Parameter Estimation in Dynamical Models Driven by Noise

MICHAIL D. VRETTAS, DAN CORNFORD AND YUAN SHEN

Neural Computing Research Group, School of Engineering and Applied Science, Aston University, Birmingham, UK
email: {vrettasm, d.cornford, y.shen2}@aston.ac.uk

Abstract

Estimating parameters in dynamical systems has many applications, yet remains a very challenging and relatively unexplored task. The problem becomes even more difficult if the system under consideration is non-linear and large in size. Recent attempts have been based mainly on variations of Monte Carlo methods or extensions to the well known Kalman filter. In this poster we present the results of estimating parameters in stochastic dynamical models with a variational approximation method, which is currently under development within the VISDEM[a] project.

Introduction

Motivated by numerical weather prediction models [3], within VISDEM, we are seeking a variational Bayesian treatment of the dynamic data assimilation problem (see Figure:1).

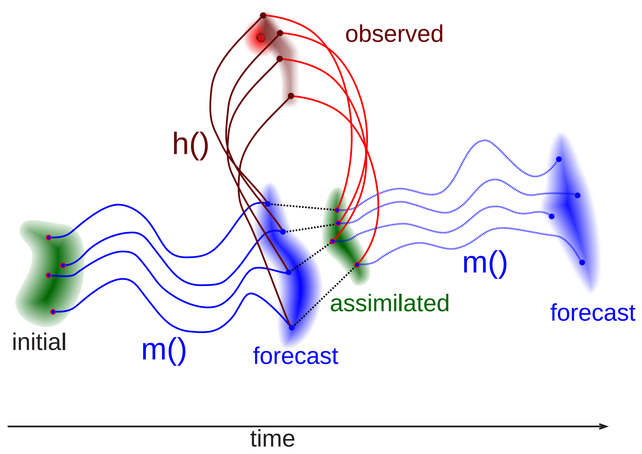

Figure 1: A sequential overview of data assimilation. Here $m(\cdot)$ represents the model, of the underlying system, and $h(\cdot)$ is the observation operator.

In particular, within the VISDEM framework, unlike most other methodologies, emphasis will be placed on estimating unknown model parameters, as well as model state, thus making full use of the observations in the posterior process.

Variational Gaussian process approximation

Our approximate inference is to the posterior measure over sample paths. In the presence of observations this is given by:

$$\frac{dp_{post}}{dp_{sde}} = \frac{1}{Z} \times \prod_{m=1}^{M} p(\mathbf{y}_m | X_m), \tag{1}$$

where the likelihood is assumed to have a multivariate Gaussian form:

$$p(\mathbf{y}_m | X_m) = \mathcal{N}(\mathbf{y}_m | \mathbf{H}X_m, \mathbf{R}), \tag{2}$$

$\mathbf{H} \in \Re^{d \times D}$ is a linear transformation between the unobservable process and the measured observations, $\mathbf{R} \in \Re^{d \times d}$ is the observational noise covariance matrix and $Z = p(Y)$ is the normalisation constant or marginal likelihood.

The idea behind our variational method [1] is to approximate the unknown distribution $p(X)$ by another distribution $q(X)$ that belongs to a family of tractable distributions (e.g. Gaussian processes). That is:

$$p(x) \rightarrow dX_t = \mathbf{f}_{\boldsymbol{\theta}}(t, X_t)dt + \sqrt{\boldsymbol{\Sigma}}dW_t, \tag{3}$$
$$q(x) \rightarrow dX_t = \mathbf{g}(t, X_t)dt + \sqrt{\boldsymbol{\Sigma}}dW_t, \tag{4}$$

where both the above SDEs are described in the Itō sense. Here $\mathbf{f}_{\boldsymbol{\theta}}(\cdot)$ is the drift function of the true process, $\mathbf{g}$ is a linear function where $\mathbf{g}(t, X_t) = -A(t)X_t + b(t) \in \Re^D$, with $A(t) \in \Re^{D \times D}$ and $b(t) \in \Re^D$ being time-dependent functions that need to be optimised. $\boldsymbol{\Sigma}$ is the system noise covariance matrix and $dW_t \sim \mathcal{N}(\mathbf{0}, dt\mathbf{I})$.

The aim is to minimise the variational free energy (or equivalently the KL divergence) between the two processes, which is defined as:

$$\mathcal{F}_{\boldsymbol{\Sigma}}(q, \boldsymbol{\theta}) = -\left\langle \ln \frac{p(Y, X|\boldsymbol{\theta}, \boldsymbol{\Sigma})}{q(X|\boldsymbol{\Sigma})} \right\rangle_q \tag{5}$$

We achieve that with a smoothing algorithm.

