WORKING PAPER SERIES

An Efficient Parallel Simulation Method for Posterior Inference on Paths of Markov Processes

Matthias Held/Marcel Omachel

Working Paper No. 10/2014

FACULTY OF ECONOMICS
AND MANAGEMENT

Impressum (§ 5 TMG)

Herausgeber:
Otto-von-Guericke-Universität Magdeburg
Fakultät für Wirtschaftswissenschaft
Der Dekan

Verantwortlich für diese Ausgabe:
Matthias Held and Marcel Omachel
Otto-von-Guericke-Universität Magdeburg
Fakultät für Wirtschaftswissenschaft
Postfach 4120
39016 Magdeburg
Germany

http://www.fww.ovgu.de/femm

Bezug über den Herausgeber
ISSN 1615-4274

An Efficient Parallel Simulation Method for Posterior Inference on Paths of Markov Processes

Matthias Held[†] **Marcel Omachel***

October 17, 2014

Abstract

In this note, we propose a method for efficient simulation of paths of latent Markovian state processes in a Markov Chain Monte Carlo setting. Our method harnesses available parallel computing power by breaking the sequential nature of commonly encountered state simulation routines. We offer a worked example that highlights the computational merits of our approach.

†OVGU Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany. *WHU – Otto Beisheim School of Management, Campus Vallendar, Burgplatz 2, 56179 Vallendar, Germany. Corresponding author: matthias.held@whu.edu, +492616509395.

Keywords: Bayesian inference, Markov Chain Monte Carlo, path simulation

In this note, we propose a novel method for efficient simulation of posterior densities of Markov processes that dominates the commonly employed sequential methods when parallel computing power is available.

To fix our idea, let $X \equiv \{X_t\}_{t=1\ldots T}$ be an unobservable N-dimensional state process, $X_t \in \mathbb{R}^N$ equipped with the Markov property, i.e. the transition probability $p(X_t \to X_{t+1})$ of X depends only on its current state. Further, let $Y \equiv \{Y_t\}_{t=1\ldots T}$ denote the set of noisy observations on $\mathbb{R}^K$ emitted by the state process, where the observation Y_t depends only on the current state, $Y_t = G(X_t, \omega_t)$, where G is a linear or non-linear function of X_t and some random (not necessarily Gaussian) noise ω_t. Without loss of generality, assume that the set of parameters Θ describing the state transition density and the observation densities are known with certainty, so that we will suppress them in our analysis.[1] Bayesian inference aims at estimating the posterior density of the latent state variable given a stream of observable information. The Markov property implies that the observation Y_t at time t is exclusively generated by X_t, the level of the states at time t. In a Bayesian framework, we are interested in the posterior density of X given all available information,

$$f(X|Y).$$

[1] Our method will usually be embedded in a larger MCMC framework, seeking inference not only on the latent states X but also on the full set of parameters Θ. As we are only considering state simulation here, we can keep Θ fixed and thus assume it to be known.

In an asset pricing context, the researcher might want to examine the level of the (latent) variance process given her model assumptions and observed option prices, or she might want to infer the level of risk neutral default intensities from observed prices of Credit Default Swaps. Usually, very large data sets are considered, and even if the state space is one-dimensional, the posterior density of X would be highly multidimensional rendering maximum likelihood inference hardly feasible. We must thus resort to simulation based methods. In a Markovian setup with linear state transition and observation equations, combined with Gaussian randomness, we can make use of the widely employed Kalman filter methods that incorporate the full set of observations such as the Forward Filter Backward Sampling (FFBS) algorithm of Früwirth-Schnatter (1994) and Carter and Kohn (1994). Here, a forward sweep of the Kalman filter estimates conditional means and covariances of X given the history of observations, $f(X_t|Y_{1\ldots t})$, whereas a backward sampler then creates a simulated path of X given all observed information, $f(X_{1\ldots T}|Y_{1\ldots T})$. The FFBS algorithm is thus able to directly – and swiftly – sample a random path of the latent process given posterior means and covariances. In cases where the transition equation or the observation equation are slightly nonlinear, extended Kalman filtering methods might be considered.

