Simulation of SDEs and mean-field SDEs: some recent results

Gonçalo dos Reis
University of Edinburgh (UK) & CMA/FCT/UNL (PT)

joint work with X. Chen & Z. Wilde (Edinb), and W. Stockinger (Imperial)

StatMath Research Seminar

WU, Wien, 09 Apr 2025

Partial funding by UIDB/00297/2020 and UIDP/00297/2020

Outline

1. Mean-field equations and Propagation of chaos
2. A setting of interest: super-linear Interaction MF kernel
   - Our results
   - Numerical results
3. Another setting of interest: Mean-field Langevin
   - Our results
   - Numerical results

Outline

1. Mean-field equations and Propagation of chaos
2. A setting of interest: super-linear Interaction MF kernel
   - Our results
   - Numerical results
3. Another setting of interest: Mean-field Langevin
   - Our results
   - Numerical results

McKean-Vlasov stochastic differential equations

MV-SDE* are SDE whose coefficients depend on the law of the solution:

$$\mathrm{d}X_t = \widehat{b}(t, X_t, \mu_t)\mathrm{d}t + \sigma(t, X_t, \mu_t)\mathrm{d}W_t, \quad X_0 \in L_0^p(\mathbb{R}^d), \qquad (MV-SDE)$$

where μ_t is the law of X_t, and W is a standard $\mathbb{R}^d$-BM. $\longrightarrow$ $\boxed{\text{All in } \mathbb{R}^d}$.

$W_2(\mu, \nu)$ is the 2-Wasserstein distance between μ, ν over space of finite 2nd moment prob. measure $\mathcal{P}_2(\mathbb{R}^d)$.

McKean-Vlasov stochastic differential equations

MV-SDE* are SDE whose coefficients depend on the law of the solution:

$$\mathrm{d}X_t = \widehat{b}(t, X_t, \mu_t)\mathrm{d}t + \sigma(t, X_t, \mu_t)\mathrm{d}W_t, \quad X_0 \in L_0^p(\mathbb{R}^d), \qquad (MV - SDE)$$

where μ_t is the law of X_t, and W is a standard $\mathbb{R}^d$-BM. $\longrightarrow$ $\boxed{\text{All in } \mathbb{R}^d}$.

$W_2(\mu, \nu)$ is the 2-Wasserstein distance between μ, ν over space of finite 2nd moment prob. measure $\mathcal{P}_2(\mathbb{R}^d)$.

Example (Convolution kernel MV-SDE)

$$X_t = X_0 + \int_0^t \Big\{ - X_s^3 + \big(\mathbb{E}[X_s] - X_s\big)\Big\} ds + \sigma W_t$$

McKean-Vlasov stochastic differential equations

MV-SDE* are SDE whose coefficients depend on the law of the solution:

$$\mathrm{d}X_t = \widehat{b}(t, X_t, \mu_t)\mathrm{d}t + \sigma(t, X_t, \mu_t)\mathrm{d}W_t, \quad X_0 \in L_0^p(\mathbb{R}^d), \qquad (MV - SDE)$$

where μ_t is the law of X_t, and W is a standard $\mathbb{R}^d$-BM. $\longrightarrow$ $\boxed{\text{All in } \mathbb{R}^d}$.

$W_2(\mu, \nu)$ is the 2-Wasserstein distance between μ, ν over space of finite 2nd moment prob. measure $\mathcal{P}_2(\mathbb{R}^d)$.

Example (Convolution kernel MV-SDE)

$$X_t = X_0 + \int_0^t \Big\{ - X_s^3 + \big(\mathbb{E}[X_s] - X_s\big)\Big\} ds + \sigma W_t$$

$$X_t = X_0 + \int_0^t b(s, X_s, \mu_s)ds + \int_0^t \int_{\mathbb{R}^d} K(X_s - y)d\mu_s(y)ds + \int_0^t \sigma(s, X_s, \mu_s)dW_s$$

In particle dynamics: b is *Confining Potential* and K is *Interaction Kernel*

Applications

These equations appear in many places.

- Controlling MV-SDE leads to Mean-field games
  - Finance, interacting agents in economics or opinion networks
  - Statistical mechanics, Molecular and fluid dynamics, Plasma Physics,
  - Dynamics of granular materials,
  - Chemistry of crystallisation
- Machine Learning:
  - MV-SDE as limits of (Deep) Neural networks
  - Generative Adversarial Networks (GAN): MFGs have the structure of GANs; and GANs are MFGs under the Pareto Optimality.

Applications

These equations appear in many places.

- Controlling MV-SDE leads to Mean-field games
  - Finance, interacting agents in economics or opinion networks
  - Statistical mechanics, Molecular and fluid dynamics, Plasma Physics,
  - Dynamics of granular materials,
  - Chemistry of crystallisation
- Machine Learning:
  - MV-SDE as limits of (Deep) Neural networks
  - Generative Adversarial Networks (GAN): MFGs have the structure of GANs; and GANs are MFGs under the Pareto Optimality.

Less trivial than it looks,

1. **No Flow property in** $\mathbb{R}^d$ but in $L^2(\Omega, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ or $\mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d)$:

$$X_t^{0,x} \neq X_t^{s, X_s^{0,x}}, \quad \text{for } t \in [0,\infty],\ r \in [0,t)$$

2. This leads to infinite dimensional calculus and difficult "PDEs"

$$[0,T] \times \mathbb{R}^d \times \mathcal{P}_2(\mathbb{R}^d) \ni (t,x,\mu) \mapsto u\big(t,x,\mu\big) \quad \Rightarrow \quad \text{What is } \partial_\mu u \,?$$

Its the kernels!

In many situations the kernel K falls short of neat assumptions:[1]

Example (Difficult kernels)

- Coloumb interaction: $K(x) = \frac{x}{|x|^d}$
- Bio-Savart law: $K(x) = \frac{x^\perp}{|x|^2}$ on $\mathbb{R}^2$
- Cucker-Smale flocking models: $K(x) = \frac{1}{|x|^\alpha}, \alpha > 0$
- Crystallisation: $K_p(x) = |x|^{-2p} - 2|x|^{-p}$ and take $p \to \infty$

Aside specific cases (e.g Coloumb in $d = 1, 2$, a general Wellposedness & PoC to all such cases is open).

In many situations a smooth and bounded approximation of the kernel is employed for modelling and theory.

[1] Harang and Mayorcas, "Pathwise Regularisation of Singular Interacting Particle Systems and their Mean Field Limits", 2020.

Approximation of MV-SDE – the IPS

LLN & Monte Carlo idea: $\mathbb{E}[X_t] \approx \frac{1}{N}\sum_{j=1}^{N} X_t^{j,N}$

This is in $(\mathbb{R}^d)^N$

[2] Sznitman, "Topics in propagation of chaos", 1991.

Approximation of MV-SDE – the IPS

$$\text{LLN \& Monte Carlo idea: } \mathbb{E}[X_t] \approx \frac{1}{N}\sum_{j=1}^{N} X_t^{j,N} \qquad \boxed{\text{This is in } (\mathbb{R}^d)^N}$$

A common technique for simulating MV-SDEs: interacting particle system:

$$\mathrm{d}X_t^{i,N} = \widehat{b}\Big(t, X_t^{i,N}, \mu_t^{X,N}\Big)\mathrm{d}t + \sigma\Big(t, X_t^{i,N}, \mu_t^{X,N}\Big)\mathrm{d}W_t^i, \quad \longrightarrow \boxed{\text{This is in } (\mathbb{R}^d)^N}$$

$$\mu_t^{X,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N} \delta_{X_t^{j,N}}(\mathrm{d}x), \qquad i = 1, \cdots, N$$

where $\delta_{X_t^{j,N}}$ is the Dirac measure at point $X_t^{j,N}$, and the Brownian motions $W^i, i = 1, \ldots, N$ are independent.

[2] Sznitman, "Topics in propagation of chaos", 1991.

Approximation of MV-SDE – the IPS

$$\text{LLN \& Monte Carlo idea: } \mathbb{E}[X_t] \approx \frac{1}{N}\sum_{j=1}^{N} X_t^{j,N} \qquad \boxed{\text{This is in } (\mathbb{R}^d)^N}$$

A common technique for simulating MV-SDEs: interacting particle system:

$$\mathrm{d}X_t^{i,N} = \widehat{b}\Big(t, X_t^{i,N}, \mu_t^{X,N}\Big)\mathrm{d}t + \sigma\Big(t, X_t^{i,N}, \mu_t^{X,N}\Big)\mathrm{d}W_t^i, \quad \longrightarrow \boxed{\text{This is in } (\mathbb{R}^d)^N}$$

$$\mu_t^{X,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N} \delta_{X_t^{j,N}}(\mathrm{d}x), \qquad i = 1, \cdots, N$$

where $\delta_{X_t^{j,N}}$ is the Dirac measure at point $X_t^{j,N}$, and the Brownian motions $W^i, i = 1, \dots, N$ are independent. "Propagation of chaos" (Sznitman '91)[2] : under appropriate conditions, as $N \to \infty$, for every i, the process $X^{i,N}$ converges to X^i, the solution of the MV-SDE driven by the Brownian motion W^i.

$$\lim_{N\to\infty} \sup_{1\le i\le N} \mathbb{E}\Big[\sup_{0\le t\le T} |X_t^{i,N} - X_t^i|^2\Big] = 0\,.$$

[2] Sznitman, "Topics in propagation of chaos", 1991.

