# Closure to “Transitional Markov Chain Monte Carlo: Observations and Improvements” by Wolfgang Betz, Iason Papaioannou, and Daniel Straub

Wolfgang Betz*, Iason Papaioannou, Daniel Straub

*Engineering Risk Analysis Group, Technische Universität München, 80333 München, Germany*

The authors appreciate the discussers’ interest in the paper and are thankful for the opportunity to clarify the role of the proposed iTMCMC method. As *Ching & Wang* the discussers correctly show, iTMCMC is not the most efficient TMCMC variant in all cases. However, iTMCMC exhibits a good performance for a larger class of problems than the original TMCMC method.

iTMCMC proposes an adaptive choice of the scaling factor $\beta$ in contrast to the use of a fixed $\beta$ suggested in [1]. If the sought posterior distribution has certain known properties, it is possible to theoretically derive a fixed optimal scaling factor. In particular, if the target distribution has iid components and obeys certain regularity conditions, the optimal scaling factor is $2.38/\sqrt{M \cdot F}$, where $M$ is the dimension and $F$ is a Fisher’s information measure of the component target density with $F = 1$ for the Gaussian case [2, 3]. The corresponding optimal acceptance rate of the candidate sample is 0.23 for sufficiently large $M$. For such situations, $\beta$ should be directly selected as the optimal value. However, the crux is that in most real applications, these conditions are not met and the optimal $\beta$ cannot be found directly. The discussers deal with this situation by postulating a fixed value of $\beta$ that works well in a variety of cases, whereas iTMCMC determines a near-optimal $\beta$ case-specifically in an adaptive manner.

In the examples investigated by the discussers, the TMCMC-1 variant with a fixed $\beta = 0.5$ appears to work well. In cases where the postulated $\beta$ is near-optimal, TMCMC-1 will inevitably outperform iTMCMC. For the comfort of not having to pre-specify a reasonable $\beta$ in iTMCMC, one has to pay with a decrease in efficiency compared to working with the optimal $\beta$ right from the start. However, as the discussers note: The optimal choice for $\beta$ in TMCMC (or TMCMC-1) remains a challenge and one cannot be sure that a $\beta$ of e.g. 0.5 will be near-optimal for the problem at hand. This is exemplified in the following multi-dimensional problem: The symmetric prior distribution is centered around the origin. The concentric likelihood has the shape $\varphi\left(\frac{\|\theta\|-d}{\epsilon}\right)$, where $\varphi(\cdot)$ denotes the PDF of the standard normal distribution, $d$ is a distance and $\epsilon \ll \mathrm{Var}[\theta]$. For the example case presented in Table 1, TMCMC-1 with a fixed $\beta = 0.2$ performs better than TMCMC-1 with $\beta = 0.5$, whereas iTMCMC outperforms both variants. As an optimal selection of $\beta$ remains a challenge, the authors regard iTMCMC as a more robust approach.

The discussers showed that for the investigated example problems, the performance of implementing iTMCMC in the underlying independent standard normal space is similar to working directly in the original random variable space. However, the authors believe that this will not necessarily be true for general problems. Working in the underlying standard normal space, the transformed target distribution is more likely to obey the conditions for which the suggested target acceptance rate $t_{\mathrm{acr}} = 0.21/M + 0.23$ is the optimal one [3]. This is especially the case in the initial sampling levels of iTMCMC where the exponent of the likelihood is small and hence the transformed target density is closer to the independent standard normal prior.

## iTMCMC algorithm

From discussions with *Ching & Wang* and other readers of the paper, the authors realized that the implementation of the *proposed modification (3)* was not clearly described in the original paper. To eliminate this ambiguity the authors summarize the implementation of iTMCMC in the following:

Initially, set $j = 0$ and $\beta = 2.4/\sqrt{M}$, where $M$ denotes the dimension of vector $\theta$. Furthermore, set $t_{\mathrm{acr}} = 0.21/M + 0.23$ and $N_{\mathrm{a}} = 100$, where $N_{\mathrm{a}}$ denotes the number of MCMC steps after which the value of $\beta$ is modified Additionally, set $N_{\mathrm{adapt}} = 1$.

