Scalable Inference in SDEs by Direct Matching of the Fokker–Planck–Kolmogorov Equation

  

Arno Solin Ella Tamir Prakhar Verma

Aalto University

NEURAL INFORMATION PROCESSING SYSTEMS

TL;DR

- Simulation-based techniques for solving stochastic differential equations (SDEs) are the *de facto* approach for inference in the machine learning community
- We advocate the use of alternative methods for solving SDEs by approximating the typically intractable Fokker–Planck–Kolmogorov equation
- We revisit classical SDE theory and directly match the moments of weak solutions, allowing us to forego sampling in lieu of more scalable approaches
- This workflow is fast, scales to high-dimensional latent spaces, and is applicable to scarce-data applications
- We demonstrate the methodology on general SDE problems and GP-SDE models, where a GP encodes prior knowledge into the SDE dynamics

Gaussian Process SDEs

- We are concerned with continuous-time dynamical modelling in machine learning, typically in the *latent* space of models
- Consider an ODE model of some latent state $\mathbf{z}(t)$ defined as

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{z}(t) = \mathbf{v}_\theta(\mathbf{z}(t), t), \qquad (1)$$

  where $\mathbf{v}_\theta(\mathbf{z}(t), t)$ is a velocity field
- Instead of a deterministic field, as introduced in [1], we set the prior to be a Gaussian process

$$\mathbf{v}(\mathbf{z}, t) \sim \mathrm{GP}(\boldsymbol{\mu}(\mathbf{z}), \boldsymbol{\kappa}(\mathbf{z}, \mathbf{z}')), \qquad (2)$$

  where μ is the GP mean and κ the kernel.
- Instead of the random ODE formulation defined by Eq. (1) and (2), we write the model as an Itô SDE matching the GP

$$\mathrm{d}\mathbf{z}(t) = \mathbf{f}(\mathbf{z}, t)\,\mathrm{d}t + \mathbf{L}(\mathbf{z}, t)\,\mathrm{d}\boldsymbol{\beta}(t)\,. \qquad (3)$$

- The GP-SDE above has its drift $\mathbf{f}(\cdot,\cdot)$ set as the GP mean, and diffusion $\mathbf{L}(\cdot,\cdot)$ as a square-root factor of the Gaussian covariance given by the GP model
- Allows for encoding prior knowledge into the dynamics, such as curl-freeness or divergence-freeness in $\mathbf{v}_\theta(\mathbf{z}(t), t)$

