A piecewise deterministic Monte Carlo method for diffusion bridges

Exposition of the kick-off research project of my PhD titled
Bayesian inference for high dimensional diffusion processes

Sebastiano Grazzi[1]

J.Bierkens[1] F. v.d. Meulen[1] M. Schauer[2]

[1]Delft Institute of Applied Mathematics (DIAM)
Delft University of Technology

[2]Mathematical Institute
University of Leiden

PhD seminars, May 7, 2019

Table of Contents

1. Introduction
   - Statement of the problem
2. Body
   - ZigZag sampler
   - Faber-Schauder expansion of diffusion bridges
   - Application
3. Conclusion

Preliminary definitions

Informal definition

A one-dimensional *stochastic differential equation* is a differential equation containing *random forces* defined in a probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$. It can be seen as

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma(t, X_t) \cdot \text{noise}, \qquad X_0 = u. \tag{1}$$

Preliminary definitions

Informal definition

A one-dimensional *stochastic differential equation* is a differential equation containing *random forces* defined in a probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$. It can be seen as

$$\frac{dX_t}{dt} = b(t, X_t) + \sigma(t, X_t) \cdot \text{noise}, \qquad X_0 = u. \tag{1}$$

- If the random force is *'white noise'* W_t, then we write $'dB_t = W_t dt'$ and (1) becomes

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t.$$

  its solution is called *diffusion process.*

- $\forall t, \quad X_t$ is a r.v.. The map $t \to X_t(\omega)$ is almost surely continuous and induces a probability measure on $C(0, T)$.

Statistical inference for diffusion processes

Problem:

Given a finite set of observations $D = (x_0, x_{t_1}, ..., x_{t_n})$ and a parametric model

$$dX_t = b_\theta(t, X_t)dt + \sigma_\theta(t, X_t)dB_t, \qquad X_0 = x_0,$$

estimate $\theta \in \Theta$.

Statistical inference for diffusion processes

Problem:

Given a finite set of observations $D = (x_0, x_{t_1}, ..., x_{t_n})$ and a parametric model

$$dX_t = b_\theta(t, X_t)dt + \sigma_\theta(t, X_t)dB_t, \qquad X_0 = x_0,$$

estimate $\theta \in \Theta$.

The quantity of interest is the likelihood function $L(\theta, D)$

$$L(\theta, D) = P_\theta(D) = p_\theta(X_0) \prod_{i=1}^{n} p_\theta(X_{t_i} | X_{t_{i-1}}),$$

where $p_\theta(\cdot|\cdot)$ is known as *transition density* which generally is not available. What do we know?

Statistical inference for diffusion processes

Problem:

Given a finite set of observations $D = (x_0, x_{t_1}, \dots, x_{t_n})$ and a parametric model

$$dX_t = b_\theta(t, X_t)dt + \sigma_\theta(t, X_t)dB_t, \qquad X_0 = x_0,$$

estimate $\theta \in \Theta$.

The quantity of interest is the likelihood function $L(\theta, D)$

$$L(\theta, D) = P_\theta(D) = p_\theta(X_0)\prod_{i=1}^{n} p_\theta(X_{t_i}|X_{t_{i-1}}),$$

where $p_\theta(\cdot|\cdot)$ is known as *transition density* which generally is not available. What do we know?

Induced measure	Sde
$\mathbb{Q}$	$dX_t = \sigma_\theta(t, X_t)dW_t$
$\mathbb{P}$	$dX_t = b_\theta(t, X_t)dt + \sigma_\theta(t, X_t)dW_t$

Statistical inference for diffusion processes (X3)

Girsanov theorem

Under regularity conditions on b_θ, σ_θ,

$$\frac{d\mathbb{P}}{d\mathbb{Q}}(X_t) = \exp\left(\int_0^t u(s, X_s) dX_s - \frac{1}{2}\int_0^t u(s, X_s)^2 ds\right), \; u(s, x) = \frac{b(s, x)}{\sigma(s, x)}.$$

