Parameter Estimation in Chaotic Stochastic Differential Equations using Hamiltonian Monte Carlo

Dallas Foster
Oregon State University

Abstract

Chaotic behavior is a well known limiting factor in the predictability of state estimation and forecasting. While the influence of instability in dynamics has a long studied history in the literature, less is known about the ability to recover parameter values in systems with chaos. This paper sets out to begin to understand the relationship between chaos and parameter fidelity by analyzing a modified Lorenz 96 model. Using this model and a Bayesian framework for parameter estimation, we are able to directly measure the impact that the mean and variance of external forcing (and hence chaotic nature of the solution) have on generating samples from the associated posterior probability distribution. We find that while increasing external forcing (and chaotic periodicity) have a positive relationship with parameter recovery, the variance of the forcing has an important stunting impact on estimates.

Introduction

The combination of model with observation to produce estimates of the current state of a system is often called Data Assimilation(Ching, Beck, and Porter, 2006). This inverse problem, concerned with optimal estimation of state variables, assumes a known and well-specified model of the dynamics in question. When parameters of the proposed model are unknown, optimal estimation must necessarily include inference on these values in addition to the state. State estimation on nonlinear dynamical systems has a long history with many treatments (Kushner, 1967), but the study of parameter estimation in such systems is more limited(Sarkka, 2013; Ghahramani, Z. & Hinton, 1996; Ching, Beck, and Porter, 2006). Since parameters are never directly observed, their inference is dependent on the indirect effect they have on the state variables.

For many problems, parameter estimation can be formulated in a deterministic framework using some variation of least squares minimization(Stock et al., 2015). This estimation method has some deficiencies. First, model or observational uncertainties cannot be appropriately accounted for in the deterministic framework and the resulting parameter estimation may fail or be widely inaccurate. Second, uncertainties in the parameter values may be mis-characterized or mis-calibrated. Lastly, optimization-based methods that return point estimates may be overly prone to numerical error and only return partial information about the parameters. Many other attempts at parameter estimation co-op methods for state estimation by including the parameters as state variables (Sarkka, 2013; Møller, Madsen, and Carstensen, 2011; Ghahramani, Z. & Hinton, 1996; Singer, 2002).

We approach the parameter estimation problem using a Bayesian framework, meaning that we will be concerned with characterizing the probability distribution of the parameter - as a random variable. This approach, in generality, is not unique (Sarkka, 2013; Ching, Beck, and Porter, 2006; Karimi and McAuley, 2016; Gelman et al., 2013). Using Bayes' theorem, we will construct the posterior distribution using a combination of distributions utilizing the observations, model, and prior beliefs about the parameter. To characterize the distribution, it is common to use the mean and standard deviations of the posterior (Maximum A Posteriori Estimation). This method, while computationally and memory cheap, may not capture the nuances of the posterior distribution. Furthermore, samples from the posterior distribution may be more important than summary statistics alone (Brooks, 2011). This may be the case if integrals with respect to the posterior are required, or if the parameters are to be used in the model to forecast the state variables.

Producing samples from the posterior distribution is often difficult, as it lacks a closed form expression. There are many methods to try and sample from an arbitrary probability distribution(Brooks, 2011). In this project, we focus on a Markov Chain Monte Carlo (MCMC) approach called Hamiltonian Monte Carlo (HMC)(Neal, 2012). Using the method in (Alexander, Eyink, and Restrepo, 2005), we construct the Hamiltonian using the model and observation distributions as position and conjugate momentum variables. The prior distribution of the parameters are included as a third term that influence the Hamiltonian. As opposed to traditional MCMC methods, the HMC approach produces less correlated samples that converge to the target distribution faster, requiring fewer samples and less memory overhead at the expense of increased computational cost.

As a final dimension to the project, we investigate the effect that chaos in the model has on parameter fidelity. Original studies of chaos and the development of the famous Lorenz 63 and 96 models (Lorenz, 1995) revolved around the question of atmospheric predictability. It is well known that chaotic behavior in model dynamics dramatically reduces the ability to produce accurate long-time forecasts.

