ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 303-308

Prediction of dynamical systems with composition networks

Yves Moreau , Joos Vandewalle

Katholieke Universiteit Leuven, Elektrotechniek-ESAT,
Kardinaal Mercierlaan 94, B-3001 Heverlee, Belgium

Abstract. In this paper we study the problem of predicting dynamical systems from discrete-time measurements of the state variables. Our approach is based on the theory of Lie algebras and the Baker-Campbell-Hausdorff formula. We present the composition network which is a multi-layer architecture which can use the a priori knowledge we have about the system. We also introduce the "MLP in dynamics space" which is a general implementation of the composition network having a universal approximation property. We demonstrate the efficiency of the proposed method on the task of predicting the Lorenz attractor.

1. Introduction

We study the problem of predicting nonlinear dynamical systems from discrete-time measurements. The systems which we consider here are continuous-time autonomous dynamical systems. This means that they can be described by some ordinary differential equation $\dot{x} = f(x)$ on the state-space. It is often the case in practice that the system is observed at discrete-time intervals Δt only. We will further assume that all the states variables can be directly measured. The data is thus a sequence of states $x(0), x(\Delta t), ..., x(N.\Delta t)$. Under those assumptions the system now takes a discrete-time form $x(k+1) = F(x(k))$. The purpose of learning is to determine an estimate for F. This writes as:

$$\hat{F} = \min_{G \in \mathcal{M}} \sum_{j=1}^{N-1} \|G(x(j.\Delta t)) - x((j+1).\Delta t)\|^2 \qquad (1)$$

where $\mathcal{M}$ is some parametrized family of admissible models. The task of learning is then to determine an optimal $\hat{F}$, that is a set of parameters which minimizes the prediction error.

This research work was carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the framework of a Concerted Action Project of the Flemish Community, entitled *Model-based Information Processing Systems* GOA-MIPS, and within the framework of the Belgian program on inter-university attraction poles (IUAP-17) initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. The scientific responsibility rests with its authors. Yves Moreau is a research assistant of the N.F.W.O. (Belgian National Fund for Scientific Research).

ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 303-308

The main drawback of using some class of neural networks as the class of admissible models is that it can be shown [6] that these classes contain many models which do not have the structure necessary for the prediction of any dynamical system. The method we present answers these objections by providing models whose structure is closely related to that of dynamical systems.
The first section of this paper presents the elements of Lie algebra theory, the Baker-Campbell-Hausdorff formula and derives the structure of the composition network. The second section presents a general implementation of the composition network called "MLP in dynamics space" which has a universal approximation property in the space of dynamical systems. The last section presents the results of the learning of the Lorenz attractor with this method.

2. Lie algebra theory

Lie algebra theory has a long history in physics, mostly in the areas of classical mechanics [1] and partial differential equations. It is also an essential part of nonlinear system theory [3]. Our approach to system prediction can be best cast in this framework.
The first element we need is the notion of the exponential map of a vector field. Suppose a system of the form $\dot{x} = A(x)$ where A is some vector field defined on the state-space. The solution of the initial value problem is denoted by:

$$y = e^{t.A} x_0 \tag{2}$$

This can be read as: "y is the image after time t of the initial condition x_0 under the flow of $\dot{x} = A(x)$". Although this notation is also valid to the linear case, the exponential $e^{t.A}$ is in general a nonlinear map from the state-space onto itself.
The other notions we will need are these of Lie algebra and Lie bracket. A Lie algebra is a vector space where we define a supplementary operation: the bracket $[.,.]$ of two elements of the algebra. The bracket is an operation which is bilinear, antisymmetric and satisfies the Jacobi identity [1]. One can then define the Lie bracket by:

$$[A, B] = \left. \frac{\partial^2}{\partial s . \partial t} \right|_{t=s=0} e^{-s.B} . e^{-t.A} . e^{s.B} . e^{t.A} \tag{3}$$

Here the product of exponentials denotes the composition of the maps. When the state-space is $\mathbb{R}^n$, analytical expressions for the bracket can be found [1].

2.1. The Baker-Campbell-Hausdorff formula

The Baker-Campbell-Hausdorff (BCH) formula gives an expansion for the product of two exponentials of elements of the Lie algebra, see [4].

$$e^A . e^B = e^{A+B+\frac{1}{2}[A,B]+\frac{1}{12}([A,[A,B]]-[B,[B,A]])+\ldots} \tag{4}$$

ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 303-308

The problem of predicting a dynamical system with vector field X then becomes that of building an approximation for $e^{\Delta t.X}$ as we have that $x(k+1) = e^{\Delta t.X}x(k)$.

