Feynman–Kac models

Stability & discretization

MCQMC 2018

Grégoire Ferré[**,⋆] Mathias Rousset[†] Gabriel Stoltz[*]

[*] École des Ponts ParisTech & INRIA Paris
[⋆] Labex Bézout
[†] INRIA Rennes & IRMAR

Wednesday July 4th, 2018

Situation

If *this meeting* is about

$$x - \text{Monte Carlo}$$

with

$$x \in \{\text{Quasi}, \text{Markov Chain}, \text{Sequential}, \text{Quantum}, \text{Hamiltonian}, \text{Multilevel}, \text{Diffusion}, \text{Langevin}, \text{Multifidelity}\ldots\},$$

Situation

If *this meeting* is about

$$x - \text{Monte Carlo}$$

with

$$x \in \{\text{Quasi}, \text{Markov Chain}, \text{Sequential}, \text{Quantum}, \text{Hamiltonian}, \text{Multilevel}, \text{Diffusion}, \text{Langevin}, \text{Multifidelity}\ldots\},$$

then *this talk* would be on

$$x - \text{Monte Carlo},$$

with

$$x \in \{\text{Sequential}, \text{Quantum}, \text{Diffusion}\},$$

summarized as *Feynman–Kac models* (also known as splitting, genealogical, branching models, etc).

A rare event problem

- Consider a dynamics over $\mathcal{X} \subset \mathbb{R}^d$:

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

  with invariant measure μ and generator $\mathcal{L}$.

- Ergodicity: for a test function f

$$\frac{1}{t}\int_0^t f(X_s)ds \xrightarrow{t\to+\infty} \int_{\mathcal{X}} f(x)\,\mu(dx).$$

A rare event problem

- Consider a dynamics over $\mathcal{X} \subset \mathbb{R}^d$:

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

with invariant measure μ and generator $\mathcal{L}$.

- Ergodicity: for a test function f

$$\frac{1}{t}\int_0^t f(X_s)ds \xrightarrow{t\to+\infty} \int_{\mathcal{X}} f(x)\,\mu(dx).$$

- Idea of *large deviations*:

$$\mathbb{P}\left[\frac{1}{t}\int_0^t f(X_s)ds = a\right] \asymp \mathrm{e}^{-tI(a)},$$

where I is the *rate function*, or *generalized entropy creation*.

Duality and importance sampling

- **Donsker-Varadhan** duality [70's]: in many cases it holds

$$I(a) = \sup_{k\in\mathbb{R}} \left\{ ka - \lambda_f(k) \right\},$$

where

$$\lambda_f(k) = \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}\left[\mathrm{e}^{k\int_0^t f(X_s)ds} \right].$$

Duality and importance sampling

- **Donsker-Varadhan** duality [70's]: in many cases it holds

$$I(a) = \sup_{k\in\mathbb{R}} \left\{ ka - \lambda_f(k) \right\},$$

where

$$\lambda_f(k) = \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}\left[\mathrm{e}^{k\int_0^t f(X_s)ds} \right].$$

- Similar to study, for an observable φ,

$$\Theta_t(\mu)(\varphi) = \frac{\mathbb{E}_\mu\left[\varphi(X_t)\, \mathrm{e}^{\int_0^t f(X_s)ds} \right]}{\mathbb{E}_\mu\left[\mathrm{e}^{\int_0^t f(X_s)ds} \right]}.$$

- This procedure *weights the trajectories* depending on f.

Natural questions

One may wonder:

- for which class of functions f does λ_f exists?
- under which conditions Θ_t admits a long time limit?
- how can we discretize in time for numerical purposes?
- how to sample efficiently the expectations?

Natural questions

One may wonder:

- for which class of functions f does λ_f exists?
- under which condition Θ_t admits a long time limit?
- how can we discretize in time for numerical purposes?
- how to sample efficiently the expectations?

