Exact inference for discretely observed diffusions

Giorgos Sermaidis[1]
joint work with
Alex Beskos[2], Omiros Papaspiliopoulos [3] and Gareth O. Roberts [1]

[1]University of Warwick

[2]University College London

[3]Universitat Pompeu Fabra, Barcelona

Diffusion process

A diffusion process V is a continuous time Markov process driven by Brownian Motion (BM) B, defined as the solution to the Stochastic Differential Equation (SDE)

$$\mathrm{d}V_s = \beta(V_s;\theta)\mathrm{d}s + \sigma(V_s;\theta)\mathrm{d}B_s\,,\, V_0 = v_0\,,\, s \geq 0.$$

The drift $\beta(.;\theta)$ and diffusion coefficient $\sigma(.;\theta)$ functionals model the instaneous mean and variance and are allowed to depend on a vector of parameters θ.

Transition density

The exact dynamics of the diffusion process are governed by its transition density (wrt the Lebesgue measure)

$$p_t(u, w; \theta) = P(V_t \in dw \mid V_0 = u; \theta)/dw, \quad t > 0,\ w,\ u \in \mathbf{R}.$$

Unfortunately, it is typically unavailable except for a few cases. An approximation is given for sufficiently small time increment dt (Euler-Maruyama)

$$V_{t+dt} \approx V_t + \beta(V_t; \theta)dt + \sigma(V_t; \theta)\sqrt{dt}Z, \ \ Z \sim N(0, 1)$$

$$p_{dt}(u, w; \theta) \approx \mathcal{N}\left\{w; u + \beta(u; \theta)dt, \sigma^2(u; \theta)dt\right\};$$

where $\mathcal{N}\left\{w; \mu, \sigma^2\right\}$ is the density of a Gaussian r.v. with mean μ and variance σ^2 evaluated at $w \in \mathbb{R}$.

The process is defined in continuous time but the available data are always sampled in discrete time

$$Y = \{V_{t_0}, V_{t_1}, ..., V_{t_n}\}, \quad 0 = t_0 < t_1 < ... < t_n \,.$$

Let $\Delta t_i = t_i - t_{i-1}$ be the time increment between consecutive observations.

GOAL: Infer on the parameter vector θ given the data,

$$\pi(\theta \mid Y) \propto \pi(\theta) \prod_{i=1}^{n} p_{\Delta t_i}(V_{t_{i-1}}, V_{t_i}; \theta)$$

but the transition density is unavailable. Intuitevily it is the marginal density where the path between two consecutive points has been integrated out.
Proceed by augmenting the data with the missing paths in between since there exists an analytic expression for the complete likelihood.

Complete path and likelihood

Without loss of generality assume only two observations, V_0 and V_t. For simplicity assume $\sigma(.;\theta) = \sigma$ and let $V = \{V_s,\, 0 \leq s \leq t\}$ be the complete path from

$$\mathrm{d}V_s = \beta(V_s;\theta)\mathrm{d}s + \sigma \mathrm{d}B_s.$$

We can write the density of a complete path w.r.t. the law $\mathbb{W}_\sigma$ of the simpler driftless process $\mathrm{d}M_s = \sigma \mathrm{d}B_s$ as

$$G(V;\theta) = \exp\left\{-\int_0^t \frac{\beta(V_s;\theta)}{\sigma^2}\mathrm{d}V_s - \frac{1}{2}\int_0^t \frac{\beta^2(V_s;\theta)}{\sigma^2}\mathrm{d}s\right\}.$$

Proceed to a Metropolis within Gibbs (MwG) algorithm

- Sample from $\pi(V \mid Y, \theta)$ - this is a path from the diffusion bridge (intractable) $\Rightarrow$ propose from the simpler driftless process, Brownian Bridge (BB)
- Sample from $\pi(\theta \mid Y, V)$.

Back

There are two problems with this approach:

- Simulation of a complete path is not possible (infinite dimensional). It is necessary to discretize the imputed path in finite number of points.
- The missing path contains infinite information on the parameters involved in the diffusion coefficient due to the quadratic variation identity

$$\lim_{m\to\infty}\sum_{i=1}^{m}(V_{ti/m}-V_{t(i-1)/m})^2 = t\sigma^2.$$

  Strong dependence $\Rightarrow$ the finer the discretization of the path gets, the worse the mixing of the chain becomes.

Following Roberts and Stramer (2001), need to transform the missing path to break this dependence and create a parameter free dominating measure. The transformation takes place in two stages.

First path transformation: Transform the original SDE to a unit diffusion coefficient one by applying $X := \eta(V;\theta) = \int^V 1/\sigma(u;\theta)\mathrm{d}u$. Itô's formula gives

$$\mathrm{d}X_s = \alpha(X_s;\theta)\mathrm{d}s + \mathrm{d}B_s, \;\; x_0(\theta) = \eta(V_0;\theta);$$

where

$$\alpha(u;\theta) = \frac{\beta\{\eta^{-1}(u;\theta);\theta\}}{\sigma\{\eta^{-1}(u;\theta);\theta\}} - \frac{\sigma'\{\eta^{-1}(u;\theta);\theta\}}{2}.$$

The driftless process now is BM starting at $x_0(\theta)$ - not parameter free measure. Back

Second path transformation: If $\dot{X}$ is the bridge constructed by conditioning X on $x_t(\theta) = \eta(V_t;\theta)$ then

$$\ddot{X}_s := \dot{X}_s - (1 - \frac{s}{t})x_0(\theta) - \frac{s}{t}x_t(\theta).$$

This transformation forces the bridge to start and finish at 0. The inverse transformation is given by

$$g_\theta(\ddot{X}_s) := \ddot{X}_s + (1 - \frac{s}{t})x_0(\theta) + \frac{s}{t}x_t(\theta).$$

We can now write a joint density for the transformed missing path $\ddot{X}$ and the data Y w.r.t. the parameter free product measure $Leb \times \mathbb{W}^{(t,0,0)}$

$$\pi(\ddot{X}, Y \mid \theta) \propto |\eta^{'}(v_t;\theta)| \mathcal{N}\left\{x_t(\theta); x_0(\theta), t\right\} G\{g_\theta(\ddot{X}); \theta\}$$

where $\mathbb{W}^{(t,0,0)}$ is the law of a BB from $(0,0)$ to $(t,0)$. 1. Girsanov

We can now write a joint density for the transformed missing path $\ddot{X}$ and the data Y w.r.t. the parameter free product measure $Leb \times \mathbb{W}^{(t,0,0)}$

$$\pi(\ddot{X}, Y \mid \theta) \propto |\eta^{'}(v_t;\theta)| \mathcal{N}\left\{x_t(\theta); x_0(\theta), t\right\} G\{g_\theta(\ddot{X}); \theta\}$$

where $\mathbb{W}^{(t,0,0)}$ is the law of a BB from $(0,0)$ to $(t,0)$. 1. Girsanov

MwG implementation:

- $\pi(\ddot{X} \mid Y, \theta)$ - discretize the path and use a Metropolis step proposing by something tractable (BB). The acceptance ratio depends on G.
- $\pi(\theta \mid Y, \ddot{X})$ - Approximate the integrals by Riemman sums and use

$$\exp\left\{-\sum_{j=0}^{M-1} \alpha\{g_\theta(\ddot{X}_j);\theta\}\Delta g_\theta(\ddot{X}_j) - \frac{\delta}{2}\sum_{j=0}^{M-1} \alpha\{g_\theta(\ddot{X}_j);\theta\}\right\}.$$

Consequence: We are sampling from an approximation to the posterior.

An example - The SINE process

Consider 1000 equidistant observations, $\Delta t = 0.5$, from the SINE process,

$$\mathrm{d}V_s = \sin(V_s - \phi)\mathrm{d}s + \sigma \mathrm{d}B_s \,,$$

with $\phi = \pi$ and $\sigma = 0.5$. We run the algorithm for three values of $m = \{5, 10, 20\}$.

