Numerical approximation of McKean-Vlasov SDEs via stochastic gradient descent

Andrea Amato[1]
(joint work with Ankush Agarwal, Stefano Pagliarani and Gonçalo dos Reis)

[1]Dipartimento di Matematica, University of Bologna, Italy

INdAM Meeting: Workshop on Irregular Stochastic Analysis, Cortona - June 24th, 2025

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA | DEPARTMENT OF MATHEMATICS

McKean-Vlasov SDEs

$$\begin{cases} dX_t = b(t, X_t, \mu_t)dt + \sigma dW_t, & t > 0, \\ X_0 = \xi \end{cases}$$

- Drift depends on X_t and on μ_t (law of X_t);
- μ_t *mean-field* limit, as $N \to +\infty$, of the law of each

$$\begin{cases} dX_t^{(i,N)} = b(t, X_t^{(i,N)}, \mu^{emp}_{X_t^N})dt + \sigma dW_t^i, \\ X_0^{(i,N)} \sim \xi \end{cases} \qquad 1 \le i \le N$$

Propagation of chaos [Sznitman, 1991, Méléard, 1996].
Interacting particle system (IPS)
[Antonelli and Kohatsu-Higa, 2002].

MKV SDEs with separable coefficients

Projection method [Belomestny and Schoenmakers, 2018]

$$\begin{cases} dX_t = \bar{\gamma}(t)\big(\alpha(t,X_t)dt + \beta(t,X_t)dW_t\big), & X_0 = \xi, \\ \bar{\gamma}(t) = \mathbb{E}[\varphi(X_t)], \end{cases} \tag{1}$$

with $T > 0$ fixed time horizon and $\alpha : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{K\times d}$, $\beta : [0,T] \times \mathbb{R}^d \to \mathbb{R}^{K\times d\times q}$, $\varphi : \mathbb{R}^d \to \mathbb{R}^K$ measurable maps;

MKV SDEs with separable coefficients

Projection method [Belomestny and Schoenmakers, 2018]

$$\begin{cases} dX_t = \bar{\gamma}(t)\big(\alpha(t, X_t)dt + \beta(t, X_t)dW_t\big), \qquad X_0 = \xi, \\ \bar{\gamma}(t) = \mathbb{E}[\varphi(X_t)], \end{cases} \tag{1}$$

with $T > 0$ fixed time horizon and $\alpha : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{K\times d}$, $\beta : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{K\times d\times q}$, $\varphi : \mathbb{R}^d \to \mathbb{R}^K$ measurable maps;

Assumption (1)

- $\alpha, \beta \in C_b([0, T] \times \mathbb{R}^d)$ *Lipschitz in space, uniformly in time;*
- φ *Lipschitz and bounded;*
- $\xi \in L^2$.

MKV SDEs with separable coefficients

Projection method [Belomestny and Schoenmakers, 2018]

$$\begin{cases} dX_t = \bar{\gamma}(t)\big(\alpha(t, X_t)dt + \beta(t, X_t)dW_t\big), \qquad X_0 = \xi, \\ \bar{\gamma}(t) = \mathbb{E}[\varphi(X_t)], \end{cases} \tag{1}$$

with $T > 0$ fixed time horizon and $\alpha : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{K\times d}$, $\beta : [0, T] \times \mathbb{R}^d \to \mathbb{R}^{K\times d\times q}$, $\varphi : \mathbb{R}^d \to \mathbb{R}^K$ measurable maps;

Assumption (1)

- $\alpha, \beta \in C_b([0, T] \times \mathbb{R}^d)$ *Lipschitz in space, uniformly in time;*
- φ *Lipschitz and bounded;*
- $\xi \in L^2$.

Assumption (2)

$\alpha, \beta \in C_b^{1,2}([0, T] \times \mathbb{R}^d)$ *and* $\varphi \in C_b^2(\mathbb{R}^d)$.

