Dimensionally Consistent Learning with Buckingham Pi

Joseph Bakarji¹, Jared Callaham¹, Steven L. Brunton¹, J. Nathan Kutz²

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, i.e. force is measured in Newtons which is $kg \cdot m/s^2$ and electric charge by the volt which is $kg \cdot m^2/(s^3 \cdot A)$. 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, physicsinformed 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.

¹ Department of Mechanical Engineering, University of Washington, Seattle, WA 98195, United States

² Department of Applied Mathematics, University of Washington, Seattle, WA 98195, United States

^{*} Corresponding author (jbakarji@uw.edu).

Buckingham Pi theory 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 inputoutput function approximations [22-24] from simulation and experimental data alone, with considerable progress in the field of fluid mechanics [25-28]. Particularly, unsupervised learning techniques have recently made significant progress in distilling physical data into interpretable dynamical models that automate the scientific process [29-33]. 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. Principal component analysis (PCA), which is a leading data analysis method [34], attempts to circumvent the idea that physical laws do not depend on the units of measurements by taking each measurement variable and setting its mean to zero and its variance to unity. However, there are more sophisticated methods for explicitly considering measurement units. Active subspaces provides a principled algorithmic method for extracting dimensionless numbers from experimental and simulation data [35], 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 [36] and to improve neural network models [37]. 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.

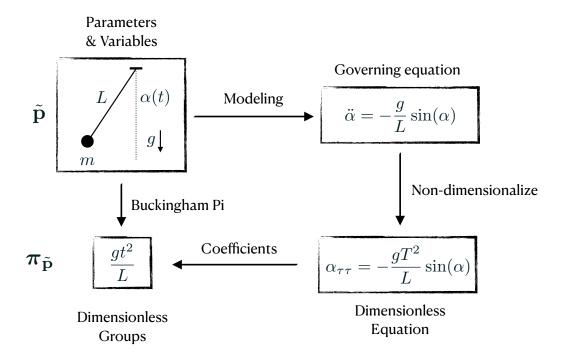


Figure 1: Two approaches for non-dimensionalization. In the absence of known governing equations, scientists use the Buckingham Pi theorem to infer dimensionless groups $\pi_{\tilde{p}}$ from units of input variables and parameters. When the equation is known, normalizing variables and parameters by system-specific constants and/or averages gives rise to dimensionless groups as equation coefficients.

Thus it relies on intuition and experience with the physical problem at hand. If the underlying equations are available, finding dimensionless groups is constrained by the form of the equation and they are computed by dividing the system average magnitude of every term by a reference term. Even then, the problem of determining the relevant dimensionless numbers depends on experience. 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 (See Fig. 1), 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 Buckingham Pi Problem Formulation

Let $\mathbf{q} = f(\mathbf{p})$ be a mapping where $\mathbf{p} \in \mathbb{R}^n$ is a vector of input measurement parameters and variables, and f is a function that maps \mathbf{p} to a set of dependent quantities of interest $\mathbf{q} \in \mathbb{R}^k$. One generally seeks a predictor f that maximizes the size of the set of possible measurable parameters and variables \mathcal{P} , for which the prediction error $\varepsilon = \|f(\mathbf{p}) - \mathbf{q}\|$ is minimized, with $\mathbf{p}, \mathbf{q} \in \mathcal{P}$. 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 $\pi = (\pi_p, \pi_q)$, also known as π -groups, such that the dimensionless input parameters/variables $\pi_p \in \mathbb{R}^{n'}$ span the full dimensionless solution space $\pi_q \in \mathbb{R}^{k'}$, with $n' \leq n$ and $k' \leq k$ [3]. Furthermore, the theorem determines the value of d' = n' + k' given the units of the measurement variables and parameters.

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 (π_p, π_q) without compromising the prediction error ε , such that,

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

where ψ is an unknown function that can be approximated from available data. However, the Buckingham Pi 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 a dimensionless basis with the generalization properties of the mapping f.

In a data-driven statistical setting, we assume that the inputs \mathbf{p} and outputs \mathbf{q} are sampled from a distribution function $F_{\mathbf{pq}}(P,Q)$, and that f is an optimal mapping in a given class of function. In this context, we generally seek the most physically meaningful transformation $(\mathbf{p},\mathbf{q}) \to (\pi_p,\pi_q)$ that provides the best fit $\pi_q = \psi(\pi_p)$, assuming ψ to be an optimal predictor over an pre-determined hypothesis class \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 π^* is the optimal coordinate transformation $\tilde{\mathbf{p}} \to \pi^*$ that satisfies the minimization problem

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

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

Note that defining *physical meaningfulness* often depends on the historical and scientific context, which makes the above hypothesis subject to interpretation. In some cases, theoretical and experimental tradition narrows a field of study down to a set of well-known dimensionless numbers with which known regimes and behaviors are associated (e.g. Reynolds number quantifying turbulence in fluid mechanics). Particularly, when analytical models and equations are available, dimensionless numbers that arise naturally from the relative measure of different terms are physically significant because they quantify changes in the qualitative nature of the solution. Hypothesis 1 proposes a new approach for defining *physical meaningfulness* by first providing evidence for its validity in Sec.3. The purpose of the proposed methods is to discover dimensionless groups in contexts where data is available but analytical models are not known.

This hypothesis is further constrained by the functional relationship between π and $\tilde{\mathbf{p}}$ 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 $\phi(\pi_j) = [\Phi_{1j}, \Phi_{2j}, \dots, \Phi_{dj}]^T$, where d = n + k.

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

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

such that

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

Here, we assume that the number of dimensional and dimensionless predictions are equal, i.e. $|\pi_q| = |\mathbf{q}| = k$, which is often the case, and we later propose methods for relaxing this assumption. The Buckingham Pi theorem determines the number of Pi-groups to be $d' = d - \operatorname{rank}(\mathbf{D})$. The main shortcoming of the theorem is its inability to provide a unique set of d' dimensionless numbers. In practice, additional heuristic constraints are required to solve for the unknowns s_{ij} .

3 Methods

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 *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.