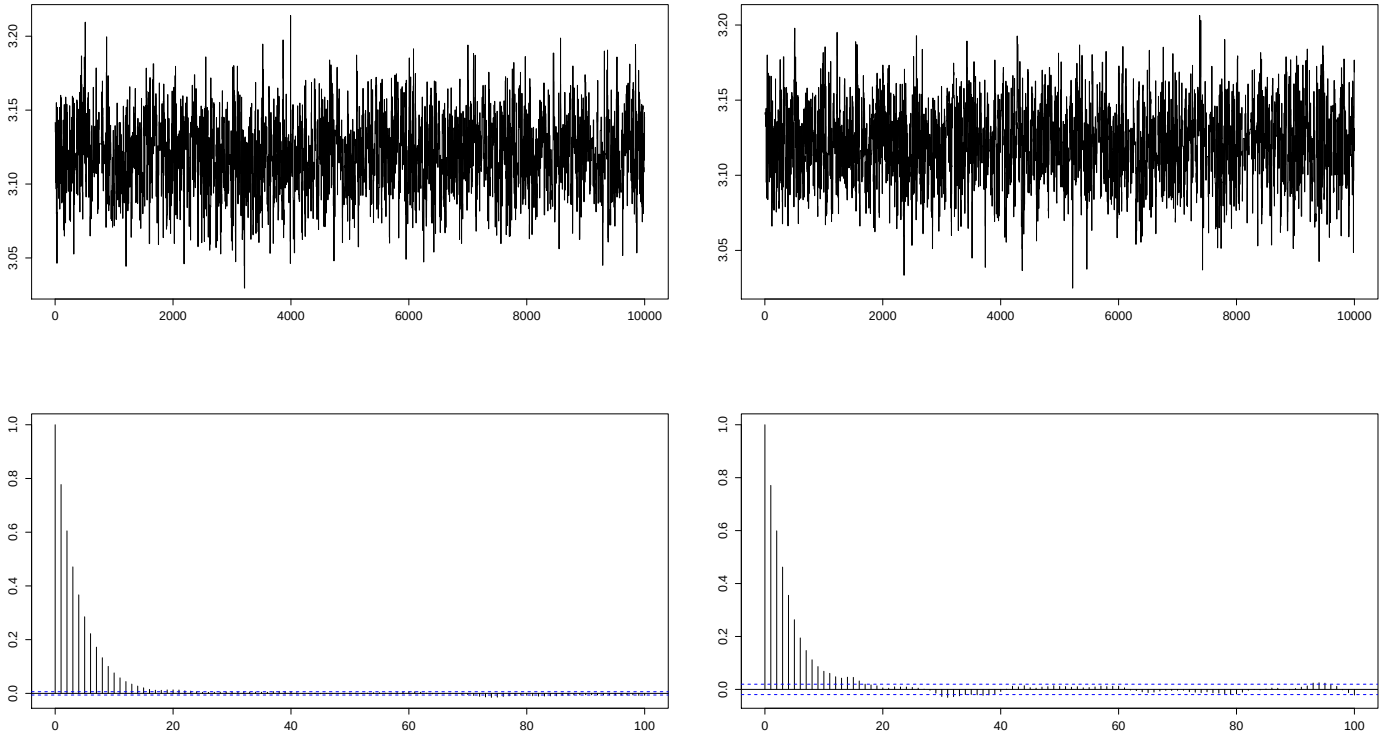

Figure: Trace plots and autocorrelation function for ϕ using $m = 5$ (left) and $m = 20$ (right).

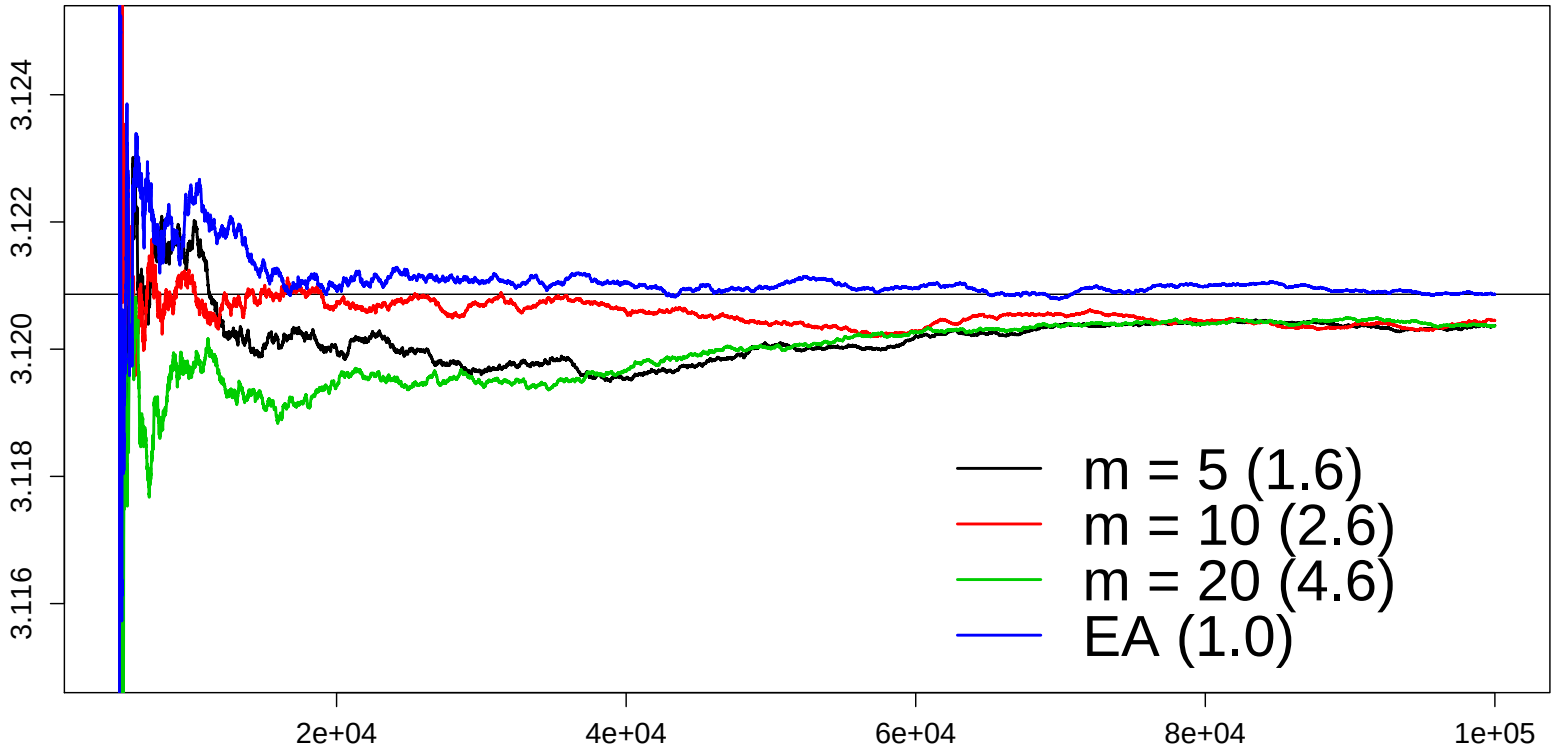

Figure: Running average for ϕ for the three discretization schemes. In parentheses are the units of time needed for 1000 iterations of the EA chain. The average number of imputed points for the EA was 2.

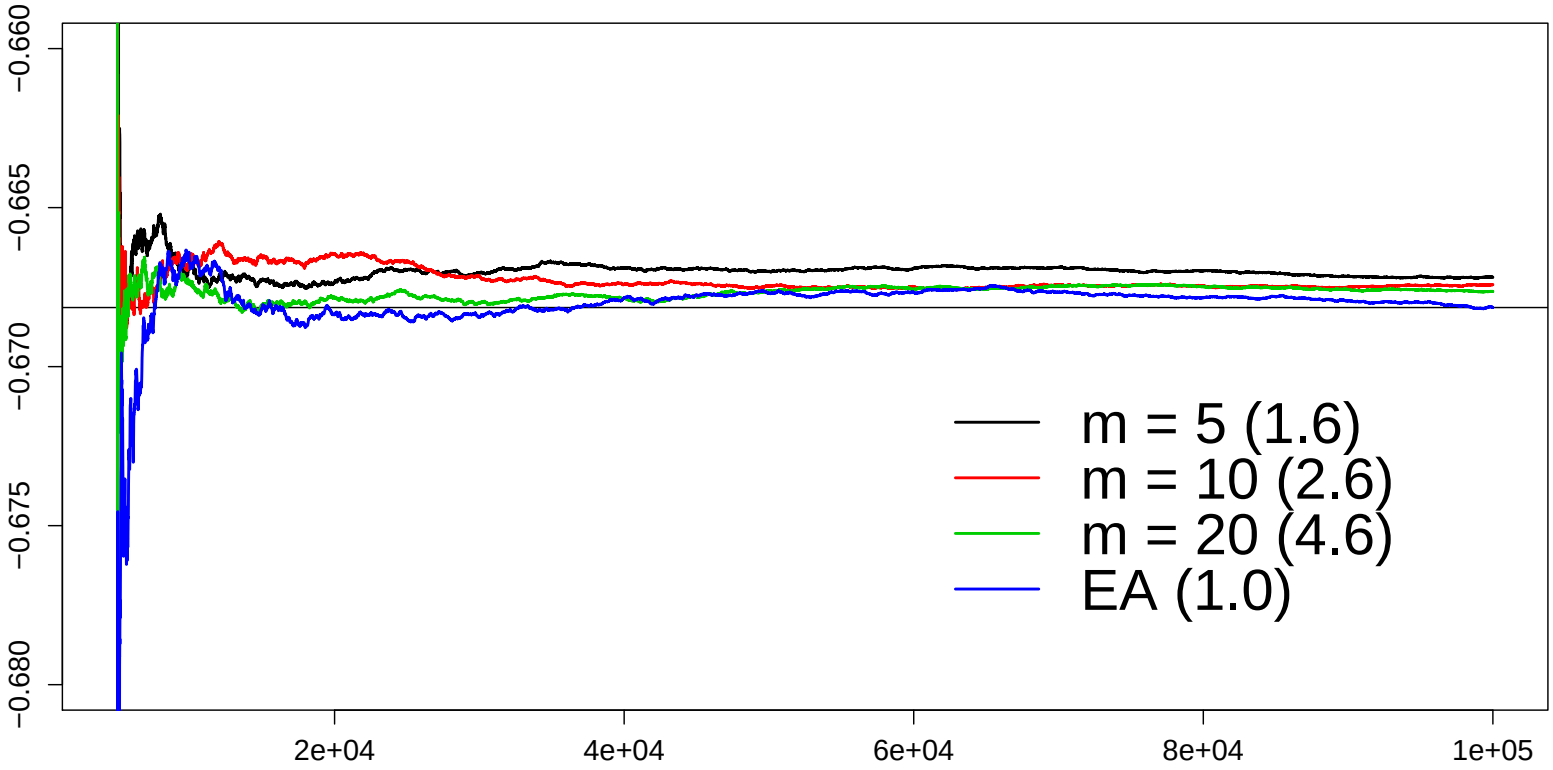

Figure: Running average for $\log(\sigma)$ for the three discretization schemes. In parentheses are the units of time needed for 1000 iterations of the EA chain. The average number of imputed points for the EA was 2.