From a fixed-point equation to a minimization problem

Equivalent fixed-point equation: find $\bar{\gamma} \in C([0,T], \mathbb{R}^K)$ that solves

$$\gamma = \mathbb{E}[\varphi(Z^{\gamma}_{\cdot})], \tag{2}$$

with

$$dZ^{\gamma}_t = \gamma(t)\big(\alpha(t, Z^{\gamma}{}_t)dt + \beta(t, Z^{\gamma}{}_t)dW_t\big), \quad Z^{\gamma}_0 = \xi.$$

From a fixed-point equation to a minimization problem

Equivalent fixed-point equation: find $\bar{\gamma} \in C([0,T], \mathbb{R}^K)$ that solves

$$\gamma = \mathbb{E}[\varphi(Z_\cdot^{\gamma})], \tag{2}$$

with

$$dZ_t^{\gamma} = \gamma(t)\big(\alpha(t, Z^{\gamma}{}_t)dt + \beta(t, Z^{\gamma}{}_t)dW_t\big), \quad Z_0^{\gamma} = \xi.$$

Minimization problem:

$$\bar{\gamma} = \mathbb{E}[\varphi(Z_\cdot^{\bar{\gamma}})] \Longleftrightarrow \min_{\gamma \in C([0,T])} F^2(\gamma) = F^2(\bar{\gamma}) = 0,$$

with

$$F : C\big([0,T], \mathbb{R}^K\big) \to \mathbb{R}, \quad F(\gamma) := \|\mathbb{E}[\varphi(Z_\cdot^{\gamma})] - \gamma\|_{L^2([0,T])}.$$

Projection on a finite-dimensional domain

Fix $g_0, \cdots, g_n \in C([0,T],\mathbb{R})$ linearly independent and define the finite-dimensional sub-space of $C([0,T],\mathbb{R})$

$$S_n := \{a_0 g_0 + \cdots + a_n g_n : a = (a_0, \cdots, a_n) \in \mathbb{R}^{n+1}\}.$$

Projection on a finite-dimensional domain

Fix $g_0, \cdots, g_n \in C([0,T], \mathbb{R})$ linearly independent and define the finite-dimensional sub-space of $C([0,T], \mathbb{R})$

$$S_n := \{a_0 g_0 + \cdots + a_n g_n : a = (a_0, \cdots, a_n) \in \mathbb{R}^{n+1}\}.$$

New minimization problem:

$$\min_{\gamma \in C([0,T])} F^2(\gamma) \approx \min_{p \in S_n} F^2(p) \Longleftrightarrow \min_{a \in \mathbb{R}^{n+1}} G(a),$$

with

$$G(a) := F^2(\mathcal{L}a) = \int_0^T \left| \mathbb{E}\left[\varphi({Z^a}_t) - (\mathcal{L}a)(t)\right]\right|^2 dt,$$

$\mathcal{L}$ is a linear lifting operator and ${Z^a}_t$ denote the solution to

$$dZ_t^a = (\mathcal{L}a)_t \big(\alpha(t, {Z^a}_t)dt + \beta(t, {Z^a}_t)dW_t\big), \qquad Z_0^a = \xi. \tag{3}$$

Heuristics

$$\partial_{a_h} G(a) = 2\int_0^T \underbrace{\left\langle \mathbb{E}\left[\varphi(Z_t^a) - (\mathcal{L}a)_t\right], \mathbb{E}[\partial_{a_h}(\varphi(Z_t^a)) - g_h(t)]\right\rangle}_{\mathbb{E}\left[\left\langle \varphi(Z_t^a(W)) - (\mathcal{L}a)_t, \partial_{a_h}(\varphi(Z_t^a(\widetilde{W}))) - g_h(t)\right\rangle\right]} dt$$

with W, $\widetilde{W}$ independent samples. Formally:

$$\nabla_a G(a) = \mathbb{E}[v(a; W, \widetilde{W})],$$

with $v(a; W, \widetilde{W}) \in \mathbb{R}^{n+1}$, whose components are given by

$$v_{h,j}(a; W, \widetilde{W}) := 2\int_0^T \left\langle \varphi\big(Z_t^a(W)\big) - (\mathcal{L}a)_t, \nabla_x \varphi\big(Z_t^a(\widetilde{W})\big) Y_t^{a;h,j}(\widetilde{W}) - g_h(t) \right\rangle dt,$$

where $Y^a := \nabla_a Z^a$ by formally differentiating (3) w.r.t. $a_{h,j}$.

Algorithm

$$\mathbf{a}_{m+1} := \mathbf{a}_m - \eta_m \mathbf{v}_{m+1}, \quad \mathbf{v}_{m+1} = v(\mathbf{a}_m; W_{m+1}, \widetilde{W}_{m+1}), \quad m \in \mathbb{N}_0.$$

Algorithm

$$\mathbf{a}_{m+1} := \mathbf{a}_m - \eta_m \mathbf{v}_{m+1}, \quad \mathbf{v}_{m+1} = v(\mathbf{a}_m; W_{m+1}, \widetilde{W}_{m+1}), \quad m \in \mathbb{N}_0.$$

