separate paper we shall present detailed comparisons of TTMCMC and RJMCMC with respect to various other variable-dimensional problems. Our investigations are on and we seek to establish TTMCMC as a far superior alternative compared to RJMCMC.

Acknowledgments

Supplementary Material

Supplement to "Transdimensional transformation based Markov chain Monte Carlo" (DOI: 10.1214/17-BJPS380SUPP; .pdf). In Sections S-1 and S-2 we prove detailed balance, irreducibility and aperiodicity of TTMCMC. In Section S-3 we provide the general TTMCMC algorithm for jumping m dimensions, and in Section S-4 we prove detailed balance for this algorithm. The TTMCMC algorithm for jumping more than one dimension at a time when several sets of parameters are related, is detailed in Section S-5. A brief discussion on label switching in mixtures is included in Section S-6, and in Section S-7, an effective method of summarization of the posterior distribution of mixture densities is considered. In Section S-8 we shed light on the important problem of convergence diagnostics in mixture problems, for both known and unknown number of components. Further simulation studies with gamma mixtures with varying data sizes are elaborated in Section S-9. In Section S-10 we include a detailed comparison between additive TTMCMC and random walk RJMCMC with respect to three real data sets, and finally, in Section S-11, we provide detailed comparisons between additive TTMCMC and RJMCMC with respect to the prior structure and the algorithm of Richardson and Green (1997) in the galaxy data context.

References

Al-Awadhi, F. H. and Jennison, C. (2004). Improving the acceptance rate of reversible-jump MCMC proposals. *Statistics and Probability Letters* **69**, 189–198. MR2086905

Bhattacharya, S. (2008). Gibbs sampling based Bayesian analysis of mixtures with unknown number of components. *Sankhya. Series B* **70**, 133–155. MR2507480

Brooks, S. P., Giudici, P. and Roberts, G. O. (2003). Efficient construction of reversible-jump Markov chain Monte Carlo proposal distributions (with discussion). *Journal of the Royal Statistical Society. Series B* **65**, 3–55. MR1959092

Cappé, O., Robert, C. P. and Rydén, T. (2003). Reversible jump, birth-and-death and more general continuous time Markov chain Monte Carlo samplers. *Journal of the Royal Statistical Society. Series B* **65**, 679–700.

Chu, J.-H., Clyde, M. A. and Liang, F. (2009). Bayesian function estimation using continuous wavelet dictionaries. *Statistica Sinica* **19**, 1419–1438. MR2589190

Das, M. and Bhattacharya, S. (2016). Nonstationary nonparametric Bayesian spatio-temporal modeling using kernel convolution of order based dependent Dirichlet process. Submitted. Available at arXiv:1405.4955.

Das, M. and Bhattacharya, S. (2017). Supplement to "Transdimensional transformation based Markov chain Monte Carlo." DOI:10.1214/17-BJPS380SUPP.

Dellaportas, P. and Forster, J. J. (1999). Markov chain Monte Carlo model determination for hierarchical and graphical log-linear models. *Biometrika* **86**, 615–633. MR1723782

Dellaportas, P., Forster, J. J. and Ntzoufras, I. (2002). On Bayesian model and variable selection using MCMC. *Statistics and Computing* **12**, 27–36. MR1877577

Dellaportas, P. and Papageorgiou, I. (2006). Multivariate mixtures of normals with unknown number of components. *Statistics and Computing* **16**, 57–68. MR2224189

Denison, D. G. T., Mallick, B. K. and Smith, A. F. M. (1998). Automatic Bayesian curve fitting. *Journal of the Royal Statistical Society. Series B* **60**, 333–350. MR1616029

Dey, K. K. and Bhattacharya, S. (2016). On geometric ergodicity of additive and multiplicative transformation based Markov chain Monte Carlo in high dimensions. *Brazilian Journal of Probability and Statistics* **30**, 570–613. Also available at arXiv:1312.0915v2.

Dey, K. K. and Bhattacharya, S. (2017a). A brief review of optimal scaling of the main MCMC approaches and optimal scaling of additive TMCMC under non-regular cases. *Brazilian Journal of Probability and Statistics*. To appear. Available at arXiv:1405.0913.

Dey, K. K. and Bhattacharya, S. (2017b). A brief tutorial on transformation based Markov chain Monte Carlo and optimal scaling of the additive transformation. *Brazilian Journal of Probability and Statistics* **31**, 569–617. Also available at arXiv:1307.1446. MR3693982

Dutta, S. and Bhattacharya, S. (2014). Markov chain Monte Carlo based on deterministic transformations. *Statistical Methodology* **16**, 100–116. Also available at arXiv:1106.5850. Supplement available at 1306.6684.

