Dimensionally Consistent Learning with Buckingham Pi

Joseph Bakarji (jbakarji@uw.edu)
University of Washington
Jared Callaham
University of Washington
Steven Brunton
University of Washington https://orcid.org/0000-0002-6565-5118
J. Nathan Kutz
University of Washington

Article

Keywords: Dimensional analysis, Buckingham Pi, dimensionality reduction, machine learning, deep learning, fluid mechanics

Posted Date: May 24th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1547348/v1

Dimensionally Consistent Learning with Buckingham Pi

Joseph Bakarji[1*], Jared Callaham[1], Steven L. Brunton[2] and J. Nathan Kutz[1]

[1*]Department of Mechanical Engineering, University of Washington, Seattle, 98195, WA, USA.
[2]Department of Applied Mathematics, University of Washington, Seattle, 98195, WA, USA.

*Corresponding author(s). E-mail(s): jbakarji@uw.edu;
Contributing authors: jc244@uw.edu; sbrunton@uw.edu; kutz@uw.edu;

Abstract

In the absence of governing equations, dimensional analysis is a robust technique for extracting insights and finding symmetries in physical systems. Given measurement variables and parameters, the Buckingham Pi theorem provides a procedure for finding a set of dimensionless groups that spans the solution space, although this set is not unique. We propose an automated approach using the symmetric and self-similar structure of available measurement data to discover the dimensionless groups that best collapse this data to a lower dimensional space according to an optimal fit. We develop three data-driven techniques that use the Buckingham Pi theorem as a constraint: (i) a constrained optimization problem with a non-parametric input-output fitting function, (ii) a deep learning algorithm (BuckiNet) that projects the input parameter space to a lower dimension in the first layer, and (iii) a technique based on sparse identification of nonlinear dynamics (SINDy) to discover dimensionless equations whose coefficients parameterize the dynamics. We explore the accuracy, robustness and computational complexity of these methods as applied to three example problems: a bead on a rotating hoop, a laminar boundary layer, and Rayleigh-Bénard convection.

Keywords: Dimensional analysis, Buckingham Pi, dimensionality reduction, machine learning, deep learning, fluid mechanics

1 Introduction

Dimensional analysis is based on the simple idea that physical laws do not depend on the units of measurements. As a consequence, any function that expresses a physical law has the fundamental property of so-called *generalized homogeneity* [1] and does not depend on the observer. Although such concepts of dimensional analysis go back to the time of Newton and Galileo [2], it was formalized mathematically by the pioneering contributions of Edgar Buckingham in 1914 [3]. Specifically, Buckingham proposed a principled method for extracting the most general form of physical equations by simple dimensional considerations of the seven fundamental units of measurement: length (meter), mass (kg), time (seconds), electric current (Ampere), temperature (Kelvin), amount of substance (mole), and luminous intensity (candela). From electromagnetism to gravitation, measurements can be directly related to the seven fundamental units, e.g. force is measured in Newtons which is $\text{kg}\cdot\text{m}/\text{s}^2$ and electric charge by the volt which is $\text{kg}\cdot\text{m}^2/(\text{s}^3\cdot\text{A})$. The resulting Buckingham Pi theorem was originally contextualized in terms of physically similar systems [3], or groups of parameters that related similar physics. In modern times, rapid advancements in measurement and sensor technologies have produced data of diverse quantities at almost any spatial and temporal scale. This has engendered a renewed consideration of the Buckingham Pi theorem in the context of data-driven modeling [4–7]. Specifically, as shown here, physics-informed machine learning algorithms can be constructed to automate the principled approach advocated by Buckingham for the discovery of the most general and parsimonious form of physical equations possible from measurements alone.

The Buckingham Pi theorem was critically important in the pre-computer era. Indeed, the discovery of key dimensionless quantities was considered a fundamental result, as it often uncovered the self-similarity structure of the solution along with parametric dependencies and symmetries. In fluid dynamics alone, for instance, there are numerous named dimensionless parameters that are discovered through the Buckingham Pi reduction, including the Reynolds, Prandtl, Froude, Weber, and Mach numbers. Scaling and dimensional analysis continue to provide a robust starting point for understanding both complex and basic physical phenomena, like the physics of wrinkling [8], chaotic patterns in Rayleigh-Bénard convection [9], the structure of drops falling from a faucet [10], spontaneous pattern formation by self-assembly [11], and osmotic spreading of biofilms [12]. In practice, many dimensionless groups naturally arise from ratios of domain averaged terms in physically meaningful equations that are required to be dimensionally homogeneous. When equations are non-dimensionalized, dimensionless terms that are small on average relative to others can be treated as perturbations, thus providing insights into the qualitative structure of the solution and simplifying the problem of finding it. More importantly, Buckingham Pi provides the potential for an *order parameter* description [13] that determines the qualitative structure of the solution and its bifurcations without recourse

to the full-state equations [14]. These order parameters form the basis for casting problems into normal forms that reveal physical symmetries and can potentially be solved analytically [15–17]. Even when governing equations are unknown, Buckingham Pi establishes a dimensionally consistent relationship between the input parameters/variables and output predictions which help constrain models and prevent over-fitting.

The rise of scientific computing in the 1980s made it possible to numerically integrate highly nonlinear ordinary and partial differential equations, without the need for analytic simplifications provided by dimensional analysis. More recently, machine learning algorithms have shown promise in discovering scientific laws [18], differential equations [19–21], and deep network input-output function approximations [22–24] from simulation and experimental data alone, with considerable progress in the field of fluid mechanics [25–33]. However, these methods do not take into account the units of their training data, which can significantly constrain the hypothesis class to physically meaningful solutions. Active subspaces provide a principled algorithmic method for extracting dimensionless numbers from experimental and simulation data [34], having successfully been demonstrated to discover dimensionless numbers in particle-laden turbulent flows [5]. In a related line of work, statistical null-hypothesis testing has been used alongside the Buckingham Pi theorem to find hidden variables in dimensional experimental data [4]. Dimensional analysis has also been used for a physics-inspired symbolic regression algorithm that discovers physics formulas from data [35] and to improve neural network models [36]. Data-driven dimensional analysis has also been applied to model turbulence data [6]. Machine learning algorithms that use sparse identification of differential equation to discover dimensionless groups and scaling laws that best collapse experimental data have also been recently proposed [7].

In this study, we address the problem of automatic discovery of dimensionless groups from available experimental or simulation data using the Buckingham Pi theorem as a constraint. Importantly, although the Buckingham Pi theorem provides a step-by-step method for finding dimensionless groups directly from the input variables and parameters, its solution is not unique. Thus it relies on intuition and experience with the physical problem at hand. Our aim is to leverage symmetries in the data to inform the algorithm of which set of dimensionless groups control the behavior (e.g. the Reynolds number controls the turbulence of a flow-field). We develop three methods that find the most physically meaningful Pi groups, assumed to be the most predictive of the dependent variables. The first method is a constrained optimization that fits independent variable and parameter inputs to predictions, while satisfying the Buckinham Pi theorem as a hard constraint. The second method uses a Buckingham Pi (nullspace) constraint on the first hidden layer of a deep network that fits input parameters to output predictions. And the third method uses SINDy [19, 20] to constrain available dimensionless groups to be coefficients in a sparse differential equation.

