the Split-step scheme for McKean-Vlasov stochastic differential equations (MV-SDEs) with super-linearity in space and measure

Xingyuan Chen

University of Edinburgh (UK)

joint work with Gonçalo dos Reis (UoE) and Wolfgang Stockinger (Oxford)

London, IMS 2022

Outline

MV-SDEs with super-linearity in space and measure

We will consider McKean-Vlasov stochastic differential equations(MV-SDEs) for $\{X_t\} \in \mathbb{R}^d$ of the following type:

$$\begin{aligned} \mathrm{d}X_t &= \hat{b}(t, X_t, \mu_t^X)\mathrm{d}t + \sigma(t, X_t, \mu_t^X)\mathrm{d}W_t, \quad X_0 \in L_0^m(\mathbb{R}^d), \quad \hat{b} = v + b \qquad (1) \\ &= \Big(\underbrace{\int_{\mathbb{R}^d} f(X_t - y)\mu_t^X(\mathrm{d}y) + u(X_t, \mu_t^X)}_{v(X_t,\mu_t^X),\ \text{contains all super-linearity}} + b(t, X_t, \mu_t^X) \Big)\mathrm{d}t + \sigma(t, X_t, \mu_t^X)\mathrm{d}W_t \end{aligned}$$

where μ_t^X is the law of X_t, and W is a l-dimensional Brownian motion.

Example (super-linearity in space and measure)

$$\mathrm{d}X_t = \Big(X_t - \frac{1}{4}X_t^3 - \int_{\mathbb{R}} (X_t - y)^3 \mu_t^X(\mathrm{d}y) \Big)\, \mathrm{d}t + X_t\, \mathrm{d}W_t, \quad X_0 \sim \mathcal{N}(0, 1).$$

Assumption 1: Assume that v, b and σ are $1/2$-Hölder continuous in time.

1. b, σ is Lipschitz in space and in law for the Wasserstein distance $W^{(2)}$.
2. For u, there exist $L_u \in \mathbb{R}$, $L_{\hat{u}} > 0$, $L_{\tilde{u}} \geq 0$, $q_1 > 0$ such that for all $x, x' \in \mathbb{R}^d$ and $\forall \mu, \mu' \in \mathcal{P}_2(\mathbb{R}^d)$, it holds that

$$\langle x - x', u(x,\mu) - u(x',\mu) \rangle \leq L_u |x - x'|^2, \qquad \text{(One-sided Lipschitz)},$$
$$|u(x,\mu) - u(x',\mu)| \leq L_{\hat{u}}(1 + |x|^{q_1} + |x'|^{q_1})|x - x'|, \text{(Polynomial growth}$$
$$|u(x,\mu) - u(x,\mu')|^2 \leq L_{\tilde{u}} W^{(2)}(\mu,\mu')^2, \qquad \text{(Lipschitz in measure)}.$$

3. For f, there exist $L_f \in \mathbb{R}$, $L_{\hat{f}} > 0$, $q_2 > 0$ such that for all $x, x' \in \mathbb{R}^d$, it holds that

$$\langle x - x', f(x) - f(x') \rangle \leq L_f |x - x'|^2, \qquad \text{(One-sided Lipschitz)},$$
$$|f(x) - f(x')| \leq L_{\hat{f}}(1 + |x|^{q_2} + |x'|^{q_2})|x - x'|, \qquad \text{(Polynomial growth)},$$
$$f(x) = -f(-x), \qquad \text{(Odd function)}.$$

Motivation

1. Existing model side,
   1. the granular media equation, $f(x) = -sign(x)|x|^2$,
      PDE form: $\partial_t \rho = \nabla \cdot [\ \nabla \rho + \rho \nabla W * \rho]$ with $W = \frac{1}{3}|x^3|$,
      $\rho(t,x)$ is the corresponding density map.
   2. the Double-Well model, $f(x) = -x^3$,
      PDE form: $\partial_t \rho = \nabla \cdot [\ \nabla (\frac{\rho|x|^2}{2}) + \rho \nabla V + \rho \nabla W * \rho]$ with $W = \frac{1}{4}|x|^4$,
      $V = \frac{1}{16}|x|^4 - \frac{1}{2}|x|^2$,
      $\rho(t,x)$ is the corresponding density map.
2. Simulation side, "particle corruptions"
   1. the standard Euler method failed to converge, see later for examples,
   2. other methods for super-linearity in space like Taming and time-adaptive method are unclear how to work on super-linearity in measure.

