

## Comment on Article by Schmidl et al.

Mark Girolami \* and Antonietta Mira †

We thoroughly enjoyed reading this paper and are delighted to contribute to its discussion. The authors have been particularly innovative in drawing upon quite different areas of statistical inference to devise efficient Markov transition kernels for complex posterior distributions. Landmark papers provoke discussion and raise many open questions leading to potentially fruitful new avenues of investigation and this paper is no exception.

### 1 Statistical inference for dynamic systems

The authors have arguably chosen one of the most challenging problems for Bayesian inference: sampling from the induced posterior of the parameters of a deterministic nonlinear dynamic system. To illustrate the problem, consider the simple dynamic system<sup>1</sup> below

$$\frac{dx}{dt} = \frac{72}{36 + y} - k_3 \quad \frac{dy}{dt} = k_4 x - 1. \quad (1)$$

Setting initial conditions of  $x(0) = 7$ ,  $y(0) = -10$  and parameters to values of  $k_3 = 2$  and  $k_4 = 1$ , the system is forward simulated in time steps of 0.5 from  $t = 0$  to  $t = 60$ . An additive zero mean Gaussian observation error with covariance  $0.5\mathbf{I}_2$  then generates the 120 observations of the system dynamics  $\mathcal{D}$ . The induced posterior under flat priors  $p(k_3, k_4 | \mathcal{D})$  is shown in Figure 1. The multiple elongated ridges show strong nonlinear correlation separated by regions of low density. This simple low dimensional model presents challenges due to the near isolated ridges and their long correlated structure, making them challenging to traverse and move between.

Statistical inference for systems described by nonlinear ordinary differential equations is a relatively recent area of interest. The non-Bayesian tradition has had major contributions from the likes of Ramsay et al. (2007); Maiwald and Timmer (2008). Bayesian approaches have been spearheaded by, for example, Wilkinson (2009); Toni et al. (2009); Raue et al. (2013). Recent work by Gutenkunst et al. (2007) has described the posterior distributions of dynamic system parameters as *sloppy*, a catchy term describing the effects of the lack of identifiability and complexity of the underlying models. The authors of this paper under discussion therefore have made an important contribution to an area of research that is much in need of innovation and new insight.

---

\*Department of Statistical Science, University College London, [m.girolami@ucl.ac.uk](mailto:m.girolami@ucl.ac.uk)

†Institute of Finance, University of Lugano, Switzerland, [antonietta.mira@usi.ch](mailto:antonietta.mira@usi.ch)

<sup>1</sup>Mark Girolami is grateful to Simon Byrne, Anne-Marie Lyne and Ben Calderhead for producing the two figures in this contribution.

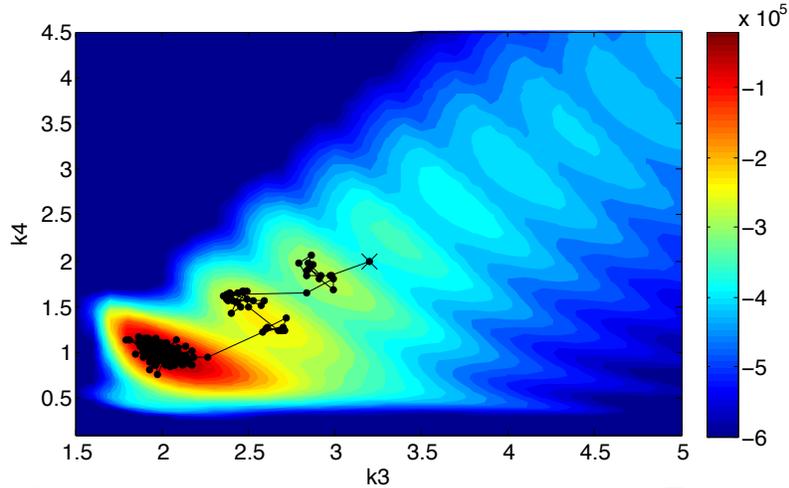


Figure 1: Posterior density for the two parameters,  $k_3$  and  $k_4$ . The multiple peaked ridges, with strong nonlinear correlation, induced by the occurrence of limit cycles in the dynamics, present a major problem for posterior simulation. The dots show a run of Parallel Tempering that is required to allow the chain to move between the ridges (limit cycles of dynamics).

## 2 Copula-based MCMC

The copula-based approach to MCMC is extremely interesting and may suggest links to other seemingly unrelated work such as the use of deterministic optimal maps as suggested by [El Moselhy and Marzouk \(2012\)](#). Once the copula is defined and parameterised, it seems that sampling will proceed quickly with little computational expense. The overhead associated with the characterisation of the copula is pushed back to a tuning prerun phase of the overall procedure. On the one hand this is an attractive feature of the methodology, on the other hand however it does raise a number of questions that it would be useful to hear the authors elaborate upon.