The latter method has the added benefit of discovering a parametric (dimensionless) differential equation that describes the data, thus simultaneously generalizing the SINDy method. We apply these methods to three problems: the rotating hoop, the Blasius boundary layer, and the Rayleigh-Bérnard problem.

2 Methods

Given a vector of n input measurement parameters and variables $\mathbf{p} \in \mathbb{R}^n$ and a set of k quantities of interest $\mathbf{q} \in \mathbb{R}^k$, one generally seeks a mapping f that minimizes the prediction error $\varepsilon = \|f(\mathbf{p}) - \mathbf{q}\|$. The function f is typically the solution of an initial and/or boundary value problem with known variables and parameters $\mathbf{p}$ and unknown solution $\mathbf{q}$, all of which have physical dimensions [units].

The Buckingham Pi theorem states that there exists a set of dimensionless quantities $\boldsymbol{\pi} = (\boldsymbol{\pi}_p, \boldsymbol{\pi}_q)$, also known as $\boldsymbol{\pi}-$groups, such that the dimensionless input parameters/variables $\boldsymbol{\pi}_p \in \mathbb{R}^{n'}$ span the full dimensionless solution space $\boldsymbol{\pi}_q \in \mathbb{R}^{k'}$, with $n' \leq n$ and $k' \leq k$ [3].

A probabilistic corollary of the Buckingham Pi theorem is that the input-output data $(\mathbf{p}, \mathbf{q})$ can be projected on a lower dimensional space with basis $(\boldsymbol{\pi}_p, \boldsymbol{\pi}_q)$ without compromising the prediction error ε, such that,

$$\|\mathbf{q} - f(\mathbf{p})\| < \varepsilon \quad \rightarrow \quad \|\boldsymbol{\pi}_q - \psi(\boldsymbol{\pi}_p)\| < \varepsilon, \tag{1}$$

where ψ is an unknown function that can be approximated from available data. However, the theorem does not provide any information on the properties of ψ or its relationship with the Pi groups. The theorem generally assumes $\mathbf{p}$, $\mathbf{q}$, and f to be deterministic without considering the optimality of their corresponding dimensionless basis.

In a data-driven statistical setting, we assume that the inputs $\mathbf{p}$ and outputs $\mathbf{q}$ are sampled from a distribution function, and that f is an optimal mapping in a given class of functions $\mathcal{H}$. Accordingly, we posit the following hypothesis

Hypothesis 1 Given a set of input and output measurement pairs $\tilde{\mathbf{p}} = (\mathbf{p}, \mathbf{q})$, the most physically meaningful dimensionless basis $\boldsymbol{\pi}^*$ is the optimal coordinate transformation $\tilde{\mathbf{p}} \rightarrow \boldsymbol{\pi}^*$ that satisfies the minimization problem

$$\boldsymbol{\pi}^* = \arg\min_{\boldsymbol{\pi}} \left(\min_{\psi} \|\boldsymbol{\pi}_q - \psi(\boldsymbol{\pi}_p)\|_2^2 \right), \tag{2}$$

where *physical meaningfulness* is defined as the dimensionless group's ability to simplify an equation in relevant regimes and act as a scaling parameter that collapses the input-output solution space to a lower dimension.

While defining *physical meaningfulness* depends on the historical and scientific context, we use known results from various scientific traditions to

validate our hypothesis. Conversely, if this hypothesis gives rise to known dimensionless numbers, it proposes a new approach for defining *physical meaningfulness* as an optimal fit in the context of non-dimensionalization.

The functional relationship between $\boldsymbol{\pi}$ and $\tilde{\mathbf{p}}$ is given by

$$\pi_j = \prod_{i=1}^{d} \tilde{p}_i^{\Phi_{ij}}, \tag{3}$$

where Φ_{ij} are unknown parameters that are chosen to make π_j dimensionless.

Let $\Omega()$ be a function that maps a measurable quantity $\tilde{p}_i$ to a vector containing the powers of its basic dimensions (i.e. mass M, length L, time T, etc.). For example, if one chooses $[M, L, T]$ as the basic dimensions, then the units of a force are $[F] = ML/T^2$, and $\Omega(F) = [1, 1, -2]^T$. Let $\phi()$ be a function that maps a dimensionless group to a vector that corresponds to the powers of the input parameters $\mathbf{p}$ such that $\phi(\pi_j) = [\Phi_{1j}, \Phi_{2j}, \ldots, \Phi_{dj}]^T$, where $d = n + k$.

The Buckingham Pi theorem [3] constrains $\phi(\pi_i)$ to be in the null-space of the units matrix

$$\boldsymbol{D} = [\boldsymbol{D}_p, \boldsymbol{D}_q] = \begin{bmatrix} | & | & & | & | & \\ \Omega(p_1) & \Omega(p_2) & \ldots & \Omega(p_n) & \Omega(q_1) & \ldots & \Omega(q_k) \\ | & | & & | & | & \end{bmatrix}, \tag{4}$$

such that

$$\boldsymbol{D}\phi(\pi_i) = \mathbf{0}, \quad \forall i \in \{1, \ldots, d\}. \tag{5}$$

The Buckingham Pi theorem determines the number of Pi-groups to be $d' = d - \text{rank}(\boldsymbol{D})$. But its main shortcoming is its inability to provide a unique set of d' dimensionless numbers. In practice, additional heuristic constraints are required to solve for the unknowns Φ_{ij}.

Given m measurement pairs of the inputs vector $\mathbf{p}^{(i)} \in \mathbb{R}^n$ and their corresponding output predictions $\mathbf{q}^{(i)} \in \mathbb{R}^k$, with index i, we define the input parameter matrix $\boldsymbol{P} \in \mathbb{R}^{m \times n}$ and the output prediction matrix $\boldsymbol{Q} \in \mathbb{R}^{m \times k}$ as the row-wise concatenation of all measurements $\mathbf{p}^{(i)}$ and $\mathbf{q}^{(i)}$ respectively. We also define $\tilde{\boldsymbol{P}} = [\boldsymbol{P}, \boldsymbol{Q}] \in \mathbb{R}^{m \times d}$ as the column-wise concatenation of $\boldsymbol{P}$ and $\boldsymbol{Q}$. The Pi-groups powers matrix is given by

$$\boldsymbol{\Phi} = \begin{bmatrix} | & | & & | \\ \phi(\pi_1) & \phi(\pi_2) & \ldots & \phi(\pi_{d'}) \\ | & | & & | \end{bmatrix} \in \mathbb{R}^{d \times d'}, \tag{6}$$

such that Eq.(5) is satisfied. Similarly, we define $\boldsymbol{\Phi}_p \in \mathbb{R}^{n \times n'}$ to contain the n' dimensionless groups corresponding to the n input parameters/variables $\mathbf{p}$.

