A posteriori error estimates for fully coupled McKean–Vlasov forward-backward SDEs

Christoph Reisinger

with Wolfgang Stockinger and Yufei Zhang

University of Oxford
Mathematical Institute

Online Seminar on SDEs

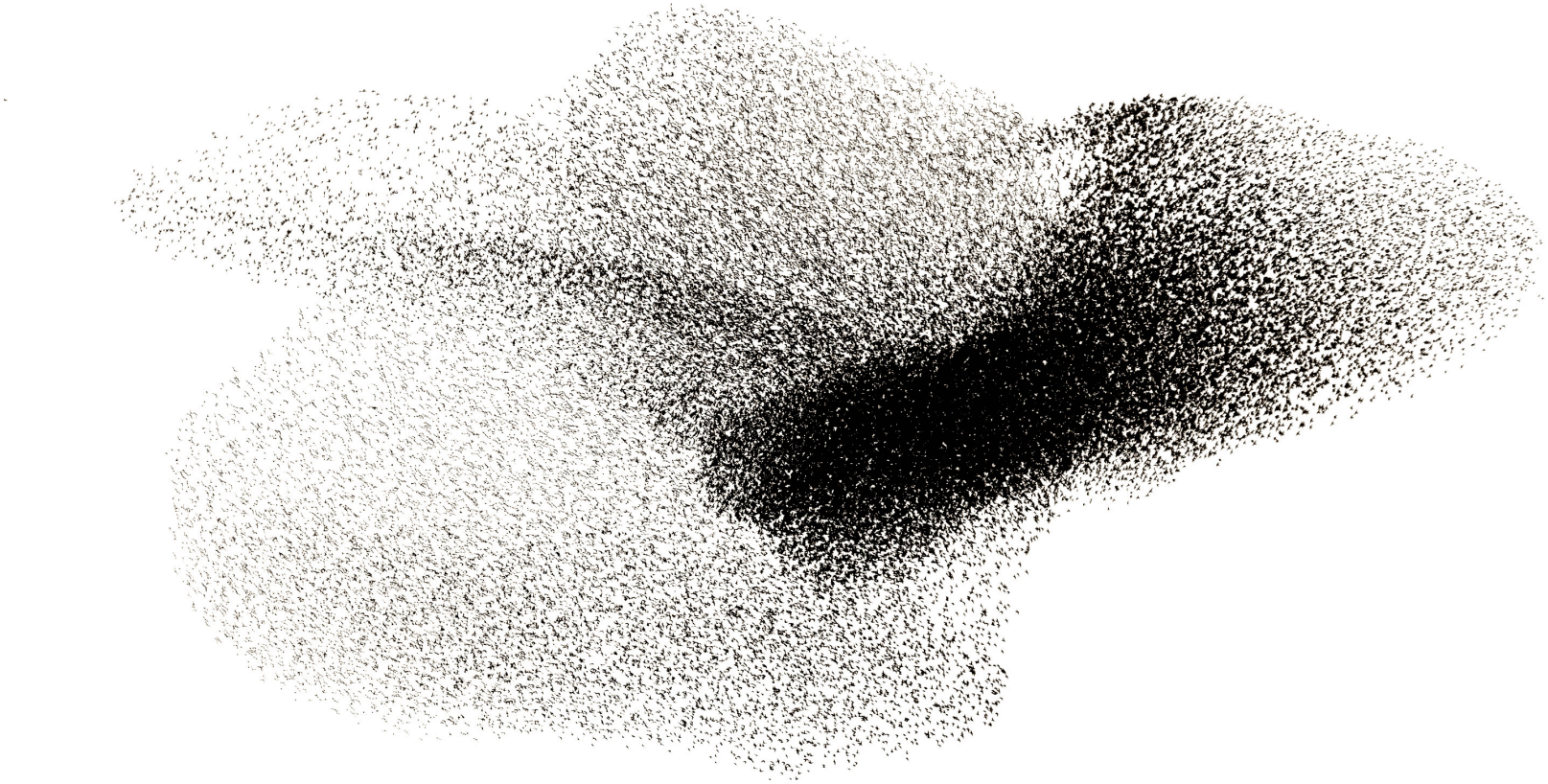
19 May 2023

Oxford Mathematics

Outline

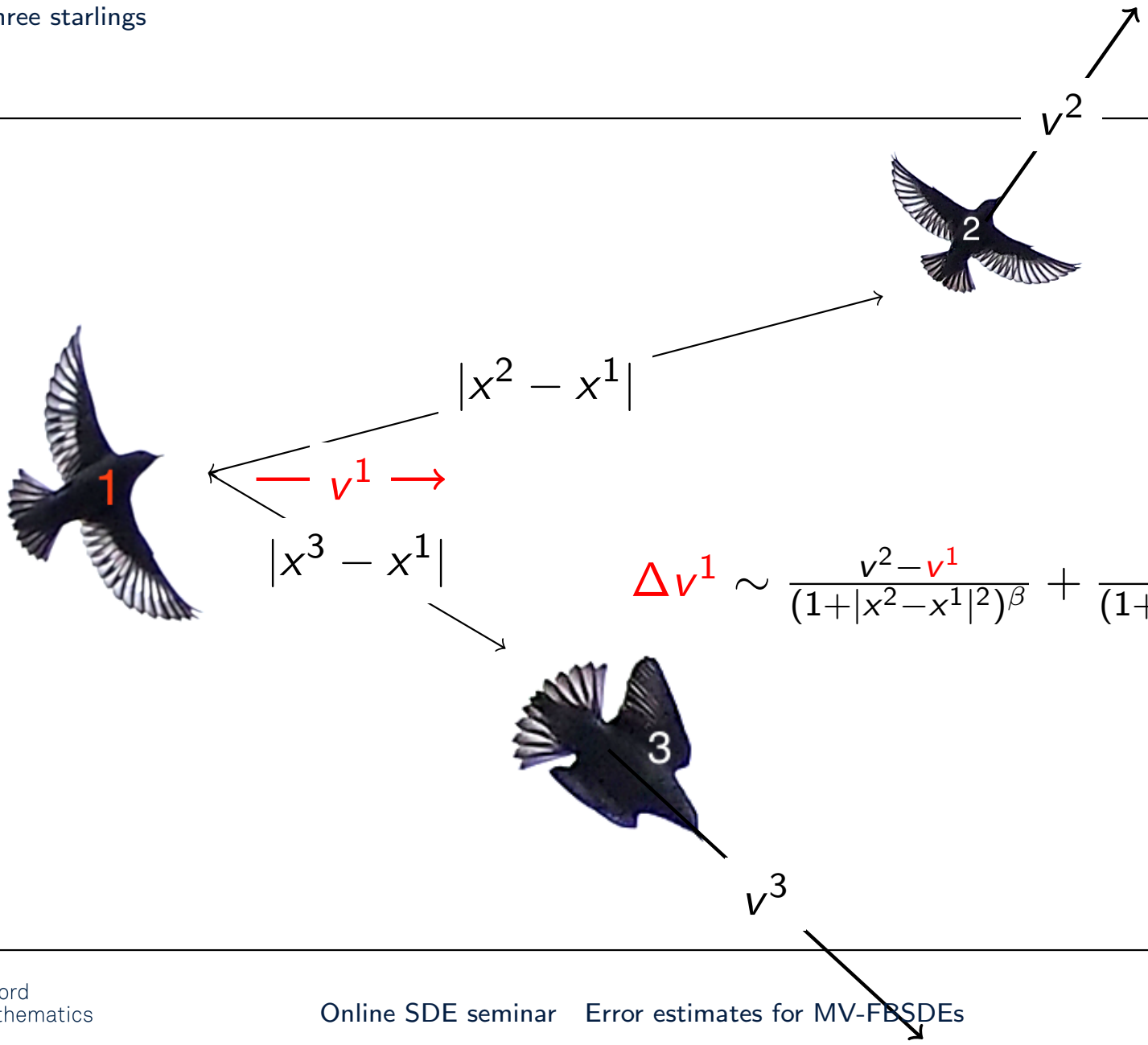
- Mean-field models and mean-field control
- Continuous and discrete MV-FBSDEs – (numerical) challenges
- A posteriori error estimates – reliability and efficiency
- Numerical demonstrations

Interacting particles



Motivating example: Cucker–Smale model

Three starlings



$$\Delta v^1 \sim \frac{v^2 - v^1}{(1+|x^2-x^1|^2)^\beta} + \frac{v^3 - v^1}{(1+|x^3-x^1|^2)^\beta}$$

Motivating example: Cucker–Smale model

Interacting kernel and noise

UNIVERSITY OF OXFORD
Mathematical Institute

More precisely:

$$v^1_{t+\Delta t} - v^1_t = K\left(\frac{v^2_t - v^1_t}{(1+|x^2_t - x^1_t|^2)^\beta} + \frac{v^3_t - v^1_t}{(1+|x^3_t - x^1_t|^2)^\beta}\right)\Delta t$$