Strong and weak Quantitative PoC

Strong PoC (based on[3])

$$(\text{in } L^p,\ p > 4) \qquad \sup_{1\le i\le N} \mathbb{E}\Big[\sup_{0\le t\le T} |X_t^i - X_t^{i,N}|^2\Big] \overset{!}{\le} C \begin{cases} N^{-1/2} & \text{if } d < 4, \\ N^{-1/2}\log(N) & \text{if } d = 4, \\ N^{-2/d} & \text{if } d > 4. \end{cases}$$

[3] Carmona and Delarue, *Probabilistic Theory of Mean Field Games with Applications I*, 2017.

[4] Chassagneux, Szpruch, and Tse, "Weak quantitative propagation of chaos via differential calculus on the space of measures", 2022.

[5] Haji-Ali, Hoel, and Tempone, "A simple approach to proving the existence, uniqueness, and strong and weak convergence rates for a broad class of McKean–Vlasov equations", 2021.

[6] Bernou and Duerinckx, "Uniform-in-time estimates on the size of chaos for interacting Brownian particles", 2024.

Strong and weak Quantitative PoC

Strong PoC (based on[3])

$$(\text{in } L^p,\ p > 4) \qquad \sup_{1\le i\le N} \mathbb{E}\Big[\sup_{0\le t\le T} |X_t^i - X_t^{i,N}|^2 \Big] \overset{!}{\le} C \begin{cases} N^{-1/2} & \text{if } d < 4, \\ N^{-1/2}\log(N) & \text{if } d = 4, \\ N^{-2/d} & \text{if } d > 4. \end{cases}$$

Weak PoC is much harder:

$$\sup_{h\in\mathfrak{F}} \Big| \mathbb{E}\big[h(X^i)\big] - \mathbb{E}\Big[\frac{1}{N}\sum_{k=1}^{N} h(X^{k,N})\Big] \Big| \overset{!}{=} \mathcal{O}\Big(\frac{1}{N}\Big) \qquad (\text{for some class } \mathfrak{F})$$

- For $T < \infty$: Chassagneux et al '22[4] and Haji-Ali et al '21[5]
- For $T \ge 0$: Bernou & Duerinckx '24[6] (so called "*Uniform in time PoC*")

[3] Carmona and Delarue, *Probabilistic Theory of Mean Field Games with Applications I*, 2017.

[4] Chassagneux, Szpruch, and Tse, “Weak quantitative propagation of chaos via differential calculus on the space of measures”, 2022.

[5] Haji-Ali, Hoel, and Tempone, “A simple approach to proving the existence, uniqueness, and strong and weak convergence rates for a broad class of McKean–Vlasov equations”, 2021.

[6] Bernou and Duerinckx, “Uniform-in-time estimates on the size of chaos for interacting Brownian particles”, 2024.

Approximation of MV-SDE: Objects of interest

Let $X_t^{i,N,n}$ be the i-th component of the particle system, discretized on $[0, T]$ over n steps. The Monte Carlo estimator of $\theta = \mathbb{E}[G(X)]$ writes

(eg pricing contracts in finance or summary statistics for statistical inference)

$$\hat{\theta}^{N,n} = \frac{1}{N} \sum_{i=1}^{N} G(X^{i,N,n}).$$

Approximation of MV-SDE: Objects of interest

Let $X_t^{i,N,n}$ be the i-th component of the particle system, discretized on $[0, T]$ over n steps. The Monte Carlo estimator of $\theta = \mathbb{E}[G(X)]$ writes
(eg pricing contracts in finance or summary statistics for statistical inference)

$$\hat{\theta}^{N,n} = \frac{1}{N}\sum_{i=1}^{N} G(X^{i,N,n}).$$

This approximation is affected by three sources of error:

- The statistical error: difference between $\hat{\theta}^{N,n}$ and $\mathbb{E}[G(X^{i,N,n})]$. The standard deviation of the statistical error is of order $\frac{1}{\sqrt{N}}$.

Approximation of MV-SDE: Objects of interest

Let $X_t^{i,N,n}$ be the i-th component of the particle system, discretized on $[0, T]$ over n steps. The Monte Carlo estimator of $\theta = \mathbb{E}[G(X)]$ writes

(eg pricing contracts in finance or summary statistics for statistical inference)

$$\hat{\theta}^{N,n} = \frac{1}{N} \sum_{i=1}^{N} G(X^{i,N,n}).$$

This approximation is affected by three sources of error:

- The statistical error: difference between $\hat{\theta}^{N,n}$ and $\mathbb{E}[G(X^{i,N,n})]$. The standard deviation of the statistical error is of order $\frac{1}{\sqrt{N}}$.
- The discretization error: difference between $\mathbb{E}[G(X^{i,N,n})]$ and $\mathbb{E}[G(X^{i,N})]$. Under Lipschitz assumptions the Euler scheme has weak error of order $\frac{1}{n}$.

Approximation of MV-SDE: Objects of interest

Let $X_t^{i,N,n}$ be the i-th component of the particle system, discretized on $[0, T]$ over n steps. The Monte Carlo estimator of $\theta = \mathbb{E}[G(X)]$ writes
(eg pricing contracts in finance or summary statistics for statistical inference)

$$\hat{\theta}^{N,n} = \frac{1}{N}\sum_{i=1}^{N} G(X^{i,N,n}).$$

This approximation is affected by three sources of error:

- The statistical error: difference between $\hat{\theta}^{N,n}$ and $\mathbb{E}[G(X^{i,N,n})]$. The standard deviation of the statistical error is of order $\frac{1}{\sqrt{N}}$.
- The discretization error: difference between $\mathbb{E}[G(X^{i,N,n})]$ and $\mathbb{E}[G(X^{i,N})]$. Under Lipschitz assumptions the Euler scheme has weak error of order $\frac{1}{n}$.
- The propagation of chaos error: difference between $\mathbb{E}[G(X^{i,N})]$ and $\mathbb{E}[G(X)]$. For G and X nice enough this error is also of order $\frac{1}{\sqrt{N}}$

Outline

1. Mean-field equations and Propagation of chaos
2. A setting of interest: super-linear Interaction MF kernel
   - Our results
   - Numerical results
3. Another setting of interest: Mean-field Langevin
   - Our results
   - Numerical results

Our setting: Super linear

Wrap up: σ is unif. Lip. in space-measure;
Drift: $\widehat{b} := b + K \star \mu$ such that: b is superlinear in space & Lip is measure;
K is odd & superlinear growth (one-sided Lipschitz)

Assumption ("super-measure-super-space")

- $\exists L > 0$ *such that for a.a.* $s \in [0, T]$, $\forall \mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ *and* $\forall x, y \in \mathbb{R}^d$,

$$\begin{aligned} &\langle b(s,x,\mu) - b(s,y,\mu), x-y \rangle \le L\|x-y\|^2, \\ &\|\sigma(s,x,\mu) - \sigma(s,y,\mu)\| \le L\|x-y\|, \\ &\|b(s,x,\mu) - b(s,x,\nu)\| + \|\sigma(s,x,\mu) - \sigma(s,x,\nu)\| \le LW_2(\mu,\nu). \end{aligned}$$

- $\exists L > 0, \exists \alpha \in (0,1]$ *such that* $\forall s, t \in [0, T]$, $\forall \mu \in \mathcal{P}_2(\mathbb{R}^d)$ *and* $\forall x \in \mathbb{R}^d$,

$$\|\sigma(t,x,\mu) - \sigma(s,x,\mu)\| \le L\|t-s\|^\alpha.$$

- $K(0) = 0$, $K(x) = -K(-x)$ *and* $\exists L \in \mathbb{R}$ *such that* $\forall x, y \in \mathbb{R}^d$,
$\big\langle K(x) - K(y), x-y \big\rangle \le L\|x-y\|^2$,
$\|K(x) - K(y)\| \le C\|x-y\|\big(1 + \|x\|^{r-1} + \|y\|^{r-1}\big), \quad \|K(x)\| \le C\big(1 + \|x\|^r\big)$.

The simulation problem

- Wellposedness//stability//PoC//invariant distribution//LDPs:
  - Growing collection of results under varied conditions[7],[8],[9]
- Numerics
  - PDE/FPE[10],[11]
  - Stochastic Euler schemes: Malrieu '03[12], Malrieu & Talay '06[13]

    Fully implicit scheme under strong structural assumptions (σ const)
  - If $\mu \mapsto \widehat{b}(\cdot,\cdot,\mu)$ is unif. Lip. then the answer is known

    ▷ Standard Euler, ▷ Taming , ▷ Time-adaptive, ▷ Split-Step methods,

    ▷ Randomised Milstein

---

[7] Zhang, "Existence and non-uniqueness of stationary distributions for distribution dependent SDEs", 2021.

[8] Dos Reis, Salkeld, and Tugaut, "Freidlin–Wentzell LDP in path space for McKean–Vlasov equations and the functional iterated logarithm law", 2019.

[9] Adams et al., "Large Deviations and Exit-times for reflected McKean-Vlasov equations with self-stabilizing terms and superlinear drifts", 2020.