Equation (3) can be written as $\pi_i = \exp\left\{\sum_{j=1}^{d} \Phi_{ij} \log(\tilde{p}_j)\right\}$, with matrix form

$$\boldsymbol{\Pi} = \exp\left(\log\left(\tilde{\boldsymbol{P}}\right)\boldsymbol{\Phi}\right), \tag{7}$$

where each row in $\boldsymbol{\Pi} \in \mathbb{R}^{m \times d'}$ corresponds to the values of the dimensionless groups $\boldsymbol{\pi}^{(i)}$ for a given experiment i, and the exponential and logarithm are both taken element-wise. We also define the matrix $\boldsymbol{\Pi}_q \in \mathbb{R}^{m \times k'}$ to be the data matrix of the output Pi-groups $\boldsymbol{\pi}_q$ and $\boldsymbol{\Pi}_p \in \mathbb{R}^{m \times n'}$ to be the data matrix of the input Pi-groups $\boldsymbol{\pi}_p$, such that $\boldsymbol{\Pi} = [\boldsymbol{\Pi}_p, \boldsymbol{\Pi}_q]$.

This section introduces three methods to simultaneously identify the nondimensionalization $\boldsymbol{\Phi}$ and an approximation of ψ from a set of experimental or simulation data. In this data-driven context, (7) can be combined with the nullspace constraint, $\boldsymbol{D\Phi} = \boldsymbol{0}$, to optimize ψ over a pre-defined hypothesis class given the measurement data.

We present three methods for discovering dimensionless groups from data. The main features that differentiate them are

1. Imposing hard, soft, or no constraints on the null-space of $\boldsymbol{D}$ to ensure that the Pi-groups are dimensionless according to equation (5).
2. The type of input-output function approximation ψ: e.g. neural network, non-parametric function, or differential equation.

2.1 Constrained optimization

Equation (2) can be cast as a constrained optimization problem, with a fitting function ψ. The input Pi-groups powers matrix $\boldsymbol{\Phi}_p$ maps the input parameters/variables to the dimensionless groups through equation (7), and the constraint of the optimization is given by equation (5). The choice of the hypothesis class of ψ is assumed to be arbitrary in this formulation, notwithstanding its effect on the accuracy of the results and the success of the method.

In this setup, we assume that the outputs, $\mathbf{q}$, can be non-dimensionalized by known constants of motion (i.e. $\boldsymbol{\Pi}_q$ is known). Accordingly, the resulting optimization problem is given by

$$\check{\boldsymbol{\Phi}}_{\boldsymbol{p}} = \arg\min_{\boldsymbol{\Phi}_p} \left\|\boldsymbol{\Pi}_q - \psi(\exp\left(\log(\boldsymbol{P})\boldsymbol{\Phi}_p\right))\right\|_2 + \lambda_1 \left\|\boldsymbol{\Phi}_p\right\|_1, \quad \text{s.t.} \quad \boldsymbol{D}_p\boldsymbol{\Phi}_p = \boldsymbol{0}, \tag{8}$$

where the ℓ_1 regularization enforces sparsity (typical in dimensionless numbers). An example application of this approach that uses kernel ridge regression as a non-parametric approximation of ψ is given in Sec. 3.2.

2.2 BuckiNet: non-dimensionalization with a neural network

The unknown function ψ is generally nonlinear and can be high dimensional. In the absence of a governing equation, a deep neural network is a natural

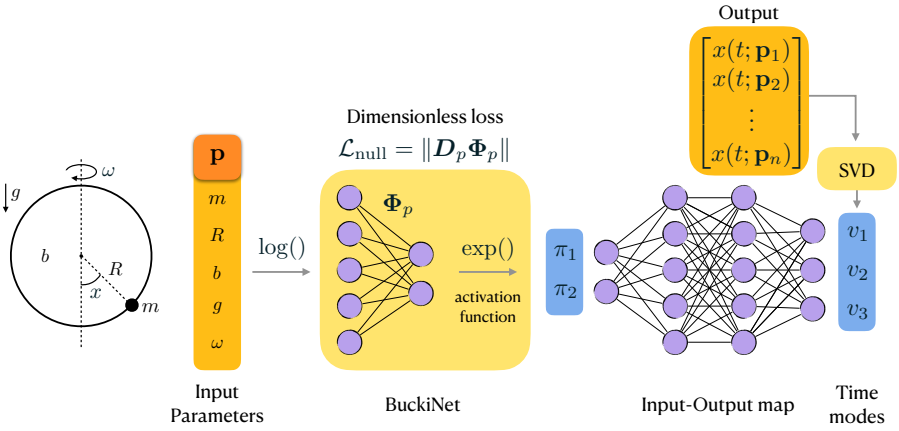

Fig. 1: Illustration of the BuckiNet layer for the rotating hoop problem (described in section 3.1). The dimensionless loss imposes a soft Buckingham Pi constraint from equation (5)) and the BuckiNet layer satisfies equation (3). In this example, the output is given by the three dominant time modes $\mathbf{v}$ of the output $x(t)$ rather than taking the time t as an input.

candidate for approximating ψ. Equation (7) can be naturally integrated in a deep learning architecture by first applying a $\log()$ transform to the inputs $\mathbf{p}$, then using an exponential activation function at the output of the first layer. The number of nodes in the first hidden layer is n' as determined by the Buckingham Pi theorem, and $\boldsymbol{\Phi}_p$ contains the fitting parameters of the first layer, as shown in Fig. 1. We call this architecture the BuckiNet. When negative input data is given, it has to be shifted to the positive domain so that the initial $\log()$ operation becomes possible.

This technique offers many advantages. First, the BuckiNet layer implicitly performs an unsupervised dimensionality reduction without any adjustment to the overall architecture of the deep network. This results in better generalization properties and a faster optimization thanks to a loss-less reduction in the number fitting parameters. Second, the BuckiNet can be easily added to deep learning algorithms that fit input-output measurement data, with a few lines of code on Tensorflow or PyTorch.

The BuckiNet architecture on its own does not guarantee a dimensionless combination of the input parameters. Whether the network discovers a dimensionless basis without any constraint is discussed in Sec.B. To explicitly account for the constraint in equation (5), we add a the loss term

$$\mathcal{L}_{\text{null}} = \|\boldsymbol{D}_p\boldsymbol{\Phi}_p\|_2^2\,, \tag{9}$$

to the total loss function. This results in

$$\mathcal{L} = \|\boldsymbol{\Pi}_q - \psi(\exp\left(\log(\boldsymbol{P})\boldsymbol{\Phi}_p\right))\|_2 + \lambda\|\boldsymbol{D}_p\boldsymbol{\Phi}_p\|_2 + \text{reg.} \tag{10}$$

In contrast to the constrained optimization problem proposed in the previous section, this method minimizes the null-space according to equation (5) but does not satisfy it exactly. However, this proves to be sufficient for finding Pi-groups with sparse and approximately rational exponents as shown in Sec.3.1. The regularization term includes the L_1 and L_2 losses on $\mathbf{\Phi}_p$ to promote simple dimensionless groups with lower powers.