Let $\{\tilde{\mathbf{p}}^{(i)}\}_{i=1}^m \equiv \{(\mathbf{p}^{(i)},\mathbf{q}^{(i)})\}_{i=1}^m$ be a set of 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 $\mathbf{P} \in \mathbb{R}^{m \times n}$ and the output prediction matrix $\mathbf{Q} \in \mathbb{R}^{m \times k}$ as the row-wise concatenation of all measurements $\mathbf{p}^{(i)}$ and $\mathbf{q}^{(i)}$ respectively. We also define $\tilde{\mathbf{P}} = [\mathbf{P}, \mathbf{Q}] \in \mathbb{R}^{m \times d}$ as the column-wise concatenation of \mathbf{P} and \mathbf{Q} . The Pi-groups powers matrix is given by

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

such the equation $D\Phi = \mathbf{0}$ is satisfied according to Eq.(5). Similarly, we define $\Phi_p \in \mathbb{R}^{n \times n'}$ to contain the n' dimensionless groups corresponding to the n input parameters/variables \mathbf{p} . Equation (3) can be written as $\pi_i = \exp\left\{\sum_{j=1}^d \Phi_{ij} \log(\tilde{p}_j)\right\}$, corresponding to the matrix form

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

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

For example, consider the pendulum shown in Fig. 1. If we define the inputs to be gravity g, mass m, length L, and time t, and the output to be the angle α , then the dimensional input and output parameter vectors \mathbf{p} and \mathbf{q} are

$$\mathbf{p} = \begin{bmatrix} g & m & L & t \end{bmatrix}^T, \qquad \mathbf{q} = \begin{bmatrix} \alpha \end{bmatrix}. \tag{8}$$

The corresponding units matrices D_p and D_q constructed with basic dimensions mass, length, and time are

$$\mathbf{D}_{p} = \begin{bmatrix} & | & | & | & | \\ \Omega(g) & \Omega(m) & \Omega(L) & \Omega(t) \\ | & | & | & | \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ -2 & 0 & 0 & 1 \end{bmatrix}$$
(9a)

$$\boldsymbol{D}_{q} = \begin{bmatrix} & & & \\ & \Omega(\alpha) & & \\ & & & \end{bmatrix} = \begin{bmatrix} & 0 & \\ & 0 & \\ & & & \end{bmatrix}. \tag{9b}$$

Note that the angle α is already dimensionless, so $\Omega(\alpha)$ is the zero vector. A possible Φ matrix consisting of columns spanning the nullspace of $D = [D_p, D_q]$ is

$$\mathbf{\Phi} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ -1 & 0 \\ 2 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{s.t.} \quad \mathbf{\Phi}_p = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 2 \\ 0 \end{bmatrix}, \quad \mathbf{\Phi}_q = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}. \tag{10}$$

Then $\pi_q = \alpha$ and $\pi_p = gt^2/L$, so that the input/output relationship is some function $\alpha = \psi(gt^2/L)$. We solve this example using a data-driven method in Sec. 4.1.

In this case we know from basic mechanics that $\psi() \equiv \sin()$ for small α , but for more complicated problems it may not be obvious what the appropriate Φ and ψ are. This section introduces three methods to simultaneously identify the nondimensionalization Φ 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, $D\Phi=0$, to optimize ψ over a pre-defined hypothesis class given the measurement data.

3.1 Constrained optimization

Equation (2) can be cast as a constrained optimization problem, with a fitting function ψ that can be either parameter or non-parametric. The input Pi-groups powers matrix Φ_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. Π_q is known). Accordingly, the resulting optimization problem is given by

$$\check{\mathbf{\Phi}}_{p} = \underset{\mathbf{\Phi}_{p}}{\operatorname{argmin}} \|\mathbf{\Pi}_{q} - \psi(\exp(\log(\mathbf{P})\mathbf{\Phi}_{p}))\|_{2} + \lambda_{1} \|\mathbf{\Phi}_{p}\|_{1} + \lambda_{2} \|\mathbf{\Phi}_{p}\|_{2}, \quad \text{s.t.} \quad \mathbf{D}_{p}\mathbf{\Phi}_{p} = \mathbf{0},$$
 (11)

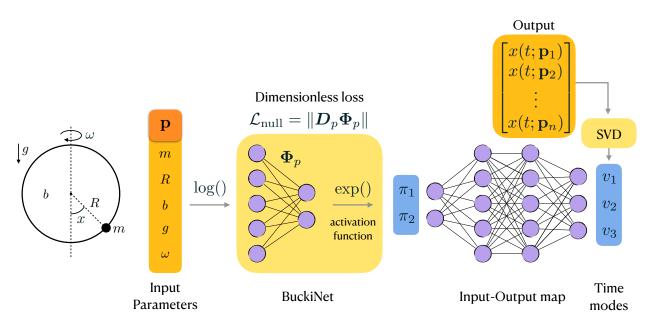


Figure 2: Illustration of the BuckiNet layer for the rotating hoop problem (described in section 4.2). The dimensionless loss imposes a soft Buckingham Pi constraint from equation (5)) and the BuckiNet layer satisfies equation (3). In this example, the output is given by the three dominant time modes ${\bf v}$ of the output x(t) rather than taking the time t as an input.

where the ℓ_1 regularization enforces sparsity (typical in dimensionless numbers) and the ℓ_2 regularization encourages smaller powers. An example application of this approach that uses kernel ridge regression as a non-parametric approximation of ψ is given in Sec. 4.3.

3.2 BuckiNet: non-dimensionalization with a neural network

The unknown function ψ is generally nonlinear and can be high dimensional. Given input-output data, in the absence of a governing equation, a deep neural network is a natural 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. This makes Φ_p the fitting weights of the first layer, as shown in Fig. 2. We call this layer the BuckiNet. When negative input data is given, it has to be pre-processed so that the initial $\log()$ operation becomes possible. In most cases, a simple shifting of the data to the positive domain is sufficient.

