Likelihood Construction for discretely observed RDEs

Anastasia Papavasiliou (Warwick University)
Joint work with
K. Taylor (National Grid) and T. Papamarkou (Glasgow)
Supported by the Leverhulme Trust

July 18, 2017

Motivation

- Data from processes modelled as diffusions does not exhibit finite quadratic variation (ex. frequently traded stocks, molecular dynamics).

Motivation

- Data from processes modelled as diffusions does not exhibit finite quadratic variation (ex. frequently traded stocks, molecular dynamics).
- Statistical inference methodology for fractional diffusions is still under developed.

Motivation

- Data from processes modelled as diffusions does not exhibit finite quadratic variation (ex. frequently traded stocks, molecular dynamics).
- Statistical inference methodology for fractional diffusions is still under developed.
- Initial motivation: speech recognition!

The problem

- We consider the following type of differential equations

$$dY_t = a(Y_t;\theta)dt + b(Y_t;\theta)dX_t, \quad Y_0 = y_0, \; t \leq T,$$

where $X \in G\Omega_p(\mathbb{R}^m)$ is a realization of a random geometric p-rough path defined as the limit of a random sequence $(\pi_n(X))_{n>0}$ of nested piecewise linear paths, that we assume to converge almost surely in the p-variation topology. We also assume that a and b satisfy the usual conditions.

The problem

- We consider the following type of differential equations

$$dY_t = a(Y_t;\theta)dt + b(Y_t;\theta)dX_t, \quad Y_0 = y_0, \ t \leq T,$$

  where $X \in G\Omega_p(\mathbb{R}^m)$ is a realization of a random geometric p-rough path defined as the limit of a random sequence $(\pi_n(X))_{n>0}$ of nested piecewise linear paths, that we assume to converge almost surely in the p-variation topology. We also assume that a and b satisfy the usual conditions.
- We want to construct the likelihood of observing

$$y_{\mathcal{D}(n)} := \{y_{t_i} \in \mathbb{R}^d; \ t_i \in \mathcal{D}(n)\},$$

  when we know the distribution of X.

The Challenge

- We need to match the infinite dimensional information on the distribution of X with the finite-dimensional information on Y.

The Challenge

- We need to match the infinite dimensional information on the distribution of X with the finite-dimensional information on Y.
- The Girsanov + MCMC approach.

The Challenge

- We need to match the infinite dimensional information on the distribution of X with the finite-dimensional information on Y.
- The Girsanov + MCMC approach.
- The Euler approach.

The Challenge

- We need to match the infinite dimensional information on the distribution of X with the finite-dimensional information on Y.
- The Girsanov $+$ MCMC approach.
- The Euler approach.
- The Projection $+$ solution of the inverse problem approach.

After projection: an approximate problem

- Let $\mathcal{D}$ be a fixed grid of $[0, T]$ and $X^{\mathcal{D}}$ a piecewise linear path on $\mathcal{D}$. Let $Y^{\mathcal{D}}$ be the corresponding response, i.e.

$$dY_t^{\mathcal{D}} = a(Y_t^{\mathcal{D}};\theta)dt + b(Y_t^{\mathcal{D}};\theta)dX_t^{\mathcal{D}}, \quad Y_0 = y_0, \ t \leq T,$$

After projection: an approximate problem

- Let $\mathcal{D}$ be a fixed grid of $[0, T]$ and $X^{\mathcal{D}}$ a piecewise linear path on $\mathcal{D}$. Let $Y^{\mathcal{D}}$ be the corresponding response, i.e.

$$dY_t^{\mathcal{D}} = a(Y_t^{\mathcal{D}}; \theta)dt + b(Y_t^{\mathcal{D}}; \theta)dX_t^{\mathcal{D}}, \quad Y_0 = y_0, \ t \leq T,$$

- As before, we assume we observe $Y^{\mathcal{D}}$ on the grid $\mathcal{D}$, denoted by $y_{\mathcal{D}}$ and we know the distribution of $X^{\mathcal{D}}$.

After projection: an approximate problem

- Let $\mathcal{D}$ be a fixed grid of $[0, T]$ and $X^{\mathcal{D}}$ a piecewise linear path on $\mathcal{D}$. Let $Y^{\mathcal{D}}$ be the corresponding response, i.e.

$$dY_t^{\mathcal{D}} = a(Y_t^{\mathcal{D}}; \theta)dt + b(Y_t^{\mathcal{D}}; \theta)dX_t^{\mathcal{D}}, \quad Y_0 = y_0, \ t \leq T,$$

- As before, we assume we observe $Y^{\mathcal{D}}$ on the grid $\mathcal{D}$, denoted by $y_{\mathcal{D}}$ and we know the distribution of $X^{\mathcal{D}}$.
- In this case, the likelihood of $y_{\mathcal{D}}$ can be constructed exactly.

Likelihood construction for the approximate problem: Main idea

- By only considering piecewise linear drivers, the problem becomes finite dimensional. The idea is to express the data as a function of the increments of the piecewise linear path $\Delta X^{\mathcal{D}}$, i.e. we can write

$$y_{\mathcal{D}} = J_{\mathcal{D}}(\Delta X^{\mathcal{D}}; \theta)$$

Likelihood construction for the approximate problem: Main idea

- By only considering piecewise linear drivers, the problem becomes finite dimensional. The idea is to express the data as a function of the increments of the piecewise linear path $\Delta X^{\mathcal{D}}$, i.e. we can write

$$y_{\mathcal{D}} = J_{\mathcal{D}}(\Delta X^{\mathcal{D}}; \theta)$$

- Then, the likelihood can be written as

$$L_{Y^{\mathcal{D}}}(y_{\mathcal{D}}|\theta) = L_{\Delta X^{\mathcal{D}}}(J_{\mathcal{D}}^{-1}(y_{\mathcal{D}}; \theta)) \cdot |DJ_{\mathcal{D}}^{-1}(y_{\mathcal{D}}; \theta)|.$$

assuming that

$$\Delta X^{\mathcal{D}} = J_{\mathcal{D}}^{-1}(y_{\mathcal{D}}; \theta)$$

exists and is uniquely defined.

Construction of the inverse Itô map $J_{\mathcal{D}}^{-1}$

- We want to compute $\Delta X^{\mathcal{D}}$ as a function of the data and the parameter θ.

Construction of the inverse Itô map $J_{\mathcal{D}}^{-1}$

- We want to compute $\Delta X^{\mathcal{D}}$ as a function of the data and the parameter θ.
- By definition, $Y^{\mathcal{D}}$ satisfies

$$dY_t^{\mathcal{D}} = \left(a(Y_t^{\mathcal{D}};\theta)dt + b(Y_t^{\mathcal{D}};\theta)\Delta X_{t_i}\right) dt$$

with initial conditions $Y_{t_i} = y_{t_i}$.