$$= \left(\kappa(x^2_t - x^1_t, v^2_t - v^1_t) + \kappa(x^3_t - x^1_t, v^3_t - v^1_t)\right)\Delta t,$$

- say $t \in \{t_0, t_1, t_2, \ldots, t_{n-1}\}$, $t_k = k\Delta t$.
- Add some noise, $Z^1_k \sim N(0,1)$ at t_k:

$$v^1_{t_{k+1}} - v^1_{t_k} = \sum_{j\in\{2,3\}} \kappa(x^j_t - x^1_t, v^j_t - v^1_t)\Delta t + Z^1_k\sqrt{\Delta t}$$

Motivating example: Cucker–Smale model

Finitely many birds

Mathematical Institute

Now let there be N birds, all like bird 1:

$$v^i_{t_{k+1}} - v^i_{t_k} = \sum_{j \neq i} \kappa(x^j_t - x^i_t, v^j_t - v^i_t)\Delta t + Z^i_k \sqrt{\Delta t}$$

Two limits:

- $\Delta t \to 0$: $\quad dx^i_t = v^i_t \, dt$ and

$$dv^i_t = \sum_{j \neq i} \kappa(x^j_t - x^i_t, v^j_t - v^i_t) \, dt + dW_t$$

- $N \to \infty$: $\quad dx_t = v_t \, dt$ and

$$\mathrm{d}v_t = \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} \kappa(x' - x_t, v' - v_t) \, \mathbb{P}_{(x_t, v_t)}(\mathrm{d}x', \mathrm{d}v') \right) \mathrm{d}t + \mathrm{d}W_t$$