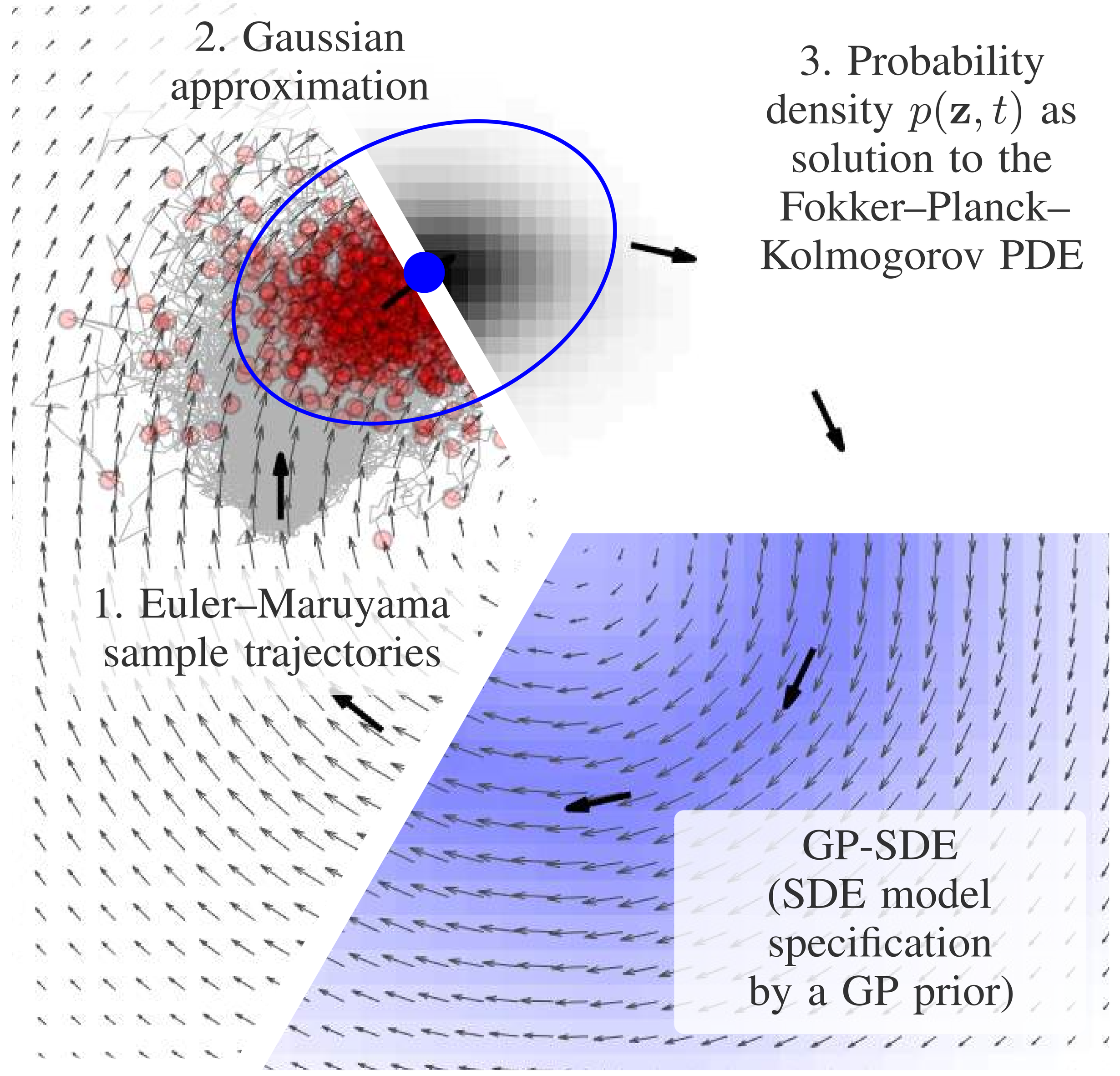

Fig. 1: Views into solutions to SDEs: simulation-based solutions, the FPK solution, and a Gaussian approximation for a GP-SDE model conditioned on the arrow observations.

Matching Moments of the Fokker–Planck–Kolmogorov (FPK) Equation

- The FPK PDE gives the weak evolution of the SDE (3), describing the development of the marginal density $p(\mathbf{z}, t)$:

$$\frac{\partial p(\mathbf{z},t)}{\partial t} = -\sum_i \frac{\partial}{\partial z_i}[f_i(\mathbf{z}, t)\, p(\mathbf{z}, t)] + \frac{1}{2}\sum_{i,j} \frac{\partial^2}{\partial z_i \partial z_j}\{[\mathbf{L}(\mathbf{z}, t)\,\mathbf{Q}\,\mathbf{L}^\top(\mathbf{z}, t)]_{ij}\, p(\mathbf{z}, t)\}\,.$$

- The PDE is typically intractable, but assuming $p(\mathbf{z}, t) \approx \mathrm{N}(\mathbf{m}(t), \mathbf{P}(t))$ is Gaussian we write down the ODE describing the evolution of the moments:

$$\frac{\mathrm{d}\mathbf{m}}{\mathrm{d}t} = \int \mathbf{f}(\mathbf{z}, t)\,\mathrm{N}(\mathbf{z} \mid \mathbf{m}, \mathbf{P})\,\mathrm{d}\mathbf{z} \quad \text{and}$$

$$\frac{\mathrm{d}\mathbf{P}}{\mathrm{d}t} = \int \mathbf{f}(\mathbf{z}, t)\,(\mathbf{z} - \mathbf{m})^\top\,\mathrm{N}(\mathbf{z} \mid \mathbf{m}, \mathbf{P})\,\mathrm{d}\mathbf{z} + \int (\mathbf{z} - \mathbf{m})\,\mathbf{f}^\top(\mathbf{z}, t)\,\mathrm{N}(\mathbf{z} \mid \mathbf{m}, \mathbf{P})\,\mathrm{d}\mathbf{z} + \int \mathbf{L}(\mathbf{z}, t)\,\mathbf{Q}\,\mathbf{L}^\top(\mathbf{z}, t)\,\mathrm{N}(\mathbf{z} \mid \mathbf{m}, \mathbf{P})\,\mathrm{d}\mathbf{z}$$