### 2.1 Questions

- The prerun Markov chain is required to define the marginals for the copula. One wonders how much effort must be expended to ensure that the marginals are estimated appropriately. If the target posterior is as complex as that discussed in the simple example above, then considerable effort will be required to ensure the chain mixes well enough to get reasonable estimates of the marginals. Have the authors any guidelines that can be provided regarding this prerun phase and the implications for the overall performance, e.g , the number of samples required to adequately identify the copula, how well should this be characterised to ensure good mixing of the chain?

- Rather than the overhead of a full blown MCMC prerun would the adoption of one of the many posterior deterministic approximations, e.g , Nested Laplace Approximations, Variational Approximations, etc , be of use?
- The choice of the functional form of the copula, the structure of the vine, its implications for the rapidity of convergence and quality of the MCMC mixing are not entirely clear. It would be interesting to hear from the authors how sensitive performance is to these choices.
- The tuning obtained by the copula is capturing the global characteristics of the multivariate target posterior. However it is clear that there is rich local micro-structure in the posteriors for deterministic dynamic systems. One then considers whether this is really optimal in the sense of efficiently exploring and covering the support of the posterior. An example of such a distribution was considered in the discussion of [Girolami and Calderhead \(2011\)](#), shown in [Figure 2](#) and it would be interesting to see how the vine-copula scheme performs on such distributions in the authors' response. We suspect this would be a much more informative example than the bivariate Gaussian example included in the paper and it would provide a valuable insight.
- We note that the simulations were initialised at the mode of the distribution which is perfectly sensible and in accordance with standard practice, again it would be interesting to hear how the vine-copula scheme performs in the transient phase of the chain and how quickly convergence occurs.
- The vine-copula method is agnostic to the underlying dynamic system save from what can be estimated via the MCMC prerun and subsequent adaption of the copula. For dynamic systems with limit cycles such as the simple oscillator described above, exploitation of the dynamics is almost a necessity to ensure efficiency of convergence, sampling, and subsequent Monte Carlo errors. It would be interesting to consider how, if at all possible, such information could be exploited in this copula-based MCMC scheme.
- The proof of convergence of CIMH (and consequently of ACIMH), relies on being able to design and sample from a heavy-tailed independence proposal ( $q_3$ , in the notation of the paper). For multi-dimensional and highly complex target distributions, this is not an easy task. For example, if the target has extreme value marginals and the dependence structure is itself defined by a non standard extreme value copula (here, by non standard, we mean hard to get i.i.d. samples from) what could be a suitable  $q_3$ ?
- We wonder how the performance indices ( $I_1$  and  $I_2$ ), scale as the dimension of the target increases in particular with reference to the first example of the paper.
- The appearance of the acceptance rate in the numerator of the performance index  $I_1$  may not provide an unbiased comparison across different MCMC methods. The RWMH and the AM algorithms are tuned to the "optimal" 23% acceptance probability whereas the vine-copula sampler naturally aims at an acceptance rate

of 100%, which is achieved if  $r_1 = 1$  and  $q_1$  is the target. Therefore an optimally tuned RWMH (or AM) will have a performance score of at most 0.23. This would seem to be a general issue for the comparison of MCMC methods (e.g. Hamiltonian Monte Carlo compared to Langevin algorithms) and some normalisation taking account of target acceptance rates seems to be required. We wonder if such a performance index could be devised to further evaluate the methods considered in the paper and if so how will this impact the conclusions one can make?

### 3 Local information structure MCMC

The direct exploitation of local structure of a distribution by way of the Fisher-Rao Metric Tensor and Levi-Civita Connection was suggested by [Girolami and Calderhead \(2011\)](#). In [Figure 2](#) the local metric tensor which induces a local covariance structure at a range of points for the banana shaped distribution is highlighted and the performance of such MCMC methods is discussed in [Girolami and Calderhead \(2011\)](#). For dynamic systems the Fisher-Rao metric tensor emerges as an inner product of the solution of the systems sensitivity equations. This explicitly captures the local posterior structure induced by the underlying dynamics at that precise point in the parameter space. This of course carries a computational overhead in that both the equations of dynamics and sensitivity have to be solved in producing the metric. If the Levi-Civita Connection is also exploited, then second-order sensitivities are required. In addition to the enhanced sampling performance achievable these gradients and metric structures can also be used directly in Monte Carlo estimate variance reduction using the Zero-Variance methodology ([Mira et al. 2012](#)). Inference over the parameters of the JAK - STAT5 cellular signalling example was addressed by such geometric MCMC methods in [Calderhead and Girolami \(2011\)](#) and it would be interesting to consider if a mixture transition kernel of the vine-copula and a metric tensor based kernel would be of use in these particularly challenging sampling problems. A clever way of “mixing” would be through the Delayed Rejection strategy first introduced in [Tierney and Mira \(1999\)](#) and further refined and extended in [Green and Mira \(2001\)](#): at the first stage a vine-Copula proposal could be used and only upon rejection the more computationally demanding gradients and metric local structures could be exploited by way of the Fisher-Rao Metric Tensor and Levi-Civita Connection. This way, global and local moves can be mixed efficiently, resorting to higher order local computation only upon demand.

