Uncovering differential equations

Agustín Somacal[1], Matthieu Jonckheere[2], Vincent Lefieux[3], Dominique Picard[4] and Ezequiel Smucler[1,5]

[1]Aristas S.R.L., [2]Instituto de Cálculo-CONICET, Universidad de Buenos Aires, [3]RTE, [4]Laboratoire de Probabilités et Modéles aléatoires Université Paris VII, [5]Universidad Torcuato Di Tella

a.somacal@aristas.com.ar

Abstract

Many branches of science and engineering require differential equations to model the dynamics of the systems under study. Traditionally, the identification of the appropriate terms in the equation has been done by experts. [2], [1] developed a method to automate this task using the data itself. In this work, we extend the applicability of this method to situations where not all variables are observed by adding higher-order derivatives to the model space search. We test the approach with known chaotic dynamical systems like the Rosseler-1974 and show that by including higher order time derivatives it is possible to obtain a differential equation which generates phase diagrams with a geometry equivalent to the original series.

Introduction

Problem setting: given empirical data drawn from a dynamical system, represented by functions $f_1(x,y,z,t)\dots f_H(x,y,z,t)$, our objective is to find a set of differential equations (1) that explains the relationship between the measurements, even though some of the functions may be unobserved.

$$\mathbb{D}\mathbf{f} = \mathbf{U}(\mathbf{f}, \mathbb{E}\mathbf{f}), \tag{1}$$

where $\mathbf{f} = (f_1,\dots,f_H)$, $\mathbb{D},\mathbb{E}$ are differential operators in space and time (x,y,z,t) variables and $\mathbf{U}:\mathbb{R}^H\to\mathbb{R}^H$ is an unknown map.

Proposal: when not all the variables can be measured, we propose to enhance the approach outlined by [1] and [2] by including higher order derivatives to account for the missing information.

Method: choose a large dictionary of functions (polynomial powers of f_h and its derivatives up to $n-1$, $h=1,\dots,H$, spatial derivatives of $f_1\dots,f_H$ and trigonometric functions of t, and so on)

$$\mathcal{A} = (A_1,\dots,A_p)$$

$$\mathcal{A}_{example} = \left\{x, y, z, x^2, xy, \dots, z^2, \frac{\partial f(x,y,z,t)}{\partial x}, f\frac{\partial f}{\partial t}, \frac{\partial^2 f}{\partial z^2}, sin(t), t, ty\right\}$$

and regress over the discretizations of higher order time derivatives

$$\frac{\partial^n f_1}{\partial t^n},\dots,\frac{\partial^n f_H}{\partial t^n}$$

Since we look for sparse representations, the Lasso [3] regression technique is used to find the combination of the elements of the dictionary of functions that adequately explains the behaviour of the response variables. That is, for all h such that f_h is observable,

$$\mathbf{c}_h^* = \arg\min_{\mathbf{c}\in\mathbb{R}^p} \sum_{i,j,k,l}\left(\frac{\partial^n f_h(x_i,y_j,z_k,t_l)}{\partial t^n} - \sum_{g=1}^{p} c_{g,h}A_g(x_i,y_j,z_k,t_l)\right)^2 + \lambda\left\|\mathbf{c}_h\right\|_1, \tag{2}$$

where $\mathbf{c}_h^* = (c_{h,1},\dots,c_{h,p})$ are the regression coefficients.