Statistical inference for diffusion processes (X3)

Girsanov theorem

Under regularity conditions on b_θ, σ_θ,

$$\frac{d\mathbb{P}}{d\mathbb{Q}}(X_t) = \exp\left(\int_0^t u(s, X_s) dX_s - \frac{1}{2}\int_0^t u(s, X_s)^2 ds\right), \; u(s,x) = \frac{b(s,x)}{\sigma(s,x)}.$$

Bayesian approach

Assume $(\theta, (X_t)_{t\in T})$ to be jointly measurable in a probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$, find $P((X_t)_{t\in T}, \theta | D)$.

The joint probability of (X, θ) has to be computed numerically with *MCMC methods* on $(\theta^{(i)}, X^{(i)})$ in which, after initialization, we alternate the following steps. for $i = 1, 2, \ldots$

1. draw $\theta^{(i)} \sim P(\theta | D, X^{(i-1)})$
2. draw $X^{(i)} \sim P(X | D, \theta^{(i)})$

Diffusion bridge

Definition

A Diffusion Bridge is a stochastic differential equation with boundary conditions on the left and on the right:

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \quad X_0 = u, \quad X_T = v.$$

Its measure $\mathbb{P}^*$ is given by the original measure $\mathbb{P}$ conditioned to hit its final point at time T.

Diffusion bridge

Definition

A Diffusion Bridge is a stochastic differential equation with boundary conditions on the left and on the right:

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \quad X_0 = u, \quad X_T = v.$$

Its measure $\mathbb{P}^*$ is given by the original measure $\mathbb{P}$ conditioned to hit its final point at time T.

Bayesian rule for diffusion bridges

$$\frac{d\mathbb{P}}{d\mathbb{Q}}(X) = \frac{d\mathbb{P}^*}{d\mathbb{Q}^*}(X)\frac{\tilde{p}(X_T = x_T)}{p(X_T = x_T)}$$

where $p(X_T = x_T)$, $\tilde{p}(X_T = x_T)$ are probability densities of the random variables X_T induced respectively by $\mathbb{P}, \mathbb{Q}$.

Outline

Remarks:

- The measure $\mathbb{P}^*$ is known up to normalization factor.
- $\mathbb{P}^*$ is defined on $C(0, T)$ (infinite dimensional).

Outline

Remarks:

- The measure $\mathbb{P}^*$ is known up to normalization factor.
- $\mathbb{P}^*$ is defined on $C(0, T)$ (infinite dimensional).

Plan:

- ZigZag sampler: a Monte Carlo method based on piecewise deterministic Markov process.
- Ciesielski construction of Brownian motion and extension to any diffusion processes $\mathbb{P}$.
- Application

Piecewise deterministic Markov process

Davis (1993)

The PDMP is a process whose behaviour is governed by random jumps at points in time, but whose evolution is deterministically governed by an ordinary differential equation between those times.

Figure: the process is fully described by a ODe (1), a Poisson rate characterizing the random jump events (2), the Markov transition kernel (3).

Can we choose (1),(2),(3) such that the process has a desired stationary distribution?

ZigZag sampler

Bierkens et al. (2019)

The One-dimensional ZigZag sampler is defined in the *augmented space* $(\xi, \theta) \in Z = (\mathbb{R} \otimes \{+1, -1\})$.

ZigZag sampler

Bierkens et al. (2019)

The One-dimensional ZigZag sampler is defined in the *augmented space* $(\xi, \theta) \in Z = (\mathbb{R} \otimes \{+1, -1\})$.

- its flow is linear with direction θ on its first argument and constant on θ: $\Phi_t(\xi, \theta) = (\xi + t\theta, \theta)$

ZigZag sampler

Bierkens et al. (2019)

The One-dimensional ZigZag sampler is defined in the *augmented space* $(\xi, \theta) \in Z = (\mathbb{R} \otimes \{+1, -1\})$.