This problem has recently been the focus of much attention in the field of numerical analysis for the integration of differential equations [4] [5]. Suppose that the vector field X is of the form $X = A + B$ where we can integrate A and B directly. We can use the BCH formula to produce a first-order approximation to the exponential map:

$$\text{BCH} : e^{\Delta t.X} = e^{\Delta t.A}.e^{\Delta t.B} + o(\Delta t^2) \tag{5}$$

This is the essence of the method as it shows that one can approximate an exponential map (that is the map arising from the solution of an ODE) by composing simpler maps. By repeated use of the BCH formula, we can show that the following leapfrog scheme is second-order. Using this leapfrog scheme as a basis element for further leapfrog schemes, Yoshida [8] showed that it was possible to produce an approximation to $e^{\Delta t.X}$ up to any order. Forest and Ruth [2] showed that approximations could be built for more than two vector fields. Combining this we can state that it is possible to build an approximation to the solution of a linear combination of vector fields as a product (i.e. compositions) of exponential maps.

$$\begin{array}{l} X = \sum_{i=1}^{m} a_i.A_i \\ \Rightarrow \exists w_{ij} : e^{\Delta t.X} = \prod_{j=1}^{k} \prod_{i=1}^{m} e^{w_{ij}.\Delta t.A_i} + o(\Delta t^{p+1}) \end{array} \tag{6}$$

Suzuki and Yamauchi [7] have further shown that some schemes could be guaranteed to converge for arbitrary value of Δt.

2.2. Example

We briefly explicit the essence of the technique by looking at the Van der Pol oscillator. This system can be written as a first-order system in state-space form. It can be seen as a linear combination of two vector fields A, B which can be solved analytically:

$$\begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{pmatrix} y \\ -x \end{pmatrix} + \alpha. \begin{pmatrix} 0 \\ (1-x^2).y \end{pmatrix} = A + \alpha.B \tag{7}$$

The flows of the vector fields X, A and B are presented in the following figure (Fig. 1). Back to the system prediction problem, let us suppose we observe a sample x_0 and want to compute an estimate for a future sample x_1. Consider the first-order case, an estimate for x_1 is $e^{\Delta t.\alpha.B}e^{\Delta t.A}x_0$. As can be seen on the next figure (Fig. 2), this corresponds to first following the flow of A for Δt units of time, then following the flow of B for $\alpha.\Delta t$ units of time. This gives an estimate for x_1. This is not the only first-order estimate, we could also use $e^{\Delta t.A}e^{\Delta t.\alpha.B}x_0$ as an estimate for x_1. By using more than two compositions steps, we can build higher-order estimates for x_1, hence more precise ones.

ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 303-308

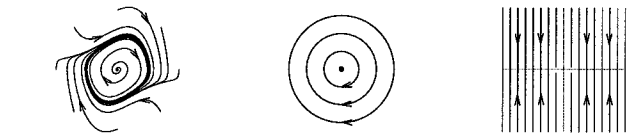

Figure 1: flow of the Van der Pol oscillator and of its splittings.

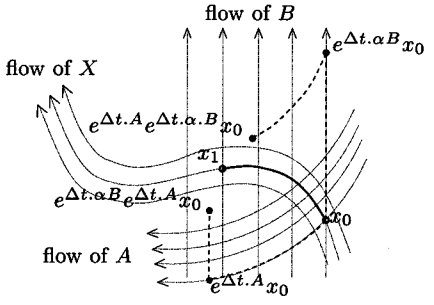

Figure 2: First-order predictions

Such an approximation scheme can be easily implemented as a multilayer composition network (Fig. 3). The problem of predicting the system can now simply be solved by minimizing the prediction error of the model by tuning the weights w_{ij} using some gradient-based optimization technique.

The most important feature of this method is that we are able to use the a priori knowledge we have about the system. This is in contrast with neural networks where it is difficult to incorporate a priori knowledge as it is a black-box model.

3. MLP in dynamics space

If we decide to approximate the vector field of our system by a multi-layer perceptron, one can also derive a composition network. The system is of the form $\dot{x} = X(x)$ where:

$$X(x) = \sum_{j=1}^{n} \vec{c}_j.\sigma(\vec{b}_j.x + b_{j_0}) = \sum_{j=1} A_{\vec{c}_j,\vec{b}_j}(x) \tag{8}$$

where $\sigma(x) = \tanh(x)$ and $\vec{c}_j, \vec{b}_j \in \mathbb{R}^{d+1}, j = 1, ..., n$.

