Parameter and Structure Inference for Nonlinear Dynamical Systems

Robin D. Morris[*], Vadim N. Smelyanskiy[†] and Mark Millonas[†]

*USRA-RIACS, NASA Ames Research Center, MS 269-2, Moffett Field, CA 94035
†NASA Ames Research Center, MS 269-2, Moffett Field, CA 94035

Abstract. A great many systems can be modeled in the non-linear dynamical systems framework, as $\dot{x} = f(x) + \xi(t)$, where $f()$ is the potential function for the system, and ξ is the excitation noise. Modeling the potential using a set of basis functions, we derive the posterior for the basis coefficients. A more challenging problem is to determine the set of basis functions that are required to model a particular system. We show that using the Bayesian Information Criteria (BIC) to rank models, and the beam search technique, that we can accurately determine the structure of simple non-linear dynamical system models, and the structure of the coupling between non-linear dynamical systems where the individual systems are known. This last case has important ecological applications.

Keywords: non-linear dynamical system; BIC; beam search; population dynamics;
PACS: 02.50.Tt, 05.45.Tp, 05.10.Gg, 87.19.Hh, 05.45.Pq, 05.45.Xt, 87.23.Cc

INTRODUCTION

Many natural systems can be modeled as noise-driven nonlinear dynamical systems. The classic example is the Lorenz system [1], intended as a simplified representation of climate variables. Other examples are the human cardiovascular system [2], where the heart-lung oscillations show physiologically important couplings, and predator-prey interactions, where one predator species feeds primarily on a single prey species, but will, when necessary, feed on others [3]. These last two are examples of an important class of problem where it is the *coupling* between subsystems which is of scientific importance.

A non-linear dynamical system is typically modeled as evolving according to

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}) + \xi(t), \tag{1}$$

where $\mathbf{f}(\mathbf{x})$ is the system's potential field, and $\xi(t)$ an additive noise process that drives the dynamics. The objects of inference are the field, $\mathbf{f}(\mathbf{x})$, and the intensity of the noise. A number of methods have been developed for these tasks. Deterministic methods fail to give accurate parameter estimates in the presence of noise [4]. A method was developed in [5] which was termed a maximum likelihood estimator, but was derived from an *ad-hoc* cost function. A Bayesian approach was formulated in [6], however this relied on MCMC, and so is too computationally intensive when the structure of $\mathbf{f}(\mathbf{x})$ must also be determined, as is the case here. A different estimation framework was suggested in [2, 7] that uses a path integral approach to derive a closed form expression for the parameter estimates. This method dramatically speeds up the convergence of the

inference process and makes it extremely stable to a large levels of dynamical noise $\xi(t)$. We use the parameter estimates obtained via this method together with the Bayesian Information Criteria (BIC) [8] to estimate the structure (in terms of basis functions) of $\mathbf{f}(\mathbf{x})$, using Beam Search, a non-complete search algorithm quite well known in the artificial intelligence community, but with little exposure in the Bayesian modeling field. Finally, we present a number of examples of learning the structure of unknown non-linear dynamical systems and the coupling between known systems.

PARAMETER ESTIMATION FOR NON-LINEAR DYNAMICAL SYSTEMS

An N-dimensional non-linear dynamical system evolves according to equation 1, and is observed at times t_k to give a time-series $\mathscr{Y} = \{(t_k, \mathbf{y}_k); k = 0 : K\}$. We assume that the observation noise is negligible compared with the excitation noise $\xi(t)$, hence $\mathbf{y}_k = \mathbf{x}_k$. Following the approach of [2, 7], we parameterize the potential field as

$$\mathbf{f}(\mathbf{x}) = \mathbf{U}(\mathbf{x})\,\mathbf{c} = \mathbf{f}(\mathbf{x};\mathbf{c})$$

where $\mathbf{U}$ is an $N \times M$ matrix of basis functions and $\mathbf{c}$ an M-dimensional vector. In this way we convert the problem of estimating a nonlinear $\mathbf{f}(\mathbf{x})$ into a problem that is linear in the parameters. We assume the excitation noise, $\xi(t)$, is stationary, white and Gaussian, and hence characterized by its covariance matrix, $\mathbf{D}$.

