71st International Astronautical Congress (IAC), The CyberSpace Edition, 12-14th October 2020.

IAC–20–C1,6,5,x59118

ORBITAL ANOMALY RECONSTRUCTION USING DEEP SYMBOLIC REGRESSION

Matteo Manzi[a*], Massimiliano Vasile[b]

[a] *Mechanical and Aerospace Department, University of Strathclyde, 75 Montrose Street, Glasgow, United Kingdom G1 1XJ*, matteo.manzi@strath.ac.uk
[b] *Mechanical and Aerospace Department, University of Strathclyde, 75 Montrose Street, Glasgow, United Kingdom G1 1XJ*, massimiliano.vasile@strath.ac.uk
[*] Corresponding author

Abstract

This work explores the combination of Sparse and Symbolic Regression, here called Deep Symbolic Regression, for the autonomous reconstruction of orbital anomalies. Orbital anomalies are detectable deviations from the state of an object that can be predicted from the propagation of some observable initial conditions. We contemplate anomalies that can derive from unmodelled natural phenomena or from intentional and unintentional orbital manoeuvres: an accurate modelling of atmospheric density fluctuations, for example, allows informing the space weather.
Leveraging the powerful modelling capacity of symbolic regression and the sparse representability of dynamical systems in orbital mechanics , the proposed approach allows one to generate a symbolic representation of orbital anomalies from state observations only. In other words, we use sparse measurements of position and velocity, in general associated with uncertainty, to derive a symbolic representation of the unmodelled part of the dynamics that can explain the deviations of the propagated states.
The advantage of such an approach, compared to more traditional filtering techniques, is twofold: it provides an explicit analytical representation of the phenomenon causing the anomaly and it provides a better long term prediction of the dynamics of the object under consideration. The use of Deep Symbolic Regression outdoes more traditional Genetic Programming-based approaches in that it is less prone to overfitting and far less computationally expensive. The explicit dependence, with respect to time, of the symbolic representation, allows one to indirectly model the evolution of unobservable states, whose behaviour can be later inferred from the analysis of the estimated equation itself. The proposed approach yields solutions that are robust against measurement noise: its estimation can be integrated into the derivation of the missing part of the dynamics.
The performance of the Deep Symbolic Regression will be assessed against a number of case studies, with a focus on the interpretability of the obtained solutions, demonstrating the performance of this new tracking data-based algorithm.
Keywords: Anomaly Reconstruction, Manoeuver Detection, Data-driven Orbital Mechanics, Sparse Regression, Symbolic Regression

1. Introduction

The context of data-driven dynamical systems is characterized by the goal of automating the discovery of governing equations from data, for example with Machine Learning techniques, by means of constructing interpretable, white-box models with both generalizing and extrapolating capabilities. We here focus on *data-driven Orbital Mechanics*; in fact, the field of Orbital Dynamics has a rich history in motivating the development of techniques in dynamical systems theory, data-driven ones included. For example, techniques for system identification from data have been developed in order to model the dynamics of the Hubble space telescope and the International Space Station [4].

In order to do this, Symbolic Regression, combined with Genetic Programming [21], has been one of the first successful techniques in determining the underlying structure of nonlinear dynamical systems from data ([2], [37]): such an approach has been shown to be successful in reconstructing common differential equations appearing in orbital mechanics [28].

Genetic Programming-based Symbolic Regression is however computationally expensive and prone to overfitting, motivating its combination with physically inspired approaches [4]. In fact, we are here generalizing the problem statement given in [28], aiming at efficiently modelling orbit anomalies (i.e., detectable deviations of

the state of an object w.r.t. an expected value) arising both from natural phenomena and human intervention. This is needed as, in order to govern the rapid changes in the space environment due to the launch of NewSpace systems, space operations timelines need to be updated [31]. In fact, more accurate dynamical models are needed in order to better predict and therefore avoid collisions.

A number of recent works have focused on informing with physical principles, such as dimensional analysis, Machine Learning techniques ([38], [15], [10]), while others combined Deep Learning with more traditional Machine Learning techniques, such as Symbolic Regression ([11], [22], [17], [32]*); in fact, the computer science community is focused on developing techniques for Anomaly Detection [35], and a lot of this can be applied for aerospace problems, in particular in Orbital Mechanics [39].

A number of works ([24], [25], [26]) aimed at developing an estimator, similar to the Kalman Filter and incorporating optimal control policies, to be used to reconstruct mismodeled dynamical parameters and perform manoeuver detection using satellite tracking data. However, this approach does not abstract the estimated state into a symbolic equation, which can be subsequently analyzed, in order to infer the causal nature of the anomaly; also, it often requires an a priori knowledge of the structure of the unknown perturbation.

In order to overcome these limitations, sparsity promoting techniques [4] can be applied. Such techniques, naturally compatibile with Symbolic Regression, are strictly related to Dynamic Mode Decomposition [5] and the Koopman Operator theory†, the latter being particularly relevant from an optimal control perspective in Orbital Mechanics [23].

2. Methodology

Two different and in may ways complementary approaches can be used to tackle the problem of interest. In this section the Sparse Regression and the Genetic Programming-based Symbolic Regression techniques will be presented, together with the proposed architecture combining them, here called Deep Symbolic Regression; having already applied the latter in [28], we are here generalizing such methodology.

2.1 Background

Orbit anomalies, such as optimal control manoeuvers performed by tracked object, atmospheric density fluctuations, and behaviours resulting from debris impacts, can be modelled as non-linear, non-autonomous dynamical systems:

$$\ddot{\mathbf{s}}(t) = \mathbf{h}(\mathbf{s}(t), \mathbf{v}(t), t) \quad [1]$$

It is commonly known [30] that we can, without loss of generality, construct a first order, autonomous differential equation, at the expenses of increasing the dimensionality of our system:

$$\dot{\mathbf{x}}(t) = \frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\mathbf{s}\\ \mathbf{v}\end{bmatrix} = \begin{bmatrix}\mathbf{v}(t)\\ \mathbf{h}(\mathbf{s}(t), \mathbf{v}(t), t)\end{bmatrix} = \mathbf{f}(\mathbf{x}(t), t) \quad [2]$$

from which:

$$\dot{\mathbf{y}}(t) = \frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\mathbf{x}\\ t\end{bmatrix} = \begin{bmatrix}\mathbf{f}(\mathbf{x}(t), t)\\ 1\end{bmatrix} = \mathbf{g}(\mathbf{y}(t)) \quad [3]$$

The structure of such differential equation is, in our context, only partially known: we assume that the mismatch is linear, so that we can estimate linear parameters appearing in the known terms; at the same time missing terms, additive in nature, can be fully reconstructed. In this context, we have access to a number of measurements, in general noisy, that can be arranged into a snapshot matrix:

$$\mathbf{Y} = [\mathbf{y}_1(t_1),\ \mathbf{y}_2(t_2),\ \ldots,\ \mathbf{y}_n(t_n)] \quad [4]$$

We assume that such values are extracted from tracklet-like observations, allowing to average the observed state over a small time interval (i.e., the computed state can then be associated to the mean observation time); because of this, we can also estimate the state derivative of the system associated to the mean observation time via finite differencing:

$$\dot{\mathbf{y}}(t) \approx \frac{\mathbf{y}(t + \Delta t) - \mathbf{y}(t + \Delta t)}{2\Delta t} \quad [5]$$

When dealing with noisy state measurements, it is important to avoid the noise amplification characterizing finite-difference methods, and different techniques are available ([9], [34]). This can be done, making again use of the fact that the tracklet-like observation is not simply characterized by two observation epochs, but by a sequence of observations, over which some kind of filtering technique, such as the Savitzky-Golay filter [36], can be applied.

