On Multiple try versions of Transformation based Markov Chain Monte Carlo

Kushal Kr. Dey[†] , Sourabh Bhattacharya[‡+]

[†] Department of Statistics, University of Chicago, USA
[‡] Interdisciplinary Statistical Research Unit, Indian Statistical Institute, Kolkata
+ Corresponding author: bhsourabh@gmail.com

1 Methods

In this paper, we present two multiple try versions of Additive Transformation based Markov Chain Monte Carlo.

Bullet points so far

- Methods: Algorithms for Versions 1 and 2 of MT-TMCMC
- Methods: To present the schematic diagram of the MT-TMCMC method (similar to Fig 1 of Martino and Read 2013, **?**).
- Methods: Show detailed balance condition holds for these methods
- Methods: Irreducibility and aperiodicity of the MT-TMCMC method (will be similar to the one for TMCMC).
- Results: Simulation comparison (KS test) between the two versions of MT-TMCMC
- Results: Simulation comparison (KS test) between general TMCMC and MT-TMCMC and RWMH.
- Results: A toy simulation study depicting how quick the space traversal is for MT-TMCMC (similar to Fig 4 of Martino and Read 2013, **?**).
- Results: Compare this approach and its speed in high dimensions to the multiple try RWMH approach as suggested by Liu, Liang and Wong 2001, **?**.
- : Discussions: Talk about the implementation of MT-TMCMC in the package **tmcmcR**.

The other possible directions worth investigating

- Optimal scaling study for MT-TMCMC method (similar in line to Bedard, Douc and Moulines 2012, **?**.
- Geometric ergodicity of the MT-TMCMC sampler.

We next prove that the detailed balance condition holds for Versions 1 and 2 of the Multiple Try TMCMC approach.

Theorem 1.1. *Detailed balance holds for Algorithm 1 and Algorithm 2, implying that the chains generated by these algorithms converges to the target density* π*.*

Algorithm 1 Multiple Try Transformation based Markov Chain Monte Carlo (MT-TMCMC): Ver. 1

- Suppose we want to draw samples from a D dimensional target density π.
- Start at the initial value x_0 and let the total number of iterations be N.
- For $t = 0, 1, 2, \cdots, N-1$,

1. Fix parameters K and L (user defined). Draw step sizes $\epsilon_1^{(t)}, \epsilon_2^{(t)}, \cdots, \epsilon_K^{(t)} \sim g(.)$ where $supp(g) = \mathbb{R}^+$. Draw step signs $b_{kl}^{(t)}$ for $k = 1, 2, \cdots, K$ and $l = 1, 2, \cdots, L$,such that

$$\begin{aligned} b_{kl}^{(t)} &= +1 \qquad prob \quad 0.5 \\ b_{kl}^{(t)} &= -1 \qquad prob \quad 0.5 \end{aligned}$$

2. Propose candidate moves $y_{kl}^{(t)}$ as follows

$$y_{kl}^{(t)} = x^{(t)} + b_{kl}^{(t)} \epsilon_k^{(t)} \qquad k = 1, 2, \cdots, K, \qquad l = 1, 2, \cdots, L$$

3. Compute target density values at each candidate move, namely π_{kl} and draw a random move $y^\star$ from $y_{kl}^{(t)}$ with probability $\frac{\pi_{kl}}{\sum k, l \pi_{kl}}$.
4. Determine the k and l corresponding to the move selected.

$$(k^\times, l^\times) := find \left\{ (k, l) : y_{kl}^{(t)} = y^\star \right\}$$

5. Draw $(K-1)$ step sizes $\epsilon_1^{(*,t)}, \epsilon_2^{(*,t)}, \cdots, \epsilon_{k^\star-1}^{(*,t)}, \epsilon_{k^\star+1}^{(*,t)}, \cdots, \epsilon_K^{(*,t)} \sim g(.)$ and let $\epsilon_{k^\times}^{(*,t)}$ deterministically determined as

$$\epsilon_{k^\times}^{(*,t)} := \left| y_d^\star - x_d^{(t)} \right| \qquad \forall d = 1, 2, \cdots, D$$