Link with filtering & Sequential Monte Carlo

In Sequential Monte Carlo/filtering, we generally consider the weighted path measure

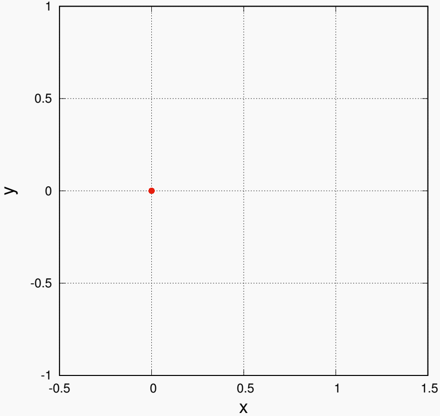
$$\mathbb{Q}_n(dx_{0:n}) = \frac{1}{L_n}\Big\{ G_0(x_0) \prod_{k=1}^{t} G_k(x_{k-1}, x_k) \Big\} \mathbb{M}_0(dx_0) \prod_{k=1}^{n} M_k(x_{k-1}, x_k).$$

Analogies:

- $G_k(x_{k-1}, x_k) \leftrightarrow \mathrm{e}^{f(x_{k-1})}$ with «potential» f,
- $M_k \leftrightarrow Q(x, \cdot) = \mathbb{E}\big[x_k \in \cdot \,\big|\, x_{k-1} = x\big]$ homogeneous transition kernel,
- $L_n \leftrightarrow (\mathrm{e}^f Q)^n \mathbb{1} = \mathbb{E}\big[\mathrm{e}^{\sum_{k=0}^{n-1} f(x_k)}\big]$ accumulated weight up to time n (sometimes written $\gamma_n(1)$).

Numerics – Sequential Monte Carlo

- Dynamics $(x_{n+1}, y_{n+1}) = (x_n, y_n) - 0.1(x_n, y_n) + G_n \in \mathbb{R}^2$,
- Weight function $f(x, y) = x$.



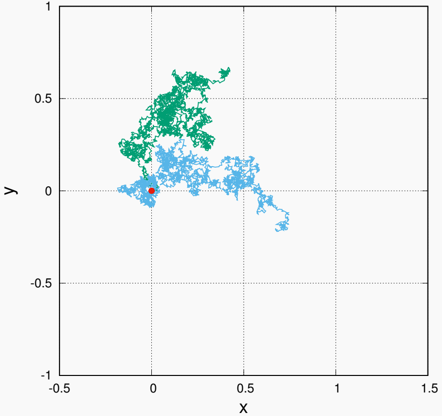
The weight

$$w_m^n = \exp\left(\sum_{k=0}^{n-1} f(x_k^m)\right).$$

of particle m at time n is used for resampling (renormalization).

Numerics – Sequential Monte Carlo

- Dynamics $(x_{n+1}, y_{n+1}) = (x_n, y_n) - 0.1(x_n, y_n) + G_n \in \mathbb{R}^2$,
- Weight function $f(x, y) = x$.



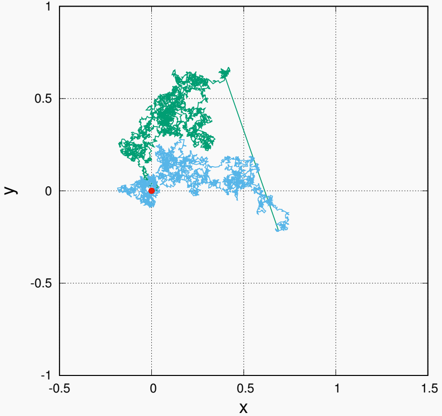
The weight

$$w_m^n = \exp\left(\sum_{k=0}^{n-1} f(x_k^m)\right).$$

of particle m at time n is used for resampling (renormalization).

Numerics – Sequential Monte Carlo

- Dynamics $(x_{n+1}, y_{n+1}) = (x_n, y_n) - 0.1(x_n, y_n) + G_n \in \mathbb{R}^2$,
- Weight function $f(x, y) = x$.