Approximation of MV-SDE: Particle system

We can approximate MV-SDEs(1) through an interacting particle system:

$$\mathrm{d}X_t^{i,N} = \hat{b}(t, X_t^{i,N}, \mu_t^{X,N})\mathrm{d}t + \sigma(t, X_t^{i,N}, \mu_t^{X,N})\mathrm{d}W_t^i \tag{2}$$

$$\mu_t^{X,N}(\mathrm{d}x) := \frac{1}{N}\sum_{j=1}^{N} \delta_{X_t^{j,N}}(\mathrm{d}x), \qquad X_0^{i,N} = X_0^i \in \mathbb{R}^d$$

where $\delta_{X_t^{j,N}}$ is the Dirac measure at point $X_t^{j,N}$, $\mu_t^{X,N}$ is the corresponding empirical measure, ,and $\{W^i\}_{i\in\{1,\cdots,N\}}$ are independent the Brownian motions.

"Propagation of chaos" guarantee that we can connect the MV-SDEs with the particle system, the bias is related to particle numbers N and dimension of each particle d, see [1, 4] for more details.

Brief introduction of different methods

Recall function $\hat{b} = v + b$, set the terminal time as T, the stepsize of the scheme as h with $Mh = T$, $t_n = nh$,the amount of the particles is N, then, for all $i \in \{1, \ldots, N\}$, $n \in \{0, \ldots, M-1\}$ with $\Delta W_n^i = W_{t_{n+1}}^i - W_{t_n}^i$

1. the Euler method:

$$X_{n+1}^{i,N} = X_n^{i,N} + \hat{b}(t_n, X_n^{i,N}, \mu_n^{X,N})h + \sigma(t_n, X_n^{i,N}, \mu_n^{X,N})\Delta W_n^i$$

2. the Taming method:

$$X_{n+1}^{i,N} = X_n^{i,N} + \frac{\hat{b}(t_n, X_n^{i,N}, \mu_n^{X,N})}{1 + M^{-0.5}|\hat{b}(t_n, X_n^{i,N}, \mu_n^{X,N})|}h + \sigma(t_n, X_n^{i,N}, \mu_n^{X,N})\Delta W_n^i$$

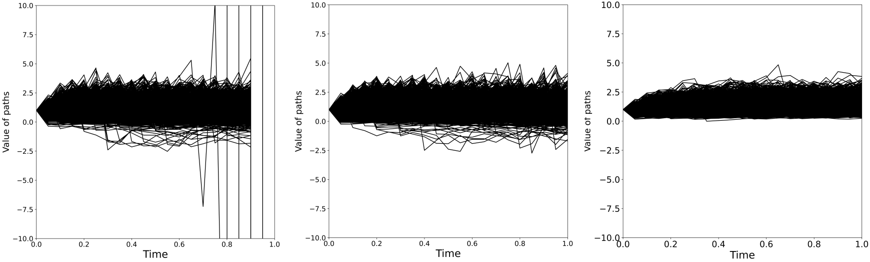
3. the Split-step method (SSM): Introduce later :>

Other methods, time adaptive method, truncated method..