For the initial $j = 0$, all $N_{\mathrm{s}}$ samples $\mathbf{u}_{(0,1)}, \ldots, \mathbf{u}_{(0,N_{\mathrm{s}})}$ are drawn from the $M$-dimensional independent standard normal distribution. For $k = 1, \ldots, N_{\mathrm{s}}$, the standard normal samples $\mathbf{u}_{(0,k)}$ are transformed to samples $\theta_{(0,k)}$ in the original parameter space, using e.g. the *Nataf* transformation or the *Rosenblatt* transformation. Note that the samples $\mathbf{u}_{(0,k)}$ follow the prior distribution. Thereafter, $j$ is set to *one*.

For all $j > 0$, the following scheme is applied:

1. Find $q_j$ through solving the minimization problem

$$q_{j+1} = \arg\min_{q} \left( \left| \mathrm{CV}_j(q) - \nu_{\mathrm{t}} \right| \right) \tag{1}$$

where $\mathrm{CV}_j(q)$ with $q \in (q_j, 1]$ is the sample coefficient of variation of the current samples, and $\nu_{\mathrm{t}}$ is the target coeffi-

*Corresponding author

*Email addresses:* wolfgang.betz@tum.de (Wolfgang Betz), iason.papaioannou@tum.de (Iason Papaioannou), straub@tum.de (Daniel Straub)

*Manuscript, published in J. Eng. Mech., 2017, 143(9): 07017002* *DOI: 10.1061/(ASCE)EM.1943-7889.0001066*

cient of variation (typically set to $v_t = 100\%$). If $q_j > 1$, then set $q_j = 1$.

2. For all samples $k = 1, \ldots, N_s$ compute a weighting coefficient $w_{(j,k)}$:

$$w_{(j,k)} = \left(L(\theta_{(j-1,k)}|\mathbf{d})\right)^{q_j - q_{j-1}} \tag{2}$$

3. Compute the mean of the weighting coefficients:

$$S_j = \frac{1}{N_s} \sum_{k=1}^{N_s} w_{(j,k)} \tag{3}$$

4. Compute the sample covariance matrix of the Gaussian proposal distribution:

$$\mathbf{\Sigma}_{\text{sample},j} = \sum_{k=1}^{N_s} \left[ \frac{w_{(j,k)}}{S_j \cdot N_s} \cdot \left(\mathbf{u}_{(j-1,k)} - \bar{\mathbf{u}}_j\right) \cdot \left(\mathbf{u}_{(j-1,k)} - \bar{\mathbf{u}}_j\right)^{\mathrm{T}} \right] \tag{4}$$

with

$$\bar{\mathbf{u}}_j = \frac{\sum_{l=1}^{N_s} w_{(j,l)} \cdot \mathbf{u}_{(j-1,l)}}{\sum_{l=1}^{N_s} w_{(j,l)}} \tag{5}$$

