High-dimensional Bayesian filtering through deep density approximation:

Joint work with Filip Rydin

Gothenburg statistics seminar, Kasper Bågmark

February 11, 2026

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

CHALMERS UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GOTHENBURG

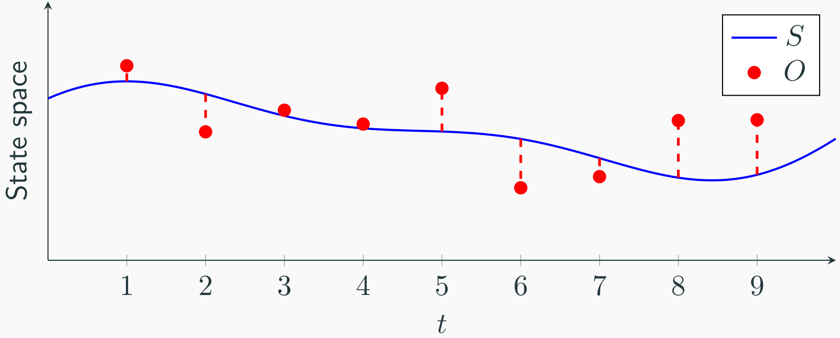
WASP WALLENBERG AI, AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

Outline

1. Problem formulation and classical filters
2. The Fokker–Planck PDE
3. Feynman–Kac representations - Optimization formulations
4. Experiments

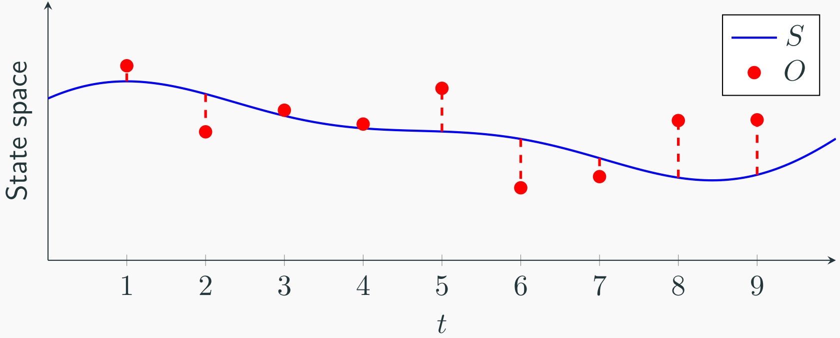
Introduction

The filtering problem



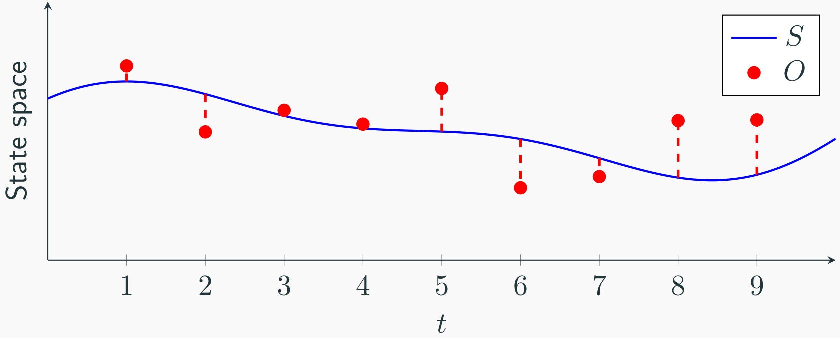
- What is S if we know O?

The filtering problem



- What is $\mathbb{P}(S_{t_k} \mid O_{1:k})$?