Particle Corruption: super space

Example (The stochastic Ginzburg Landau type process)

$$\mathrm{d}X_t = \left(\frac{1}{2}\mathbb{E}[X_t] + \frac{9}{8}X_t - X_t^3\right)\mathrm{d}t + \frac{3}{2}X_t\mathrm{d}W_t, \quad X_0 = 1.$$



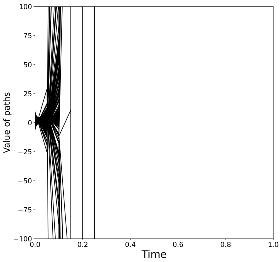
(a) paths of the Euler (b) paths of the Taming (c) paths of the SSM

Figure: Simulation with $N = 5000, \ h = 0.05, \ X_0 = 1$

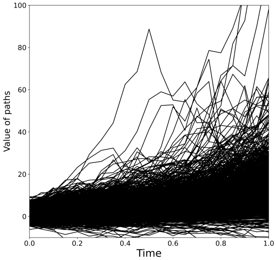
Particle Corruption: super measure

Example (The Double-Well Model)

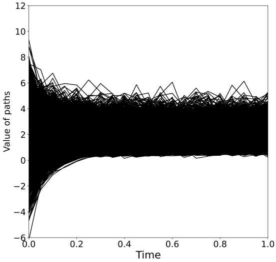
$$\mathrm{d}X_t = \left(X_t - \frac{1}{4}X_t^3 + \int_{\mathbb{R}} -(X_t - y)^3 \mu_t^X(\mathrm{d}y)\right)\mathrm{d}t + X_t \mathrm{d}W_t, \quad X_0 \sim \mathcal{N}(2,4).$$



(a) paths of the Euler



(b) paths of the Taming



(c) paths of the SSM

Figure: Simulation with $N = 5000,\ h = 0.05,\ X_0 \sim \mathcal{N}(2,4)$

Outline

The Split-step method (SSM)

Inspired by [3], the Split-step method for MV-SDEs is define as: the uniform partition as $\pi := \{t_n := nh : n \in \{0, \cdots, M\}, h := T/M\}$ on $[0, T]$ for a prescribed $M \in \mathbb{N}$. Define recursively the split-step method to approximate (2) as follows: for $i \in \{1, \cdots, N\}$ set $\hat{X}_0^{i,N} = X_0^i$, $V = (\dots, v(x_i, \mu^x), \dots)$, then

$$Y_n^{\star,N} = \hat{X}_n^N + hV\left(Y_n^{\star,N}\right), \quad \hat{X}_n^N = (\dots, \hat{X}_n^{i,N}, \dots), \quad Y_n^{\star,N} = (\dots, Y_n^{i,\star,N}, \dots),$$

$$Y_n^{i,\star,N} = \hat{X}_n^{i,N} + hv(Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N}), \qquad \hat{\mu}_n^{Y,N}(dx) := \frac{1}{N}\sum_{j=1}^{N} \delta_{Y_n^{j,\star,N}}(dx),$$

$$\hat{X}_{n+1}^{i,N} = Y_n^{i,\star,N} + b(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})h + \sigma(t_n, Y_n^{i,\star,N}, \hat{\mu}_n^{Y,N})\Delta W_n^i.$$

Convergence results

The main result of the SSM, more details and proof in the paper.

Theorem

Let our assumptions hold and choose h as above. Let $i \in \{1, \cdots, N\}$, take $X^{i,N}$ as the solution (2) and let $\hat{X}^{i,N}$ be the continuous-time extension of the SSM. If $m \geq 4q+4$, where $X_0 \in L_0^m(\mathbb{R}^d)$, q is define in Assumption, then

$$\sup_{i\in\{1,\cdots,N\}} \sup_{0\leq t\leq T} \mathbb{E}\left[|X_t^{i,N} - \hat{X}_t^{i,N}|^2\right] \leq Ch.$$

$$\sup_{i\in\{1,\cdots,N\}} \sup_{0\leq t\leq T} \mathbb{E}\left[|\hat{X}_t^{i,N}|^{2p}\right] \leq C\Big(1 + \mathbb{E}\left[|\hat{X}_0|^{2p}\right]\Big) < \infty, \textit{ With } m \geq 2p.$$

$$\sup_{i\in\{1,\cdots,N\}} \mathbb{E}\left[\sup_{0\leq t\leq T} |X_t^{i,N} - \hat{X}_t^{i,N}|^2\right] \leq Ch^{1-\epsilon}, \textit{ With } m \geq \max\{4q+4, 2+q+q/\epsilon\}.$$