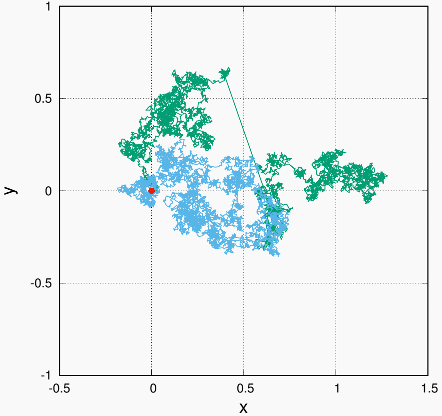
The weight

$$w_m^n = \exp\left(\sum_{k=0}^{n-1} f(x_k^m)\right).$$

of particle m at time n is used for resampling (renormalization).

Numerics – Sequential Monte Carlo

- Dynamics $(x_{n+1}, y_{n+1}) = (x_n, y_n) - 0.1(x_n, y_n) + G_n \in \mathbb{R}^2$,
- Weight function $f(x, y) = x$.



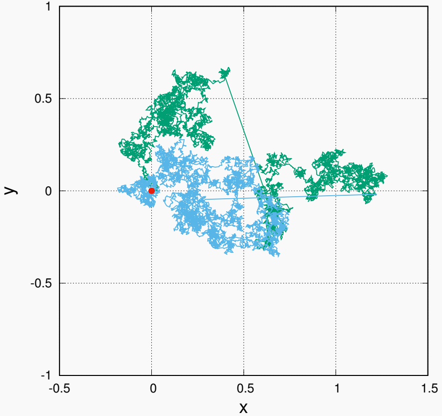
The weight

$$w_m^n = \exp\left(\sum_{k=0}^{n-1} f(x_k^m)\right).$$

of particle m at time n is used for resampling (renormalization).

Numerics – Sequential Monte Carlo

- Dynamics $(x_{n+1}, y_{n+1}) = (x_n, y_n) - 0.1(x_n, y_n) + G_n \in \mathbb{R}^2$,
- Weight function $f(x, y) = x$.



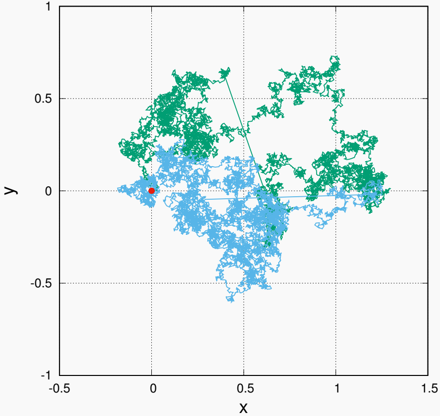
The weight

$$w_m^n = \exp\left(\sum_{k=0}^{n-1} f(x_k^m)\right).$$

of particle m at time n is used for resampling (renormalization).

Numerics – Sequential Monte Carlo

- Dynamics $(x_{n+1}, y_{n+1}) = (x_n, y_n) - 0.1(x_n, y_n) + G_n \in \mathbb{R}^2$,
- Weight function $f(x, y) = x$.



The weight

$$w_m^n = \exp\left(\sum_{k=0}^{n-1} f(x_k^m)\right).$$

of particle m at time n is used for resampling (renormalization).

Conclusion: the particles are selected towards the right.

Markov chains and ergodicity

A homogeneous Markov chain $(x_n)_{n\in\mathbb{N}}$ is defined through its *evolution operator* Q:

$$Q\varphi(x) := \mathbb{E}[\varphi(x_1) \,|\, x_0 = x] = \int_{\mathcal{X}} \varphi(y) Q(x, dy).$$

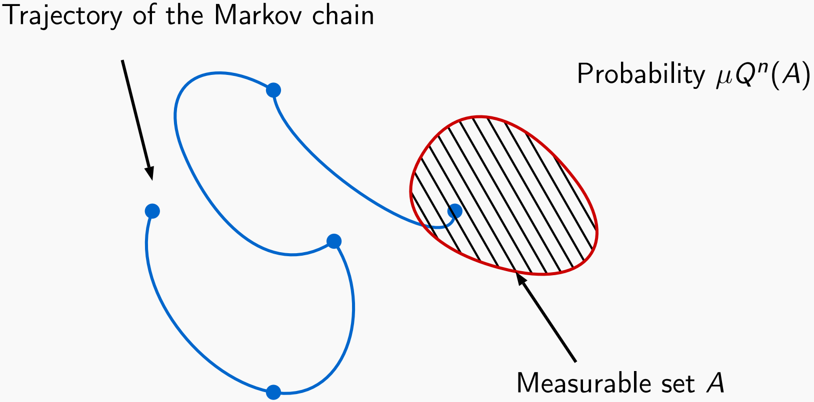
A measure $\mu^* \in \mathcal{P}(\mathcal{X})$ is *invariant* if $\forall\, A \subset \mathcal{X}$:

$$\mu^* Q(A) := \int_{\mathcal{X}} \mu^*(dx) Q(x, A) = \mu(A).$$

The process is *ergodic* when, for any initial distribution μ,

$$\mu Q^n \xrightarrow{n\to+\infty} \mu^*.$$

Big picture I



An ergodic theorem

Theorem [M. Hairer & J. Mattingly]

Assume there exist $W : \mathcal{X} \to \mathbb{R}_+$, $\gamma \in (0,1)$ and $C > 0$ such that

$$(L) \quad PW \le \gamma W + C,$$

and $\alpha > 0$, $\eta \in \mathcal{P}(\mathcal{X})$ such that

$$(M) \quad \inf_{x \in \mathcal{C}} P(x,\cdot) \ge \alpha\eta(\cdot),$$

for $\mathcal{C}$ a large enough level set of W. Then, there exist a unique $\mu^* \in \mathcal{P}(\mathcal{X})$, $C > 0$ and $\bar{\alpha} \in (0,1)$ such that for any $\mu \in \mathcal{P}(\mathcal{X})$

$$\|P^n\mu - \mu^*\|_W \le C\bar{\alpha}^n \|\mu - \mu^*\|_W.$$

Here: $\|f\| = \sup_{x\in\mathcal{X}} \dfrac{|f(x)|}{1+W(x)}$, $\quad \|\mu-\nu\|_W = \sup_{\|\varphi\|\le 1} \int_{\mathcal{X}} \varphi(x)(\mu-\nu)(dx).$

Feynman–Kac models

We consider:

- a kernel operator Q^f with $Q^f \neq \mathbb{1}$;
- a dynamics $\Phi_n : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{X})$ defined by, for tests functions φ

$$\Phi_n(\mu)(\phi) = \frac{\mu\big((Q^f)^n \varphi\big)}{\mu\big((Q^f)^n \mathbb{1}\big)} = (\Phi^n)(\mu)(\varphi) \quad \text{with } \Phi(\mu)(\varphi) = \frac{\mu\big(Q^f \varphi\big)}{\mu\big(Q^f \mathbb{1}\big)}.$$

Feynman–Kac models

We consider:

- a kernel operator Q^f with $Q^f \neq \mathbb{1}$;
- a dynamics $\Phi_n : \mathcal{P}(\mathcal{X}) \to \mathcal{P}(\mathcal{X})$ defined by, for tests functions φ

$$\Phi_n(\mu)(\phi) = \frac{\mu\big((Q^f)^n\varphi\big)}{\mu\big((Q^f)^n\mathbb{1}\big)} = (\Phi^n)(\mu)(\varphi) \quad \text{with } \Phi(\mu)(\varphi) = \frac{\mu\big(Q^f\varphi\big)}{\mu\big(Q^f\mathbb{1}\big)}.$$

Typically

- $Q^f = \mathrm{e}^f Q$, where Q defines a Markov chain $(x_n)_{n\in\mathbb{N}}$;
- then

$$\Phi_n(\mu)(\phi) = \frac{\mathbb{E}_\mu\left[\varphi(x_n)\,\mathrm{e}^{\sum_{k=0}^{n-1} f(x_k)}\right]}{\mathbb{E}_\mu\left[\mathrm{e}^{\sum_{k=0}^{n-1} f(x_k)}\right]}.$$

Ergodicity for Feynman–Kac dynamics

Theorem [G.F., M. Rousset & G. Stoltz, yesterday]

Assumptions:

- Lyapunov condition: there exist $W : \mathcal{X} \to [1, +\infty)$, $\gamma_n \xrightarrow{\infty} 0$, $b_n \geq 0$ and compact sets K_n s.t.

$$(L) \quad Q^f W(x) \leq \gamma_n W(x) + b_n \mathbb{1}_{K_n};$$

- Minorization condition: $\forall\, n \geq 1$, there exist $\alpha_n > 0$, $\eta_n \in \mathcal{P}(\mathcal{X})$ s.t.