Some studies have investigated the efficiency of parameter estimation in chaotic systems, but not in imperfect model systems and not as a function of increasingly chaotic behavior. Therefore, in this study, we investigate the Lorenz 96 model, of which we modify to include an unknown parameter, because we can control the level of chaos.

Problem Framework

Let $\mathbf{x}(t) = (x_1, \ldots, x_D) \in \mathbb{R}^D \times \mathbb{T}$ be the state variables of interest. The model for these variables is taken to be a modified Lorenz 96 system, defined as

$$\frac{dx_i}{dt} = a\left(x_{i+1} - x_{i-2}\right)x_{i-1} - x_i + F, \tag{1}$$

for $i = 1, \ldots, D > 3$, where we assume that $x_{-1} = x_{D-1}$, $x_0 = x_D$, and $x_{D+1} = x_1$. F is a parameter that controls the level of external forcing of the system, and as a result, the level of chaos. The additional variable a is the focus of our parameter estimation efforts. The model (1) may be thought of as some atmospheric quantity in D sectors of a latitude circle. For $a = 1$, it is known that, for small F, all solutions decay to the steady state solution $x_1 = \cdots = x_D = F$. For larger values of F, the solutions become periodic and increasingly chaotic. A typical value of F for which the system exhibits chaotic behavior depends on D, but usually taken to be 8. In this study, we will use $D = 10$ and F ranging from 1 to 15. For shorthand, we will write (1) as

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}; a)$$

To incorporate model uncertainty, we are concerned not with the deterministic (1), but the Itô stochastic differential equation

$$d\mathbf{x} = f(\mathbf{x}; a)dt + Qd\mathbf{W}_t, \tag{2}$$

where $\mathbf{W}_t$ is standard Brownian Motion, and $Q \in R^{D\times D}$ is some symmetric positive definite matrix taken roughly as representing the uncertainty between variables and the model. The SDE (2) is simulated using the Euler-Maryuama discretization

$$\begin{aligned} \mathbf{x}_n &= \mathbf{x}_{n-1} + f(\mathbf{x}_{n-1}; a)\Delta t + \sqrt{\Delta t Q}\Delta\mathbf{W}_n, \\ &= \mathcal{F}(\mathbf{x}_{n-1}; a), \end{aligned} \tag{3}$$

for $n = 1, \ldots, N$. Draws from this SDE for a variety of forcing parameter values is shown in **Figure 1**

To construct the parameter estimation problem, consider observations of the form

$$\mathbf{d}(t_m) = \mathbf{x}(t_m) + \epsilon_m, \quad m = 1, \ldots, M, \tag{4}$$

where $\epsilon \sim \mathcal{N}(0, R)$. It is assumed that observations are on a much sparser grid than the model, i.e. $M \ll N$. In the experiments in this paper, we will observe the impact of the frequency of observations on parameter fidelity by using $t_m = 5\Delta t + t_{m-1}$ and $t_m = 10\Delta t + t_{m-1}$. In general, one could also assume a more general observation operator of the form $h(\mathbf{x}(t_m))$ with little change to the theory or implementation of the algorithm.

The model dynamics and observation relations induce the conditional probability densities $p(\mathbf{x}_n|\mathbf{x}_{1:n-1}, a) = \exp\left(-H_{\text{dynamics}}\right)$ and $p(\mathbf{d}(t_{1:M})|\mathbf{x}(t_{1:M})) = \exp\left(-H_{\text{obs}}\right)$, where H_{dynamics} and H_{obs}, defined as