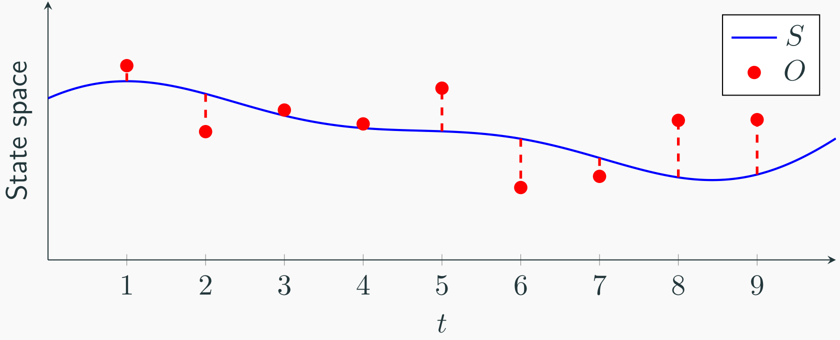
The filtering problem



- The filtering density p_{t_k} satisfies, for $B \subset \mathbb{R}^d$

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

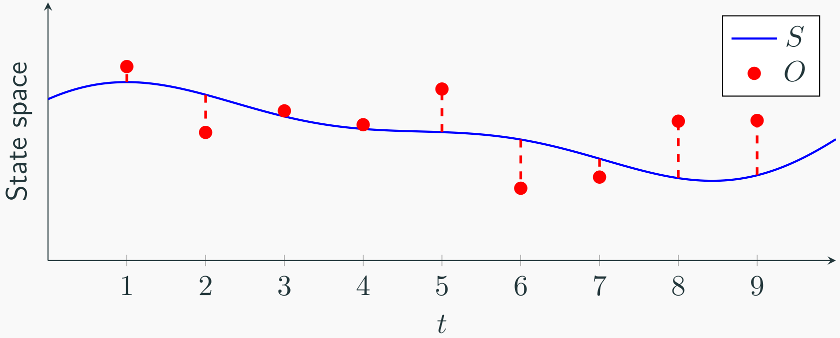
The filtering problem



- Goal: Find the density p_{t_k}

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

The filtering problem

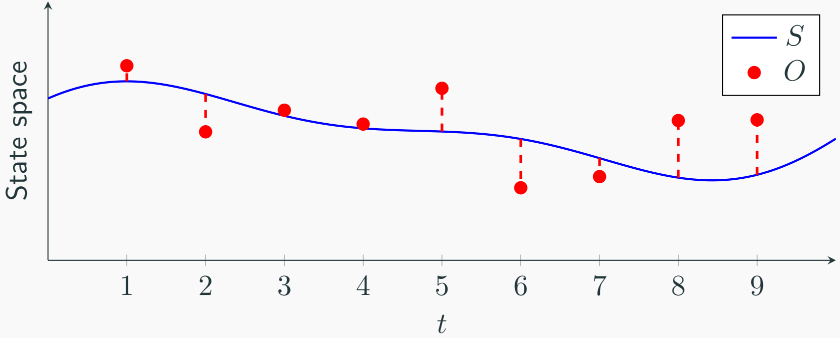


- Goal: Find the density p_{t_k}

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

Nonlinear dynamics: Particle filters (PF) $\sim O(n^d), \quad S_t \in \mathbb{R}^d$

