



HAL
open science

Learning Constrained Dynamical Embeddings for Geophysical Dynamics

Said Ouala, Steven L Brunton, Duong Nguyen, Lucas Drumetz, Ronan Fablet

► **To cite this version:**

Said Ouala, Steven L Brunton, Duong Nguyen, Lucas Drumetz, Ronan Fablet. Learning Constrained Dynamical Embeddings for Geophysical Dynamics. CI 2019 : 9th International Workshop on Climate Informatics, 2019, Paris, France. hal-02285700

HAL Id: hal-02285700

<https://imt-atlantique.hal.science/hal-02285700>

Submitted on 18 Nov 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



LEARNING CONSTRAINED DYNAMICAL EMBEDDINGS FOR GEOPHYSICAL DYNAMICS

Said Ouala¹, Steven L. Brunton², Duong Nguyen¹, Lucas Drumetz¹ and Ronan Fablet¹

Abstract—In this work, we investigate the implementation of physical constraints for the regularization of linear quadratic dynamical representations of partially observed systems. We focus on energy preserving quadratic terms and propose to enforce this constraint within the learning criterion of the models. We further demonstrate on the Lorenz 63 system that the generalization performance is significantly improved to states beyond the attractor spanned by the observation data when this constraint is satisfied.

I. INTRODUCTION

Recent advances in data driven modeling, especially in optimization techniques, machine learning and neural networks address the learning of data-driven representations of dynamical systems as relevant alternatives to model driven strategies for applications ranging from system identification [1], forecasting [2], reconstruction [3] and control [4]. When considering observation data issued from an a priori complex field as encountered in ocean, atmosphere and climate science, these powerful tools should be considered with care to account for the proper specifications of the underlying dynamics. For instance, when considering the data-driven identification of an Ordinary Differential Equation (ODE) from a set of observations $\mathbf{x}_t \in \mathbb{R}^n$, where $t \in \{t_0, \dots, T\}$ is the temporal sampling and n the dimension of our observation space, the first question to answer is the existence (or not) of an appropriate ODE mapping in the observation space. For fully-observed systems, *i.e.* when observed variables \mathbf{x}_t are governed by an ODE or are related to some underlying states \mathbf{z}_t that are governed by an ODE according to a diffeomorphic mapping, recent advances [1], [5], [6] have shown that one can identify the governing equations of the dynamics of \mathbf{z} from a representative dataset of observations $\{\mathbf{x}_{t_i}\}_i$. However, in the more general cases, it

is more likely that our observations depend (possibly in a non-linear fashion) on unobserved latent variables that make the underlying dynamical model evolve in a higher dimensional space \mathbb{R}^s with $s > n$. Under the assumption that the relationship between the observed and unobserved variables can not be decoupled, it is rigorously impossible to find an appropriate *one-to-one* mapping governed by an ODE in the observation space \mathbb{R}^n . In the latter case, classical approaches do not apply since no ODE or, more generally, no one-to-one mapping defined in the observation space can represent the time evolution of the observations.

In this context, Takens's theorem states the conditions under which a delay embedding representation guarantees the existence of governing equations in the embedded space [7]. This technique was initially used as a geometrical reconstruction technique of the higher dimensional unobserved limit-cycle. The derivation of a dynamical system from such a representation, on the other hand, encountered large disparities since no explicit relationships between the defined phase space and an ODE formulation have been clearly identified.