The posterior distribution of interest is

$$p(\mathbf{c}, \mathbf{D}|\mathscr{Y}) \propto p(\mathscr{Y}|\mathbf{c}, \mathbf{D})p(\mathbf{c})p(\mathbf{D})$$

where independent priors have been assumed. We use a diffuse Gaussian prior on the elements of $\mathbf{c}$ and a uniform prior on the elements of $\mathbf{D}$. The likelihood can be written in the form of a path integral over the random trajectories

$$p(\mathscr{Y}|\mathbf{c}, \mathbf{D}) = \int_{\mathbf{x}(t_\mathrm{i})}^{\mathbf{x}(t_f)} p(\mathscr{Y}|\mathscr{X})\,\mathscr{F}[\mathbf{x}(t)]\,\mathscr{D}\mathbf{x}(t),$$

where $\mathscr{X} = \{\mathbf{x}(t)\}$ and $t_i \ll t_0 < t_K \ll t_f$ so that the likelihood does not depend on the initial and final states. The probability functional $\mathscr{F}[\mathbf{x}(t)]$ is determined by the properties of the excitation noise $\xi(t)$ [9, 10]. Under the assumption of negligible observation noise, the log-likelihood is

$$-\frac{2}{K}\log p(\mathscr{Y}|\mathbf{c}, \mathbf{D}) = \ln\det\mathbf{D} + \frac{h}{K}\sum_{k=0}^{K-1}\left[\,\mathrm{tr}\hat{\Phi}(\mathbf{y}_k;\mathbf{c})\right.$$
$$\left. + (\dot{\mathbf{y}}_k - \mathbf{f}(\mathbf{y}_k;\mathbf{c}))^T\mathbf{D}^{-1}(\dot{\mathbf{y}}_k - \mathbf{f}(\mathbf{y}_k;\mathbf{c}))\right] + N\ln(2\pi h)$$

where $\dot{\mathbf{y}}_k = (\mathbf{y}_{k+1} - \mathbf{y}_k)/h$ and the matrix $\hat{\Phi}(\mathbf{x}) = \{\partial f_i(\mathbf{x};\mathbf{c})/\partial x_j\}$. The posterior is thus $p(\mathbf{c}, \mathbf{D}|\mathscr{Y}) = \frac{1}{Z}\exp[-S(\mathbf{c}, \mathbf{D}|\mathscr{Y})]$ where

$$S(\mathbf{c}, \mathbf{D}|\mathscr{Y}) = \frac{1}{2}\rho(\mathbf{D}) - \mathbf{c}^T\mathbf{w}(\mathbf{D}) + \frac{1}{2}\mathbf{c}^T\hat{\Xi}(\mathbf{D})\mathbf{c}$$

and

$$\rho(\mathbf{D}) \;=\; h \sum_{k=0}^{K-1} \dot{\mathbf{y}}_k^T \mathbf{D}^{-1} \dot{\mathbf{y}}_k + K \ln(\det \mathbf{D})$$

$$\mathbf{w}(\mathbf{D}) \;=\; \hat{\Sigma}_{\mathrm{pr}}^{-1} \mathbf{c}_{\mathrm{pr}} + h \sum_{k=0}^{K-1} \left[\mathbf{U}_k^T \mathbf{D}^{-1} \dot{\mathbf{y}}_k - \frac{1}{2} \mathbf{v}(\mathbf{y}_k) \right]$$

$$\hat{\Xi}(\mathbf{D}) \;=\; \hat{\Sigma}_{\mathrm{pr}}^{-1} + h \sum_{k=0}^{K-1} \mathbf{U}_k^T \mathbf{D}^{-1} \mathbf{U}_k$$

where $\mathbf{c}_{pr}$ and $\hat{\Sigma}_{pr}$ are the mean and covariance of the prior on $\mathbf{c}$, $\mathbf{U}_k \equiv \mathbf{U}(\mathbf{y}_k)$ and the components of vector $\mathbf{v}(\mathbf{x})$ are:

$$v_m(\mathbf{x}) = \sum_{n=1}^{N} \frac{\partial U_{nm}(\mathbf{x})}{\partial x_n}, \quad m = 1 : M.$$