[10] Baladron et al., "Mean-field description and propagation of chaos in networks of Hodgkin-Huxley and FitzHugh-Nagumo neurons", 2012.

[11] Goddard et al., "Noisy bounded confidence models for opinion dynamics: the effect of boundary conditions on phase transitions", 2022.

[12] Malrieu, "Convergence to equilibrium for granular media equations and their Euler schemes", 2003.

[13] Malrieu and Talay, "Concentration inequalities for Euler schemes", 2006.

MV-SDEs with super linear growth and standard Euler

The MV-SDE in $\mathbb{R}^d$ for $p \geq 2$

$$\mathrm{d}X_t = \widehat{b}(t, X_t, \mu_t^X)\mathrm{d}t + \sigma(t, X_t, \mu_t^X)\mathrm{d}W_t, \quad X_0 \in L_0^p(\mathbb{R}^d),$$

The particle approximation in $(\mathbb{R}^d)^N$

$$\mathrm{d}X_t^{i,N} = \widehat{b}\Big(t, X_t^{i,N}, \mu_t^{X,N}\Big)\mathrm{d}t + \sigma\Big(t, X_t^{i,N}, \mu_t^{X,N}\Big)\mathrm{d}W_t^i, \quad \mu_t^{X,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N} \delta_{X_t^{j,N}}(\mathrm{d}x)$$

where $\delta_{X_t^{j,N}}$ is the Dirac measure at point $X_t^{j,N}$, and the Brownian motions $W^i, i = 1, \dots, N$ are independent.

MV-SDEs with super linear growth and standard Euler

The MV-SDE in $\mathbb{R}^d$ for $p \geq 2$

$$\mathrm{d}X_t = \widehat{b}(t, X_t, \mu_t^X)\mathrm{d}t + \sigma(t, X_t, \mu_t^X)\mathrm{d}W_t, \quad X_0 \in L_0^p(\mathbb{R}^d),$$

The particle approximation in $(\mathbb{R}^d)^N$

$$\mathrm{d}X_t^{i,N} = \widehat{b}\Big(t, X_t^{i,N}, \mu_t^{X,N}\Big)\mathrm{d}t + \sigma\Big(t, X_t^{i,N}, \mu_t^{X,N}\Big)\mathrm{d}W_t^i, \quad \mu_t^{X,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^N \delta_{X_t^{j,N}}(\mathrm{d}x)$$

where $\delta_{X_t^{j,N}}$ is the Dirac measure at point $X_t^{j,N}$, and the Brownian motions $W^i, i = 1, \dots, N$ are independent.

Given a time partition $\{t_k\}_{k=0,\cdots,M}$, the **explicit** Euler scheme:

$$\bar{X}_{t_{k+1}}^{i,N,M} = \bar{X}_{t_k}^{i,N,M} + \widehat{b}\Big(t_k, \bar{X}_{t_k}^{i,N,M}, \bar{\mu}_{t_k}^{X,N}\Big)h + \sigma\Big(t_k, \bar{X}_{t_k}^{i,N,M}, \bar{\mu}_{t_k}^{X,N}\Big)\Delta W_{t_k}^i,$$

where $\bar{\mu}_{t_k}^{X,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^N \delta_{\bar{X}_{t_k}^{j,N,M}}(\mathrm{d}x)$, $\Delta W_{t_k}^i := W_{t_{k+1}}^i - W_{t_k}^i$ and $\bar{X}_0^{i,N,M} := X_0^i$.

Euler goes wrong

The stochastic Ginzburg Landau equation and with added mean field term,

$$\mathrm{d}X_t = \Big(\frac{\sigma^2}{2}X_t - X_t^3 + c\mathbb{E}[X_t]\Big)\mathrm{d}t + \sigma X_t \mathrm{d}W_t, \quad X_0 = x.$$

We simulate $N = 5000$ particles with a time step $h = 0.05$, $T = 2$ and $X_0 = 1$, we also take $\sigma = 3/2$ and $c = 1/2$.

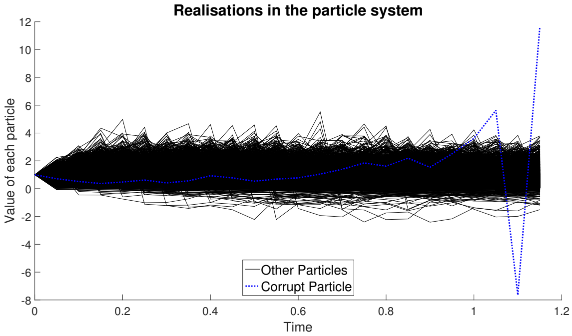


Figure: ‘Particle corruption’: the dashed particle is starting to oscillate and is taking larger values than its surrounding particles.

Split-Step method (SSM)

$$\mathrm{d}X_t = \Big[b(t, X_t, \mu_t^X) + v(t, X_t, \mu_t^X)\Big]\mathrm{d}t + \sigma(t, X_t, \mu_t^X)\mathrm{d}W_t, \quad X_0 \in L_0^p(\mathbb{R}^d),$$

with $v(t, x, \mu) = (K \star \mu)(x)$ conv. kernel.

Split-Step method (SSM)

$$\mathrm{d}X_t = \Big[b(t, X_t, \mu_t^X) + v(t, X_t, \mu_t^X)\Big]\mathrm{d}t + \sigma(t, X_t, \mu_t^X)\mathrm{d}W_t, \quad X_0 \in L_0^p(\mathbb{R}^d),$$

with $v(t, x, \mu) = (K \star \mu)(x)$ conv. kernel.

The $\boxed{\textbf{Split-Step method (SSM)}}$ scheme

$$Y_{t_k}^{i,\star,N} = \hat{X}_{t_k}^{i,N} + hv(t_k, Y_{t_k}^{i,\star,N}, \mu_{t_k}^{X,N}), \qquad \hat{\mu}_{t_k}^{Y,N}(dx) := \frac{1}{N}\sum_{j=1}^{N} \delta_{Y_{t_k}^{j,\star,N}}(dx) \tag{1}$$

$$\hat{X}_{t_{k+1}}^{i,N} = Y_{t_k}^{i,\star,N} + b(t_k, Y_{t_k}^{i,\star,N}, \hat{\mu}_{t_k}^{Y,N})h + \sigma(t_k, Y_{t_k}^{i,\star,N}, \hat{\mu}_{t_k}^{Y,N})\Delta W_n^i. \tag{2}$$

In a nutshell: solve super-linear/convolution component implicitly, then in (2), use the empirical measure of $Y_{t_k}^{i,\star,N}$ and deal with other terms.

Split-Step method (SSM)

$$\mathrm{d}X_t = \Big[b(t, X_t, \mu_t^X) + v(t, X_t, \mu_t^X)\Big]\mathrm{d}t + \sigma(t, X_t, \mu_t^X)\mathrm{d}W_t, \quad X_0 \in L_0^p(\mathbb{R}^d),$$

with $v(t, x, \mu) = (K \star \mu)(x)$ conv. kernel.

The $\boxed{\textbf{Split-Step method (SSM)}}$ scheme

$$Y_{t_k}^{i,\star,N} = \hat{X}_{t_k}^{i,N} + hv(t_k, Y_{t_k}^{i,\star,N}, \mu_{t_k}^{X,N}), \qquad \hat{\mu}_{t_k}^{Y,N}(dx) := \frac{1}{N}\sum_{j=1}^{N} \delta_{Y_{t_k}^{j,\star,N}}(dx) \tag{1}$$

$$\hat{X}_{t_{k+1}}^{i,N} = Y_{t_k}^{i,\star,N} + b(t_k, Y_{t_k}^{i,\star,N}, \hat{\mu}_{t_k}^{Y,N})h + \sigma(t_k, Y_{t_k}^{i,\star,N}, \hat{\mu}_{t_k}^{Y,N})\Delta W_n^i. \tag{2}$$

In a nutshell: solve super-linear/convolution component implicitly, then in (2), use the empirical measure of $Y_{t_k}^{i,\star,N}$ and deal with other terms.