From here, we can construct the snapshot matrix given in Equation [6]:

$$\dot{\mathbf{Y}} = [\dot{\mathbf{y}}_1(t_1),\ \dot{\mathbf{y}}_2(t_2),\ \ldots,\ \dot{\mathbf{y}}_n(t_n)] \quad [6]$$

Starting with such data, together with a partial knowledge of the system of interest, our goal is to autonomously identify anomalies in orbital mechanics. These include:

* The name of the proposed approach has been inspired by this work.

† In particular, this work is strictly related to methods aiming at reconstructing Koopman eigenfunctions from data ([27], [7]).

- optimal control manoeuvers;
- environmental anomalies (e.g., atmospheric density fluctuations caused by space weather events);
- dynamics resulting from impacts with debris.

Such behaviours are modelled via non-autonomous terms appearing in the dynamics: these can be related to time-varying unobservable states, such as the atmospheric density, the spacecraft attitute or its ballistic coefficient.

2.2 Sparse Regression

Focusing on interpretability, sparse regression aims at constructing parsimonious models, making use of sparsity-promoting techniques. More formally, it makes use of the fact that governing equations of physical systems of interest are sparse in the space of all possible functions of the state, making the governing equations *sparse* in a high-dimensional nonlinear function space [19]. This is true even for the augmented system given in Equation [3], allowing us to maintain this formulation without loss of generality.

Making use of the sparse identification of nonlinear dynamics (SINDy) algorithm, we seek to approximate the unmodeled term of the dynamics (the modelled one being $\mathbf{g_m}(\mathbf{Y})$) by the generalized linear model given in Equation [7]:

$$\boldsymbol{\delta}\dot{\mathbf{Y}} = \dot{\mathbf{Y}} - \mathbf{g_m}(\mathbf{Y}) \approx \boldsymbol{\Theta}(\mathbf{Y})\boldsymbol{\Xi} \qquad [7]$$

In it, the state-dependent matrix is constructed from an a priori selected library of candidate functions, which in general are nonlinear and are characterized by nonlinear parameters; moreover, $\boldsymbol{\Xi}$ is a sparse matrix of constant coefficients. We solve for $\boldsymbol{\Xi}$ in Equation [7] via the Sequential Thresholded Least Squares (STLSQ) algorithm, by which we iteratively cancel out the small coefficients and optimize for the remaining ones, in a least square sense.

This technique, robust with respect to measurement noise, allows one to efficiently estimate the value of the linear parameters appearing in the $\boldsymbol{\Xi}$ matrix; the method is convex and scales well with the dimensionality of the problem, as opposed to brute-force combinatorial alternatives. The limiting assumption, however, is that the a priori selected library of functions used to perform the regression has to contain the correct terms describing the dynamics. Moreover, different algorithms have been developed to deal with nonlinear parameter estimation, such as the Sparse Relaxed Regularized Regression (SR3) ([40], [8]): however, these are based on local optimality conditions and are therefore strongly dependent on the initial guess. Also, optimizing for the initial guess does not lead to global convergence, because of the nature of the fitness function associated to such routine.

Because of such limitations, we also make use of symbolic regression, here briefly outlined.

2.3 Genetic Programming-based Symbolic Regression

Symbolic regression is a method by which data are fitted using compact, closed-form analytical expressions. Via the Polish notation, such expressions can be mapped into tree-like representations (Figure 1) relating "tokens" of elementary functions, variables and constants [32].

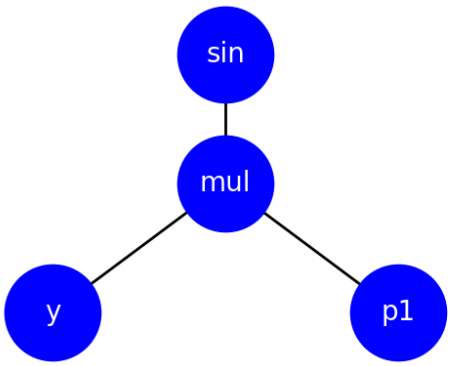


Fig. 1: Representation of the expression $\sin(p_1 \cdot y)$, as seen by the Symbolic Regression algorithm.

Because of the combinatorial nature of the optimization problem associated to identifying the best symbolic expression (i.e., the best tree), Symbolic Regression is often combined with Genetic Programming: in this way, algebraic expressions are combined stochastically.

2.3.1 Genetic Operations

Each tree can be constructed using *Strongly Typed* Genetic Programming (STGP) [13]: every function and terminal node (position, velocity, time and parameters) are assigned to a specific type. For example, the norm operator can be applied to vectors and its output is a scalar. At the same time, the output of the SR expression is forced to be a vector, whose size is given by the degrees of freedom of the problem under investigation. Trees are therefore initialized randomly satisfying these constraints, and they can be evaluated in some way; the selection process is then performed, at the beginning of each generation.

The key operations associated to the evolution, performed with a given probability, are *mutation* and *crossover*. The used mutation operations, randomly applied to modify the structure of a selected individual, are:

- Shrink;

- Insert;
- Replace node.

In the crossover step, subtrees of two different individuals are exchanged at a randomly selected location, leading to two new individuals to be then evaluated. The fitness used for this part of the algorithm makes use of the state observation spanning over the time interval of interest: it is therefore not related to the local, differential behaviour of the system, as it is the case in the sparse regression approach.

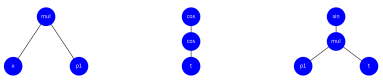
In this framework, it is possible to perform nonlinear parameter estimation, optimizing the values of the symbolic constants appearing in the expressions. However, while being naturally parallelizable (i.e., individuals of the same generation can be evaluated in parallel), the process is computationally expensive and prone to overfitting.

2.4 Deep Symbolic Regression

At this point, it should be clear why the two approaches are in many ways complementary. Before putting them together in a framework here called Deep Symbolic Regression, we further generalize the Genetic Programming-based part of the algorithm: in particular, insipred by multi-gene Genetic Programming [14], we identify an individual as a string of n symbolic expressions, from now on called sub-individuals. The genetic operations can be easily generalized, performing them on the sub-individuals identifying the $i-th$ component of the individual.