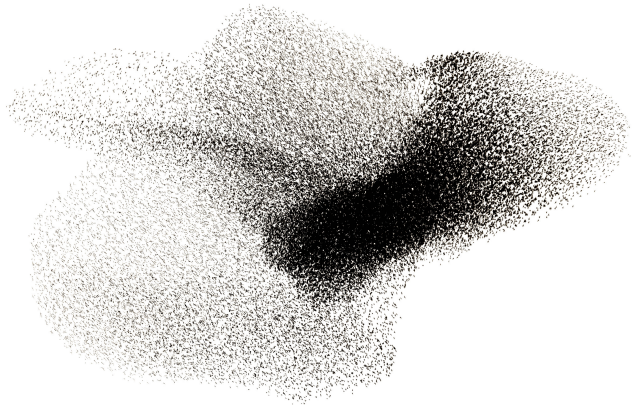
Motivating example: Cucker–Smale model

Infinitely many birds

Mathematical Institute

Now let there be N birds:

$$v^i_{t_{k+1}} - v^i_{t_k} = \sum_{j \neq i} \kappa(x^j_t - x^i_t, v^j_t - v^i_t)\Delta t + Z^i_k \sqrt{\Delta t}$$



Two limits:

- $\Delta t \to 0$: $\quad dx^i_t = v^i_t \, dt$ and

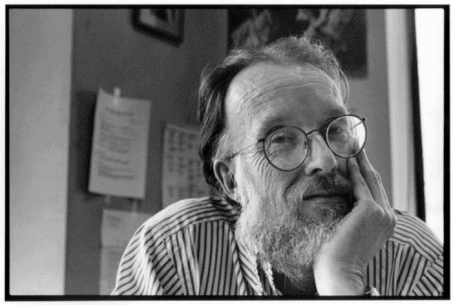
$$dv^i_t = \sum_{j \neq i} \kappa(x^j_t - x^i_t, v^j_t - v^i_t) \, dt + dW_t$$

- $N \to \infty$: $\quad dx_t = v_t \, dt$ and

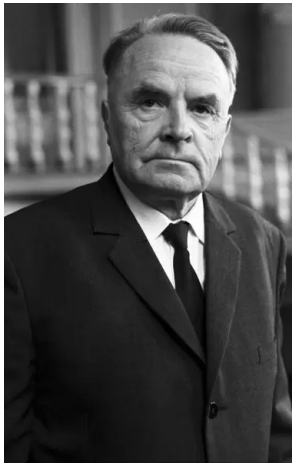
$$\mathrm{d}v_t = \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} \kappa(x' - x_t, v' - v_t) \, \mathbb{P}_{(x_t, v_t)}(\mathrm{d}x', \mathrm{d}v') \right) \mathrm{d}t + \mathrm{d}W_t$$

McKean–Vlasov equations

Classical setting



Henry McKean



Anatoly Vlasov

$$dX_t = b(t, X_t, \nu_t)\,\mathrm{d}t + \sigma(t, X_t, \nu_t)\,\mathrm{d}W_t$$
$$\nu_t = \mathrm{Law}(X_t)$$

- μ, σ Hölder $1/2$ in t,
- Lipschitz in x and ν;
- $b(0,0,\nu)$, $\sigma(0,0,\nu)$ bounded
- $X_0 \sim \nu_0 \in \mathcal{P}_2$ (bounded 2nd moments)

A.-S. Sznitman, Ecole d'Eté de Probabilités de Saint-Flour XIX, 1989
S. Méléard, Probab mod nonlin partial differential equations, 1996

Propagation of chaos

Convergence in N

Consider further, for some $N \geq 1$, the *empirical measure*

$$\nu_t^N = \frac{1}{N} \sum_{i=1}^{N} \delta_{X_t^i};$$

the interacting particle system

$$dX_t^{i,N} = b(t, X_t^{i,N}, \nu_t^N)\,\mathrm{d}t + \sigma(t, X_t^{i,N}, \nu_t^N)\,\mathrm{d}W_t^i,$$

and the non-interacting particle system

$$dX_t^i = b(t, X_t^i, \nu_t)\,\mathrm{d}t + \sigma(t, X_t^i, \nu_t)\,\mathrm{d}W_t^i,$$

with $X_0^i = X_0^{i,N} \sim \nu_0$. Then

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

Generic mean-field control problem

For any admissible control process α we consider $X_0^\alpha = \xi_0$ and

$$\mathrm{d}X_t^\alpha = b(t, X_t^\alpha, \alpha_t, \mathbb{P}_{X_t^\alpha})\,\mathrm{d}t + \sigma(t, X_t^\alpha, \alpha_t, \mathbb{P}_{X_t^\alpha})\,\mathrm{d}W_t.$$

Example: Cucker–Smale model

$b_1 = 1,$

$b_2 = \mathbb{E}'\left[\kappa(X_t - X'_t)\right] + \alpha_t,$

$\sigma_1 = 0,$

$\sigma_2 = 1.$

Minimise

$$F(\alpha) = \mathbb{E}\left[\int_0^T f(t, X_t^\alpha, \alpha_t, \mathbb{P}_{X_t^\alpha})\,\mathrm{d}t + g(X_T^\alpha, \mathbb{P}_{X_T^\alpha})\right].$$