Some advantages

- Implicit method for the bad drift components $\rightarrow$ more stable than explicit method.
- Time step restriction for solvability of implicit step is *artificial*: just $\pm\gamma x$
- (This is a type of Lie-Trotter splitting method)

Outline

Convergence results: Lipschitz diffusion

Theorem (Chen & GdR '22: SSM's MSE Conv (I))

Under monotonicity + Holder in time hold + $X_0 \in L^m(\mathbb{R}^d)$ *and* $\boxed{\sigma \text{ unif. Lip}}$

Let X^i *be the solution to the MV-SDE (driven by* W^i*), and* $X^{i,N,M}$ *be the SSM scheme. Then we obtain the following convergence result*

$$MSE := \sup_{1 \le i \le N} \mathbb{E}\Big[\sup_{0 \le t \le T} |X_t^{i,N} - X_t^{i,N,M}|^2\Big] \le Ch^{1-\varepsilon}, \quad \varepsilon > 0.$$

- Its very difficult to obtain L^p-moment bounds ($p > 2$) for the scheme.
  - critical to have $\sup_{\text{time}}$ inside expectation is that somewhere we use: $\mathbb{1}_{|X^{i,N,M}|>R} + \mathbb{1}_{|X^{i,N,M}|\le R}$

Convergence results: Lipschitz diffusion

Theorem (Chen & GdR '22: SSM's MSE Conv (I))

Under monotonicity + Holder in time hold + $X_0 \in L^m(\mathbb{R}^d)$ *and* $\boxed{\sigma \text{ unif. Lip}}$

Let X^i *be the solution to the MV-SDE (driven by* W^i*), and* $X^{i,N,M}$ *be the SSM scheme. Then we obtain the following convergence result*

$$MSE := \sup_{1\leq i\leq N} \mathbb{E}\big[\sup_{0\leq t\leq T} |X_t^{i,N} - X_t^{i,N,M}|^2\big] \leq Ch^{1-\varepsilon}, \quad \varepsilon > 0.$$

- Its very difficult to obtain L^p-moment bounds ($p > 2$) for the scheme.
  - critical to have $\sup_{\text{time}}$ inside expectation is that somewhere we use: $\mathbb{1}_{|X^{i,N,M}|>R} + \mathbb{1}_{|X^{i,N,M}|\leq R}$
- Exploit convolution structure but use that K is an odd function ☹

Convergence results: super linear growth diffusion

Theorem (Chen, GdR, & Stockinger '23: SSM's MSE Conv (II))

Under monotonicity + Holder in time hold + $X_0 \in L^m(\mathbb{R}^d)$ *and* $\boxed{\sigma \text{ polynomial} ☺}$

Let X^i *be the solution to the MV-SDE (driven by* W^i*), and* $X^{i,N,M}$ *be the SSM scheme. Then we obtain the following convergence result*

$$MSE_{sup\ outside} := \sup_{1 \leq i \leq N} \sup_{0 \leq t \leq T} \mathbb{E}[|X_t^{i,N} - X_t^{i,N,M}|^2] \leq Ch.$$

- Its much easier to obtain this result. One gets away with just L^2 estimates.
- We can have additionally a polynomial growth diffusion map

Convergence results: super linear growth diffusion

Theorem (Chen, GdR, & Stockinger '23: SSM's MSE Conv (II))

Under monotonicity + Holder in time hold + $X_0 \in L^m(\mathbb{R}^d)$ *and* $\boxed{\sigma \text{ polynomial} ☺}$

Let X^i *be the solution to the MV-SDE (driven by* W^i*), and* $X^{i,N,M}$ *be the SSM scheme. Then we obtain the following convergence result*

$$MSE_{sup\ outside} := \sup_{1 \le i \le N} \sup_{0 \le t \le T} \mathbb{E}[|X_t^{i,N} - X_t^{i,N,M}|^2] \le Ch.$$

- Its much easier to obtain this result. One gets away with just L^2 estimates.
- We can have additionally a polynomial growth diffusion map

let's see some comparative numerics (Euler, taming, time-adaptive, SSM)

Other schemes: Tamed Euler scheme & Time-adaptive

- **Taming:** *tamed* Euler explicit scheme.[14] With the notation above consider the following scheme $h := T/M$

$$\bar{X}^{i,N,M}_{t_{k+1}} = \bar{X}^{i,N,M}_{t_k} + \frac{\widehat{b}\Big(t_k, \bar{X}^{i,N,M}_{t_k}, \bar{\mu}^{X,N}_{t_k}\Big)}{1 + h^{\alpha}\Big|\widehat{b}\Big(t_k, \bar{X}^{i,N,M}_{t_k}, \bar{\mu}^{X,N}_{t_k}\Big)\Big|} h$$
$$+ \sigma\Big(t_k, \bar{X}^{i,N,M}_{t_k}, \bar{\mu}^{X,N}_{t_k}\Big)\Delta W^i_{t_k},$$

  where $\bar{\mu}^{X,N}_{t_k}(\mathrm{d}x) = \frac{1}{N}\sum_{j=1}^{N} \delta_{\bar{X}^{j,N,M}_{t_k}}(\mathrm{d}x)$ and $\alpha \in (0, 1/2]$ with $\bar{X}^{i,N,M}_0 = X^i_0$.

- **Time-adaptive.**[15]
  Just like standard explicit Euler. Timestep h is now $h(x)$ such that $|\widehat{b}(t, x, \mu)h(x)| \leq C(1 + |x|)$.

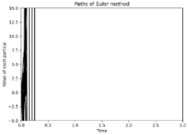
[14] Reis, Engelhardt, and Smith, "Simulation of McKean-Vlasov SDEs with super-linear growth", Jan. 2021.
[15] Reisinger and Stockinger, "An adaptive Euler-Maruyama scheme for McKean SDEs with super-linear growth and application to the mean-field FitzHugh-Nagumo model", 2020.

Outline

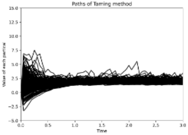
A stylised example

$$dX_t = \Big(v(X_t, \mu_t^X) + \mathbb{E}[X_t] \Big) dt + \frac{3}{10}(1 - X_t^2) dW_t, \quad X_0 \sim \mathcal{N}(2,2),$$

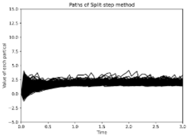
$$v(x,\mu) = -\frac{1}{4}x^3 + \int_{\mathbb{R}^d} -(x-y)^3 \mu(dy),$$



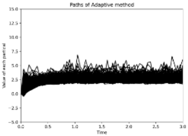
(a) Paths of the Euler method



(b) Paths of the Taming method



(c) Paths of the SSM



(d) Paths of the Adaptive

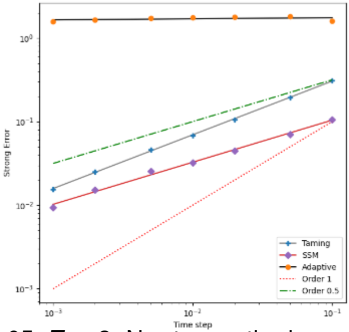
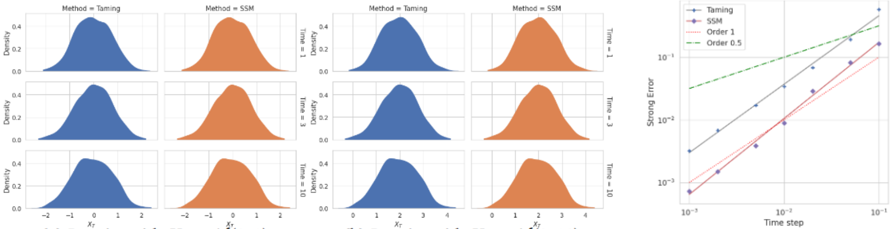


Figure: Simulations: $N = 100$ particles, $h = 0.05$, $T = 2$. Newton method $w = \sqrt{h}$.

Granular media type equation with additive noise

$$dX_t = v(X_t, \mu_t^X)dt + \sqrt{2}\, dW_t \text{ with } v(x,\mu) = \int_{\mathbb{R}^d} \Big(-\operatorname{sign}(x-y)|x-y|^2 \Big)\mu(dy),$$

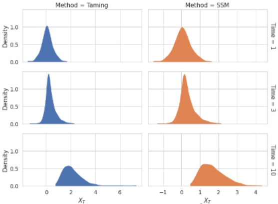


(a) Density with $X_0 \sim \mathcal{N}(0,1)$ (b) Density with $X_0 \sim \mathcal{N}(2,16)$ (c) Strong error

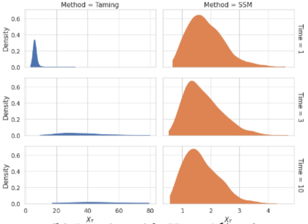
Figure: $N = 1000$ particles, $h = 0.01$. Density maps at $T = 1, 3, 10$ and strong convergence rates with $X_0 \sim \mathcal{N}(2, 16)$.

Double-well with Multiplicative noise

$$dX_t = \big(v(X_t, \mu_t^X) + X_t\big)dt + {\color{blue}X_t}dW_t \text{ with } v(x,\mu) = -\tfrac{1}{4}x^3 + \int_{\mathbb{R}^d} -(x-y)^3\mu(dy)$$



(a) Density with $X_0 \sim \mathcal{N}(0,1)$



(b) Density with $X_0 \sim \mathcal{N}(2,1)$

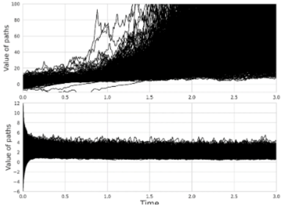


Figure: $N = 1000<$ $h = 0.01$ at times $T = 1, 3, 10$. Last Fig $t \in [0,3]$ and with $X_0 \sim \mathcal{N}(3,9)$.

Outline

Mean-field Langevin equations

We consider the 1-d mean-field Langevin (MFL) equation for $(X_t)_{t\geq 0} \in \mathbb{R}^1$:

$$X_t = \xi - \int_0^t \Big(\nabla U(X_s) + \nabla V * \mu_s(X_s)\Big)\mathrm{d}s + \sigma W_t, \tag{3}$$

where μ_t is the law of X_t, and W is a 1-d Brownian motion.