We find the MAP estimates by alternately optimizing for $\mathbf{D}$ assuming $\mathbf{c}$ is known, and then for $\mathbf{c}$ assuming $\mathbf{D}$ is known. The maximizations can be done analytically, giving

$$\hat{\mathbf{D}}^{nn'}(\mathbf{c}) = \frac{1}{K} \sum_{k=0}^{K-1} [\dot{\mathbf{y}}_k - \mathbf{U}(\mathbf{y}_k)\,\mathbf{c}]_n\, [\dot{\mathbf{y}}_k - \mathbf{U}(\mathbf{y}_k)\,\mathbf{c}]_{n'}^T$$

and

$$\hat{\mathbf{c}} = \hat{\Xi}^{-1}(\mathbf{D})\mathbf{w}(\mathbf{D}).$$

Further details are given in [7].

STRUCTURE ESTIMATION USING BEAM SEARCH AND BIC

In general, which basis functions to include in $\mathbf{U}(\mathbf{x})$ will be unknown. Sometimes the *type* of basis function will be known from physical considerations (see later), but often it will be necessary to use a very general basis. In this case it is important to determine which of the basis terms are required to model the system. A very general basis is the Volterra expansion [11], where $V(N,k)$ is an expansion up to order k of an N-dimensional vector. For example

$$V(3,2) = \{1, x_1, x_2, x_3, x_1^2, x_1 x_2, x_1, x_3, x_2^2, x_2, x_3, x_3^2\}.$$

A model, $\mathscr{M}$, is then specified by which of the basis terms are included.

The formal Bayesian solution is to compute the integrated likelihood, or evidence, $p(\mathscr{M}|\mathscr{Y})$. In practice, for large datasets, the BIC can provide a good approximation [8],

$$BIC(\mathscr{M}) = \log p(\mathscr{Y}|\hat{\mathbf{c}}, \hat{\mathbf{D}}) - \frac{1}{2} \log N_m$$

where N_m is the number of terms in model $\mathscr{M}$. In either case the search is over 2^{N_m} models, and it rapidly becomes impossible to examine all models to find the model
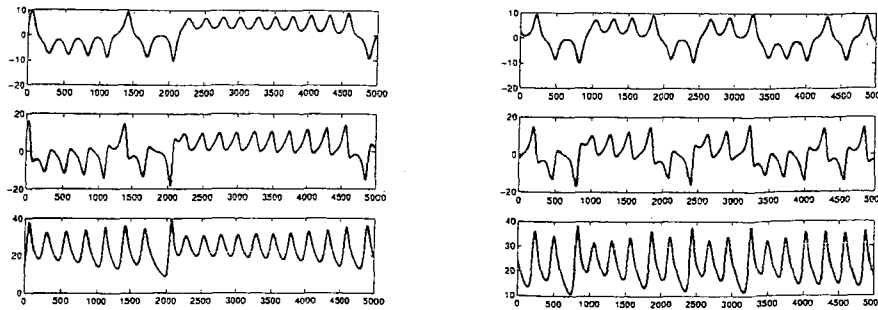
FIGURE 1. (left) Simulated trajectory from the Lorenz attractor. (right) Simulated trajectory from the system with the x_1^2 term included in the equation for $\dot{x}_3$..

with the highest BIC score. Instead, some non-exhaustive search must be used, with no guarantee of finding the optimum model. Instead, we hope to choose a search technique which returns a *good enough* model in an acceptable length of time.