Fully coupled MV-FBSDEs

by stochastic maximum principle

We consider the McKean–Vlasov Forward-Backward-SDE

$$X_t = \xi_0 + \int_0^t b(s, \Theta_s, \mathbb{P}_{\Theta_s})\, ds + \int_0^t \sigma(s, \Theta_s, \mathbb{P}_{\Theta_s})\, dW_s,$$

$$Y_t = g(X_T, \mathbb{P}_{X_T}) + \int_t^T f(s, \underbrace{X_s, Y_s, Z_s}_{\Theta_s}, \underbrace{\mathbb{P}_{(X_s, Y_s, Z_s)}}_{\mathbb{P}_{\Theta_s}})\, ds - \int_t^T Z_s\, dW_s,$$

where

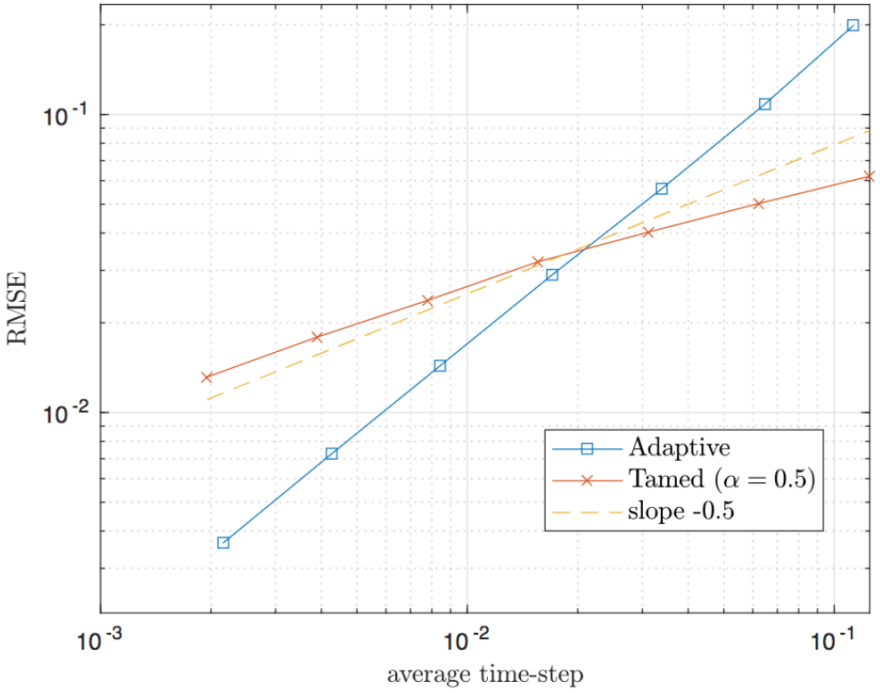
- X, Y, Z are unknown solution processes taking values in $\mathbb{R}^n, \mathbb{R}^m, \mathbb{R}^{m\times d}$,
- $T > 0$ is an arbitrary given finite number,
- ξ_0 is a given n-dimensional random variable,
- W is a d-dimensional standard Brownian motion,
- $\mathbb{P}_{(X_t, Y_t, Z_t)}$ is the marginal law of the process (X, Y, Z) at time $t \in [0, T)$,
- $\mathbb{P}_{X_T}$ is the marginal law of the process X at the terminal time T, and
- b, σ, g, h are given functions with appropriate dimensions.

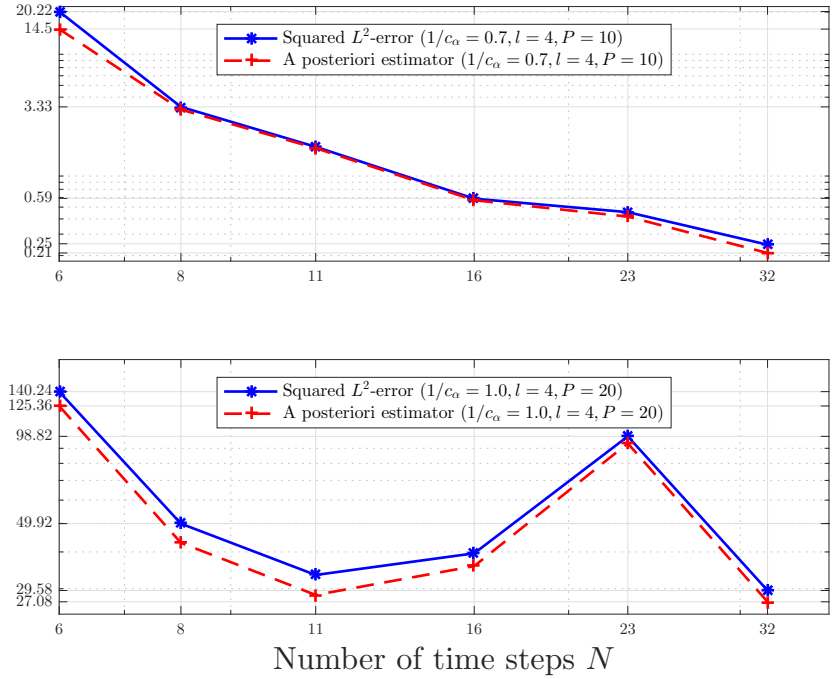
Solving MV-FBSDEs

- Time-stepping scheme, such as Euler–Maruyama.
- Solve discrete (MV-)FBSDE:
  - Projection on polynomial bases: Gobet, Lemor, & Warin (2006)
  - Picard iterations, Markovian iterations: Bender & Zhang (2008)
  - Picard iterations, continuation, trees: Chassagneux, Crisan, & Delarue (2019)
  - Neural networks and gradient descent: eg, Carmona &Laurière (2019), Germain, Mikael, & Warin (2022)

In all cases, iterations and projections.

