Revisiting nonlinear filtering through deep BSDE methods:

Joint work with Adam Andersson, Stig Larsson, & Filip Rydin

Center for Dynamics and Computation, Vrije Universiteit, Amsterdam

Kasper Bågmark

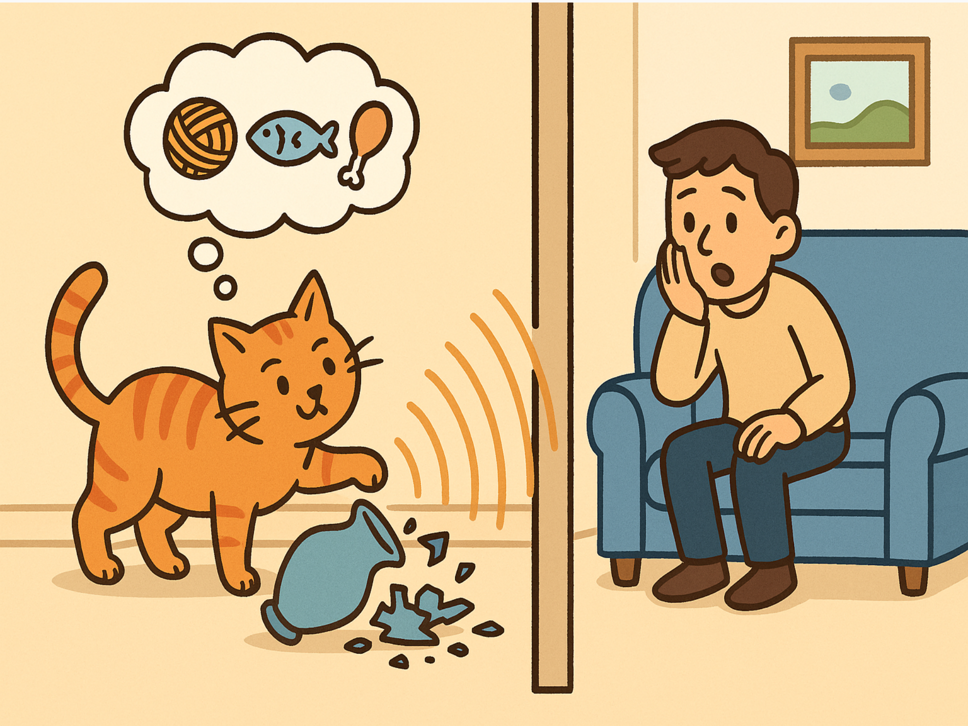
October 16, 2025

Chalmers University of Technology and University of Gothenburg,
Department of Mathematical Sciences

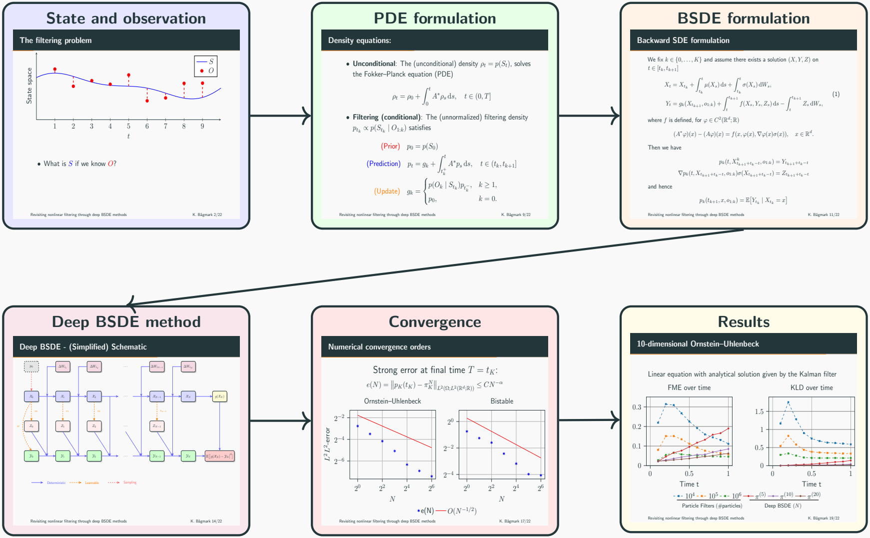
CHALMERS
UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GOTHENBURG