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. The number of nodes in the first hidden layer is n' as determined by the Buckingham Pi theorem. Second, the optimal weights of the first layer, Φ_p , correspond to the powers of dimensionless numbers that are most predictive of the solution. Third, 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.4.1. To explicitly account for the constraint in equation (5), we add a null-space loss on Φ_p

$$\mathcal{L}_{\text{null}} = \|\boldsymbol{D}_{p}\boldsymbol{\Phi}_{p}\|_{2}^{2},\tag{12}$$

in the loss function. Therefore, the total loss is given by

$$\mathcal{L} = \|\mathbf{\Pi}_q - \psi(\exp(\log(\mathbf{P})\mathbf{\Phi}_p))\|_2 + \lambda \|\mathbf{D}_p\mathbf{\Phi}_p\|_2 + \text{reg}.$$
 (13)

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.4.2. The regularization term includes the L_2 loss $\|\mathbf{\Phi}_p\|_2$ that promotes dimensionless groups with lower powers, and the L_1 loss $\|\mathbf{\Phi}_p\|_1$ that promotes simple groups that have as little input variables/parameters as possible.

3.3 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 π . 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 differential equations (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, which is also the intuition often used in classical 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 $\pi_q(t)$, our goal is to learn an equation of the form

$$\frac{\mathrm{d}\boldsymbol{\pi}_q}{\mathrm{d}t} \equiv \frac{1}{T} \frac{\mathrm{d}\boldsymbol{\pi}_q}{\mathrm{d}\tau} = \mathcal{F}(\boldsymbol{\pi}_q; \boldsymbol{\pi}_p),\tag{14}$$

where T is a characteristic time scale, $\tau = t/T$ is the corresponding dimensionless time, π_p is a vector of dimensionless parameters, 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 $\{\pi_p, \pi_q\}$ sampled from a time series, SINDy minimizes the loss

$$\mathcal{L}_{\text{SINDy}}(\boldsymbol{\pi}_{q}, \boldsymbol{\pi}_{p}, T; \boldsymbol{\Xi}) = \left\| \frac{1}{T} \frac{\mathrm{d}\boldsymbol{\pi}_{q}}{\mathrm{d}\tau} - \boldsymbol{\Theta} \left(\boldsymbol{\pi}_{q}, \boldsymbol{\pi}_{p}; T \right) \boldsymbol{\Xi} \right\|_{2}^{2} + \lambda \left\| \boldsymbol{\Xi} \right\|_{0}, \tag{15}$$

where Ξ contains unknown fitting coefficients and Θ is a pre-determined library of potential candidate functions and differential operators, a linear combination of which makes up the approximation $\mathcal{F} = \Theta\left(\pi_q, \pi_p\right) \Xi$. For dimensional consistency, derivatives in the dependent variable t are scaled by an unknown problem specific timescale T, which we assume can be expressed as a function of the input parameters: $T = T(\mathbf{p})$. In essence, the timescale T is treated in the same manner as the dimensionless groups π , except that its dimensions are constrained to be those of time.

Assuming that $\mathcal{F}(\pi_q; \pi_p)$, and its corresponding dictionary Θ , are *separable*, we can write

$$\Theta(\boldsymbol{\pi}_q; \boldsymbol{\pi}_p, T) = \boldsymbol{g}(\boldsymbol{\pi}_p) \otimes \hat{\boldsymbol{\Theta}}(\boldsymbol{\pi}_q, T), \tag{16}$$

where $\hat{\Theta}$ is a dictionary of s derivatives in the dimensionless quantity of interest π_q , π_p is a vector of dimensionless input parameter excluding time, g() is a dictionary of polynomial functions, and \otimes is the Kronecker product which is a vectorization of the outer product of two vectors. The assumption of separability is not strictly necessary; candidate functions may be chosen that are functions of both π_p and π_q . However, SINDy libraries are often constructed as multinomials, so that the variables are separable. Moreover, the parameters typically appear as coefficients of the state variable in both bifurcation normal forms and Taylor series approximations of dynamical systems, so the assumption of separability is natural.

The resulting feature space dimension is $\Theta(\pi_q; \pi_p) \in \mathbb{R}^{ns}$. Having m measurements, we define the full dictionary has dimensions $\Theta \in \mathbb{R}^{m \times ns}$, where every row is an examples (i.e. a sample in time) and every column a potential term in the differential equation. Accordingly, $\Xi \in \mathbb{R}^{ns \times k'}$ where k' is the dimension of π_q .

While equation (15) takes dimensionless pairs as input data, we would also like to optimize over candidate non-dimensionalizations. That is, we would like to minimize the following loss function

$$\mathcal{L}_{\text{dSINDv}}(\mathbf{q}, \mathbf{p}; \mathbf{\Xi}, \mathbf{\Phi}) = \mathcal{L}_{\text{SINDv}}(\boldsymbol{\pi}_{p}(\mathbf{p}; \mathbf{\Phi}), \boldsymbol{\pi}_{q}(\mathbf{q}; \mathbf{\Phi}), T(\mathbf{p}, \mathbf{\Phi}); \mathbf{\Xi}), \tag{17}$$

where the input-output pairs (\mathbf{p}, \mathbf{q}) are sampled from the dimensional data, and Φ is the Pi-groups powers matrix that defines the mapping $\pi(\mathbf{x}; \Phi) = \exp(\log(\mathbf{x})\Phi)$ given by equation (7).

Equation (17) does not include a constraint on Φ (Eq. (5)) to ensure that π is dimensionless. To address this issue, we generate a finite number of candidate dimensionless numbers up to a predetermined fractional power that satisfy equation (5). The resulting set of non-dimensionalizations correspond to a set of power matrices $\bar{\Phi} = \{\bar{\Phi}_1, \bar{\Phi}_2, \dots, \bar{\Phi}_r\}$ over which we minimize the loss

$$\check{\mathbf{\Xi}}, \check{\mathbf{\Phi}} = \operatorname*{argmin}_{\mathbf{\Xi},i} \mathcal{L}_{dSINDy}(\mathbf{p}, \mathbf{q}; \mathbf{\Xi}, \bar{\mathbf{\Phi}}_i), \tag{18}$$

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).

To avoid this "brute-force" search, the dimensionless SINDy method could be combined with the constrained optimization approach described in Sec. 3.1. In general, this tactic would be more flexible, since it also allows for non-integer powers in the dimensionless groups. We use brute-force enumeration in this work because we expect integer powers, there are relatively few nullspace vectors, and the sparse regression problem can be solved efficiently, so this method scales relatively well in this particular case. We explore a more efficient approach in future work. The general computational efficiency of the brute-force search approach will be discussed in the results.