Escobar, M. D. and West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* **90**, 577–588. MR1340510

Fan, Y. and Sisson, S. A. (2011). Reversible jump MCMC. In *Handbook of Markov Chain Monte Carlo* (S. Brooks, A. Gelman, G. L. Jones and X.-L. Meng, eds.) 67–87. New York: Chapman & Hall/CRC. MR2858445

Gilks, W. R., Roberts, G. O. and George, E. I. (1994). Adaptive direction sampling. *Journal of the Royal Statistical Society. Series D (The Statistician)* **43**, 179–189. MR1276441

Godsill, S. J. (2003). Discussion of "Trans-dimensional Markov chain Monte Carlo". In *Highly Structured Stochastic Systems* (P. J. Green, N. L. Hjort and S. Richardson, eds.) 199–203. Oxford, UK: Oxford University Press. MR2082410

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711–732. MR1380810

Green, P. J. (2003). Trans-dimensional Markov chain Monte Carlo. In *Highly Structured Stochastic Systems* (P. J. Green, N. L. Hjort and S. Richardson, eds.) 179–198. Oxford, UK: Oxford University Press. MR2082410

Guan, Y. and Krone, S. M. (2007). Small-world MCMC and convergence to multi-modal distributions: From slow mixing to fast mixing. *The Annals of Applied Probability* **17**, 284–304. MR2292588

Jain, S. and Neal, R. M. (2004). A split-merge Markov chain Monte Carlo procedure for the Dirichlet process mixture model. *Journal of Computational and Graphical Statistics* **13**, 158–182.

Jain, S. and Neal, R. M. (2007). Splitting and merging components of a nonconjugate Dirichlet process mixture model. *Bayesian Analysis* **2**, 445–472. MR2342168

Liu, J. (2001). *Monte Carlo Strategies in Scientific Computing*. New York: Springer. MR1842342

Liu, J. S., Liang, F. and Wong, W. H. (2000). The multiple-try method and local optimization in metropolis sampling. *Journal of the American Statistical Association* **95**, 121–134. MR1803145

Liu, J. S. and Yu, Y. N. (1999). Parameter expansion for data augmentation. *Journal of the American Statistical Association* **94**, 1264–1274. MR1731488

Lopes, H. F. and West, M. (2004). Bayesian model assessment in factor analysis. *Statistica Sinica* **14**, 41–67. MR2036762

Martino, L., Elvira, V., Luengo, D., Corander, J. and Louzada, F. (2016). Orthogonal parallel MCMC methods for sampling and optimization. *Digital Signal Processing* **58**, 64–84.

Møller, J. and Waagepetersen, R. P. (2004). *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton, Florida: Chapman & Hall/CRC. MR2004226

Mukhopadhyay, S. and Bhattacharya, S. (2013). Cross-validation based assessment of a new Bayesian palaeoclimate model. *Environmetrics* **24**, 550–568. MR3161979

Richardson, S. and Green, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society. Series B* **59**, 731–792. MR1483213

Robert, C. P. (2003). Advances in MCMC: A discussion. In *Highly Structured Stochastic Systems* (P. J. Green, N. L. Hjort and S. Richardson, eds.) 167–171. Oxford, UK: Oxford University Press. MR2082404

Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods*, 2nd edn. New York: Springer. MR2080278

Sisson, S. A. (2005). Transdimensional Markov chains: A decade of progress and future perspectives. *Journal of the American Statistical Association* **100**, 1077–1089. MR2201033

Storvik, G. (2011). On the flexibility of Metropolis-Hastings acceptance probabilities in auxiliary variable proposal generation. *Scandinavian Journal of Statistics* **38**, 342–358. MR2829604

Vermaak, J., Andrieu, C., Doucet, A. and Godsill, S. J. (2004). Reversible jump Markov chain Monte Carlo strategies for Bayesian model selection in autoregressive processes. *Journal of Time Series Analysis* **25**, 785–809. MR2099640

Wiper, M., Insua, D. R. and Ruggeri, F. (2001). Mixtures of gamma distributions with applications. *Journal of Computational and Graphical Statistics* **10**, 440–454. MR1939034

Interdisciplinary Statistical Research Unit
Indian Statistical Institute
203, B. T. Road
Kolkata 700108
India
E-mail: sourabh@isical.ac.in