WASP
WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

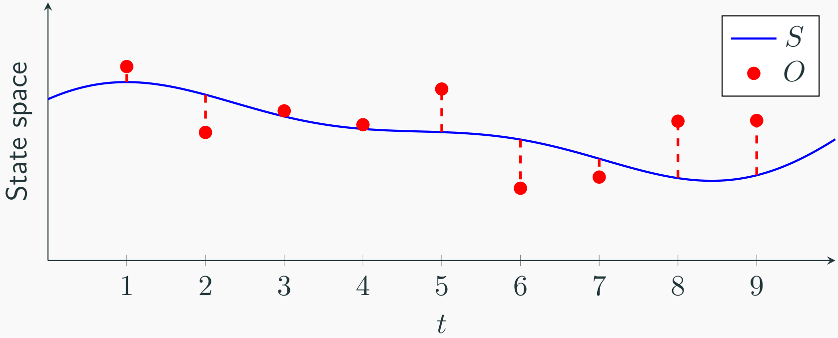


Summary


State and observation
The filtering problem
What is S if we know O?
PDE formulation
Density equations:
BSDE formulation
Backward SDE formulation
Deep BSDE method
Deep BSDE - (Simplified) Schematic
Convergence
Numerical convergence orders
Strong error at final time $T = t_K$:
Results
10-dimensional Ornstein–Uhlenbeck
Linear equation with analytical solution given by the Kalman filter

Introduction

The filtering problem



- Goal: Find the density p_{t_k}

$$\mathbb{P}(S_{t_k} \in B \mid O_{1:k}) = \int_B p_{t_k}(x \mid O_{1:k})\,\mathrm{d}x$$

Contributed method

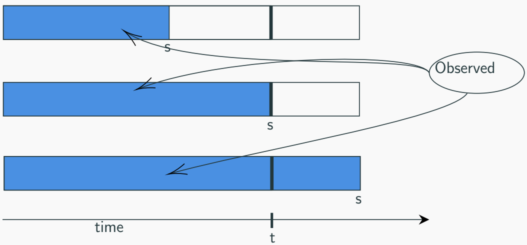
- Learns the map $O_{1:k} \to p(S_k \mid O_{1:k})$
  - Assumes known SDE coefficients and measurement model
- PDE based method, scalable in dimension of S
  - No spatial discretization required
  - Numerically stable for toy example in 100 dimensions
- Convergence
  - Strong convergence in the probabilistic representation
  - Convergence order empirically verified

Setting

The filtering problem

Goal: find $p(S_t \mid O_{0:s})$

- Prediction: $s < t$
- Filtering: $s = t$
- Smoothing: $s > t$



Continuous state

S solves the Stochastic Differential Equation (SDE)

$$S_t = S_0 + \int_0^t \mu(S_s)\,\mathrm{d}s + \int_0^t \sigma(S_s)\,\mathrm{d}B_s, \quad t \in [0, T].$$

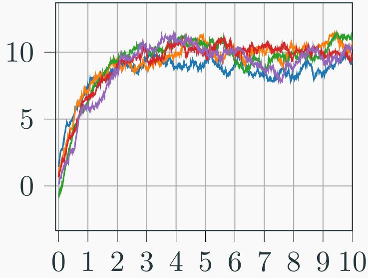
Infinitesimal generator A

$$A\varphi = \frac{1}{2}\sum_{i,j=1}^{d} a_{ij}\,\frac{\partial^2 \varphi}{\partial x_i \partial x_j} + \sum_{i=1}^{d} \mu_i\,\frac{\partial \varphi}{\partial x_i}$$

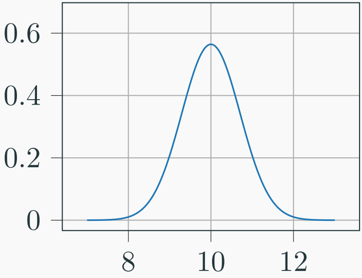
where $a := \sigma\sigma^\top$

Example: Ornstein–Uhlenbeck process (linear SDE)

$$S_t = S_0 - \int_0^t (S_s - 10)\,\mathrm{d}s + B_t, \quad t \in [0, T],$$
$$S_0 \sim p_0 = \mathcal{N}(0, 1)$$