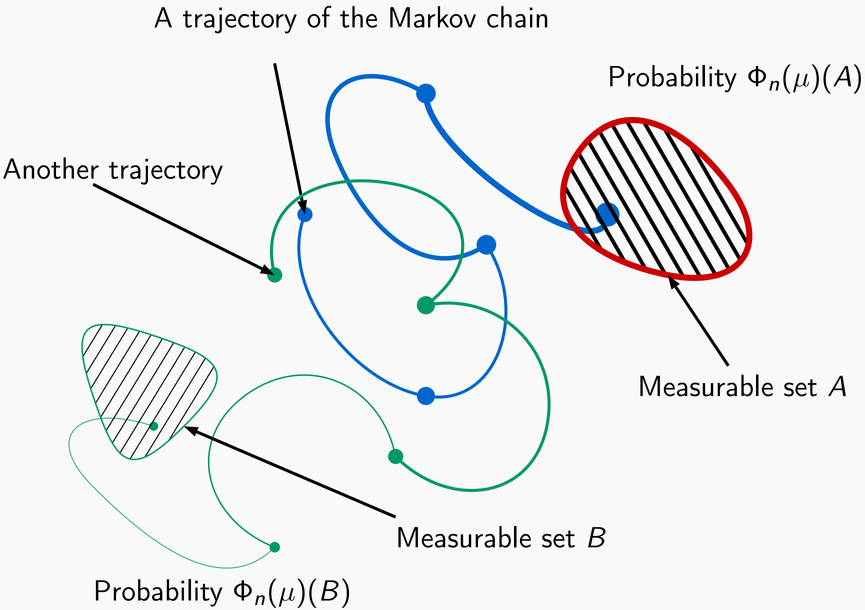
$$(M) \quad \forall\, x \in K_n, \quad Q^f(x, \cdot) \geq \alpha_n \eta_n(\cdot);$$

- $\mathcal{X}$ is Polish and Q^f is «locally strong Feller».

Then, there exist a unique μ_f^* and $\bar{\alpha} \in (0,1)$ such that for any $\mu \in \mathcal{P}(\mathcal{X})$, $\exists C_\mu > 0$ for which

$$\|\Phi_n(\mu) - \mu_f^*\|_W \leq C_\mu \bar{\alpha}^n, \quad \text{and} \quad \Phi(\mu_f^*) = \mu_f^*.$$

Big picture II



Sketch of proof

The steps of the proof are as follows:

- Q^f is quasi-compact in $L^\infty_W(\mathcal{X}) = \{\varphi \, : \, \sup |\varphi|/W < +\infty\}$;
- the spectral radius Λ is an eigenvalue with «positive» eigenvector h:

$$Q^f h = \Lambda h\,;$$

Sketch of proof

The steps of the proof are as follows:

- Q^f is quasi-compact in $L^\infty_W(\mathcal{X}) = \{\varphi : \sup |\varphi|/W < +\infty\}$;
- the spectral radius Λ is an eigenvalue with «positive» eigenvector h:
$$Q^f h = \Lambda h;$$
- define the *Markovian* kernel $Q_h = \Lambda h^{-1} Q^f h$;
- Φ_n is rewritten
$$\Phi_n(\mu)(\phi) = \frac{\mu\big(h(Q_h)^n h^{-1}\varphi\big)}{\mu\big(h(Q^h)^n h^{-1}\big)};$$
- the kernel $Q_h = \Lambda h^{-1} Q^f h$ satifies the conditions of Hairer and Mattingly's theorem.