Outline

Example: the double-well

$$dX_t = \Big(X_t - \frac{1}{4}X_t^3 + \int_{\mathbb{R}} -(X_t - y)^3 \mu_t^X(dy)\Big)dt + X_t dW_t,$$

the correspond Fokker-Plank equation is

$\partial_t \rho = \nabla \cdot [\, \nabla(\frac{\rho|x|^2}{2}) + \rho \nabla V + \rho \nabla W * \rho]$ with $W = |x|^4$, $V = \frac{1}{16}|x|^4 - \frac{1}{2}|x|^2$, $\rho(t, x)$ is the corresponding density map. There are three expected stable states $\{-2, 0, 2\}$ of this model.

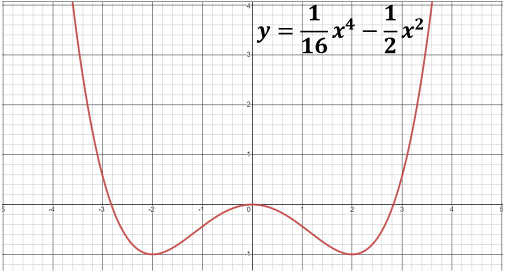
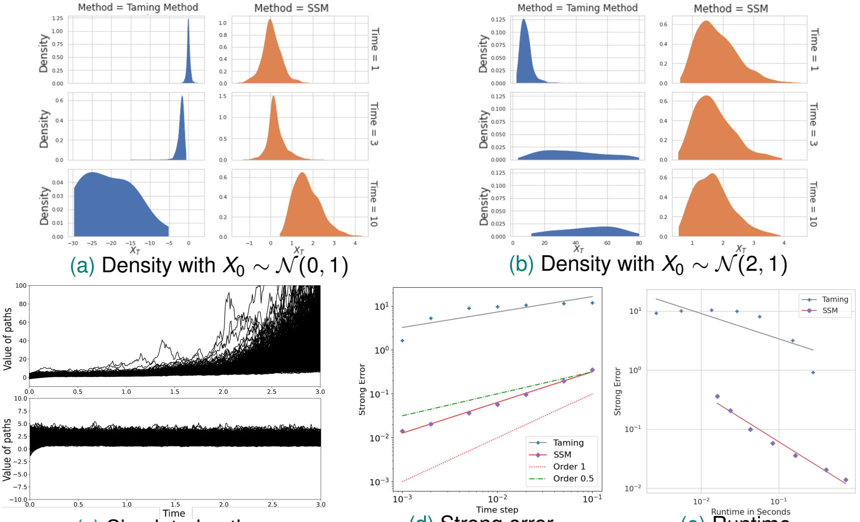


Figure: the Double-Well



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

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

(c) Simulated paths

(d) Strong error

(e) Runtime

Figure: Simulation of the Double-Well model with $N = 1000$ particles.

Initial Distribution	Negative	Zero	Positive
$\mathcal{N}$ (-0.10,1)	100	0	0
$\mathcal{N}$(-0.05,1)	93	0	7
$\mathcal{N}$(-0.02,1)	75	0	25
$\mathcal{N}$(-0.01,1)	57	0	43
$\mathcal{N}$(0.00,1)	46	2	52
$\mathcal{N}$(0.01,1)	44	2	54
$\mathcal{N}$(0.02,1)	24	1	75
$\mathcal{N}$(0.05,1)	5	0	95
$\mathcal{N}$(0.10,1)	0	0	100

Table: Number of cluster states with different initial distribution.