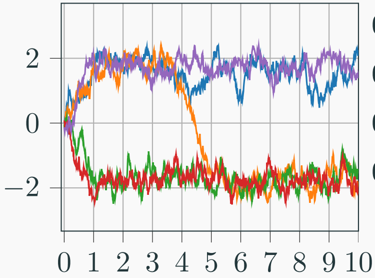
State $(S_t)_{t\in[0,T]}$



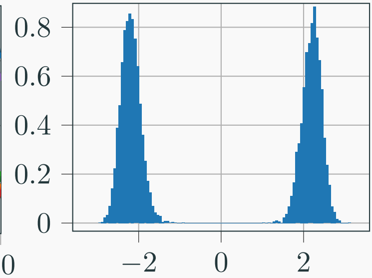
Distribution $p(S_T)$

Example: Bistable process (nonlinear SDE)

$$S_t = S_0 + \int_0^t (5S_s - S_s^3)\,\mathrm{d}s + B_t, \quad t \in [0,T],$$
$$S_0 \sim p_0 = \mathcal{N}(0,1)$$



State $(S_t)_{t\in[0,T]}$



Distribution $p(S_T)$

Discrete observations

- **Measurement model**:

$$O_k = h(S_{t_k}) + U_k, \quad k = 1, \ldots, K,$$

where $h\colon \mathbb{R}^d \to \mathbb{R}^{d'}$ and $U_k \sim \mathcal{N}(0, R_k)$.

- **Goal**: Find $p(S_{t_k} \mid O_{1:k})$ for $k = 1, \ldots, K$.

Classical methods

Linear and Gaussian case

Exact solutions:

- Prediction and filtering:
  The Kalman filter (Kalman–Bucy 1960)
- Smoothing:
  Rauch–Tung–Striebel smoother (1965)

Nonlinear case

Approximations:

- Approximative Kalman filter (Extended, Unscented, Ensemble)
- Particle filters (sequential Monte Carlo)

PDE formulation

Density equations:

- **Unconditional**: The (unconditional) density $\rho_t = p(S_t)$, solves the Fokker–Planck equation (PDE)

$$\rho_t = \rho_0 + \int_0^t A^* \rho_s \, \mathrm{d}s, \quad t \in (0, T]$$

- **Filtering (conditional)**: The (unnormalized) filtering density $p_{t_k} \propto p(S_{t_k} \mid O_{1:k})$ satisfies

$$\text{(Prior)} \quad p_0 = p(S_0)$$

$$\text{(Prediction)} \quad p_t = g_k + \int_{t_k^+}^t A^* p_s \, \mathrm{d}s, \quad t \in (t_k, t_{k+1}]$$

$$\text{(Update)} \quad g_k = \begin{cases} p(O_k \mid S_{t_k}) p_{t_k^-}, & k \geq 1, \\ p_0, & k = 0. \end{cases}$$

BSDE approach - brief outline

Backward SDE formulation

Define f, for $\varphi \in C^2(\mathbb{R}^d; \mathbb{R})$, by

$$(A^*\varphi)(x) - (A\varphi)(x) = f(x, \varphi(x), \sigma(x)^\top \nabla\varphi(x)), \quad x \in \mathbb{R}^d.$$

We fix $k \in \{0, \ldots, K\}$ and assume there exists a solution (X, Y, Z) on $t \in [t_k, t_{k+1}]$

$$\begin{aligned} X_t &= X_{t_k} + \int_{t_k}^{t} \mu(X_s)\,\mathrm{d}s + \int_{t_k}^{t} \sigma(X_s)\,\mathrm{d}W_s, \\ Y_t &= g_k(X_{t_{k+1}}, o_{1:k}) + \int_{t}^{t_{k+1}} f(X_s, Y_s, Z_s)\,\mathrm{d}s - \int_{t}^{t_{k+1}} Z_s^\top\,\mathrm{d}W_s. \end{aligned} \tag{1}$$

Then we have

$$p_k(t, X^k_{t_{k+1}+t_k-t}, o_{1:k}) = Y_{t_{k+1}+t_k-t} \tag{2}$$

$$\sigma(X_{t_{k+1}+t_k-t})^\top \nabla p_k(t, X_{t_{k+1}+t_k-t}, o_{1:k}) = Z_{t_{k+1}+t_k-t} \tag{3}$$

and hence

$$p_k(t_{k+1}, x, o_{1:k}) = \mathbb{E}\big[Y_{t_k} \mid X_{t_k} = x\big] \tag{4}$$