- $(\eta_m)_{m\in\mathbb{N}}$ scalars satisfying (Robbins-Monro)

$$\sum_{m=0}^{\infty} \eta_m = +\infty \quad \text{and} \quad \sum_{m=0}^{\infty} \eta_m^2 < +\infty.$$

- $(\bar{\Omega}, \bar{\mathcal{F}}, (\bar{\mathcal{F}}_m)_{m\in\mathbb{N}}, \bar{\mathbb{P}})$ supporting independent
  - Brownian samples $(W_m)_{m\in\mathbb{N}}$ and $(\widetilde{W}_m)_{m\in\mathbb{N}}$
  - ξ-samples $(\xi_m)_{m\in\mathbb{N}}$, $(\widetilde{\xi}_m)_{m\in\mathbb{N}}$

Algorithm

$$\mathbf{a}_{m+1} := \mathbf{a}_m - \eta_m \mathbf{v}_{m+1}, \quad \mathbf{v}_{m+1} = v(\mathbf{a}_m; W_{m+1}, \widetilde{W}_{m+1}), \quad m \in \mathbb{N}_0.$$

- $(\eta_m)_{m\in\mathbb{N}}$ scalars satisfying (Robbins-Monro)

$$\sum_{m=0}^{\infty} \eta_m = +\infty \quad \text{and} \quad \sum_{m=0}^{\infty} \eta_m^2 < +\infty.$$

- $(\bar{\Omega}, \bar{\mathcal{F}}, (\bar{\mathcal{F}}_m)_{m\in\mathbb{N}}, \bar{\mathbb{P}})$ supporting independent
  - Brownian samples $(W_m)_{m\in\mathbb{N}}$ and $(\widetilde{W}_m)_{m\in\mathbb{N}}$
  - ξ-samples $(\xi_m)_{m\in\mathbb{N}}$, $(\widetilde{\xi}_m)_{m\in\mathbb{N}}$

Theorem (Convergence to a stationary point)

Under Assumptions 1 and 2 the map G is differentiable and there exists an $\mathbb{R}^{(n+1)K}$-valued r.v. $\mathbf{a}_\infty$ s.t.

$$\lim_{m\to+\infty} \mathbf{a}_m = \mathbf{a}_\infty \quad \bar{\mathbb{P}}\text{-almost surely} \quad \text{and} \quad \nabla_a G(\mathbf{a}_\infty) = 0.$$

Numerical test: Polynomial drift model

$$dX_t = \big(\mathbb{E}[X_t] - X_t\mathbb{E}[X_t^2] + \delta X_t\big)\, dt + X_t dW_t, \quad X_0 = x_0.$$

The above is in the form of (1) with $K = 3$, $d = q = 1$ and

$$\varphi(x) = (x, x^2, 1),$$
$$\alpha(t,x) = (1, -x, \delta x)^\top,$$
$$\beta(t,x) = (0, 0, x)^\top.$$

Numerical test: Polynomial drift model

$$dX_t = \big(\mathbb{E}[X_t] - X_t\mathbb{E}[X_t^2] + \delta X_t\big)\,dt + X_t dW_t, \quad X_0 = x_0.$$

The above is in the form of (1) with $K = 3$, $d = q = 1$ and

$$\varphi(x) = (x, x^2, 1),$$
$$\alpha(t, x) = (1, -x, \delta x)^\top,$$
$$\beta(t, x) = (0, 0, x)^\top.$$

Benchmark:

- Monte Carlo $\bar{\gamma}_\cdot^{\mathrm{MC}} = \frac{1}{N}\sum_{i=1}^{N} \varphi(X_\cdot^{N,i})$ with 10^6 particles;

Numerical test: Polynomial drift model

$$dX_t = \left(\mathbb{E}[X_t] - X_t\mathbb{E}[X_t^2] + \delta X_t\right) dt + X_t dW_t, \quad X_0 = x_0.$$

The above is in the form of (1) with $K = 3$, $d = q = 1$ and

$$\begin{aligned}\varphi(x) &= (x, x^2, 1),\\ \alpha(t,x) &= (1, -x, \delta x)^\top,\\ \beta(t,x) &= (0,0,x)^\top.\end{aligned}$$

<u>Benchmark:</u>

- Monte Carlo $\bar{\gamma}^{\mathrm{MC}}_{\cdot} = \frac{1}{N}\sum_{i=1}^N \varphi(X^{N,i}_{\cdot})$ with 10^6 particles;

<u>SGD:</u>

- Basis of S_n: Lagrange polynomials centered at the Chebyshev nodes;
- Learning rates: $\eta_m = \frac{r_0}{(m+1)^\rho}, \quad r_0 > 0, \quad \rho \in (0.5, 1]$;
- Relative error after m iterations: $\varepsilon_m := \frac{\|\mathcal{L}\mathbf{a}_m - \bar{\gamma}^{\mathrm{MC}}\|_{L^2}}{\|\bar{\gamma}^{\mathrm{MC}}\|_{L^2}} < 1\%$.