[a] VISDEM stands for Variational Inference in Stochastic Dynamic Environmental Models.

Smoothing Algorithm

We illustrate the VGPA smoothing algorithm on the one dimensional noisy double well system (see Figure:2).

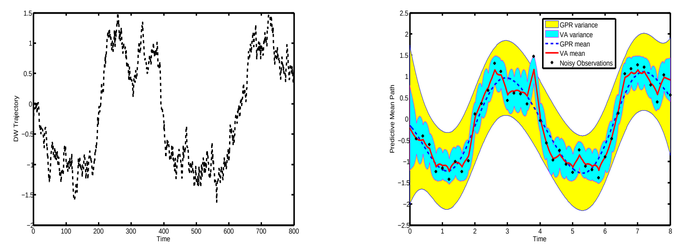

Figure 2: Example of VGPA algorithm. On the left we have the double well system with three rather unlikely transitions between the two wells. On the right we have the result of the variational approximation compared with a Gaussian process regression, at convergence.

We start with some initial values for the variational parameters $\mathbf{A}$ and $\mathbf{b}$ where we have discretised time. We then propagate these values forward in time (forward sweep), to produce the predictive mean $\mathbf{m}$ and variance $\mathbf{S}$.

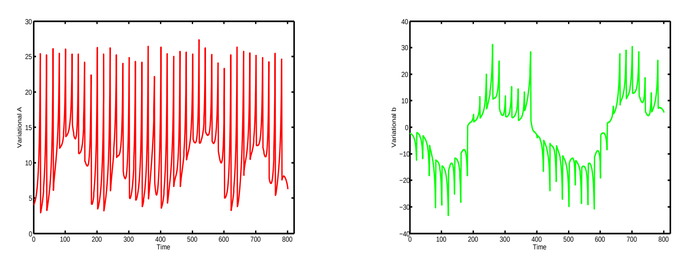

Figure 3: Example of variational parameters, obtained when the algorithm has converged.

Subsequently, we calculate the energy that comes from both the SDE and the observations and then we propagate the variational parameters $\mathbf{A}$ backward in time (backward sweep), to produce the Lagrange multipliers (see Figure:4, on the left). This ensures consistency with the constraints imposed by the Gaussian nature of our approximate process.

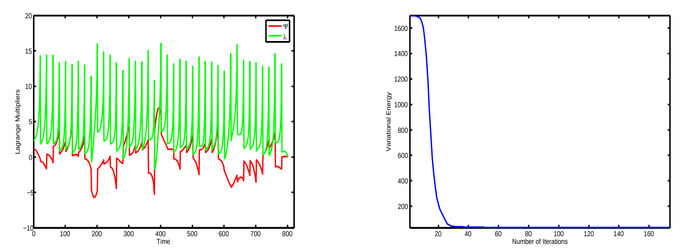

Figure 4: Example of the Lagrange multipliers and the energy evolution, obtained on the convergence of the algorithm.

Finally we iterate the above procedure until the desired level of convergence has been achieved (see Figure:4, on the right).

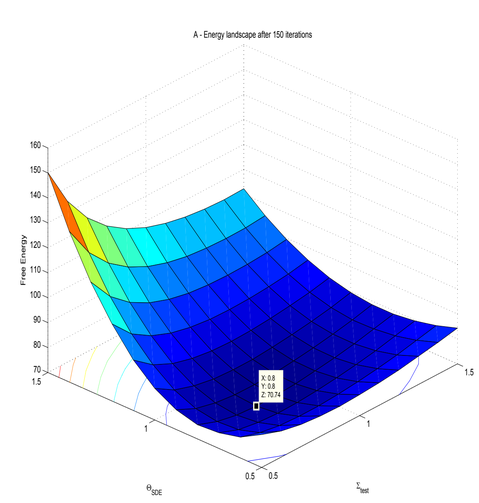

Figure 5: Profile of the free energy on the parameter space, at convergence.

In Figure (5) we can see the "landscape" of the free energy, on the parameter space, for a range of different values of the forcing (drift) parameter $\boldsymbol{\theta}$, as well as the system noise covariance matrix $\boldsymbol{\Sigma}$.