For functions U, V with some suitable regularity and convexity then

- X_t admits a unique stationary distribution μ^*, i.e., $\mathrm{Law}(X_t) \xrightarrow{d} \mu^*$ as $t \to \infty$
- μ^* has well-known implicit form

$$\mu^*(x) \propto \exp\Big(-\frac{2}{\sigma^2}U(x) - \frac{2}{\sigma^2}\int_{\mathbb{R}} V(x-y)\mu^*(\mathrm{d}y)\Big). \tag{4}$$

Thus,

▷ how sample from μ^* better than Euler/Milstein? (What is "better"?)

Preparation for main result

The IPS to (3) is for $i = 1, \cdots, N$

$$X_t^{i,N} = \xi^{i,N} - \int_0^t \Big(\nabla U(X_s^{i,N}) + \frac{1}{N} \sum_{j=1}^{N} \nabla V(X_s^{i,N} - X_s^{j,N}) \Big) \mathrm{d}s + \sigma W_t^i .$$

Preparation for main result

The IPS to (3) is for $i = 1, \cdots, N$

$$X_t^{i,N} = \xi^{i,N} - \int_0^t \Big(\nabla U(X_s^{i,N}) + \frac{1}{N}\sum_{j=1}^N \nabla V(X_s^{i,N} - X_s^{j,N})\Big)\mathrm{d}s + \sigma W_t^i.$$

Or written as a $\mathbb{R}^N$-valued map B as

$$\mathbb{R}^N \ni \boldsymbol{x} = (x_1, \ldots, x_N) \mapsto \boldsymbol{B}(\boldsymbol{x}) := \big(B_1(x_1, \ldots, x_N), \ldots, B_N(x_1, \ldots, x_N)\big),$$

$$\text{with} \quad B_i(\boldsymbol{x}) = B_i(x_1, \ldots, x_N) := -\nabla U(x_i) - \frac{1}{N}\sum_{j=1}^N \nabla V(x_i - x_j),$$

and we **re-write the IPS** for $(\boldsymbol{X}_t^N)_{t\geq 0} := (X_t^{1,N}, \ldots, X_t^{N,N})_{t\geq 0}$ as

$$\boldsymbol{X}_t^N = \boldsymbol{\xi} + \int_0^t B(\boldsymbol{X}_s^N)\mathrm{d}s + \sigma \boldsymbol{W}_t \tag{5}$$

$$\text{(Euler Scheme)} \Rightarrow \quad \boxed{\boldsymbol{X}_{i+1}^{N,h} = \boldsymbol{X}_i^{N,h} + hB(\boldsymbol{X}_i^{N,h}) + \sigma \Delta \boldsymbol{W}_{i+1}.} \tag{6}$$

The non-Markovian Euler scheme

The scheme introduced in Leimkuhler et al '14[16] for our IPS as a $\mathbb{R}^N$-valued SDE

$$\boldsymbol{X}_t^N = \boldsymbol{\xi} + \int_0^t B(\boldsymbol{X}_s^N)\mathrm{d}s + \sigma \boldsymbol{W}_t$$

$$\text{(n-ME Scheme)} \Rightarrow \quad \boxed{\boldsymbol{X}_{i+1}^{N,h} = \boldsymbol{X}_i^{N,h} + hB(\boldsymbol{X}_i^{N,h}) + \sigma\frac{1}{2}(\Delta\boldsymbol{W}_{i+1} + \Delta\boldsymbol{W}_i).} \quad (7)$$

[16] Leimkuhler, Matthews, and Tretyakov, "On the long-time integration of stochastic gradient systems", 2014.

The results for standard SDEs

Results for SDEs[17] $\rightarrow$ setting $\nabla V = 0$ in our case; $U \in C^7$ (in $\mathbb{R}^d$)

$(\sigma = cI_d)$	Strong $(T < \infty)$	Weak $(T < \infty)$	Weak $(T = \infty)$
Euler / Milstein	1	1	1
non-ME			

$$\text{Weak Error}^{\text{Euler}}(h; T) = C_T h + \mathcal{O}(h^2) \quad \text{where} \quad \lim_{T \to \infty} C_T = \text{Const} > 0.$$

[17] Leimkuhler, Matthews, and Tretyakov, "On the long-time integration of stochastic gradient systems", 2014.
[18] Ibid.

The results for standard SDEs

Results for SDEs[17] $\rightarrow$ setting $\nabla V = 0$ in our case; $U \in C^7$ (in $\mathbb{R}^d$)

$(\sigma = cI_d)$	Strong $(T < \infty)$	Weak $(T < \infty)$	Weak $(T = \infty)$
Euler / Milstein	1	1	1
non-ME		1	2

$$\text{Weak Error}^{\text{Euler}}(h; T) = C_T h + \mathcal{O}(h^2) \quad \text{where} \quad \lim_{T\to\infty} C_T = \text{Const} > 0.$$

but **for the non Markovian scheme** (Theorem 3.4[18])

$$\lim_{T\to\infty} C_T = 0 \quad \Rightarrow \quad \lim_{T\to\infty} \text{Weak Error}^{\text{non-Mark. Euler}}(h; T) = \mathcal{O}(h^2),$$

[17] Leimkuhler, Matthews, and Tretyakov, “On the long-time integration of stochastic gradient systems”, 2014.
[18] Ibid.

The results for standard SDEs

Results for SDEs[17] $\rightarrow$ setting $\nabla V = 0$ in our case; $U \in C^7$ (in $\mathbb{R}^d$)

$(\sigma = cI_d)$	Strong $(T < \infty)$	Weak $(T < \infty)$	Weak $(T = \infty)$
Euler / Milstein	1	1	1
non-ME	1/2	1	2

$$\text{Weak Error}^{\text{Euler}}(h; T) = C_T h + \mathcal{O}(h^2) \quad \text{where} \quad \lim_{T\to\infty} C_T = \text{Const} > 0.$$

but **for the non Markovian scheme** (Theorem 3.4[18])

$$\lim_{T\to\infty} C_T = 0 \quad \Rightarrow \quad \lim_{T\to\infty} \text{Weak Error}^{\text{non-Mark. Euler}}(h; T) = \mathcal{O}(h^2),$$

Lemma (Proposition 2.2)

[a] *Under Lip. the non-ME pointwise strong error is* $1/2$ *(also when* $\nabla V \neq 0$*)*

[a] Chen et al., "Improved weak convergence for the long time simulation of Mean-field Langevin equations", 2024.

[17] Leimkuhler, Matthews, and Tretyakov, "On the long-time integration of stochastic gradient systems", 2014.
[18] Ibid.

How to understand the results?

The SDE

$$dX(t) = B\big(X(t)\big)dt + \sigma dW(t), \quad X(0) = X_0$$

New view: Vilmart '15[19] conceptualised "*Postprocessed Integrators*" to study algorithms as $T \to \infty$. Instead of

$$\bar{X}_{n+1} = \bar{X}_n + hB\left(\bar{X}_n\right) + \frac{1}{2}\sigma\sqrt{h}\left(\xi_n + \xi_{n+1}\right)$$

[19] Vilmart, "Postprocessed integrators for the high order integration of ergodic SDEs", 2015.

How to understand the results?

The SDE

$$dX(t) = B\big(X(t)\big)dt + \sigma dW(t), \quad X(0) = X_0$$

New view: Vilmart '15[19] conceptualised "*Postprocessed Integrators*" to study algorithms as $T \to \infty$. Instead of

$$\bar{X}_{n+1} = \bar{X}_n + hB\left(\bar{X}_n\right) + \frac{1}{2}\sigma\sqrt{h}\left(\xi_n + \xi_{n+1}\right)$$

rewrite it as a "predictor-corrector" (postprocessed) method

$$X_{n+1} = X_n + hB\left(X_n + \frac{1}{2}\sigma\sqrt{h}\xi_n\right) + \sigma\sqrt{h}\xi_n,$$
$$\bar{X}_{n+1} = X_{n+1} + \frac{1}{2}\sigma\sqrt{h}\xi_{n+1}$$

Intuition... and our case

[19] Vilmart, "Postprocessed integrators for the high order integration of ergodic SDEs", 2015.

Outline

Assumptions

Assumption 1:
Let The potentials $U, V \in \mathcal{C}^2(\mathbb{R})$. Further suppose that

1. U is uniformly convex : there exists $\lambda > 0$ such that for all $x, y \in \mathbb{R}$,
$$\big(\nabla U(x) - \nabla U(y)\big)\big(x-y\big) \geq \lambda |x-y|^2. \tag{8}$$
2. V is even (thus ∇V is odd), and convex, i.e., for all $x, y \in \mathbb{R}$,
$$\big(\nabla V(x) - \nabla V(y)\big)(x-y) \geq 0,$$
and there exists $K_V > 0$ such that $|\nabla^2 V|_\infty \leq K_V$.

Assumption 2:

1. The potentials $U, V \in \mathcal{C}^7(\mathbb{R})$, and all derivatives of $\nabla U, \nabla V$ are uniformly bounded, with λ, K_V satisfy $\lambda \geq 7K_V$.
2. Let $N \in \mathbb{N}$ with $N \gg 6$. For any $n \leq 6$ and $(\gamma_1, \dots, \gamma_{|\gamma|}) = \gamma \in \bigcup_{k=1}^n \Pi_k^N$, with integers $\gamma_j \in \{1, \dots, N\}$, the function $g : \mathbb{R}^N \to \mathbb{R}$, satisfies $|\partial^{|\gamma|}_{x_{\gamma_1}, \dots, x_{\gamma_{|\gamma|}}} g|_\infty = \mathcal{O}(N^{-\hat{\mathcal{O}}(\gamma)})$, with an implied constant independent of N.

