Rejoinder

Daniel Schmidl *, Claudia Czado †, Sabine Hug ‡, and Fabian J. Theis §

First of all, we would like to thank both D. Woodard and M. Girolami & A. Mira for their excellent and detailed comments on our paper. Their remarks and questions have given us quite a number of new ideas for improving our sampling procedure. As suggested by both, we have conducted new sampling runs for two additional examples, which illustrate the usefulness of CIMH and ACIMH and answer some of the questions brought forward. Since Woodard's comments focus on one particular aspect of CIMH/ACIMH, while Girolami & Mira point out several different considerations, we will first reply to Woodard's comments.

1 Rejoinder on Woodard's comments

While our paper focuses primarily on a practical evaluation of the proposed new sampling schemes, Woodard's comments provide a nice theoretical addition. As pointed out in Cowles and Carlin (1996), both approaches complement each other and are equally important. Theoretical convergence results are often not easy or straightforward to assess, so we thank D. Woodard for the effort to improve the theoretical understanding of our algorithm.

Woodard shows that the efficiency should degrade exponentially in the dimension d of the target distribution. It is true that our algorithm seems to suffer from a decaying performance in higher-dimensional systems, as many other popular MCMC algorithms like the independence sampler or the Metropolis-Hastings algorithm do, cf. Hairer et al. (2011). However, we have already shown that the algorithm still performed well in two medium-sized systems of 12 and 15 dimensions, see Schmidl et al. (2012).

To see if we can assess the dimensionality issue more thoroughly also in practice, we chose a simple and straightforward example, which we sampled in various dimensions from 2 to 50. Concretely, we chose to sample from a strongly correlated normal

*Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München and Institute of Mathematical Sciences, Technische Universität München dschmidl@ma.tum.de

†Institute of Mathematical Sciences, Technische Universität München cczado@ma.tum.de

‡Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München and Institute of Mathematical Sciences, Technische Universität München sabine.hug@helmholtz-muenchen.de

§Institute of Bioinformatics and Systems Biology, Helmholtz Zentrum München and Institute of Mathematical Sciences, Technische Universität München
fabian.theis@helmholtz-muenchen.de

distribution with mean zero and the tridiagonal, positive definite covariance matrix:

$$\Sigma_d = \begin{pmatrix} 1 & 0.49 & 0 & \cdots & \cdots & 0 \\ 0.49 & 1 & 0.49 & 0 & \cdots & 0 \\ 0 & 0.49 & 1 & 0.49 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 0.49 & 1 & 0.49 \\ 0 & \cdots & 0 & 0 & 0.49 & 1 \end{pmatrix}. \tag{1}$$

The example was run 5 times for 50,000 MCMC iterations for each examined dimension d. The same set-up was applied for prerun sampling. The copula update parameters for ACIMH were set to R = 10,000 and S = 4. Furthermore the copulas were fitted on 1,000 equally spaced prerun samples. The time for the prerun was added to the sampling times of IM, RWMH, CIMH and ACIMH. Canonically, the cdf's of a normal distribution were used to transform the prerun samples to $[0,1]^d$. The independence proposal density q_3 was taken to be a uniform distribution. Furthermore, we set $r_1 = 0.99$ and $r_2 = 0$. All samplers were started at the origin. While CIMH and ACIMH approximated the d-dimensional normal distributions very well, especially the AM algorithm ran into severe problems. Even for $d = 20$, its acceptance rate was $< 0.1\%$ for all five runs. This is because we did not readjust the parameter that controls the initial step size of AM. We chose instead to leave all algorithms as they were adjusted for the two-dimensional example in the main paper, to check if the algorithms scale well. Consequently, we show only results for CIMH and ACIMH. However, we also feel that CIMH and ACIMH would definitely benefit from having more prerun samples as the dimension of the problem gets higher. As it was, CIMH and ACIMH showed a linear decay of acceptance rates with dimension, while the performance indices (I_1) and (I_2) showed an exponentially decaying behavior, see Figure 1. This confirms Woodard's argument that the efficiency of the algorithms should degrade exponentially in the dimension d.