If the transition equation or the observation equation are highly nonlinear or very computationally intensive to evaluate, or if the process innovations are nonnormal, one must usually resort to MCMC methods such as the Metropolis Hastings (MH) sampling algorithm (Metropolis et al. (1953)). In a nutshell, the MH-algorithm draws proposals for the variables of interest from a distribution that allows for fast sampling and rejects or accepts these proposals with a certain probability. Let $g(x^*|x_m)$ denote the (simple) proposal density that samples x^* given its level obtained from previous simulation sweep, x_m. The algorithm accepts this proposal with probability:

$$\alpha = \min\left(\frac{f(x^*|Y)g(x_m|x^*)}{f(x_m|Y)g(x^*|x_m)}, 1\right)$$

If the sample is rejected by the MH-algorithm, set $x_{m+1} = x_m$, else set $x_{m+1} = x^*$. This algorithm assures that the samples x_m form a Markov chain which converges to the stationary distribution $f(X|Y)$ after a sufficient iteration count M_0 has been reached. All iterations from $M_0 + 1$ onward can be regarded as samples from the posterior density, and inference about the posterior distribution of X can then be drawn from them.

A further milestone in posterior analysis is the theorem of Clifford and Hammersley (Hammersley and Clifford (1971), Besag (1974)) who show that it is possible to break the highly multivariate joint posterior $f(X|Y)$ into the complete set of conditional distributions

$$\begin{gathered} f(X_1|X_{\sim 1}, Y) \\ f(X_2|X_{\sim 2}, Y) \\ \ldots \\ f(X_T|X_{\sim T}, Y), \end{gathered}$$

where $X_{\sim s}$ denotes the set of all latent states except for X_s, i.e. $X_{\sim s}$=$\{X_t\}_{t=1\ldots s-1,s+1\ldots T}$.

When we combine the Metropolis-Hastings algorithm with the theorem of Clifford and Hammersley, we can infer the posterior density $f(X|Y)$ by iteratively sampling the state in each point in time and holding all other states and observations constant. In the extreme case,

this means sampling each state variable at each point in time, which requires $N \times T$ MH-steps per sweep of the main MCMC algorithm.

We have so far only encountered the naive sequential fragmentation of the state space in the literature (Carlin et al. (1992), Kim et al. (1998), Kitagawa (1996), Eraker (2001)). Here, each single iteration of the MCMC algorithm consecutively simulates the state variable for each observation point in time, i.e. T consecutive MH-steps are considered. When the posterior density is highly complicated, this approach requires plenty of computation time and offers no leeway for parallelization as it is inherently sequential in nature.[2]

To remedy this weakness, we propose a different application of the Clifford Hammersley theorem which is particularly advantageous when applied to Markovian systems. The next section states the decomposition of the conditional posteriors $f(X_t|X_{\sim t},Y)$ and presents our method. Section 2 presents a worked example.

1 The method

Given the Markovian nature of our state and observation processes, the following lemma holds:

Lemma 1. *Let $X \equiv \{X_t\}_{t=1\ldots T}$ be an unobservable N-dimensional Markovian state process that emits the stream of K-dimensional observations $Y \equiv \{Y_t\}_{t=1\ldots T}$, where Y_t only depends on X_t. Then, for the posterior density of X_t, $t = 2\ldots T-1$, given all information Y and all other latent states $X_{\sim t}$ it holds*

$$f(X_t|X_{\sim t},Y) \propto f(Y_t|X_t)f(X_{t+1}|X_t)f(X_t|X_{t-1})$$

where $\propto$ denotes proportionality, $f(Y|X) = \frac{f(Y,X)}{f(X)} \propto f(Y,X)$.