Weak error and the test functions g

We analyse the weak error:

$$\mathbb{E}[g(\boldsymbol{X}_T^N)] - \mathbb{E}[g(\boldsymbol{X}_T^{N,h})], \qquad \boldsymbol{X}_T^N, \boldsymbol{X}_T^{N,h} \in \mathbb{R}^N$$

Typical test functions g are

$$g(\boldsymbol{x}) = \tilde{g}\left(\frac{1}{N}\sum_{i=1}^{N} f(x_i)\right), \qquad \text{for some nice diff } f, \tilde{g},$$

using the associated Backward Kolmogorov equation[20],[21]

How does g behave? *(more difficult than the weak PoC test functions)*

- $|\partial^3_{x_1,x_2,x_3} g|_\infty = \mathcal{O}(N^{-3})$
- $|\partial^3_{x_1,x_1,x_3} g|_\infty = \mathcal{O}(N^{-2})$.
- If $f = \mathrm{id}$ then for any $|\gamma|$-order derivative, one has automatically $|\partial^{|\gamma|}_{x_{\gamma_1},\dots,x_{\gamma_{|\gamma|}}} g|_\infty = \mathcal{O}(N^{-|\gamma|})$.

[20] Talay and Tubaro, "Expansion of the global error for numerical schemes solving stochastic differential equations", 1990.

[21] Milstein and Tretyakov, *Stochastic numerics for mathematical physics*, 2004.

Main result

Theorem

Let Assumptions hold, let $\xi \in L^{10}(\Omega, \mathbb{R})$ *and let* $0 < h \ll \min\{1/2\lambda, 1\}$*. Then*

$$\Big|\mathbb{E}[g(\boldsymbol{X}_T^N)] - \mathbb{E}[g(\boldsymbol{X}_T^{N,h})]\Big| \approx K \exp(-\lambda_0 T)h + Kh^{3/2} + \mathcal{O}(h^2),$$

where $g : \mathbb{R}^N \to \mathbb{R}$ *is the weak-error test function for some positive constants* λ_0, K *independent of* h, T, M *and* N*.*

▷ **Main difficulties**:
Start point: $\mathbb{R}^N \ni \boldsymbol{x} \mapsto u(t, \boldsymbol{x}) = \mathbb{E}\big[g(\boldsymbol{X}_T^{N,t,\boldsymbol{x}}) \mid X_t^{N,t,\boldsymbol{x}} = \boldsymbol{x}\big]$.
▷ Taylor expansions

(a) K, λ_0 independent of N, T + exponentially decay over time and
(b) across 6-variation orders of $u(t, \boldsymbol{x})$
thus

$$\mathbb{R}^N \ni \boldsymbol{x} \mapsto \boldsymbol{X}_T^{N,x}, \text{ i.e., } \nabla_x \boldsymbol{X}_T^{N,x}, \nabla_{xx}^2 \boldsymbol{X}_T^{N,x} \ldots$$

Some results

Proposition

$$|\partial^2_{x_j,x_k} u(t,\boldsymbol{x})|^2$$

$$= \left| \mathbb{E}\Big[\sum_{i=1}^N \partial_{x_i} g(\boldsymbol{X}_T^{t,\boldsymbol{x},N}) X_{T,x_j,x_k}^{t,x_i,i,N} \Big] + \mathbb{E}\Big[\sum_{i=1}^N \sum_{i'=1}^N \partial^2_{x_i,x_{i'}} g(\boldsymbol{X}_T^{t,\boldsymbol{x},N}) X_{T,x_j}^{t,x_i,i,N} X_{T,x_k}^{t,x_{i'},i',N} \Big] \right|^2$$

$$|\partial^n_{x_{\gamma_1},\dots,x_{\gamma_n}} u(t,\boldsymbol{x})|^2$$

$$= \left| \mathbb{E}\left[\sum_{\substack{\alpha,\beta\in\bigcup_{k=0}^{n-1}\Pi_k^N,\\ \gamma\setminus(\gamma_1)\in\alpha\sqcup\beta}} \sum_{i=1}^N \Big(\partial_{x_i} g(\boldsymbol{X}_T^{t,\boldsymbol{x},N}) \Big)_{x_{\alpha_1},\dots,x_{\alpha_{|\alpha|}}} \big(X_{T,x_{\gamma_1}}^{t,x_i,i,N} \big)_{x_{\beta_1},\dots,x_{\beta_{|\beta|}}} \right] \right|^2 .$$

For the first variation process (K indep. of N)

$$\sum_{i=1}^N \mathbb{E}\Big[|X_{s,x_j}^{t,x_i,i,N}|^p \Big] \le K e^{-\lambda p(s-t)}, \ \textit{and} \ \sum_{i=1,i\neq j}^N \mathbb{E}\Big[|X_{s,x_j}^{t,x_i,i,N}|^p \Big] \le \frac{K}{N^{p-1}} e^{-\lambda_1 p(s-t)}.$$

Outline

1. Mean-field equations and Propagation of chaos
2. A setting of interest: super-linear Interaction MF kernel
   - Our results
   - Numerical results
3. Another setting of interest: Mean-field Langevin
   - Our results
   - Numerical results

A basic example

Take the linear example:

$$\mathrm{d}X_t = \Big(-\alpha\big(X_t - \mathbb{E}[X_t]\big) - X_t \Big)\mathrm{d}t + \sigma \mathrm{d}W_t, \quad X_0 \in L^{10}(\Omega, \mathbb{R}), \tag{9}$$

where $\alpha, \sigma > 0$. We have $\mathbb{E}[X_t] = \mathbb{E}[X_0]e^{-t}$ and

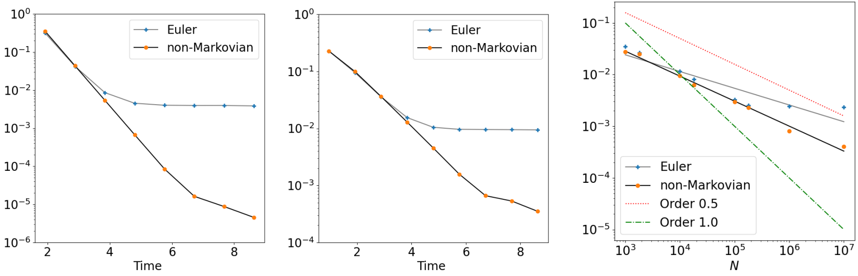
$$\mu^*(x) = \frac{1}{Z}\exp\Big(-\frac{\alpha+1}{\sigma^2}x^2 \Big), \qquad Z := \int_{\mathbb{R}} \mu^*(x)\mathrm{d}x. \tag{10}$$

We compute the relative entropy error and the L_2-Error (of the density)

$$\begin{aligned} \text{Relative Entropy Error } &= \sum_{i=1}^{N_{\text{bins}}} \mu_i^{\text{true}} \ln\Big(\frac{\mu_i^{\text{true}}}{\mu_i^{\text{approx}}}\Big) \\ L_2(\mathbb{R})\text{-Error } &= \sqrt{\sum_{i=1}^{N_{\text{bins}}} |\mu_i^{\text{true}} - \mu_i^{\text{approx}}|^2}, \end{aligned}$$

where $N_{\text{bins}} \sim 100$ is partition of $\mathbb{R}$.

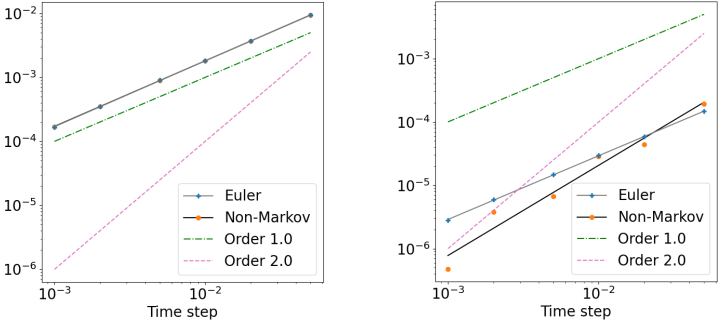
Numerical results in a stylised (linear) example



(a) Relative Entropy Error (b) L_2-Error (c) PoC L_2-Error (log-scale)

Figure: Simulation of the linear MV-SDE with $\alpha = 0.5, \sigma = 0.8$, $N = 10^7$, $h = 0.16$, and $X_0 \sim \mathcal{N}(\pi, 1)$. (a) Entropy Error of the Euler method and non-Markovian method in log-scale over time. (b) L_2-Error of the Euler method and non-Markovian method in log-scale over time. (c) L_2-Error in particle size N of the Euler method and non-Markovian method in log-scale with different N at $T = 9$.