To take up the suggestion in Woodard's conclusion, we think that a block update sampling algorithm is a very good idea in higher-dimensional systems. This was also pointed out by Roberts and Sahu (1997) and Wilkinson and Yeung (2002). Inference in dynamical systems with over one hundred parameters is challenging, but possible, see Hug et al. (2013). Parameters from dynamical systems often come with a natural "hierarchy": kinetic parameters, parameters which correspond to initial conditions, noise parameters and the so-called nuisance parameters like scaling and offset parameters. This hierarchy could be used for designing block updates. Another intuitive variant for finding blocks of parameters would be to choose a kinetic parameter and add to that update block all parameters that are directly linked with this parameter, i.e. the corresponding scaling, offset, initial condition and noise parameters of the players in the ODE system that are directly influenced by the kinetic rate at question, and maybe also "neighboring" kinetic parameters. Many variations are possible, but we feel that the idea is very promising and we are eager to try it out in future applications.
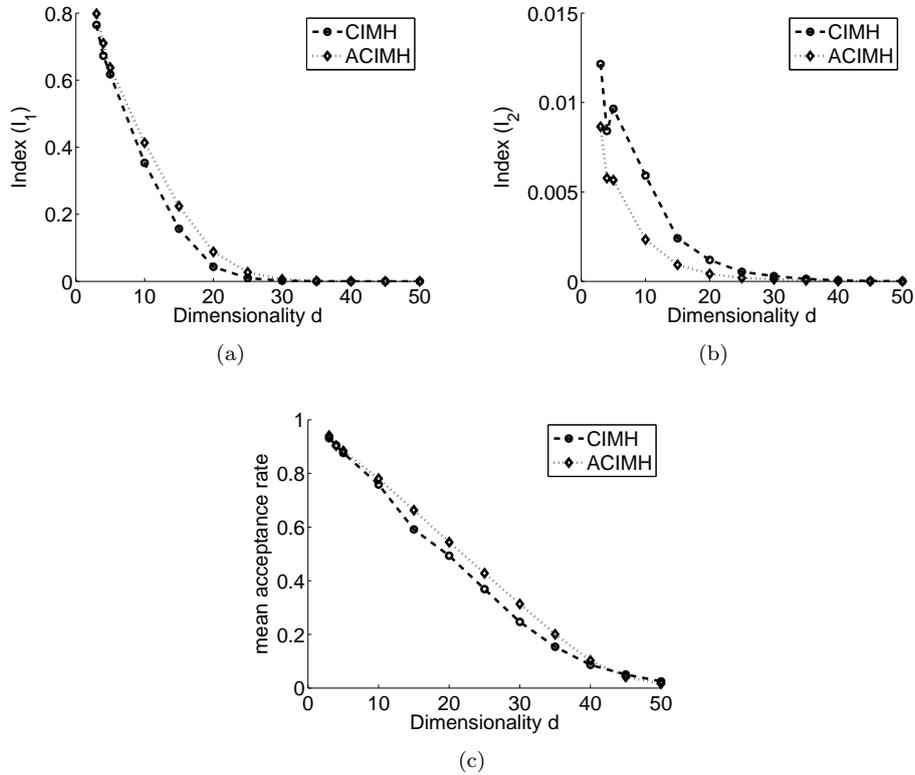
Figure 1: Results for the normal distribution in various dimensionalities. Figure (a): Quotient of acceptance rate and INEFF, (I_1), plotted over different dimensionalities. Figure (b): Effective Sampling Size per second, (I_2), plotted over different dimensionalities. Figure (c): Acceptance rate, plotted over different dimensionalities. A decay of sampling efficiency with rising dimensionality is clearly visible. Shown are results for CIMH and ACIMH.

2   Rejoinder on Girolami and Mira's comments

To our great pleasure, Girolami & Mira pointed out that inference in nonlinear deterministic dynamic systems is amongst the most challenging problems that currently feature prominently in the literature of Bayesian inference. We believe that this is rightly so since inference in these systems, while highly non-trivial, still is a very promising way of discovering new biological insights. We will now address each of the points mentioned in section 2.1 of the comment by Girolami & Mira.

- **Fitting of the marginals** The first point raised by Girolami & Mira is the fitting of the marginals, for which the prerun Markov chain is required. We refer the reader to section 5.4 of the original paper, where we showcased the deteriorating

effect a wrongly chosen marginal distribution can have. The ESS per second dropped by a factor of 2.9, if we specify exponential distributions as marginal types instead of the correct normal distributions. Furthermore, we would like to suggest the following guidelines for the prerun phase to ensure adequate performance of the algorithm:

– We strongly recommend applying standard MCMC convergence statistics to the prerun, such as the Geweke test (see e.g. Cowles and Carlin (1996)). Also visual inspection of the chains by an expert is often very helpful. We recommend that the prerun chain should not be in a visible burn-in phase.

– Like the number of burn-in samples to be discarded in standard MCMC algorithms, the required number of prerun samples strongly depends on the posterior to be sampled and can hardly be determined *a priori*. We recommend adding more prerun samples if the performance of the algorithm is unsatisfactory.