5. For $k = 1, \ldots, N_s$, set: $\mathbf{u}^{\mathrm{c}}_{(j,k)} = \mathbf{u}_{(j-1,k)}$ and $\theta^{\mathrm{c}}_{(j,k)} = \theta_{(j-1,k)}$.
6. Set $n_a = 0$ and acr $= 0$.
7. For $k = 1, \ldots, N_s$ do:
   (a) Select index $l$ from the set $\{1, \ldots, N_s\}$ at random, where each index $i \in \{1, \ldots, N_s\}$ of the set is assigned probability
   $$\frac{w_{(j,i)}}{\sum_{n=1}^{N_s} w_{(j,n)}} \tag{6}$$
   (b) Propose a new sample: draw $\mathbf{u}^*$ from a normal distribution that is centered at $\mathbf{u}^{\mathrm{c}}_{(j,l)}$ and has covariance matrix $\mathbf{\Sigma}_j = \beta^2 \cdot \mathbf{\Sigma}_{\text{sample},j}$.
   (c) Transform the sample $\mathbf{u}^*$ to sample $\theta^*$ in original parameter space.
   (d) Generate a sample $r$ from a uniform distribution on $[0, 1]$.
   (e) If $r \leq \frac{p_j(\theta^*)}{p_j(\theta^{\mathrm{c}}_{(j,l)})}$ then set $\mathbf{u}^{\mathrm{c}}_{(j,l)} = \mathbf{u}^*$, $\theta^{\mathrm{c}}_{(j,l)} = \theta^*$ and acr $=$ acr $+ 1$, otherwise do nothing.
   (f) Set $\mathbf{u}_{(j,k)} = \mathbf{u}^{\mathrm{c}}_{(j,l)}$ and $\theta_{(j,k)} = \theta^{\mathrm{c}}_{(j,l)}$.
   (g) Set $w_{(j,l)} = \left(L(\theta^{\mathrm{c}}_{(j,l)}|\mathbf{d})\right)^{q_j - q_{j-1}}$
   (h) Increase $n_a$ by *one*; i.e., $n_a = n_a + 1$.
   (i) If $n_a \geq N_a$, then
      - Compute the average acceptance-rate $p_{\text{acr}}$ of the last $N_a$ MCMC steps; i.e., $p_{\text{acr}} = \text{acr}/N_a$
      - Evaluate coefficient $c_a = (p_{\text{acr}} - t_{\text{acr}}) / \sqrt{N_{\text{adapt}}}$.
      - Update the value of $\beta = \beta \cdot \exp(c_a)$.
      - Set $n_a = 0$, $N_{\text{adapt}} = N_{\text{adapt}} + 1$ and acr $= 0$.
8. If $q_j = 1$ then stop the iteration, otherwise set $j = j + 1$ and continue with 1.

A Matlab and C++ version of this algorithm is available from www.era.bgu.tum.de/software.

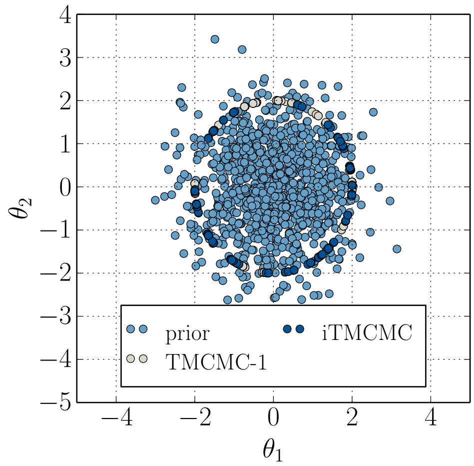


Figure 1: $10^3$ samples of the prior and the posterior. The posterior samples are generated with iTMCMC and TMCMC-1 ($\beta = 0.5$). The example problem is presented in Table 1.

### Reference solution for EX4

In addition to the original examples, the discussers investigated the performance of iTMCMC for a problem where the likelihood is based on a one-dimensional stationary normal random field. This is an informative and easily implementable example. Based on the measurement data used by the discussers, the reference solution for this problem is listed in Table 2. The reference solution is provided here, because it is required for assessment of the performance of the methods and is not given in the discussion.

### Erratum to the original paper

In the typesetting process, the expressions *prior belief*, *prior distribution*, and *prior PDF* where transformed to *previous belief*, *previous distribution*, and *previous PDF* – which are misnomers. All expression of *previous* (expect for two occurrences) should be replaced with *prior*.

### References


[1] Ching, J., Chen, Y.-C., 2007. Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging. Journal of Engineering Mechanics 133 (7), 816–832.