The identification of an embedding of the observations parametrized by an ODE was proposed in [8] and appears to be an interesting trade-off between reconstructing the phase space of the unseen dynamical system and forecasting the observations through the parametric ODE. However, this formulation is very limited when considering generalization issues above the limit-cycle described by the observations. From a topological point of view, and without loss of generality, one can expect the ODE representation to i) be bounded, ii) only include the limit-cycle describing the observations in a higher dimensional space with a reasonable attracting region. Unfortunately, those characteristics relate to some physical constraints that define trapping regions of limit-cycles. The optimization criterion as proposed in [8] does not guarantee those elementary constraints which severely affects the generalization quality of the models. In this work, we propose a new implementation of the learning algorithm that allows to enforce prior

Corresponding author: S. Ouala, said.ouala@imt-atlantique.fr
¹IMT-Atlantique, Lab STICC, UMR CNRS 6285, F-29238, France,
²Department of Mechanical Engineering, University of Washington, Seattle, WA 98195, USA



knowledge such as physical constraints. We focus on energy preserving non-linearities and illustrate on a toy model whose the long term boundedness and the attracting region of the revealed limit cycle are highly influenced when this type of constraints are satisfied. Regarding the data driven identification of climate and ocean dynamics, we believe that this work provides an initial playground for learning consistent models in terms of long term forecast through the implementation of physical constraints issued from prior knowledge of the conservation laws governing the dynamics.

This paper is organized as follows. Section II reviews the Neural Embedding for Dynamical Systems technique as proposed in [8]. Section III introduces the new optimization criterion of the model which includes energy preservation constraints. The numerical experiments are presented in Section IV. We further discuss our contributions in section V.

II. NEURAL EMBEDDING OF DYNAMICAL SYSTEMS

This section summaries the Neural Embedding of Dynamical Systems —NbedDyn— proposed in [8].

Let us Consider a dynamical system governed by and autonomous ODE:

$$\dot{\mathbf{z}}_t = f_H(\mathbf{z}_t) \quad (1)$$

For most applications, The true state $\mathbf{z}_t \in \mathbb{R}^s$ of the system is unknown and we are only provided a series of observations $\{\mathbf{x}_t\}$:

$$\mathbf{x}_t = h(\mathbf{z}_t) \quad (2)$$

Where $h : \mathbb{R}^s \rightarrow \mathbb{R}^n$ is an observation operator that does not satisfy the conditions [9] under which the predictable deterministic dynamics expressed in the space of \mathbf{z} is still deterministic in the observation space.

The NbedDyn technique tackles this problem by searching an augmented latent space, where the latent states are governed by diffeomorphic flows and can be mapped to the observations \mathbf{x}_t . For any given operator h of a deterministic dynamical system, Takens's theorem [7] guarantees that such augmented space exists. However, instead of using a delay embedding, NbedDyn defines a d_E -dimensional augmented latent space with states ($d_E > n$) $\mathbf{X}_t \in \mathbb{R}^{d_E}$ as follows:

$$\mathbf{X}_t^T = [\mathbf{x}_t^T, \mathbf{y}_t^T] \quad (3)$$

where $\mathbf{y}_t \in \mathbb{R}^{d_E-n}$ presents the information of the unobserved components of the true latent state \mathbf{z}_t .

The corresponding dynamics and observation operator are defined as:

$$\dot{\mathbf{X}}_t = f_\theta(\mathbf{X}_t) \quad (4)$$

$$\mathbf{x}_t = G(\mathbf{X}_t) \quad (5)$$

where the dynamical operator f_θ belongs to a family of operators parametrized by a parameter vector θ . Using an integration scheme, we can associate f_θ with an one-step-ahead diffeomorphic mapping:

$$\Phi_{\theta,t}(\mathbf{X}_{t-1}) = \mathbf{X}_{t-1} + \int_{t-1}^t f_\theta(\mathbf{X}_{t-1}) \quad (6)$$

From Eqs. (4), (5) and (6), we have a state space model:

$$\begin{cases} \mathbf{X}_t = \Phi_{\theta,t}(\mathbf{X}_{t-1}) \\ \mathbf{x}_t = G(\mathbf{X}_t) \end{cases} \quad (7)$$