- **Posterior deterministic approximation** We very much like the idea of replacing the MCMC prerun with a posterior deterministic approximation. We could imagine, e.g., approximating the posterior distribution with a variational approximation and fitting the copula to that. This would certainly save computational time. However, if the dependency structure is very rich or non-standard, the results might be poor compared to CIMH with an MCMC prerun.

- **Choice of copula** Concerning the choice of the functional form of the copula, the structure of the vine and their ramifications for the performance of the algorithm, we can offer the following insights. The vine structure for the chosen D-vine was selected using the procedure described in Dißmann et al. (2013). There a general sequential tree by tree procedure was developed for the regular (R) vine case. It is driven by modeling strong dependencies early in the vine structure using a maximum spanning tree algorithm to reduce the influence of rounding errors in later trees. As weights the absolute value of corresponding pairwise Kendall's tau values are used. For the subclass of D-vines this approach requires finding a Hamiltonian path and this has been implemented in the R package VineCopula() within the function RVineStructureSelect(). A small simulation study in Dißmann et al. (2013) shows a reasonable performance with regard to model fit. The general problem of model selection in R-vines is a current research area and the results obtained so far are contained in Czado et al. (2013). We expect that after choosing a good fitting R vine model the implications for the rapidity of convergence and quality of the MCMC algorithm to be minor, but this will have to be investigated in more detail in the future.

- **Sampling from a banana-shaped distribution** We gladly take up the suggestion of trying out our copula-based algorithms on a "banana-shaped" distribution as shown in Figure 2(a). This shape of distribution for example appears as the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$ of the parameters $\boldsymbol{\theta} = (\theta_1, \theta_2)$ based on data $\mathbf{y}$ for

the following model:

$$\mathbf{y}|\boldsymbol{\theta} \sim \mathcal{N}\left(\theta_1 + \theta_2^2, \sigma_y^2\right) \tag{2}$$

$$\theta_i \sim \mathcal{N}\left(0, \sigma_\theta^2\right), i = 1, 2. \tag{3}$$

As in Lan et al. (2012), we generate 100 data points with $\theta_1 + \theta_2^2 = 1$, $\sigma_y = 2$ and set $\sigma_\theta = 1$. As in the main paper, the example was run 100 times for 50,000 MCMC iterations. The same set-up was applied for prerun sampling. The copula update parameters for ACIMH were set to R = 10,000 and S = 4. Furthermore the copulas were fitted on 1,000 equally spaced prerun samples. The time for the prerun was added to the sampling times of IMH, RWMH, CIMH, and ACIMH. Canonically, the cdf's of the prior distribution were used to transform the prerun samples to $[0, 1]^2$. The independence proposal density q_3 was taken to be a uniform distribution. Furthermore, we set $r_1 = 0.99$ and $r_2 = 0$. All samplers were started at the origin and approximated the two-dimensional distribution very well. We found that our sampling algorithm performed very well on this demanding problem. More specifically, CIMH and ACIMH performed nicely when speaking of acceptance rate, CIMH yielding $27.00 \pm 0.07\%$, while ACIMH performs slightly better at $27.09 \pm 0.06\%$. This led to effective sample sizes of 3138 ± 125 for CIMH and 2897 ± 109 for ACIMH. Also the other algorithms performed quite well and we especially have to point out that AM performed very well with an acceptance rate of $22.56 \pm 0.09\%$ and an effective sample size of 1030 ± 15. CIMH and ACIMH by far outperformed the other three algorithms with respect to (I_1) (Figure 2(b)). For (I_2) the situation is a bit different, here AM performed best with a mean ESS per second of 95 samples per second, while the mean ESS per second for CIMH was entirely comparable to that at 86 samples per second (Figure 2(c)). ACIMH performed worse than that with a mean ESS per second of 46 samples per second, since the extra time for the copula updates did not dramatically increase the quality of the fit, which seems to have been covered very well from the prerun already. We offer two reasons for the performance: first, we think that AM is particularly easy and fast to implement, giving it a computational advantage in such a low dimensional case, especially since it does not require any prerun. Second, as can be seen from Figure 2(d), the uniformization with normally distributed marginals might not be optimal. We suggest that applying for example a kernel density estimate to the prerun samples might increase performance of CIMH. This, however, is work in progress.

- **Initialization of the algorithm** Concerning the initialization of the algorithm, we would like to point out that the copula proposal function is independent of the current state of the chain, so we do not expect the initialization of the chain to have any influence on performance. However, if r_2 is rather large, i.e. if a lot of random walk samples are generated, we naturally expect the chain to remain in a transient phase much longer. The random walk portion is also the main reason for initializing the chain in the mode of the distribution in the first place.