Back

The Exact Algorithm

The problem can be tackled using a novel methodology for simulating the skeleton of a diffusion bridge with no discretization error (Beskos *et al* 2006).

Let $Z = \{Z_s\}, 0 \leq s \leq t$ be a unit diffusion coefficient diffusion satisfying

$$\mathrm{d}Z_s = \alpha(Z_s; \theta)\mathrm{d}s + \mathrm{d}B_s,$$

starting at $(0, x)$ and finishing at (t, y); denote its law by $\mathbb{Q}_\theta^{(t,x,y)}$.

2. Transformation

We can write the density of the diffusion bridge w.r.t that of a BB

$$\frac{\mathrm{d}\mathbb{Q}_\theta^{(t,x,y)}}{\mathrm{d}\mathbb{W}^{(t,x,y)}}(\omega) \quad \propto \quad \exp\left\{-\int_0^t \alpha(\omega_s; \theta)\mathrm{d}\omega_s - \frac{1}{2}\int_0^t \alpha^2(\omega_s; \theta)\mathrm{d}s\right\}.$$

The Exact Algorithm

The problem can be tackled using a novel methodology for simulating the skeleton of a diffusion bridge with no discretization error (Beskos *et al* 2006).
Let $Z = \{Z_s\}, 0 \leq s \leq t$ be a unit diffusion coefficient diffusion satisfying

$$\mathrm{d}Z_s = \alpha(Z_s; \theta)\mathrm{d}s + \mathrm{d}B_s,$$

starting at $(0, x)$ and finishing at (t, y); denote its law by $\mathbb{Q}_\theta^{(t,x,y)}$.

2. Transformation

We can write the density of the diffusion bridge w.r.t that of a BB

$$\frac{\mathrm{d}\mathbb{Q}_\theta^{(t,x,y)}}{\mathrm{d}\mathbb{W}^{(t,x,y)}}(\omega) \quad \propto \quad \exp\left\{-\int_0^t \alpha(\omega_s; \theta)\mathrm{d}\omega_s - \frac{1}{2}\int_0^t \alpha^2(\omega_s; \theta)\mathrm{d}s\right\}.$$

IDEA: Proceed to simulation by bounding the density and use Rejection Sampling.

Assumptions:

- $\alpha(.;\theta)$ is continuously differentiable. Let $\alpha^{'}(.;\theta)$ be its derivative.
- $(\alpha^2 + \alpha^{'})(.;\theta)$ is bounded below (mild, usually satisfied).

$$l(\theta) \leq \inf_{u \in \mathbf{R}} \left\{ \frac{1}{2}(\alpha^2 + \alpha^{'})(u;\theta) \right\}$$

- For a given ω define the range as

$$r(\omega;\theta) \geq \sup_{s \in [0,t]} \left\{ \frac{1}{2}(\alpha^2 + \alpha^{'})(\omega_s;\theta) - l(\theta) \right\}.$$

  and assume that $r(\omega;\theta) = r\{H(\omega);\theta\}$ where $H(\omega)$ is a finite dimensional path characteristic.

Define the non-negative function $0 \leq \phi \leq 1$ as

$$\phi(\omega_s;\theta) = \frac{1}{r(\omega;\theta)} \left\{ \frac{1}{2}(\alpha^2 + \alpha^{'})(\omega_s;\theta) - l(\theta) \right\}.$$

Using Itô's formula we can rewrite the density of the diffusion bridge as

$$\frac{\mathrm{d}\mathbb{Q}_\theta^{(t,x,y)}}{\mathrm{d}\mathbb{W}^{(t,x,y)}}(\omega) \propto \exp\left\{-r(\omega;\theta)\int_0^t \phi(\omega_s;\theta)\mathrm{d}s\right\} \leq 1;$$

Ideally, we would simulate a complete BB path and evaluate this density.

We can construct a Bernoulli experiment which has acceptance probability equal to the bounded density using an auxiliary random variable.

Theorem: Let Φ be a homogeneous Poisson process of intensity $r(\omega;\theta)$ on $[0,t]x[0,1]$ and N the number of points of Φ below the graph $s \longrightarrow \phi(\omega_s;\theta)$. Then

$$\mathbb{P}(N=0 \mid \omega) = \exp\left\{-r(\omega;\theta)\int_0^t \phi(\omega_s;\theta)\mathrm{d}s\right\}.$$

A rejection sampling would be as follows

1. Simulate a complete path $\omega \sim \mathbb{W}^{(x,y)}$.
2. Evaluate $H(\omega)$ and calculate $r(\omega;\theta)$. Draw $\kappa \sim \mathrm{Po}\{r(\omega;\theta)t\}$ and generate a marked Poisson process $\Phi = \{\Psi, \Upsilon\}$, where $\Psi = \{\psi_1, \ldots, \psi_\kappa\}$ are uniformly distributed points on $[0, t]$ and $\Upsilon = \{\upsilon_1, \ldots, \upsilon_\kappa\}$ are uniform marks on $[0, 1]$.
3. Simulate a skeleton of the proposed path $S(\omega) := \{\omega_{\psi_j}, j = 1, ..., \kappa\}$ at finite number of points Ψ and compute the acceptance indicator

$$I := \prod_{j=1}^{\kappa} \mathbb{I}[\phi(\omega_{\psi_j}) < \upsilon_j].$$

4. If $I = 1$ accept the proposed BB, otherwise repeat.

Implemenation is impossible. Note though that decision of acceptance depends only on finite information $\{H(\omega), \Phi, S(\omega)\} \Rightarrow$ change the order of steps $(1, 2)$. The technique is called Retrospective Sampling, introduced in Papaspiliopoulos and Roberts (2008) and the resulting algorithm is the Exact Algorithm (EA).

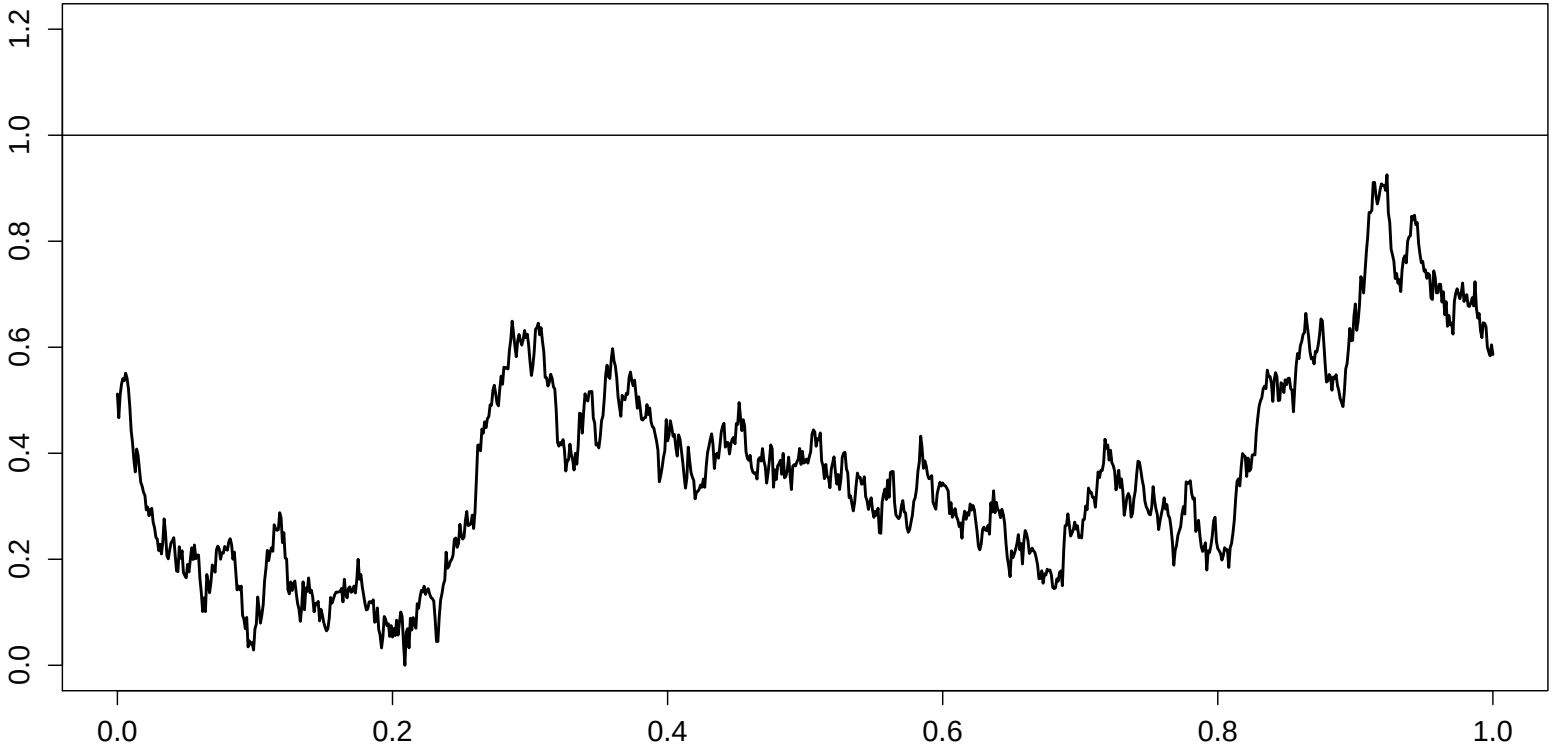

Figure: The graph of $s \to \phi(\omega_s; \theta)$.