Further, $f(X_1|X_{\sim 1},Y) \propto f(Y_1|X_1)f(X_2|X_1)$, and $f(X_T|X_{\sim T},Y) \propto f(Y_T|X_T)f(X_T|X_{T-1})$.

Proof. By repeatedly invoking Bayes' theorem and the Markovian, conditional independence nature of the state space, we obtain

$$\begin{aligned}
f(X_t|X_{\sim t},Y) &\propto f(Y_t|X_t,X_{\sim t},Y_{\sim t})f(X_t|X_{\sim t},Y_{\sim t}) \\
&\propto f(Y_t|X_t,X_{\sim t})f(X_t|X_{\sim t}) \\
&\propto f(Y_t|X_t)f(X_t|X_{\sim t}) \\
&= f(Y_t|X_t)f(X_t|X_{1\ldots t-1,t+1\ldots T}) \\
&\propto f(Y_t|X_t)f(X_{t+2\ldots T}|X_{1\ldots t+1})f(X_t|X_{1\ldots t-1},X_{t+1}) \\
&= f(Y_t|X_t)f(X_{t+2\ldots T}|X_{t+1})f(X_t|X_{1\ldots t-1},X_{t+1}) \\
&\propto f(Y_t|X_t)f(X_t|X_{1\ldots t-1},X_{t+1}) \\
&= f(Y_t|X_t)f(X_t|X_{t-1},X_{t+1}) \\
&\propto f(Y_t|X_t)f(X_{t+1}|X_{t-1},X_t)f(X_t|X_{t-1}) \\
&= f(Y_t|X_t)f(X_{t+1}|X_t)f(X_t|X_{t-1})
\end{aligned}$$

By analogous derivation, we obtain $f(X_1|X_{\sim 1},Y)$ and $f(X_T|X_{\sim T},Y)$. ■

[2] At $t = \tau$, simulate x_τ^m given $x_{1\ldots\tau-1}^m, x_{\tau+1\ldots T}^{m-1}$ and Y. Thus, this algorithm requires consecutive results of the MH-steps, prohibiting parallelization.

The standard route to sampling a path of X is to iteratively propose x_t^*, $t = 1 \dots T$, and reject or accept the proposals given the proposal and the posterior densities stated in lemma 1. As the lemma holds irrespective of the choice of *how* we apply the theorem of Clifford and Hammersley, we suggest the following sampling method. During the $(m+1)$th iteration sweep of the MCMC algorithm, do the following:

Sampling algorithm

Given the simulation $x^m \equiv \{x_t^m\}_{t=1\dots T}$ obtained from the current, mth iteration of the MCMC chain

1. Simultaneously, propose draws x_t^* for all *even* dates, $t = 2, 4, \dots, T$, from independent candidate densities $g_t(x_t^*|x_t^m)$.

2. Accept or reject each single proposal using the Metropolis Hastings algorithm and the conditional densities stated in lemma 1, with probability

$$\alpha_t = \min\left(\frac{f(Y_t|x_t^*)f(x_t^*|x_{t-1}^m)f(x_{t+1}^m|x_t^*)g(x_t^m|x_t^*)}{f(Y_t|x_t^m)f(x_t^m|x_{t-1}^m)f(x_{t+1}^m|x_t^m)g(x_t^*|x_t^m)}, 1\right)$$

   If the draw gets accepted, set $x_t^{m+1} = x_t^*$, else set $x_t^{m+1} = x_t^m$.