$N = 1000$, $h = 0.01$, $T = 10s$, we test 100 times for each different initial distribution, "Negative" means particles cluster around -2, "Zero" means cluster around 0, "Positive" means cluster around 2.

Example: 2d degenerate Van der Pol (VdP) oscillator

We consider the Van der Pol (VdP) model with added super-linearity in measure and non-constant diffusion. We study the following MV-SDE dynamics, set $x = (x_1, x_2) \in \mathbb{R}^2$, we define the functions f, u, b, σ as

$$f(x) = -x|x|^2, \; u(x) = \begin{bmatrix} -\frac{4}{3}x_1^3 \\ 0 \end{bmatrix}, \; b(x) = \begin{bmatrix} 4(x_1 - x_2) \\ \frac{1}{4}x_1 \end{bmatrix}, \; \sigma(x) = \begin{bmatrix} x_1 & 0 \\ 0 & x_2 \end{bmatrix}.$$

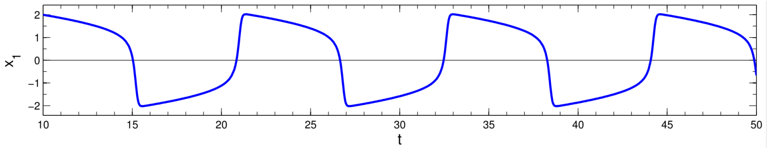
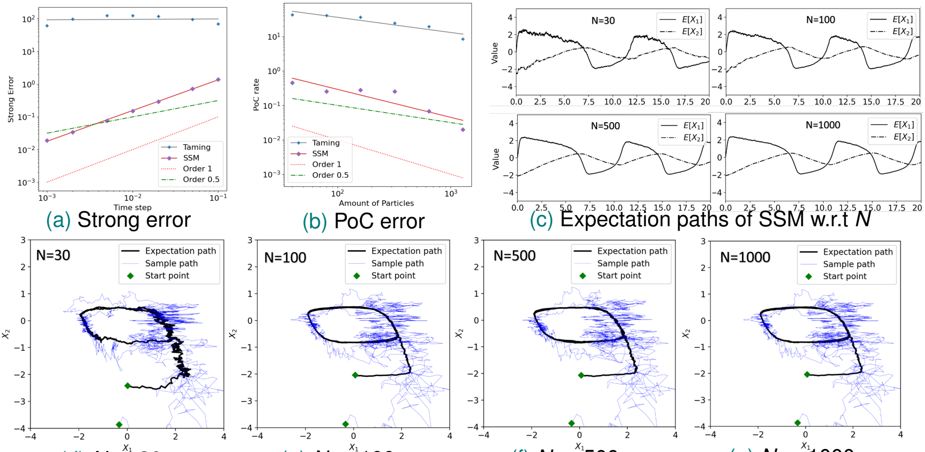


Figure: the Vdp oscillator path of x_1 without additional stochastic terms.



(a) Strong error (b) PoC error (c) Expectation paths of SSM w.r.t N

(d) $N = 30$ (e) $N = 100$ (f) $N = 500$ (g) $N = 1000$

Figure: Simulation of the Vdp model with $X_1 \sim \mathcal{N}(0, 4), X_2 \sim \mathcal{N}(-2, 4)$.

Conclusion

Thank you for listening! :>

References I

G. dos Reis, S. Engelhardt, and G. Smith.
Simulation of mckean-vlasov sdes with super-linear growth.
IMA Journal of Numerical Analysis, 01 2021.
draa099.

W. Fang and M. B. Giles.
Adaptive Euler–Maruyama method for SDEs with nonglobally Lipschitz drift.
Annals of Applied Probability, 30(2):526–560, 2020.

D. J. Higham, X. Mao, and A. M. Stuart.
Strong convergence of Euler-type methods for nonlinear stochastic differential equations.
SIAM Journal on Numerical Analysis, 40(3):1041–1063, 2002.