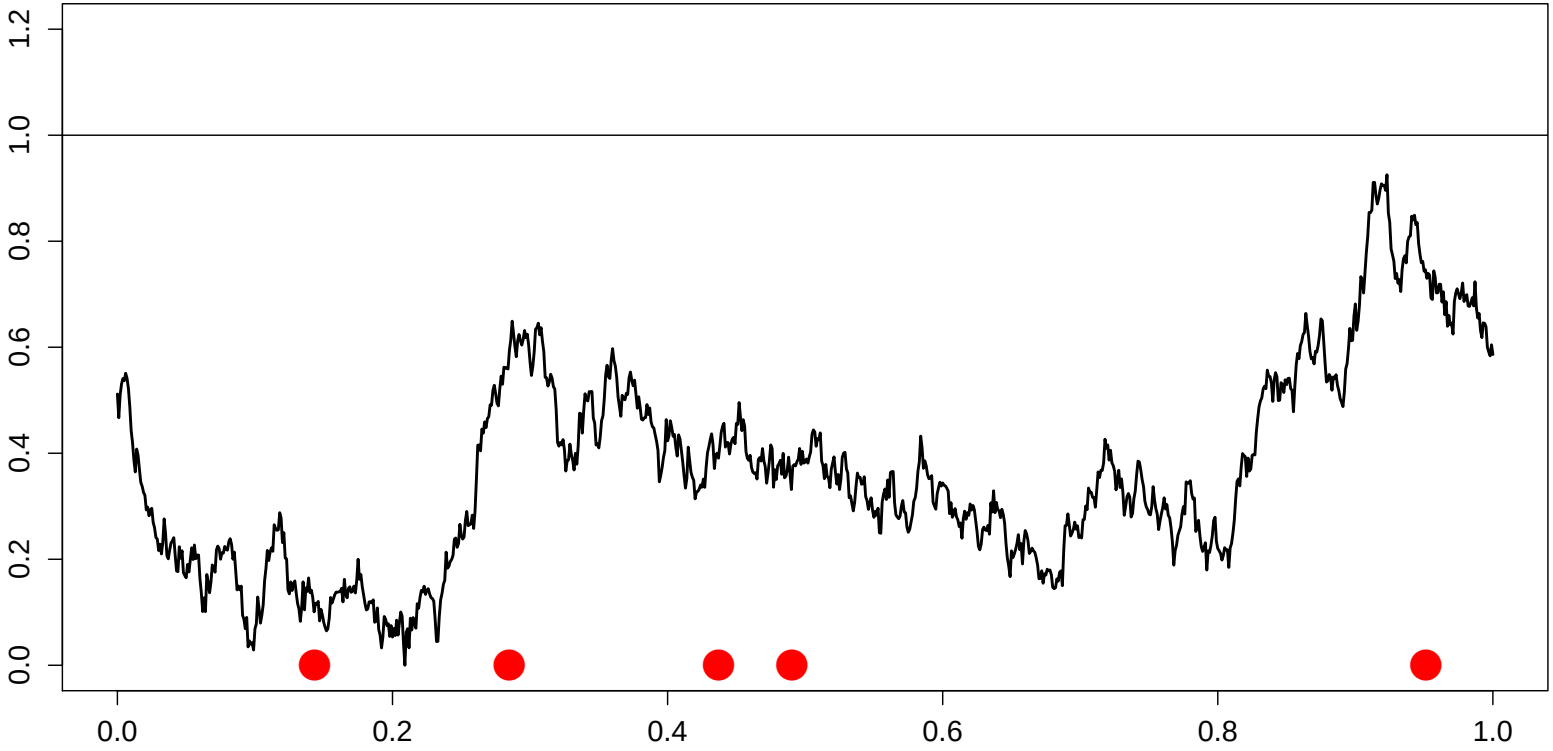

Figure: The graph of $s \longrightarrow \phi(\omega_s; \theta)$.

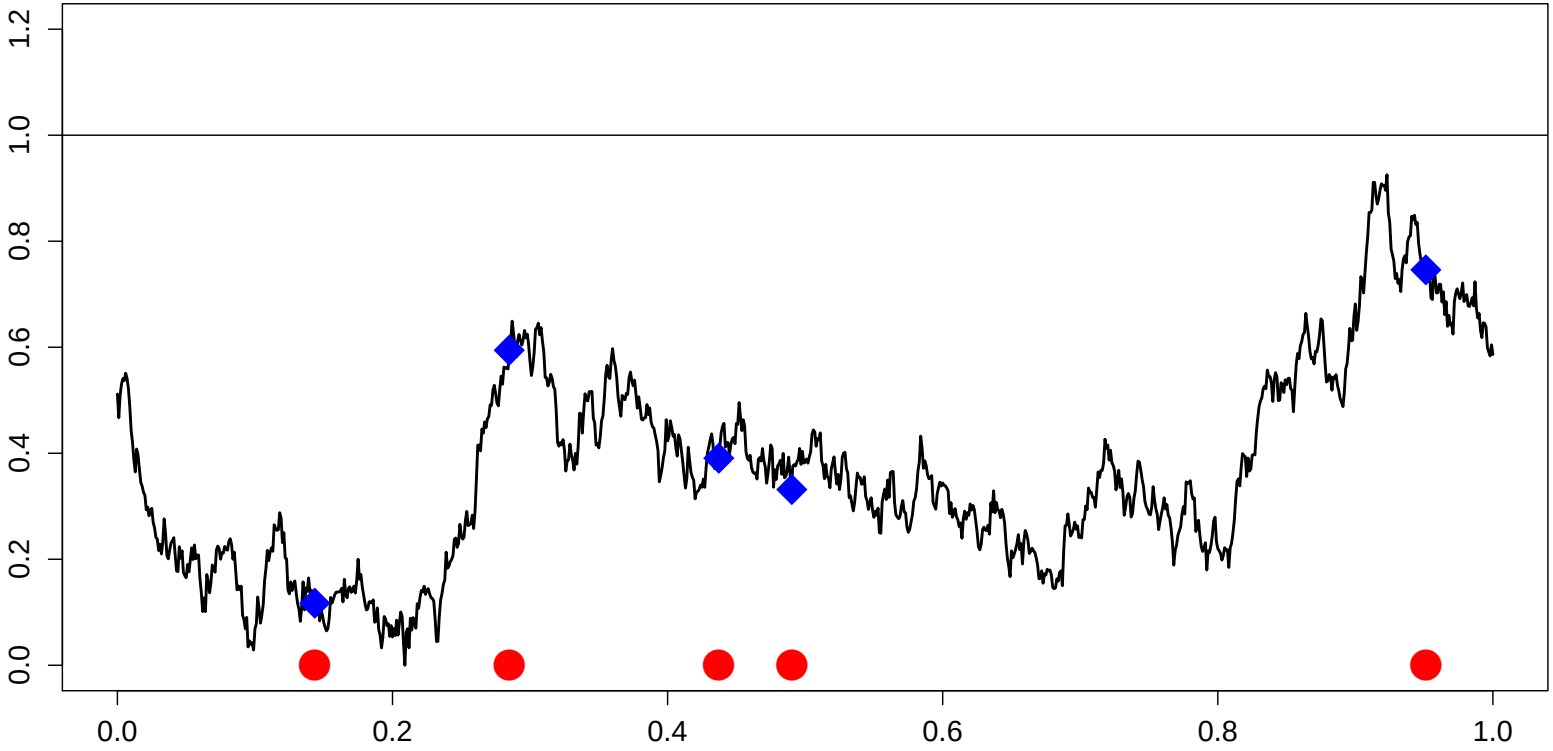

Figure: The graph of $s \longrightarrow \phi(\omega_s; \theta)$.