Continuous optimization formulation

The solution p is the solution to the following minimization problem

$$\min_{u \in C([t_k, t_{k+1}] \times \mathbb{R}^d \times \mathbb{R}^{d' \times k}; \mathbb{R})} \mathbb{E}\left[\left|Y_{t_{k+1}}^{(u,o_{1:k})} - g_k(X_{t_{k+1}}, o_{1:k})\right|^2\right]$$

$$Y_t^{(u,o_{1:k})} = u(t, X_t, o_{1:k}), \quad Z_t^{(u,o_{1:k})} = \sigma(X_t)^\top \nabla u(t, X_t, o_{1:k})$$

$$X_t = X_{t_k} + \int_{t_k}^t \mu(X_s)\,\mathrm{d}s + \int_{t_k}^t \sigma(X_s)\,\mathrm{d}W_s, \quad t \in [t_k, t_{k+1}],$$

$$Y_t^{(u,o_{1:k})} = Y_{t_k}^{(u,o_{1:k})} - \int_{t_k}^t f(X_s, Y_s^{(u,o_{1:k})}, Z_s^{(u,o_{1:k})})\,\mathrm{d}s + \int_{t_k}^t (Z_s^{(u,o_{1:k})})^\top\,\mathrm{d}W_s, \quad t \in [t_k, t_{k+1}].$$

The discrete optimization problem

- Define a finer time partition $t_k = t_{k,0} < t_{k,1} < \cdots < t_{k,N} = t_{k+1}$
- Approximate $(X_t, Y_t)_{t \in [t_k, t_{k+1}]}$ with Euler–Maruyama $(\mathcal{X}_n, \mathcal{Y}_n)_{n=0}^N$
- Approximate $u(t_k, x) \approx w(x)$ and $\nabla u(t_{k,n}, x) \approx v_n(x)$, $n = 0, \ldots, N-1$

$$\min_{\substack{w \in C(\mathbb{R}^d \times \mathbb{R}^{d' \times k}; \mathbb{R}) \\ (v_n)_{n=0}^{N-1} \in C(\mathbb{R}^d \times \mathbb{R}^{d' \times k}; \mathbb{R}^d)^N}} \mathbb{E}\Big[\Big|\mathcal{Y}_N^{o_{1:k}} - \bar{g}_k(\mathcal{X}_N, o_{1:k})\Big|^2\Big]$$

$$\mathcal{Y}_0^{o_{1:k}} = w(\mathcal{X}_0, o_{1:k}),$$

for $n = 0, \ldots, N-1$

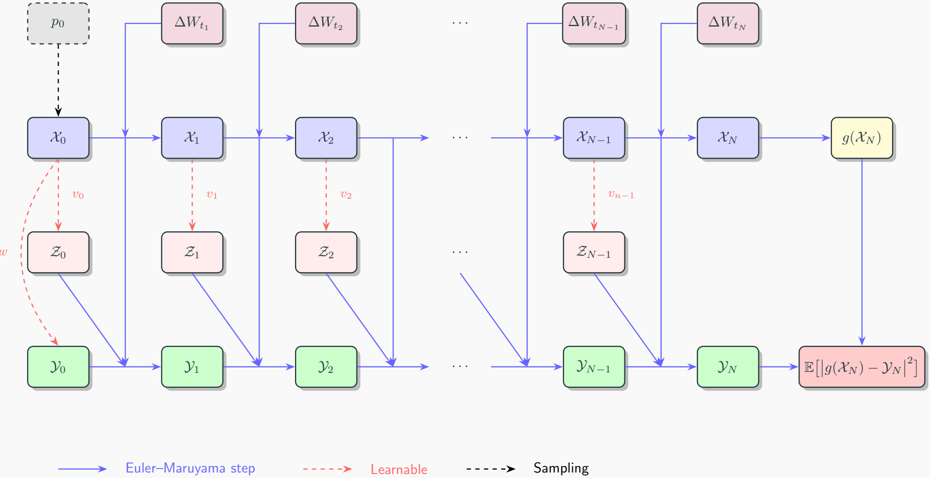
$$\mathcal{Z}_n^{o_{1:k}} = \sigma(\mathcal{X}_n)^\top v_n(\mathcal{X}_n, o_{1:k}),$$

$$\mathcal{X}_{n+1} = \mathcal{X}_n + \mu(\mathcal{X}_n)(t_{k,n+1} - t_{k,n}) + \sigma(\mathcal{X}_n)(W_{t_{k,n+1}} - W_{t_{k,n}}),$$