C. Reisinger and W. Stockinger.
An adaptive Euler-maruyama scheme for McKean SDEs with super-linear growth and application to the mean-field Fitzhugh-Nagumo model.
Journal of Computational and Applied Mathematics, 400:113725, 2022.

A recipe of Newton's method-1

Theorem

Denote $\hat{\Psi}_i : \mathbb{R}^d \times \mathbb{R}^{Nd} \times [0, T] \to \mathbb{R}^d$ *be a mapping for the first step of the SSM as*

$$\hat{\Psi}_i(\hat{X}_n^{i,N}, \hat{X}_n^N, h) = Y_n^{i,\star,N}, \quad \hat{\Psi} = (\hat{\Psi}_1, \dots, \hat{\Psi}_N). \tag{3}$$

if $\sup_i \mathbb{E}[|\hat{\Psi}_i(x_i, x, h) - \overline{\Psi}_i(x_i, x, h)|^2] \leq Ch$ *for all* $x = (x_1, \dots, x_N) \in L_0^2(\mathbb{R}^{Nd})$ *and some constant* C *independent of* N*, then*

$$\sup_{n \in \{0,\dots,M\}} \sup_{i \in \{1,\dots,N\}} \mathbb{E}\big[|\hat{X}_n^{i,N} - \bar{X}_n^{i,N}|^2\big] \leq Ch \tag{4}$$

A recipe of Newton's method-2

$$F(y) = y - x - hV(y) = 0,\ V_i(y) = u(y_i) + \frac{1}{N}\sum_{j=1}^{N} f(y_i - y_j),\ V = (V_1, .., V_N)$$

$$[\nabla F](y) = I_{Nd} - hA(y) + \frac{h}{N}\Gamma(y),$$

$$A(y) = \begin{bmatrix} \nabla u(y_1) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \nabla u(y_N) \end{bmatrix} + \begin{bmatrix} \frac{1}{N}\sum_{j=1}^{N} \nabla f(y_1 - y_j) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{N}\sum_{j=1}^{N} \nabla f \end{bmatrix}$$

$$\Gamma(y) = \begin{bmatrix} \nabla f(y_1 - y_1) & \cdots & \nabla f(y_1 - y_n) \\ \vdots & \ddots & \vdots \\ \nabla f(y_n - y_1) & \cdots & \nabla f(y_n - y_n) \end{bmatrix},$$

We stop the Newton's iteration at step κ when the error tolerance rule $\|y^\kappa - y^{\kappa-1}\|_\infty < \sqrt{h}$ is satisfied.

Discussion on the complexity

By using the Newton's method described above, for each timestep, the computation cost for the interacting particle system is of order $\mathcal{O}(N^2\kappa)$, where κ is 2 to 4 on average in practice, and there are M steps ($Mh = T$) in total, thus the whole computation cost is of order $\mathcal{O}(N^2M\kappa)$, which is similar to the explicit method like Taming $\mathcal{O}(N^2M)$.
Also, the random batch method is an explicit type Euler-based method and works for mild assumptions and can reduce the computation cost for each timestep from $\mathcal{O}(N^2)$ to $\mathcal{O}(N)$. But for our settings, the explicit type Euler method failed to work, thus there is a question on how to apply the SSM-based random batch method to this type of system. This is out of the scope of this paper and we leave it for future research.