- its flow is linear with direction θ on its first argument and constant on θ: $\Phi_t(\xi, \theta) = (\xi + t\theta, \theta)$
- at event time τ,the process jumps only on θ changing its sign: $Q(\xi_{\tau-}, \theta_{\tau-}) = (\xi_\tau, -\theta_\tau)$

ZigZag sampler

Bierkens et al. (2019)

The One-dimensional ZigZag sampler is defined in the *augmented space* $(\xi, \theta) \in Z = (\mathbb{R} \otimes \{+1, -1\})$.

- its flow is linear with direction θ on its first argument and constant on θ: $\quad \Phi_t(\xi, \theta) = (\xi + t\theta, \theta)$
- at event time τ,the process jumps only on θ changing its sign: $Q(\xi_{\tau-}, \theta_{\tau-}) = (\xi_\tau, -\theta_\tau)$
- the event rate is defined by a Poisson rate $\lambda : Z \to \mathbb{R}^+$ such that $P(\tau \in [t, t+\epsilon]) = \lambda(\xi_t, \theta_t) + o(\epsilon)$

Condition for the process to target the stationary distribution

Assume the target distribution to have strictly positive density with respect to the Lebesgue measure, i.e. the its density can be written as $\pi(\xi) = exp(-\psi(\xi)), \forall \xi$.

Theorem

The ZigZag process satisfying

$$\lambda(\xi, \theta) - \lambda(\xi, -\theta) = \theta \partial_\xi \psi(\xi), \quad \forall (\xi, \theta) \in \mathbb{R} \otimes \{1, -1\} \tag{2}$$

has π as stationary density.

Remarks:

- condition (2) $\iff \lambda(\xi, \theta) = (\theta \partial_\xi \psi(\xi))^+ + \gamma(\xi)$ with $\gamma(x) \geq 0$
- the algorithm uses only local knowledge of the probability distribution
- stationary condition does not rely on *detailed balance*

Example: Gaussian density

- $\pi(\xi) \propto \exp(\frac{(\xi-\mu)^2}{2\sigma^2})$
- $\partial_\xi \psi(\xi) = \frac{(\xi-\mu)}{\sigma^2}$
- $\lambda(\xi, \theta) = (\frac{\theta(\xi-\mu)}{\sigma^2})^+$

Extensions: d-dimensional ZigZag process:

- Define the d-dimensional ZigZag process in $(\xi, \theta) \in \mathbb{R}^d \otimes \{+1, -1\}^d$

Extensions: d-dimensional ZigZag process:

- Define the d-dimensional ZigZag process in $(\xi, \theta) \in \mathbb{R}^d \otimes \{+1, -1\}^d$
- Define the flipping function as $F_i(\theta) = (\theta_1, \theta_2, ..., -\theta_i, ..., \theta_d)$

Extensions: d-dimensional ZigZag process:

- Define the d-dimensional ZigZag process in $(\xi, \theta) \in \mathbb{R}^d \otimes \{+1, -1\}^d$
- Define the flipping function as $F_i(\theta) = (\theta_1, \theta_2, ..., -\theta_i, ..., \theta_d)$
- Condition (2) becomes:

$$\lambda(\xi, \theta) - \lambda(\xi, -\theta) = \langle \theta \nabla \psi(\xi) \rangle,$$

$$\sum_{i=1}^{n} \lambda_i(\xi, \theta) - \lambda_i(\xi, F_i(\theta)) = \sum_{i=1}^{n} \theta_i \partial_{\xi_i} \psi(\xi), \quad \forall \xi, \theta \tag{3}$$

A sufficient condition for (3) to hold is $\lambda_i(\xi, \theta) = (\theta_i \partial_{\xi_i} \psi(\xi))^+$.