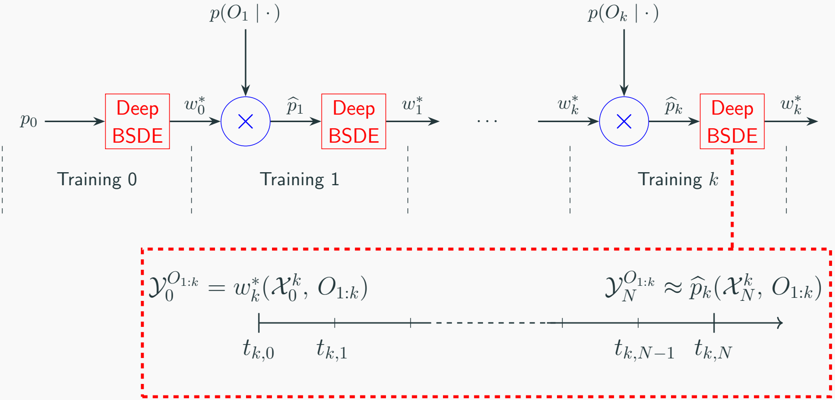
$$\mathcal{Y}_{n+1}^{o_{1:k}} = \mathcal{Y}_n^{o_{1:k}} - f\big(\mathcal{X}_n, \mathcal{Y}_n^{o_{1:k}}, \mathcal{Z}_n^{o_{1:k}}\big)(t_{k,n+1} - t_{k,n}) + (\mathcal{Z}_n^{o_{1:k}})^\top (W_{t_{k,n+1}} - W_{t_{k,n}}).$$

Let $\widehat{p}_k^N(x, o_{1:k}) = w^*(x, o_{1:k-1}) p(O_k = o_k \mid S_{t_k} = x)$ define our deep BSDE filter

Deep BSDE - (Simplified) Schematic



deep BSDE filter



Numerical convergence

Error bound

- p - true filter solution
- $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$ - deep BSDE approximation
- $\widehat{p}$ - filter approximation

Under sufficient conditions there exists a constant C such that, for all $k = 1, \ldots, K$

$$\left\|p_k(t_k) - \widehat{p}_k^N\right\|_{L^\infty(\mathbb{O};L^\infty(\mathbb{R}^d;\mathbb{R}))} \leq C\left(N^{-\frac{1}{2}} + \sum_{j=0}^{K-1} \sup_{o\in\mathbb{O}} \sup_{x\in\mathbb{R}^d} \mathbb{E}\left[\left|\overline{g}_j(\mathcal{X}_N^{j,x}, o_{1:j}) - \mathcal{Y}_N^{j,x}\right|^2\right]^{\frac{1}{2}}\right)$$

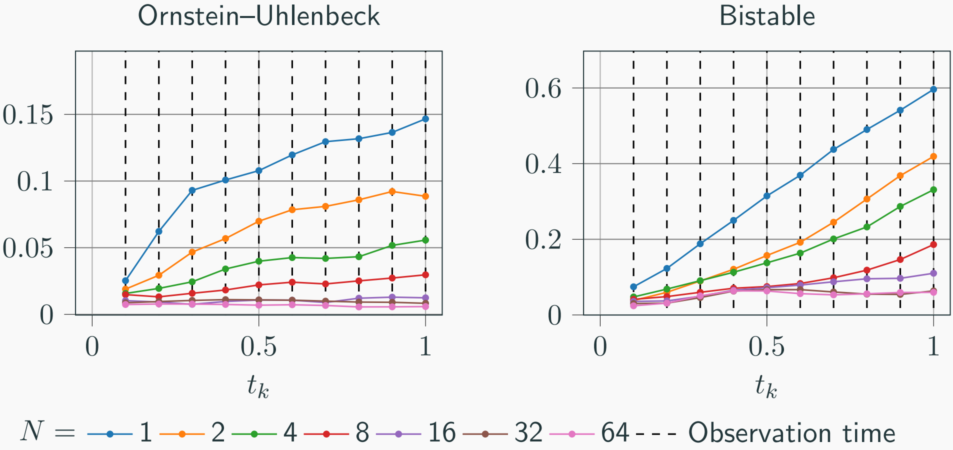
Define

$$e_k(N) = \left\|p_k(t_k) - \widehat{p}_k^N\right\|_{L^\infty(\mathbb{O};L^\infty(\mathbb{R}^d;\mathbb{R}))}$$