Experiments - Results

Running a relatively large number of simulations on different double well trajectories (see Figure:2, on the left) and different realisations of every sample path, the aim of these experiments was to test the robustness to different initialisations of the variational approximation framework. In addition, the convergence properties of the algorithm were investigated. In all the results that follows the initial values for the parameters were : $\boldsymbol{\theta}_{init} = 1.5$ and $\boldsymbol{\Sigma}_{init} = 1.5$, while the true values were $\boldsymbol{\theta}_{true} = 1.0$ and $\boldsymbol{\Sigma}_{true} = 1.0$.

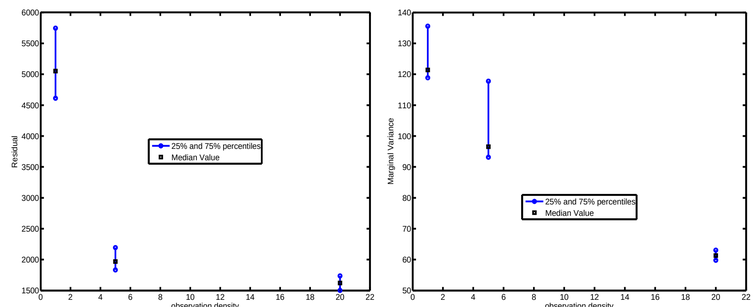

Figure 6: Results of the residual and the optimal marginal variance of the optimization process, as a function of the observation density, at convergence.

In Figure (6) we can see the averaged results for the residual and the marginal variance. The residual is given by:

$$I = \sum_{t=1}^{T} \frac{(m_t - x_t)^2}{S_t}, \tag{6}$$

where m_t, S_t are the mean value and the variance at time t, x_t is the true value at time t and T is the discretisation set. In addition to the state estimation, obtained by the smoothing algorithm, as seen in Figure (2, on the right) we can calculate gradients of the free energy [2], with respect to the parameters that we want to optimise, resulting in an outer optimisation loop.

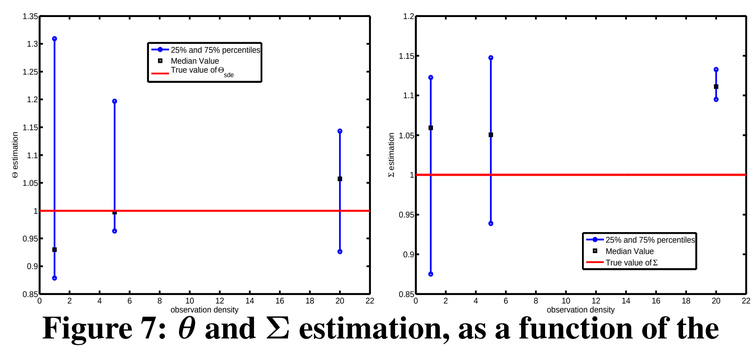

Figure 7: $\boldsymbol{\theta}$ and $\boldsymbol{\Sigma}$ estimation, as a function of the observation density, at convergence. The red horizontal line corresponds to the true value of the parameters.

In Figure (7) we can see the results for the estimation of the parameters. In both cases the results show that the estimated values are close to the true ones.

Conclusions & Future Work

A novel variational Gaussian approximation algorithm was presented, which allows estimation of the system forcing and noise parameters, using gradient techniques. The initial results are promising. However, many things need to be done to make the algorithm more robust and efficient (e.g. better optimisation method, suboptimal approximations of the variational parameters, variational Bayes estimation of parameters). Also a more broad range of higher dimensional dynamical systems needs to be considered.

Acknowledgements

I wish to acknowledge the help and support of all the VISDEM colleagues and especially Dr. Cédric Archambeau, for providing useful MATLAB code, and Prof. Manfred Opper for forming and explaining the theoretical part of the variational approximation framework.

References

[1] C. Archambeau, D. Cornford, M. Opper, and J. Shawe-Taylor. Gaussian process approximations of stochastic differential equations. *Journal of Machine Learning Research (JMLR)*, pages 1–16, 2007.

[2] C. Archambeau, M. Opper, Y. Shen, D. Cornford, and J. Shawe-Taylor. Variational inference for diffusion processes. In *Annual Conference on Neural Information Processing Systems (NIPS)*, 2007.

[3] E. Kalnay. *Atmospheric Modeling, Data Assimilation and Predictability.* Cambridge University Press., 2003.

Address for correspondence: Neural Computing Research Group (NCRG), School of Engineering and Applied Science, Aston University, Aston Triangle, Birmingham, B4 7ET