Extensions: d-dimensional ZigZag process:

- Define the d-dimensional ZigZag process in $(\xi, \theta) \in \mathbb{R}^d \otimes \{+1, -1\}^d$
- Define the flipping function as $F_i(\theta) = (\theta_1, \theta_2, ..., -\theta_i, ..., \theta_d)$
- Condition (2) becomes:

$$\lambda(\xi, \theta) - \lambda(\xi, -\theta) = \langle \theta \nabla \psi(\xi) \rangle,$$

$$\sum_{i=1}^{n} \lambda_i(\xi, \theta) - \lambda_i(\xi, F_i(\theta)) = \sum_{i=1}^{n} \theta_i \partial_{\xi_i} \psi(\xi), \quad \forall \xi, \theta \tag{3}$$

A sufficient condition for (3) to hold is $\lambda_i(\xi, \theta) = (\theta_i \partial_{\xi_i} \psi(\xi))^+$.

Superposition theorem

Given countable many independent Poisson processes $\Pi_1, \Pi_2, ...$ with rates $\lambda_1, \lambda_2, ...$ If $\sum_{i=1}^{\infty} \lambda_i < \infty$, then

$$\Pi = \cup_{i=1}^{\infty} \Pi_i$$

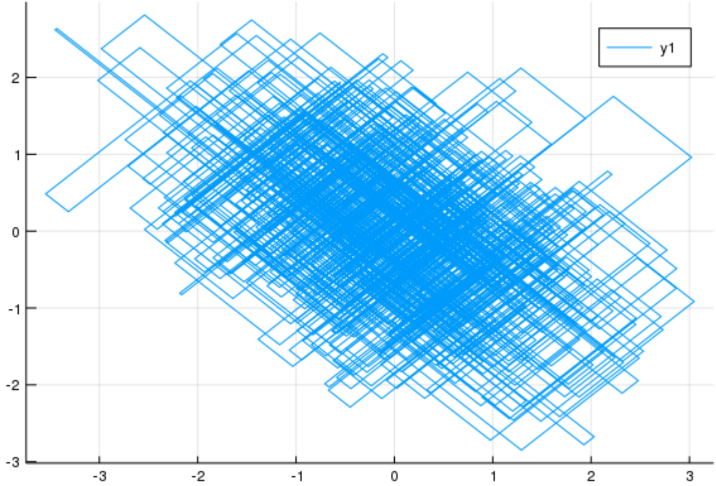
is again a Poisson Process with rate function equal to $\lambda = \sum_{i=1}^{\infty} \lambda_i$.

Algorithm

Algorithm 1 ZigZag sampler

procedure ZZSAMPLER(T)
  Inizialize $t, \boldsymbol{\theta}^d, \boldsymbol{\xi}^d, \boldsymbol{\tau}^d$, i $= 1$
  while $t \leq T$ **do**
    $\tau_0, i_0 \leftarrow$ findmin$(\boldsymbol{\tau}^d)$
    Update: $\boldsymbol{\xi}^d \leftarrow \boldsymbol{\xi}^d + \boldsymbol{v}^d \tau_0$
    Save ξ_i, t_i
    $\boldsymbol{v} \leftarrow F_{i_0}(\boldsymbol{v}^{\boldsymbol{d}})$
    for j in $(1 : d)$ **do**
      Draw $\tau[j]$ from IPP$(\lambda_j(\xi, v))$ ▷ Inhomogeneous Poisson process (IIP)
    end for
    $i \leftarrow i + 1$
  end while
  return Skeletons $((\xi_i, t_i)_{i=1,2,\dots N})$.
end procedure

Plots



two-dimensional Gaussian random variable with negative correlation

Ciesielski construction of Brownian motion:
Davis (2004)

Denote with $\eta_1, \eta_2, ...$ a complete orthonormal basis in $L_2(0,1)$. Define $\xi_1, \xi_2, ...$ to be i.i.d Gaussian r.v defined in a probability space $(\Omega, \mathcal{F}, P)$ and

$$X_N(t) = \sum_{i=1}^{N} \xi_i \int_0^t \eta_i(s)ds$$

Ciesielski construction of Brownian motion:
Davis (2004)