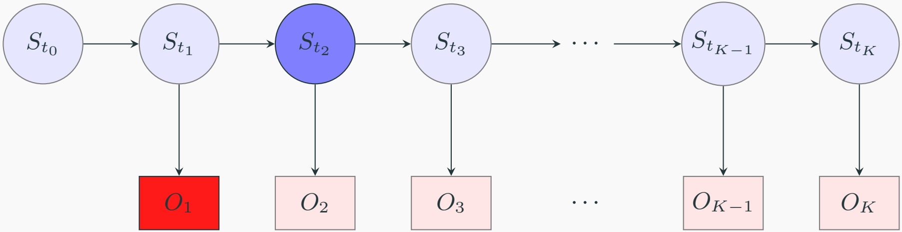
The filtering problem



- **Rephrased Goal:** Find p_t without the curse of dimensionality

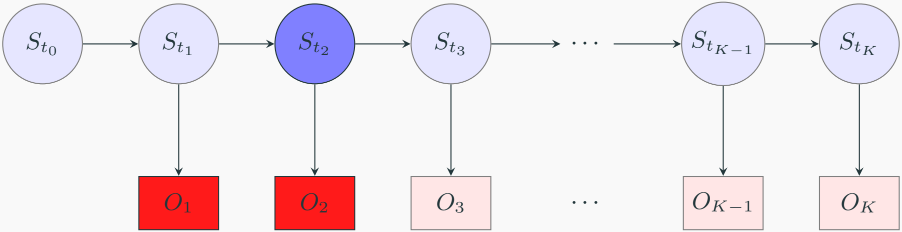
Setting

Prediction: $\mathbb{P}(S_{t_2} \mid O_1)$



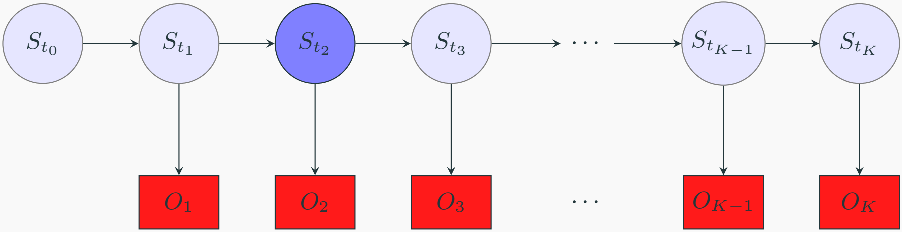
Prediction - Filtering - Smoothing

Filtering: $\mathbb{P}(S_{t_2} \mid O_{1:2})$



Prediction - Filtering - Smoothing

Smoothing: $\mathbb{P}(S_{t_2} \mid O_{1:K})$



Continuous-discrete setting

S solves the d-dimensional 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, with $a := \sigma\sigma^\top$, given by

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

- **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 $k = 1,\ldots,K$.

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 $p_k(t_k) \propto p(S_{t_k} \mid O_{1:k})$ satisfies

$$\text{(Prior)} \quad p_0(0) = p(S_0)$$

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

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

Feynman–Kac representations

Backward Kolmogorov

Define f such that for $\varphi \in D(A^*)$ we have

$$A^*\varphi - A\varphi = f(\varphi, \nabla\varphi).$$

Then for $q(t) = p_k(t_k + t_{k+1} - t)$, $t \in [t_k, t_{k+1}]$ it holds

$$\frac{\partial}{\partial t} q(t) + Aq(t) = -f(q(t), \nabla q(t)), \quad t \in [t_k, t_{k+1}],$$

$$q(t_{k+1}) = g_k.$$

Let X be an auxiliary process with the generator A,

$$X_t = X_0 + \int_0^t \mu(X_s)\,\mathrm{d}s + \int_0^t \sigma(X_s)\,\mathrm{d}W_s, \quad X_0 \sim p_0.$$

By Feynman–Kac formulas we derive two optimization schemes:

- Deep Splitting
- Deep Backward SDE

Deep Splitting

Deep splitting

Feynman–Kac on a time partition $t_k = t_{k,0} < \cdots < t_{k,N} = t_{k+1}$, recursively defined for $n = 0, 1 \ldots, N-1$

$$p_k(t_{k,n+1}, x) = \mathbb{E}\Big[p_k(t_{k,n}, X^{t_{k,n},x}_{t_{k,nn+1}}) + \int_{t_{k,n}}^{t_{k,n+1}} f\big(p_k(s, X^{t_{k,n},x}_{t_{k,n+1}+t_{k,n}-s}), \nabla p_k(s, X^{t_{k,n},x}_{t_{k,n+1}+t_{k,n}-s})\big)\,\mathrm{d}s\Big]$$

where for $t \in [t_{k,n}, t_{k,n+1}]$

$$X^{t_{k,n},x}_t = x + \int_{t_{k,n}}^{t} \mu(X^{t_{k,n},x}_s)\,\mathrm{d}s + \int_{t_{k,n}}^{t} \sigma(X^{t_{k,n},x}_s)\,\mathrm{d}W_s$$

Deep splitting

Feynman–Kac on a time partition $t_k = t_0 < \cdots < t_N = t_{k+1}$, recursively defined for $n = 0, 1 \ldots, N-1$

$$p_k(t_{n+1}, x) = \mathbb{E}\Big[p_k(t_n, X^x_{t_{n+1}}) + \int_{t_n}^{t_{n+1}} f\big(p_k(s, X^x_{t_{n+1}+t_n-s}), \nabla p_k(s, X^x_{t_{n+1}+t_n-s})\big)\,\mathrm{d}s\Big]$$

where for $t \in [t_n, t_{n+1}]$

$$X^x_t = x + \int_{t_n}^{t} \mu(X^x_s)\,\mathrm{d}s + \int_{t_n}^{t} \sigma(X^{t_n,x}_s)\,\mathrm{d}W_s$$

Deep splitting

Feynman–Kac on a time partition $t_k = t_0 < \cdots < t_N = t_{k+1}$, recursively defined for $n = 0, 1 \ldots, N-1$

$$p_k(t_{n+1}, x) \approx \mathbb{E}\Big[p_k(t_n, X^x_{t_{n+1}}) + f\big(p_k(t_n, X^x_{t_{n+1}}), \nabla p_k(t_n, X^x_{t_{n+1}})\big)\Delta t\Big]$$

where for $t \in [t_n, t_{n+1}]$

$$X^x_t = x + \int_{t_n}^t \mu(X^x_s)\,\mathrm{d}s + \int_{t_n}^t \sigma(X^x_s)\,\mathrm{d}W_s$$