2.3 Sparse identification of dimensionless equations

The previous two methods addressed the problem of data-driven non-dimensionalization by simultaneously optimizing for the fitting function ψ and the dimensionless groups $\boldsymbol{\pi}$. In light of the fact that ψ is usually the solution of a differential equation, we propose casting the problem as a sparse identification of a differential equation (SINDy) [19, 20] with candidate dimensionless groups as coefficients. This constrains the dimensionless groups to be physically meaningful according to their associated differential operators that act on the unknown output variables. This is the intuition often used in classical ways of non-dimensionalization.

In the following formulation, we assume that the time t is the only dependent variable, and non-dimensionalize it separately from the rest of the input parameters. However, the method can be generalized to any number of dependent variables. For a dimensionless quantity of interest $\boldsymbol{\pi}_q(t)$, our goal is to learn an equation of the form

$$\frac{\mathrm{d}\boldsymbol{\pi}_q}{\mathrm{d}t} \equiv \frac{1}{T}\frac{\mathrm{d}\boldsymbol{\pi}_q}{\mathrm{d}\tau} = \mathcal{F}(\boldsymbol{\pi}_q;\ \boldsymbol{\pi}_p), \tag{11}$$

where T is a characteristic time scale, $\tau = t/T$ is the corresponding dimensionless time, and $\mathcal{F}$ is a differential operator. In the absence of a governing equation, SINDy approximates $\mathcal{F}$ by a sum of differential operators with fitting coefficients that are optimized for both sparsity and prediction error [19]. That is, given input-output pairs $\{\boldsymbol{\pi}_p, \boldsymbol{\pi}_q\}$, SINDy minimizes the loss

$$\mathcal{L}_{\text{SINDy}}(\boldsymbol{\pi}_q, \boldsymbol{\pi}_p, T;\ \mathbf{\Xi}) = \left\| \frac{1}{T}\frac{\mathrm{d}\boldsymbol{\pi}_q}{\mathrm{d}\tau} - \mathbf{\Theta}\left(\boldsymbol{\pi}_q, \boldsymbol{\pi}_p;\ T\right)\mathbf{\Xi} \right\|_2^2 + \lambda \left\|\mathbf{\Xi}\right\|_0, \tag{12}$$

where $\mathbf{\Xi}$ contains unknown fitting coefficients and $\mathbf{\Theta}$ is a pre-determined library of potential candidate functions and differential operators, a linear combination of which makes up the approximation $\mathcal{F} = \mathbf{\Theta}\left(\boldsymbol{\pi}_q, \boldsymbol{\pi}_p\right)\mathbf{\Xi}$.

Assuming that $\mathcal{F}(\boldsymbol{\pi}_q;\ \ \boldsymbol{\pi}_p)$, and its corresponding dictionary $\mathbf{\Theta}$, are *separable*, we write

$$\mathbf{\Theta}(\boldsymbol{\pi}_q;\ \boldsymbol{\pi}_p, T) = \boldsymbol{g}(\boldsymbol{\pi}_p) \otimes \hat{\mathbf{\Theta}}(\boldsymbol{\pi}_q, T), \tag{13}$$

where $\hat{\mathbf{\Theta}}$ is a dictionary of s derivatives in the dimensionless quantity of interest $\boldsymbol{\pi}_q$, $\boldsymbol{g}(\mathbf{x}) = [\mathbf{x}, \mathbf{x}^2, \ldots]$ is a dictionary of simple fractional powers, and $\otimes$ is

the Kronecker product which is a vectorization of the outer product of two vectors.

Having access to dimensional data, we optimize over candidate non-dimensionalizations by minimizing the following loss function

$$\mathcal{L}_{\text{dSINDy}}(\mathbf{q}, \mathbf{p};\ \boldsymbol{\Xi}, \boldsymbol{\Phi}) = \mathcal{L}_{\text{SINDy}}\left(\boldsymbol{\pi}_p(\mathbf{p};\ \boldsymbol{\Phi}), \boldsymbol{\pi}_q(\mathbf{q};\ \boldsymbol{\Phi}), T(\mathbf{p}, \boldsymbol{\Phi});\ \boldsymbol{\Xi}\right), \tag{14}$$

where the input-output pairs $(\mathbf{p}, \mathbf{q})$ are sampled from the dimensional data.

To account for the dimensionless constraint, we generate a finite number of candidate dimensionless groups up to a predetermined fractional power that satisfy equation (5). The resulting set of non-dimensionalizations correspond to a set of power matrices $\bar{\boldsymbol{\Phi}} = \{\bar{\boldsymbol{\Phi}}_1, \bar{\boldsymbol{\Phi}}_2, \ldots, \bar{\boldsymbol{\Phi}}_r\}$ over which we minimize the loss

$$\check{\boldsymbol{\Xi}}, \check{\boldsymbol{\Phi}} = \arg\min_{\boldsymbol{\Xi}, i} \mathcal{L}_{\text{dSINDy}}(\mathbf{p}, \mathbf{q};\ \boldsymbol{\Xi}, \bar{\boldsymbol{\Phi}}_i), \tag{15}$$

with $i \in [1, r]$. r depends on the range of predetermined powers from which dimensionless numbers are sampled according to the null-space condition in (5).

This method solves two problems at once, providing

1. Dimensionless input parameters $\boldsymbol{\pi}_p$ and dimensionless dependent variables $\tau = t/T$.
2. A sparse and parametric dynamical system $\dot{\boldsymbol{\pi}}_q = \mathcal{F}(\boldsymbol{\pi}_q;\ \boldsymbol{\pi}_p)$.

While we only consider time as a dependent variable in this section, a generalization to spatial and other dependent variables is straightforward. The method also allows for combining known dimensionless numbers with unknown ones as often encountered in practical problems.

3 Results

In this section, we apply the three methods presented above on three non-dimensionalization problems: the bead on a rotating hoop in Sec. 3.1, the Blasius boundary layer in Sec. 3.2, and the Rayleigh-Bénard problem in Sec. 3.3. We discuss the advantages and shortcomings of each method in terms of accuracy, robustness, and speed in the context of our proposed hypothesis.

3.1 Bead on a rotating hoop

Consider a wire hoop with radius R, rotating about a vertical axis coinciding with its diameter at an angular velocity ω, as shown in Fig. 1. A bead with mass m slides along the wire with tangential damping coefficient b. The equation governing the dynamics of the angular position x of the bead with

respect to the vertical axis is known to be ([37] sec. 3.5)

$$mR\ddot{x} = -b\dot{x} - mg\sin x + mR^2\omega\sin x\cos x. \tag{16}$$

A traditional dimensional analysis leads to the following Pi-groups [37]

$$\gamma = \frac{R\omega^2}{g}, \qquad \epsilon = \frac{m^2gR}{b^2}, \qquad \tau = \frac{mg}{b}t, \tag{17}$$

where ϵ controls the inertial term and γ is a pitchfork bifurcation parameter that accounts for two additional fixed points at $x^* = \pm\arccos(\gamma)$, for $\gamma > 1$. Non-dimensionlizing equation (16) gives

$$\epsilon\frac{d^2x}{d\tau^2} = -\frac{dx}{d\tau} - \sin x + \gamma\sin x\cos x. \tag{18}$$

For $\epsilon \ll 1$ and $\gamma = \mathcal{O}(1)$, the system is overdamped and approximately first-order.