Construction of the inverse Itô map $J_{\mathcal{D}}^{-1}$

- We want to compute $\Delta X^{\mathcal{D}}$ as a function of the data and the parameter θ.
- By definition, $Y^{\mathcal{D}}$ satisfies

$$dY_t^{\mathcal{D}} = \left(a(Y_t^{\mathcal{D}};\theta)dt + b(Y_t^{\mathcal{D}};\theta)\Delta X_{t_i}\right) dt$$

  with initial conditions $Y_{t_i} = y_{t_i}$.
- This is an ODE and we have already assumed sufficient regularity on a and b for existence and uniqueness of its solutions. The general form of the ODE is given by

$$d\tilde{Y}_t = \left(a(\tilde{Y}_t;\theta) + b(\tilde{Y}_t;\theta)\cdot c\right) dt,\, Y_0 = y_0$$

  and is solution is denoted by $F_t(y_0, c;\theta)$. Then,

$$Y_t^{\mathcal{D}} = F_{t-t_i}(y_{t_i}, \Delta X_{t_i};\theta),\ \forall t \in [t_i, t_{i+1}).$$

Construction of the inverse Itô map $J_{\mathcal{D}}^{-1}$ cont'd

- The next step is to solve for ΔX_{t_i}, using the terminal value, i.e. solve

$$F_{t_{i+1}-t_i}(y_{t_i}, \Delta X_{t_i}; \theta) = y_{t_{i+1}}$$

  for $\Delta X_{t_i}(y_{t_i}, y_{t_{i+1}}; \theta)$. So, for every interval $[t_i, t_{i+1})$, we need to solve an independent system of d equations and m unknowns.

Construction of the inverse Itô map $J_{\mathcal{D}}^{-1}$ cont'd

- The next step is to solve for ΔX_{t_i}, using the terminal value, i.e. solve
$$F_{t_{i+1}-t_i}(y_{t_i}, \Delta X_{t_i}; \theta) = y_{t_{i+1}}$$
for $\Delta X_{t_i}(y_{t_i}, y_{t_{i+1}}; \theta)$. So, for every interval $[t_i, t_{i+1})$, we need to solve an independent system of d equations and m unknowns.
- A natural question to ask here is whether solutions to this system exist and are unique.

Construction of the inverse Itô map $J_{\mathcal{D}}^{-1}$ cont'd

- The next step is to solve for ΔX_{t_i}, using the terminal value, i.e. solve
$$F_{t_{i+1}-t_i}(y_{t_i}, \Delta X_{t_i}; \theta) = y_{t_{i+1}}$$
for $\Delta X_{t_i}(y_{t_i}, y_{t_{i+1}}; \theta)$. So, for every interval $[t_i, t_{i+1})$, we need to solve an independent system of d equations and m unknowns.
- A natural question to ask here is whether solutions to this system exist and are unique.
- We are going to assume existence of solution, by requiring that $y_{t_{i+1}} \in \cap_{\theta \in \Theta} \mathcal{M}_{t_{i+1}-t_i}(y_{t_i}; \theta)$, where
$$\mathcal{M}_\delta(y_0; \theta) = \{F_\delta(y_0, c; \theta); c \in \mathbb{R}^m\}.$$

The auxilliary process Z

- We define a new auxiliary process as $Z_t(c) = D_c F_t(y_0, c; \theta) \in \mathbb{R}^{d \times m}$, or,

$$Z_t^{i,\alpha}(c) = \frac{\partial}{\partial c_\alpha} F_t^i(y_0, c; \theta), \text{ for } i = 1, \ldots, d, \ \alpha = 1, \ldots, m.$$

The auxilliary process Z

- We define a new auxiliary process as $Z_t(c) = D_c F_t(y_0, c; \theta) \in \mathbb{R}^{d \times m}$, or,

$$Z_t^{i,\alpha}(c) = \frac{\partial}{\partial c_\alpha} F_t^i(y_0, c; \theta), \text{ for } i = 1, \ldots, d, \ \alpha = 1, \ldots, m.$$

- Then, under suitable regularity conditions, $Z_t(c)$ satisfies

$$\frac{d}{dt} \bar{Z}_t^\alpha(c) = \bigtriangledown (a + b \cdot c)(F_t) \cdot \bar{Z}_t^\alpha(c) + \bar{b}_\alpha(F_t),$$

with initial conditions $Z_0(c) \equiv 0$, where by $\bar{Z}_t^\alpha(c)$ and $\bar{b}^\alpha(y)$ we denote column $\alpha \in \{1, \ldots, m\}$ of matrix $Z_t(c)$ and $b(y)$ respectively.

The auxilliary process Z

- We conclude that

$$\bar{Z}_t^{\alpha}(c) = \int_0^t \exp\left(\mathbf{A}\right)_{s,t} \bar{b}_{\alpha}(F_s) ds,$$

where by $\exp\left(A\right)_{s,t}$ we denote the sum of iterated integrals

$$\exp\left(\mathbf{A}\right)_{s,t} = \sum_{k=0}^{\infty} \mathbf{A}_{s,t}^k$$

and

$$\mathbf{A}_{s,t}^k = \int \cdots \int_{s<u_1<\cdots<u_k<t} A(F_{u_1}) \cdots A(F_{u_k}) du_1 \dots du_k$$

for $A(y) = \triangledown\left(a + b \cdot c\right)(y)$.

Uniqueness

- We are studying the uniqueness of the solution to the general system of equation

$$F_\delta(y, c_1; \theta) = y'.$$

Uniqueness

- We are studying the uniqueness of the solution to the general system of equation

$$F_\delta(y, c_1; \theta) = y'.$$

- Note that if c_1 and c_2 are both solutions, then we can write

$$\left(\int_0^1 D_c F_\delta(y, c_1 + s(c_2 - c_1); \theta) ds \right) (c_2 - c_1) = 0.$$

Uniqueness

- We are studying the uniqueness of the solution to the general system of equation

$$F_\delta(y, c_1; \theta) = y'.$$

- Note that if c_1 and c_2 are both solutions, then we can write

$$\left(\int_0^1 D_c F_\delta(y, c_1 + s(c_2 - c_1); \theta) ds \right) (c_2 - c_1) = 0.$$

- If the rank of $d \times m$ matrix $D_c F_\delta$ is always d, then we have $m - d$ degrees of freedom. I.e. if we specify $m - d$ coordinates of c, then the rest of the coordinates are uniquely defined.