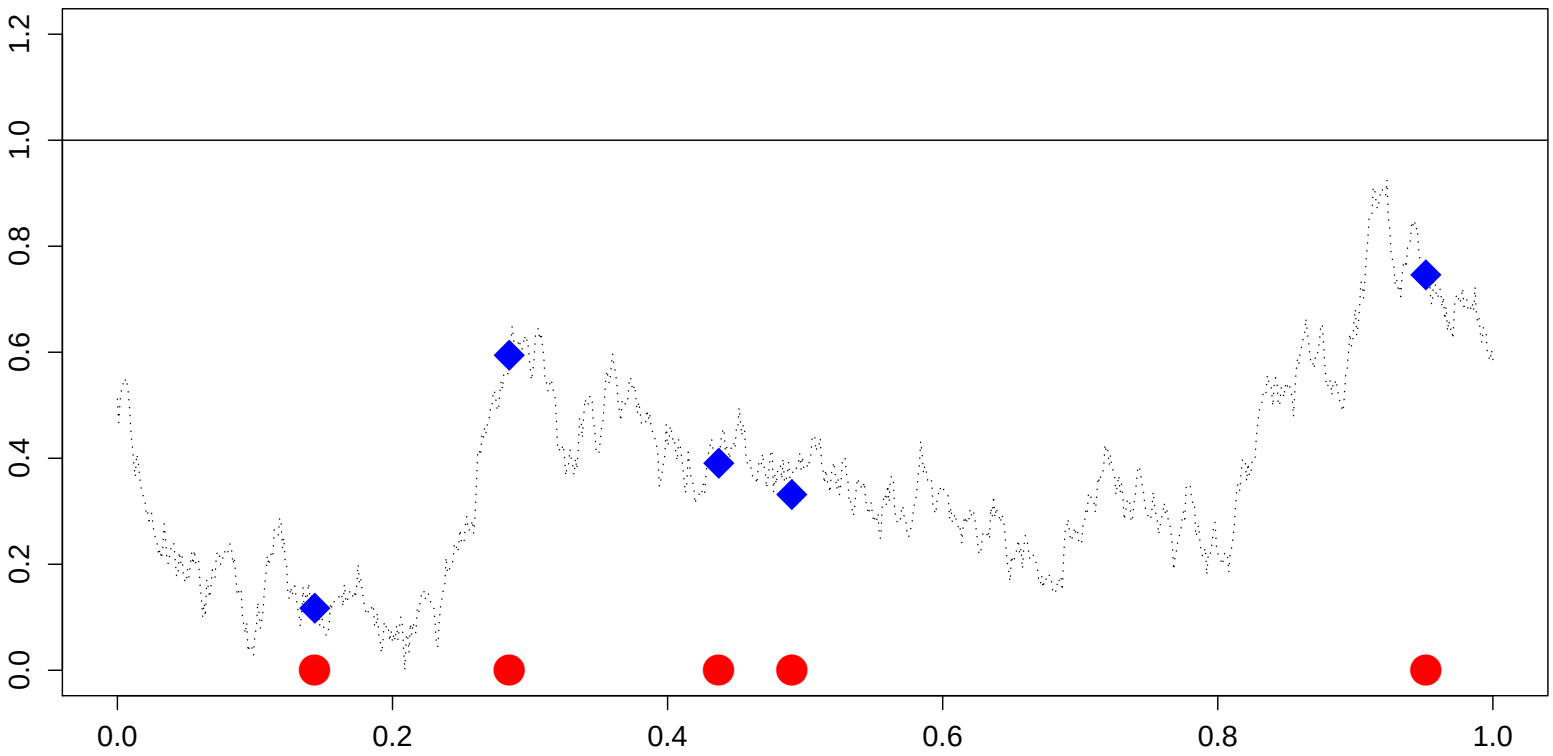

Figure: The graph of $s \longrightarrow \phi(\omega_s; \theta)$.

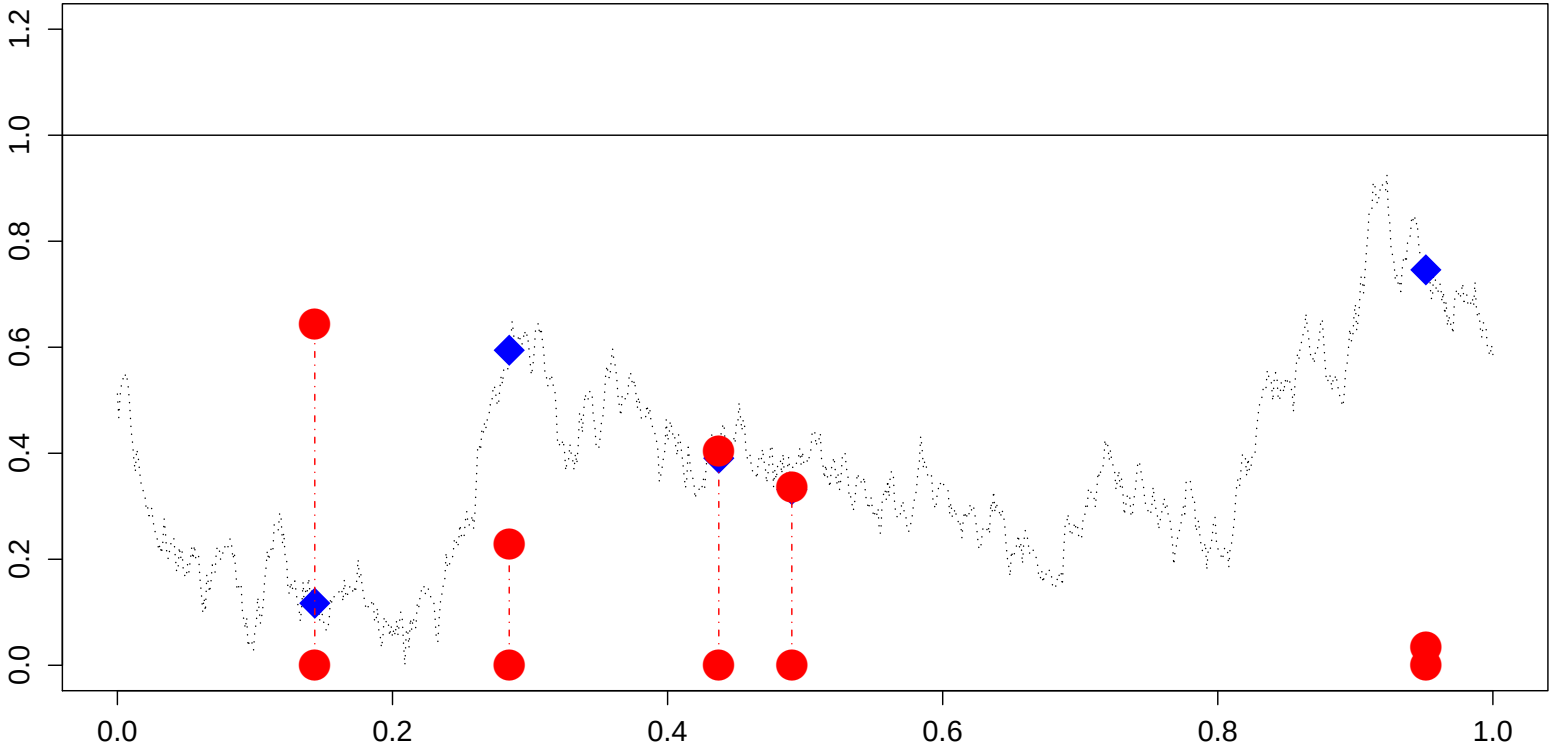

Figure: The graph of $s \longrightarrow \phi(\omega_s; \theta)$.

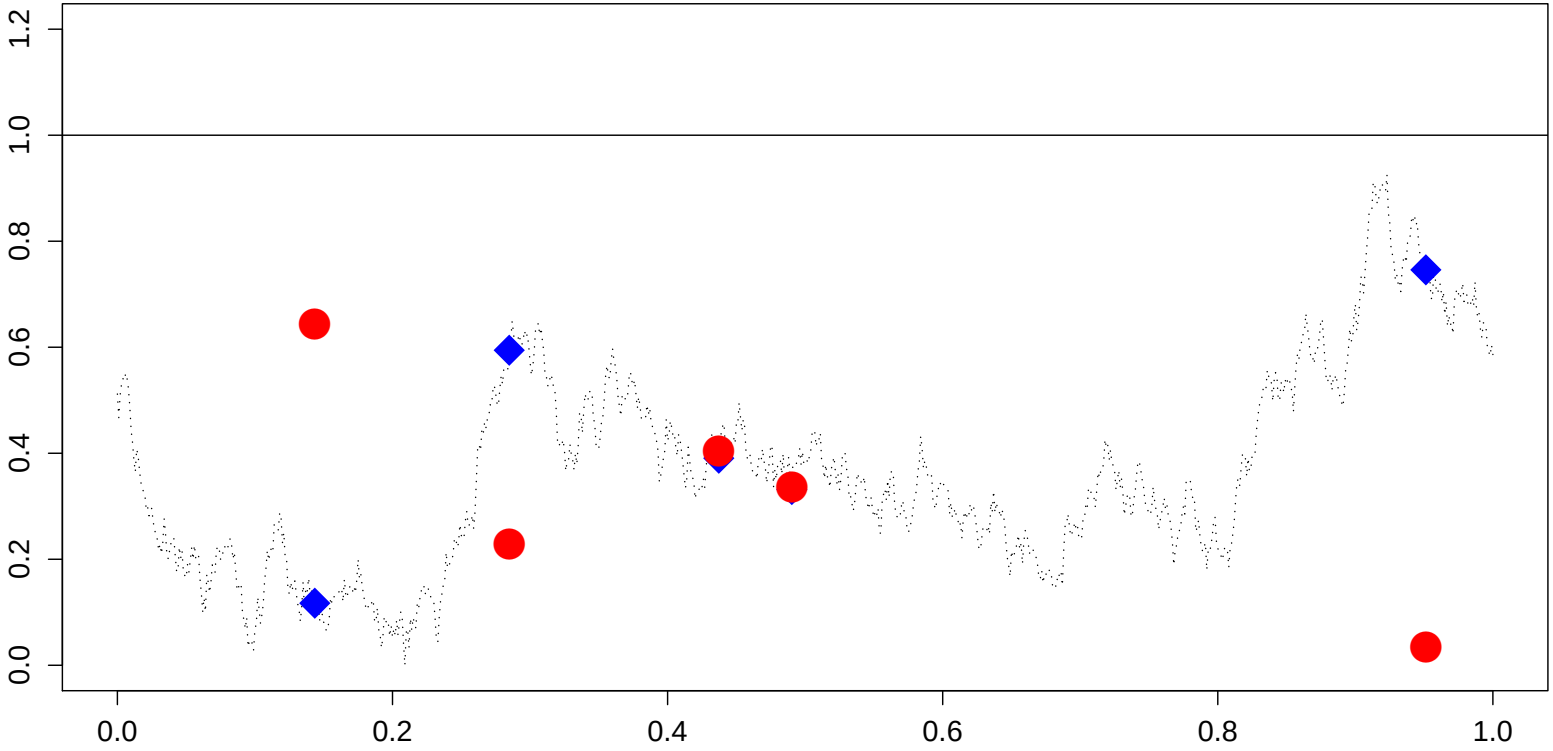

Figure: The graph of $s \longrightarrow \phi(\omega_s; \theta)$.