To test the BuckiNet and the constrained optimization algorithms, we solve the governing equations numerically to obtain 3000 solutions with mass, radius, damping coefficient, and angular velocity sampled from a uniform distribution. In order to recover γ and ϵ without explicitly accounting for the time scale, we use the dominant modes (obtained by a principle component analysis) of the time series solution matrix (where each row corresponds to a different parameter combination, and each column to a different time sample), so that the output $\boldsymbol{\pi}_q$ is the set of coefficients for the leading r principal components as shown in Fig. 1. Here $\boldsymbol{\pi}_q \equiv \mathbf{q} = x$ because the angle x is dimensionless. A sample of the time-series solutions for different parameters is shown in Fig. 2.

After optimizing over neural network hyperparameters – given the optimal architecture, with enough modes to represent the solution – BuckiNet recovers the correct dimensionless numbers shown in table 1. The discovered Pi-groups are scaled to make the power of R unity. Sub-optimal linear combinations of known dimensionless groups are often discovered (see discussion in appendix **??**).

Using the dimensionless SINDy approach, we generate 20 parameter combinations, sampled from a uniform distribution – in this case, accounting for the time t as an input. The library $\boldsymbol{\Theta}$ contains low-order polynomials in x and $\dot{x}$ that can approximate the trigonometric nonlinearity in (16) for relatively small values of x. For each candidate nondimensionalization generated from the nullspace of $\boldsymbol{D}$, the nonconvex optimization problem (12) is

$\boldsymbol{\Phi}$	m	R	b	g	ω
$\phi(\pi_1)$	0.0011	1.000	0.0001	-0.997	1.990
$\phi(\pi_2)$	1.991	1.000	-1.990	0.998	0.002

Table 1: Discovered Pi-groups powers with BuckiNet.

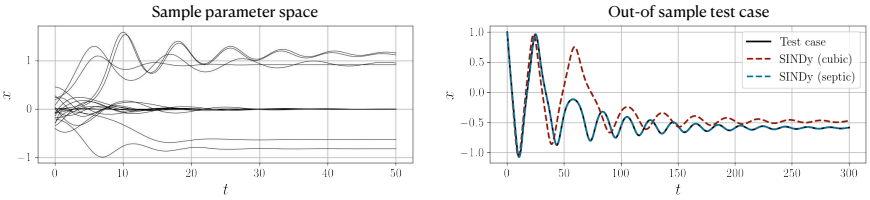

Fig. 2: Dimensionless SINDy applied on the rotating hoop problem. The discovered SINDy model reproduces the data and identifies the most physically relevant dimensionless groups.

approximated with the simple sequentially thresholded least squares algorithm [19], where the only tuning parameter is a threshold that approximates the ℓ_0 regularization loss.

Fig. 3 shows that the method identifies the same time scale $T = b/mg$ as that proposed by Strogatz [37] in equation (17), along with ratios of ϵ and γ which is consistent with equation (18) – divided by ϵ – such that

$$\pi_1 = \frac{b^2}{Rgm^2} = \frac{1}{\epsilon}, \qquad \pi_2 = \frac{\omega^2 g^2}{m^2 g^2} = \frac{\gamma}{\epsilon}. \tag{19}$$

Depending on the sparsity threshold in SINDy, models of varying fidelity can be identified. For example, with a threshold of 10^{-1} the algorithm selects a cubic model

$$\frac{\mathrm{d}^2 x}{\mathrm{d}\tau^2} = -0.96\pi_1 \frac{\mathrm{d}x}{\mathrm{d}\tau} - 0.94\pi_1 x + \pi_2(0.86x - 0.34x^3), \tag{20}$$

while reducing the threshold to 10^{-3} results in the seventh-order model

$$\frac{\mathrm{d}^2 x}{\mathrm{d}\tau^2} = -\pi_1 \frac{\mathrm{d}x}{\mathrm{d}\tau} - \pi_1(x - 0.16x^3 + 0.01x^5) + \pi_2(x - 0.66x^3 + 0.13x^5 - 0.01x^7). \tag{21}$$

This closely matches the Taylor-series approximation of the true dynamics (18). Accordingly, an appropriate model can be selected based on a desired tradeoff between accuracy and simplicity. However, higher order models increase the accuracy of the discovered dimensionless groups.

Fig. 2 tests the generalization of the two models on a set of parameters chosen to be outside the range of the training data (i.e. an extrapolation task). Although the cubic model captures the qualitative behavior reasonably well, the seventh-order approximation closely tracks the true solution.

3.2 Laminar boundary layer: identifying self-similarity

Non-dimensionalization often arises in the context of scaling and collapsing experimental results to lower dimensions by revealing the self-similarity