MV-SDEs versus MV-FBSDEs





Mean-field FitzHugh–Nagumo MV-SDE

R, Stockinger (2021)

Linear MV-FBSDE

R, Stockinger, Zhang (2023)

MV-FBSΔE

For each $N \in \mathbb{N}$, consider on the time grid π_N: for all $i \in \mathcal{N}_{<N}$,

$$\Delta X_i^\pi = b(t_i, \Theta_i^\pi, \mathbb{P}_{\Theta_i^\pi})\tau_N + \sigma(t_i, \Theta_i^\pi, \mathbb{P}_{\Theta_i^\pi})\,\Delta W_i,$$

$$\Delta Y_i^\pi = -f(t_i, \underbrace{X_i^\pi, Y_i^\pi, Z_i^\pi}_{\Theta_i^\pi}, \underbrace{\mathbb{P}_{(X_i^\pi, Y_i^\pi, Z_i^\pi)}}_{\mathbb{P}_{\Theta_i^\pi}})\tau_N + Z_i^\pi\,\Delta W_i + \Delta M_i^\pi,$$

$$X_0^\pi = \xi_0, \quad Y_N = g(X_N^\pi, \mathbb{P}_{X_N^\pi}),$$

where

- $\xi_0 \in L^2(\mathcal{F}_0; \mathbb{R}^n)$,
- the solution processes X^π, Y^π, Z^π, M^π take values in $\mathbb{R}^n$, $\mathbb{R}^m$, $\mathbb{R}^{m\times d}$, $\mathbb{R}^m$,
- the coefficients (b, σ, f, g) are (possibly random) functions with appropriate dimensions, and
- $W = (W_t)_{t\in[0,T]} \in \mathcal{M}^2(0, T; \mathbb{R}^d)$ is a given martingale process satisfying for all $i \in \mathcal{N}_{<N}$, $\mathbb{E}_i[\Delta W_i(\Delta W_i)^*] = \tau_N \mathbb{I}_d$.

A computable error estimator

cf. Bender & Steiner (2013) for BSDEs; Han & Long (2022) for (weakly coupled) FBSDEs

Given $(\hat{X}, \hat{Y}, \hat{Z}, \hat{M})$ from any numerical scheme on π_N. Consider

$$\mathcal{E}_\pi(\hat{X}, \hat{Y}, \hat{Z}, \hat{M}) := \mathbb{E}[|\hat{X}_0 - \xi_0|^2] + \mathbb{E}[|\hat{Y}_N - g(\hat{X}_N, \mathbb{P}_{\hat{X}_N})|^2]$$

$$+ \max_{i \in \mathcal{N}_{<N}} \mathbb{E}\left[\left|\hat{X}_{i+1} - \hat{X}_0 - \sum_{j=0}^{i} \left(b(t_j, \hat{\Theta}_j, \mathbb{P}_{\hat{\Theta}_j})\tau_N + \sigma(t_j, \hat{\Theta}_j, \mathbb{P}_{\hat{\Theta}_j})\,\Delta W_j\right)\right|^2\right]$$

$$+ \max_{i \in \mathcal{N}_{<N}} \mathbb{E}\left[\left|\hat{Y}_{i+1} - \hat{Y}_0 + \sum_{j=0}^{i} \left(f(t_j, \hat{\Theta}_j, \mathbb{P}_{\hat{\Theta}_j})\tau_N - \hat{Z}_j\,\Delta W_j\right) - \hat{M}_{i+1}\right|^2\right].$$

- We demonstrate the reliability and efficiency of the proposed *a posteriori* error estimator for the FBSΔE.
- We estimate the error between discrete and continuous solutions *a priori*.

Reliability and Efficiency

R, Stockinger, Zhang (2023)

- Reliability: error $\lesssim$ tolerance
- Efficiency: error $\gtrsim$ tolerance

Under assumptions (see next slide), there is $C > 0$ such that

$$\mathcal{E}_\pi(\hat{X}, \hat{Y}, \hat{Z}, \hat{M})/C \leq$$

$$\max_{i \in \mathcal{N}} \left(\mathbb{E}[|\hat{X}_i - X_i^\pi|^2] + \mathbb{E}[|\hat{Y}_i - Y_i^\pi|^2] \right) + \sum_{i=0}^{N-1} \mathbb{E}[|\hat{Z}_i - Z_i^\pi|^2]\tau_N$$

$$+\mathbb{E}[|\hat{M}_N - M_N^\pi|^2]$$

$$\leq C\mathcal{E}_\pi(\hat{X}, \hat{Y}, \hat{Z}, \hat{M}).$$

Well-posedness of FBSΔE

see Peng & Wu (1999) and Bensoussan, Yam, & Zhang (2015) for continuous (MV-)FBSDEs

UNIVERSITY OF OXFORD
Mathematical Institute

We need the following assumptions on the generator (b, σ, f, g):

- Monotonicity:
  There exists a full-rank matrix $G \in \mathbb{R}^{m\times n}$ and $\alpha \geq 0$, $\beta_1, \beta_2 \geq 0$ with $\alpha + \beta_1 > 0$ and $\beta_1 + \beta_2 > 0$ such that $\beta_1 > 0$ (resp. $\alpha > 0, \beta_2 > 0$) when $m < n$ (resp. $m > n$), it holds $\mathbb{P}$-a.s., $\forall t \in [0, T]$, $\Theta_i \in L^2(\mathcal{F}; \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{m\times d})$,

$$\begin{aligned}
&\mathbb{E}[\langle b(t,\Theta_1,\mathbb{P}_{\Theta_1}) - b(t,\Theta_2,\mathbb{P}_{\Theta_2}), G^*(\delta Y)\rangle] + \mathbb{E}[\langle \sigma(t,\Theta_1,\mathbb{P}_{\Theta_1}) - \sigma(t,\Theta_2,\mathbb{P}_{\Theta_2}), G^*(\delta Z)\rangle] \\
&\quad + \mathbb{E}[\langle -f(t,\Theta_1,\mathbb{P}_{\Theta_1}) + f(t,\Theta_2,\mathbb{P}_{\Theta_2}), G(\delta X)\rangle] \\
&\quad \leq -\beta_1(\mathbb{E}[|G^*(\delta Y)|^2] + \mathbb{E}[|G^*(\delta Z)|^2]) - \beta_2\mathbb{E}[|G(\delta X)|^2], \\
&\mathbb{E}[\langle g(X_1,\mathbb{P}_{X_1}) - g(X_2,\mathbb{P}_{X_2}), G(\delta X)\rangle] \geq \alpha\,\mathbb{E}[|G(\delta X)|^2].
\end{aligned}$$