[2] Gelman, A., Roberts, G., Gilks, W., 1996. Efficient metropolis jumping rules. Bayesian statistics 5 (599-608), 42.

[3] Roberts, G. O., Gelman, A., Gilks, W. R., et al., 1997. Weak convergence and optimal scaling of random walk metropolis algorithms. The annals of applied probability 7 (1), 110–120.

Table 1: Example problem that has a non-linear covariance structure: The prior distribution is a two-dimensional independent standard normal distribution with mean *zero* and standard deviation *one*; i.e. $p(\theta_1, \theta_2) = \varphi(\theta_1) \cdot \varphi(\theta_2)$, where $\varphi(\cdot)$ denotes the PDF of the standard normal distribution. The likelihood is concentric and defined as $L(\theta_1, \theta_2) = \varphi\left(\left(\sqrt{\theta_1^2 + \theta_2^2} - 4\right)/\epsilon\right)$, with $\epsilon = 10^{-3}$. The reference solution can be computed through numerical integration as: $c_{\mathtt{E}} = 0.27067$, $\mathrm{E}[\theta_1] = 0$ and $\sigma[\theta_1] = 1.41421$.
The results were obtained by repeatedly ($8 \cdot 10^6$ times) generating $10^3$ samples by means of the specified TMCMC variant. In this example, iTMCMC performs similar to TMCMC-1 with $\beta = 0.2$. Samples of the prior and the posterior distribution are illustrated in Fig. 1.

| statistical quantity | iTMCMC | TMCMC-1 ($\beta = 0.5$) | TMCMC-1 ($\beta = 0.2$) | TMCMC ($\beta = 0.5$) | TMCMC ($\beta = 0.2$) |
|---|---|---|---|---|---|
| $\mathrm{bias}_{c_{\mathtt{E}}}$ | 0.01 | $3 \cdot 10^{-3}$ | $2 \cdot 10^{-3}$ | 0.14 | 0.17 |
| $\kappa_{c_{\mathtt{E}}}$ | 0.42 | 0.55 | 0.36 | 0.57 | 0.41 |
| $N_{\mathrm{eff}}$ | 8.2 | 4.1 | 8.0 | 4.0 | 7.6 |
| $\mathrm{E}[a_g]$ | $2 \cdot 10^{-4}$ | $2 \cdot 10^{-4}$ | $7 \cdot 10^{-5}$ | $2 \cdot 10^{-4}$ | $1 \cdot 10^{-4}$ |
| $\mathrm{bias}_{s_g}$ | $3 \cdot 10^{-5}$ | $8 \cdot 10^{-5}$ | $5 \cdot 10^{-5}$ | $2 \cdot 10^{-5}$ | $8 \cdot 10^{-5}$ |

Table 2: Reference solution for EX4 with the measurement data used by the discussers. The solution was obtained by means of rejecton sampling and $N_{\mathrm{samples}}$ statistically independent posterior samples.

| statistical quantity | EX4-1 | EX4-2 | EX4-3 |
|---|---|---|---|
| $c_{\mathtt{E}}$ | $7.161 \cdot \exp(-260)$ | $2.150 \cdot \exp(-745)$ | $3.148 \cdot \exp(-1845)$ |
| $\mathrm{E}[\theta_1]$ | 1051.2 | 952.26 | 991.29 |
| $\sigma[\theta_1]$ | 93.9 | 86.5 | 34.2 |
| $\mathrm{E}[\theta_2]$ | 4.997 | 5.296 | 5.220 |
| $\sigma[\theta_2]$ | 0.323 | 0.244 | 0.101 |
| $\mathrm{E}[\theta_3]$ | −1.55 | −0.540 | −0.725 |
| $\sigma[\theta_3]$ | 0.770 | 0.521 | 0.223 |
| $N_{\mathrm{samples}}$ | $2.2 \cdot 10^6$ | $1.5 \cdot 10^5$ | $1.5 \cdot 10^5$ |