Comparison of variational Bayes and classical SDE approximations for Bayesian smoothing in continuous-time stochastic systems with discrete-time measurements

Juha Ala-Luhtala[1] **and Simo Särkkä**[2]
[1] Tampere University of Technology, Finland
[2] Aalto University, Finland

In this work, we study approximate Bayesian smoothing for the following continuous-time stochastic system with discrete-time noisy measurements

$$
\begin{aligned}
dx &= f(x,t)\,dt + L(t)d\beta(t) \\
y_k &= h(x(t_k)) + v_k,
\end{aligned}
$$

where x is the state, f is the drift or dynamic function, L is the dispersion matrix and $\beta(t)$ is a Brownian motion stochastic process. The measurements y_k are taken at discrete-time instants t_k. The measurement noise $\{v_k\}$ is a zero-mean Gaussian white noise stochastic process.

Given a set of measurements $\{y_1,\ldots,y_K\}$, the solution to the Bayesian smoothing problem is the conditional distribution $p(x,t\,|\,y_1,\ldots,y_K)$. The exact computation of the smoothing distribution involves solving the Fokker-Planck-Kolmogorov (FPK) partial differential equation, which is intractable for most nonlinear models (Särkkä, 2006). In this work, we consider approximate smoothing algorithms that result in a Gaussian approximation for the smoothing distribution:

$$
p(x,t\,|\,y_1,\ldots,y_K) \approx N(x(t)\,|\,m(t),P(t)).
$$

The smoothing problem now reduces to computing the mean and covariance matrix of the Gaussian approximation. Two approaches found from the literature are considered: classical Gaussian approximation (Särkkä and Sarmavuori, 2013) and variational Gaussian approximation (Archambeau *et al.*, 2007).

The first approach is called classical approach in this work, since it is based on the classical Gaussian approximation for the filtering distribution (Jazwinski, 1970; Ito and Xiong, 2000). Särkkä and Sarmavuori (2013) derive two Gaussian smoothing algorithms, called type I and type II Gaussian smoothers, using the Gaussian approximation for the filtering distribution. For both methods, the Gaussian approximation is found by first solving the forward differential equations for the mean and covariance of the filtering distribution, and then using these results to solve the backward differential equations for the smoothed mean and covariance.

The second approach, called variational Gaussian smoother, was derived by Archambeau *et al.* (2007). It is based on minimizing the Kullback-Leibler divergence between the approximating Gaussian process and the true smoothing process. This results in a boundary value problem that can be solved using an iterative fixed-point algorithm.

Implementation of the smoothers relies on a method to compute Gaussian expectations over arbitrary functions. These methods are well known in the literature and include the Taylor series based linearization methods, Cubature rule and Unscented transform based sigma-point methods and Gauss-Hermite quadrature based methods. For this work we have chosen to use the Gauss-Hermite quadrature since it can be made exact for monomials up to arbitrary fixed degree (Särkkä and Sarmavuori, 2013).

The smoothing methods are compared in terms of computational properties and quality of the resulting approximation for the double-well stochastic system described by

$$\begin{aligned} dx &= 4x(1 - x^2)\, dt + \sigma\, d\beta, \\ y_k &= x(t_k) + v_k. \end{aligned}$$

The double-well system is highly nonlinear and its stationary distribution has two modes at $x = \pm 1$. With sufficiently large value for process noise parameter σ, there is frequent jumping between the two modes. For this one dimensional example a relatively accurate reference smoothing solution can be obtained by solving the exact Bayesian smoothing equations on a dense grid using a finite difference approximation. The type I and II smoothers are computationally less complex than the variational smoother, since each iteration of the variational smoother requires approximately the same amount of computations as the two passes needed in type I and II smoothers. However, the extra effort results in better approximation of the reference smoothing solution, especially in estimating the jumps between the two modes.

Keywords: Bayesian smoothing, stochastic differential equation, Gaussian approximation, variational Bayes

References:

Archambeau, C., Cornford, D., Opper, M., Shawe-Taylor, J. (2007). Gaussian Process Approximations of Stochastic Differential Equation, *JMLR Workshop & Conference Proceedings*, 1:1-16

Ito, K., Xiong, K. (2000). Gaussian filters for nonlinear filtering problems. *IEEE Transactions on Automatic Control*, 45(5):910-927

Jazwinski, A.H. (1970). *Stochastic processes and filtering theory.* Dover

Särkkä, S., Sarmavuori, J. (2013). Gaussian filtering and smoothing for continuous-discrete dynamic systems, *Signal Processing*, 93(2):500-510

Särkkä S. (2006). *Recursive Bayesian Inference on Stochastic Differential Equations.* Doctoral dissertation, Helsinki University of Technology