Uniqueness

- We are studying the uniqueness of the solution to the general system of equation

$$F_\delta(y, c_1; \theta) = y'.$$

- Note that if c_1 and c_2 are both solutions, then we can write

$$\left(\int_0^1 D_c F_\delta(y, c_1 + s(c_2 - c_1); \theta) ds\right)(c_2 - c_1) = 0.$$

- If the rank of $d \times m$ matrix $D_c F_\delta$ is always d, then we have $m - d$ degrees of freedom. I.e. if we specify $m - d$ coordinates of c, then the rest of the coordinates are uniquely defined.
- Since $Z_t(c) = D_c F_t(y, c; \theta)$, we see that the rank will be equal to the rank of b.

Likelihood Construction

- We will construct the likelihood for the case $m = d$. If $m > d$, then the construction is similar, conditioning on the values of the free coordinates of c.

Likelihood Construction

- We will construct the likelihood for the case $m = d$. If $m > d$, then the construction is similar, conditioning on the values of the free coordinates of c.
- We can write

$$L_{Y^{\mathcal{D}}}\left(y_{\mathcal{D}}|\theta\right) = L_{\Delta X^{\mathcal{D}}}\left(J_{\mathcal{D}}^{-1}(y_{\mathcal{D}})\right)|DJ_{\mathcal{D}}^{-1}(y_{\mathcal{D}})|,$$

  where by $L_{\Delta X^{\mathcal{D}}}(\Delta x_{\mathcal{D}(n)})$ we denote the likelihood of observing a realisation of the piecewise linear path $X^{\mathcal{D}}$ with increments $\{\Delta x_{t_i},\ t_i \in \mathcal{D}\}$.

Likelihood Construction

- We will construct the likelihood for the case $m = d$. If $m > d$, then the construction is similar, conditioning on the values of the free coordinates of c.
- We can write

$$L_{Y^{\mathcal{D}}}\left(y_{\mathcal{D}}|\theta\right) = L_{\Delta X^{\mathcal{D}}}\left(J_{\mathcal{D}}^{-1}(y_{\mathcal{D}})\right)|DJ_{\mathcal{D}}^{-1}(y_{\mathcal{D}})|,$$

  where by $L_{\Delta X^{\mathcal{D}}}(\Delta x_{\mathcal{D}(n)})$ we denote the likelihood of observing a realisation of the piecewise linear path $X^{\mathcal{D}}$ with increments $\{\Delta x_{t_i},\ t_i \in \mathcal{D}\}$.
- Since ΔX_{t_i} only depends on y_{t_i} and $y_{t_{i+1}}$, the Jacobian matrix will be block lower triangular and consequently:

$$|DJ_{\mathcal{D}}^{-1}(y_{\mathcal{D}(n)})| = \prod_{t_i \in \mathcal{D}(n)} \left| \nabla \Delta X_{t_i}(y_{t_i}, y; \theta)|_{y=y_{t_{i+1}}} \right|.$$

Likelihood Construction cont'd

- Note that, by definition,

$$F_{t_{i+1}-t_i}(y_{t_i}, \Delta X_{t_i}(y_{t_i}, y; \theta); \theta) \equiv y.$$

Thus,

$$D_c F_{t_{i+1}-t_i}(y_{t_i}, c; \theta)|_{c=\Delta X_{t_i}(y_{t_i}, y_{t_{i+1}}; \theta)} \cdot \nabla \Delta X_{t_i}(y_{t_i}, y; \theta)|_{y=y_{t_{i+1}}} \equiv I_d$$

and, consequently,

$$\begin{aligned} \nabla \Delta X_{t_i}(y_{t_i}, y; \theta)|_{y=y_{t_{i+1}}} &= \left(D_c F_{t_{i+1}-t_i}(y_{t_i}, c; \theta)|_{c=\Delta X_{t_i}(y_{t_i}, y_{t_{i+1}}; \theta)} \right)^{-1} \\ &= \left(Z_{t_{i+1}-t_i}(\Delta X_{t_i}(y_{t_i}, y_{t_{i+1}}; \theta)) \right)^{-1}. \end{aligned}$$

Likelihood Construction cont'd

- Note that, by definition,

$$F_{t_{i+1}-t_i}(y_{t_i}, \Delta X_{t_i}(y_{t_i}, y; \theta); \theta) \equiv y.$$

Thus,

$$D_c F_{t_{i+1}-t_i}(y_{t_i}, c; \theta)|_{c=\Delta X_{t_i}(y_{t_i}, y_{t_{i+1}}; \theta)} \cdot \nabla \Delta X_{t_i}(y_{t_i}, y; \theta)|_{y=y_{t_{i+1}}} \equiv I_d$$

and, consequently,

$$\begin{aligned} \nabla \Delta X_{t_i}(y_{t_i}, y; \theta)|_{y=y_{t_{i+1}}} &= \left(D_c F_{t_{i+1}-t_i}(y_{t_i}, c; \theta)|_{c=\Delta X_{t_i}(y_{t_i}, y_{t_{i+1}}; \theta)} \right)^{-1} \\ &= \left(Z_{t_{i+1}-t_i}(\Delta X_{t_i}(y_{t_i}, y_{t_{i+1}}; \theta)) \right)^{-1}. \end{aligned}$$

- So, $L_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right)$ can be written as

$$L_{X^{\mathcal{D}}}\left(J_{\mathcal{D}}^{-1}(y_{\mathcal{D}(n)})\right) \left(\prod_{t_i \in \mathcal{D}} \left| Z_{t_{i+1}-t_i}\left(J_{\mathcal{D}}^{-1}(y_{\mathcal{D}})_{t_i}\right) \right| \right)^{-1}.$$

Example: linear system

- Consider the equation

$$dY_t^{\mathcal{D}} = -\lambda Y_t^{\mathcal{D}} dt + \sigma X_t^{\mathcal{D}}, \quad Y_0^{\mathcal{D}} = 0,$$

where $X_t^{\mathcal{D}}$ is the piecewise linear interpolation to a fractional Brownian path with Hurst parameter h on a homogeneous grid $\mathcal{D} = \{k\delta; k = 0, \ldots, N\}$ where $N\delta = T$.