- **Including information about the dynamic system** As rightly pointed out by Girolami & Mira, the vine-copula based algorithm is completely agnostic to
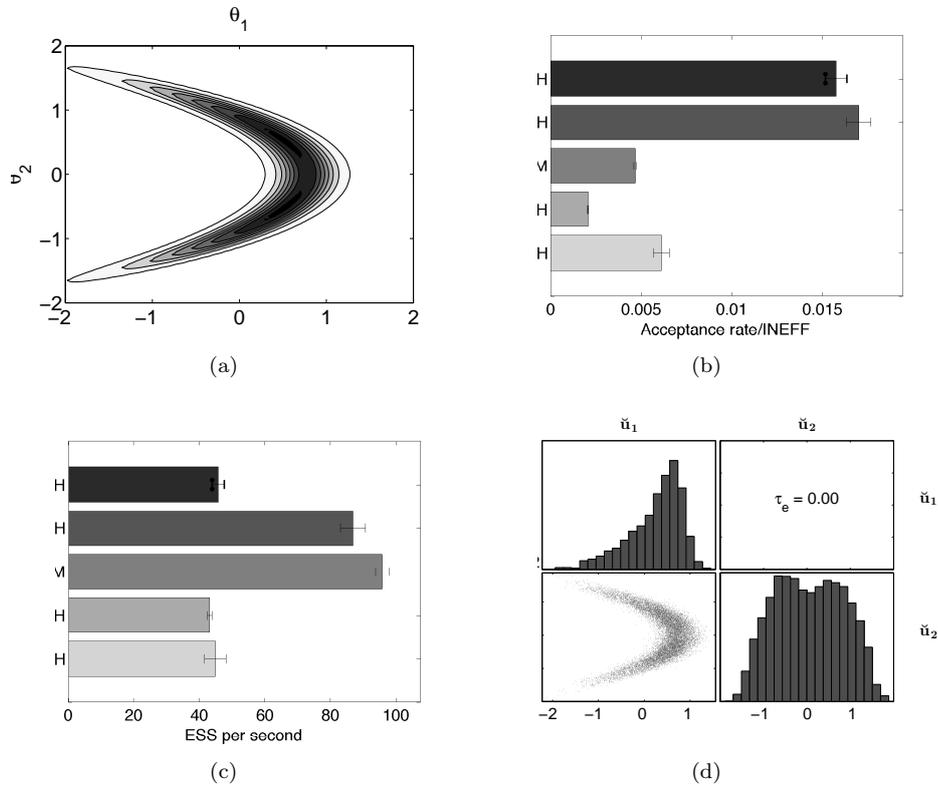
Figure 2: Results for the banana-shaped distribution. Figure (a): Posterior density surface for the banana-shaped distribution. Figure (b): Quotient of acceptance rate and INEFF, (I_1). Error bars show the estimated standard errors based on 100 runs. Figure (c): ESS per second, (I_2). Error bars show the estimated standard errors based on 100 runs. Figure (d): Marginal copula data used to fit the CIMH and ACIMH copula of the first run. The diagonal displays the histograms of the MCMC sample marginals and τ_e the corresponding empirical Kendall's τ. In the lower left panel, the banana-shaped dependency structure is clearly visible.

the underlying dynamic system. While we are convinced that this is worthwhile and sufficient in many cases, we nevertheless agree that including such information would most certainly improve the performance of the sampling algorithm. We also agree that pursuing this approach is very interesting. A straightforward way to do this might be to combine the copula-based approach with the Riemann manifold-based MCMC method MALA of Girolami and Calderhead (2011) instead of a regular random walk Metropolis-Hastings as a second proposal function q_2 for CIMH. We will pursue research in this direction and hope to combine the benefits of both approaches (see also the last item of this list).

- **Heavy-tailed proposal function** Being able to sample from and design a heavy-tailed independence proposal is indeed not always an easy task and requires some expertise. For our applications in systems biology, the parameter values can be truncated at some sensible bounds naturally. We feel that this usually should make it possible to find a suitable uniform distribution that is more heavy-tailed than the posterior. Otherwise, this certainly requires some expertise and is more of a general problem. Having a uniform distribution as a heavy-tailed proposal function worked fine in our examples and of course has the advantage that it is easy to sample from.

- **Performance in high dimensions** As for how the indices (I_1) and (I_2) perform in higher dimensions, we refer to our reply to Woodard's comments in Section 1, where we showed that the indices indeed decay exponentially as the dimensionality of the target distribution increases, see also Figure 1. However, since these results were obtained without readjusting the parameters of the algorithms, we feel that nevertheless a better performance in high dimensions than shown here can be achieved. Still, both indices should always be seen in the context of the dimensionality of the problem.