Deep splitting

Feynman–Kac on a time partition $t_k = t_0 < \cdots < t_N = t_{k+1}$, recursively defined for $n = 0, 1 \ldots, N-1$

$$p_k(t_{n+1}, x) \approx \mathbb{E}\Big[p_k(t_n, \mathcal{X}^x_{n+1}) + f\big(p_k(t_n, \mathcal{X}^x_{n+1}), \nabla p_k(t_n, \mathcal{X}^x_{n+1})\big)\Delta t\Big]$$

where

$$\mathcal{X}^x_{n+1} = x + \mu(\mathcal{X}^x_n)\Delta t + \sigma(\mathcal{X}^x_n)(W_{t_{n+1}} - W_{t_n})$$

Deep splitting

Feynman–Kac on a time partition $t_k = t_0 < \cdots < t_N = t_{k+1}$, recursively defined for $n = 0, 1 \ldots, N-1$

$$\overline{\pi}_{k,n+1} = \underset{\phi \in C(\mathbb{R}^d;\mathbb{R})}{\arg\min}\, \mathbb{E}\Bigg[\bigg|\phi(\mathcal{X}_n) - \Big(\overline{\pi}_{k,n}(\mathcal{X}_{n+1}) + f\big(\overline{\pi}_{k,n}(\mathcal{X}_{n+1}), \nabla\overline{\pi}_{k,n}(\mathcal{X}_{n+1})\big)\Delta t\Big)\bigg|^2\Bigg]$$

where

$$\mathcal{X}_{n+1} = \mathcal{X}_n + \mu(\mathcal{X}_n)\Delta t + \sigma(\mathcal{X}_n)(W_{t_{n+1}} - W_{t_n})$$

Deep splitting

Feynman–Kac on a time partition $t_k = t_0 < \cdots < t_N = t_{k+1}$, recursively defined for $n = 0, 1 \ldots, N-1$

$$\pi_{k,n+1} = \underset{\phi \in C(\mathbb{R}^d \times \mathbb{R}^{d' \times k}; \mathbb{R})}{\arg\min} \mathbb{E}\Big[\Big|\phi(\mathcal{X}_n, O_{1:k}) - \Big(\pi_{k,n}(\mathcal{X}_{n+1}, O_{1:k}) + f\big(\pi_{k,n}(\mathcal{X}_{n+1}, O_{1:k}), \nabla\pi_{k,n}(\mathcal{X}_{n+1}, O_{1:k})\big)\Delta t\Big)\Big|^2\Big]$$

where

$$\mathcal{X}_{n+1} = \mathcal{X}_n + \mu(\mathcal{X}_n)\Delta t + \sigma(\mathcal{X}_n)(W_{t_{n+1}} - W_{t_n})$$

Deep Backward SDE

Deep Backward SDE

Nonlinear Feynman–Kac:

$$p_k(t_{k+1}, x) = \mathbb{E}\left[g_k(X^x_{t_{k+1}}) + \int_{t_k}^{t_{k+1}} f(X^x_s, Y_s, Z_s)\,\mathrm{d}s\right],$$

where (X, Y, Z), with $t \in [t_k, t_{k+1}]$, satisfies

$$X^x_t = x + \int_{t_k}^{t} \mu(X^x_s)\,\mathrm{d}s + \int_{t_k}^{t} \sigma(X^x_s)\,\mathrm{d}W_s,$$

$$Y_t = g_k(X^x_{t_{k+1}}, o_{1:k}) + \int_t^{t_{k+1}} f(X^x_s, Y_s, Z_s)\,\mathrm{d}s - \int_t^{t_{k+1}} Z_s^\top\,\mathrm{d}W_s.$$

Deep Backward SDE

Nonlinear Feynman–Kac:

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

where (X, Y, Z), with $t \in [t_k, t_{k+1}]$, satisfies

$$X_t^x = x + \int_{t_k}^{t} \mu(X_s^x)\,\mathrm{d}s + \int_{t_k}^{t} \sigma(X_s^x)\,\mathrm{d}W_s,$$

$$Y_t = g_k(X_{t_{k+1}}^x, o_{1:k}) + \int_t^{t_{k+1}} f(X_s^x, Y_s, Z_s)\,\mathrm{d}s - \int_t^{t_{k+1}} Z_s^\top\,\mathrm{d}W_s.$$