- The integrals above are not tractable: further approximation is required, such as linearization or Gaussian quadrature methods such as the 3rd order cubature (see [2])

Reduced Computational Cost

- One of the key advantages of weak solution concepts are lower computational costs, especially in high-dimensional problems
- Faithfully representing the underlying distribution through sampling methods, such as Euler–Maruyama, often requires multiple trajectories.
- In contrast, a single step in the linearization moment ODEs can be completed with $O(1)$ evaluations of the drift, diffusion and the Jacobian, and 3rd order cubature evaluates drift and diffusion $O(d)$ times
- We empirically test the runtime of approximations: see plots below for wall-clock timing results for a multi-dimensional Beneš SDE, with a setting where the number of E-M trajectories is selected to match the accuracy of the approximations for comparability

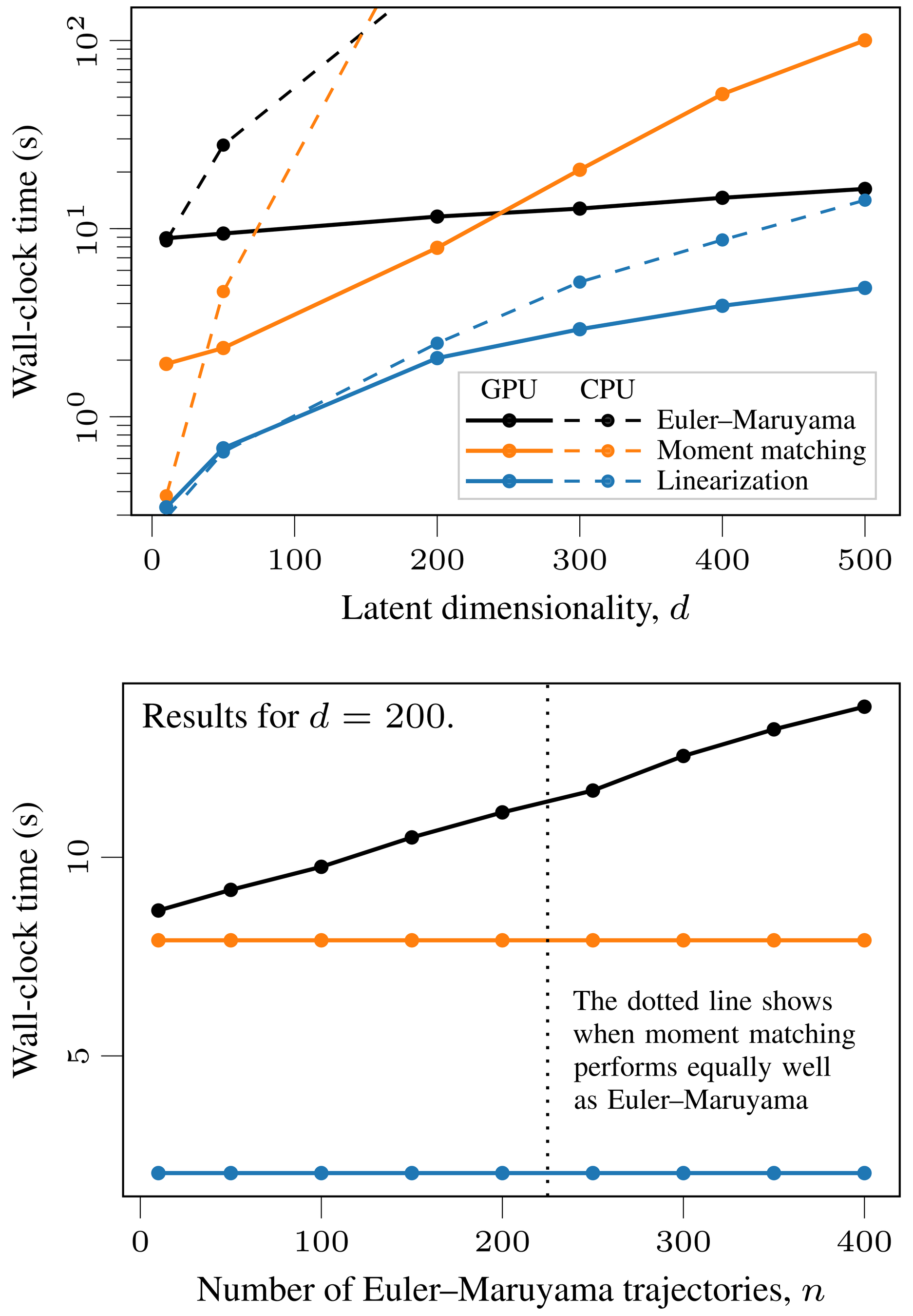

Fig. 2: Empirical timing experiments with error of final margins matched.