- Lipschitz continuity
- Integrability

Then, the FBSΔE has a unique solution.

Continuous solution

Define the error

$$\mathrm{ERR}(\hat{X}, \hat{Y}, \hat{Z})$$

$$:= \max_{i \in \mathcal{N}_{<N}} \max_{t \in [t_i, t_{i+1}]} \Big(\mathbb{E}[|X_t - \hat{X}_i|^2] + \mathbb{E}[|Y_t - \hat{Y}_i|^2] \Big) + \sum_{i=0}^{N-1} \mathbb{E}\Big[\int_{t_i}^{t_{i+1}} |Z_t - \hat{Z}_i|^2] \, dt \Big],$$

a measure of regularity:

$$\mathcal{R}_\pi(X, Y, Z)$$

$$:= \max_{i \in \mathcal{N}_{<N}} \max_{t \in [t_i, t_{i+1}]} \big(\mathbb{E}[|X_t - X_i|^2] + \mathbb{E}[|Y_t - Y_i|^2] \big) + \sum_{i=0}^{N-1} \mathbb{E}\Big[\int_{t_i}^{t_{i+1}} |Z_t - \bar{Z}_i|^2 \, dt \Big],$$

and $\overline{\omega}$ the modulus of continuity in t of b, σ, f.

Error estimators

R, Stockinger, Zhang (2023)

Under the previous assumptions, there exists $C > 0$ such that, for *any* $(\hat{X}, \hat{Y}, \hat{Z})$,

$$\mathrm{ERR}(\hat{X}, \hat{Y}, \hat{Z}) \leq C\big(\overline{\omega}(\tau_N)^2 + \mathcal{R}_\pi(X, Y, Z) + \mathcal{E}_\pi(\hat{X}, \hat{Y}, \hat{Z})\big),$$

$$\mathcal{E}_\pi(\hat{X}, \hat{Y}, \hat{Z}) \leq C\big(\overline{\omega}(\tau_N)^2 + \mathcal{R}_\pi(X, Y, Z) + \mathrm{ERR}(\hat{X}, \hat{Y}, \hat{Z})\big).$$

Return to Cucker–Smale model

We consider $\mathrm{d}X_t = V_t\,\mathrm{d}t$,

$$\mathrm{d}V_t = \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} \kappa(X_t, V_t, x', v')\, \mathcal{L}_{(X_t, V_t)}(\mathrm{d}x', \mathrm{d}v') + \alpha_t \right) \mathrm{d}t + \sigma\, \mathrm{d}W_t,$$

with initial state $(X_0, V_0) \in L^2(\Omega; \mathbb{R}^d \times \mathbb{R}^d)$, and

$$\kappa(x, v, x', v') = \frac{K(v' - v)}{(1 + |x - x'|^2)^\beta}, \quad \text{with some } \beta, K \geq 0.$$

Find an n-dimensional adapted control $(\alpha_t)_{t \in [0,T]}$ to minimize

$$F(\alpha) = \mathbb{E}\left[\int_0^T \left(|V_t - \mathbb{E}[V_t]|^2 + \gamma |\alpha_t|^2 \right) \mathrm{d}t + |V_T - \mathbb{E}[V_T]|^2 \right].$$

Numerical illustration (Cucker-Smale model, 2d)