Beam Search

Beam search is a heuristic where a set of models (the "beam") are maintained at each step, and it is hoped that the diversity of models retained will allow the algorithm to escape from local minima. Applied to the non-linear dynamical system structure learning problem, the beam search algorithm is as follows:

1. Choose the size, N_B, of the beam (the number of models to retain at each step).
2. Index the models by a binary vector.
3. Put models $1 \ldots N_B$ on the beam and compute the BIC for each model.
4. For each model on the beam
   - flip the next bit
   - compute the BIC
5. Of the $2N_B$ models now under consideration, keep the best N_B.
6. Goto 4, until all bits have been flipped

Instead of examining 2^{N_m} models, this algorithm examines $N_m \times N_B$. It is often useful to randomize the order in which the basis terms are added to each of the beam entries, to give a better exploration of the search space.

Results for the Lorenz Attractor

The Lorenz attractor is a three dimensional system governed by the equations

$$\dot{x}_1 = \sigma(x_2 - x_1)$$
$$\dot{x}_2 = rx_1 - x_2 - x_1 x_3$$
$$\dot{x}_3 = x_1 x_2 - bx_3$$

TABLE 1. Frequency of inclusion of each basis term in 100 repetitions.

$\dot{x}_1$	x_1	x_2			
	100	100			

$\dot{x}_2$	x_1	x_2	x_3		
	100	100	100		

$\dot{x}_3$	1	x_2	x_3	x_1^2	$x_1 x_2$	x_3^2
	6	5	94	100	100	6

where $\sigma = 10$, $r = 28$ and $b = 8/3$. A sample trajectory is shown in figure 1 for $h = 0.005$ and the variance of the excitation noise being 0.001. A suitable basis for this system is the $V(3,2)$ basis discussed above. This has 10 terms for each of the 3 dimensions, resulting in a search over $2^{30} \simeq 10^9$ models. Using beam search with a beam of size 16 results in fewer than 800 models being examined. The algorithm was run 100 times, using 5000 data points generated from the Lorenz system with a random initial start point, after 2000 data points were ignored to ensure that the trajectory had converged to the attractor. The frequency of inclusion of each term in the model is shown in table 1. Apart from the x_3 term in the equation for $\dot{x}_3$, which was included 94% of the time, the other correct terms were always included. In the equation for $\dot{x}_3$, the term x_1^2 was also always included incorrectly. Typically the value of the coefficient for this term was -0.04, which is an order of magnitude smaller than the smallest of the other coefficients.

Figure 1 (right) shows data simulated from the system with the x_1^2 term included in the equation for $\dot{x}_3$. Even with the extra term it is in good qualitative agreement with the data in figure 1 (left).

Coupled Lorenz Attractors

In many applications the system of interest is constructed by coupling together a number of simpler systems, where each of the simpler systems is either well known or can be studied in isolation. In these cases it is the structure of the *coupling* between the systems that is of interest.

In [12] the authors studied a system of three coupled Lorenz attractors. Inspired by their system, we modify the second equation in each oscillator to be

$$\dot{x}_2 = rx_1 - x_2 - x_1 x_3 + c\bar{x}_1$$

where $\bar{x}_1$ is the x_1 component from the previous oscillator (in a circularly coupled manner). Figure 2 shows a simulation of this system with $c = 28$. To model this system with no knowledge of the underlying Lorenz attractors would require the use of the $V(9,2)$ basis, which has 55 terms for each dimension, a total of 495. A search over a space this size is infeasible. Instead, we search only over the coupling between oscillators, including in all models the terms known to be included, and excluding terms that are completely within a single oscillator that are not in the Lorenz equations. This reduces the search space down to 54 terms.
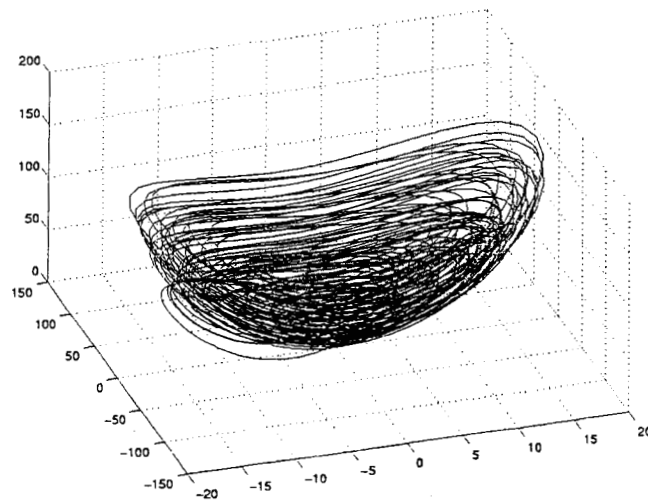
FIGURE 2. Simulation of the system of three coupled Lorenz oscillators