Denote with $\eta_1, \eta_2, ...$ a complete orthonormal basis in $L_2(0,1)$. Define $\xi_1, \xi_2, ...$ to be i.i.d Gaussian r.v defined in a probability space $(\Omega, \mathcal{F}, P)$ and

$$X_N(t) = \sum_{i=1}^{N} \xi_i \int_0^t \eta_i(s) ds$$

Theorem 1

For all t, $(X^N(t))_{N=1,2,...}$ is a Cauchy sequence in $L_2(\Omega)$.
$X_t = \lim_{N\to\infty} X^N(t)$ is a Gaussian random variable centered in 0 with $E(X_t^2) = t$ and for any s, t we have $E[X_s, X_t] = s \wedge t$

Ciesielski construction of Brownian motion:

Davis (2004)

Denote with $\eta_1, \eta_2, ...$ a complete orthonormal basis in $L_2(0,1)$. Define $\xi_1, \xi_2, ...$ to be i.i.d Gaussian r.v defined in a probability space $(\Omega, \mathcal{F}, P)$ and

$$X_N(t) = \sum_{i=1}^{N} \xi_i \int_0^t \eta_i(s) ds$$

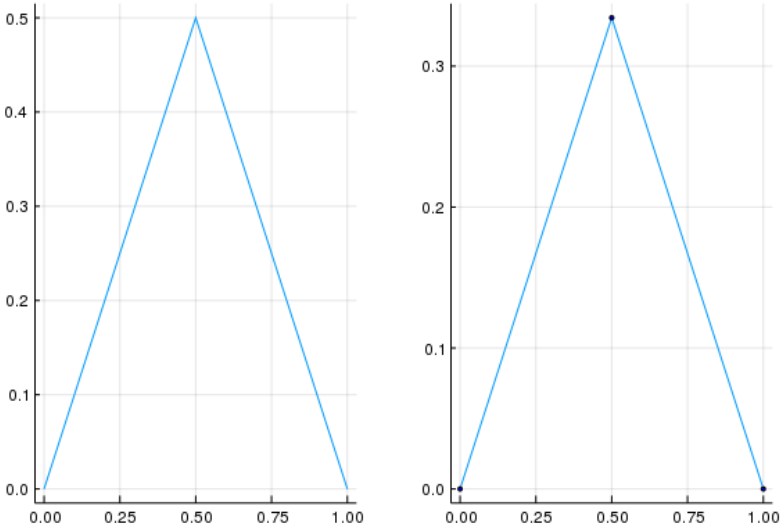
Theorem 1

For all t, $(X^N(t))_{N=1,2,...}$ is a Cauchy sequence in $L_2(\Omega)$.
$X_t = \lim_{N \to \infty} X^N(t)$ is a Gaussian random variable centered in 0 with $E(X_t^2) = t$ and for any s, t we have $E[X_s, X_t] = s \wedge t$

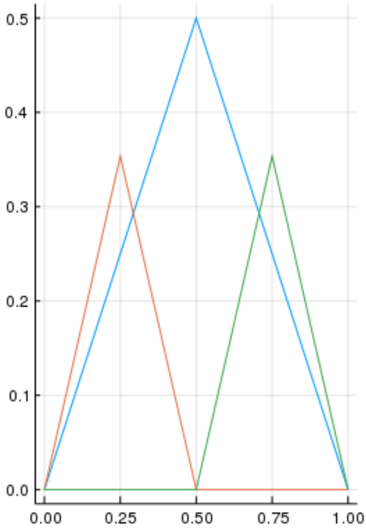
Theorem 2

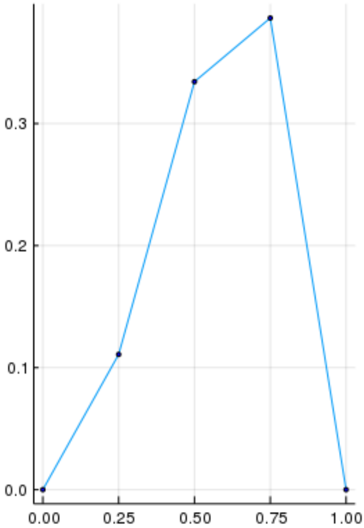
If we choose $\eta_1, \eta_2, ...$ to be the Haar functions. Then $t \to X_t$ is almost surely continuous.