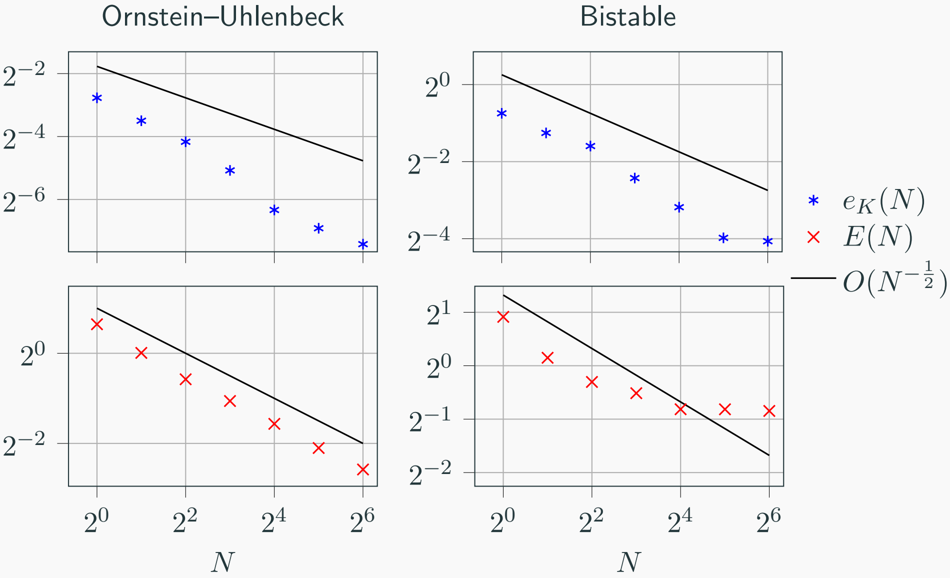
$$E(N) = \sum_{j=0}^{K-1} \sup_{o\in\mathbb{O}} \sup_{x\in\mathbb{R}^d} \mathbb{E}\left[\left|\overline{g}_j(\mathcal{X}_N^{j,x}, o_{1:j}) - \mathcal{Y}_N^{j,x}\right|^2\right]^{\frac{1}{2}}$$

Numerical convergence

Strong error $e_k(N) = \left\| p_k(t_k) - \widehat{p}_k^N \right\|_{L^\infty(\mathbb{O};L^\infty(\mathbb{R}^d;\mathbb{R}))}$ for $k = 1, \ldots, 10$



Numerical convergence orders



Comparison to particle filters

Metrics

Let μ and p denote the true mean and density, $\widehat{\mu}$ and $\widehat{p}$ a generic approximation

First Moment Error

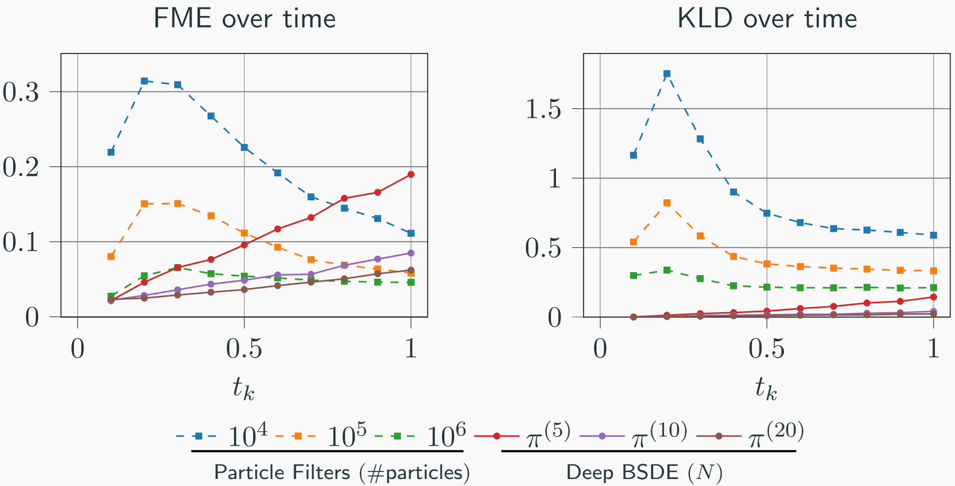
$$\mathsf{FME} = \frac{1}{M} \sum_{m=1}^{M} \left\| \mu_k^{(m)} - \widehat{\mu}_k^{(m)} \right\|, \quad \text{for } k = 1, \dots, K.$$