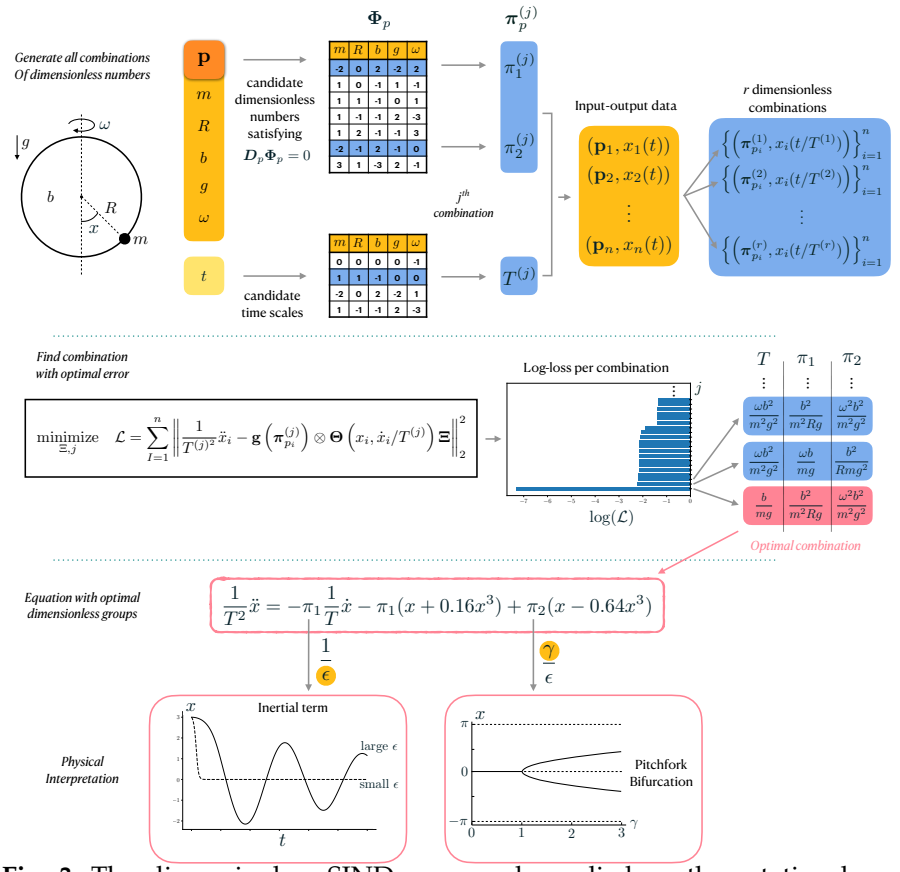

Fig. 3: The dimensionless SINDy approach applied on the rotating hoop problem. i) generate candidate dimensionless numbers and time scales from the nullspace of the input powers matrix $\boldsymbol{D}_p$, ii) choose a combination of Pi-groups and a time scale, iii) cast as a SINDy problem with the chosen Pi-groups as coefficients, iv) choose the combination with the lowest SINDy loss.

structure of the solution space. The discovery of self-similarity is considered an important result in many applications because it reveals universality properties.

Using scaling analysis, Prandtl showed that the Navier-Stokes equation describing an incompressible laminar boundary layer flow can be simplified to the *boundary layer equations* in the streamfunction $\Psi(x, y)$

$$\Psi_y \Psi_{xy} - \Psi_x \Psi_{yy} = \nu \Psi_{yyy}, \tag{22}$$

where the subscripts denote partial differentiation in x and y (with units of L), ν is the kinematic viscosity (with units of L^2/T), with $u(x, y) = \Psi_y$ and $v(x, y) = -\Psi_x$.

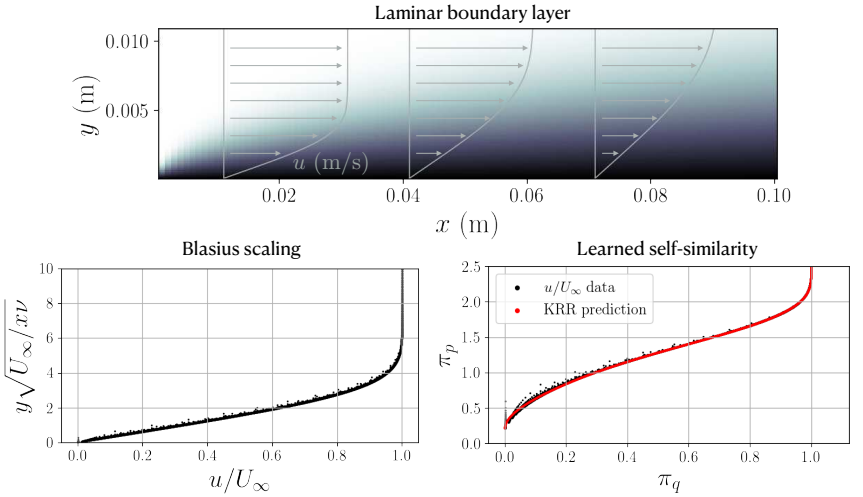

Fig. 4: Constrained optimization method for Blasius boundary layer problem, identifying the dimensionless group that collapses the input-output map to a single curve.

Blasius took the scaling one step further by showing that if $\Psi(x, y)$ is a solution of (22), then so is $\tilde{\Psi}(x, y) = \alpha\Psi(\alpha^2 x, \alpha y)$ for any dimensionless constant α, leading to the discovery of the dimensionless similarity variable η and streamfunction $f = f(\eta)$

$$\eta = y\sqrt{\frac{U_\infty}{\nu x}}, \qquad f(\eta) = \frac{\Psi(x, y)}{\sqrt{\nu U_\infty x}}, \tag{23}$$

with freestream velocity U_∞. The boundary layer equations reduce to the nonlinear boundary value problem

$$f'''(\eta) + \frac{1}{2}f''(\eta)f(\eta) = 0, \tag{24a}$$

$$f(0) = f'(0) = 0, f'(\infty) = 1. \tag{24b}$$

In this example, we show that the constrained optimization technique (in Sec. 2.1) can replace the long history of analytical transformations that have led to the discovery of the dimensionless variable η and its associated ordinary differential equation in $f(\eta)$.

Defining the output quantity of interest as the nondimensional streamwise velocity $\pi_q = u/U_\infty$, we seek to learn a model for $\pi_q = \psi(\pi_p)$, where π_p is an input dimensionless group that can depend on x, y, U_∞, and ν.

Here, we use kernel ridge regression (KRR) with a non-parametric radial basis function to approximate ψ.

We generate data for this example by solving the boundary layer equations (22) via a shooting method applied to (24). The free-stream velocity is chosen to be $U_\infty = 0.01m/s$ with viscosity $\nu = 10^{-6}m^2/s$, close to that of water at room temperature. The resulting two-dimensional profile $u(x, y)$ is shown in Fig. 4 (top), along with profiles at selected locations of x. 100 points in this field are selected randomly as training data and (8) is solved with a constrained trust region method implemented in Scipy.

Here, we use kernel ridge regression (KRR) with a non-parametric radial basis function to approximate ψ. The KRR model (implemented in scikit-learn) uses a ridge (ℓ_2) penalty of 10^{-4}, an ℓ_1 penalty of 10^{-4}, and a scale factor of 1. Moreover, since only 100 points are used in training, the performance of the final model can be evaluated against the entire field (10^4 data points). The optimization problem is inexpensive but non-convex and sensitive to the initial guess, which results in sub-optimal results. To address this issue, we run multiple optimizations with different initial guesses (around 20) and return the solution that has the minimum cost.

Fig. 4 compares the discovered nondimensionalization, $y^{0.46}U_\infty^{0.24}/(x^{0.22}\nu^{0.24}) \approx \sqrt{\eta}$, to the Blasius solution. When scaled to make the power of y unity, the discovered dimensionless number is

$$\pi_p = \frac{yU_\infty^{0.51}}{x^{0.49}\nu^{0.51}} \approx \eta. \tag{25}$$

The constrained optimization learns a different but equivalent model, in the sense that $\psi(\pi_p)$ is one-to-one with $f(\eta)$.

The scale of the solution is a balance between the ℓ_1 and ℓ_2 penalties and the scale factor in the radial basis functions. Setting the scale equal to one with small penalties thus biases the algorithm towards $\mathcal{O}(1)$ exponents, which is usually desirable in dimensionless groups.