Given an observation time series $\{\mathbf{x}_0, \dots, \mathbf{x}_T\}$, the Neural Embedding of Dynamical Systems model minimizes the forecasting error of the observations with respect to the model parameters and the augmented states as follow

$$\begin{aligned} \hat{\theta}, \mathbf{y}_{1:T} = \arg \min_{\theta} \min_{\{\mathbf{y}_t\}_t} & \sum_{t=1}^T \|\mathbf{x}_t - G(\Phi_{\theta,t}(\mathbf{X}_{t-1}))\|^2 \\ & + \lambda \|\mathbf{X}_t - \Phi_{\theta,t}(\mathbf{X}_{t-1})\|^2 \end{aligned} \quad (8)$$

with λ a trade-off parameter.

The ODE operator f_θ is stated as a linear quadratic neural network and the corresponding flow map $\Phi_{\theta,t}$ is a neural network based on a numerical integration scheme formulation (typically a 4th-order Runge-Kutta scheme).

III. CONSTRAINED DYNAMICAL EMBEDDING

The dynamical model f_θ is expressed as a linear quadratic model. This particular architecture is suitable for the identification of reduced order models of incompressible flows as it can be seen as a low dimensional approximation of the Navier-Stokes equation. Formally, we can formulate the operator f_θ as follow

$$\dot{\mathbf{X}}_i = c_i + \sum_{j=1}^{d_E} l_{i,j} \mathbf{X}_j + \sum_{j=1}^{d_E} \sum_{k=1}^{d_E} b_{i,j,k} \mathbf{X}_j \mathbf{X}_k \quad (9)$$

with $c_i, l_{i,j}, b_{i,j,k}$ are the trainable coefficients of the dynamical operator (θ), the time index t of \mathbf{X} is omitted for simplicity.

Regarding the data-driven identification of the parameters θ , the minimization of the cost function in (8) does not guarantee f_θ to satisfy elementary conservation constraints present in the true underlying system which severely affects the generalization performance of the model. This is a classical issue in most data driven representations. If the provided data is not big enough for the model to learn these constraints one should explicitly enforce them within the optimization criterion



of the model. Here, we will focus on energy preserving non-linearities as this constraint is known to help long-term boundness of reduced order models of non-compressible flow [10], however, the general framework proposed here applies to any prior knowledge (known coefficients, existing symmetries ...etc.) on this specific parametrization of the network f_θ .

Let us consider the evolution of the fluctuation energy $K = \frac{1}{2} \sum_{i=1}^{d_E} X_i^2$ of the system described by f_θ . The time derivative of this quantity can be written as:

$$\begin{aligned} \dot{K} &= [\nabla_X \mathbf{K}]^T \dot{\mathbf{X}} = \sum_{i=1}^{d_E} \mathbf{X}_i f_{\theta,i} \\ &= \sum_{i=1}^{d_E} c_i \mathbf{X}_i + \sum_{i,j=1}^{d_E} l_{i,j} \mathbf{X}_i \mathbf{X}_j + \sum_{i,j,k=1}^{d_E} b_{i,j,k} \mathbf{X}_i \mathbf{X}_j \mathbf{X}_k \end{aligned} \quad (10)$$

An energy preserving quadratic non linearity satisfies the constraint:

$$\sum_{i,j,k=1}^{d_E} b_{i,j,k} \mathbf{X}_i \mathbf{X}_j \mathbf{X}_k = 0 \quad (11)$$

i.e. the contribution of the quadratic terms of f_θ to the fluctuation energy should sum up to zero. In this case, the quadratic coefficients are responsible for redistributing the perturbation energy in directions of positive and negative energy growth that are defined by the eigenvalues of the matrix $(l_{i,j})$, $i, j = 1, \dots, d_E$ [11]. It can be shown that for constraint (11) to hold, the sums of the quadratic coefficients over index permutations must be zero:

$$b_{i,j,k} + b_{i,k,j} + b_{j,i,k} + b_{j,k,i} + b_{k,i,j} + b_{k,j,i} = 0, \quad (12)$$