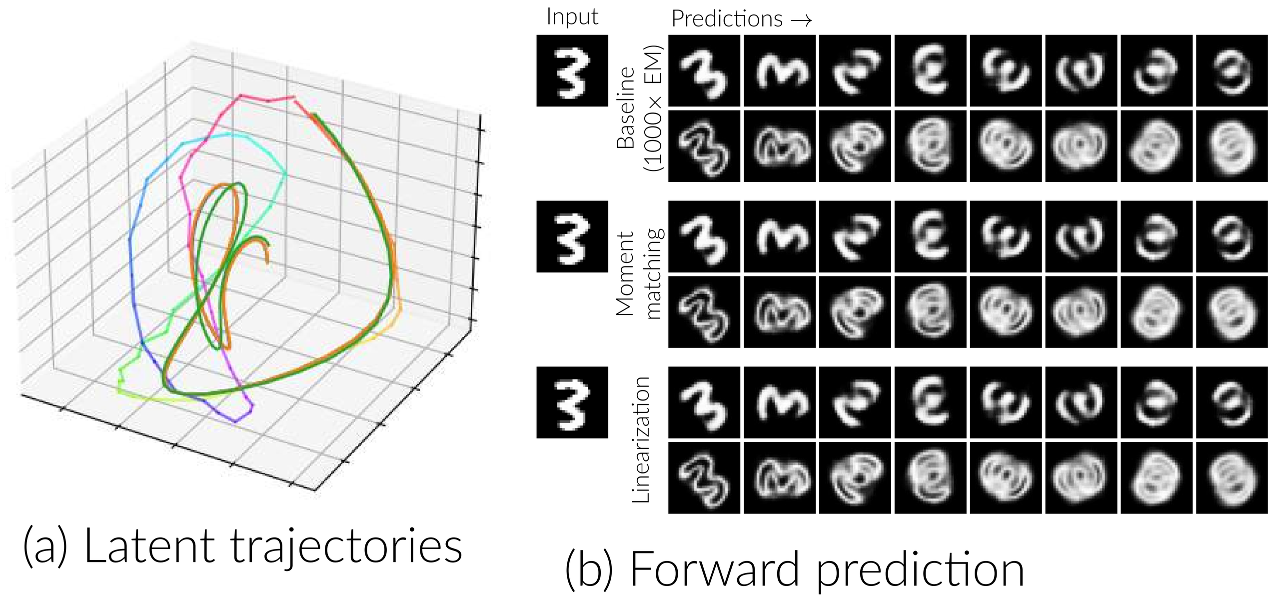

Fig. 3: Results on rotating MNIST. In (a), true and approximated latent trajectories, and in (b), progression of the rotating MNIST prediction at varying angles.

Outlook

- We highlight the usefulness of weak solutions in rotating MNIST and motion capture examples
- For the MNIST example, we encoded the observations using a VAE, designed as in [3], further lowering the computational cost related to solving the SDEs defined by the GP-SDE model
- In Fig. 3, both moment matching methods (quadrature) and linearization are able to produce a faithful representation of the distribution defined by Euler–Maruyama sampling with multiple trajectories
- In the paper, we include results for a motion capture example that demonstrates that weak solutions can perform close to state-of-the-art, while being considerably more efficient

References

[1] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, "Neural ordinary differential equations," in *Advances in Neural Information Processing Systems 31 (NeurIPS)*, pp. 6571–6583, Curran Associates, Inc., 2018.

[2] S. Särkkä and A. Solin, *Applied Stochastic Differential Equations*. Cambridge University Press, 2019.

[3] C. Yıldız, M. Heinonen, and H. Lahdesmaki, "ODE²VAE: Deep generative second order ODEs with Bayesian neural networks," in *Advances in Neural Information Processing Systems 32 (NeurIPS)*, pp. 13412–13421, Curran Associates, Inc., 2019.

Code and resources available:
https://github.com/AaltoML/scalable-inference-in-sdes

arno.solin@aalto.fi