The differential equation $\dot{x} = \sigma(x)$ can be solved explicitly in the one-dimensional case. We can use this to explicitly integrate the multidimensional

ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 303-308

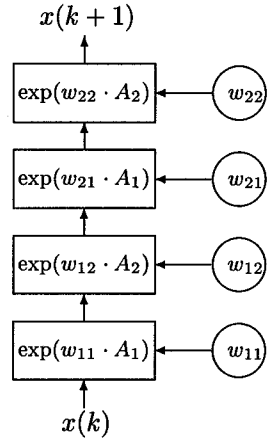

Figure 3: Implementation of the composition network.

system $\dot{x} = A_{\vec{c},\vec{b}}(x)$ for any value of the parameters $\vec{c}_j, \vec{b}_j$. So, we can design a network of the following form:

$$\hat{x}(k+1) = F(x(k)) = e^{w_k . A^{\vec{c}_k, \vec{b}_k}} \dots e^{w_1 . A^{\vec{c}_1, \vec{b}_1}} x(k) \tag{9}$$

We call this an "MLP in dynamics space" as the MLP is implicitly used to parametrize the vector field. As the MLP possesses universal approximation properties in the space of vector fields, the "MLP in dynamics space" will have universal approximation properties in the space of dynamical systems.

4. Prediction of the Lorenz attractor

To demonstrate the efficiency of this method, we build a network and train it for the prediction of the Lorenz attractor. The Lorenz attractor is of the form:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} -\beta & 0 & 0 \\ 0 & -\sigma & \sigma \\ 0 & \rho & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} xy \\ 0 \\ -yz \end{bmatrix} = C_{\beta,\rho,\sigma} + D \tag{10}$$

The linear vector field $C_{\beta,\rho,\sigma}$ can be solved for any values of β, ρ, σ by exponentiation of the matrix. The nonlinear vector field D can also be solved analytically. The training data consists in a set of 1000 data points $x(k), y(k), z(k)$ on a trajectory sampled at time-interval $\Delta t = 0.01$. The parameters β, ρ, σ have values equal to respectively $8/3$, 28 and 10. To build a predictor we consider the parameters of the Lorenz attractor β, ρ, σ and the sampling-time Δt to be unknown. We compare a first-, second- and fourth-order method [8].

ESANN'1996 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-25-26 April 1996, D-Facto public., ISBN 2-9600049-6-5, pp. 303-308

Table 1: Performance of composition networks

	SSE	$\hat{\beta}$	$\hat{\rho}$	$\hat{\sigma}$	$\hat{\Delta t}$
first-order	20.6819	3.1344	10.4087	27.5655	0.0092
second-order	0.0177	2.6788	10.0767	28.1132	0.0100
fourth-order	$2.09.10^{-5}$	2.6673	10.0019	28.0072	0.0100

We can see that the error (SSE sum of squared errors) can rapidly be made small by increasing the order of the model. This guarantees an accurate estimation of the parameters and a parcimonious parametrization of the model.

5. Conclusion

We have presented a multi-layer architecture for the prediction of dynamical systems from discrete-time measurements. The method was derived from the theory of Lie algebras. We presented an implementation which made use of a priori knowledge about the system. We successfully applied it to the prediction of the Lorenz attractor. We also presented a more general implementation having a universal approximation property in the space of dynamical systems.

References

[1] V. Arnold. *Mathematical methods of classical mechanics.* Springer-Verlag, New York, NY, 1989.

[2] E. Forest and R. Ruth. Fourth-order symplectic integration. *Physica D*, 43:105–117, 1990.

[3] A. Isidori. *Nonlinear control theory.* Springer-Verlag, New York, NY, 1989.

[4] P.-V. Koseleff. *Calcul formel pour les méthodes de Lie en mécanique hamiltonienne.* PhD thesis, Ecole Polytechnique, Paris, 1993.

[5] R. I. McLachlan. On the numerical integration of ordinary differential equations by symmetric composition methods. *SIAM J. Sci. Comput.*, 16(1):151–168, 1995.

[6] Y. Moreau and J. Vandewalle. Dynamical system prediction: a lie algebraic approach for a novel neural architecture. To appear in *Annals of Mathematics and Artificial Intelligence*, 1996.

[7] M. Suzuki and T. Yamauchi. Convergence of unitary and complex decompositions of exponential operators. *J. Math. Phys.*, 34(10), October 1993.

[8] H. Yoshida. Construction of higher order symplectic integrators. *Phys. Lett. A.*, 150:262–269, 1990.