Deep Backward SDE

Nonlinear Feynman–Kac:

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

where (X, Y, Z), with $t \in [t_k, t_{k+1}]$, satisfies

$$X_t = X_{t_k} + \int_{t_k}^{t} \mu(X_s) \, \mathrm{d}s + \int_{t_k}^{t} \sigma(X_s) \, \mathrm{d}W_s,$$

$$Y_t = u(t_k, X_{t_k}) - \int_{t_k}^{t} f(X_s, Y_s, Z_s) \, \mathrm{d}s + \int_{t_k}^{t} Z_s^{\top} \, \mathrm{d}W_s,$$

$$Z_t = \nabla u(t, X_t).$$

Deep Backward SDE

Nonlinear Feynman–Kac:

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

where $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$, with $n = 0, \ldots, N-1$, satisfies

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

$$Y_t = u(t_k, X_{t_k}) - \int_{t_k}^{t} f(X_s, Y_s, Z_s)\, \mathrm{d}s + \int_{t_k}^{t} Z_s^\top\, \mathrm{d}W_s,$$

$$Z_t = \nabla u(t, X_t).$$

Deep Backward SDE

Nonlinear Feynman–Kac:

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

where $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$, with $n = 0, \ldots, N-1$, satisfies

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

$$\mathcal{Y}_{k,n+1} = u(t_k, \mathcal{X}_{k,0}) - \sum_{\ell=0}^{n} f\big(\mathcal{X}_{k,\ell}, \mathcal{Y}_{k,\ell}, \mathcal{Z}_{k,\ell}\big)\Delta t + \sum_{\ell=0}^{n} \mathcal{Z}_{k,\ell}^{\top}(W_{t_{k,\ell+1}} - W_{t_{k,\ell}})$$

$$\mathcal{Z}_{k,\ell} = \nabla u(t_{k,\ell}, \mathcal{X}_{k,\ell}).$$

Deep Backward SDE

Nonlinear Feynman–Kac:

$$p_k \approx \underset{u \in C([t_k, t_{k+1}] \times \mathbb{R}^d \times \mathbb{R}^{d' \times k}; \mathbb{R})}{\arg\min} \mathbb{E}\left[\left| \mathcal{Y}_{k,N} - g_k(\mathcal{X}_{k,N}, O_{1:k}) \right|^2 \right]$$

where $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$, with $n = 0, \ldots, N-1$, satisfies

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

$$\mathcal{Y}_{k,n+1} = u(t_k, \mathcal{X}_{k,0}) - \sum_{\ell=0}^{n} f\left(\mathcal{X}_{k,\ell}, \mathcal{Y}_{k,\ell}, \mathcal{Z}_{k,\ell}\right)\Delta t + \sum_{\ell=0}^{n} \mathcal{Z}_{k,\ell}^{\top}(W_{t_{k,\ell+1}} - W_{t_{k,\ell}})$$

$$\mathcal{Z}_{k,\ell} = \nabla u(t_{k,\ell}, \mathcal{X}_{k,\ell}).$$

Deep Backward SDE

Nonlinear Feynman–Kac:

$$\pi_k = \mathop{\arg\min}_{\substack{w \in C(\mathbb{R}^d \times \mathbb{R}^{d' \times k}; \mathbb{R}) \\ (v_n)_{n=0}^{N-1} \in C(\mathbb{R}^d \times \mathbb{R}^{d' \times k}; \mathbb{R}^d)^N}} \mathbb{E}\left[\left|\mathcal{Y}_{k,N} - g_k(\mathcal{X}_{k,N}, O_{1:k})\right|^2\right]$$

where $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$, with $n = 0, \ldots, N-1$, satisfies