The estimated equations using the data in figure 2, and a beam of size 128 are

$$\dot{x}_1 = x_1, x_2$$
$$\dot{x}_2 = x_2, x_2, x_1 x_3, (x_4), x_7$$
$$\dot{x}_3 = x_3, x_1 x_2, (x_9)$$

$$\dot{x}_4 = x_4, x_5$$
$$\dot{x}_5 = x_1, (x_2), x_4, x_5, x_4 x_6$$
$$\dot{x}_6 = x_6, x_4 x_5, (x_9)$$

$$\dot{x}_7 = x_7, x_8$$
$$\dot{x}_8 = (x_1), (x_2), x_4, (x_5), x_7, x_8, x_7 x_9$$
$$\dot{x}_9 = (x_3), (x_6), x_9, x_7 x_8$$

where the terms not in the original equations are in parentheses.

PREDATOR-PREY SYSTEMS

The structure of predator-prey interactions in natural ecosystems is of great ecological interest. The structure can reveal which prey species are chosen preferentially by which predator species, and changes in this structure can be indicative of changes in the ecosystem due to external factors.
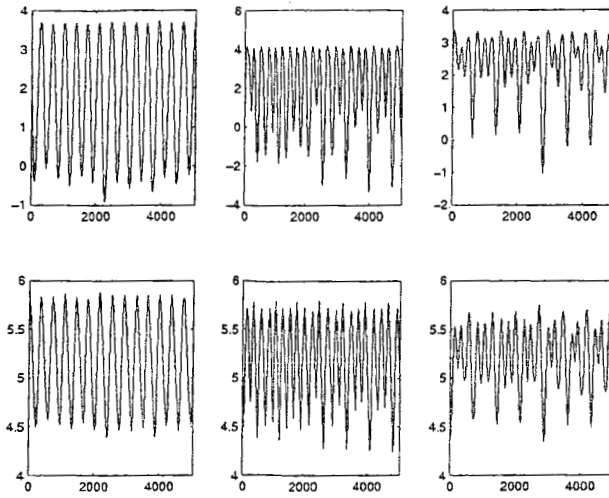
FIGURE 3. Synthetic data from a system describing three coupled predator (bottom) - prey (top) pairs

A multidimensional 2-trope predator-prey system can be modeled as

$$\frac{dH_i}{dt} = H_i \left[a_i - b_i H_i - \frac{c_i P_i}{1 + H_i} + \sum_{j \neq i} \frac{\alpha_{ij} P_j}{1 + H_i} + \xi_i(t) \right] \tag{2}$$

$$\frac{dP_i}{dt} = P_i \left[-d_i + e_i H_i + \sum_{j \neq i} \beta_{ij} H_j + \zeta_i(t) \right] \tag{3}$$

where P_i are the predators and H_i the prey. We assume the same number of predator species as prey species, and that predator P_i preys predominantly on a single prey species H_i. The remaining interactions, specified by the sparse matrices α_{ij} and β_{ij}, are weak. Also $\alpha_{ii} = \beta_{ii} = 0$. For this system we have also the constraint that for realism, if $\alpha_{ij} = 0$ then so does β_{ij}; if species P_i preys on H_i, then P_i shows some increase in quality of life.