3. Simultaneously, propose draws x_t^* for all *odd* dates, $t = 1, 3, \dots, T-1$, from independent candidate densities $g_t(x_t^*|x_t^m)$

4. Accept or reject each single proposal using the Metropolis Hastings algorithm and the conditional densities stated in lemma 1 with probability

$$\alpha_t = \min\left(\frac{f(Y_t|x_t^*)f(x_t^*|x_{t-1}^{m+1})f(x_{t+1}^{m+1}|x_t^*)g(x_t^m|x_t^*)}{f(Y_t|x_t^m)f(x_t^m|x_{t-1}^{m+1})f(x_{t+1}^{m+1}|x_t^m)g(x_t^*|x_t^m)}, 1\right)$$

This algorithm breaks the serial nature of standard approaches and allows for efficient use of parallel computing power: Given the Markovian setup, the time t posterior of each state only depends on its immediate neighbors as well as the observation made in time t. Thus, it is feasible to simulate the state process in two sequential computations, where each computation is inherently parallel in nature. The total number of simulations required per iteration is still $N \times T$, but it can now be distributed on several CPUs as the simulations are not of sequential nature anymore. For example, if the observation equation is computationally intensive, the Markovian setup in combination with our algorithm allows for computing the observation likelihoods in parallel, thereby dramatically decreasing total CPU time required, as the example in the next section shows.

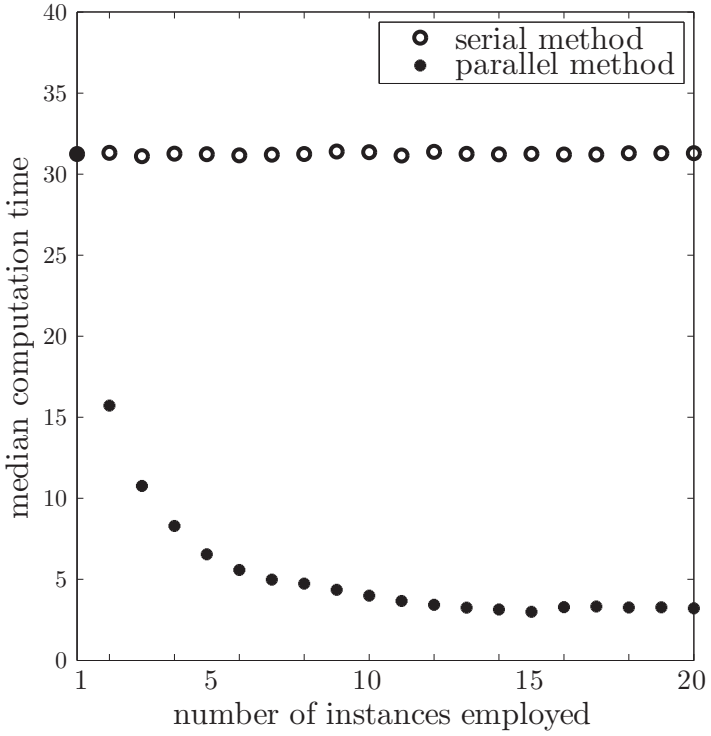
2 Example

To give an example, we consider a linear state transition model with nonlinear observations and Gaussian innovations. The nonlinearity of the system implies that we must resort to MCMC

methods. We compare our parallel approach to the naive sequential fragmentation found in the literature.

To add tangibility to our example, consider the inverse problem of Credit Default Swap (CDS) pricing: Given the set of observed CDS spread quotes for a period $t = 1 \dots T$, we want to find the posterior distribution of the latent default intensities in X, where we have discretized the state evolution equations.[3] As the CDS pricing equation is highly nonlinear, we cannot resort to linear, i.e. Kalman like, methods. Even linearization of the observation equation is not feasible, as the linearization coefficients have to be obtained repeatedly for every single observation date. The requried computational time would still be prohibitive. When we apply our method, on the other hand, the computational time can be reduced considerably.

Figure 2.1: Comparison of computational time required



This figure depicts the time requirements of both, a serial and our parallel simulation algorithm for posterior state inference in latent state Markov models. Each point represents the median time requirement for one sweep through $T = 1000$ time points in a model with Gaussian linear state transitions and Gaussian, nonlinear, observation equations. Each median is computed from 100 sweeps. The algorithms ran on workstation equipped with twelve Intel Xeon E5-2630 CPUs, 2.30 GHz each, and a maximum of 20 possible instances, 64 GB RAM.