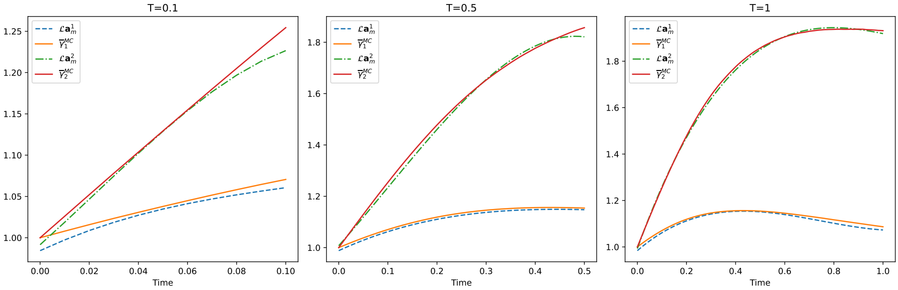
Polynomial drift model: number of iterations and execution time

Numerical study details:

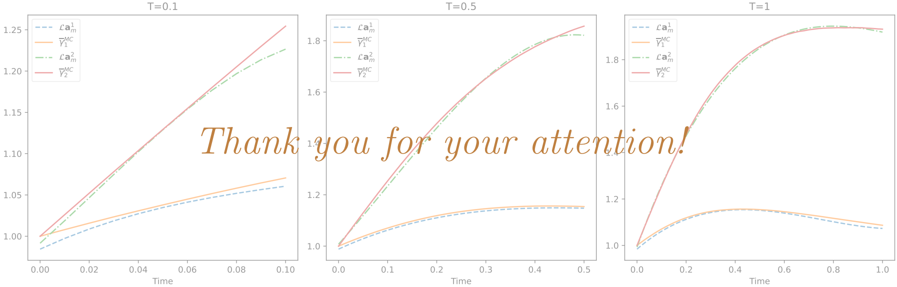
- 1000 independent runs of the algorithm;
- $n = 3$, timestep size $h = 10^{-2}$, $x_0 = 1$ and $\delta = 0.8$;
- Monte Carlo benchmark execution time: 2.57 seconds for $T = 0.1$, 9.78 seconds for $T = 0.5$, and 19.53 seconds for $T = 1$.

	$M = 1$	$M = 10$	$M = 100$	$M = 1000$
$T = 0.1$	258 (0.06)	52.3 (0.02)	31.5 (0.01)	26.9 (0.03)
$T = 0.5$	3456.7 (3.52)	590.3 (0.72)	131.2 (0.20)	79.7 (0.39)
$T = 1.0$	4995.9 (9.91)	3582 (8.54)	1003.9 (3.07)	776.8 (8.25)

Polynomial drift model: plots with $M = 1000$



Polynomial drift model: plots with $M = 1000$



Thank you for your attention!

Bibliography I

Agarwal, A., Amato, A., Reis, G. d., and Pagliarani, S. (2023).
Numerical approximation of mckean-vlasov sdes via stochastic gradient descent.

arXiv preprint arXiv:2310.13579.

Antonelli, F. and Kohatsu-Higa, A. (2002).
Rate of convergence of a particle method to the solution of the mckean–vlasov equation.
The Annals of Applied Probability, 12(2):423–476.

Belomestny, D. and Schoenmakers, J. (2018).
Projected particle methods for solving mckean–vlasov stochastic differential equations.
SIAM Journal on Numerical Analysis, 56(6):3169–3195.

Bertsekas, D. P. and Tsitsiklis, J. N. (2000).
Gradient convergence in gradient methods with errors.
SIAM Journal on Optimization, 10(3):627–642.

Bibliography II

Méléard, S. (1996).
Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models, pages 42–95.
Springer Berlin Heidelberg, Berlin, Heidelberg.

Patel, V. and Zhang, S. (2021).
Stochastic gradient descent on nonconvex functions with general noise models.
arXiv preprint arXiv:2104.00423.

Sznitman, A.-S. (1991).
Topics in propagation of chaos.
Ecole d'été de probabilités de Saint-Flour XIX—1989, 1464:165–251.