In summary, this method solves two problems at once, providing

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

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.

4 Results

In this section, we apply the three methods presented above on four non-dimensionalization problems: the harmonic oscillator in Sec. 4.1, the bead on a rotating hoop in Sec. 4.2, the Blasius boundary layer in Sec. 4.3, and the Rayleigh-Bénard problem in Sec. 4.4. We discuss the advantages and shortcomings of each method in terms of accuracy, robustness, and speed in the context of our proposed hypothesis.

4.1 Non-dimensionalization as optimal change of variables: harmonic oscillator

Dimensionless groups can act as scaling parameters (i.e. revealing self-similarity), as they often arise in analytical solutions. In some problems, Pi-groups can be deduced from an optimal change of variables without the need for a Buckingham Pi constraint. While this is not always the case, it supports the hypothesis that an optimal mapping between input and output data gives rise to a physically meaningful change of variables. We demonstrate this point on the simple pendulum problem described in Sec.3 and shown in Fig. 1.

We evaluate the output prediction, α , at 100 parameter combinations of the inputs L, m and g sampled from a uniform distribution, and at 100 time steps, t. In this problem, we employ the BuckiNet architecture without the Buckingham Pi constraint (i.e. without $\mathcal{L}_{\text{null}}$) with four inputs [L, m, g, t], one perceptron in the first hidden layer, a dimensionless output α , 3 layers with 8 perceptrons each for ψ , and an exponential linear unit (ELU) activation function. The resulting dimensionless group in the BuckiNet layer is

$$\pi_p = \frac{gt^2}{L},\tag{19}$$

which appears in the solution of the linear approximation of a harmonic oscillator as

$$\alpha(t) = \alpha_0 \cos(\sqrt{\pi_p} + \theta), \tag{20}$$

where θ is a constant that depends on the initial condition. This is a case where a change of variables that gives the simplest analytical solution is dimensionless. However, this is not generally the case, especially in higher dimensional problems. To fully take advantage of the Buckingham Pi constraint, we apply the three methods proposed in Sec.3 in the following examples.

4.2 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. 2. 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 given by ([38] sec. 3.5)

$$mR\ddot{x} = -b\dot{x} - mg\sin x + mR^2\omega\sin x\cos x. \tag{21}$$

A traditional dimensional analysis leads to the following Pi-groups [38]

$$\gamma = \frac{R\omega^2}{a}, \qquad \epsilon = \frac{m^2gR}{b^2}, \qquad \tau = \frac{mg}{b}t, \qquad (22)$$

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 (21) gives

$$\epsilon \frac{d^2x}{d\tau^2} = -\frac{dx}{d\tau} - \sin x + \gamma \sin x \cos x. \tag{23}$$

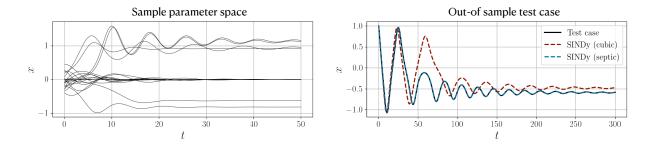


Figure 3: Dimensionless SINDy applied on the rotating hoop problem. The discovered SINDy model reproduces the data and identifies the most physically relevant dimensionless groups.

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 Principal Component Analysis of 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 π_q is the set of coefficients for the leading r principal components as shown in Fig. 2. Here $\pi_q \equiv \mathbf{q} = x$ because the angle x is dimensionless. A sample of the time-series solutions for different parameters is shown in Fig. 3.

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. In order to understand the limitations of the method, it is informative to look at sub-optimal solutions. For instance, for some hyperparameters (e.g. network architecture, regularization etc.) BuckiNet and the constrained optimization method give a product of either ϵ or γ and a Pi-group that is closely related to τ as a solution, such that

$$\phi(\pi_1) = \phi(\pi_2) + c\phi\left(\frac{mg}{\omega b}\right),\tag{24}$$

where c is an arbitrary constant, and $mg/(\omega b)$ is the closest approximation of $\tau=mgt/b$, given that t is not included as an input. This shows that a product of multiple Pi-groups is a local minimum that satisfies the Buckingham Pi and spans the solution space in a different coordinate system. However, a hyperparameter optimization over multiple trials will consistently result in Pi-groups that closely approximate γ and ϵ . This motivates the use of SINDy to further constrain the dimensionless numbers to be coefficients in a differential equation where they acquire a concrete mathematical meaning as controllers of the dominant balance.

Using the dimensionless SINDy approach, we generate 20 parameter combinations, sampled

Φ	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.

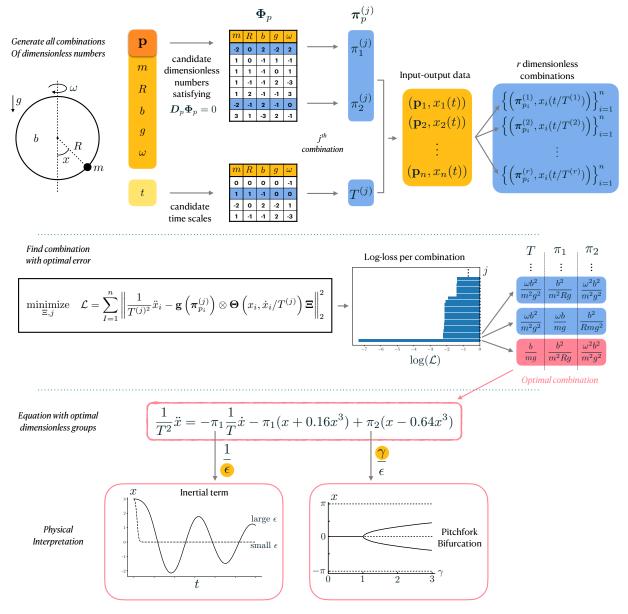


Figure 4: The dimensionless SINDy approach applied on the rotating hoop problem. i) generate candidate dimensionless numbers and time scales from the nullspace of the input powers matrix D_p , ii) choose a combination of Pi-groups and a time scale, iii) cast as a SINDy problem with the chosen Pi-groups as coefficients, iv) choose the combination with the lowest SINDy loss.