Sketch of proof

The steps of the proof are as follows:

- Q^f is quasi-compact in $L^\infty_W(\mathcal{X}) = \{\varphi : \sup |\varphi|/W < +\infty\}$;
- the spectral radius Λ is an eigenvalue with «positive» eigenvector h:
$$Q^f h = \Lambda h;$$
- define the *Markovian* kernel $Q_h = \Lambda h^{-1} Q^f h$;
- Φ_n is rewritten
$$\Phi_n(\mu)(\phi) = \frac{\mu\big(h(Q_h)^n h^{-1}\varphi\big)}{\mu\big(h(Q^h)^n h^{-1}\big)};$$
- the kernel $Q_h = \Lambda h^{-1} Q^f h$ satifies the conditions of Hairer and Mattingly's theorem.

Toy model application

Consider the following Markov chain in $\mathbb{R}^d$:

$$x_{n+1} = \rho x_n + \sigma G_n,$$

with $\rho \in (-1, 1)$, $(G_n)_{n\in\mathbb{N}}$ standard Gaussian r.v.

Toy model application

Consider the following Markov chain in $\mathbb{R}^d$:

$$x_{n+1} = \rho x_n + \sigma G_n,$$

with $\rho \in (-1, 1)$, $(G_n)_{n\in\mathbb{N}}$ standard Gaussian r.v.

Then, the Feynman–Kac dynamics is stable if

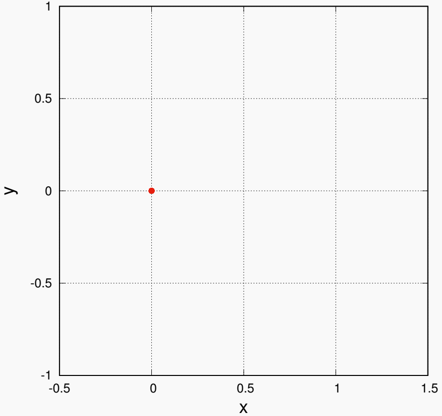
$$f(x) \leq a|x|^p + c,$$

with $a > 0$, $c \in \mathbb{R}$, and $0 \leq p < 2$. The *Lyapunov function* is

$$W(x) = \mathrm{e}^{\beta x^2}, \qquad \text{with } 0 < \beta < \frac{1-\rho^2}{2\sigma^2}.$$

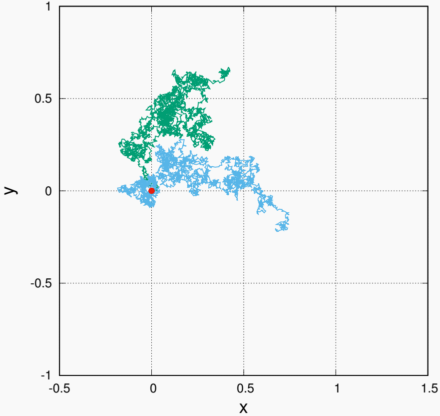
Back to O.U. process

- Dynamics $(x_{n+1}, y_{n+1}) = (x_n, y_n) - 0.1(x_n, y_n) + G_n \in \mathbb{R}^2$,
- Weight function $f(x, y) = a|x|^\alpha$, $a > 0$.



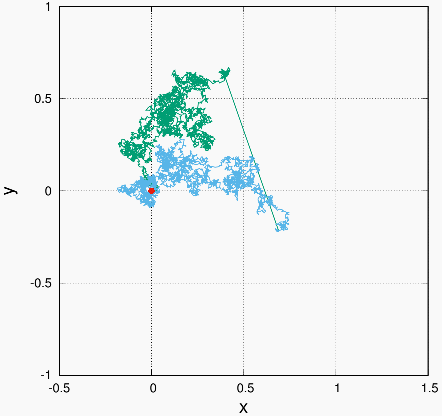
Back to O.U. process

- Dynamics $(x_{n+1}, y_{n+1}) = (x_n, y_n) - 0.1(x_n, y_n) + G_n \in \mathbb{R}^2$,
- Weight function $f(x, y) = a|x|^{\alpha}$, $a > 0$.



