Adversarial Sequential Monte Carlo

Kira Kempinska and John Shawe-Taylor

UCL

Abstract

How can we perform efficient inference in directed probabilistic models with intractable posterior distributions? We introduce a new technique for improving finite-sample inference approximations by learning highly flexible proposal distributions for sequential importance samplers, such as particle filters. We represent proposal distributions as implicit generative models, that is models that you can sample from but which do not have an explicit parametric form, and train them using variational inference rephrased as a two-player game, hence establishing a principled connection between Sequential Monte Carlo (SMC) and Generative Adversarial Networks (GANs). Our approach achieves state-of-the-art performance on a range of inference problems.

Sequential Monte Carlo

- Consider a directed graphical model comprising of N latent variables x and N observed variables y, whose joint distribution factorises as

$$p(x_{1:N}, y_{1:N}) = p(x_1)p(y_1|x_1)\prod_{n=2}^{N} p(x_n|x_{1:n-1})p(y_n|x_{1:n}, y_{1:n-1})$$

This form subsumes common state-space models, such as Hidden Markov Models (HMMs), as well as non-Markovian models, such as Gaussian processes.

- We are interested in the intractable posterior distribution of the unobserved signal $p(x|y) \equiv p(x_{1:N}, y_{1:N})$.
- Sequential Monte Carlo (SMC) approximates $p(x|y)$ with a weighted set of samples drawn from a presumably simpler proposal distribution $q(x|y)$.
- SMC performs the approximation in a sequential fashion using a single forward pass (*filtering*).
- **Initially designed for online inference in temporal state-space models, SMC methods are increasingly used to perform inference in a wide range of graphical models.**
- **SMC methods are scalable, easy to implement and flexible, but their performance crucially depends on the quality of the proposal distribution, i.e. how well it matches the target posterior.**

Case Study 1: polynomial regression

Consider a non-conjugate polynomial regression model proposed by [5], with global-only latent variables:

$$p(w_d) = Laplace(0, 10^{1-d}) \qquad d = 0,1,2$$
$$p(t_n|w_0, w_1, w_2, z_n) = t_v(w_0 + w_1 z_n + w_2 z_n^2, \epsilon^2) \qquad n = 1, \ldots, N$$

for fixed $v = 4$, $\epsilon = 1$ and $z_n \in (-10, 10)$ uniformly. Our goal is to infer latent variables $x \equiv \{w_0, w_1, w_2\}$ given observed variables $y \equiv \{z_n, y_n\}_{n=1}^{N}$.