Example: linear system

- Consider the equation

$$dY_t^{\mathcal{D}} = -\lambda Y_t^{\mathcal{D}} dt + \sigma X_t^{\mathcal{D}}, \quad Y_0^{\mathcal{D}} = 0,$$

  where $X_t^{\mathcal{D}}$ is the piecewise linear interpolation to a fractional Brownian path with Hurst parameter h on a homogeneous grid $\mathcal{D} = \{k\delta; k = 0, \ldots, N\}$ where $N\delta = T$.
- In this case, we can solve both the ODE and the system of equations explicitely and we get

$$J_{\mathcal{D}}^{-1}(y_{\mathcal{D}}; \theta)_{k+1} := \Delta x_{k+1} = \frac{\lambda\delta\left(y_{(k+1)\delta} - y_{k\delta} e^{-\lambda\delta}\right)}{\sigma\left(1 - e^{-\lambda\delta}\right)}$$

Example: linear system

- Consider the equation

$$dY_t^{\mathcal{D}} = -\lambda Y_t^{\mathcal{D}} dt + \sigma X_t^{\mathcal{D}}, \quad Y_0^{\mathcal{D}} = 0,$$

where $X_t^{\mathcal{D}}$ is the piecewise linear interpolation to a fractional Brownian path with Hurst parameter h on a homogeneous grid $\mathcal{D} = \{k\delta; k = 0, \ldots, N\}$ where $N\delta = T$.

- In this case, we can solve both the ODE and the system of equations explicitely and we get

$$J_{\mathcal{D}}^{-1}(y_{\mathcal{D}};\theta)_{k+1} := \Delta x_{k+1} = \frac{\lambda\delta\left(y_{(k+1)\delta} - y_{k\delta}e^{-\lambda\delta}\right)}{\sigma\left(1 - e^{-\lambda\delta}\right)}$$

- Moreover,

$$Z_t = \int_0^t \exp(-\lambda(t-s))\frac{\sigma}{\delta}ds = \frac{\sigma}{\lambda\delta}(1 - e^{-\lambda t}).$$

Example: linear system cont'd

- We can now write down the likelihood:

$$L_{Y^{\mathcal{D}}}(y_{\mathcal{D}}|\ \theta) = L_{\Delta X_{\mathcal{D}}}\left(J_{\mathcal{D}}^{-1}(y_{\mathcal{D}};\theta)\right)\left(\frac{\lambda\delta}{\sigma(1-e^{-\lambda\delta})}\right)^{N}.$$

Example: linear system cont'd

- We can now write down the likelihood:

$$L_{Y^{\mathcal{D}}}\left(y_{\mathcal{D}} \mid \theta\right) = L_{\Delta X_{\mathcal{D}}}\left(J_{\mathcal{D}}^{-1}(y_{\mathcal{D}};\theta)\right)\left(\frac{\lambda\delta}{\sigma(1-e^{-\lambda\delta})}\right)^{N}.$$

- In particular, for X fBM, this becomes

$$\frac{1}{\sqrt{|2\pi\Sigma_h^{\mathcal{D}}|}}\exp\left(-\frac{1}{2}J_{\mathcal{D}}^{-1}(y_{\mathcal{D}};\theta)\left(\Sigma_h^{\mathcal{D}}\right)^{-1}J_{\mathcal{D}}^{-1}(y_{\mathcal{D}};\theta)^{*}\right)\left(\frac{\lambda\delta}{\sigma(1-e^{-\lambda\delta})}\right)^{N},$$

where $\Sigma_h^{\mathcal{D}}$ is the coveriance matrix of fGN with Hurst parameter h.

Example: linear system cont'd

- Using also the expression for $J_{\mathcal{D}}^{-1}$, the log-likelihood becomes

$$\ell_Y\left(y_{\mathcal{D}} \middle|\, \lambda, \sigma\right) \quad \propto \quad -\frac{\lambda^2\delta^2}{2\sigma^2(1-e^{-\lambda\delta})^2}\left(\Delta^\lambda y\right)_{\mathcal{D}}\left(\Sigma_h^{\mathcal{D}}\right)^{-1}\left(\Delta^\lambda y\right)_{\mathcal{D}}^* + N\log\left(\frac{\lambda\delta}{\sigma(1-e^{-\lambda\delta})}\right),$$

where $\Delta^\lambda y_{k\delta} = y_{(k+1)\delta} - y_{k\delta}e^{\lambda\delta}$.

Main Theorem

- Let $\ell_{Y(n)}(\cdot|\theta)$ be the approximate likelihood, y be the response to a p-rough path x and $y(n)$ be the response to $\pi_n(x)$, where $\pi_n(x)$ is the piecewise linear interpolation of x on grid $\mathcal{D}(n) = \{k2^{-n}T,\ k = 0, \ldots, N\}$ for $N = 2^n T$. Then, assuming that the determinant of b is uniformly bounded from below over both parameters y and θ and that

$$\left|\ell_{\Delta X_{\mathcal{D}(n)}}\left(\Delta x_{\mathcal{D}(n)}\right) - \ell_{\Delta X_{\mathcal{D}(n)}}\left(\Delta \tilde{x}_{\mathcal{D}(n)}\right)\right| \leq \omega\left(d_p(x, \tilde{x})\right),$$

we get that

$$\lim_{n\to\infty} \sup_{\theta} \left|\ell_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right) - \ell_{Y(n)}\left(y(n)_{\mathcal{D}(n)}|\theta\right)\right| = 0.$$

Lemma 1

For $Z_{t_{i+1}-t_i}$ and $J^{-1}_{\mathcal{D}(n)}$ defined as before and under the additional assumption on b that

$$\inf_{y,\theta} ||b(y;\theta)|| = \frac{1}{M_b} > 0,$$

for some $M_b > 0$, it holds that

$$|\sum_{t_i \in \mathcal{D}(n)} \log |Z_{t_{i+1}-t_i}\left(J^{-1}_{\mathcal{D}(n)}(y_{\mathcal{D}(n)})_{t_i}\right)| \\ - \sum_{t_i \in \mathcal{D}(n)} \log |Z_{t_{i+1}-t_i}\left(J^{-1}_{\mathcal{D}(n)}(\tilde{y}_{\mathcal{D}(n)})_{t_i}\right)| \leq \\ C \cdot \omega(d_p(J^{-1}_{\mathcal{D}(n)}(y_{\mathcal{D}(n)}), J^{-1}_{\mathcal{D}(n)}(\tilde{y}_{\mathcal{D}(n)})))$$

for some $C \in \mathbb{R}_+$ and modulus of continuity function ω.

Lemma 2

Let $J_{\mathcal{D}(n)}^{-1}$ be the inverse Itô map previously defined. Moreover, let $Y(n, I_{\theta_0}(x)_{\mathcal{D}(n)})$ and $Y(n, I_{\theta_0}(\pi_n(x))_{\mathcal{D}(n)})$ be the responses to the piecewise linear map, parametrised by its values on the grid $\mathcal{D}(n)$, given by $I_{\theta_0}(x)_{\mathcal{D}(n)}$ and $I_{\theta_0}(\pi_n(x))_{\mathcal{D}(n)}$ respectively, where x is a fixed rough path in $G\Omega_p(\mathbb{R}^d)$ and $\theta_0 \in \Theta$. Then,

$$\lim_{n\to\infty} d_p \left(I_\theta^{-1} \left(Y(n, I_{\theta_0}(x)_{\mathcal{D}(n)}) \right), I_\theta^{-1} \left(Y(n, I_{\theta_0}(\pi_n(x))_{\mathcal{D}(n)}) \right) \right) = 0,$$

provided that $d_p(\pi_n(x), x) \to 0$ as $n \to \infty$.