It is now possible to combine Symbolic Regression with Sparse Regression: the symbolic expressions identifying the individual can be used to construct the library of functions introduced in Equation [7]. Evaluating them against the available state observations leads to the state-dependent matrix given in Equation [8]

$$\Theta(\mathbf{Y}) = \left[\quad \right] \quad [8]$$



In this way, no a priori choice has to be made about the structure of the state-dependent matrix, since this is varying for each evaluation of the Genetic Programming part of the algorithm. Only the number of genes of the algorithm has to be identified: as long as the number of unmodeled perturbing forces is smaller than the number of genes, the sparsity-promoting section of the algorithm is in principle able to swich-off the meaningless components of the individual. Because of the same reason, together with initializing sub-individuals with a small size, overfitting is avoided: sparse regression works as a bottleneck for the complexity of the estimated solution associated to the Genetic Programming.

Moreover, the nonlinear parameters appearing inside the symbolic expressions can be estimated via an intermediate optimizer: the Symbolic Regression treats them as symbolic constants, without estimating their actual values. Finally, we are able to make use of the efficiency of sparse regression when dealing with the linear aspects of our problem, while we maintain the power of Symbolic Regression when dealing with the non-linear and combinatorial aspects of our problem.

A more detailed description of the Deep Symbolic Regression algorithm, combining pySINDy [12] with DEAP [13], is given in Algorithm [1]:

Algorithm 1 Deep Symbolic Regression

```
Initialize k individuals of n symbolic expressions
for i = 1 : n_gen do
    for j = 1 : k do
        Nonlinear parameters Optimization:
            Construct Θ(Y)
            Estimate Ξ via the STLSQ
            Compute Local Fitness
        Propagate estimated model
        Compute Global Fitness
    end for
    Perform Crossover Operations
    Perform Mutation Operations
end for
```

The main steps of Algorithm 1 are therefore the following: first, the Genetic Programming initialises and evolves, via mutation and crossover, the population of list of functions. From there, an optimization is performed to estimate the value of the set of nonlinear parameter appearing inside an individual: to do this, for each evaluation the library matrix is built and the sparse matrix estimated, leading to a set of differential equations.

The structure of such algorithm underlines how the Genetic Programming part does not deal with evaluating the value of the nonlinear parameters appearing inside the sub-individuals. These are handled with a separate global optimizer, not dealing with the stochastic combinatory search performed by the outer one. It is also worth underlining how the fitness used to evaluate the outer optimizer is different from the metric used to evaluate the sparse regression (Local Fitness): the latter only deals with local, differential aspects of the system, while the former deals with its long term behaviour. This is a further source of robustness with respect to measurement noise.

The metric function used to evaluate the performance of the regression and to compute the Local Fitness is the

coefficient of determination R^2 [12]. The Global Fitness function associated to the outer optimizer is given by:

$$\mathcal{F}_1 = \int_{t_0}^{t_f} (\mathbf{x}_{obs}(s) - \mathbf{x}_{est}(s))^2 \mathrm{d}s \qquad [9]$$

which is computed numerically, making use of the available state observations, via the Simpson's rule. A propagation is therefore necessary, to evaluate an individual. Nevertheless, we are able to efficiently perform the optimization, as the propagation is just used to optimize the structure of the symbolic expressions.

2.5 Optimal Control

Up to now, no assumption has been made about the nature of the unmodelled component(s) of the dynamics, beside the requirement for it to match the available state observations. Particularly in the low-data limit, however, including additional assumptions enables one to reconstruct different families of possible underlying dynamics, all equally compatible with the available tracking data [16].

A natural choice is to assume that the unmodeled term in the dynamics is due to human intervetion (i.e., it is a manoeuver), so that it respects some optimality criterion [25]. Following previous works, we incorporate Optimal Control Policies into our framework, which is naturally compatible with predictive control formulations ([29], [6], [20]). In particular, we assume that the manoeuver minimizes fuel consumption (equivalent to minimizing the integral of the magnitude of the control effort across the complete maneuver period). We introduce the quadratic cost function:

$$\mathcal{F}_2 = \int_{t_0}^{t_f} ||\mathbf{u}(s)||^2 \mathrm{d}s \qquad [10]$$

And we modify the global fitness of Algorithm [1] using Equation [11]:

$$\mathcal{F} = \mathcal{F}_1 + w\mathcal{F}_2 \qquad [11]$$

where w is a weighting parameter. The quantity given in Equation [10] is computed in two different ways, respectively for the optimal manoeuver construction and for its estimation from tracking data. In the former case, we make use of Direct Finite Elements in Time (DFET) to solve optimal control problems [33], so that the fitness is computed using Simpson's rule from the estimated nodal values. In the estimation part, the approximation of Equation [10] is computed evaluating the model at the available state observation epochs.

3. Results

In this section, the presented algorithm will be applied to a number of dynamical systems, ranging from simple problems, based on the driven oscillator, to ones relevant in Orbital Mechanics: this is to underline how general and powerful the proposed approach is.

3.1 Driven harmonic oscillators

3.1.1 Time-dependent Perturbation

We start from a driven, damped harmonic oscillators in which the externally applied force $F(t)$ is not a function of the state of the system. Assuming unitary mass, such a system is described by the following:

$$\ddot{x}(t) = f(x, \dot{x}, t) = -kx(t) - c\dot{x}(t) + F_0 \sin(\omega t) \quad [12]$$

in which the various parameters are given by:

$$\begin{array}{rcll} k & = & 4.518 & kg/s^2 \\ c & = & 0.376 & kg/s \\ F_0 & = & 8.865 & kg\,m/s^2 \\ \omega & = & 1.440 & s^{-1} \end{array} \qquad [13]$$

Propagating the initial condition:

$$\begin{array}{rcll} x_0 & = & 2 & m \\ \dot{x}_0 & = & 3 & m/s \end{array} \qquad [14]$$

we obtain the reference trajectory given in Figure 2:

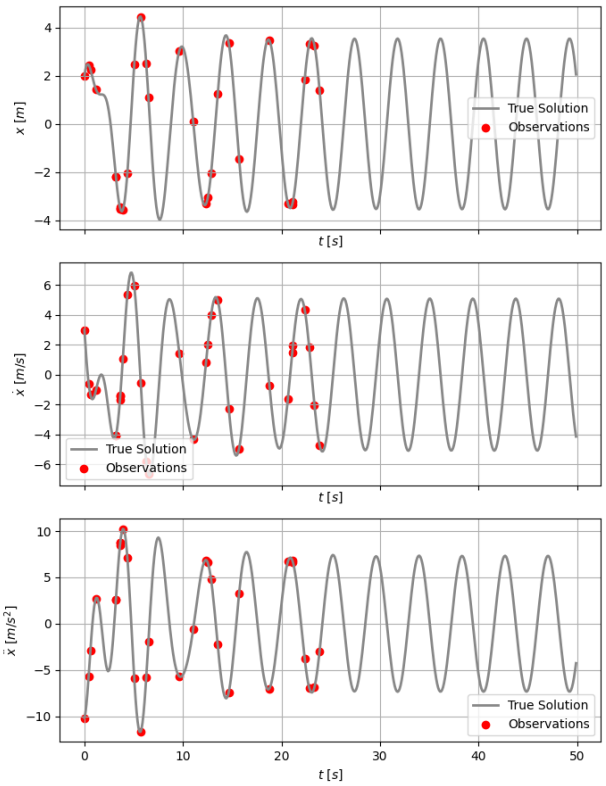


Fig. 2: State and state derivative associated to the system given in Eq. [12].