from a uniform distribution – in this case, accounting for the time t as an input. The library Θ contains low-order polynomials in x and \dot{x} that can approximate the trigonometric nonlinearity in (21) for relatively small values of x. For each candidate nondimensionalization generated from the nullspace of D, the nonconvex optimization problem (15) is 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. 4 shows that the method identifies the same time scale T = b/mg as that proposed by Strogatz [38] in equation (22), along with ratios of ϵ and γ which is consistent with equation (23) –

divided by ϵ – such that

$$\pi_1 = \frac{b^2}{Rgm^2} = \frac{1}{\epsilon}, \qquad \qquad \pi_2 = \frac{\omega^2 g^2}{m^2 g^2} = \frac{\gamma}{\epsilon}.$$
(25)

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

$$\frac{\mathrm{d}^2 x}{\mathrm{d}\tau^2} = -0.96\pi_1 \frac{\mathrm{d}x}{\mathrm{d}\tau} - 0.94\pi_1 x + \pi_2 (0.86x - 0.34x^3),\tag{26}$$

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

$$\frac{\mathrm{d}^2 x}{\mathrm{d}\tau^2} = -\pi_1 \frac{\mathrm{d}x}{\mathrm{d}\tau} - \pi_1 (x - 0.16x^3 + 0.01x^5) + \pi_2 (x - 0.66x^3 + 0.13x^5 - 0.01x^7). \tag{27}$$

This closely matches the Taylor-series approximation of the true dynamics (23). 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. 3 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.

4.3 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 structure of the solution space. The discovery of self-similarity is considered an important result in many applications because it reveals universality properties. There are plenty of such examples in the history of fluid dynamics, particularly in understanding turbulence [1, 39].

Boundary layer theory is an example where of nondimensionalization and self-similarity have played a central role in this understanding [40]. 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 Ψ

$$\Psi_y \Psi_{xy} - \Psi_x \Psi_{yy} = \nu \Psi_{yyy},\tag{28}$$

where the subscripts denote partial differentiation in x and y, ν is the kinematic viscosity (with units of L^2/T). Ψ is defined so that the streamwise and wall normal velocities are respectively given by

$$u(x,y) = \Psi_y, \qquad v(x,y) = -\Psi_x. \tag{29}$$

Although Prandtl's boundary layer equations are themselves a significant simplification and triumph of scaling analysis, they can be further simplified and expressed as an ordinary differential equation with the help of self-similarity. Blasius took the scaling one step further by showing that if $\Psi(x,y)$ is a solution of (28), then so is $\tilde{\Psi}(x,y) = \alpha \Psi(\alpha^2 x, \alpha y)$ for any dimensionless constant α . This in turn implies that Ψ itself is not an independent function of x and y, but of the combination x/\sqrt{y} .

Defining the dimensionless similarity variable η and streamfunction $f = f(\eta)$ as

$$\eta = y\sqrt{\frac{U_{\infty}}{\nu x}}, \qquad f(\eta) = \frac{\Psi(x,y)}{\sqrt{\nu U_{\infty} x}},$$
(30)

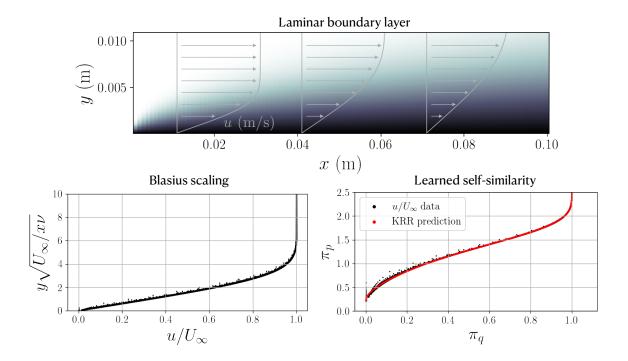


Figure 5: Constrained optimization method for Blasius boundary layer problem, identifying the dimensionless group that collapses the input-output map to a single curve.

with freestream velocity U_{∞} , the boundary layer equations reduce to the nonlinear boundary value problem

$$f'''(\eta) + \frac{1}{2}f''(\eta)f(\eta) = 0, (31a)$$

$$f(0) = f'(0) = 0, f'(\infty) = 1.$$
 (31b)

Although there is no known closed-form solution to this problem, it can be analyzed with asymptotic perturbation methods or solved numerically (e.g. with a shooting method to identify an appropriate initial value for f''(0)).

In contrast to the rotating hoop example, in this case the velocity profile is more important than the differential equation itself. For instance, once the Blasius solution $f(\eta)$ is known, the boundary layer profile can be found by undoing the scaling, e.g. $u(x,y) = U_{\infty}f'(\eta)$. The most natural method to directly identify a dimensionless group associated with the boundary layer is therefore the constrained optimization approach introduced in Sec. 3.1.

Defining the output quantity of interest as the nondimensional streamwise velocity $\pi_q = u/U_{\infty}$, the problem is 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 . In this problem we seek to rediscover the Blasius scaling which was discovered analytically by Blasius and Prandtl with $\pi_p = \eta$ and $\psi = f'(\eta)$.

We use kernel ridge regression (KRR) with a non-parametric radial basis function to approximate ψ . Accordingly, the constrained optimization (11) must learn a dimensionless group π_p in the nullspace of the units matrix \boldsymbol{D} that leads to a good KRR approximation of π_q as a function of π_q .

We generate data for this example by solving the boundary layer equations (28) via a shooting method applied to (31). 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. 5 (top), along with profiles at selected locations of x. 100 points in this field are selected randomly as training data and (11) is solved with a constrained trust region method implemented in Scipy.

Specifically, for each candidate nondimensionalization, the kernel ridge regression model with radial basis function kernels (implemented in scikit-learn) is trained and evaluated on the dimensionless parameters computed from the 100 test points. The KRR model uses ridge (ℓ_2) penalty of 10^{-4} , and ℓ_1 penalty of 10^{-4} , and a scale factor of 1. Note that in this method, generalizing on a test case is not crucial because the model is not used to make prediction but only as a proxy for the mutual information between the candidate nondimensionalization and the quantity of interest. In other words, the method does not have to address problems of overfitting. Moreover, since only 100 points are used in training, the performance of the final model can be evaluated against the entire field (10^4 points). The optimization problem is inexpensive but non-convex and sensitive to the initial guess, thus problem to converging to local minima. To address this issue, we run multiple optimizations with different initial guesses (around 20) and return the solution that has the minimum cost.

