Data-driven discovery of dimensionless numbers and scaling laws from experimental measurements

Xiaoyu Xie
Northwestern University
Wing Kam Liu (w-liu@northwestern.edu)
Northwestern University https://orcid.org/0000-0001-7725-8438
Zhengtao Gan
Northwestern University https://orcid.org/0000-0002-4309-0929

Article

Keywords: Dimensional analysis, Buckingham's Pi theorem, Physics-informed machine learning, Fluid mechanics, Sparse identification, Additive manufacturing

Posted Date: December 10th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1122326/v1

DATA-DRIVEN DISCOVERY OF DIMENSIONLESS NUMBERS AND SCALING LAWS FROM EXPERIMENTAL MEASUREMENTS

Xiaoyu Xie
Department of Mechanical Engineering
Northwestern University
Evanston, IL 60208
xiaoyuxie2020@u.northwestern.edu

Wing Kam Liu*
Department of Mechanical Engineering
Northwestern University
Evanston, IL 60208
w-liu@northwestern.edu

Zhengtao Gan*
Department of Mechanical Engineering
Northwestern University
Evanston, IL 60208
zhengtao.gan@northwestern.edu

November 29, 2021

ABSTRACT

Dimensionless numbers and scaling laws provide elegant insights into the characteristic properties of physical systems. Classical dimensional analysis and similitude theory fail to identify a set of unique dimensionless numbers for a highly-multivariable system with incomplete governing equations. In this study, we embed the principle of dimensional invariance into a two-level machine learning scheme to automatically discover dominant and unique dimensionless numbers and scaling laws from data. The proposed methodology, called dimensionless learning, can be treated as a physics-based dimension reduction. It can reduce high-dimensional parameter spaces into descriptions involving just a few physically-interpretable dimensionless parameters, which significantly simplifies the process design and optimization of the system. We demonstrate the algorithm by solving several challenging engineering problems with noisy experimental measurements (not synthetic data) collected from the literature. The examples include turbulent Rayleigh-Bénard convection, vapor depression dynamics in laser melting of metals, and porosity formation in 3D printing. We also show that the proposed approach can identify dimensionally homogeneous differential equations with minimal parameters by leveraging sparsity-promoting techniques.

Keywords Dimensional analysis · Buckingham's Pi theorem · Physics-informed machine learning · Fluid mechanics · Sparse identification · Additive manufacturing

1 Introduction

All physical laws can be expressed as dimensionless relationships with fewer dimensionless numbers and a more compact form [1]. A dimensionless number is a power-law monomial of some physical quantities [2]. There is no physical dimension (such as mass, length, or energy) assigned to a dimensionless number, which provides a property of scale invariance, i.e., the dimensionless numbers are invariant when the length scale, time scale, or energy scale of the system varies. More than 1,200 dimensionless numbers have been discovered in an extremely wide range of fields, including physics and physical chemistry, fluid and solid mechanics, thermodynamics, electromagnetism, geophysics and ecology, and engineering [3]. There are several significant advantages to describing a physical process or system using dimensionless numbers. Using dimensionless numbers can considerably simplify a problem by reducing the

*Corresponding authors: w-liu@northwestern.edu (W. Liu), zhengtao.gan@northwestern.edu (Z. Gan)

number of variables that describe the physical process, thereby reducing the number of experiments (or simulations) required to understand and design the physical system. For example, the Reynolds number (Re) is a well-known dimensionless number in fluid mechanics named after Osborne Reynolds, who investigated the fluid flow through pipes in 1883 [4]. The Reynolds number is defined as a power law of four physical quantities, i.e., the fluid density, the average fluid velocity, the diameter of the pipe, and the dynamic fluid viscosity. The flow characteristics (laminar or turbulent) in a pipe is best determined by the Re, rather than the four individual dimensional quantities. Moreover, the scale-invariance property of the dimensionless numbers plays a critical role in similitude theory [5]. Many small-scale experiments have been designed to understand and predict the behaviors of the full-scale applications in aerospace [6], nuclear [7], and marine engineering [8], where the full-scale applications are typically extremely expensive and even dangerous. All the dimensionless numbers should be identical between the small-scale and full-scale experiments, which yields perfect geometry, dynamic, and kinematic similarities between the two scales. Furthermore, dimensionless numbers are ratios of two forces, energies, or mechanisms. Thus, they are physically interpretable and can provide fundamental insights into the behavior of complex systems. For example, the Péclet number (Pe) represents the ratio of the convection rate of a physical quantity by the flow to the gradient-driven diffusion rate, which enables order-of-magnitudes analysis for the transport phenomena of a process.