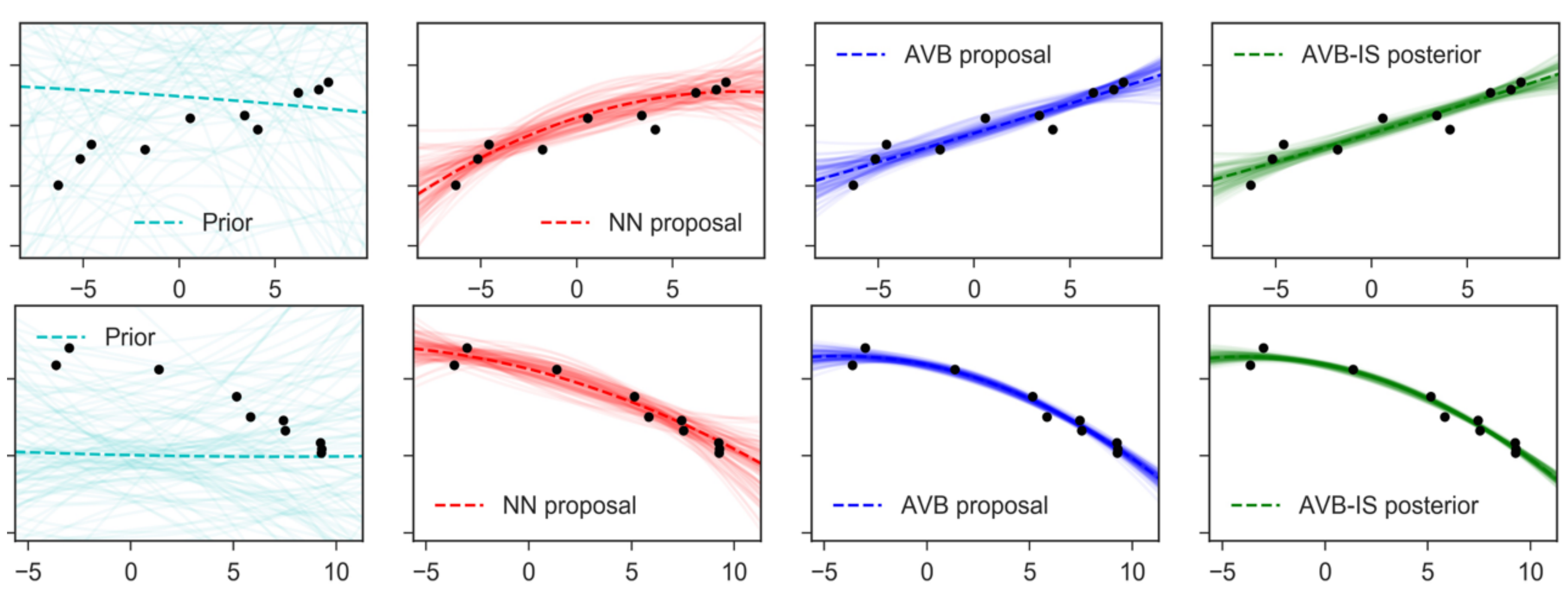


Figure 1: Exemplary output in the polynomial regression case study. Plots show 100 regression curves, each curve estimated from one sample of weights, and the mean curve marked as a dashed line. Our proposal (AVB) yields estimates close to proposal by [5] (NN), but slightly less diffused. Black dots represent true observations.

Our Method

- We train highly flexible proposal distributions for SMC by representing them as implicit models, that is, models that we can sample from but whose density we cannot evaluate.
- Implicit proposal models use an input noise ϵ and a deep network architecture parametrised by ϕ to flexibly characterise a wide range of proposal distributions.
- We find optimal proposal parameters phi by minimising KL divergence between the proposal $q_\phi(x|y)$ and the target posterior $p(x|y)$.
- The implicit formulation of the proposal requires adversarial training to perform the optimisation [4].
- We perform the training off-line and then use the trained proposals q_ϕ within an SMC algorithm called particle filters [3].