Therefore, the EA keeps proposing (ω, Φ) and decides its acceptance based only on partial information (finite) on the path

$$X = \{H(\omega), \Phi, S(\omega)\}.$$

The r.v. X can be thought of as a random sufficient statistic for the diffusion bridge; can construct the complete path simulating a BB given X.

Depending on the drift functional we have three EA types

- Case where $(\alpha^2 + \alpha^{'})(.;\theta)$ is bounded above (EA1).
  No need to simulate $H(\omega)$ since $r(\omega;\theta) = r(\theta)$.

- Case where $\limsup_{u\to\infty}(\alpha^2 + \alpha^{'})(u;\theta) < \infty$ (EA2).
  Evaluation of $r(\omega;\theta)$ requires simulation of $m = \inf\{\omega_s; s \in [0,t]\}$ and $\tau = \sup\{s \in [0,t] : \omega_s = m\}$.

- Otherwise (EA3).
  Evaluation of $r(\omega;\theta)$ requires simulation of a lower and upper bound of the BB.

Using the EA for an exact MCMC

Fix two points x and y. Let $\mathbb{P}$ be the measure of a unit rate Poisson process on $[0,t] \times [0,1]$. The density of an accepted (ω^*, Φ^*) from x to y is

$$\pi(\omega^*, \Phi^* \mid x, y, \theta) = \frac{e^{-r(\theta)t} r^{\kappa^*}(\theta)}{k(x,y;\theta)} \prod_{j=1}^{\kappa^*} \left[1 - \phi \left\{ \omega^*_{\psi^*_j} + (1 - \frac{\psi^*_j}{t})x + \frac{\psi^*_j}{t} y \right\} \right]$$

w.r.t. the product measure $\mathbb{P} \times \mathbb{W}^{(t,0,0)}$ (parameter free). The term $k(x,y;\theta) \propto p_t(V_0, V_t; \theta)$ is intractable.

- This density can be evaluated using partial information $X = \{\Phi^*, S(\omega^*)\}$.
- No need for any discretization. The method is exact.
- Augment the observed data with the accepted elements of the EA and proceed to a MwG.

The corresponding hierarchical model is

$$\begin{aligned}
\theta &\sim \pi(\theta), \\
V_t \mid V_0, \theta &\sim p_t(V_0, V_t; \theta), \\
(\omega^*, \Phi^*) \mid \theta, x_0(\theta), x_t(\theta) &\sim \pi\{\omega^*, \Phi^* \mid \theta, x_0(\theta), x_t(\theta)\}.
\end{aligned}$$

Note that the observed data are in the middle of the hierarchy, ensuring that the output of the EA is indeed from $\mathbb{Q}_\theta^{(t, x_0(\theta), x_t(\theta))}$. Also allows to work with likelihood functions that have intractable constants, the constant $k(x_0(\theta), x_t(\theta); \theta)$ and the transition density have been cancelled out.

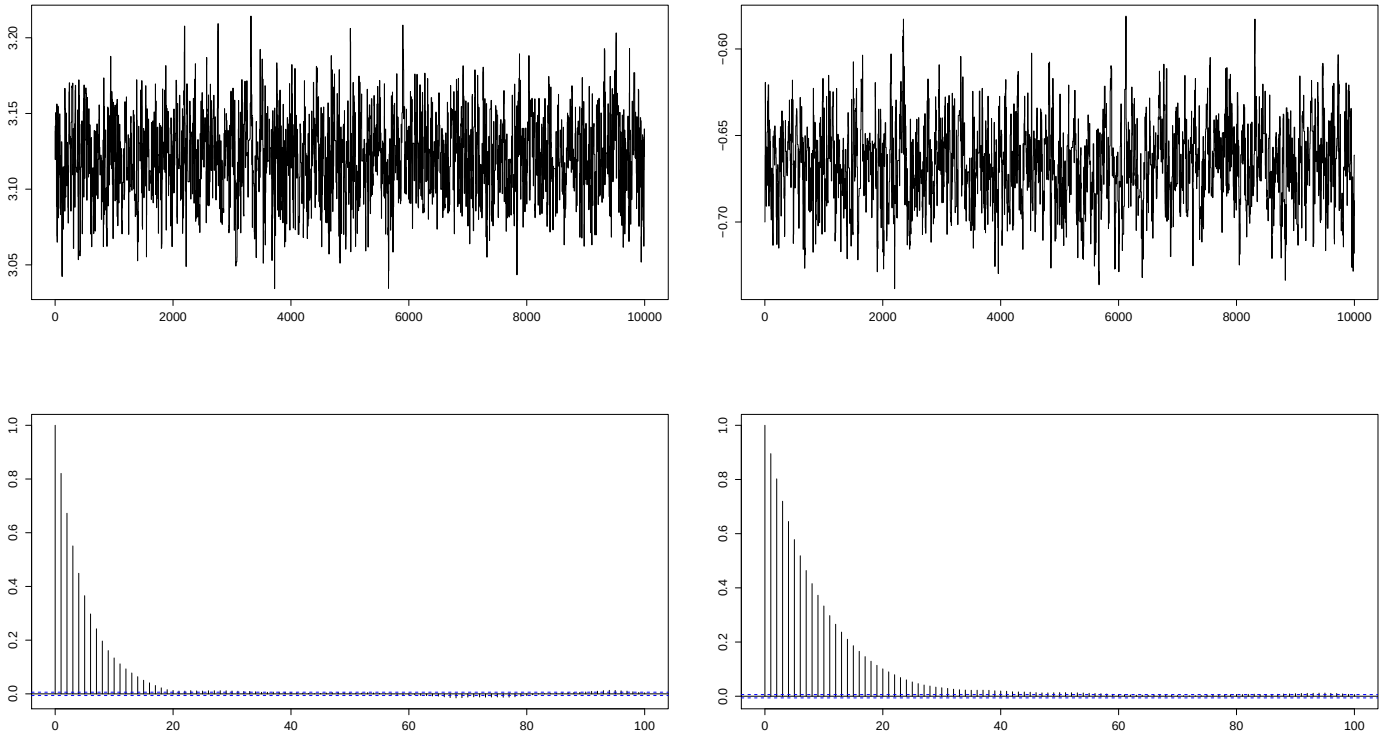

Figure: Trace plots and autocorrelation function for ϕ (left) and $\log(\sigma)$ (right).

3. Convergence

Non-centring of the EA1

Motivation:

Consider again 100 equidistant observations, $\Delta t = 5$, from the SINE process,

$$\mathrm{d}V_s = \sin(V_s - \phi)\mathrm{d}s + \sigma \mathrm{d}B_s \,,$$

with $\phi = \pi$ and $\sigma = 0.5$.