In the plot, the 27 observations computed using the interpolation and finite differencing techniques given in Section 2.1 are also represented. These can hence be fed into the Deep Symbolic Regression algorithm in order to reconstruct the governing equation of the system. Since in this case the position and velocity of the system are scalar quantities, there is no need to introduce Strongly-Type Genetic Programming, being the elementary fuctions used $+, \cdot, \sin(), \cos(), ()^2, e^{()}$; the number of genes has been set to 4.

The estimated system, obtained using a population of 15 individuals evolved for 20 generations, is given by:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} x \\ v \\ t \end{bmatrix} = \begin{bmatrix} v \\ -kx(t) - cv(t) + 8.857 f_2(x, v, t) \\ 1.000 \end{bmatrix} \quad [15]$$

and the associated Global Fitness is:

$$\mathcal{F}_1 \approx 2.357 \cdot 10^{-5} \;\; m^2 s \quad [16]$$

The correct estimation of the f_2 function is represented in Figure 3, together with the associated value of the estimated parameter. The sparse regression part of the algorithm has been able to correctly associate to the meaningless sub-individuals a linear parameter of zero, so that among the symbolic expression only one is active, as desired.

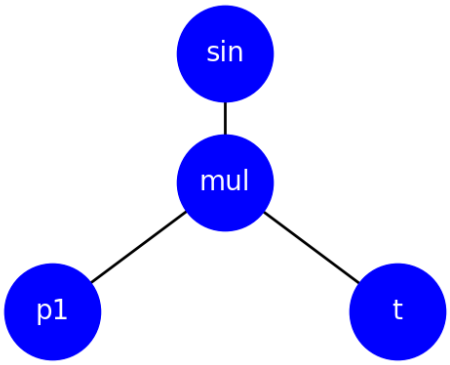


Fig. 3: $f_2(x, v, t)$, $p_1 = 1.44000943$

Finally, making use of the autonomously constructed Equation [15], we can propagate the initial condition associated to the first observation, as given in Figure 4. Being interested in investigating the extrapolation performance of the algorithm, we propagate beyond the last observation epoch: this is particularly important, from a theoretical point of view, for chaotic systems and for systems in which an inaccurate estimation of both linear and nonlinear parameters leads to bifurcation. Application-wise, the inferred white-box model can be used to perform collision assessment and its extrapolating performance is crucial for this.

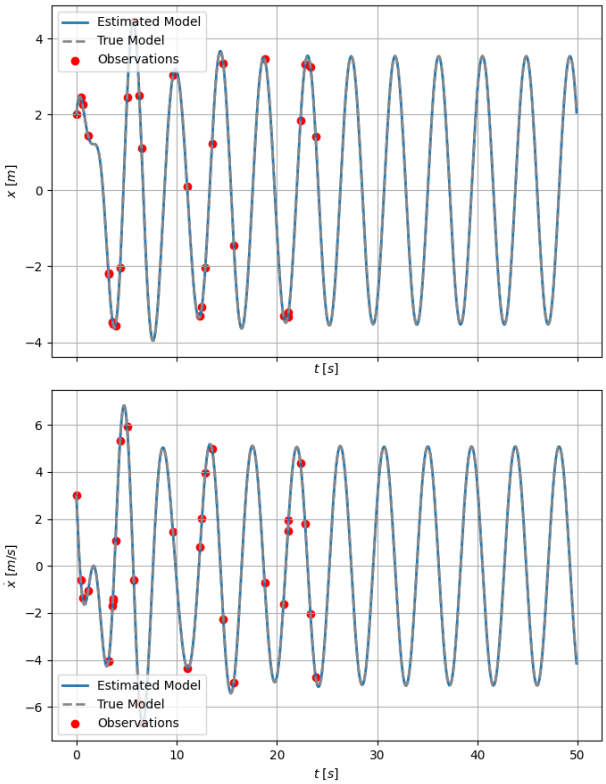


Fig. 4: State estimation associated to the system given in Eq. [15].

3.1.2 *Time- and State-dependent Perturbation*

We now modify Equation [12] in order to introduce an unmodelled term which is both time- and state-dependent. This allows us to underline a qualitative difference between the Sparse Regression and Symbolic Regression part of the algorithm: while the former makes use of the formulation given in Equation [3], the latter works with quantities with physical meaning (i.e., position, velocity and time separately): this is necessary, in order to satisfy the sparsity assumption. This new system is given in Equation [17]:

$$\ddot{x}(t) = -kx(t) - cv(t) + F_0 v(t) \sin(\omega t) \quad [17]$$

with

$$\begin{array}{rcll} k & = & 4.518 & kg/s^2 \\ c & = & 0.376 & kg/s \\ F_0 & = & 2.865 & kg\, m/s^2 \\ \omega & = & 1.447 & s^{-1} \end{array} \quad [18]$$

Making use of the same initial conditions given in Equation [14] and the same observation epochs, we obtain the reference results in Figure 5:

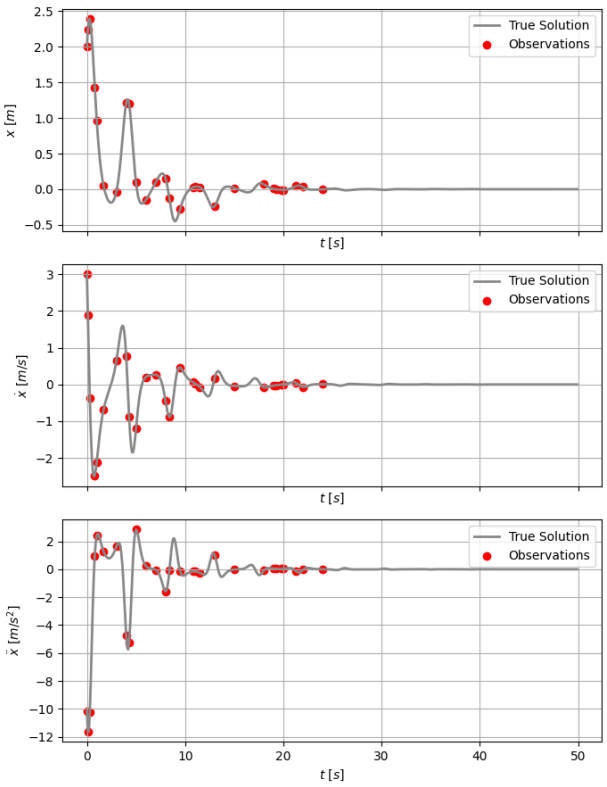


Fig. 5: State and state derivative associated to the system given in Eq. [17].

With the same elementary symbols and number of genes used in the previous case, the algorithm is again successful in inferring the system's governing equation (Equation [19]), obtained using the same number of individuals and generations as before. The missing term, represented in Figure 6, is correctly reconstructed and associated to good estimates of both the linear and non-linear parameters.