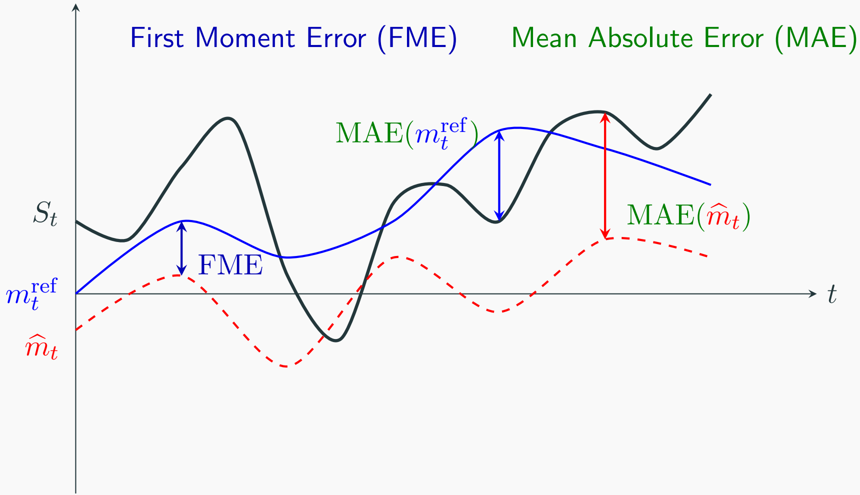
$$\mathcal{X}_{k,n+1} = \mathcal{X}_{k,n} + \mu(\mathcal{X}_{k,n})(t_{k,n+1} - t_{k,n}) + \sigma(\mathcal{X}_n)(W_{t_{k,n+1}} - W_{t_{k,n}})$$

$$\mathcal{Y}_{k,n+1} = w(\mathcal{X}_{k,0}) - \sum_{\ell=0}^{n} f\big(\mathcal{X}_{k,\ell}, \mathcal{Y}_{k,\ell}, \mathcal{Z}_{k,\ell}\big)\Delta t + \sum_{\ell=0}^{n} \mathcal{Z}_{k,\ell}^\top (W_{t_{k,\ell+1}} - W_{t_{k,\ell}})$$

$$\mathcal{Z}_{k,\ell} = v_\ell(\mathcal{X}_{k,\ell}).$$

Numerical experiments

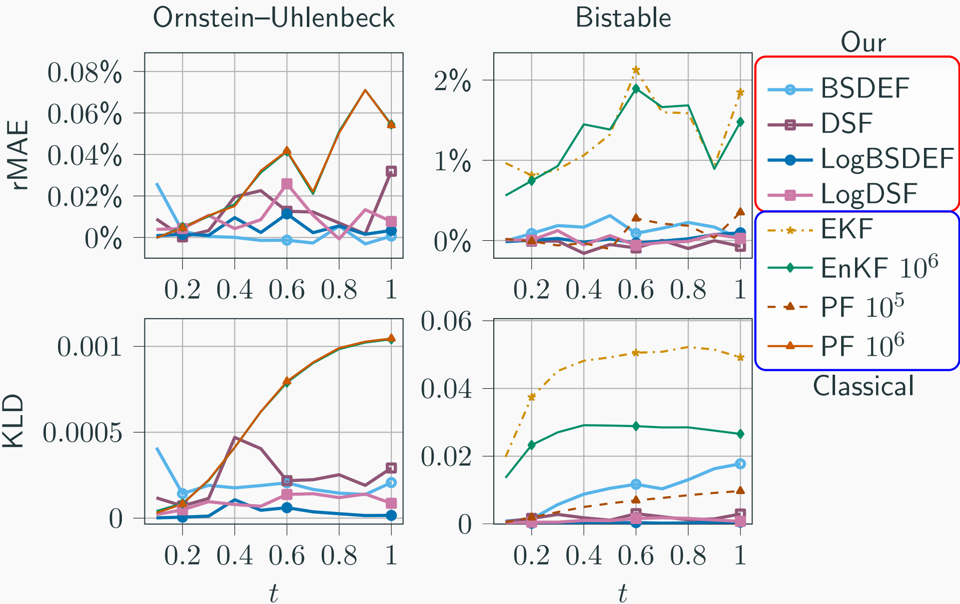
Moment errors



Tons of abbreviations ending with "F"

- BSDEF - deep Backward SDE Filter
- DSF - Deep Splitting Filter
- LogBSDEF and LogDSF - log versions
- EKF - Extended Kalman Filter
- EnKF - Ensemble Kalman Filter
- PF - bootstrap Particle Filter

Toy example



Ornstein–Uhlenbeck
Bistable
rMAE
KLD
0.08%
0.06%
0.04%
0.02%
0%
2%
1%
0.001
0.0005
0
0.06
0.04
0.02
0.2 0.4 0.6 0.8 1
t
Our
BSDEF
DSF
LogBSDEF
LogDSF
EKF
EnKF 10^6
PF 10^5
PF 10^6
Classical