Equations 2 and 3 are not in the form of equation 1, so we make the transformations $X_i = \ln H_i$ and $Y_i = \ln P_i$, giving

$$\frac{dX_i}{dt} = a_i - b_i \exp X_i - \frac{c_i \exp Y_i}{1 + \exp X_i} + \sum_{j \neq i} \frac{\alpha_{ij} \exp Y_j}{1 + \exp X_i} + \xi_i(t)$$

$$\frac{dY_i}{dt} = -d_i + e_i \exp X_i + \sum_{j \neq i} \beta_{ij} \exp X_j + \zeta_i(t)$$

The Volterra basis used in the earlier examples is not suitable here. Instead, we construct a basis of terms of the form $\exp X_i$ and $\frac{\exp Y_i}{1 + \exp X_i}$. Figure 3 shows synthetic data from a system of three coupled predator-prey pairs. Table 2 shows the system of equations used to generate this data, and also the frequency of inclusion of each term in the estimated model for ten data sets generated with different initial conditions. None of the terms not represented in table 2 was included in any of the estimated models.

TABLE 2. Frequency of inclusion of each basis term in 10 repetitions.

$\dot{X}_1 =$	$\dfrac{8}{10}$	$-0.14\exp X_1 \quad \dfrac{}{10}$	$-0.09\dfrac{\exp Y_1}{20+\exp X_1} \quad \dfrac{}{10}$	
$\dot{X}_2 =$	$\dfrac{20}{10}$	$-0.26\exp X_2 \quad \dfrac{}{10}$	$+0.08\dfrac{\exp Y_1}{20+\exp X_2} \quad \dfrac{}{3}$	$-2.5\dfrac{\exp Y_2}{20+\exp X_2} \quad \dfrac{}{10}$
$\dot{X}_3 =$	$\dfrac{11}{10}$	$-0.29\exp X_3 \quad \dfrac{}{10}$	$+0.05\dfrac{\exp Y_2}{20+\exp X_3} \quad \dfrac{}{7}$	$-1.2\dfrac{\exp Y_3}{20+\exp X_3} \quad \dfrac{}{10}$
$\dot{Y}_1 =$	$\dfrac{-0.8}{10}$	$+0.06\exp X_1 \quad \dfrac{}{10}$		
$\dot{Y}_2 =$	$\dfrac{-1.25}{10}$	$+0.01\exp X_1 \quad \dfrac{}{3}$	$+0.05\exp X_2 \quad \dfrac{}{10}$	
$\dot{Y}_3 =$	$\dfrac{-1.1}{10}$	$+0.02\exp X_2 \quad \dfrac{}{7}$	$+0.05\exp X_3 \quad \dfrac{}{10}$	

CONCLUSIONS

We have presented Bayesian inference methods for the potential of a non-linear dynamical system when the potential is described in terms of basis functions. We have shown that the search method known as "beam search" can be applied with BIC as the scoring function to the problem of determining which of the basis terms are required for a particular system. We have presented examples representative of an important class of system, and demonstrated reasonable performance in those cases.

ACKNOWLEDGMENTS

REFERENCES

1. E. Lorenz, *Journal of Atmospheric Science* **20**, 130–141 (1963).
2. V. Smelyanskiy, D. Luchinsky, A. Stfanovska, and P. McClintock, *Physical Review Letters* **94**, 098101 (2005).
3. N. Stenseth, W. Falk, O. Bjornstad, and C. Krebsdagger, *Procdeedings of the National Academy of Sciences* **94**, 5147–5152 (1997).
4. H. Kantz, and T. Schreiber, *Nonlinear Time Series Analysis*, CUP, 1997.
5. P. McSharry, and L. Smith, *Phyiscal Review Letters* **83**, 4285–4288 (1999).
6. R. Meyer, and N. Christensen, *Physical Review E* **62**, 3535–3542 (2000).
7. V. Smelyanskiy, D. Luchinsky, D. Timuccin, and A. Bandrivskyy, *Physical Review E* **72**, 026202 (2005).
8. B. Carlin, and T. Louis, *Bayes and Empirical Bayes Methods for Data Analysis*, Chapman and Hall/CRC, 2000, 2nd edn.
9. R. Graham, *Z. Physik B* **26**, 281–290 (1977).
10. M. Dykman, *Physical Review A* **42**, 2020–2029 (1990).
11. M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*, John Wiley and Sons, 1980.
12. E. Sanchez, and M. A. Matias, *International Journal of Bifurcation and Chaos* **9**, 2335–2343 (1999).