Example: likelihood and parameter estimation

- According to our previous analysis, the approximate likelihood of discrete observations of an OU process becomes

$$-\tfrac{\lambda^2\delta^2}{2\sigma^2(1-e^{-\lambda\delta})^2}\tfrac{1}{\delta}\sum_{k=1}^{\frac{T}{\delta}}(y_{t_k}-y_{t_{k-1}}e^{\lambda\delta})^2+\tfrac{T}{\delta}\log\left(\tfrac{\lambda\delta}{\sigma(1-e^{-\lambda\delta})}\right)=$$
$$\tfrac{1}{\delta}\left(-T\log\sigma-\tfrac{1}{2\sigma^2}\sum_k\Delta y_k^2\right)$$
$$-\left(\tfrac{\lambda}{\sigma^2}\sum_k y_{t_k}(y_{t_{k+1}}-y_{t_k})+\tfrac{\lambda^2}{\sigma^2}\sum_k y_{t_k}^2\delta\right)+o(\delta).$$

Example: likelihood and parameter estimation

- According to our previous analysis, the approximate likelihood of discrete observations of an OU process becomes

$$-\frac{\lambda^2\delta^2}{2\sigma^2(1-e^{-\lambda\delta})^2}\frac{1}{\delta}\sum_{k=1}^{\frac{T}{\delta}}(y_{t_k}-y_{t_{k-1}}e^{\lambda\delta})^2+\frac{T}{\delta}\log\left(\frac{\lambda\delta}{\sigma(1-e^{-\lambda\delta})}\right)=$$
$$\frac{1}{\delta}\left(-T\log\sigma-\frac{1}{2\sigma^2}\sum_k \Delta y_k^2\right)$$
$$-\left(\frac{\lambda}{\sigma^2}\sum_k y_{t_k}(y_{t_{k+1}}-y_{t_k})+\frac{\lambda^2}{\sigma^2}\sum_k y_{t_k}^2\delta\right)+o(\delta).$$

- The $o(\frac{1}{\delta})$ term converges to

$$-T\log\sigma-\frac{1}{2\sigma^2}<Y,Y>_T\ .$$

Example: likelihood and parameter estimation

- According to our previous analysis, the approximate likelihood of discrete observations of an OU process becomes

$$-\tfrac{\lambda^2\delta^2}{2\sigma^2(1-e^{-\lambda\delta})^2}\tfrac{1}{\delta}\sum_{k=1}^{\frac{T}{\delta}}(y_{t_k}-y_{t_{k-1}}e^{\lambda\delta})^2+\tfrac{T}{\delta}\log\left(\tfrac{\lambda\delta}{\sigma(1-e^{-\lambda\delta})}\right)=$$
$$\tfrac{1}{\delta}\left(-T\log\sigma-\tfrac{1}{2\sigma^2}\sum_k \Delta y_k^2\right)$$
$$-\left(\tfrac{\lambda}{\sigma^2}\sum_k y_{t_k}(y_{t_{k+1}}-y_{t_k})+\tfrac{\lambda^2}{\sigma^2}\sum_k y_{t_k}^2\delta\right)+o(\delta).$$

- The $o(\frac{1}{\delta})$ term converges to

$$-T\log\sigma-\frac{1}{2\sigma^2}<Y,Y>_T\ .$$

- The $o(1)$ term converges to

$$-\frac{\lambda}{\sigma^2}\int_0^T y_u dy_u-\frac{\lambda^2}{2\sigma^2}\int_0^T y_u^2 du.$$

Approximate likelihood and MLEs

- To avoid losing information about certain parameters in the limit, we construct a canonical expansion of the log-likelihood as

$$\ell_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right) = \sum_{k=0}^{M} \ell^{(k)}_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right) n^{-\alpha_k} + R_M(y_{\mathcal{D}(n)}, \theta)$$

  for $M > 0$ and $-\infty < \alpha_0 < \alpha_1 < \cdots < \alpha_M < \infty$, where $\ell^{(k)}_{Y(n)}\left(y_{\mathcal{D}(n)}|\theta\right)$ converges to a non-trivial limit (finite and non-zero) for every $k = 0, \ldots, M$ and the remainder $R_M(y_{\mathcal{D}(n)}, \theta)$ satisfies $\lim_{n\to\infty} n^{\alpha_M} R_M(y_{\mathcal{D}(n)}, \theta) = 0$.

Going Further

- We have assumed that X is approximated by piecewise linear paths. Is this really necessary?

Going Further

- We have assumed that X is approximated by piecewise linear paths. Is this really necessary?
- We have assumed that the inverse problem can be solved explicitly. What if it cannot?

Generalisation of the previous set-up

- We assume that $X(n)$ are such that $d_p(X(n)), X) \to 0$ as $n \to \infty$ w.p. 1 and that $X(n)$ live in a finite-dimensional sub-manifold $\mathcal{X}_n \subset \Omega_p(\mathbb{R}^m)$. Moreover, the response $Y(n) = I(X(n))$ also belongs to a finite-dimensional sub-manifold $\mathcal{Y}_n \subset \Omega_p(\mathbb{R}^d)$ that is in $1-1$ correspondence with $\mathbb{R}^{d|\mathcal{D}(n)|}$.

Generalisation of the previous set-up

- We assume that $X(n)$ are such that $d_p(X(n)), X) \to 0$ as $n \to \infty$ w.p. 1 and that $X(n)$ live in a finite-dimensional sub-manifold $\mathcal{X}_n \subset \Omega_p(\mathbb{R}^m)$. Moreover, the response $Y(n) = I(X(n))$ also belongs to a finite-dimensional sub-manifold $\mathcal{Y}_n \subset \Omega_p(\mathbb{R}^d)$ that is in $1-1$ correspondence with $\mathbb{R}^{d|\mathcal{D}(n)|}$.
- The goal is to express a finite-dimensional representation of $X(n)$ (whose distribution we assume we know) in terms of the data.

Solving the 'inverse problem'

- We want to find $X(n) \in \mathcal{X}_n$ such that $I(X(n))_{t_i} = y_{t_i}, \ \forall t_i \in \mathcal{D}(n)$.