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} x \\ v \\ t \end{bmatrix} = \begin{bmatrix} v \\ -kx(t) - cv(t) + 2.865 \quad f_0(x,v,t) \\ 1.000 \end{bmatrix} \tag{19}$$

The estimated model is associated to the following Global Fitness:

$$\mathcal{F}_1 \approx 3.06 \cdot 10^{-6} \ \ m^2 s \tag{20}$$

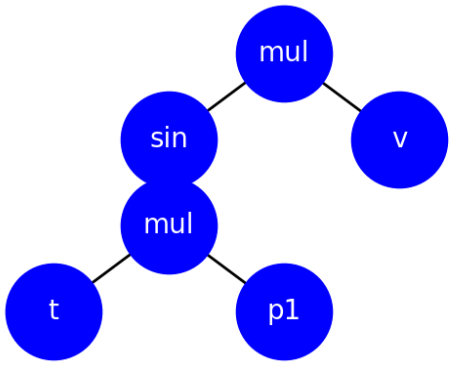


Fig. 6: $f_0(x, v, t)$, $p_1 = 1.44704886$

Finally, the dynamics resulting from the estimated system is represented in Figure 7.

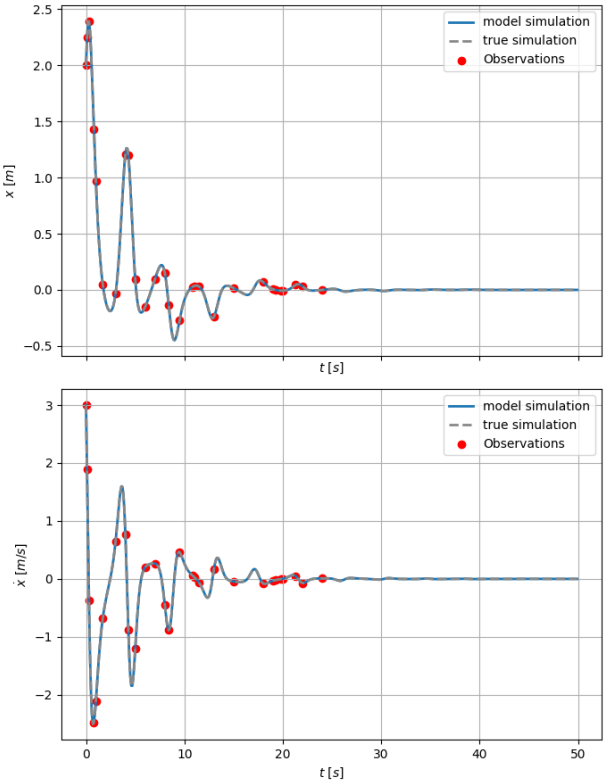


Fig. 7: State estimation associated to the system given in Eq. [19]

3.2 Two-Body problem with Drag

We now apply the proposed methodology to an Orbital Mechanics problem: we assume to be tracking an inactive satellite at the early phases of its re-entry, in or-

der to be able to predict its impact location. A simplified dynamics, given in Equation [21] is used to model such a scenario.

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} r \\ \theta \\ v_r \\ v_t \end{bmatrix} = \begin{bmatrix} v_r \\ v_t/r \\ -\mu/r^2 + v_t^2/r + a_r^{DRAG} \\ -(v_t v_r)/r + a_t^{DRAG} \end{bmatrix} \quad [21]$$

We assume that the drag acceleration is time-dependent due to a time-varying effective area (Equation [22]), associated to the tumbling behaviour of the satellite resulting from a debris impact (Equation [23]).

$$a_i^{DRAG} = -\frac{1}{2}\frac{A(t)}{m}\overline{\rho} C_D ||\mathbf{v}|| v_i \quad [22]$$

Being able to reconstruct the symbolic structure of this unmodeled force, allows one to infer the causal nature of the observed phenomenon and make predictions about it; moreover, automating such a task allows one to reduce the analysis time.

$$A(t) = \overline{A}[1 + \frac{1}{2}\sin(\omega_t t)] \quad [23]$$

We assume an average effective area of 10 dm^2, a ballistic coefficient $C_D = 2.2$, a mass of 1 kg and a constant atmospheric density, obtained from the exponential model evaluated at the initial altitude: the trajectory is such that, during the tracking interval, the altitude is approximately constant. Finally, the nonlinear parameter is given by:

$$\omega_t = \frac{2\pi}{60} \quad [24]$$

The initial state used for the propagation is the following:

$$\begin{aligned} r_0 &= 6521 \quad km \\ \theta_0 &= 0 \quad rad \\ v_{r0} &= 0.05 \quad km/s \\ v_{t0} &= \sqrt{\tfrac{\mu}{r0}} \end{aligned} \quad [25]$$

The propagated reference state and state derivative associated to such a model, together with 30 randomly distributed observations computed in the usual way, are represented in Figure 8:

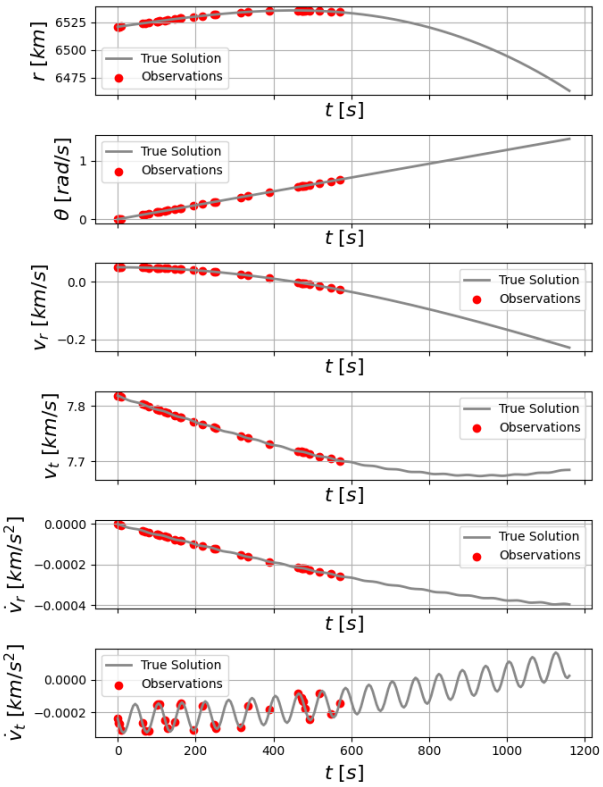


Fig. 8: State and state derivative associated to the system given in Eq. [21].

In order to reconstruct the missing term, together with the previously introduced operations, also the norm operator $||()||$ is introduced for the Symbolic Regression: here Strongly-Typed Genetic Programming is necessary, since both scalars and vectors are taken as inputs.