$i, j, k = 1, \dots, d_E$

From the above definition of the energy-preserving non-linearity and the corresponding energy function, one can think of two distinct ways to enforce this constraint on the approximate model, either through enforcing the constraint described in (12) as a penalty term over the quadratic coefficients in the loss function (8) or penalizing the quadratic energy expressed by (11). The latter implementation of the constraint is avoided in this work since the loss function is optimized with respect to both the parameters of the model and the latent states. This will result in an expression (11) that is not necessarily minimized due to the constraint over the quadratic weights (12) as the latent states in (11) are also trained to minimize the quadratic energy. This represents an issue in the sense that our energy preserving constraint will depend on the latent states and thus, for large deviations from the spanned

manifold, this constraint will no longer be satisfied. Finally, the following criterion is considered:

$$\begin{aligned} \hat{\theta}, \mathbf{y}_{1,\dots,T} &= \arg \min_{\theta} \min_{\{\mathbf{y}_t\}_t} \sum_{t=1}^T \|\mathbf{x}_t - G(\Phi_{\theta,t}(\mathbf{X}_{t-1}))\|^2 \\ &\quad + \lambda \|\mathbf{X}_t - \Phi_{\theta,t}(\mathbf{X}_{t-1})\|^2 \\ \text{s.t. } &\{ b_{i,j,k} + b_{i,k,j} + b_{j,i,k} + b_{j,k,i} + b_{k,i,j} + b_{k,j,i} = 0 \end{aligned} \quad (13)$$

the constrained optimization problem is solved by using the equality constraint as a penalty term in the loss function.

IV. NUMERICAL EXPERIMENTS

Considered system : Lorenz-63 dynamical system is a 3-dimensional model governed by the following ODE:

$$\begin{cases} \frac{dz_{t,1}}{dt} = \sigma(z_{t,2} - z_{t,1}) \\ \frac{dz_{t,2}}{dt} = \rho z_{t,1} - z_{t,2} - z_{t,1} z_{t,3} \\ \frac{dz_{t,3}}{dt} = z_{t,1} z_{t,2} - \beta z_{t,3} \end{cases} \quad (14)$$

Under parametrization $\sigma = 10$, $\rho = 28$ and $\beta = 8/3$, this system involves chaotic dynamics with a strange attractor [12].

We simulate Lorenz-63 state sequences using the LOSDA ODE solver [13] with an integration step of 0.01. We assume that only the first Lorenz-63 variable is observed $\mathbf{x}_t = \mathbf{z}_{t,1}$. We apply the proposed framework to this experimental setting using a training sequence of 10000 time-steps.

Proposed model : Regarding the proposed framework, we tested the model for a dimension of the latent space equal to 3. The neural-network parametrization for operator f_θ is a simple linear quadratic model. We compare in this work the model optimized using the initial criterion (8) as proposed in [8] and the new criterion with the energy preserving constraint (13).

Forecasting performances of the proposed data-driven models: We further evaluate the performances of the learning criterion based on the comparison of the forecasted limit-cycles. Table I reports the Lyapunov spectrum and the Lyapunov dimension of the data driven models of the proposed NbedDyn representation compared with the true spectrum and dimension of the Lorenz 63 system. As demonstrated in [8], when the initial condition is inside the spanned manifold of the augmented states, the dynamical model optimized using criterion (8) gives trajectories that are bounded (the sum of the Lyapunov exponents is negative) and with topological characteristics that are very similar to the true Lorenz 63 model. However, When the initial condition is far from the spanned manifold, the model optimized by the equation (8) diverges to