Propogation of Chaos

Let the assumptions hold for some $m > 2(q+1)$. Let X^i be the solution to the orgin MV-SDEs. Then, there exists a unique solution $X^{i,N}$ of the particle system and for any $1 \le p \le m$ there exists $C > 0$ independent of N such that

$$\sup_{t\in[0,T]} \sup_{i\in\{1,\ldots,N\}} \mathbb{E}\big[|X_t^{i,N}|^p\big] \le C(1+\mathbb{E}\big[\,|X_0^{\cdot}|^p\big]). \tag{5}$$

Moreover, suppose that $m > \max\{2(q+1), 4\}$, then there exist constant $C(T)$ depends on T

$$\sup_{i\in\{1,\ldots,N\}} \mathbb{E}\big[\sup_{0\le t\le T} |X_t^i - X_t^{i,N}|^2\big] \le C(T) \begin{cases} N^{-1/2}, & d<4 \\ N^{-1/2}\log N, & d=4 \\ N^{\frac{-2}{d+4}}, & d>4 \end{cases}. \tag{6}$$

Short Conclusions

We have proposed a Split-step method for simulation of MV-SDEs which can be applied to super-linearity in space and measure.

- The SSM deals with the super-linear growth component separately by using implicit method. Its structure enable us to compute the implicit part flexibly.
- For super-space, compare to the Taming method and the the Adaptive method, the strong error for SSM is of same order but smaller. There are more properties like stability and efficiency advantage for SSM.
- For super-measure, SSM has already been proofed to work. Adaptive and taming methods are still unclear how to work. And we see that SSM do a good job in the examples (error rate, distribution).
- Although we need to implicitly solve the first step in the SSM, we can always compute them in parallel, and thus save a lot of time.

Example: the granular media equation

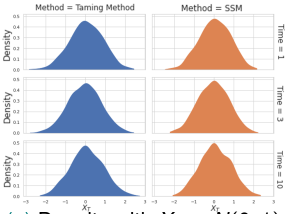
The first example is the granular media equation taking the form $\partial_t \rho = \nabla \cdot [\nabla \rho + \rho \nabla W * \rho]$ with $W(x) = |x|^3$ (symmetric double well potential) and $\rho(t, x)$ is the correspondent probability measure. Cast in SDE form we have

$$dX_t = v(X_t, \mu_t^X)dt + \sqrt{2}\, dW_t, \quad X_0 \in L_0^m(\mathbb{R}^d) \tag{7}$$

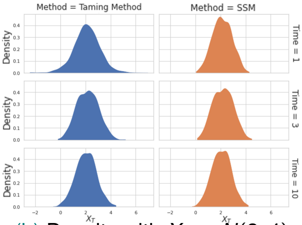
$$v(x, \mu) = \int_{\mathbb{R}^d} \Big(- sign(x - y)|x - y|^2 \Big) \mu(dy). \tag{8}$$

where $sign(\cdot)$ is the standard sign function, μ_t^X is the law of the solution process X at time t.

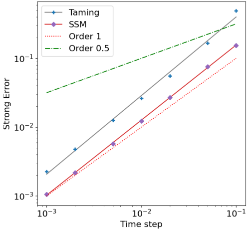
Example: the granular media equation



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



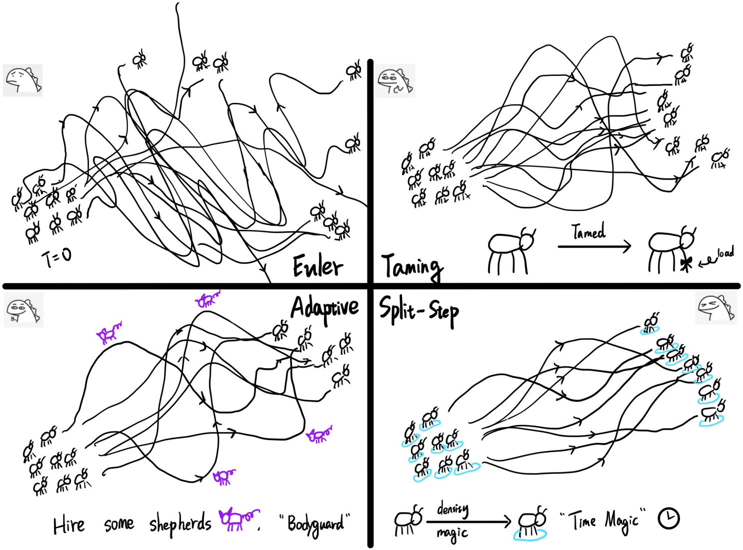
(b) Density with $X_0 \sim N(2, 4)$



(c) Strong error

Figure: Simulations of the granular media equation with $N = 1000$ particles. (a) and (b) show the density map for Taming (blue) and SSM (orange) with $h = 0.01$ at times $T = 1, 3, 10$ seen top to bottom with different initial distribution. (c) Strong error of different methods with $X_0 \sim \mathcal{N}(2, 4)$.