6. Similarly draw $b_{kl}^{(*,t)}$ for $k = 1, 2, \cdots, K$ and $l = 1, 2, \cdots, (L-1)$, similar to $b_{kl}^{(t)}$ and define

$$b_{k^\times l^\times}^{(*,t)} := -\frac{y_d^\star - x_d^{(t)}}{\epsilon_{k^\times}^{(*,t)}}$$

7. Draw reverse step candidates $x_{kl}^{(*,t)}$ as follows

$$x_{kl}^{(*,t)} := y_{kl}^{(t)} + b_{kl}^{(*,t)} \epsilon_k^{(*,t)}$$

8. By definition of $\epsilon^{(*,t)}$ and $b^{(*,t)}$, we have

$$x_{kl}^{(*,t)} = x^{(t)}$$

9. Accept the candidate move $y^{(\star)}$ with acceptance rate defined as follows

$$acc := min \left\{ 1, \frac{\sum_{kl} y_{kl}^{(t)}}{\sum_{kl} x_{kl}^{(*,t)}} \right\}$$

10. Draw $u \sim \mathcal{U}(0,1)$. If $u < acc$,

$$x^{(t+1)} := y^{(t)}$$

else

$$x^{(t+1)} := x^{(t)}$$

- End For

Algorithm 2 Multiple Try Transformation based Markov Chain Monte Carlo(MT-TMCMC): Ver. 2

- Suppose we want to draw samples from a D dimensional target density π.
- Start at the initial value x_0 and let the total number of iterations be N.
- For $t = 0, 1, 2, \cdots, N-1$,

1. Fix parameters K and L (user defined). Draw step sizes $\epsilon_1^{(t)}, \epsilon_2^{(t)}, \cdots, \epsilon_K^{(t)} \sim g(.)$ where $supp(g) = \mathbb{R}^+$. Draw step signs $b_{kl}^{(t)}$ for $k = 1, 2, \cdots, K$ and $l = 1, 2, \cdots, L$,such that

$$\begin{aligned} b_{kl}^{(t)} &= +1 \qquad prob \quad 0.5 \\ b_{kl}^{(t)} &= -1 \qquad prob \quad 0.5 \end{aligned}$$

2. Propose candidate moves $y_{kl}^{(t)}$ as follows

$$y_{kl}^{(t)} = x^{(t)} + b_{kl}^{(t)} \epsilon_k^{(t)} \qquad k = 1, 2, \cdots, K, \qquad l = 1, 2, \cdots, L$$

3. Compute target density values at each candidate move, namely π_{kl} and draw a random move $y^\star$ from $y_{kl}^{(t)}$ with probability $\frac{\pi_{kl}}{\sum k, l \pi_{kl}}$.
4. Define

$$(k^{'}, l^{'}) := find\left\{(k, l) : y_{kl}^{(t)} = y^\star\right\}$$

5. Similarly draw $b_{kl}^{(*,t)}$ for $k = 1, 2, \cdots, K$, $l = 1, 2, \cdots, L$, $(k, l) \neq (k^{'}, l^{'})$ and similar to $b_{kl}^{(t)}$, define

$$b_{k'l'}^{(*,t)} := -b_{k'l'}^{(t)}$$

6. Draw reverse step candidates $x_{kl}^{(*,t)}$ as follows

$$x_{kl}^{(*,t)} := y_{kl}^{(t)} + b_{kl}^{(*,t)} \epsilon_k^{(*,t)}$$

7. By definition of $\epsilon^{(*,t)}$ and $b^{(*,t)}$, we have

$$x_{k'l'}^{(*,t)} = x^{(t)}$$

8. Accept the candidate move $y^{(\star)}$ with acceptance rate defined as follows

$$acc := min\left\{1, \frac{\sum_{kl} y_{kl}^{(t)}}{\sum_{kl} x_{kl}^{(*,t)}}\right\}$$