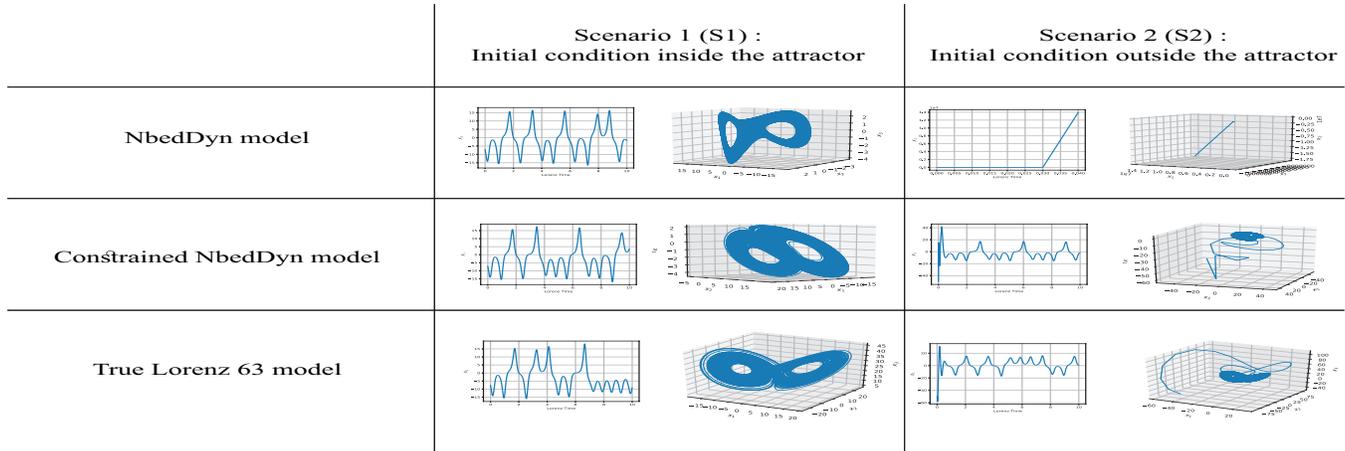


Fig. 1: *Forecasting performances of the data driven models under different initial conditions: first row, NbedDyn model as proposed in [8]; second row, proposed constrained NbedDyn model; third row, True Lorenz 63 model.*

infinity (the sum of the Lyapunov exponents is positive). From a machine learning perspective, this is the direct consequence of a poor generalization performance to states that are far from the attractor spanned by the training data. From a dynamical systems point of view, our model contains several attracting regions of chaotic and unstable solutions and when the initial condition is far from the spanned attractor, the state evolution is dominated by positive energy growth which makes our model diverge to infinity. The constrained model in the other hand, satisfies elementary preservation constraints that are present in the actual Lorenz 63 system and leads to a much more stable behavior with a larger attracting region of the chaotic limit-cycle.

Qualitative analysis of the proposed schemes: We also illustrate these conclusions through the forecasting examples in Figure 1. When starting from an initial condition inside the attractor, both the NbedDyn and the Constrained NbedDyn models end up with a forecasted limit cycle that is similar to the true Lorenz attractor. When starting from an initial condition that is far from the spanned attractor, the classical NbedDyn as proposed in [8] diverge to infinity. By contrast, adding energy preserving constraints to the model significantly improves the generalization performances to states beyond the attractor spanned by the training data.

V. CONCLUSION

In this work, we address the data-driven identification of dynamical representations of partially observed systems. We propose to include physical constraints to the data driven models as prior knowledge of the dynamics. The reported forecasting performance for Lorenz-63 dynamics illustrates clearly the importance

Model	Exponents	Dimension
NbedDyn	S1 (0.889, 0.0, -14.21)	2.063
	S2 NaN	NaN
Constrained NbedDyn	S1 (0.811, 0.0, -12.66)	2.064
	S2 (0.828, 0.0, -12.67)	2.064

TABLE I: *Forecasting performance the data-driven models: full Lyapunov spectrum and Lyapunov dimension of the NbedDyn model as proposed in [8] and with the additional energy preserving constraint on the quadratic terms of f_θ . The simulation S1 is carried with respect to an initial condition inside of the spanned attractor of the augmented states \mathbf{x}_t , the simulation S2 is performed with respect to an initial condition far from the attractor. The Lyapunov spectrum of the true Lorenz 63 system is (0.91, 0.0, -14.57) and it's dimension is estimated to be 2.064 [14]*

of such an approach as only enforcing energy preserving nonentities constraints significantly improves the generalization performances of the model far from the attractor spanner by the training data.

