Variational Inference and Learning for Continuous-Time Nonlinear State-Space Models

Antti Honkela, Markus Harva, Tapani Raiko, Juha Karhunen
Helsinki University of Technology
http://www.cis.hut.fi/projects/bayes/

Inference in continuous-time stochastic dynamical models is a challenging problem. To complement existing sampling-based methods [2], variational methods have recently been developed for this problem [1].

Our approach, which was first introduced in [3], solves the variational continuous-time inference problem by discretisation that essentially reduces it to a discrete-time problem previously considered in [8]. While this approach is not as elegant as that of [1], our framework makes learning the model in addition to inference easy. Other extensions such as heteroscedastic models are also relatively easy to consider within this framework.

The discrete-time model in [8] is based on using multi-layer perceptron (MLP) networks to model the nonlinearities. While it may be difficult to use them to specify complex prior information, their functional form seems quite reasonable in many computational biology applications [6], for instance.

The discrete-time state-space model studied in [8] assumes that the observations $\mathbf{x}(t)$ are generated by

$$\mathbf{s}(t+1) = \mathbf{s}(t) + \mathbf{g}_{dt}(\mathbf{s}(t), \boldsymbol{\theta}_\mathbf{g}) + \mathbf{m}(t) \tag{1}$$

$$\mathbf{x}(t) = \mathbf{f}(\mathbf{s}(t), \boldsymbol{\theta}_\mathbf{f}) + \mathbf{n}(t) \tag{2}$$

with states $\mathbf{s}(t)$, Gaussian innovation $\mathbf{m}$ and noise $\mathbf{n}$, and MLP networks to model the nonlinearities $\mathbf{f}$ and $\mathbf{g}_{dt}$. Inference and learning in the model can be made more reliable and efficient than in [8] by using new linearisation [5] and state inference techniques [4].

The general formulation of the state evolution in a continuous-time nonlinear state-space model is given by a stochastic differential equation (SDE)

$$d\mathbf{s} = \mathbf{g}(\mathbf{s})dt + \sqrt{\boldsymbol{\Sigma}}d\mathbf{W}, \tag{3}$$

where $d\mathbf{W}$ is the differential of a Wiener process [7]. The model by Valpola and Karhunen can be turned into a continuous-time nonlinear state-space model by using Eq. (3) to model state evolution instead of Eq. (1). The nonlinear mapping $\mathbf{g}$ is again modelled by a MLP network. The observation equation Eq. (2) remains unchanged.

We use variational Bayes (VB) to approximate the otherwise intractable posterior appearing in the problem. In particular, we find a Gaussian approximation q of all the parameters and the states at observation times, similar to that in [8], by optimising the VB free energy using conjugate gradient.

The continuous-time NSSM is demonstrated with a data set generated by a Lorenz process. A Lorenz process has a three-dimensional state-space with non-linear chaotic dynamics. The data set was generated by drawing 1000 samples uniformly at random time instants between 0 and 50 (Figure 1, left panel). Additive Gaussian observation noise was added (Figure 1, middle panel). To make learning more challenging and to demonstrate the benefits of the latent state-space, only two components of the observations were used in this experiment. A three dimensional state-space was used to learn this data set. The MLPs for both the observation and the dynamical mapping had 10 hidden units. In the right panel of Figure 1, we show $\langle \mathbf{f}(\mathbf{s}(t), \boldsymbol{\theta}_\mathbf{f}) \rangle$ after 1050 iterations (which took approximately one day). Given that the observations are rather noisy, the method has been able to learn the underlying process quite well.

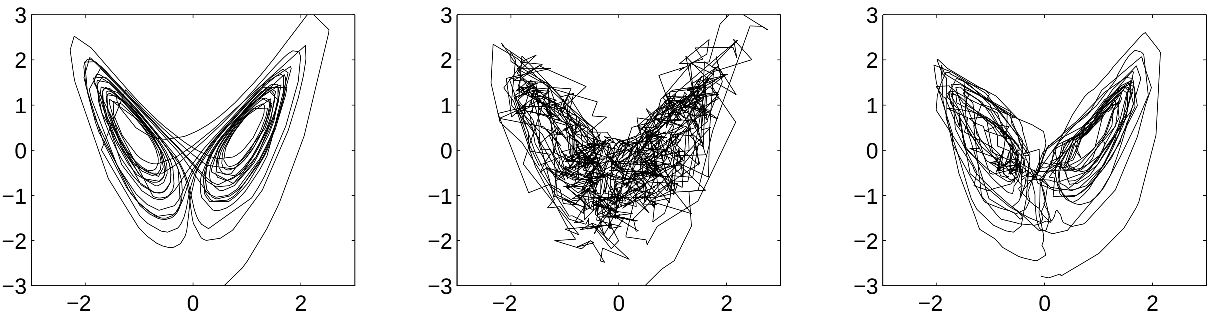

Figure 1: Left: The original data set without noise. Middle: The noisy data set used in the experiment. Right: The reconstruction of the data set by the model.

We have considered several variations to the basic setup as well. In particular, we intend to incorporate heteroscedastic priors to the model. In the discrete-time case we already have some initial results with a model where the variance of the observation noise is conditioned on the states. Some of the states can thus be associated with the dynamics of the variance which is a reasonable assumption in certain applications such as financial time series. Putting a heteroscedastic prior on the innovation process is another interesting variation of the basic model. Bringing these extensions from discrete-time to continuos-time should not pose tremendous difficulties and we hope to implement these in the future.

References

[1] C. Archambeau, M. Opper, Y. Shen, D. Cornford, and J. Shawe-Taylor. Variational inference for diffusion processes. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 17–24. MIT Press, Cambridge, MA, 2008.

[2] A. Golightly and D. J. Wilkinson. Bayesian inference for nonlinear multivariate diffusion models observed with error. *Computational Statistics and Data Analysis*, 52(3):1674–1693, 2008.

[3] A. Honkela, M. Tornio, and T. Raiko. Variational Bayes for continuous-time nonlinear state-space models. In *NIPS*2006 Workshop on Dynamical Systems, Stochastic Processes and Bayesian Inference*, Whistler, B.C., Canada, 2006.

[4] A. Honkela, M. Tornio, T. Raiko, and J. Karhunen. Natural conjugate gradient in variational inference. In *Proceedings of the 14th International Conference on Neural Information Processing (ICONIP 2007)*, Kitakyushu, Japan, November 2007.

[5] A. Honkela and H. Valpola. Unsupervised variational Bayesian learning of nonlinear models. In L. Saul, Y. Weiss, and L. Bottou, editors, *Advances in Neural Information Processing Systems 17*, pages 593–600. MIT Press, Cambridge, MA, USA, 2005.

[6] E. Mjolsness. On cooperative quasi-equilibrium models of transcriptional regulation. *Journal of Bioinformatics and Computational Biology*, 5(2b):467–490, 2007.

[7] B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications.* Springer, Berlin, fifth edition, 2000.

[8] H. Valpola and J. Karhunen. An unsupervised ensemble learning method for nonlinear dynamic state-space models. *Neural Computation*, 14(11):2647–2692, 2002.