Not so precise intuitions for all four simulation methods



Three proved to work for super-space methods: Taming, Adaptive, Split-step.

Brief introduction of the Taming method

We denote $\widehat{b} = v + b$ in MV-SDEs(1), $n \in \{1, \cdots, N\}$, $Mh = T$

The taming method [1] approximates MV-SDEs as follows (see also [4, Section 4]):

$$\bar{X}_{n+1}^{i,N,M} = \bar{X}_n^{i,N,M} + \underbrace{\frac{\widehat{b}\Big(t_n, \bar{X}_n^{i,N,M}, \bar{\mu}_n^{X,N}\Big)}{1 + M^{-\alpha}\Big|\widehat{b}\Big(t_n, \bar{X}_n^{i,N,M}, \bar{\mu}_n^{X,N}\Big)\Big|}}_{|\cdot| \text{ is bounded}} h + \sigma\Big(t_n, \bar{X}_n^{i,N,M}, \bar{\mu}_n^{X,N}\Big)\Delta W_n^i, \tag{9}$$

where $\bar{\mu}_n^{X,N}(\mathrm{d}x) = \frac{1}{N}\sum_{j=1}^N \delta_{\bar{X}_n^{j,N,M}}(\mathrm{d}x)$, $\Delta W_n^i = W_{t_{n+1}}^i - W_{t_n}^i$ with $\bar{X}_0^{i,N,M} = X_0^i$. The parameter $\alpha \in (0, 1]$ where setting $\alpha = 1/2$ delivers a rMSE convergence rate of order $1/2$ while setting $\alpha = 1$ delivers a rMSE convergence rate of order 1 (for a constant diffusion σ).

Brief introduction of the Adaptive method

The adaptive method from [4] controls the timestep to have acceptable result, it approximates MV-SDEs as follows for $t_n \in [k_n h, (k_n+1)h), k_n \in \mathbb{N}$ and

$$\bar{X}^{i,N}_{t_{n+1}} = \bar{X}^{i,N}_{t_n} + \widehat{b}\Big(t_n, \bar{X}^{i,N}_{t_n}, \bar{\mu}^{X,N}_{k_n h}\Big) h^i_n + \sigma\Big(t_n, \bar{X}^{i,N}_{t_n}, \bar{\mu}^{X,N}_{k_n h}\Big) \Delta W^i_{t_n}, \tag{10}$$

where $\bar{\mu}^{X,N}_{k_n h}(\mathrm{d}x) = \frac{1}{N}\sum_{j=1}^{N} \delta_{\bar{X}^{j,N,M}_{k_n h}}(\mathrm{d}x)$, $t_{n+1} = t_n + h^i_n$, $\Delta W^i_{t_n} = W^i_{t_{n+1}} - W^i_{t_n}$ with $\bar{X}^{i,N,M}_0 = X^i_0$ and for a map $\mathbf{h}^\delta(x) : \mathbb{R}^d \to [0,h]$

$$h^i_n = \min\{\mathbf{h}^\delta(\bar{X}^{i,N}_{t_n}), (k_n+1)h - t_n\}.$$

The function $\mathbf{h}^\delta$ is specified at each example and is to be understood as similar technique to the taming method. There the drift $\widehat{b}$ is modified to control the growth across the application of the scheme, for the adaptive scheme, one modifies instead the time-step h in a dynamic fashion to control the growth of $\widehat{b}$, see [2] & [4].