Back to O.U. process

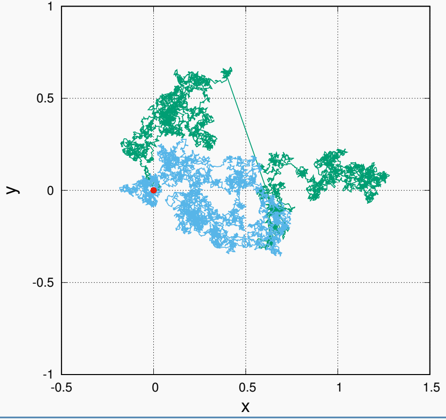
- Dynamics $(x_{n+1}, y_{n+1}) = (x_n, y_n) - 0.1(x_n, y_n) + G_n \in \mathbb{R}^2$,
- Weight function $f(x,y) = a|x|^\alpha$, $a > 0$.



Conclusion: for $\alpha > 2$, the algorithm blows up!

Back to O.U. process

- Dynamics $(x_{n+1}, y_{n+1}) = (x_n, y_n) - 0.1(x_n, y_n) + G_n \in \mathbb{R}^2$,
- Weight function $f(x, y) = a|x|^\alpha$, $a > 0$.

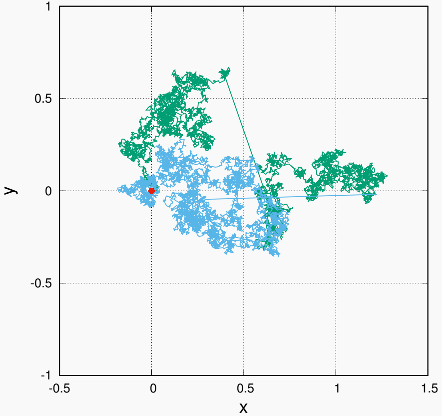


Conclusion: for $\alpha > 2$, the algorithm blows up!

For $\alpha = 2$ there is a limit value a^*.

Back to O.U. process

- Dynamics $(x_{n+1}, y_{n+1}) = (x_n, y_n) - 0.1(x_n, y_n) + G_n \in \mathbb{R}^2$,
- Weight function $f(x, y) = a|x|^{\alpha}$, $a > 0$.

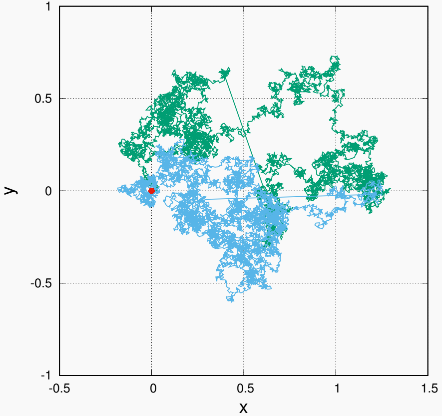


Conclusion: for $\alpha > 2$, the algorithm blows up!

For $\alpha = 2$ there is a limit value a^*.

Back to O.U. process

- Dynamics $(x_{n+1}, y_{n+1}) = (x_n, y_n) - 0.1(x_n, y_n) + G_n \in \mathbb{R}^2$,
- Weight function $f(x, y) = a|x|^{\alpha}$, $a > 0$.



Conclusion: for $\alpha > 2$, the algorithm blows up!

For $\alpha = 2$ there is a limit value a^*.

Discretization

Goal: discretize in time the quantity

$$\Theta_t(\mu)(\varphi) = \frac{\mathbb{E}_\mu\left[\varphi(X_t)\,\mathrm{e}^{\int_0^t f(X_s)ds}\right]}{\mathbb{E}_\mu\left[\mathrm{e}^{\int_0^t f(X_s)ds}\right]}.$$

Discretization

Goal: discretize in time the quantity

$$\Theta_t(\mu)(\varphi) = \frac{\mathbb{E}_\mu\left[\varphi(X_t)\,\mathrm{e}^{\int_0^t f(X_s)ds}\right]}{\mathbb{E}_\mu\left[\mathrm{e}^{\int_0^t f(X_s)ds}\right]}.$$