The number of genes, individuals and generations have been set to 2, 40, 50, respectively, leading to an estimated mismodelled term given in Equation [26].

$$\delta\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} r \\ \theta \\ v_r \\ v_t \\ t \end{bmatrix} = \begin{bmatrix} 0.000 \\ 0.000 \\ 0.000 \\ -2.9314 \cdot 10^{-6}\sin(0.1052 \cdot t)||\mathbf{v}||v_t \\ 0.000 \end{bmatrix} \quad [26]$$

As a comparison, the explicit values of the linear and non-linear parameters to be estimated are $-2.9017 \cdot 10^{-6}$, 0.1047. The Global Fitness associated to the estimated system is given by:

$$\mathcal{F}_1 \approx 0.2503 \quad [27]$$

It can be seen how the perturbation in the radial component is not recostructed, due to its low amplitude. In fact, for this case, a rescaling of the unmodeled state

derivative array has been necessary in order for the coefficients of the sparse regression to be above the threshold defined in the STLSQ algorithm. Measurements at different timescales should be available, in order to reconstruct low amplitude perturbations such as the unmodeled radial acceleration [3]. At the same time, high levels of signal-to-noise ratio are desirable.

Nevertheless, the structure of the unmodeled force in the tangential component is correctly identified, together with a good estimate of the linear and nonlinear parameters, leading to the dynamics represented in Figure 9:

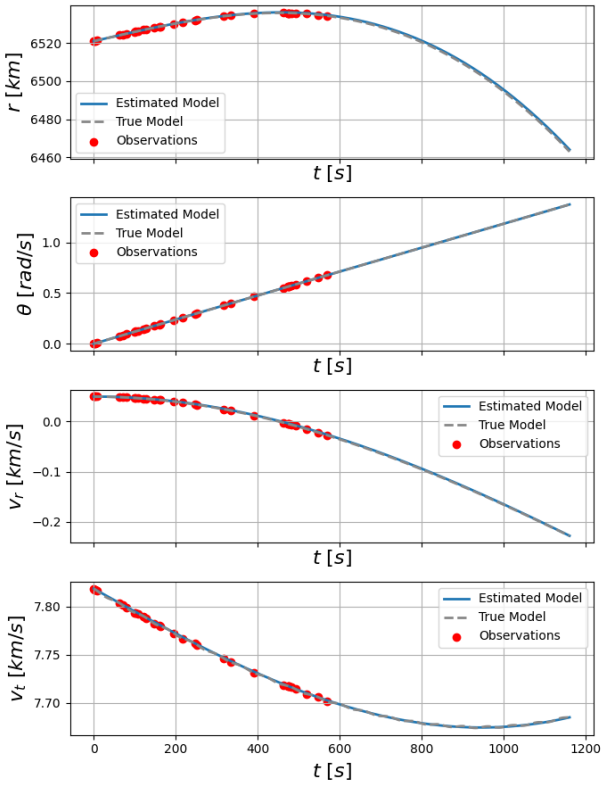


Fig. 9: State estimation associated to the system given in Eq. [26]

It is worth underlying how it is not always possible to infer the causal nature of the underlying physical process from the symbolic expression representing its effect. In this problem, for example, there is no obvious way of autonomously concluding the oscillatory behaviour to be a consequence of the motion of the spacecraft or of the behaviour of the atmosphere, for example‡.

‡ As an additional example, we can think about a spacecraft orbiting the Sun, mainly affected by its gravitational attraction and the Solar Radiation Pressure. In the presence of a time-varying parameter, it is not obvious how to associate its effect on the dynamics to one force or the other.

3.3 Optimal Control Manoeuvre

We now make use of the formulation introduced in Section 2.5, applying it to reconstruct the acceleration associated to an optimal control manoeuvre in the sense given in Equation [10]. Starting again from the differential equation associated to the two-body problem, written in polar coordinates, we introduce a control (Equation [28]) identifying the acceleration due to a manoeuver.

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} r \\ \theta \\ v_r \\ v_t \end{bmatrix} = \begin{bmatrix} v_r \\ v_t/r \\ -\mu/r^2 + v_t^2/r + u_r(t) \\ -(v_t v_r)/r + u_t(t) \end{bmatrix} \qquad [28]$$

We compute the optimal control maneouver leading to an increase in altitude of 20 km, making use of 7 nodal points associated to the DFET method. From the initial condition given by:

$$\begin{aligned} r_0 &= 7000 \quad km \\ \theta_0 &= 0.0 \quad rad \\ v_{r0} &= 0.0 \quad km/s \\ v_{t0} &= \sqrt{\tfrac{\mu}{r0}} \end{aligned} \qquad [29]$$

we can compute the trajectory and the 16 equally spaced observations given in Figure 10:

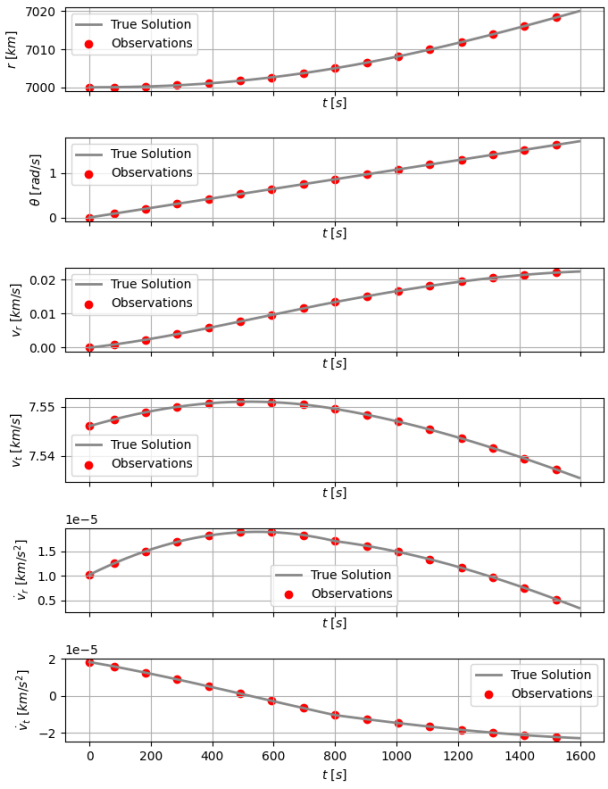


Fig. 10: State and state derivative associated to the system given in Eq. [28].

We make use of 3 genes, 50 individuals and 20 generations; we change the Global Fitness function using Equation [11], using a weighting parameter of 1. Also to show the possibility of constructing the state-dependent matrix of the Sparse Regression combining a priori selected functions with expressions associated to Symbolic Regression, we include a polynomial basis up to degree 2.

With this formulation, we estimate the control law (Equation [30]) and the the system's behaviour associated to it (Figure 11).

$$\begin{aligned} u_r &= 10^{-14}(-2.796t + 19.749r^2 + \\ &\quad -1633.064t^2 + 0.882t^3)\ km/s^2 \\ u_t &= 10^{-14}(-6.041t + 34.725r^2 + \\ &\quad -3528.112t^2 + 1.877t^3)\ km/s^2 \end{aligned} \quad [30]$$

The algorithm reconstruct the control law using a polynomial representation: this is a consequence of its sparsity-promoting nature. The estimated expressions contain a state dependent term: again because we are interested in obtaining a compact symbolic representation of the manoeuver, the algorithm makes use of available time-dependent functions (i.e. $r(t)$), if these are useful as building blocks for the optimal control law.