9. Draw $u \sim \mathcal{U}(0, 1)$. If $u < acc$,

$$x^{(t+1)} := y^{(t)}$$

else

$$x^{(t+1)} := x^{(t)}$$

- End For

The proof of detailed balance is motivated by the approach followed by **?** and **?**. We first prove this result for Algorithm 1. To prove this, if x be the current state of the Markov chain and y be the subsequent move, let $q(.|x)$ be the transition probability density function given the current location being x. Then we have to show

$$\pi(x)\mathcal{Q}(y|x) = \pi(y)\mathcal{Q}(x|y)$$

Note that we can write

$$\mathcal{Q}(y|x) = KL\mathcal{Q}(y = y_{lk}|x, k, l)$$

where $k \in \{1, 2, \cdots, K\}$ be the set of step-sizes and $l \in \{1, 2, \cdots, L\}$ be the set of step signs per step size.

So, we have to ideally show for each k and l,

$$\pi(x)\mathcal{Q}(y|x, k, l) = \pi(y)\mathcal{Q}(x|y, k, l) \tag{1}$$

$$\begin{aligned}
\pi(x)\mathcal{Q}(y|x,k,l) &= \pi(x) \sum_{\left\{\substack{b_{k',l'} \in \{-1,+1\} \\ (k,l)\neq(k',l') \\ b^{(r)}_{k',l'} \in \{-1,+1\} \\ (k,l)\neq(k',l')}\right\}} \int_{\left\{\substack{\epsilon_k>0, \\ k\neq k' \\ \epsilon^{(r)}_k>0 \\ k\neq k'}\right\}} \frac{\pi(y)}{\pi(y) + \sum_{(k',l')\neq(k,l)} \pi(y_{k',l'})} \\
&\times min\left\{1, \frac{\pi(y) + \sum_{(k',l')\neq(k,l)} \pi(y_{k',l'})}{\pi(x) + \sum_{(k',l')\neq(k,l)} \pi(x_{*,k',l'})}\right\} d\epsilon_1 \cdots d\epsilon_{k-1} d\epsilon_{k+1} \cdots \epsilon_K d\epsilon^{(r)}_1 \cdots d\epsilon^{(r)}_{k-1} d\epsilon^{(r)}_{k+1} \cdots \epsilon^{(r)}_K \\
&= \sum_{\left\{\substack{b_{k',l'} \in \{-1,+1\} \\ (k,l)\neq(k',l') \\ b^{(r)}_{k',l'} \in \{-1,+1\} \\ (k,l)\neq(k',l')}\right\}} \int_{\left\{\substack{\epsilon_k>0, \\ k\neq k' \\ \epsilon^{(r)}_k>0 \\ k\neq k'}\right\}} min\left\{\frac{1}{\pi(y) + \sum_{(k',l')\neq(k,l)} \pi(y_{k',l'})}, \frac{1}{\pi(x) + \sum_{(k',l')\neq(k,l)} \pi(x_{*,k',l'})}\right\} \\
&\times \pi(x)\pi(y) d\epsilon_1 \cdots d\epsilon_{k-1} d\epsilon_{k+1} \cdots \epsilon_K d\epsilon^{(r)}_1 \cdots d\epsilon^{(r)}_{k-1} d\epsilon^{(r)}_{k+1} \cdots \epsilon^{(r)}_K \\
&= \pi(x)\pi(y) \sum_{\left\{\substack{b_{k',l'} \in \{-1,+1\} \\ (k,l)\neq(k',l') \\ b^{(r)}_{k',l'} \in \{-1,+1\} \\ (k,l)\neq(k',l')}\right\}} \int_{\left\{\substack{\epsilon_k>0, \\ k\neq k' \\ \epsilon^{(r)}_k>0 \\ k\neq k'}\right\}} min\left\{\frac{1}{\pi(y) + \sum_{(k',l')\neq(k,l)} \pi(x + b_{k'l'}\epsilon_{k'})}, \right. \\
&\left. \frac{1}{\pi(x) + \sum_{(k',l')\neq(k,l)} \pi(y + b^{(r)}_{k'l'}\epsilon^{(r)}_{k'})}\right\} d\epsilon_1 \cdots d\epsilon_{k-1} d\epsilon_{k+1} \cdots \epsilon_K d\epsilon^{(r)}_1 \cdots d\epsilon^{(r)}_{k-1} d\epsilon^{(r)}_{k+1} \cdots \epsilon^{(r)}_K
\end{aligned}$$