Our toy model is as follows. We consider a two-dimensional linear state process model on $\mathbb{R}^2$ with $T = 1000$ time points. At each point in time, the state emits a noisy, three-dimensional

[3] In a Credit Default Swap contract two parties exchange cash flows contingent on the occurrence of a credit event. The contract consists of a premium and a protection leg. See Duffie and Singleton (2003) for details.

observation Y_t. As stated above, we can consider the coefficients in the transition and observation equations fixed and known. As the transition is linear and Gaussian, the state evolution densities are easy to evaluate. The observation equation, on the other hand, is assumed computationally intensive, i.e. we assume that the computation of the time-t observation likelihood given state X_t takes 0.0005 seconds each.

Employing one up to 20 instances on a multi-CPU machine with Mathworks' Matlab distributed computing toolbox installed, we run the mth MCMC sweep for 100 times and evaluate the median time required per sweep. Figure 2 compares the time requirement of our algorithm compared with the standard sequential simulation routine. We clearly find that our algorithm dominates the standard approach as soon as two CPU instances can be linked, and for a sufficient number of instances (i.e. cores), the time can be reduced by approximately 90% in our example and on this particular workstation. In a companion paper, Held and Omachel (2014) apply this method and are thus able to reduce the required simulation time by over 92%. We thus urge researchers using MCMC methods to seriously consider parallelization possibilities, especially the method we propose.

3 Conclusion

In this note, we present a novel state simulation method to be applied within a larger MCMC framework. Our method harnesses the power of parallel computing, thus enabling research in latent state dynamics to break the curse of sequential algorithms and employ a parallel method instead. In an example, we show that this approach can save a substantial amount of computational (calendar) time, where a reduction factor of 90% can be obtained without further ado.

References

Besag, Julian. Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society, Series B (Methodological)*, 36(2), 1974.

Carlin, Bradley; Polson, Nicholas, and Stoffer, David. A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling. *Journal of the American Statistical Association*, 87 (418):493–500, 1992.

Carter, C. K. and Kohn, R. On Gibbs Sampling for State Space Models. *Biometrika*, 81(3): 541–553, 1994.

Duffie, Darrell and Singleton, Kenneth J. *Credit Risk: Pricing, Measurement, and Management*. Princeton University Press, 2003.

Eraker, Bjørn. MCMC Analysis of Diffusion Models With Applications to Finance. *Journal of Business & Economic Statistics*, 19(2):177–191, 2001.

Früwirth-Schnatter, Sylvia. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15:183–202, 1994.

Hammersley, John and Clifford, Peter. Markov fields on finite graphs and lattices. *Unpublished*, 1971.

Held, Matthias and Omachel, Marcel. The Relationship between Sovereign Credit and EUR/USD Risk Premia in the Eurozone. *Working paper*, 2014.

Kalman, Rudolph. A new approach to linear filtering and prediction problems. *Journal of Fluid Engineering*, 82(1):35–45, 1960.

Kim, Sangjoon; Shephard, Neil, and Chib, Siddhartha. Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models. *Review of Economic Studies*, 65(3):361–393, 1998.

Kitagawa, Genshiro. Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models. *Journal of Computational and Graphical Statistics*, 5(1):1–25, 1996.

Metropolis, Nicholas; Rosenbluth, Arianna W.; Rosenbluth, Marshall N., and Teller, Augusta H. Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21(6):1087–1092, 1953.

Otto von Guericke University Magdeburg
Faculty of Economics and Management
P.O. Box 4120 | 39016 Magdeburg | Germany

Tel.: +49 (0) 3 91 / 67-1 85 84
Fax: +49 (0) 3 91 / 67-1 21 20

www.fww.ovgu.de/femm

ISSN 1615-4274