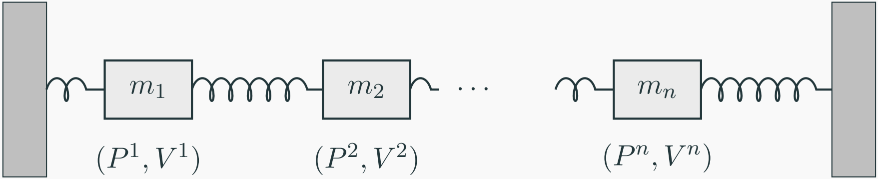
Linear spring-mass - model

Let the state S be split into position P and velocity V so that

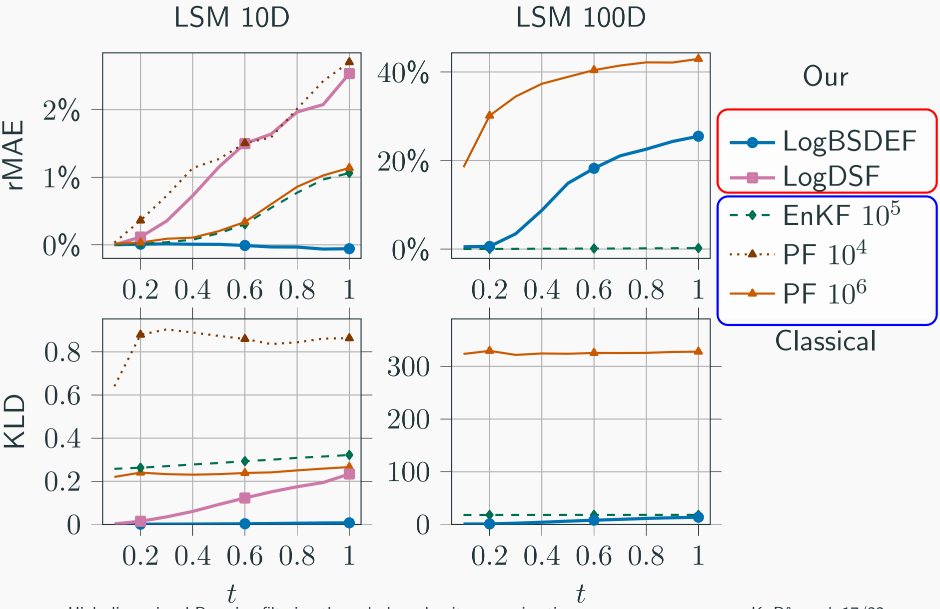
$$P_t = P_0 + \int_0^t V_s \, \mathrm{d}s + \int_0^t \sigma \, \mathrm{d}B_s^{(1)}$$
$$V_t = V_0 + \int_0^t (A_{21} P_s + A_{22} V_s) \, \mathrm{d}s + \int_0^t \sigma \, \mathrm{d}B_s^{(2)}$$

where A_{21} and A_{22} are stiffness and damping matrices with uniform sampled parameters, and we have partial observations:

$$O_k = P_{t_k} + \eta_k, \quad \eta_k \sim N(0, R)$$



Linear spring-mass - results



Lorenz-96 - model

$$\frac{\mathrm{d}x_i(t)}{\mathrm{d}t} = (x_{i+1}(t) - x_{i-2}(t))x_{i-1}(t) - x_i(t) + F, \quad i = 1, \ldots, d$$

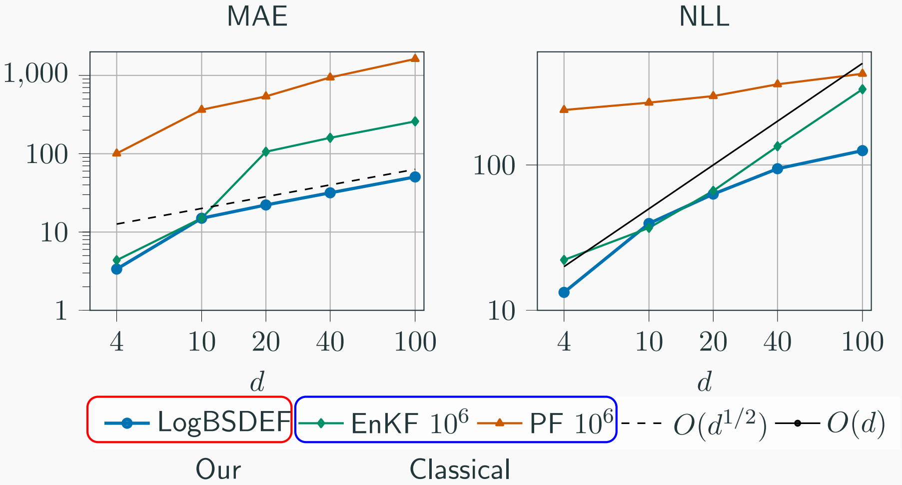
where $x_0 = x_d$ and $x_{-1} = x_{d-1}$