3.3 Rayleigh-Bénard convection: learning a normal form

Characterizing the onset and behavior of instabilities is crucial for understanding large-scale dynamical systems. In this example, we use the SINDy method introduced in Sec. 2.3 to learn a dimensionless normal form of a Rayleigh-Bénard convection problem from limited time-series data, along with its corresponding dimensional parameters; particularly, the Rayleigh number. Rayleigh-Bénard convection is a prototypical example of a system with a global instability that has been used in a range of studies of nonequilibrium dynamics [13, 38, 39]. The system typically consists of a Boussinesq fluid between two plates, where the lower plate is at a higher temperature than the upper plate, given by the boundary conditions $T(z = 0) = T_0$, $T(z = L_z) = T_0 + \Delta T$. Below a critical Rayleigh number Ra_c, the only stable solution is steady conduction between the plates; above it, the density gradient becomes unstable to convection, as shown in the lower panels of Fig. 5. For a fluid in the Boussinesq approximation linearized about density

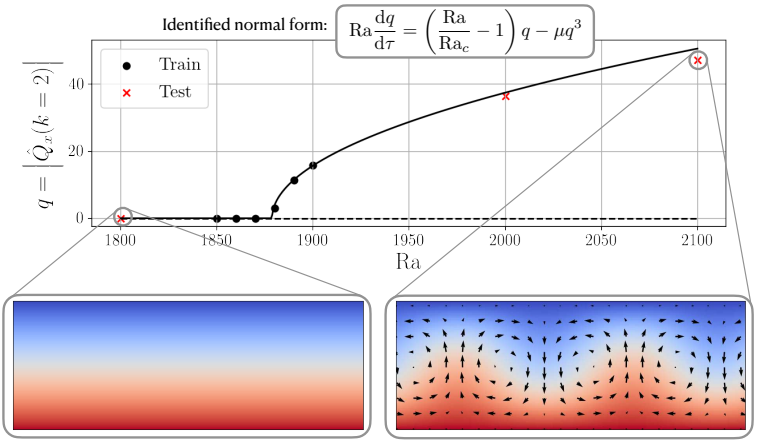

Fig. 5: The dimensionless SINDy method discovers the normal form of Rayleigh-Bénard convection problem with the analytical form of the Rayleigh as a cofficient.

ρ_0 and temperature T_0, the governing equations consist of the conservation of momentum, mass, and energy (see appendix). Besides the primitive variables, the system includes gravity g, coefficient of thermal expansion α, kinematic viscosity ν, and thermal diffusivity κ. Although the Prandtl number $\mathrm{Pr} = \nu/\kappa$ has a significant impact on the behavior of the flow above the threshold of instability [39], the onset of instability itself is not sensitive to it [40].

Having generated only 8 simulations (see implementation details in appendix), we use 5 points close to the bifurcation (black dots in Fig. 5) as a training set in the dimensionless SINDy problem and withhold the other 3 points (red crosses) as a test set. We use the wall-normal integral of the negative temperature gradient (proportional to heat flux) $q(t)$ as a quantity of interest. Since we expect a Landau-type dynamics model for the symmetry-breaking behavior, as a quantity of interest, we choose a cubic polynomial library in q and search for one dimensionless group π_p, along with an appropriate dimensionless time τ. Having 6 independent dimensional quantities in this problem $(H, g, \alpha, \Delta T, \nu, \kappa)$, it is computationally inexpensive to perform the optimization over all dimensionless numbers composed of integer powers up to ± 2.

The optimal dimensionless number discovered by the algorithm is the inverse of the Rayleigh number,

$$\pi_p = \mathrm{Ra}^{-1} = \frac{\nu\kappa}{g\alpha\Delta T L_z^3} \tag{26}$$

with dimensionless time

$$\tau = \frac{t\kappa^2}{\alpha^2 \nu (\Delta T)^2}, \tag{27}$$

as part of a normal form for a pitchfork bifurcation,

$$\mathrm{Ra}\frac{\mathrm{d}q}{\mathrm{d}\tau} = \left(\frac{\mathrm{Ra}}{\mathrm{Ra}_c} - 1\right) q - \mu q^3, \tag{28}$$

with estimated critical Rayleigh number $\mathrm{Ra}_c \approx 1878$, growth rate $\lambda \approx 1.7 \times 10^{-3}$ and Landau parameter $\mu \approx 4.6 \times 10^{-5}$.

The fixed points of this model can be easily found and compared to the steady states of the simulation, as shown in Fig. 5. The model closely matches the steady states not only of the training, but also of the withheld test points much farther from the bifurcation (shown as red crosses). While this is a simple model of a well-understood instability, the ability to directly derive normal forms from dimensional data opens the possibility of modeling more complex bifurcations, including cases where the control parameters are not well established.

4 Discussion

Finally, we've shown that data can be used to discover an optimal set of dimensionless numbers that honor the Buckingham Pi theorem according to an optimal input-output fit with dimensionless latent variables. In particular, we have shown that depending on the objective, there are three distinct methods by which we can produce a model through dimensional analysis. Each method is framed as an optimization problem which either imposes hard, soft or no constraints on the null-space of the dimensional matrix, favouring the robustness, accuracy and/or speed of the optimization procedure. However, the definition of that optimality is non-unique. Discovering a dimensionless equation through SINDy encodes an inductive bias towards low-order polynomials, which is often appropriate in dynamical systems applications or asymptotic expansions. On the other hand, the optimal dimensionless parameters as seen by more flexible and general representations such as neural networks or kernel regression may be completely different than those derived by classical analytic approaches giving insight into scaling Pi-groups that cannot be discovered by inspection. We demonstrated the application of each method on a set of well-known physics problems, showing that without any prior knowledge of the underlying physics, the various architectures are able to extract all the key concepts from the data alone. This includes extracting in an automated fashion, the system's symmetries, parametric dependencies and potential bifurcation parameters. Although modern machine learning can simply learn accurate representations of input and output relations, the imposition of Buckingham Pi theory allows for interpretable or explainable models.

Supplementary information. See attached file

Acknowledgments. The authors acknowledge support from the Army Research Office (ARO W911NF-19-1-0045) and the National Science Foundation AI Institute in Dynamic Systems (Grant No. 2112085). JLC acknowledges funding support from the Department of Defense (DoD) through the National Defense Science & Engineering Graduate (NDSEG) Fellowship Program.

Declarations

- Funding: Army Research Office (ARO W911NF-19-1-0045) and the National Science Foundation AI Institute in Dynamic Systems (Grant No. 2112085)
- There are no conflict of interests
- Data and code are available at: https://github.com/josephbakarji/buckidata
- J.B, J.L.C., S.L.B. and J.N.K. designed research; J.B. and J.L.C. performed research; and J.B, J.L.C., S.L.B. and J.N.K. wrote the paper.

References

[1] Barenblatt, G. I. *Scaling, self-similarity, and intermediate asymptotics: dimensional analysis and intermediate asymptotics* 14 (Cambridge University Press, 1996).

[2] Sterrett, S. G. in *Physically similar systems-a history of the concept* 377–411 (Springer, 2017).