In closing we are delighted to have been able to discuss this highly innovative and interesting paper and are of the opinion it will spark off other investigations in this exciting area of Bayesian inference in dynamic systems.

### References

- Calderhead, B. and Girolami, M. (2011). “Statistical analysis of nonlinear dynamical systems using differential geometric sampling methods.” *Interface Focus*, 1(6): 821–835. [30](#)

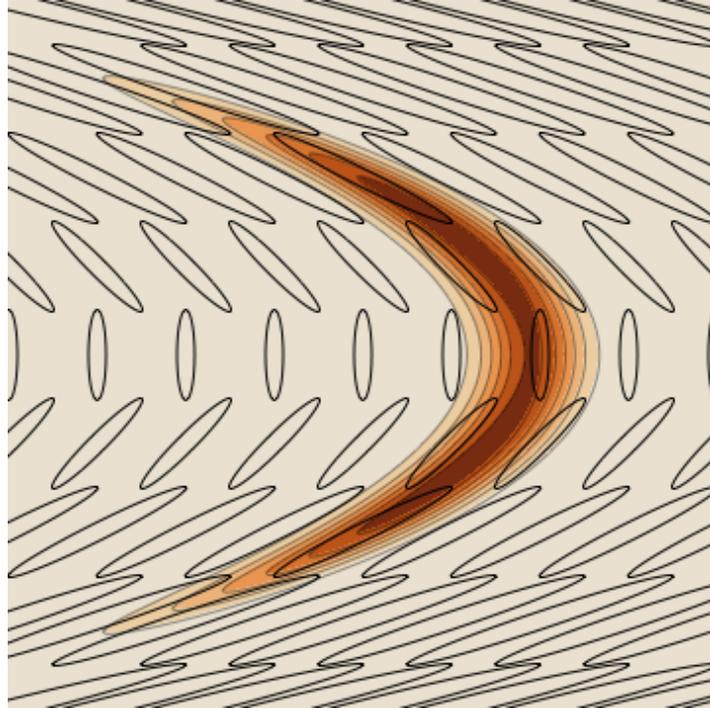


Figure 2: The banana shaped distribution as used in [Girolami and Calderhead \(2011\)](#). The metric tensor and hence the proposal covariance at each of the points is highlighted. It is clear that the local adaptivity comes naturally from the underlying geometric structure of the statistical model without the need for online estimation of a global proposal covariance.

- El Moselhy, T. A. and Marzouk, Y. M. (2012). “Bayesian inference with optimal maps.” *Journal of Computational Physics*, 231(23): 7815–7850. [28](#)
- Girolami, M. and Calderhead, B. (2011). “Riemann manifold Langevin and Hamiltonian Monte Carlo methods.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2): 123–214. [29](#), [30](#), [31](#)
- Green, P. J. and Mira, A. (2001). “Delayed rejection in reversible jump Metropolis-Hastings.” *Biometrika*, 88(4): 1035–1053. [30](#)
- Gutenkunst, R. N., Waterfall, J. J., Casey, F. P., Brown, K. S., Myers, C. R., and Sethna, J. P. (2007). “Universally Sloppy Parameter Sensitivities in Systems Biology Models.” *PLOS Computational Biology*, 3(10): 1871 – 1878. [27](#)
- Maiwald, T. and Timmer, J. (2008). “Dynamical modeling and multi-experiment fitting with PottersWheel.” *Bioinformatics*, 24(18): 2037–2043. [27](#)

- Mira, A., Solgi, R., and Imparato, D. (2012). “Zero Variance Markov Chain Monte Carlo for Bayesian Estimators.” *Statistics and Computing*, DOI 10.1007/s11222-012-9344-6. 30
- Ramsay, J. O., Hooker, G., Campbell, D., and Cao, J. (2007). “Parameter estimation for differential equations: a generalized smoothing approach.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(5): 741–796. 27
- Raue, A., Kreutz, C., Theis, F. J., and Timmer, J. (2013). “Joining Forces of Bayesian and Frequentist Methodology: A Study for Inference in the Presence of Non-Identifiability.” *Journal of the Philosophical Transactions of the Royal Society*, to appear, <http://arxiv.org/abs/1202.4605>. 27
- Tierney, L. and Mira, A. (1999). “Some adaptive Monte Carlo methods for Bayesian inference.” *Statistics in Medicine*, 18: 2507–2515. 30
- Toni, T., Welch, D., Strelkowa, N., Ipsen, A., and Stumpf, M. P. H. (2009). “Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems.” *Journal of The Royal Society Interface*, 6(31): 187–202. 27
- Wilkinson, D. J. (2009). “Stochastic modelling for quantitative description of heterogeneous biological systems.” *Nature Reviews: Genetics*, 10(2): 122–133. 27