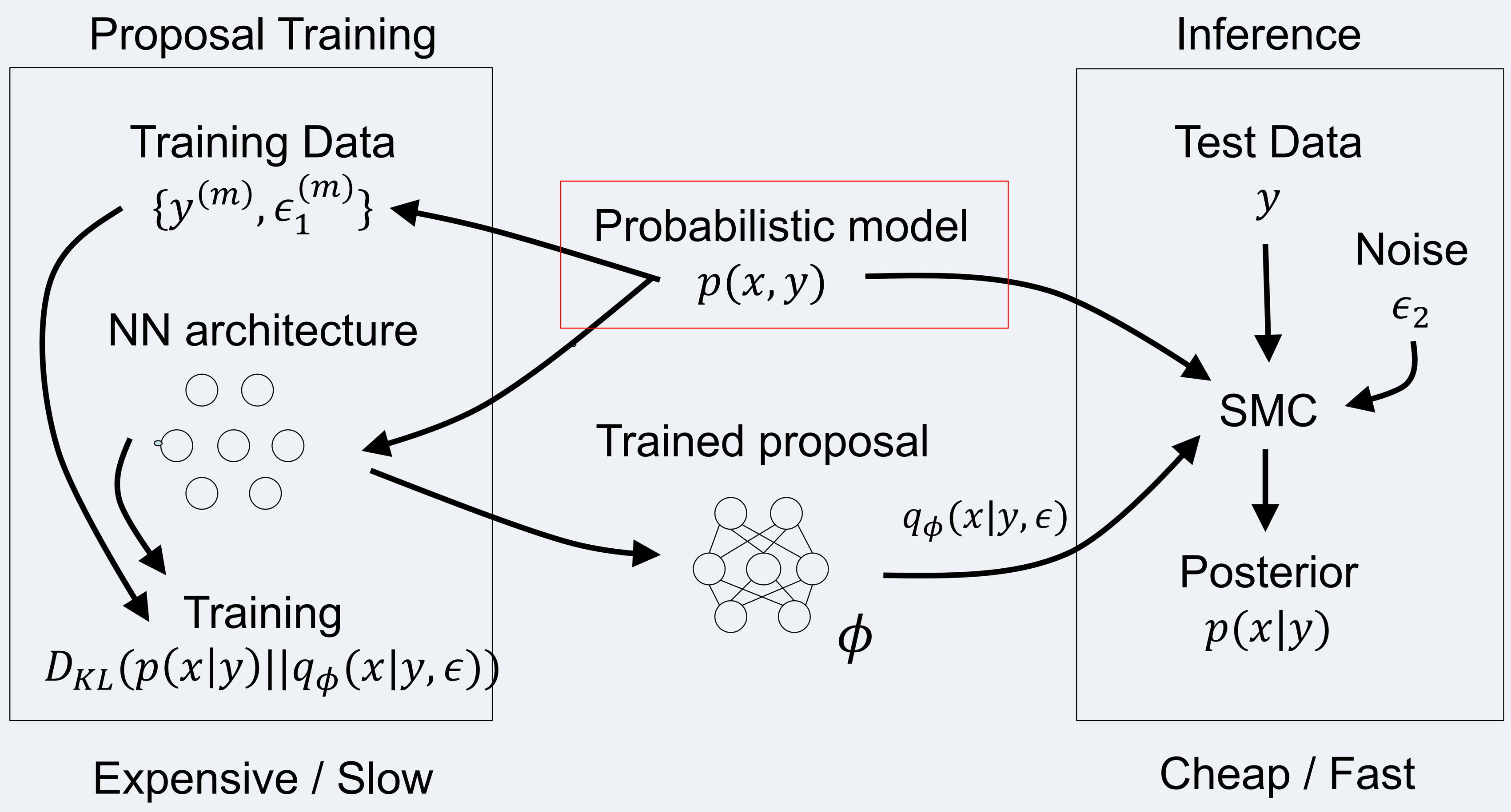


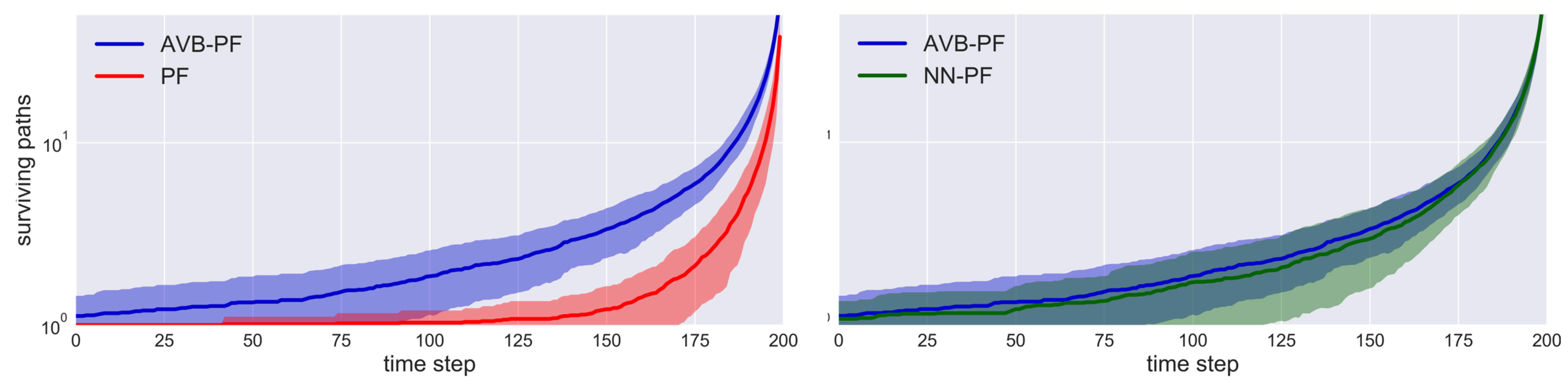
Figure 2: Schematic overview of our method.