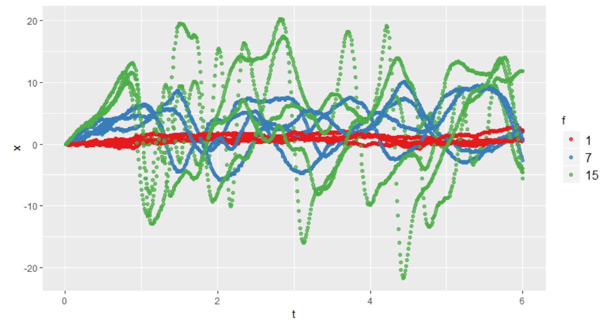

Figure 1: Draws from Lorenz SDE Model (2) using the Euler-Maryuama scheme (3) with parameter values $a = 0.75, Q = \frac{3}{4}I, dt = 0.01$ for various forcing values f.

$$\begin{aligned} H_{\text{dynamics}} &= \sum_{n=1}^{N} \frac{1}{2\Delta t}\left(\mathbf{x}_n - \mathbf{x}_{n-1} - \Delta t f(\mathbf{x}_{n-1}; a)\right)^T Q^{-1} \\ &\quad \times \left(\mathbf{x}_n - \mathbf{x}_{n-1} - \Delta t f(\mathbf{x}_{n-1}; a)\right), \\ H_{\text{obs}} &= \frac{1}{2}\sum_{m=1}^{M}\left(\mathbf{d}(t_m) - \mathbf{x}(t_m)\right)^T R^{-1}\left(\mathbf{d}(t_m) - \mathbf{x}(t_m)\right), \end{aligned} \tag{5}$$

represent factors of a Hamiltonian. The Hamiltonian being constructed comes from the application of Bayes' theorem to the conditional posterior distribution for the parameters,

$$\begin{aligned} p(a|\mathbf{x}_{1:N}, \mathbf{d}(t_{1:M})) &= \tfrac{1}{Z} p(\mathbf{x}_n|\mathbf{x}_{1:n-1}, a) p(\mathbf{d}(t_{1:M})|\mathbf{x}(t_{1:M})) p(a) \\ &= \tfrac{1}{Z}\exp(-H_{\text{dynamics}} - H_{\text{obs}} - H_{\text{prior}}). \end{aligned} \tag{6}$$

The choice of prior distribution $p(a)$, and subsequent term to the Hamiltonian H_{prior} depends on the nature of the parameter and any expert knowledge about its value. The use of informative vs. non-informative priors is a large discussion in the Bayesian community and beyond the scope of this paper. Here, we simply take a to be half-Cauchy distributed, ensuring positivity but neglecting to infer scale or location information.

As is implicit in our notation, the parameters F, Q and R are assumed known a priori, and are not being treated as random variables akin to the unknown parameter a. In our analysis, we will vary the values of F and Q so to discern their impact on the fidelity of a. Because of the form of the external forcing in (1), the noise in (2) simply acts to modify that forcing by increasing the variance. That is, modifications in F and Q simply modify the mean and variance, respectively, of the forcing in Lorenz 96 model. We will assume that the observation error is small in comparison to model error, R will be taken to be $\frac{1}{100}I$.

This formulation, writing the probability distributions in the language of statistical mechanics and Hamiltonians, will be useful when we detail the Hamiltonian Monte Carlo methodology for producing samples from the conditional posterior distribution for the parameter a. Once these samples are produced we have completed the parameter estimation problem, since we will have the ability to infer statistical knowledge about the random variable a from the samples.

Hamiltonian Monte Carlo

Sampling techniques such as Markov Chain Monte Carlo (MCMC) (Brooks, 2011; Kemp, 2003) rely on constructing a Markov chain of random variables, $\{\epsilon_1, \epsilon_2, \dots\}$ whose frequency well-describes the probability distribution of interest. A basic MCMC algorithms contains the followings steps: (i) specify an initial distribution, (ii) propose a new ϵ_k from ϵ_{k-1} using a transition distribution, and (iii) accept or reject the proposed ϵ_k using some criterion. For the MCMC method to generate a proper Markov chain that converges to the sought posterior distribution as its stationary distribution, the transition density and acceptance criterion must meet certain theoretical conditions, see (Brooks, 2011).