We consider an SDE extension with additive noise given by $\sigma(x) = \sigma I$

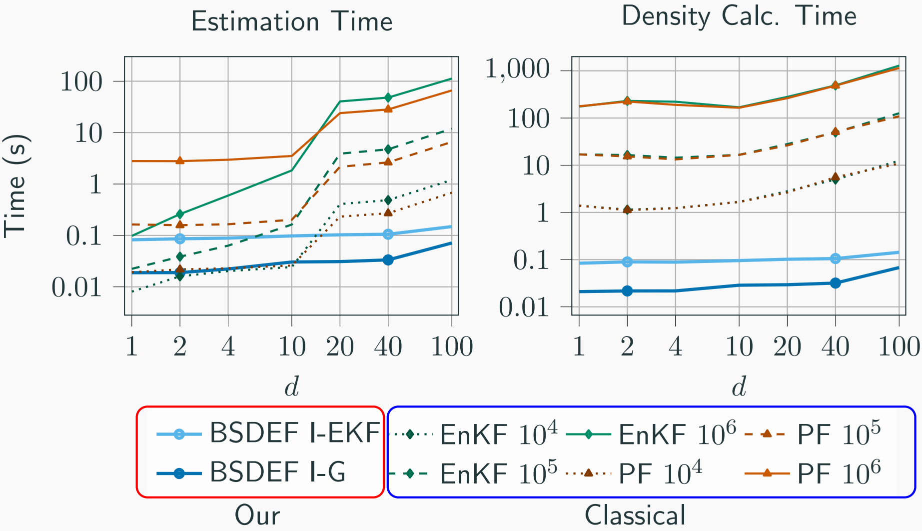
Let $d = [4, 10, 20, 40, 100]$ and consider partial observations in $d' = [4, 5, 5, 10, 25]$-dimensions through the measurement function

$$\big(h(x)\big)_i = \begin{cases} x_i, & (d, d') = (4, 4), \\ x_{2i}, & (d, d') = (10, 5), \\ x_{4i}, & (d, d') = (20, 5),\ (40, 10),\ (100, 25). \end{cases} \tag{1}$$

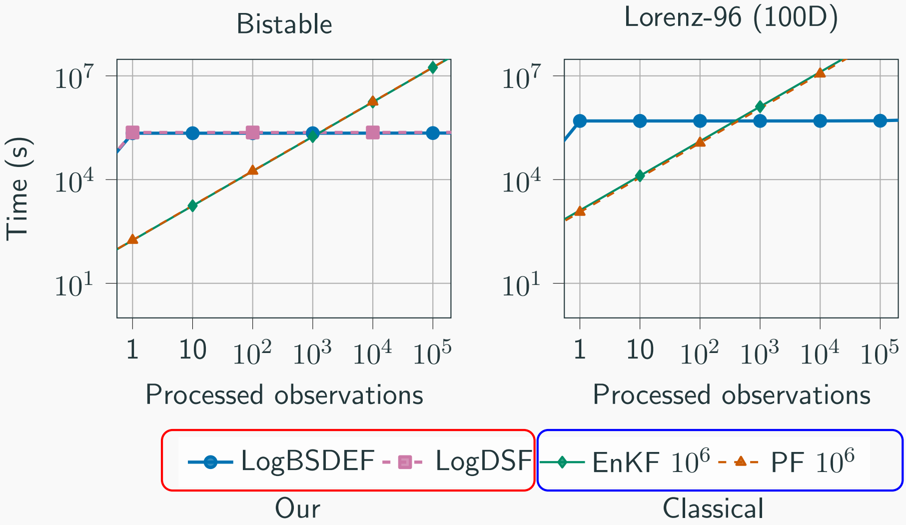
Lorenz-96 - results



Computational inference time



Trade-off including training time



What did we learn?

- The deep BSDE approach outperforms the splitting method
- Solving the log-density Fokker–Planck improves performance
- Satisfactory performance for nonlinear and high-dimensional problems without curse of dimensionality

Outlook:

- Parameter inference, introduce a probability measure over a set Θ of parameters modelling μ, σ, h, p_0
- Explore other architectures

See the paper