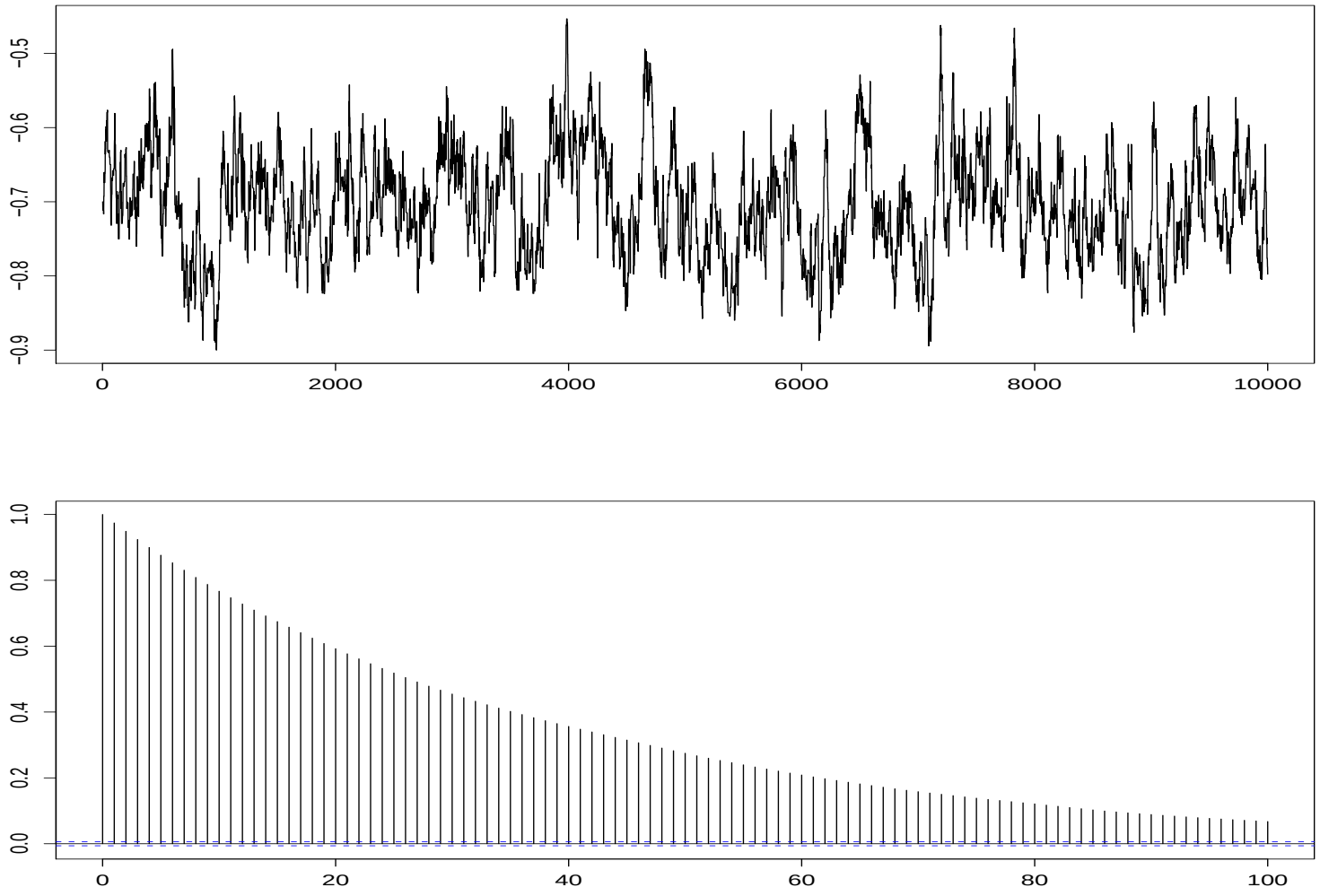

Figure: Trace plots and autocorrelation function for $\log(\sigma)$.

The performace of the chain deteriorates with large time increments.

Intuition:

- The number of points at which the missing path is evaluated increases linearly with the time increment, the EA requires $\kappa \sim \mathrm{Po}[r(\omega, \theta)\Delta t]$.
- The amount of information on the missing path increases - the missing path has to be evaluated at more points.
- The data (start and end of the missing path) do not provide enough information on this missing skeleton.
- Strong posterior dependence between the missing path and the parameters.

Construct a noncentered reparametrization for the Poisson process (based on thinning properties), Roberts *et al* (2004).

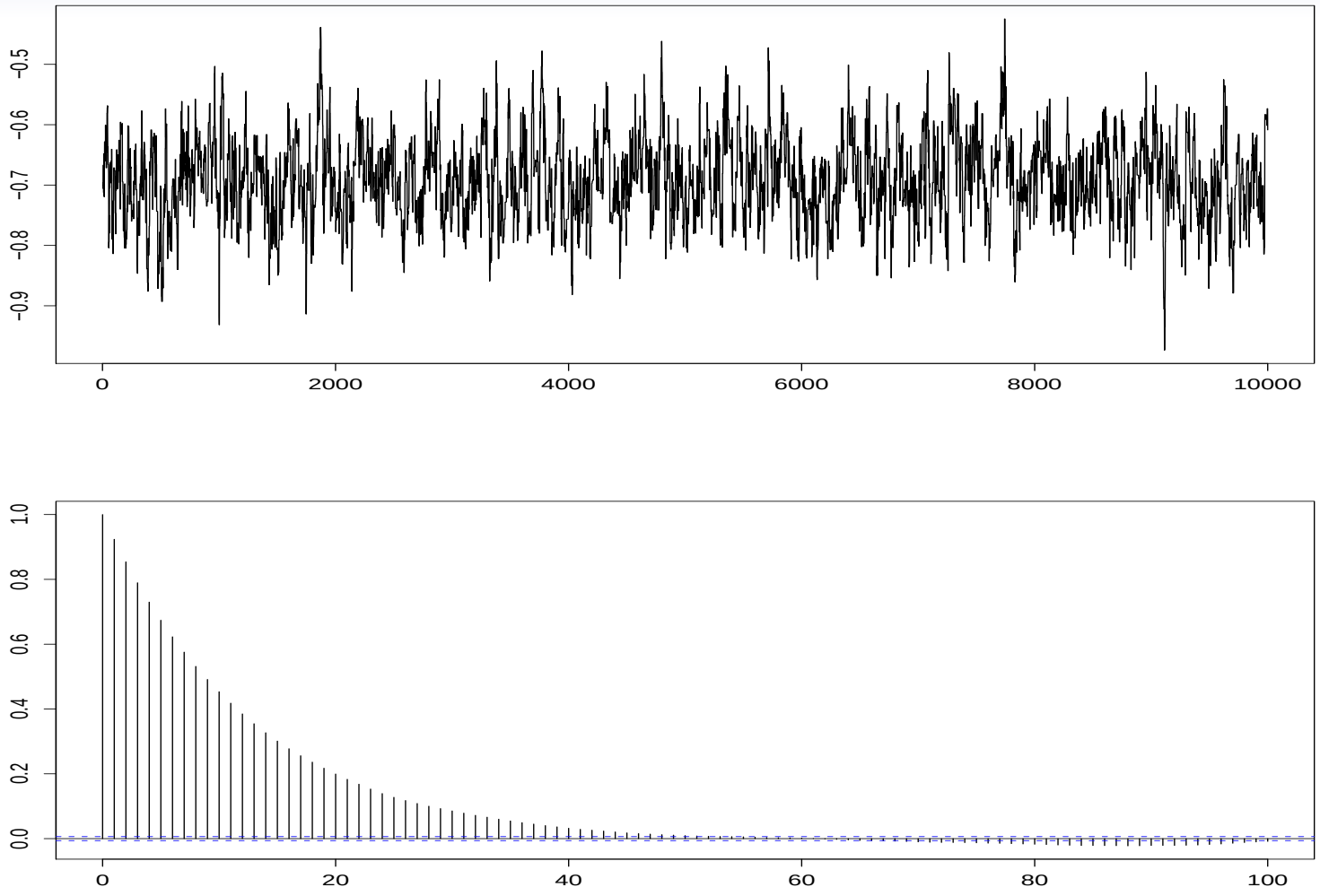

Figure: Trace plots and autocorrelation function for $\log(\sigma)$ using the noncentred reparametrization.