The Hamiltonian Monte Carlo (HMC) method is a specific MCMC algorithm that uses an evolution of a Hamiltonian dynamical system to propose new elements of the Markov chain, and accepts/rejects those proposals based on the change in total energy. Developed in the 1980's, its use had early on been confined to physics problems in statistical mechanics and lattice gauge theory (Alexander, Eyink, and Restrepo, 2005). For more information on applied probabilistic problems, or the implementation of HMC, see (Neal, 2012).

Consider a system with T degrees of freedom $\mathbf{q}_1, \dots \mathbf{q}_T$, representing the combined state and parameter variables. The HMC algorithm assigns conjugate generalized momentum variables $\mathbf{p}_i$ to each generalized coordinate $\mathbf{q}_i$. The momenta $\mathbf{p}_i$ induce the kinetic energy

$$H_K = \sum_{i=1}^{T} \frac{1}{2} \mathbf{p}_i^T M \mathbf{p}_i$$

where M is some mass matrix. For our purposes, the total Hamiltonian is given by

$$\begin{aligned} H &= H_{\text{dynamics}} + H_{\text{obs}} + H_{\text{prior}} + H_K, \\ &= H_U + H_K, \end{aligned} \tag{7}$$

where H_U, representing the potential energy portion of the total Hamiltonian, accounts for the posterior distribution (5). The standard dynamics of this Hamiltonian system is

$$\begin{aligned} \frac{d\mathbf{q}_i}{d\tau} &= \mathbf{p}_i \\ \frac{d\mathbf{p}_i}{d\tau} &= -\frac{\partial H_U}{d\mathbf{q}_i} \end{aligned} \tag{8}$$

where τ is some fictitious time. Given a current value of the generalized coordinates $(\mathbf{q}^0, \mathbf{p}^0)$ (the previous accepted member of the Markov chain), the dynamics are discretized using L leapfrog steps of size $\Delta\tau$,

$$\begin{aligned} \mathbf{p}^{k+\frac{1}{2}} &= \mathbf{p}^k - \frac{\Delta\tau}{2} \nabla_{\mathbf{q}} H_U(\mathbf{q}^k) \\ \mathbf{q}^{k+1} &= \mathbf{q}^k + \Delta\tau M \mathbf{p}^{k+\frac{1}{2}} \\ \mathbf{p}^{k+1} &= \mathbf{p}^{k+\frac{1}{2}} - \frac{\Delta\tau}{2} \nabla_{\mathbf{q}} H_U(\mathbf{q}^{k+1}) \end{aligned} \tag{9}$$

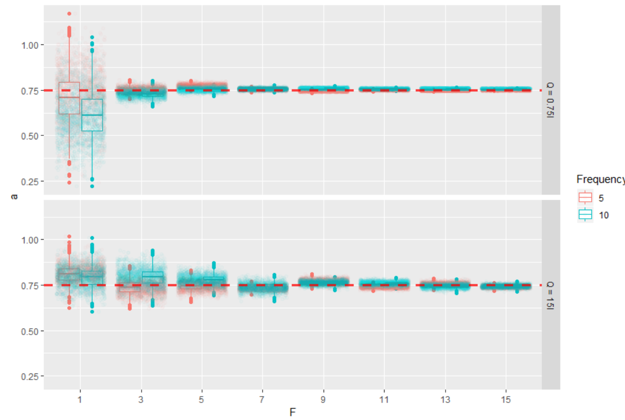

Figure 2: Samples of a from the conditional posterior distribution $p(a|\mathbf{x}_{1:N}, \mathbf{d}(t_{1:M}))$ as a function of the mean external forcing F and variance of the Weiner process Q. Red dashed line represents the 'true' parameter value. Scatter is offset to prevent over-plotting, i.e. we only considered odd integer valued F.

to arrive at the proposed sample $(\mathbf{q}^L, \mathbf{p}^L)$. This proposed sample is accepted with probability

$$\alpha = \min\left(1, \exp\{H(\mathbf{q}^0, \mathbf{p}^0) - H(\mathbf{q}^L, \mathbf{p}^L)\}\right). \tag{10}$$

If the proposed sample is accepted, then $(\mathbf{q}^0, \mathbf{p}^0) \leftarrow (\mathbf{q}^L, \mathbf{p}^L)$, otherwise $(\mathbf{q}^0, \mathbf{p}^0)$ is unchanged and is evolved again.