Fig. 5 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

$$\pi_p = \frac{yU_{\infty}^{0.51}}{x^{0.49}\nu^{0.51}} \approx \eta. \tag{32}$$

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

In contrast to the brute force-type search over candidate nondimensionalizations generated from a search over vectors of integers, the exponents in this method are floating-point numbers and are arbitrary up to an overall constant. The scale of the solution is therefore 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.

4.4 Rayleigh-Bénard convection: learning a normal form

Characterizing the onset and behavior of instabilities is crucial for understanding large-scale dynamical systems. Near a critical point, a reduced-order amplitude equation called the normal form can model the qualitative change in dynamics associated with a bifurcation. Although the normal form can be deduced analytically [41], numerically [42, 43], or via symmetry arguments [44], these methods become progressively more challenging with more complex systems. With the exception of the symmetry analysis, they are also invasive because they require direct access to the (discretized) governing equations. In this example, we use the SINDy method introduced in Sec. 3.3 to learn a dimensionless normal form from limited time-series data, along with the corresponding dimensional parameters.

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, including pattern formation [13], chaos [45], and coherent structures in turbulence [46]. The system typically consists of a Boussinesq fluid between two plates, where the lower plate is at a higher temperature than the upper plate. Heat transfer can take place via either conduction or convection, whose relative strength is encapsulated in the dimensionless Rayleigh number. Below a critical Rayleigh number, the only

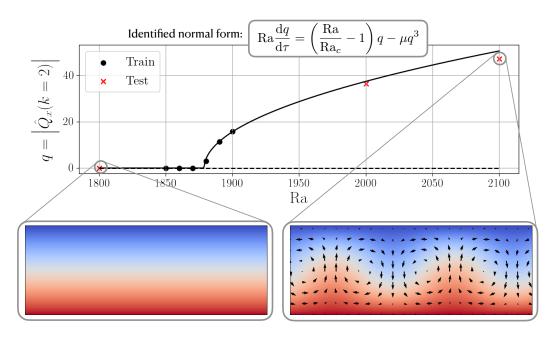


Figure 6: The dimensionless SINDy method discovers the normal form of Rayleigh-Bénard convection problem with the analytical form of the Rayleigh as a cofficient.

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. 6.

For a fluid in the Boussinesq approximation linearized about density ρ_0 and temperature T_0 , the governing equations consist of the conservation of momentum, mass, and energy:

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} = -\rho_0^{-1} \nabla p + \nu \nabla^2 \mathbf{u} - g\alpha (T - T_0)$$
(33a)

$$\nabla \cdot \mathbf{u} = 0 \tag{33b}$$

$$\frac{\partial T}{\partial t} + (\mathbf{u} \cdot \nabla)T = \kappa \nabla^2 T, \tag{33c}$$

along with the boundary conditions that $T(z=0)=T_0$, $T(z=L_z)=T_0+\Delta T$. Besides the primitive variables, the system includes gravity g, coefficient of thermal expansion α , kinematic viscosity ν , and thermal diffusivity κ . The flow is typically analyzed in terms of the Rayleigh number $\mathrm{Ra}=g\alpha\Delta TL_z^3/\nu\kappa$ and Prandtl number $\mathrm{Pr}=\nu/\kappa$. Although the Prandtl number has a significant impact on the behavior of the flow above the threshold of instability [46], the onset of instability itself is not sensitive to the Prandtl number [47].

We simulate the nondimensional flow in two dimensions with no-slip boundary conditions on the wall and periodic boundary conditions in the wall-parallel (x) direction using the Shenfun spectral Galerkin library [48]. We use a Fourier-Chebyshev discretization with (N_x, N_z) = (256, 100) and choose $L_x = 2\pi L_z$. The critical point in this simulation is at $\mathrm{Ra}_c \approx 1875$, significantly above the true value of $\mathrm{Ra}_c \approx 1708$. This is likely due to differences in boundary conditions, wall-parallel domain extent, and the restriction to two dimensions.

We simulate the system at a range of Rayleigh numbers (shown in Fig. 6) and Pr = 0.7. For each simulation, we dimensionalize the data as follows. We randomly select a value for κ in the range $(0.5, 0.7) \text{m}^2/\text{s}$, ΔT in the range (50, 80) °C, and α in the range $(1 \times 10^{-4}, 5 \times 10^4) \text{m}/\text{°C}$.

The separation L_z is then computed for consistency with the simulation Rayleigh number, with $g=9.8 \mathrm{m/s^2}$. Likewise, ν is computed for consistency with the Prandtl number. Finally, time is dimensionalized based on a free-fall time scale $\sqrt{L_z/g\alpha\Delta T}$. From eight initial simulation cases, we retain five points close to the bifurcation (black dots in Fig. 6) to solve the dimensionless SINDy problem and withold the other three points (red crosses) as test data.

One option for learning a model of the flow near the critical point would be to input the spatiotemporally varying data and learn a PDE model [20]. This might yield something similar to a Ginzburg-Landau or Swift-Hohenberg model [13]. Another approach is to approximate some quantity of interest with a linear modal decomposition such as proper orthogonal decomposition and fit an ODE model of the temporal coefficients [49]. Given the simple globally coherent spatial structure of the flow, we proceed with the latter.

To account for translational invariance, we first represent the temperature field with a Fourier series in the *x*-direction:

$$T(x,z,t) = \sum_{k} \hat{T}_k(z,t)e^{ikx}.$$
(34)

Recognizing that the weakly supercritical solution has only $k=\pm 2$ Fourier components (see Fig. 6), we define a real-valued observable as the wall-normal integral of the negative temperature gradient (proportional to heat flux) at k=2:

$$q(t) = \left| \int_0^{L_z} -ik\hat{T}_2(z,t) dz \right|. \tag{35}$$

Note that this choice is arbitrary; we could for instance model the coefficient of a particular Fourier-Chebyshev basis function, or that of a leading POD mode. All would illustrate qualitatively similar behavior, but this definition of q(t) is a simple and representative global observable.