Interweaving strategy

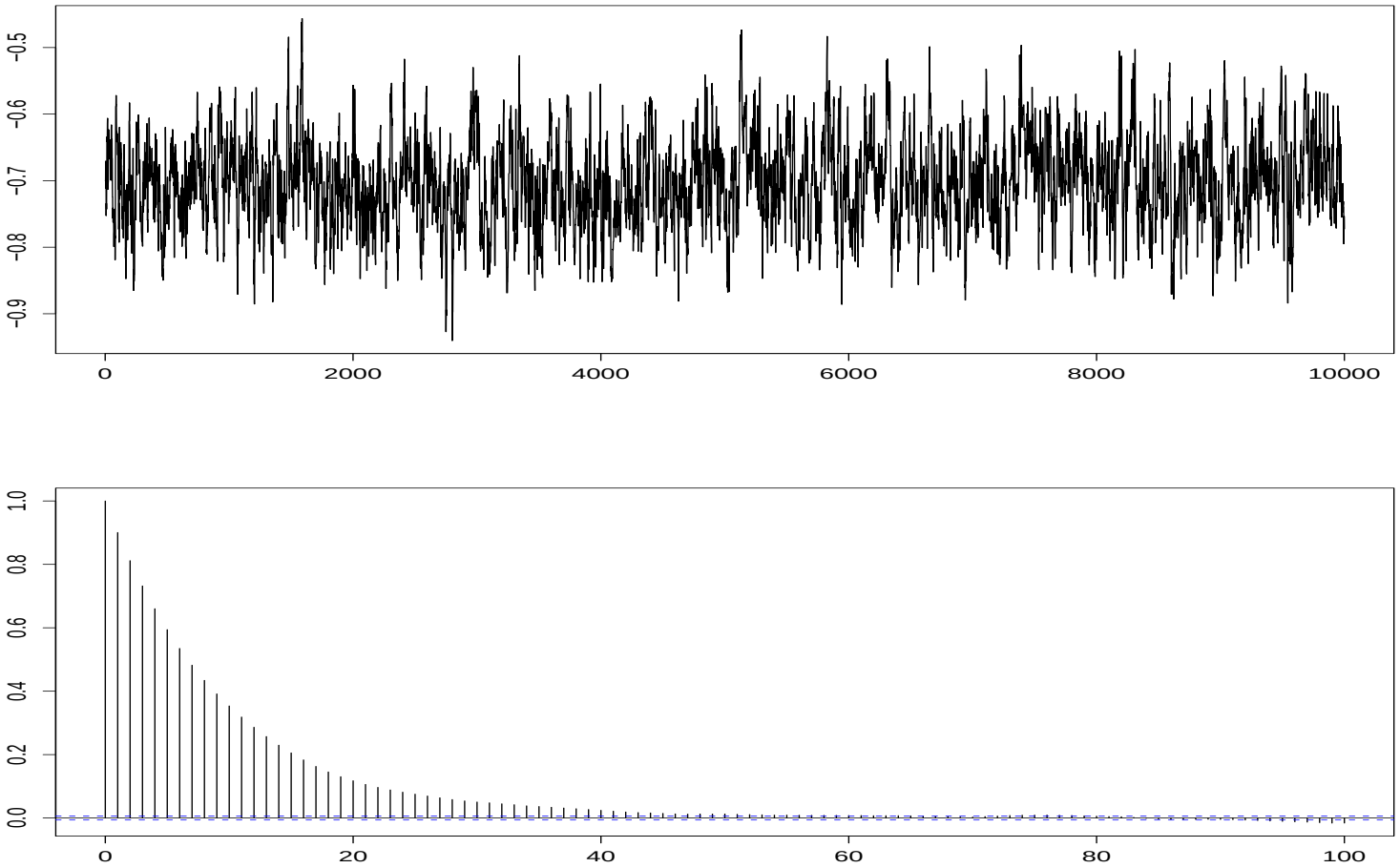

Figure: Trace plots and autocorrelation function for $\log(\sigma)$ using the interweaving strategy.

A diffusion process in the EA2 class

Consider 1000 equidistant observations, $\Delta t = 0.5$, from the log growth process,

$$\mathrm{d}V_s = \rho V_s(1 - \frac{V_s}{\lambda})\mathrm{d}s + \sigma V_s \mathrm{d}B_s,$$

with $(\rho, \lambda, \sigma) = (0.1, 1000, 0.1)$.

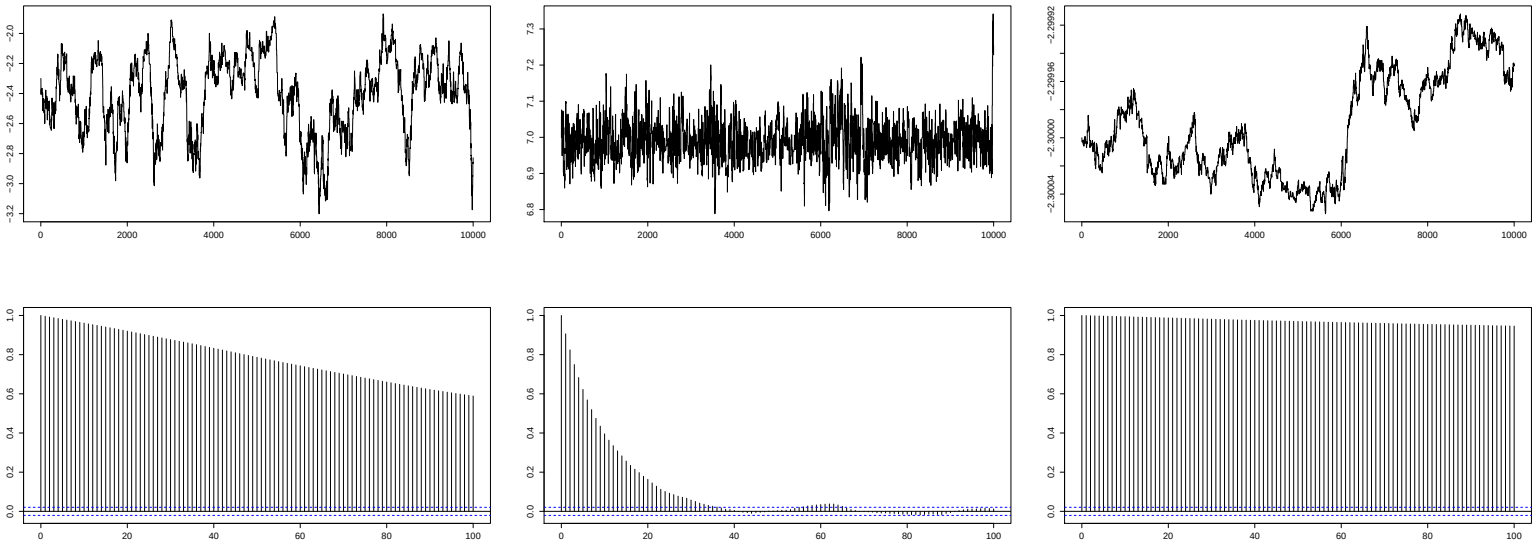

Figure: Trace plots and autocorrelation function for $\log(\rho)$, $\log(\lambda)$, $\log(\sigma)$ using the centred parametrization.

Slow mixing due to:

- The minimum of the path for a current value θ between the observations $x_0(\theta), x_t(\theta)$ is very informative on θ.
- The proposal θ^* is very likely to be rejected.
- Noncentred reparametrization for the minimum and its location $\{m(\theta), \tau(\theta)\}$ is straightforward. These r.v.s are easily simulated in terms of an Exponential, a standardized Gaussian and a Uniform r.v.

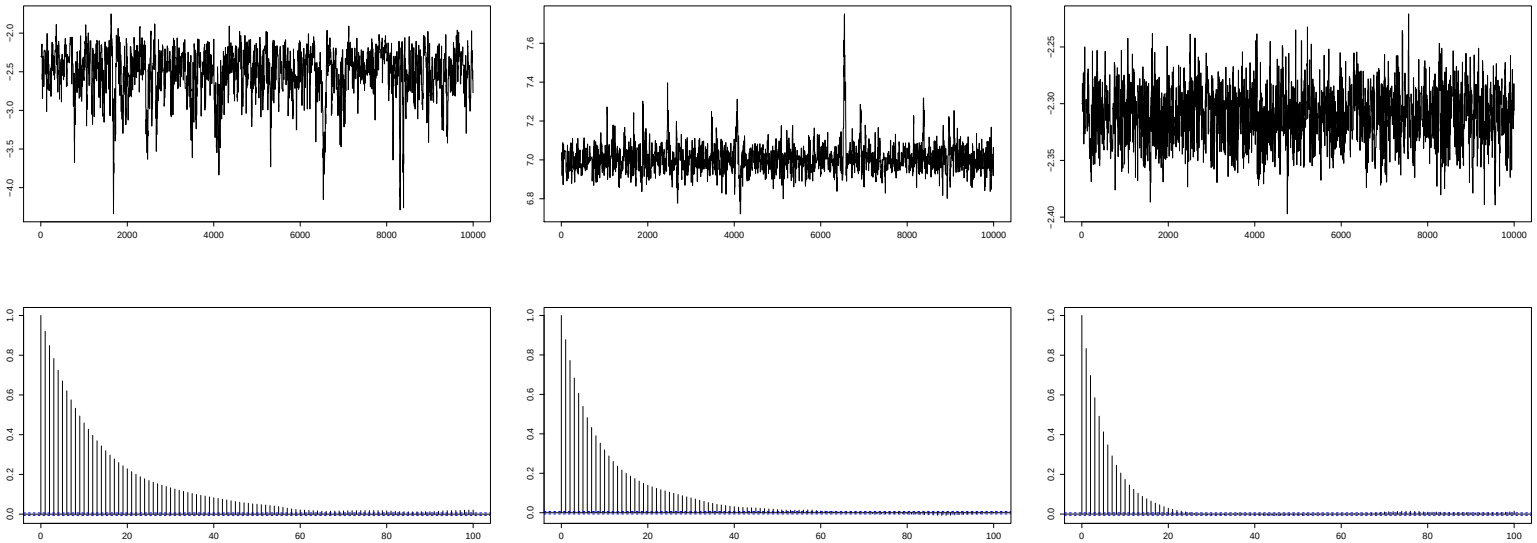

Figure: Trace plots and autocorrelation function for $\log(\rho)$, $\log(\lambda)$, $\log(\sigma)$ using the noncentred parametrization.

Discussion

- The method is exact and the only source of error is from the Monte Carlo simulations.
- Computationally efficient since the missing data are easy to simulate and usually low dimensional.
- Implementation of noncentered reparametrizations and interweaving strategy is straightforward and can significantly increase the speed of convergence.

References

Beskos, A., Papaspiliopoulos, O. and Roberts, G. O. (2006)
Retrospective exact simulation of diffusion sample paths with applications.
Bernoulli, **12**, 1077–1098.

Papaspiliopoulos, O. and Roberts, G. O. (2008)
Retrospective Markov chain Monte Carlo methods for Dirichlet process hierarchical models.
Biometrika, **95**, 169–186.

Papaspiliopoulos, O., Roberts, G. O. and Sköld, M. (2007)
A general framework for the parametrization of hierarchical models.
Statistical Science, **22**, 59–73.

Roberts, G. O., Papaspiliopoulos, O. and Dellaportas, P. (2004) Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility processes.
J. R. Stat. Soc. Ser. B Stat. Methodol., **66**, 369–393.

Roberts, G. O. and Stramer, O. (2001)
On inference for partially observed nonlinear diffusion models using the Metropolis-Hastings algorithm.
Biometrika, **88**, 603–621.