Results

Using the methods mentioned heretofore, a parameter estimation experiment was performed. A 'true' parameter value of $a = 0.75$ was used to evolve the system (3) with $\Delta t = 0.01$ to construct the sequence $\mathbf{x}_n \in \mathbb{R}^{10}$ for $n = 1, \dots, 600$. Observations of this system were performed at a frequency of either $5\Delta t$ or $10\Delta t$ and variance taken to be $\frac{3}{4}I$ or $15I$ depending on the experiment.

Using the observations, 2000 samples were generated using HMC on the posterior distribution (6). These samples are displayed, along with a boxplot describing their distribution, as a function of the mean forcing and noise variance in **Figure 2**.

The mean of samples of a roughly target the 'true' parameter value for all tested values of F and Q, which indicates that the parameter estimation was largely successful. The size of the variance of samples for small forcing values (e.g. $F = 1$ and $Q = \frac{3}{4}I$) is noticeably larger than for samples taken from a system with large external forcing (e.g. $F \geq 3$ or $Q = 15I$). In conjunction with an observation about the scale of changes in dynamics from **Figure 1**, it seems to be that more chaotic dynamics induces a more informative likelihood function. That is, small changes in a were more likely to produce larger changes in the model when the system was already in a chaotic regime and therefore be more penalized by the probability distribution. Therefore, it appears that parameter fidelity is enhanced by chaotic dynamics.

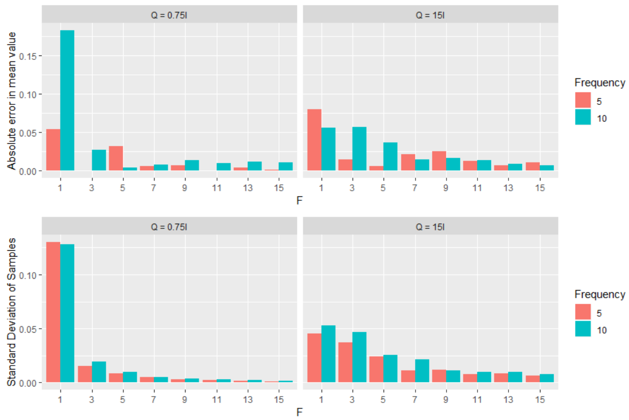

Figure 3: (TOP) Absolute errors in mean of $p(a|\mathbf{x}_{1:N}, (t_{1:M}))$ from 'True' value of $a = 0.75$ as a function of forcing parameter F, variance Q, and frequency of observations. (BOTTOM) Standard deviation of sampled posterior distribution.

This result can be made more qualitative by the results in **Figure 3**, which compares the absolute error between the mean of the samples with the 'true' parameter value and standard deviation of samples as a function of the external forcing. The largest absolute error, on the order of 20%, and standard deviation occurs when the forcing is extremely weak. For all other combinations of Q and F, the fidelity is very good ($\ll 5\%$ relative error), regardless of the frequency of observations.

Figure 3 also shows that increased variance in the external forcing diminishes the positive impact that chaos has on parameter estimation. When the mean external forcing is large, but model uncertainty is low, the standard deviation of the samples quickly decays as a function of F. This is in contrast to the case when $Q = 15I$, where there seems to be marginal improvement as a function of F. The result now appears that the chaotic nature of the dynamics may be helpful to narrow the probability distribution, noisy model error naturally inflates the posterior distribution.