We set up the dimensionless SINDy problem as described in Secs. 3.3. Since we expect a Landau-type dynamics model for the symmetry-breaking behavior, we choose a linear-cubic polynomial library in q and search for one dimensionless group π_p , along with an appropriate dimensionless time τ . Since there are only six independent dimensional quantities in this problem $(H,g,\alpha,\Delta T,\nu,\kappa)$, it is inexpensive to perform the optimization over all dimensionless numbers composed of integer powers up to ± 2 . This has the advantage of producing a simple number suitable for comparison to the classical result, but is not strictly necessary; the brute force search could be avoided with a constrained optimization approach as described in Sec. 3.1 and the boundary layer example in Sec. 4.3.

The optimal dimensionless number selected by the algorithm is the inverse of the Rayleigh number, $\pi_p = \mathrm{Ra}^{-1}$, with dimensionless time $\tau = t\kappa^2/\alpha^2\nu(\Delta T)^2$. The dynamics model is a normal form for a pitchfork bifurcation:

$$Ra\frac{dq}{d\tau} = \left(\frac{Ra}{Ra_c} - 1\right)q - \mu q^3,\tag{36}$$

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 found easily and compared to the steady states of the simulation, as shown in Fig. 6. 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). Of course, this is a simple model of a well-understood instability, but 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.

5 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 hypothesis 1. However, the definition of that optimality is non-unique, leaving the open question of how to find the best way to integrate the data with the Buckingham Pi theorem. In particular, the methods introduced in Sec. 3 all seek to identify dimensionless groups for which the corresponding function approximation (KRR, BuckiNet, or SINDy) can best represent the input-output relationship $\pi_q = \psi(\pi_p)$. Choosing a SINDy representation of this relationship encodes an inductive bias towards low-order polynomials, which is often appropriate in dynamical systems applications or asymptotic expansions. In these contexts, the identified parameters are more likely to have a familiar structure, such as the ratio of different terms in a governing equation. 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 derived by hand. For instance, the Blasius solution cannot be easily represented with low-order polynomials, but at the cost of interpretability of the function, kernel regression is able to accurately model the boundary layer with the nonstandard nondimensionalization in equation (32).

More generally, dimensional analysis is a commonly used method for physics discovery and characterization. Importantly, it highlights the fact that physical laws do not depend on any specific units of measurements. Rather, physical laws are fundamental and have the property of *generalized homogeneity* [1], thus they are independent of the observer. Advancements across scientific disciplines have used such concepts to develop governing equations, from gravitation to electromagnetism. Specifically, dimensionality reduction has identified critical symmetries, parametrizations, and bifurcation structures underlying a given model. As such, dimensionality reduction has always been a critical component of establishing the canonical models across disciplines. However, when encountering new physics-based systems where the governing equations are unknown or their parametrizations undetermined, dimensionality reduction once again is critical for determining, the general form to which any physical equation is reducible [3].

Edgar Buckingham provided the rigorous mathematical foundations on which dimensional analysis is accomplished. It is a constructive procedure which significantly constrains the space of hypothesized models. Although greatly constraining the allowable model space, Buckingham Pi theory does not produce a unique model, but rather a small subset of parametrizations which are then typically chosen from expert knowledge or asymptotic considerations [1]. With modern machine learning, we have shown that the Buckingham Pi theory procedure can be automated to discovery a diversity of important physical features and properties. 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 units matrix, or optimizes the robustness, accuracy and speed of the model selected. Specifically, 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 sparse identification of nonlinear dynamics (SINDy) based technique to discover dimensionless equations whose coefficients determine important features of the dynamics, such as inertia and bifurcations. Such regularizations solve the ill-posed nature of Buckingham Pi theory and its non-unique solution space. 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.

The suite of algorithms introduced provide a set of physics-informed learning tools that can help characterize new systems and the underlying general form to which the physical system is reducible, regardless of whether governing equations are known or not. 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. Explainable AI, especially in the context of physics-based systems, has grown in importance as it is imperative to understand the feature space on which modern AI empowered autonomy, for instance, makes decisions with guaranteed safety margins.

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.

References

- [1] Grigory Isaakovich Barenblatt. Scaling, self-similarity, and intermediate asymptotics: dimensional analysis and intermediate asymptotics. Number 14. Cambridge University Press, 1996.
- [2] Susan G Sterrett. Physically similar systems-a history of the concept. In *Springer handbook of model-based science*, pages 377–411. Springer, 2017.
- [3] Edgar Buckingham. On physically similar systems; illustrations of the use of dimensional equations. *Physical review*, 4(4):345, 1914.
- [4] Zachary del Rosario, Minyong Lee, and Gianluca Iaccarino. Lurking variable detection via dimensional analysis. *SIAM/ASA Journal on Uncertainty Quantification*, 7(1):232–259, 2019.
- [5] Lluís Jofre, Zachary R del Rosario, and Gianluca Iaccarino. Data-driven dimensional analysis of heat transfer in irradiated particle-laden turbulent flow. *International Journal of Multiphase Flow*, 125:103198, 2020.
- [6] Kai Fukami and Kunihiko Taira. Robust machine learning of turbulence through generalized buckingham pi-inspired pre-processing of training data. In APS Division of Fluid Dynamics Meeting Abstracts, pages A31–004, 2021.
- [7] Xiaoyu Xie, Wing Kam Liu, and Zhengtao Gan. Data-driven discovery of dimensionless numbers and scaling laws from experimental measurements. *arXiv* preprint arXiv:2111.03583, 2021.
- [8] E. Cerda and L. Mahadevan. Geometry and physics of wrinkling. Phys. Rev. Letters, 90(7):074302, 2003.
- [9] S. W. Morris, E. Bodenschatz, D. S. Cannell, and G. Ahlers. Spiral defect chaos in large aspect ratio Rayleigh-Bènard convection. *Physical Review Letters*, 71(13):2026, 1993.
- [10] X. D. Shi, M. P. Brenner, and S. R. Nagel. A cascade of structure in a drop falling from a faucet. *Science*, 265(5169):219–222, 1994.
- [11] B. Grzybowski, H. A. Stone, and G. M. Whitesides. Dynamic self-assembly of magnetized, millimetre-sized objects rotating at a liquid-air interface. *Nature*, 405:1033–1036, 2000.
- [12] A. Seminara, T. E. Angelini, J. N. Wilking, H. Vlamakis, S. Ebrahim, R. Kolter, D. A. Weitz, and M. P. Brenner. Osmotic spreading of bacillus subtilis biofilms driven by an extracellular matrix. *Proceedings of the National Academy of Sciences*, 109(4):1116–1121, 2012.
- [13] M. C. Cross and P. C. Hohenberg. Pattern formation outside of equilibrium. *Reviews of modern physics*, 65(3):851, 1993.