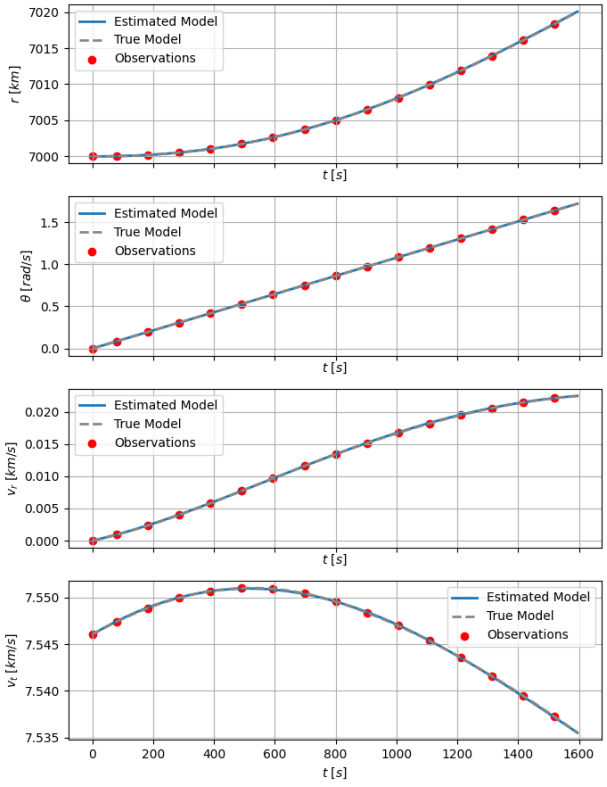


Fig. 11: State estimation associated to Eq. [30]; 16 observations.

We can also look at what happens reducing the number of available observations. Making use of $\mathcal{F}_1$ to estimate the underlying dynamics, one would have a number of systems equally compatible with the observed constraints; introducing the optimality constrait greatly reduces this freedom. In fact, as given in Figure 12, with just 4 state observations we are able to have a good estimate of the overall dynamics associated to the manoeuver.

$$\begin{aligned} u_r &= 10^{-14}(-1150306.278t + \\ &\quad +21.792r^2 + 429.116t^2)\ km/s^2 \\ u_t &= 10^{-14}(-2458647.369t + \\ &\quad +39.043r^2 + 869.123t^2)\ km/s^2 \end{aligned} \quad [31]$$

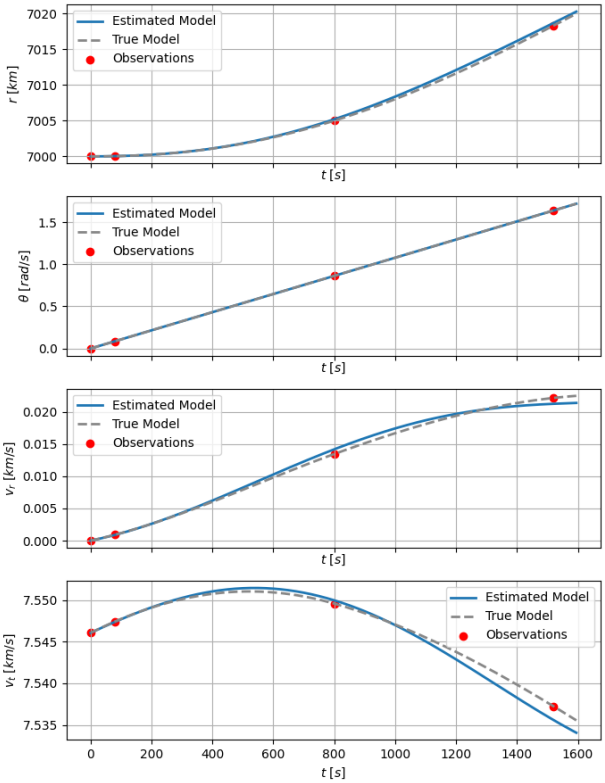


Fig. 12: State estimation associated to Eq. [31]; 4 observations.

In this section, we have been assuming that the dynamics is continuous inside the observation interval and, moreover, that the manoeuver's optimality condition is associated to the observation interval. Additional steps are required in the envisioned manoeuver detection pipeline in order to identify the time interval (in general a subset of the observation interval) over which there is a mismatch between the modelled and observed dynamics.

4. Conclusions & Recommendations

In conclusion, Genetic Programming-based Symbolic Regression can be successfully combined with Sparse Regression techniques into what we have here called Deep Symbolic Regression, in order to estimate governing equations from data and reconstruct unmodeled components, such as anomalies in Orbital Mechanics.

In the proposed architecture, nonlinear parameter estimation is performed in tandem with the construction of a sparse dynamical model, to be later used in inferring the causal nature of the observed phenomenon and used for generalization and extrapolation.

Deep Symbolic Regression has been shown to be naturally compatible with Optimal Control Policies, which can be used as additional constraints, particularly useful in the context of manoeuver detection in the low-data limit.

In future works, we aim at testing the algorithm against real Radar Tracking data. Moreover, in validating the methodology against real mission scenarios, the number of data will be greatly reduced, while the length of each tracklet-like observation will be increased. Assessing the performance of the algorithm in the low-data limit will be the focus of following works. We will also focus on investigating the robustness of the presented approach with respect to noisy measurements and on hyperparameters estimation.

As discussed in Section 3.3, in an autonomous manoeuver detection pipeline, the presented approach should be coupled with algorithms identifying time intervals associated to different dynamical regimes.

Finally, following the Kalman Filter assumption, by which both measurement noise and modelling mismatch are stochastic processes (in particular zero-mean Gaussian ones), future works should focus on generalizing the presented approach in order to deal with Markov processes ([18], [1]).

5. Acknowledgment

This work was supported through the MSCA ETN Stardust-R grant agreement number 813644.

6. Data and source code

The Python Implementation developed in the context of this work will be made available, together with the data used to produce the main figures, at `https://github.com/strath-ace-labs/CASSANDRA` soon after the submission of this paper.

References

[1] Hassan Arbabi and Themistoklis Sapsis. Data-driven modeling of strongly nonlinear chaotic systems with non-Gaussian statistics. *arXiv: Dynamical Systems*, 2019.

[2] Josh Bongard and Hod Lipson. Automated reverse engineering of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 104(24):9943–9948, 2007.

[3] Jason J. Bramburger, Daniel Dylewsky, and J. Nathan Kutz. Sparse identification of slow timescale dynamics. *Physical Review E*, 102:022204, 2020.

[4] Steven Brunton, Joshua Proctor, and J. Nathan Kutz. Discovering governing equations from data: Sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.

[5] Steven L. Brunton and J. Nathan Kutz. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press, 2019.