Ciesielski construction of Brownian motion

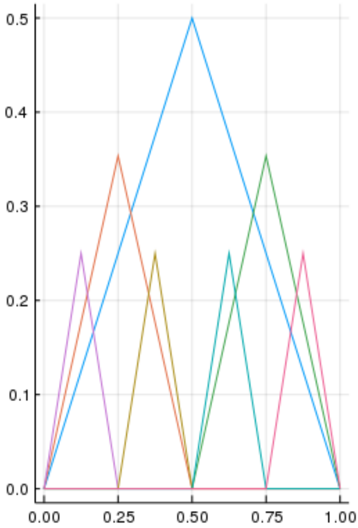


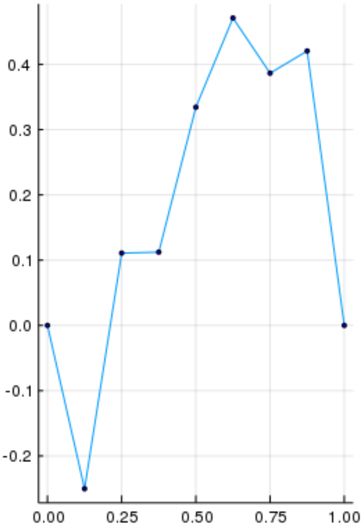
Ciesielski construction of Brownian motion



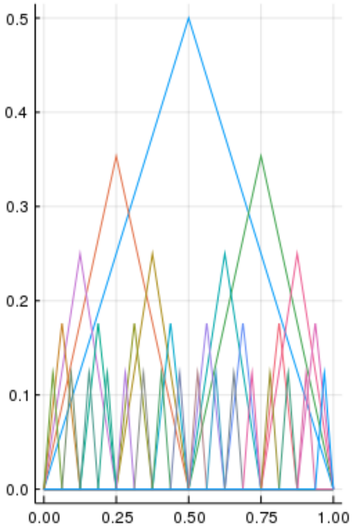


Ciesielski construction of Brownian motion





Ciesielski construction of Brownian motion



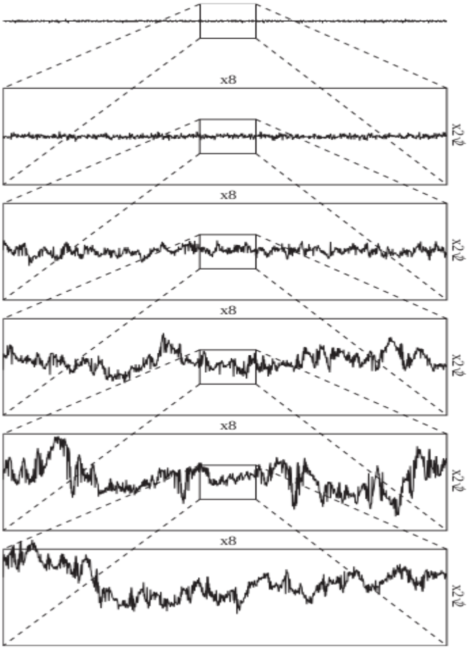


Figure: (image taken from Taillefumier and Magnasco (2008)). Under regularity assumption on the b and σ, the diffusion process behaves locally as a Brownian motion. If we expand the diffusion process with the Faber-Schauder functions, we expect $(\xi_{N+1}, \xi_{N+2}, \ldots)$ to remain i.i.d Gaussian for a large N.