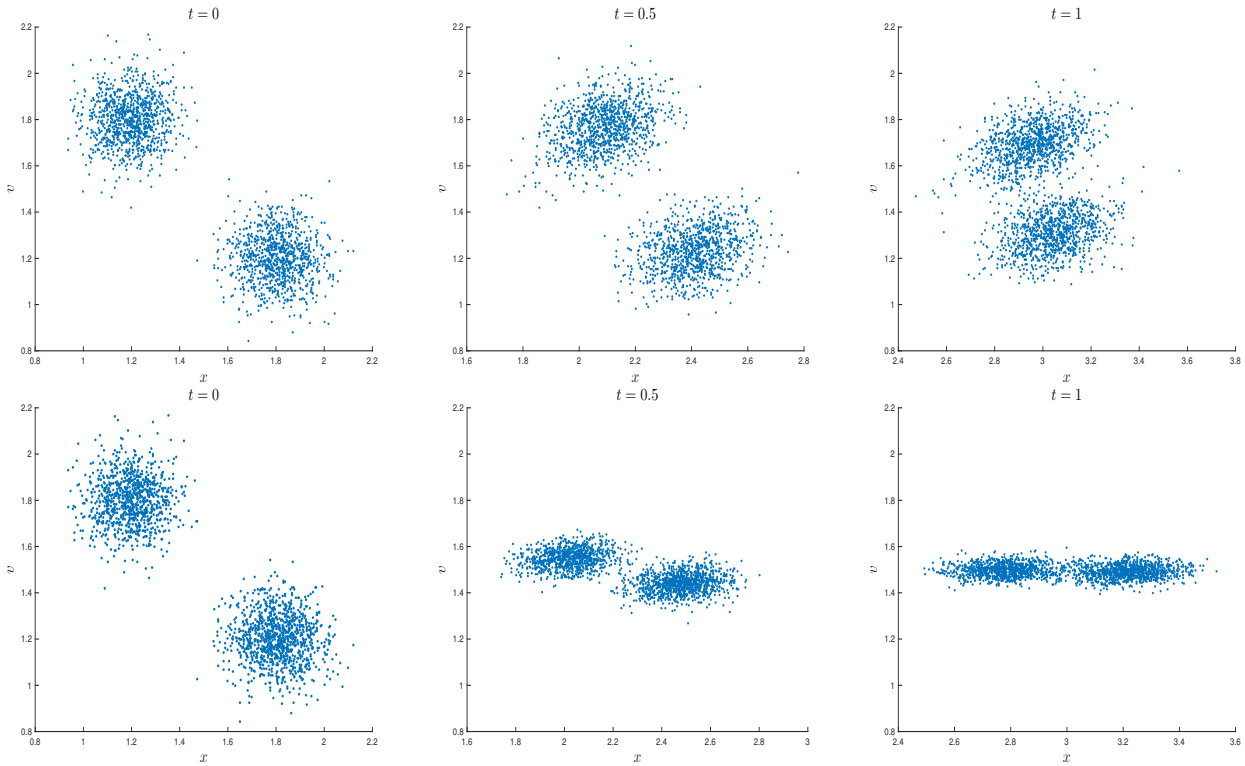


Figure: Uncontrolled (top) and controlled (bottom) models with $\beta = 10$.

Stochastic maximum principle

Value function and derivative are characterised by the MV-FBSDE

$$dX_t = V_t\,dt, \quad dV_t = \left(\mathbb{E}[\kappa(x,v,X_t,V_t)]\big|_{(x,v)=(X_t,V_t)} - \frac{1}{2\gamma}Y_t^2\right)dt + \sigma\,dW_t,$$

$$dY_t^1 = -\left(\mathbb{E}[\partial_x\kappa(x,v,X_t,V_t)]\big|_{(x,v)=(X_t,V_t)}Y_t^2 + \mathbb{E}[\partial_{x'}\kappa(X_t,V_t,x,v)Y_t^2]\big|_{(x,v)=(X_t,V_t)}\right)dt + Z_t^1\,dW_t,$$

$$dY_t^2 = -\Big(Y_t^1 + \mathbb{E}[\partial_v\kappa(x,v,X_t,V_t)]\big|_{(x,v)=(X_t,V_t)}Y_t^2 + \mathbb{E}[\partial_{v'}\kappa(X_t,V_t,x,v)Y_t^2]\big|_{(x,v)=(X_t,V_t)}$$

$$+ 2(V_t - \mathbb{E}[V_t])\Big)\,dt + Z_t^2\,dW_t,$$

$$X_0 = x_0, \quad V_0 = v_0, \quad Y_T^1 = 0, \quad Y_T^2 = 2(V_T - \mathbb{E}[V_T]),$$

where x_0, v_0 $\mathbb{R}^n$-valued square integrable RVs, W is an n-dimensional SBM.

Deep BSDE Solver

$n = 2$

- Use a neural network $f_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ (with one hidden layer of width 20 and the sigmoid activation function) to approximate the decoupling field of (Y_0^1, Y_0^2),
- and a NN $g_\vartheta : \mathbb{R}^3 \to \mathbb{R}^2$ (with width 110) to approximate the decoupling fields of (Z^1, Z^2) at all times.
- On a uniform grid π of $[0, 1]$ with stepsize $1/32$, compute $\hat{\Theta} = (\hat{X}, \hat{V}, \hat{Y}^1, \hat{Y}^2, \hat{Z}^1, \hat{Z}^2)$ by forward Euler scheme.
- Estimate the law $\mathbb{P}_{\hat{\Theta}}$ by a particle approximation of size 500.
- Minimise the terminal loss via the Adam algorithm:

$$\mathcal{E}(\hat{\Theta}) = \mathbb{E}[|\hat{Y}_N^1|^2] + \mathbb{E}[|\hat{Y}_N^2 - 2(\hat{V}_N - \mathbb{E}[\hat{V}_N])|^2].$$

Benchmark

We compute a reference solution with a gradient iteration proposed in *R, Stockinger, Zhang (2021)*:

- Define the gradient $(\nabla F)(\alpha)$.
- Compute a sequence of feedback controls ϕ^m as follows:
  - perform gradient iterations