ACKNOWLEDGMENTS

This work was supported by GERONIMO project (ANR-13-JS03-0002), Labex Cominlabs (grant SEACS), CNES (grant OSTST-MANATEE), Microsoft (AI EU Ocean awards) and by MESR, FEDER, Région Bretagne, Conseil Général du Finistère, Brest Métropole and Institut Mines Télécom in the framework of the VIGISAT program managed by "Groupement Bretagne Télédétection" (BreTel).

REFERENCES

- [1] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Discovering governing equations from data by sparse identification of nonlinear dynamical systems," *Proceedings of the National Academy of Sciences*, vol. 113, pp. 3932–3937, Apr. 2016.



- [2] A. Braakmann-Folgmann, R. Roscher, S. Wenzel, B. Uebbing, and J. Kusche, “Sea Level Anomaly Prediction using Recurrent Neural Networks,” *arXiv:1710.07099 [cs]*, oct 2017. arXiv: 1710.07099.
- [3] H. C. C. B. P. A. C. F. G. L. Ouala Said, Fablet Ronan, “Neural network based kalman filters for the spatio-temporal interpolation of satellite-derived sea surface temperature,” 2018.
- [4] S. L. Brunton, J. L. Proctor, and J. N. Kutz, “Sparse identification of nonlinear dynamics with control (sindyc),” *IFAC-PapersOnLine*, vol. 49, no. 18, pp. 710–715, 2016.
- [5] R. Fablet, S. Ouala, and C. Herzet, “Bilinear residual neural network for the identification and forecasting of geophysical dynamics,” in *2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 1477–1481, Sep. 2018.
- [6] D. Nguyen, S. Ouala, L. Drumetz, and R. Fablet, “Em-like learning chaotic dynamics from noisy and partial observations,” *SciRate*, Mar. 2019.
- [7] F. Takens, “Detecting strange attractors in turbulence,” in *Dynamical Systems and Turbulence, Warwick 1980* (D. Rand and L.-S. Young, eds.), (Berlin, Heidelberg), pp. 366–381, Springer Berlin Heidelberg, 1981.
- [8] S. Ouala, D. Nguyen, L. Drumetz, B. Chapron, A. Pascual, F. Collard, L. Gaultier, and R. Fablet, “Learning Latent Dynamics for Partially-Observed Chaotic Systems,” *arXiv e-prints*, p. arXiv:1907.02452, Jul 2019.
- [9] T. Sauer, J. A. Yorke, and M. Casdagli, “Embedology,” *Journal of Statistical Physics*, vol. 65, pp. 579–616, Nov 1991.
- [10] J.-C. Loiseau and S. L. Brunton, “Constrained sparse galerkin regression,” *Journal of Fluid Mechanics*, vol. 838, pp. 42–67, 2018.
- [11] M. Schlegel and B. R. Noack, “On long-term boundedness of galerkin models,” *Journal of Fluid Mechanics*, vol. 765, pp. 325–352, 2015.
- [12] E. N. Lorenz, “Deterministic Nonperiodic Flow,” *Journal of the Atmospheric Sciences*, vol. 20, pp. 130–141, Mar. 1963.
- [13] A. C. Hindmarsh, “ODEPACK, a systematized collection of ODE solvers,” *IMACS Transactions on Scientific Computation*, vol. 1, pp. 55–64, 1983.
- [14] J. C. Sprott, *Chaos and Time-Series Analysis*. New York, NY, USA: Oxford University Press, Inc., 2003.