Solving the ‘inverse problem’

- We want to find $X(n) \in \mathcal{X}_n$ such that $I(X(n))_{t_i} = y_{t_i}, \ \forall t_i \in \mathcal{D}(n)$.
- Most standard numerical techniques are local while we need to control the error in p-variation.

Solving the 'inverse problem'

- We want to find $X(n) \in \mathcal{X}_n$ such that $I(X(n))_{t_i} = y_{t_i}, \ \forall t_i \in \mathcal{D}(n)$.
- Most standard numerical techniques are local while we need to control the error in p-variation.
- Idea: work on the signature space.

A general algorithm

- Initialization: $Y(n, 0)$ is a linear interpolation of observations.

A general algorithm

- Initialization: $Y(n, 0)$ is a linear interpolation of observations.
- Step $Y(n, k) \to Y(n, k+1)$, $k \geq 0$.

A general algorithm

- Initialization: $Y(n, 0)$ is a linear interpolation of observations.
- Step $Y(n, k) \to Y(n, k+1)$, $k \geq 0$.
  - $X(n, k) = I^{-1}(Y(n, k))$.

A general algorithm

- Initialization: $Y(n,0)$ is a linear interpolation of observations.
- Step $Y(n,k) \to Y(n,k+1)$, $k \geq 0$.
  - $X(n,k) = I^{-1}(Y(n,k))$.
  - $\tilde{X}(n,k) = \arg\min_{\tilde{X} \in \mathcal{X}(n)} d_p(X(n,k)\tilde{X})$.

A general algorithm

- Initialization: $Y(n,0)$ is a linear interpolation of observations.
- Step $Y(n,k) \to Y(n,k+1)$, $k \geq 0$.
  - $X(n,k) = I^{-1}(Y(n,k))$.
  - $\tilde{X}(n,k) = \arg\min_{\tilde{X} \in \mathcal{X}(n)} d_p(X(n,k)\tilde{X})$.
  - $\tilde{Y}(n,k) = I\left(\tilde{X}(n,k)\right)$.

A general algorithm

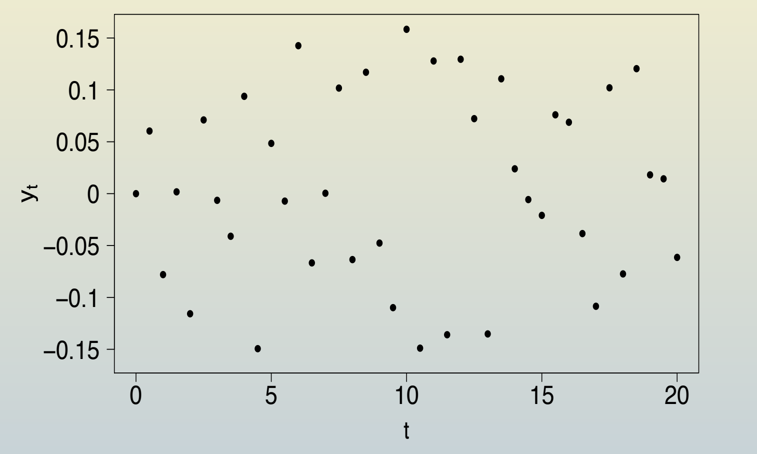
- Initialization: $Y(n,0)$ is a linear interpolation of observations.
- Step $Y(n,k) \to Y(n,k+1)$, $k \geq 0$.
  - $X(n,k) = I^{-1}(Y(n,k))$.
  - $\tilde{X}(n,k) = \arg\min_{\tilde{X} \in \mathcal{X}(n)} d_p(X(n,k)\tilde{X})$.
  - $\tilde{Y}(n,k) = I\left(\tilde{X}(n,k)\right)$.
  - $Y(n,k+1) = Y(n,k) +$ tree-like path going through observations.

A general algorithm

- Initialization: $Y(n,0)$ is a linear interpolation of observations.
- Step $Y(n,k) \to Y(n,k+1)$, $k \geq 0$.
  - $X(n,k) = I^{-1}(Y(n,k))$.
  - $\tilde{X}(n,k) = \arg\min_{\tilde{X} \in \mathcal{X}(n)} d_p(X(n,k)\tilde{X})$.
  - $\tilde{Y}(n,k) = I\left(\tilde{X}(n,k)\right)$.
  - $Y(n,k+1) = Y(n,k) +$ tree-like path going through observations.
- Intuition: try to correct the path so that it satisfies required conditions by changing the signature as little as possible.

First iteration

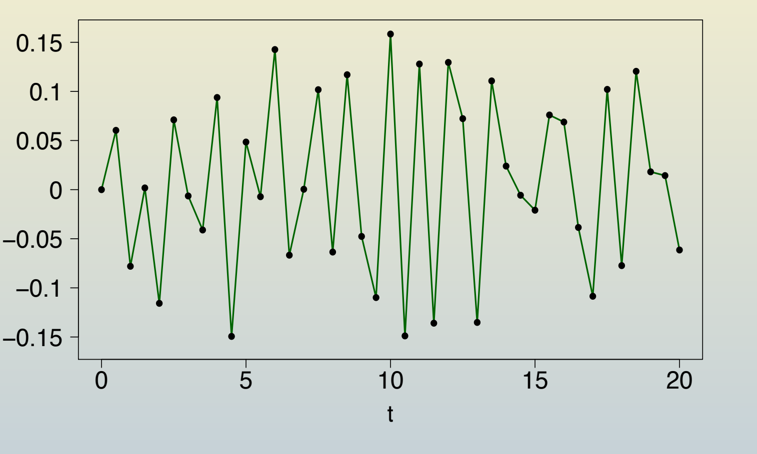
- The data:



-

First iteration

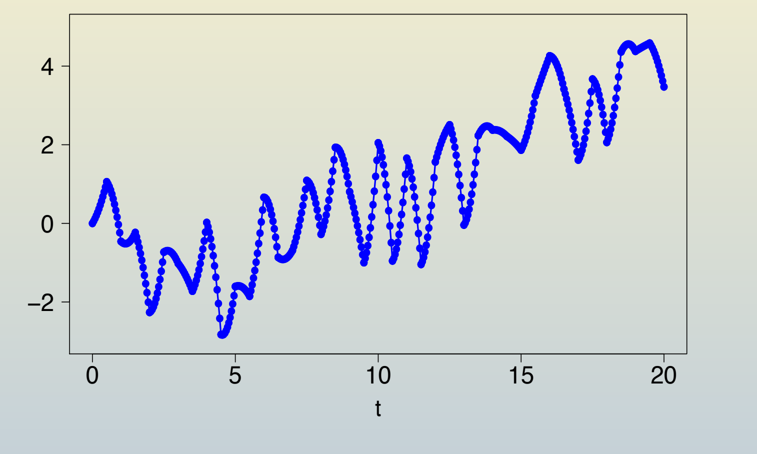
- Initialization. $Y(0)$:



-

First iteration

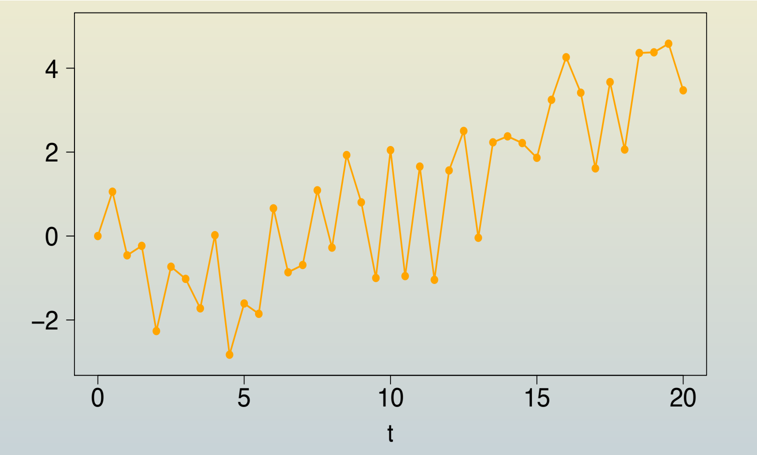
- Solving for the noise. $X(0) = I_\theta^{-1}(Y(0))$:



-

First iteration

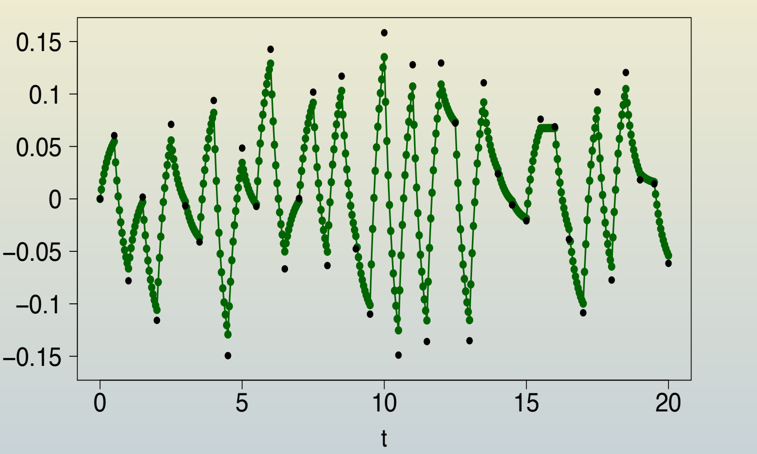
- Linear interpolation of $X(0)$. $\tilde{X}(0)$:



-

First iteration

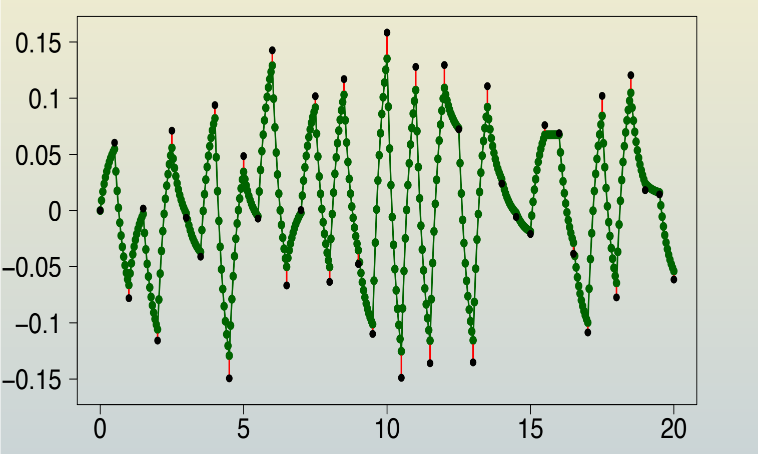
- $\tilde{Y}(0) = I_\theta(\tilde{X}(0))$:



-

First iteration

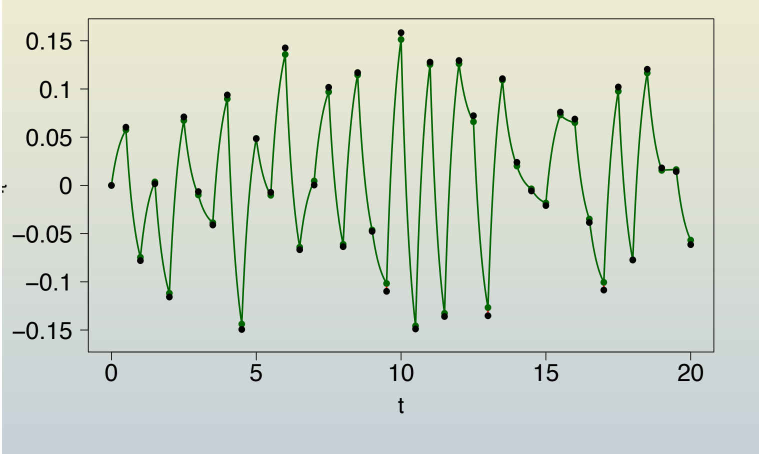
- $Y(1)$ connects $\tilde{Y}(0)$ to observations:



-

Evolution of Y

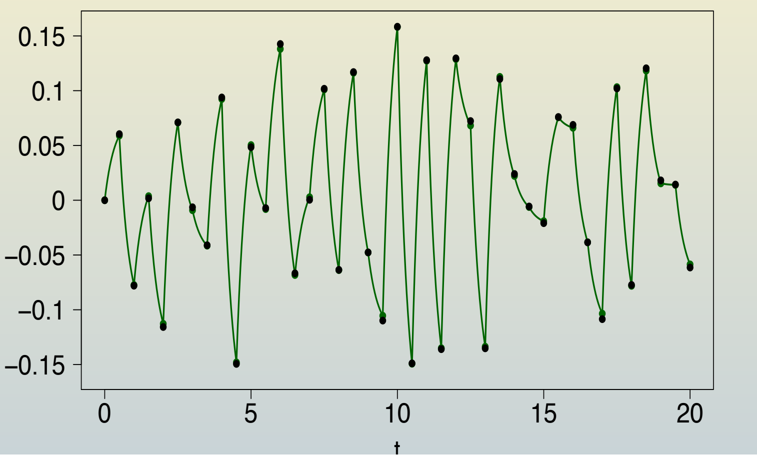
- Second iteration:



-

Evolution of Y

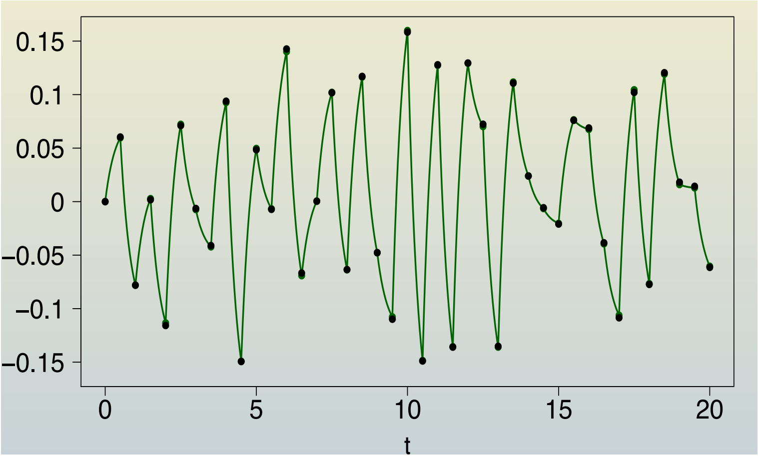
- Third iteration:



-

Evolution of Y

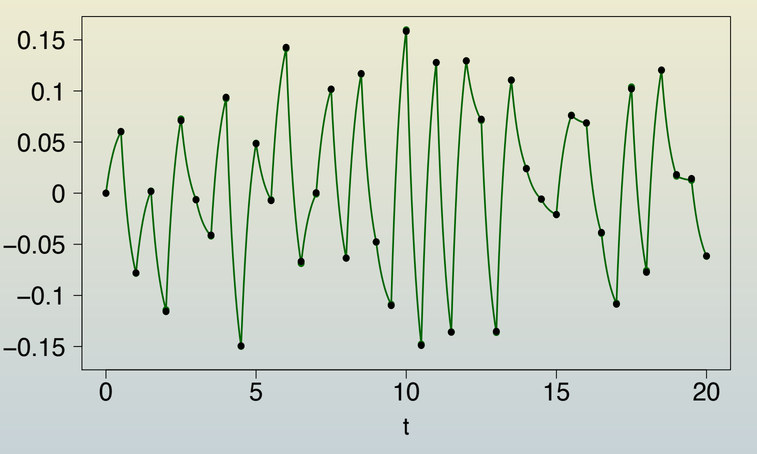
- Fourth iteration:



-

Evolution of Y

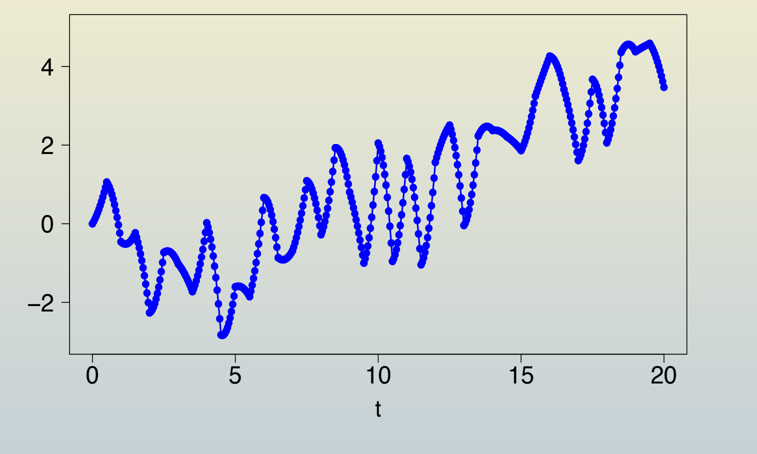
- Fifth iteration:



-

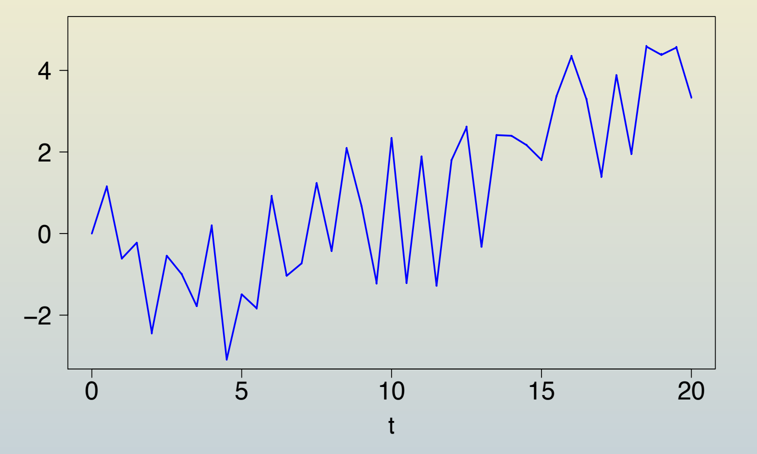
Evolution of X

- First iteration:



-

Evolution of X

- Fifth iteration:


-

A conjecture

- The map $Y(n,k) \to Y(n,k+1)$ is a contraction in the signature space, i.e.

$$d\left(S(Y(n,k+1))_{0,T}, S(Y(n,k))_{0,T}\right) < c \cdot d\left(S(Y(n,k))_{0,T}, S(Y(n,k-1))\right.$$

for $c < 1$.

A conjecture

- The map $Y(n,k) \to Y(n,k+1)$ is a contraction in the signature space, i.e.

$$d\left(S(Y(n,k+1))_{0,T}, S(Y(n,k))_{0,T}\right) < c \cdot d\left(S(Y(n,k))_{0,T}, S(Y(n,k-1))\right.$$

for $c < 1$.

- Convergence will also imply convergence in p-variation.

Summary

- We have constructed a general framework for constructing an approximate likelihood for discretely observed RDEs.

Summary

- We have constructed a general framework for constructing an approximate likelihood for discretely observed RDEs.
- There are still numerous issues to be addressed:

Summary

- We have constructed a general framework for constructing an approximate likelihood for discretely observed RDEs.
- There are still numerous issues to be addressed:
  - Solving the inverse problem in the general framework.

Summary

- We have constructed a general framework for constructing an approximate likelihood for discretely observed RDEs.
- There are still numerous issues to be addressed:
  - Solving the inverse problem in the general framework.
  - ‘Exact’ construction.

Summary

- We have constructed a general framework for constructing an approximate likelihood for discretely observed RDEs.
- There are still numerous issues to be addressed:
  - Solving the inverse problem in the general framework.
  - ‘Exact’ construction.
  - Properties of estimators.