Numerical results in a stylized (linear) example



(a) Weak err. at $t = 1$

(b) Weak err. at $t = 7$

Figure: Simulation of the linear MV-SDE with $\alpha = 0.5, \sigma = 0.8$, $N = 10^7$, $h = 0.16$, and $X_0 \sim \mathcal{N}(\pi, 1)$. (a) Weak error in particle size N of the Euler method and non-Markovian method in log-scale with different N at $T = 1$ (b) L_2-Error in particle size N of the Euler method and non-Markovian method in log-scale with different N at $T = 7$.

Error in number of particles

α	σ	a	b	N_{bins}	N	Entropy Error		L_2-Error	
						Euler	NM	Euler	NM
0.5	0.8	-1.8	1.8	72	10^3	-	-	2.89E-02	3.28E-02
					10^4	-	-	1.01E-02	1.04E-02
					10^5	8.21E-04	4.83E-04	4.29E-03	3.10E-03
					10^6	2.74E-04	4.66E-05	2.31E-03	1.26E-03
					10^7	2.33E-04	4.71E-06	2.37E-03	3.56E-04

Table: Simulation results for MV-SDE (9) with $h = 0.04$ and $T = 8.64$ for increasing numbers of particles N. (As for Fig. 5: $X_0 \sim \mathcal{N}(\pi, 1)$ and both schemes run on the exact same samples of the initial condition and Brownian increments.)

Thank you!

Thank you for your time!

[22] Chen, Xingyuan, and GdR, (2024) *Euler simulation of interacting particle systems and McKean–Vlasov SDEs with fully super-linear growth drifts in space and interaction.* IMA Journal of Numerical Analysis 44, no. 2 (2024): 751-796.

[23] Chen, Xingyuan, GdR, Wolfgang Stockinger, and Zac Wilde, (2024) *Improved weak convergence for the long time simulation of Mean-field Langevin equations*,

▷ preprint arXiv:2405.01346

[22] Chen and Dos Reis, “Euler simulation of interacting particle systems and McKean–Vlasov SDEs with fully super-linear growth drifts in space and interaction”, 2024.

[23] Chen et al., “Improved weak convergence for the long time simulation of Mean-field Langevin equations”, 2024.

Extra Slides

The Wasserstein metric

Wasserstein distance $W^{(2)}(\mu, \nu)$.

Over $\mathbb{R}^d$, set the space of probability measures as $\mathcal{P}(\mathbb{R}^d)$ and its subset $\mathcal{P}_2(\mathbb{R}^d)$ of those with finite second moment.

The Wasserstein distance metricizes the weak convergence of probability measures and is defined as

$$W_2(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \Big(\int_{\mathbb{R}^d \times \mathbb{R}^d} |x - y|^2 \pi(dx, dy) \Big)^{\frac{1}{2}}, \quad \mu, \nu \in \mathcal{P}_2(\mathbb{R}^d),$$

where $\Pi(\mu, \nu) \subset \mathcal{P}(\mathbb{R}^d \times \mathbb{R}^d)$ is the set of couplings for μ and ν such that $\pi \in \Pi(\mu, \nu)$ is a probability measure on $\mathbb{R}^d \times \mathbb{R}^d$ such that $\pi(\cdot \times \mathbb{R}^d) = \mu$ and $\pi(\mathbb{R}^d \times \cdot) = \nu$.

Weak error methodologies

How does one go about showing weak errors?

- Talay-Tubaro[24] but see Milstein Tretyakov book (2nd edition 2021)[25]
  - ▷ Feynman-Kac and exogenous PDE result
- Itô-Taylor expansions[26]
  - ▷ Expansions of drift and diffusion using the SDE itself and over a simplex
- Malliavin calculus + Duality[27]
  - ▷ Integration by parts, and pathwise analysis
- Parametrix expansions[28]
  - ▷ Expansion of the densities
- ad-hoc // by hand

[24] Talay and Tubaro, "Expansion of the global error for numerical schemes solving stochastic differential equations", 1990.

[25] Milstein and Tretyakov, *Stochastic numerics for mathematical physics*, 2004.

[26] Kloeden and Platen, *Numerical solution of stochastic differential equations*, 1992.

[27] Clément, Kohatsu-Higa, and Lamberton, "A duality approach for the weak approximation of stochastic differential equations", 2006.

[28] Konakov and Menozzi, "Weak error for stable driven stochastic differential equations: Expansion of the densities", 2011.

Back to the Analysis: Kolmogorov backward equation

Kolmogorov backward equation

We introduce $\boldsymbol{X}_s^{t,\boldsymbol{x},N} = (X_s^{t,x_1,1,N}, \ldots, X_s^{t,x_N,N,N})$, where for $i \in \{1, \ldots, N\}$

$$X_s^{t,x_i,i,N} = x_i + \int_t^s B_i(X_u^{t,x_1,1,N}, \ldots, X_u^{t,x_N,N,N})\mathrm{d}u + \sigma(W_s^i - W_t^i).$$

The generator for is defined by

$$\mathcal{L}_N = \sum_{i=1}^{N} B_i \partial_{x_i} + \frac{1}{2}\sigma^2 \partial^2_{x_i,x_i},$$

We introduce the Kolmogorov backward equation:

$$\partial_t u + \mathcal{L}_N u = 0, \quad t \in [0, T), \quad u(T, \boldsymbol{x}) = g(\boldsymbol{x}), \tag{11}$$

for the above test function $g : \mathbb{R}^N \to \mathbb{R}$, by the Feynman-Kac formula the solution of the above PDE is given by

$$u(t, \boldsymbol{x}) = \mathbb{E}\left[g(\boldsymbol{X}_T^N) \mid X_t^{i,N} = x_i, i \in \{1, \ldots, N\}\right]. \tag{12}$$

Weak-Error expansions

$$\mathbb{E}\Big[g(\boldsymbol{X}_T^N)\Big] - \mathbb{E}\Big[g(\boldsymbol{X}_T^{N,h})\Big] = h^2\mathbb{E}\left[\sum_{m=0}^{M-1} L(t_m, \boldsymbol{X}_{t_m}^{N,h})\right] + \mathbb{E}\left[\sum_{m=0}^{M-1} R(t_m, \boldsymbol{X}_{t_m}^{N,h})\right], \tag{13}$$

where the map $L : \mathbb{R}_+ \times \mathbb{R}^N \to \mathbb{R}$ is defined via the maps u and $(B_i)_{i\in\{1,\dots,N\}}$:

$$\begin{aligned} L(t,\boldsymbol{x}) = \frac{1}{2}\Big[\sum_{i,j=1}^{N} B_j(\boldsymbol{x})\partial_{x_j} B_i(\boldsymbol{x})\partial_{x_i} u(t,\boldsymbol{x}) + \frac{\sigma^2}{2}\sum_{i,j=1}^{N} \partial_{x_j} B_i(\boldsymbol{x})\partial^2_{x_i,x_j} u(t,\boldsymbol{x}) \\ + \frac{\sigma^2}{2}\sum_{i,j=1}^{N} \partial^2_{x_j,x_j} B_i(\boldsymbol{x})\partial_{x_i} u(t,\boldsymbol{x})\Big]. \end{aligned} \tag{14}$$

The remainder term $R(\cdot,\cdot)$ will later be written as a linear combination of 8 remainder terms, we need to control all the summations...

Kolmogorov backward equation Examples

Consider the first derivatives, by chain rule, we need to analysis the derivatives of g and the variation processes

$$
\begin{aligned}
&|\partial_{x_j} u(t,\boldsymbol{x})|^2 \\
&\quad = \Big|\mathbb{E}\Big[\sum_{i=1}^{N} \big(\partial_{x_i} g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})\big)\cdot\big(X_{T,x_j}^{t,x_i,i,N}\big)\Big]\Big|^2 \\
&\quad \leq 2\Big|\mathbb{E}\Big[|\partial_{x_j} g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})|\,|X_{T,x_j}^{t,x_j,j,N}|\Big]\Big|^2 + 2\Big|\mathbb{E}\Big[\sum_{i=1,\ i\neq j}^{N} \big(\partial_{x_i} g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})\big)\cdot\big(X_{T,x_j}^{t,x_i,i,N}\big)\Big]\Big|^2 \\
&\quad \leq \frac{K}{N^2}\mathbb{E}\Big[|X_{T,x_j}^{t,x_j,j,N}|^2\Big] + KN\sum_{i=1,\ i\neq j}^{N}\mathbb{E}\Big[\Big||\partial_{x_i} g(\boldsymbol{X}_T^{t,\boldsymbol{x},N})|\,|X_{T,x_j}^{t,x_i,i,N}|\Big|^2\Big] \\
&\quad \leq \frac{K}{N^2}\mathbb{E}\Big[|X_{T,x_j}^{t,x_j,j,N}|^2\Big] + \frac{K}{N}\sum_{i=1,\ i\neq j}^{N}\mathbb{E}\Big[|X_{T,x_j}^{t,x_i,i,N}|^2\Big],
\end{aligned}
$$

where we want $\partial_{x_j} u(t,\boldsymbol{x}) \sim \mathcal{O}(1/N)$ so that $|\partial_{x_j} u(t,\boldsymbol{x})|^2 \sim \mathcal{O}(1/N^2)$