Idea: discretize (X_t) into $x_n \approx X_{n\Delta t}$, and consider *e.g.*

$$\Phi_{n,\Delta t}(\mu)(\phi) = \frac{\mathbb{E}_\mu\left[\varphi(x_n)\,\mathrm{e}^{\sum_{k=0}^{n-1} f(x_k)\Delta t}\right]}{\mathbb{E}_\mu\left[\mathrm{e}^{\sum_{k=0}^{n-1} f(x_k)\Delta t}\right]}.$$

Other possible choice:

$$\Phi_{n,\Delta t}(\mu)(\phi) = \frac{\mathbb{E}_\mu\left[\varphi(x_n)\,\mathrm{e}^{\sum_{k=0}^{n-1} \frac{f(x_k)+f(x_{k+1})}{2}\Delta t}\right]}{\mathbb{E}_\mu\left[\mathrm{e}^{\sum_{k=0}^{n-1} \frac{f(x_k)+f(x_{k+1})}{2}\Delta t}\right]}.$$

Discretization

Under *verifiable conditions* and if $\mathcal{X}$ is bounded, the following theorem holds.

Theorem: error estimate on the invariant measure [G.F., G. Stoltz, 2017]

There exist $\mu_f^*, \mu_{f,\Delta t}^* \in \mathcal{P}(\mathcal{X})$ such that $\forall\, \mu \in \mathcal{P}(\mathcal{X})$

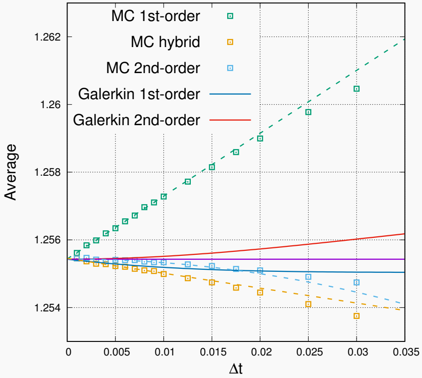
$$\Theta_t(\mu) \xrightarrow{t\to+\infty} \mu_f^*, \quad \Phi_n(\mu) \xrightarrow{n\to+\infty} \mu_{f,\Delta t}^*.$$

Moreover, there exist $p \geq 1$ and ψ solution to a Poisson equation such that

$$\int_{\mathcal{X}} \varphi \, d\mu_{f,\Delta t}^* = \int_{\mathcal{X}} \varphi \, d\mu_f^* + \Delta t^p \int_{\mathcal{X}} \varphi\psi \, d\mu_f^* + O(\Delta t^{p+1}).$$

Application

Overdamped Langevin dynamics on a one dimensional torus.



Average estimate for:

- $dX_t = (-V'(X_t) + 1)dt + dB_t$,
- $V(x) = 0.02\cos(2\pi x)$,
- $f = |V|^2$,
- $\varphi = \mathrm{e}^f$
- Euler-Maruyama scheme and 2nd order modified scheme,
- comparison to Galerkin discretization.

Conclusion

Take home message:

- new results on the ergodicity of Feynman–Kac dynamics;
- easily verifiable conditions;
- natural extension of the theory for Markov chains;
- numerical analysis for discretizations of SDE's;
- possible directions: applications to SPDE's, link with large deviations theory, application to systems of interacting particles...

References

Pierre Del Moral and Alice Guionnet.
On the stability of interacting processes with applications to filtering and genetic algorithms.
Annales de l'IHP Probabilités et statistiques, 37(2):155–194, 2001.

G. F., Mathias Rousset, and Gabriel Stoltz.
More on the stability of Feynman–Kac semigroups.
arXiv:1807.00390, 2018.

G. F. and Gabriel Stoltz.
Error estimates on ergodic properties of Feynman–Kac semigroups.
arXiv:1712.04013, 2017.

Martin Hairer and Jonathan C. Mattingly.
Yet another look at Harris' ergodic theorem for Markov chains.
In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, pages 109–117. Springer, 2011.

Luc Rey-Bellet.
Ergodic properties of Markov processes.
In *Open Quantum Systems II*, pages 1–39. Springer, 2006.

Spectrum of quasi-compact operators