Extension: approximated stochastic integral

Consider the Radon Nikodym derivative:

$$\frac{d\mathbb{P}}{d\mathbb{Q}}(X_t) = \exp\left(\int_0^t u(s, X_s) dX_s - \frac{1}{2}\int_0^t u(s, X_s)^2 ds\right),\ u(s, x) = \frac{b(s, x)}{\sigma(s, x)}.$$

Extension: approximated stochastic integral

Consider the Radon Nikodym derivative:

$$\frac{d\mathbb{P}}{d\mathbb{Q}}(X_t) = \exp\left(\int_0^t u(s, X_s) dX_s - \frac{1}{2}\int_0^t u(s, X_s)^2 ds\right), \; u(s, x) = \frac{b(s, x)}{\sigma(s, x)}.$$

- Naive approach: define the approximated stochastic process $\mathbb{P}_N$ through the change of measure

$$\frac{d\mathbb{P}_N}{d\mathbb{Q}}(\xi_1, \ldots, \xi_N) = \exp\left(\int_0^t u(s, X_s^N) d(X_s^N) - \frac{1}{2}\int_0^t u(s, X_s^N)^2 ds\right),$$

$$u(s, x) = \frac{b(s, x)}{\sigma(s, x)}, \qquad X_t^N = \sum_{i=1}^N \xi_i \int_0^t \eta_i(s) ds.$$

Extension: approximated stochastic integral

Consider the Radon Nikodym derivative:

$$\frac{d\mathbb{P}}{d\mathbb{Q}}(X_t) = \exp\left(\int_0^t u(s, X_s)dX_s - \frac{1}{2}\int_0^t u(s, X_s)^2 ds\right), \; u(s, x) = \frac{b(s, x)}{\sigma(s, x)}.$$

- Naive approach: define the approximated stochastic process $\mathbb{P}_N$ through the change of measure

$$\frac{d\mathbb{P}_N}{d\mathbb{Q}}(\xi_1, ..., \xi_N) = \exp\left(\int_0^t u(s, X_s^N)d(X_s^N) - \frac{1}{2}\int_0^t u(s, X_s^N)^2 ds\right),$$

$$u(s, x) = \frac{b(s, x)}{\sigma(s, x)}, \qquad X_t^N = \sum_{i=1}^N \xi_i \int_0^t \eta_i(s)ds.$$

- Improved approached: assume $u(s, X_s) = u(X_s)$ to be differentiable and bounded. Apply Ito differentiation rule to $f(x_t) = \int_0^{X_t} u(s)ds$.

Independence structure created by $\mathbb{P}_N$

Define $\phi_i(t) = \int_0^t \eta_i(s)ds$, being the integrated orthonormal basis and

$$S_i = \mathrm{support}(\phi_i(t))$$

and the indexes of the basis which overlap the support of the basis ϕ_{ij} as

$$N_i = \{j : S_i \cap S_j \neq \emptyset\},$$

Independence structure created by $\mathbb{P}_N$

Define $\phi_i(t) = \int_0^t \eta_i(s)ds$, being the integrated orthonormal basis and

$$S_i = \text{support}(\phi_i(t))$$

and the indexes of the basis which overlap the support of the basis ϕ_{ij} as

$$N_i = \{j : S_i \cap S_j \neq \emptyset\},$$

then the joint probability $\mathbb{P}_n$ can be factorized as

$$\mathbb{P}_n(\xi_1, ..., \xi_N) = \prod_{j=1}^{N} \exp(f_j((\xi_i)_{i\in N_j}))$$

creating a dependence structure among the vector $(\xi_1, \xi_2, ..., \xi_N)$ such that: $\xi_i \perp \xi_j \quad \text{if} \quad S_i \cap S_j = \emptyset$.