Case Study 2: benchmark non-linear state-space model

We now consider a more complex temporal model given by the following dynamics:

$$p(x_n|x_{n-1}) = \mathbb{N}(x_n; f(x_{n-1}, \sigma_v^2), \qquad p(x_1) = \mathbb{N}(x_1; 0,5)$$
$$p(y_n|x_n) = \mathbb{N}(y_n; g(x_n), \sigma_w^2)$$
$$f(x_{n-1}) = x_{n-1}/2 + 25x_{n-1}/(1 + x_{n-1}^2), \qquad g(x_n) = x_n^2/20$$

for fixed $\theta = (\sigma_v, \sigma_w) = (\sqrt{10}, 1)$. The nonlinear model is often used to assess the performance of SMC methods [2, 1]. Its posterior density $p_\theta(x_{1:N}, y_{1:N})$ is highly multimodal due uncertainty about the sign of the state x_n which is only observed through its square. Our goal is to infer a sequence of hidden states $x_{1:N}$ given a sequence of observations $y_{1:N}$.



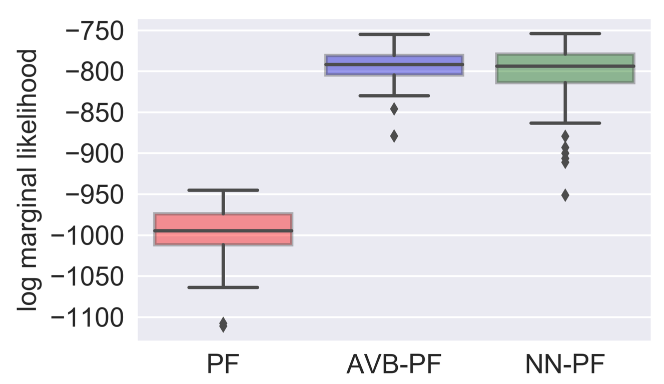


Figure 3: Adversarially learnt proposals reduce particle degeneracy and explain data better. Plots show mean and standard deviation of the number of unique particles (top) and log-marginal likelihood (bottom) over 100 sampled sequences of length $N = 200$, using 100 particles.

References

[1] Christophe Andrieu, Arnaud Doucet, and Roman Holenstein. Particle markov chain monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(3):269–342, 2010.

[2] Arnaud Doucet, Nando De Freitas, and NJ Gordon. Sequential monte carlo methods in practice. series statistics for engineering and information science, 2001.

[3] Arnaud Doucet and Adam M Johansen. A tutorial on particle filtering and smoothing: Fifteen years later. *Handbook of nonlinear filtering*, 12(656-704):3, 2009.

[4] Lars Mescheder, Sebastian Nowozin, and Andreas Geiger. Adversarial variational bayes: Unifying variational autoencoders and generative adversarial networks. *arXiv preprint arXiv:1701.04722*, 2017.

[5] Brooks Paige and Frank Wood. Inference networks for sequential monte carlo in graphical models. In *International Conference on Machine Learning*, pages 3040–3049, 2016.