Kolmogorov backward equation Examples-3

Similarly for the second derivatives

$$
\begin{aligned}
&|\partial^2_{x_j,x_k} u(t,\boldsymbol{x})|^2 \\
&= \left| \mathbb{E}\Big[\sum_{i=1}^{N} \partial_{x_i} g(\boldsymbol{X}_T^{t,\boldsymbol{x},N}) X_{T,x_j,x_k}^{t,x_i,i,N} \Big] + \mathbb{E}\Big[\sum_{i=1}^{N} \sum_{i'=1}^{N} \partial^2_{x_i,x_{i'}} g(\boldsymbol{X}_T^{t,\boldsymbol{x},N}) X_{T,x_j}^{t,x_i,i,N} X_{T,x_k}^{t,x_{i'},i',N} \Big] \right|^2
\end{aligned}
$$

The n-th derivatives

$$
\begin{aligned}
&|\partial^n_{x_{\gamma_1},\dots,x_{\gamma_n}} u(t,\boldsymbol{x})|^2 \\
&= \left| \mathbb{E}\Bigg[\sum_{\substack{\alpha,\beta \in \bigcup_{k=0}^{n-1} \Pi_k^N, \\ \gamma \backslash (\gamma_1) \in \alpha \sqcup \beta}} \sum_{i=1}^{N} \Big(\partial_{x_i} g(\boldsymbol{X}_T^{t,\boldsymbol{x},N}) \Big)_{x_{\alpha_1},\dots,x_{\alpha_{|\alpha|}}} \big(X_{T,x_{\gamma_1}}^{t,x_i,i,N} \big)_{x_{\beta_1},\dots,x_{\beta_{|\beta|}}} \Bigg] \right|^2
\end{aligned}
$$

Basically, we need to analysis and take many summations so to match all the orders in derivatives of g and the variation processes....

Orders: Properly grouping + Jensen's inequality

Consider now the specific two-dimensional example of $x_{\gamma_1,\gamma_2} = N^{1-\hat{\mathcal{O}}(\gamma)}$ (corresponding to a 2×2 matrix with diagonal entries 1 and otherwise $1/N$).

$$
\begin{aligned}
\Big| \sum_{\gamma \in \Pi_2^N} x_{\gamma_1,\gamma_2} \Big|^2 &= N^4 \Big| \frac{1}{N^2} \sum_{\gamma \in \Pi_2^N} x_{\gamma_1,\gamma_2} \Big|^2 \leq N^2 \sum_{i,j=1}^{N} |x_{i,j}|^2 \\
&= N^2 \sum_{i=1}^{N} |x_{i,i}|^2 + N^2 \sum_{i,j=1, i\neq j}^{N} |x_{i,j}|^2 = N^3 + N^2 \leq 2N^3.
\end{aligned}
$$

This estimate is too naive and can be improved, as we can instead consider

$$
\begin{aligned}
\Big| \sum_{\gamma \in \Pi_2^N} x_{\gamma_1,\gamma_2} \Big|^2 &\leq 2 \Big| \sum_{i=1}^{N} x_{i,i} \Big|^2 + 2 \Big| \sum_{i,j=1, i\neq j}^{N} x_{i,j} \Big|^2 \leq 2N \sum_{i=1}^{N} |x_{i,i}|^2 + 2N^2 \sum_{i,j=1, i\neq j}^{N} |x_{i,j}|^2 \\
&= 2N^2 + \frac{2N^3(N-1)}{N^2} \leq 4N^2,
\end{aligned}
$$

which is indeed a sharper upper bound.

The variation processes

The first variation process of $(\boldsymbol{X}_s^{t,\boldsymbol{x},N})_{s\geq t\geq 0}$ is given by

$$X_{s,x_j}^{t,x_i,i,N} = \delta_{i,j} + \int_t^s \sum_{l=1}^N \partial_{x_l} B_i(\boldsymbol{X}_u^{t,\boldsymbol{x},N}) X_{u,x_j}^{t,x_l,l,N} \mathrm{d}u,$$

The $n-$variation process of $(\boldsymbol{X}_s^{t,\boldsymbol{x},N})_{s\geq t\geq 0}$ is given by

$$X_{s,x_{\gamma_1},\dots,x_{\gamma_n}}^{t,x_i,i,N} = \int_t^s \Big(\sum_{l=1}^N \partial_{x_l} B_i(\boldsymbol{X}_u^{t,\boldsymbol{x},N}) X_{u,x_{\gamma_1}}^{t,x_l,l,N} \Big)_{x_{\gamma_2},\dots,x_{\gamma_n}} \mathrm{d}u \tag{15}$$

$$= \int_t^s \sum_{l=1}^N \partial_{x_l} B_i(\boldsymbol{X}_u^{t,\boldsymbol{x},N}) X_{u,x_{\gamma_1},\dots,x_{\gamma_n}}^{t,x_l,l,N} \mathrm{d}u$$

$$+ \sum_{\substack{\alpha,\beta\in\bigcup_{k=0}^{n-1}\Pi_k^N, \\ |\alpha|>0,\ \gamma\setminus(\gamma_1)\in\alpha\sqcup\beta}} \int_t^s \sum_{l=1}^N \Big(\partial_{x_l} B_i(\boldsymbol{X}_u^{t,\boldsymbol{x},N}) \Big)_{x_{\alpha_1},\dots,x_{\alpha_{|\alpha|}}} \Big(X_{u,x_{\gamma_1}}^{t,x_l,l,N} \Big)_{x_{\beta_1},\dots,x_{\beta_{|\beta|}}} \mathrm{d}u,$$

Some interesting results of the variation processes

Under the assumptions we have with the the starting positions $x_i \in L^2(\Omega, \mathbb{R})$ are $\mathcal{F}_t$-measurable random variables that are identically distributed over all $i \in \{1, \dots, N\}$. For each $1 \leq n \leq 6$, there exist constants $\lambda_0^{(n)} \in (0, \lambda)$ and $K > 0$ (both independent of s, t, T and N) such that for any $m \in \{1, \dots, n+1\}$, we have

$$\sum_{\gamma \in \Pi_{n+1}^N,\ \hat{\mathcal{O}}(\gamma)=m} \mathbb{E}\Big[|X_{s,x_{\gamma_2},\dots,x_{\gamma_{n+1}}}^{t,x_{\gamma_1},\gamma_1,N}|^p\Big] \leq \frac{K}{N^{p(m-1)-m}} e^{-\lambda_0^{(n)} p(s-t)}.$$

This implies that, for all $\gamma \in \Pi_{n+1}^N$, such that $\hat{\mathcal{O}}(\gamma) = m,\ m \in \{1, \dots, n+1\}$:

$$\mathbb{E}\Big[|X_{s,x_{\gamma_2},\dots,x_{\gamma_{n+1}}}^{t,x_{\gamma_1},\gamma_1,N}|^p\Big] \leq \frac{K}{N^{p(m-1)}} e^{-\lambda_0^{(n)} p(s-t)}.$$

Example (The first variation process)

$$\sum_{i=1}^{N} \mathbb{E}\Big[|X_{s,x_j}^{t,x_i,i,N}|^p\Big] \leq K e^{-\lambda p(s-t)}, \text{ and } \sum_{i=1, i\neq j}^{N} \mathbb{E}\Big[|X_{s,x_j}^{t,x_i,i,N}|^p\Big] \leq \frac{K}{N^{p-1}} e^{-\lambda_1 p(s-t)}.$$

More results

There exists a constant $K > 0$ (independent of t, T, N), such that for any $n \in \mathbb{N}, 1 \leq n \leq 6$, $\gamma \in \Pi_n^N$, and $\boldsymbol{x} \in \mathbb{R}^N$

$$
\begin{aligned}
&|\partial^n_{x_{\gamma_1},\dots,x_{\gamma_n}} u(t,\boldsymbol{x})|^2 \\
&\quad \leq K \sum_{m=0}^{n} \sum_{\substack{\ell \in \bigcup_{k=1}^n \Pi_k^N, \\ \hat{\mathcal{O}}(\ell \bigcup \gamma) = \hat{\mathcal{O}}(\gamma)+m}} N^{m-2\hat{\mathcal{O}}(\ell)} \sum_{\substack{\alpha_1,\dots,\alpha_{|\ell|} \in \bigcup_{k=1}^n \Pi_k^N, \\ \bigcup_{i=1}^{|\ell|} \alpha_i \simeq \gamma}} \mathbb{E}\Big[\prod_{i=1}^{|\ell|} \Big| X^{t,x_{\ell_i},\ell_i,N}_{T,\alpha_{i,1},\dots,\alpha_{i,|\alpha_i|}} \Big|^2 \Big],
\end{aligned}
$$

where $\alpha_i = (\alpha_{i,1}, \dots, \alpha_{i,|\alpha_i|})$ and $\alpha_{i,j} \in \{1, \dots, N\}$ for $j \in \{1, \dots, |\alpha_i|\}$.

Further, assuming that the starting points x_i are $\mathcal{F}_t$-measurable random variables in $L^2(\Omega, \mathbb{R})$ sampled from the same distribution for all $i \in \{1, \dots, N\}$, we have

$$
\mathbb{E}\Big[\big| \partial^n_{x_{\gamma_1},\dots,x_{\gamma_n}} u(t,\boldsymbol{x}) \big|^2 \Big] \leq K e^{-\lambda_0 (T-t)} N^{-2\hat{\mathcal{O}}(\gamma)}.
$$