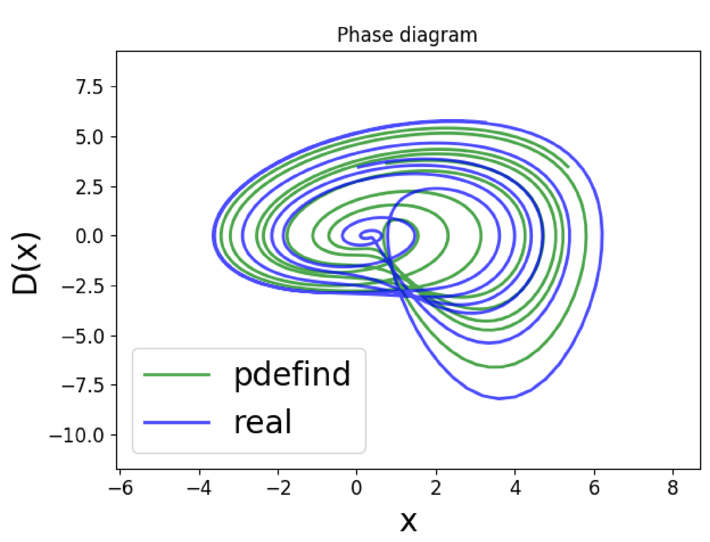
Dealing with unobserved data

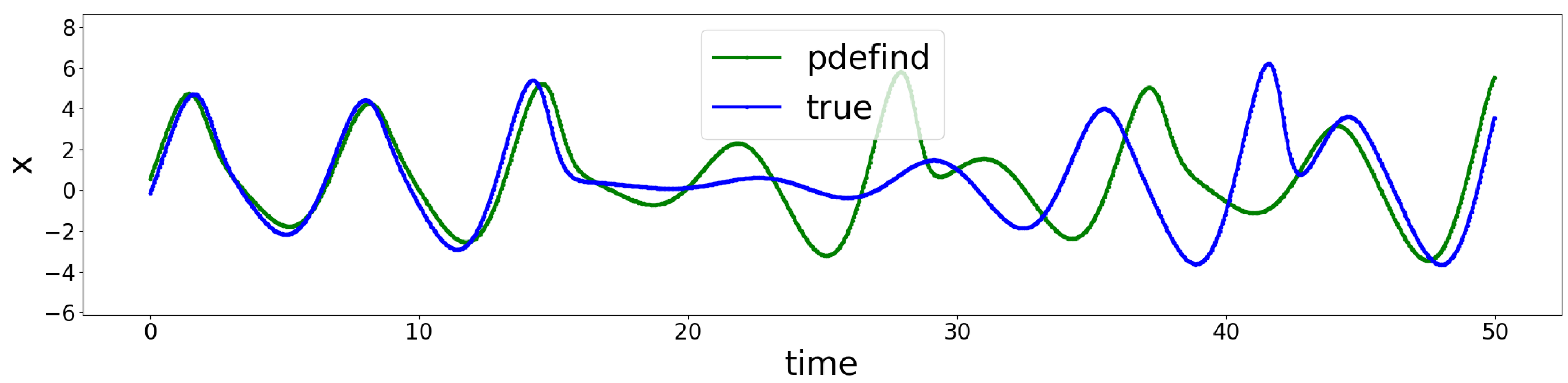
The proposed approach leads, in many cases, to equivalent systems that capture most of the underlying dynamics In the plots below phase diagrams and time series of the true dynamics (blue) and the integrated dynamics (green) of the differential equations found are shown for the variables x and y of the Rosseler system,

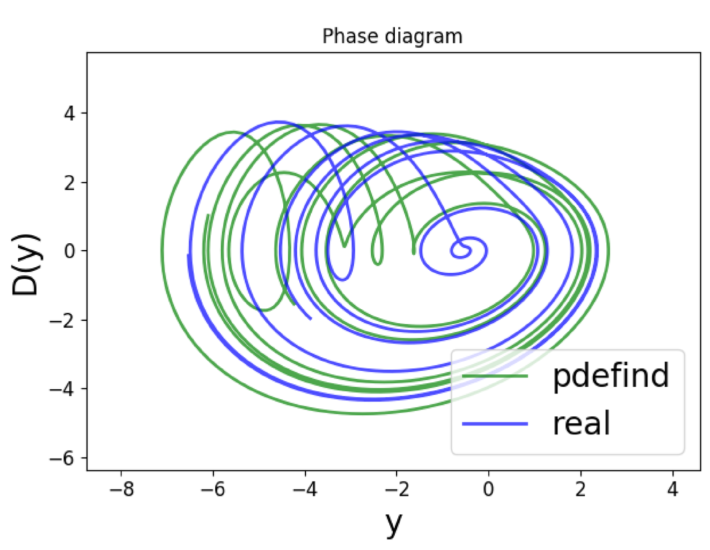
$$\frac{dx}{dt} = -z - y; \qquad \frac{dy}{dt} = 0.52y + x; \qquad \frac{dz}{dt} = 2 - 4z + xz.$$

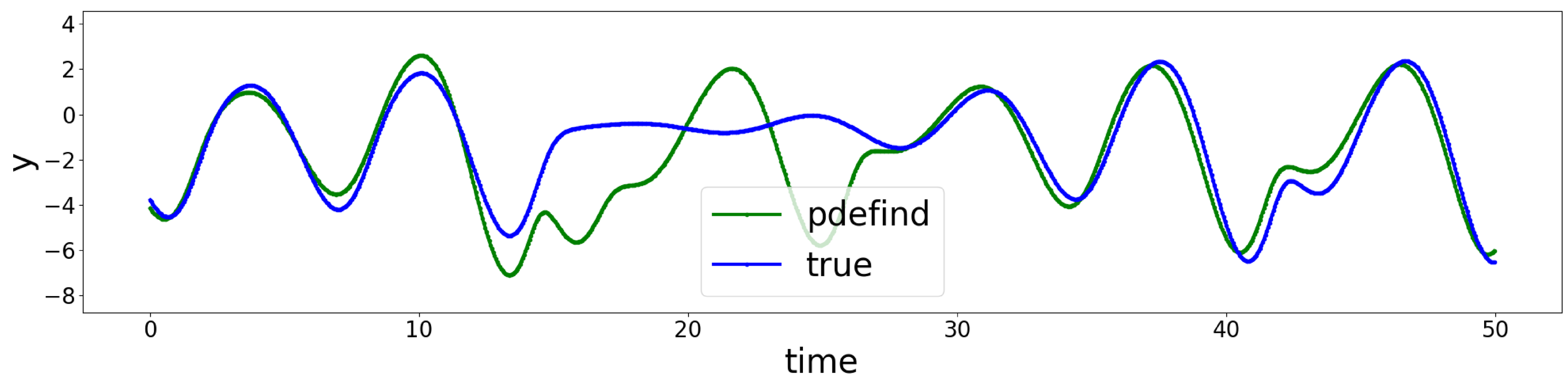
The equations found using only one variable in each case, though having many terms, are able to reproduce the geometry of the trajectories in the phase diagram.

$$\frac{d^3x}{dt^3} = \frac{1}{10}\left(0.9 - 41x + 5x^2 + x^3 - 8x\frac{dx}{dt} + 6x\frac{d^2x}{dt^2} + x\frac{dx}{dt}^2 - 2x^2\frac{dx}{dt} + x^2\frac{d^2x}{dt^2} + 11\frac{dx}{dt} + 3\frac{dx}{dt}^2 - 3\frac{dx}{dt}\frac{d^2x}{dt^2} - 34\frac{d^2x}{dt^2}\right)$$









Bibliography

References

[1] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. "Discovering governing equations from data by sparse identification of nonlinear dynamical systems". In: *Proceedings of the National Academy of Sciences* (2016). DOI: 10.1073/pnas.1517384113.

[2] Samuel H. Rudy et al. "Data-driven discovery of partial differential equations". In: *Science Advances* (2017). DOI: 10.1126/sciadv.1602614.

[3] Robert Tibshirani. "Regression shrinkage and selection via the lasso: a retrospective". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* (2011).