- **Appraisal of index (I_1)** We agree with Girolami & Mira that the performance index (I_1) is in some sense not optimal, as the result is not completely unbiased. This is mostly because the vine-copula based sampling algorithms do not directly regulate or tune the acceptance rate but strive for an acceptance rate of 100 %, whereas the step size of random walk algorithms like RWMH and AM are tuned to yield acceptance rates of approximately 23 %. We would like to suggest that it should be possible to replace the acceptance rate with a measure of how far the acceptance rate of the sampling algorithm is off from what is deemed optimal, thus achieving a normalization. However, we find that this topic is severely underrepresented in the literature and should be the focus of further research, since having sound criteria for comparisons can only be beneficial. Until then, we would like to suggest that for our contribution (I_1) should be taken more as a goodness-of-fit index for the copula-based samplers and as additional information, while the index (I_2), i.e. the ESS per second, is a more common and more easily interpretable index.

- **Combination with local structure MCMC methods** We think that the idea of combining local information structure MCMC as suggested by Girolami and Calderhead with our copula-based algorithm is very promising. The copula-based algorithm strives to approximate the global structure of the posterior by modeling the complete parameter dependencies. Girolami and Calderhead's algorithm contrastingly tries to capture the local posterior structure at the current precise point in the parameter space. A delayed rejection strategy should combine the best of both approaches and provide an efficient global exploration of parameter space at simultaneously high acceptance rates through the combination of delayed rejection with local exploration. Furthermore, delayed rejection and the local exploration should be especially beneficial for the adaptive copula-based sampling algorithm

ACIMH, since this approach should generate a lot of "new" information in the
form of accepted samples that can be used to improve the fitting of the pair-copula
decomposition. A combination of both the global and local methodologies in such
a fashion should definitely provide new insights and should be heavily pursued.

In closing, we thank the referees for their work on our paper and their fruitful
comments. Furthermore we would like to thank the editor-in-chief for suggesting our
paper for this great interactive format. This discussion has opened new avenues for us
and we sincerely hope that it will also broaden the understanding of our paper for the
readers.

References

Cowles, M. and Carlin, B. (1996). "Markov chain Monte Carlo convergence diagnostics:
a comparative review." *Journal of the American Statistical Association*, 91(434):
883–904. 33, 36

Czado, C., Brechmann, E., and Gruber, L. (2013). "Selection of Vine Copulas." In P. Ja-
worski, F. D. and Härdle, W. K. (eds.), *Copulae in Mathematical and Quantitative
Finance*. New York: Springer. To appear. 36

Dißmann, J., Brechmann, E., Czado, C., and Kurowicka, D. (2013). "Selecting and
estimating regular vine copulae and application to financial returns." *Computational
Statistics & Data Analysis*, 59: 52–69. 36

Girolami, M. and Calderhead, B. (2011). "Riemann Manifold Langevin and Hamiltonian
Monte Carlo methods." *Journal of the Royal Statistical Society - Series B*, 123–214.
38

Hairer, M., Stuart, A. M., and Vollmer, S. J. (2011). "Spectral Gaps for a Metropolis-
Hastings Algorithm in Infinite Dimensions." *arXiv preprint arXiv:1112.1392v1*, 1–38.
33

Hug, S., Raue, A., Hasenauer, J., Bachmann, J., Klingmüller, U., Timmer, J., and
Theis, F. J. (2013). "High-Dimensional Bayesian Parameter Estimation : Case Study
for a Model of JAK2 / STAT5 Signaling." *submitted to Mathematical Biosciences*.
34

Lan, S., Stathopoulos, V., Shahbaba, B., and Girolami, M. (2012). "Lagrangian Dy-
namical Monte Carlo." *arXiv preprint arXiv:1211.3759*. 37

Roberts, G. and Sahu, S. (1997). "Updating schemes, correlation structure, blocking
and parameterisation for the Gibbs sampler." *Journal of the Royal Statistical Society,
Series B*, 59(2): 291—317. 34

Schmidl, D., Hug, S., Li, W., Greiter, M. B., and Theis, F. J. (2012). "Bayesian model
selection validates a biokinetic model for zirconium processing in humans." *BMC
Systems Biology*, 6(95). 33

Wilkinson, D. and Yeung, S. (2002). "Conditional simulation from highly structured Gaussian systems, with application to blocking-MCMC for the Bayesian analysis of very large linear models." *Statistics and Computing*, 12(3): 287–300. 34