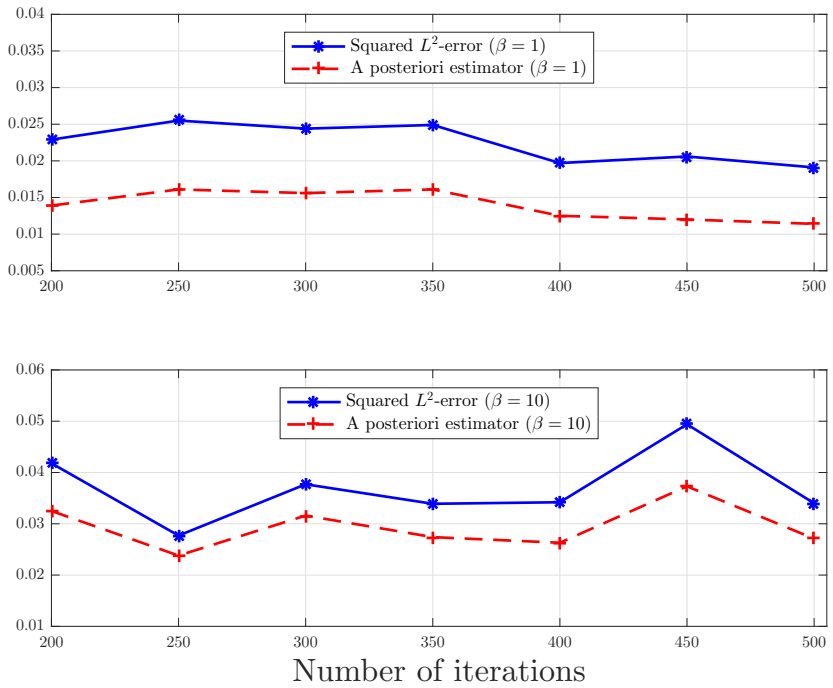
$$\phi^{m+1}(t,x,v) = \phi^m(t,x,v) - \tau(\nabla F)(\phi^m)(t,x,v);$$

  - find gradient by PDEs for decoupling fields and particle approximation.

Two-dimensional, nonlinear

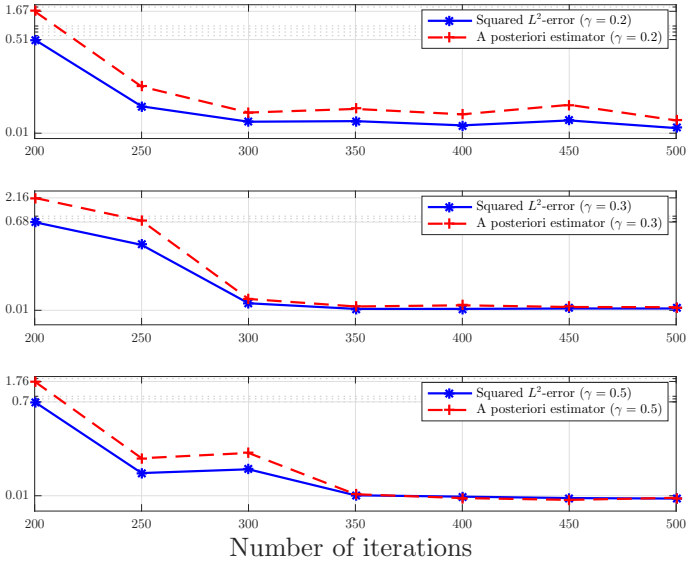
$$\mathcal{E}(\hat{\Theta}) = \mathbb{E}[|\hat{Y}_N^1|^2] + \mathbb{E}[|\hat{Y}_N^2 - 2(\hat{V}_N - \mathbb{E}[\hat{V}_N])|^2]$$

$$\mathrm{ERR}(\hat{X}, \hat{Y}, \hat{Z}) = \max_{i \in \mathcal{N}_{<N}} \left(\mathbb{E}[|X_i - \hat{X}_i|^2] + \mathbb{E}[|Y_i - \hat{Y}_i|^2] \right) + \sum_{i \in \mathcal{N}_{<N}} \mathbb{E}[|Z_i - \hat{Z}_i|^2]$$

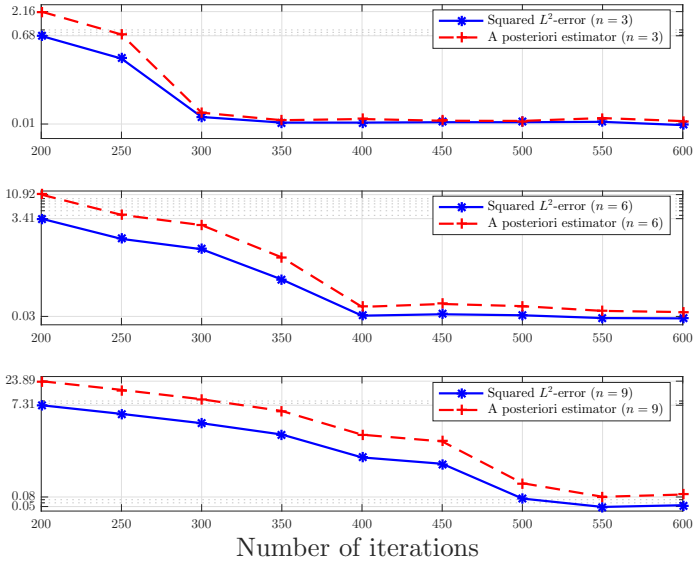


Comparison between the squared L^2-error and the a posteriori estimator with $\beta \in \{1, 10\}$ and different Adam iterations.

Higher dimensional, linear



$n = 3$, $\gamma \in \{0.2, 0.3, 0.5\}$



$\gamma = 0.3$, $n \in \{3, 6, 9\}$

Comparison between the squared L^2-error and the a posteriori estimator with different Adam iterations and values of γ and n (with $\beta = 0$).

Conclusions and references

- We show that an a posteriori error estimator is efficient and reliable for fully coupled MV-FBSDEs.
- We demonstrate its effectiveness for a number of test cases.

- C. Reisinger, W. Stockinger, Y. Zhang. A posteriori error estimates for fully coupled McKean–Vlasov forward-backward SDEs, arXiv.
- C. Reisinger, W. Stockinger, Y. Zhang. A fast iterative PDE-based algorithm for feedback controls of nonsmooth mean-field control problems, arXiv.