[3] Buckingham, E. On physically similar systems; illustrations of the use of dimensional equations. *Physical review* **4** (4), 345 (1914) .

[4] del Rosario, Z., Lee, M. & Iaccarino, G. Lurking variable detection via dimensional analysis. *SIAM/ASA Journal on Uncertainty Quantification* **7** (1), 232–259 (2019) .

[5] Jofre, L., del Rosario, Z. R. & Iaccarino, G. Data-driven dimensional analysis of heat transfer in irradiated particle-laden turbulent flow. *International Journal of Multiphase Flow* **125**, 103198 (2020) .

[6] Fukami, K. & Taira, K. *Robust machine learning of turbulence through generalized buckingham pi-inspired pre-processing of training data*, A31–004 (2021).

[7] Xie, X., Liu, W. K. & Gan, Z. Data-driven discovery of dimensionless numbers and scaling laws from experimental measurements. *arXiv preprint arXiv:2111.03583* (2021) .

[8] Cerda, E. & Mahadevan, L. Geometry and physics of wrinkling. *Phys. Rev. Letters* **90** (7), 074302 (2003) .

[9] Morris, S. W., Bodenschatz, E., Cannell, D. S. & Ahlers, G. Spiral defect chaos in large aspect ratio Rayleigh-Bènard convection. *Physical Review Letters* **71** (13), 2026 (1993) .

[10] Shi, X. D., Brenner, M. P. & Nagel, S. R. A cascade of structure in a drop falling from a faucet. *Science* **265** (5169), 219–222 (1994) .

[11] Grzybowski, B., Stone, H. A. & Whitesides, G. M. Dynamic self-assembly of magnetized, millimetre-sized objects rotating at a liquid-air interface. *Nature* **405**, 1033–1036 (2000) .

[12] Seminara, A. *et al.* Osmotic spreading of bacillus subtilis biofilms driven by an extracellular matrix. *Proceedings of the National Academy of Sciences* **109** (4), 1116–1121 (2012) .

[13] Cross, M. C. & Hohenberg, P. C. Pattern formation outside of equilibrium. *Reviews of modern physics* **65** (3), 851 (1993) .

[14] Callaham, J. L., Koch, J. V., Brunton, B. W., Kutz, J. N. & Brunton, S. L. Learning dominant physical processes with data-driven balance models. *Nature communications* **12** (1), 1–10 (2021) .

[15] Holmes, P. & Guckenheimer, J. *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields* Vol. 42 of *Applied Mathematical Sciences* (Springer-Verlag, Berlin, Heidelberg, 1983).

[16] Yair, O., Talmon, R., Coifman, R. R. & Kevrekidis, I. G. Reconstruction of normal forms by learning informed observation geometries from data. *Proceedings of the National Academy of Sciences* 201620045 (2017) .

[17] Kalia, M., Brunton, S. L., Meijer, H. G., Brune, C. & Kutz, J. N. Learning normal form autoencoders for data-driven discovery of universal, parameter-dependent governing equations. *arXiv preprint arXiv:2106.05102* (2021) .

[18] Schmidt, M. & Lipson, H. Distilling free-form natural laws from experimental data. *science* **324** (5923), 81–85 (2009) .

[19] Brunton, S. L., Proctor, J. L. & Kutz, J. N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences* **113** (15), 3932–3937 (2016) .

[20] Rudy, S. H., Brunton, S. L., Proctor, J. L. & Kutz, J. N. Data-driven discovery of partial differential equations. *Science Advances* **3** (4), e1602614 (2017) .

[21] Lu, L., Jin, P., Pang, G., Zhang, Z. & Karniadakis, G. E. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature Machine Intelligence* **3** (3), 218–229 (2021) .

[22] Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics* **378**, 686–707 (2019) .

[23] Karniadakis, G. E. *et al.* Physics-informed machine learning. *Nature Reviews Physics* **3** (6), 422–440 (2021) .

[24] Noé, F., Olsson, S., Köhler, J. & Wu, H. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science* **365** (6457), eaaw1147 (2019) .

[25] Brenner, M., Eldredge, J. & Freund, J. Perspective on machine learning for advancing fluid mechanics. *Physical Review Fluids* **4** (10), 100501 (2019) .

[26] Duraisamy, K., Iaccarino, G. & Xiao, H. Turbulence modeling in the age of data. *Annual Reviews of Fluid Mechanics* **51**, 357–377 (2019) .

[27] Brunton, S. L., Noack, B. R. & Koumoutsakos, P. Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics* **52**, 477–508 (2020) .

[28] Sonnewald, M. *et al.* Bridging observations, theory, and numerical simulation of the ocean using machine learning. *Environmental Research Letters* **16** (7), 073008 (2021) .

[29] Kaiser, B. E., Saenz, J. A., Sonnewald, M. & Livescu, D. Objective discovery of dominant dynamical processes with intelligible machine learning. *arXiv:2106.12963* (2021) .

[30] Sonnewald, M., Wunsch, C. & Heimbach, P. Unsupervised learning reveals geography of global ocean regimes. *Earth and Space Science* **6**, 784–794 (2019) .

[31] Wu, H., Mardt, A., Pasquali, L. & Noe, F. Deep generative markov state models. *32nd Conference on Neural Information Processing Systems (NeurIPS)* (2018) .

[32] Champion, K., Lusch, B., Kutz, J. N. & Brunton, S. L. Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences* **116** (45), 22445–22451 (2019) .

[33] Bakarji, J., Champion, K., Kutz, J. N. & Brunton, S. L. Discovering governing equations from partial measurements with deep delay autoencoders. *arXiv preprint arXiv:2201.05136* (2022) .

[34] Constantine, P. G., del Rosario, Z. & Iaccarino, G. Data-driven dimensional analysis: algorithms for unique and relevant dimensionless groups. *arXiv preprint arXiv:1708.04303* (2017) .

[35] Udrescu, S.-M. & Tegmark, M. Ai feynman: A physics-inspired method for symbolic regression. *Science Advances* **6** (16), eaay2631 (2020) .

[36] Gunaratnam, D. J., Degroff, T. & Gero, J. S. Improving neural network models of physical systems through dimensional analysis. *Applied Soft Computing* **2** (4), 283–296 (2003) .

[37] Strogatz, S. H. *Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering* (CRC press, 2018).

[38] Lorenz, E. N. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences* **20** (2), 130–141 (1963) .

[39] Pandey, A., Scheel, J. D. & Schumacher, J. Turbulent superstructures in Rayleigh-Bènard convection. *Nature Communications* **9** (2118) (2018) .

[40] Chandrasekhar, S. *Hydrodynamic and Hydromagnetic Stability* (Clarendon Press, Oxford, 1961).

[41] Mortensen, M. Shenfun: High performance spectral galerkin computing platform. *Journal of Open Source Software* **3** (31), 1071 (2018) .

[42] Loiseau, J.-C. & Brunton, S. L. Constrained sparse Galerkin regression. *Journal of Fluid Mechanics* **838**, 42–67 (2018) .

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.pdf