Forward averaged Kullback–Leibler Divergence

$$\mathsf{KLD} = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} \log \left(\frac{p_{t_k}^{(m)}\left(x^{(n,m)}\right)}{\widehat{p}_k^{(m)}\left(x^{(n,m)}\right)} \right), \;\; x^{(n,m)} \sim p_{t_k}^{(m)}, \;\; \text{for } k = 1, \dots, K$$

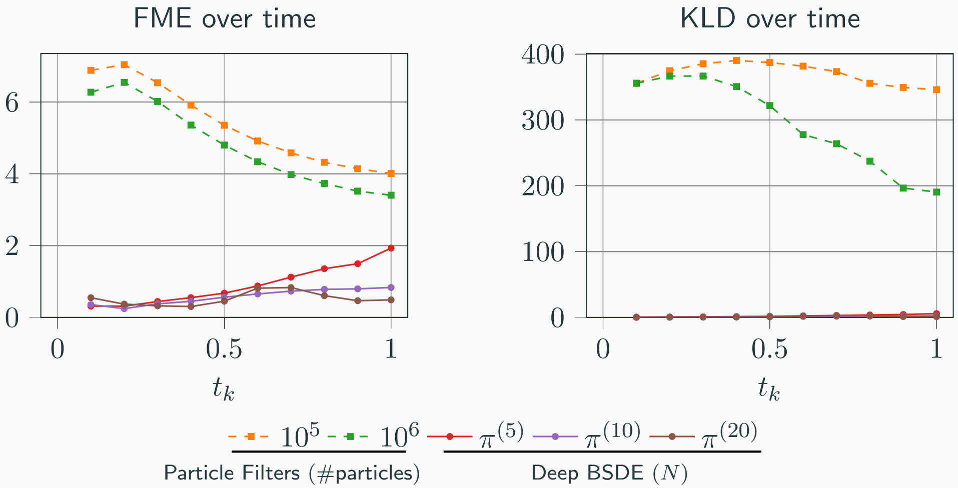
10-dimensional Ornstein–Uhlenbeck

Linear equation with analytical solution given by the Kalman filter



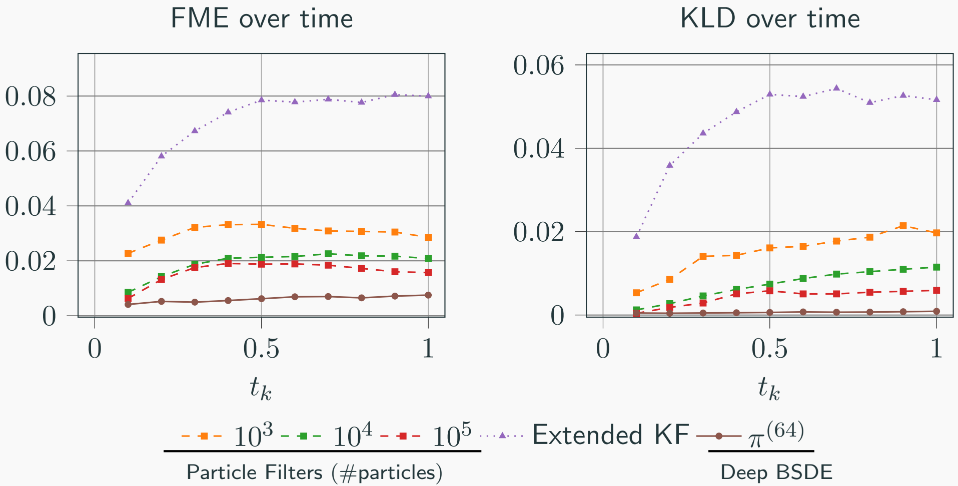
100-dimensional Ornstein–Uhlenbeck

Linear equation with analytical solution given by the Kalman filter



Bistable process

Nonlinear equation with a particle filter as reference solution



Summary

Goal: Solve the filtering problem without the curse of dimensionality

1. PDE formulation - the Fokker–Planck equation with updates
2. Reformulate as a BSDE
3. Euler–Maruyama and neural network approximation

Obtained objectives:

	Low-dimensional	**High-dimensional**
Linear	Particle Filter ✓ Proposed ✓	Particle Filter ✗ Proposed ✓
Nonlinear	Particle Filter ✓ Proposed ✓	Particle Filter ✗ Proposed (?)