Independence structure created by $\mathbb{P}_N$

Define $\phi_i(t) = \int_0^t \eta_i(s)ds$, being the integrated orthonormal basis and

$$S_i = \text{support}(\phi_i(t))$$

and the indexes of the basis which overlap the support of the basis ϕ_{ij} as

$$N_i = \{j : S_i \cap S_j \neq \emptyset\},$$

then the joint probability $\mathbb{P}_n$ can be factorized as

$$\mathbb{P}_n(\xi_1, ..., \xi_N) = \prod_{j=1}^{N} \exp(f_j((\xi_i)_{i \in N_j}))$$

creating a dependence structure among the vector $(\xi_1, \xi_2, ..., \xi_N)$ such that: $\xi_i \perp \xi_j \quad \text{if} \quad S_i \cap S_j = \emptyset$.

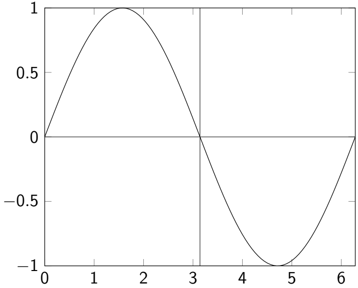
The complexity of the algorithm grows sub-linearly with the dimensions!.

First results of the experiment

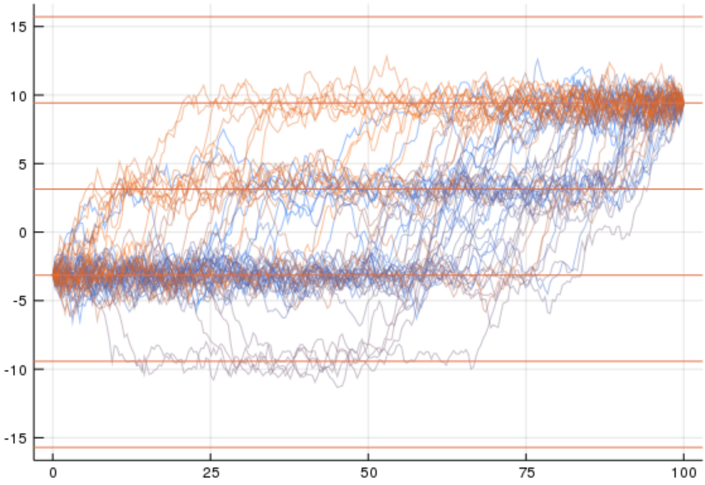
We considered a high non-linear diffusion bridge:

$$dX_t = \alpha \sin(X_t)dt + \sigma dW_t, \qquad X_0 = u, X_T = v$$

The relative ordinary differential equation has attraction points on $(2k-1)\phi$ with $k \in \mathbb{N}$.



Results (X1)



$dX_t = \alpha \sin(X_t)dt + \sigma dW_t$ starting from $-\pi$ and hitting 2ϕ at $t = 100$. $\alpha = 0.7, \sigma = 1.0$

Results (X2)

Open from:

https://media.giphy.com/media/KGZGoVCKgfUSYLtZeu/giphy.gif

Conclusion

What we did:

- set up a new methodology for computing diffusion bridges which looks promising.
- found and exploited the Independence structure implied by the Faber Schauder expansion in the algorithm allowing for high dimensional representations of the diffusion Bridge.

What has to be done:

- proving convergence of $\mathbb{P}_N \to \mathbb{P}$
- how to set the velocities in order to converge faster
- generalize to diffusion bridges in $\mathbb{R}^d$
- understand limits and applicability of this algorithm

Thank you for the attention! Questions?

References I

Bierkens, J., Fearnhead, P., and Roberts, G. (2019). The zig-zag process and super-efficient sampling for bayesian analysis of big data. *Ann. Statist.*, 47(3):1288–1320.

Davis, M. (1993). *Markov Models & Optimization*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis.

Davis, M. (2004). Construction of Brownian Motion. University Lecture.

Taillefumier, T. and Magnasco, M. O. (2008). A haar-like construction for the ornstein uhlenbeck process. *Journal of Statistical Physics*, 132(2):397–415.