Discussion and Conclusion

Chaos poses a problematic feature in many applications of state estimation, like weather and climate forecasting. Because small perturbations in initial conditions can lead to large discrepancies in solutions at later times, forecasting of state variables will invariable have a limit to their accuracy. This nature, however, encodes a lot of information into the probability distribution when attempting a parameter estimation problem. The small disturbances induced by the wrong parameter value are heavily penalized and are given low weight in the posterior distribution of the parameter given the state values and system observations. From this perspective, it is not necessarily surprising that our results give confirmation to this reasoning.

By tuning the mean and variance of the external forcing in a modified Lorenz 96 model (1), we can directly measure the two dimensions of chaotic uncertainty on parameter fidelity. Indeed, when there is a lot of structure in a system, the standard deviation of parameter samples is quite large. The nuance of the ability of chaotic behavior to induce increased parameter fidelity lies in the distinction between the mean and variance of the extern forcing. Increased mean forcing resulted largely in increased fidelity as long as the variance of the forcing was not also large. This increased variability damped parameter fidelity to some degree, naturely as an extension of the increased uncertainty in the system.

In total, the Hamiltonian Monte Carlo scheme (defined by the dynamics 8) was an effective tool in drawing from the posterior distribution (6). The number of time steps included in the algorithm is the largest constriction and future advances in state estimation should look to ameliorate this restriction. Future work can still be done to analyze the impact of chaos on parameter estimation, including a study into larger forcing values - limited in our study due to time step restrictions. Furthermore, additional types of nonlinearity and parameters should be investigated to try and determine the role that the nature of the parameter has on estimation.

References

Alexander, F. J.; Eyink, G. L.; and Restrepo, J. M. 2005. Accelerated Monte Carlo for optimal estimation of time series. *Journal of Statistical Physics* 119(5-6):1331–1345.

Brooks, S. 2011. *Handbook of Markov chain Monte Carlo*. CRC Press/Taylor & Francis.

Ching, J.; Beck, J. L.; and Porter, K. A. 2006. Bayesian state and parameter estimation of uncertain dynamical systems. *Probabilistic Engineering Mechanics* 21(1):81–96.

Gelman, A.; Carlin, J. B.; Stern, H. S.; Dunson, D. B.; Vehtari, A.; and Rubin, D. B. 2013. *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis.

Ghahramani, Z. & Hinton, G. 1996. Parameter Estimation for Linear Dynamical Systems. Technical report.

Karimi, H., and McAuley, K. B. 2016. Bayesian Estimation in Stochastic Differential Equation Models via Laplace Approximation. *IFAC-PapersOnLine* 49(7):1109–1114.

Kemp, F. 2003. An Introduction to Sequential Monte Carlo Methods. *Journal of the Royal Statistical Society: Series D (The Statistician)* 52(4):694–695.

Kushner, H. 1967. Dynamical equations for optimal nonlinear filtering. *Journal of Differential Equations* 3(2):179–190.

Lorenz, E. N. 1995. Predictability: a problem partly solved. In *Seminar on Predictability, 4-8 September 1995*, volume 1, 1–18. Shinfield Park, Reading: ECMWF.

Møller, J. K.; Madsen, H.; and Carstensen, J. 2011. Parameter estimation in a simple stochastic differential equation for phytoplankton modelling. *Ecological Modelling* 222:1793–1799.

Neal, R. M. 2012. MCMC using Hamiltonian dynamics.

Sarkka, S. 2013. *Bayesian Filtering and Smoothing*. Cambridge: Cambridge University Press.

Singer, H. 2002. Parameter estimation of nonlinear stochastic differential equations: Simulated maximum likelihood versus extended kalman filter and itô-taylor expansion. *Journal of Computational and Graphical Statistics* 11(4):972–995.

Stock, C. A.; Pegion, K.; Vecchi, G. A.; Alexander, M. A.; Tommasi, D.; Bond, N. A.; Fratantoni, P. S.; Gudgel, R. G.; Kristiansen, T.; O'Brien, T. D.; Xue, Y.; and Yang, X. 2015. Seasonal sea surface temperature anomaly prediction for coastal ecosystems. *Progress in Oceanography* 137:219–236.