Note that this function is symmetric is symmetric in x and y as the distribution of b and $b^{(r)}$ are iid and also $\epsilon_l, l = 1, 2, \cdots, k-1, k+1, \cdots, K$ and $\epsilon^{(r)}_l, l = 1, 2, \cdots, k-1, k+1, \cdots, K$ are all independent and identically distributed. Also in the $\sum_{(k',l')\neq(k,l)}$, there are also terms that contain k and $l \neq l'$, and ϵ_k is fixed.

$$\epsilon_k = |y_d - x_d| \qquad d = 1, 2, \cdots, D$$

However ϵ_k occurs $L-1$ times in both the sums in the above expression and $b_{kl'}\epsilon_k$ has the same distribution as $b^{(r)}_{kl'}\epsilon_k$. This implies that **Eqn 1** holds true and the detailed balance holds for Version 1 of the Multiple Try TMCMC algorithm.

For Algorithm 2, the approach is primarily similar, and we can write

$$
\begin{aligned}
\pi(x)\mathcal{Q}(y|x,k,l) &= \pi(x) \sum_{\left\{\substack{b_{k',l'}\in\{-1,+1\} \\ (k,l)\neq(k',l') \\ b^{(r)}_{k',l'}\in\{-1,+1\} \\ (k,l)\neq(k',l')}\right\}} \int_{\left\{\substack{\epsilon_k>0, \\ k\neq k'}\right\}} \frac{\pi(y)}{\pi(y)+\sum_{(k',l')\neq(k,l)}\pi(y_{k',l'})} \\
&\times min\left\{1, \frac{\pi(y)+\sum_{(k',l')\neq(k,l)}\pi(y_{k',l'})}{\pi(x)+\sum_{(k',l')\neq(k,l)}\pi(x_{*,k',l'})}\right\} d\epsilon_1\cdots d\epsilon_{k-1}d\epsilon_{k+1}\cdots\epsilon_K \\
&= \pi(x)\pi(y) \sum_{\left\{\substack{b_{k',l'}\in\{-1,+1\} \\ (k,l)\neq(k',l') \\ b^{(r)}_{k',l'}\in\{-1,+1\} \\ (k,l)\neq(k',l')}\right\}} \int_{\left\{\substack{\epsilon_k>0, \\ k\neq k'}\right\}} min\left\{\frac{1}{\pi(y)+\sum_{(k',l')\neq(k,l)}\pi(x+b_{k'l'}\epsilon_{k'})},\right. \\
&\left.\frac{1}{\pi(x)+\sum_{(k',l')\neq(k,l)}\pi(y+b^{(r)}_{k'l'}\epsilon_{k'})}\right\} d\epsilon_1\cdots d\epsilon_{k-1}d\epsilon_{k+1}\cdots\epsilon_K
\end{aligned}
$$

The major modification in Algorithm 2 compared to Algorithm 1 is that the reverse step-sizes $\epsilon^{(r)}$ are same as forward step sizes ϵ. By the similar logic as above, the expression is symmetric in x and y and therefore, detailed balance holds.

MT-TMCMC chains are both λ-irreducible and aperiodic. Proof is similar to the one for standard TMCMC approach in **?**, since ultimately the steps are along a similar manifold as in standard TMCMC and that is what drives λ irreducibility and aperiodicity.