[6] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Sparse Identification of Nonlinear Dynamics with Control (SINDYc). *IFAC-PapersOnLine*, 49(18):710–715, 2016.

[7] Kathleen Champion, Bethany Lusch, J. Nathan Kutz, and Steven L. Brunton. Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences*, 116(45):22445–22451, 2019.

[8] Kathleen Champion, Peng Zheng, Aleksandr Y. Aravkin, Steven L. Brunton, and J. Nathan Kutz. A unified sparse optimization framework to learn parsimonious physics-informed models from data. *IEEE Access*, 8:169259–169271, 2020.

[9] Rick Chartrand. Numerical differentiation of noisy, nonsmooth data. *ISRN Applied Mathematics*, 2011.

[10] Anshul Choudhary, John Lindner, Elliott Holliday, Scott Miller, Sudeshna Sinha, and William Ditto. Physics-enhanced neural networks learn order and chaos. *Physical Review E*, 101, 2020.

[11] Miles Cranmer, Alvaro Sanchez-Gonzalez, Peter Battaglia, Rui Xu, Kyle Cranmer, David Spergel, and Shirley Ho. Discovering Symbolic Models from Deep Learning with Inductive Biases. *arXiv*, 2020.

[12] Brian M. de Silva, Kathleen Champion, Markus Quade, Jean-Christophe Loiseau, J. Nathan Kutz,

and Steven L. Brunton. PySINDy: A Python package for the Sparse Identification of Nonlinear Dynamics from Data, 2020.

[13] Félix-Antoine Fortin, François-Michel De Rainville, Marc-André Gardner, Marc Parizeau, and Christian Gagné. DEAP: Evolutionary algorithms made easy. *Journal of Machine Learning Research*, pages 2171–2175, 2012.

[14] Amir Gandomi and Amir Alavi. A new multi-gene genetic programming approach to nonlinear system modeling. Part I: Materials and structural engineering problems. *Neural Computing and Applications*, 21:171–187, 02 2012.

[15] Sam Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian Neural Networks. *arXiv*, 2019.

[16] Marcus Holzinger, Daniel Scheeres, and Kyle Alfriend. Object Correlation, Maneuver Detection, and Characterization Using Control Distance Metrics. *Journal of Guidance, Control, and Dynamics*, 35:1312–1325, 2012.

[17] Hitoshi Iba. Evolutionary Approach to Machine Learning and Deep Neural Networks. *Neuro-Evolution and Gene Regulatory Networks*, 2018.

[18] Kadierdan Kaheman, Steven L. Brunton, and J. Nathan Kutz. Automatic Differentiation to Simultaneously Identify Nonlinear Dynamics and Extract Noise Probability Distributions from Data. *arXiv*, 2020.

[19] Kadierdan Kaheman, Eurika Kaiser, Benjamin Strom, J. Nathan Kutz, and Steven L. Brunton. Learning Discrepancy Models From Experimental Data. *ArXiv*, abs/1909.08574, 2019.

[20] Eurika Kaiser, J. Nathan Kutz, and Steven L. Brunton. Sparse identification of nonlinear dynamics for model predictive control in the low-data limit. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 474, 2017.

[21] John R. Koza. Genetic programming as a means for programming computers by natural selection. *Statistics and Computing*, 4:87–112, 1994.

[22] Li Li, Minjie Fan, Rishabh Singh, and Patrick Riley. Neural-Guided Symbolic Regression with Asymptotic Constraints. *arXiv*, 2019.

[23] Richard Linares. Koopman Operator Theory applied to the Motion of Satellites. *AAS/AIAA Astrodynamics Specialist Conference*, 2019.

[24] Daniel P. Lubey and Daniel J. Scheeres. Combined optimal control and state estimation for the purposes of maneuver detection and reconstruction. In *2014 American Control Conference*, pages 1749–1754, 2014.

[25] Daniel P. Lubey and Daniel J. Scheeres. Identifying and Estimating Mismodeled Dynamics via Optimal Control Policies and Distance Metrics. *Journal of Guidance, Control, and Dynamics*, 37(5):1512–1523, 2014.

[26] Daniel P. Lubey, Daniel J. Scheeres, and Richard Erwin. Maneuver Detection and Reconstruction of Stationkeeping Spacecraft at GEO using the Optimal Control-Based Estimator. *IFAC-PapersOnLine*, 48:216–221, 2015.

[27] Bethany Lusch, J. Nathan Kutz, and Steven L. Brunton. Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications*, 9, 2018.

[28] Matteo Manzi and Massimiliano Vasile. Discovering Unmodeled Components in Astrodynamics with Symbolic Regression. In *2020 IEEE Congress on Evolutionary Computation (CEC)*. IEEE, 2020.

[29] Francesco Marchetti, Edmondo Minisci, and Annalisa Riccardi. Towards intelligent control via genetic programming. In *IEEE World Congress on Computational Intelligence*, 2020.

[30] Andrea Milani and Giovanni Federico Gronchi. *Theory of Orbit Determination*. Cambridge University Press, 2009.

[31] Theodore J. Muelhaupt, Marlon E. Sorge, Jamie Morin, and Robert S. Wilson. Space traffic management in the new space era. *Journal of Space Safety Engineering*, 6(2):80–87, 2019.

[32] Brenden K. Petersen. Deep symbolic regression: Recovering mathematical expressions from data via risk-seeking policy gradients. *arXiv*, 2019.

[33] Lorenzo Ricciardi and Massimiliano Vasile. Direct Transcription of Optimal Control Problems with Finite Elements on Bernstein Basis. *Journal of Guidance Control and Dynamics*, pages 1–15, 2018.

[34] Samuel Rudy, J. Nathan Kutz, and Steven L. Brunton. Deep learning of dynamics and signal-noise decomposition with time-stepping constraints. *Journal of Computational Physics*, 396, 2019.

[35] Lukas Ruff, Jacob R. Kauffmann, Robert A. Vandermeulen, Grégoire Montavon, Wojciech Samek,

Marius Kloft, Thomas G. Dietterich, and Klaus-Robert Müller. A Unifying Review of Deep and Shallow Anomaly Detection, 2020.

[36] Abraham Savitzky and Marcel J. E. Golay. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, 36(8):1627–1639, 1964.

[37] Michael Schmidt and Hod Lipson. Distilling Free-Form Natural Laws from Experimental Data. *Science*, 324(5923):81–85, 2009.

[38] Silviu-Marian Udrescu and Max Tegmark. AI Feynman: A physics-inspired method for symbolic regression. *Science Advances*, 6(16), 2020.

[39] Yiran Wang, Hao Peng, Xiaoli Bai, Genshe Chen, Dan Shen, and Erik Blasch. Data-Driven Anomaly Detection for Resident Space Objects using Autoencoder with Binary Classification. *AAS/AIAA Astrodynamics Specialist Conference*, 2020.

[40] Peng Zheng, Travis Askham, Steven L. Brunton, J. Nathan Kutz, and Aleksandr Y. Aravkin. A Unified Framework for Sparse Relaxed Regularized Regression: SR3. *IEEE Access*, 7:1404–1423, 2019.