- [14] Jared L Callaham, James V Koch, Bingni W Brunton, J Nathan Kutz, and Steven L Brunton. Learning dominant physical processes with data-driven balance models. *Nature communications*, 12(1):1–10, 2021.
- [15] Philip Holmes and John Guckenheimer. *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields,* volume 42 of *Applied Mathematical Sciences*. Springer-Verlag, Berlin, Heidelberg, 1983.
- [16] Or Yair, Ronen Talmon, Ronald R Coifman, and Ioannis G Kevrekidis. Reconstruction of normal forms by learning informed observation geometries from data. *Proceedings of the National Academy of Sciences*, page 201620045, 2017.
- [17] Manu Kalia, Steven L Brunton, Hil GE Meijer, Christoph Brune, and J Nathan Kutz. Learning normal form autoencoders for data-driven discovery of universal, parameter-dependent governing equations. arXiv preprint arXiv:2106.05102, 2021.
- [18] Michael Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *science*, 324(5923):81–85, 2009.
- [19] Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. 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] Samuel H Rudy, Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. Data-driven discovery of partial differential equations. *Science Advances*, 3(4):e1602614, 2017.
- [21] Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning non-linear operators via deeponet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021.
- [22] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. 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] George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, 2021.
- [24] Frank Noé, Simon Olsson, Jonas Köhler, and Hao Wu. Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning. *Science*, 365(6457):eaaw1147, 2019.
- [25] MP Brenner, JD Eldredge, and JB Freund. Perspective on machine learning for advancing fluid mechanics. *Physical Review Fluids*, 4(10):100501, 2019.
- [26] Karthik Duraisamy, Gianluca Iaccarino, and Heng Xiao. Turbulence modeling in the age of data. *Annual Reviews of Fluid Mechanics*, 51:357–377, 2019.
- [27] Steven L. Brunton, Bernd R. Noack, and Petros Koumoutsakos. Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics*, 52:477–508, 2020.
- [28] M. Sonnewald, R. Lguensat, D. C. Jones, P. D. Dueben, J. Brajard, and V. Balaji. Bridging observations, theory, and numerical simulation of the ocean using machine learning. *Environmental Research Letters*, 16(7):073008, 2021.
- [29] B. E. Kaiser, J. A. Saenz, M. Sonnewald, and D. Livescu. Objective discovery of dominant dynamical processes with intelligible machine learning. *arXiv*:2106.12963, 2021.
- [30] M. Sonnewald, C. Wunsch, and P. Heimbach. Unsupervised learning reveals geography of global ocean regimes. *Earth and Space Science*, 6:784–794, 2019.
- [31] Hao Wu, Andreas Mardt, Luca Pasquali, and Frank Noe. Deep generative markov state models. 32nd Conference on Neural Information Processing Systems (NeurIPS), 2018.
- [32] K. Champion, B. 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.
- [33] Joseph Bakarji, Kathleen Champion, J Nathan Kutz, and Steven L Brunton. Discovering governing equations from partial measurements with deep delay autoencoders. *arXiv preprint arXiv:2201.05136*, 2022.
- [34] S. L. Brunton and J. N. Kutz. Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control. Cambridge University Press, 2019.

- [35] Paul G Constantine, Zachary del Rosario, and Gianluca Iaccarino. Data-driven dimensional analysis: algorithms for unique and relevant dimensionless groups. *arXiv preprint arXiv:1708.04303*, 2017.
- [36] Silviu-Marian Udrescu and Max Tegmark. Ai feynman: A physics-inspired method for symbolic regression. *Science Advances*, 6(16):eaay2631, 2020.
- [37] David J Gunaratnam, Taivas Degroff, and John S Gero. Improving neural network models of physical systems through dimensional analysis. *Applied Soft Computing*, 2(4):283–296, 2003.
- [38] Steven H Strogatz. Nonlinear dynamics and chaos: With applications to physics, biology, chemistry, and engineering. CRC press, 2018.
- [39] S. B. Pope. Turbulent Flows. Cambridge University Press, 2000.
- [40] H. Schlichting. Boundary-Layer Theory. McGraw-Hill, 1955.
- [41] J. Guckenheimer and P. Holmes. *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields.* Springer, 1983.
- [42] P. Meliga, J.-M. Chomaz, and D. Sipp. Global mode interaction and pattern selection in the wake of a disk: a weakly nonlinear expansion. *Journal of Fluid Mechanics*, 633:159–189, 2009.
- [43] M. Carini, F. Auteri, and F. Giannetti. Centre-manifold reduction of bifurcating flows. *J. Fluid Mech.*, 767:109–145, feb 2015.
- [44] M. Golubitsky and W. Langford. Pattern formation and bistability in flow between counterrotating cylinders. *Physica D*, 32:362–392, 1988.
- [45] E. N. Lorenz. Deterministic nonperiodic flow. Journal of the Atmospheric Sciences, 20(2):130–141, Mar 1963.
- [46] A. Pandey, J. D. Scheel, and J. Schumacher. Turbulent superstructures in Rayleigh-Benard convection. *Nature Communications*, 9(2118), 2018.
- [47] S. Chandrasekhar. Hydrodynamic and Hydromagnetic Stability. Clarendon Press, Oxford, 1961.
- [48] Mikael Mortensen. Shenfun: High performance spectral galerkin computing platform. *Journal of Open Source Software*, 3(31):1071, 2018.
- [49] J.-C. Loiseau and S. L. Brunton. Constrained sparse Galerkin regression. Journal of Fluid Mechanics, 838:42–67, 2018.