Despite the scientific significance and widespread use of dimensionless numbers, it is still challenging to discover new dimensionless numbers and their relationships (i.e., scaling laws) from experiments, especially for a complex physical system without complete governing equations. A traditional solution is dimensional analysis [2] based on Buckingham's Pi theorem [9], which provides a systematic approach to examining the units of a physical system and forming a set of dimensionless numbers that satisfy the principle of dimensional invariance [10]. However, dimensional analysis has several well-known limitations. First, the derived dimensionless numbers are not unique. The Buckingham's Pi theorem [9], from the viewpoint of mathematics, provides a linear subspace of exponents that produces dimensionless numbers. Any basis for the subspace is equally valid. Thus, it fails to identify the dimensionless numbers, given a particular choice of basis, that are dominant for the physical system. Second, the mathematical relation between dimensionless numbers (i.e., the scaling law) remains unknown by using dimensional analysis alone. A common approach to establishing the scaling law is to leverage the results of the dimensional analysis with experimental measurements of the physical system. The experimental measurements are transformed into the dimensionless numbers obtained by dimensional analysis and fitted onto a high-dimensional response surface to represent the scale-invariant relationship. However, since dimensional analysis does not provide unique dimensionless numbers, this procedure is very time-consuming and relies strongly on the experience of domain experts to select a set of appropriate dimensionless numbers through a long process of trial-and-error.

Theses limitations could be overcame by integrating dimensional analysis with advanced data science and artificial intelligence (AI). Mendez and Ordonez proposed an algorithm called SLAW (i.e., Scaling LAWs) to identify the form of a power law from experimental data (or simulation data) [11]. The proposed SLAW combines dimensional analysis with multivariate linear regressions. This approach has been applied to some engineering areas, such as ceramic-to-metal joining [11] and plasma confinement in Tokamaks [12]. However, for the sake of simplification, the proposed algorithm assumes that the relationship between the dimensionless numbers obeys a power law, which is invalid in many applications. Constantine, Rosario, and Iaccarino proposed a rigorous mathematical framework to estimate unique and relevant dimensionless groups [13, 14]. Active subspace methods are connected to dimensional analysis, which reveals that all physical laws are ridge functions [14]. Their method is applicable to idealized physical systems meaning that the experiments can be conducted for arbitrary values of the independent input variables (or dependent input variables with a known probability density function), and noises or errors in the input and output are negligible.

In this study, we propose a mechanistic data-driven approach, called dimensionless learning, in which the principle of dimensional invariance (i.e., physical laws cannot depend on an arbitrary choice of basic units of measurements [1]) is embedded into a two-level machine learning scheme to automatically discover dominant dimensionless numbers and scaling laws from noisy experimental measurements of complex physical systems. We add the physical constraint of dimensional invariance into the learning algorithm, which steers the learning toward scale-invariant and physically-interpretable low-dimensional patterns of complex high-dimensional systems. We demonstrate the proposed approach by solving multiple challenging problems in science and engineering. The examples include turbulent Rayleigh-Benard convection, vapor depression dynamics, and porosity formation in 3D printing. The experimental datasets are collected from the literature. Furthermore, we demonstrate the potential of the proposed approach by combining it with sparsity-promoting techniques (such as SINDy [15]) to identify dimensionally homogeneous differential equations from data, which involves natural coordinates of spatial and temporal variables.

2 Results

2.1 Turbulent Rayleigh-Bénard convection

In this section, we demonstrate the proposed dimensionless learning using the example of a classical fluid mechanics problem: turbulent Rayleigh-Bénard convection. The goal is to rediscover the Rayleigh number (Ra) directly from experimental measurements. The Ra is named after Lord Rayleigh, who investigated a non-isothermal buoyancy-driven flow in 1916 [16], which is now named Rayleigh-Bénard convection. Turbulent Rayleigh-Bénard convection is a paradigmatic system to study turbulent thermal flow, occurring in a planar horizontal layer of fluid in a container heated from below. The internal fluid could develop complex turbulent dynamics due to the effects of buoyancy, fluid viscosity, and gravity (Fig. 1a).

Experimentally, the heat flux through the container, q, can be measured, which depends on the height of the container h, the temperature difference between the top and bottom surfaces ΔT, the gravitational acceleration g, and the properties of the fluid, including the thermal conductivity λ, the thermal expansion coefficient α, the viscosity ν, and the thermal diffusivity κ. To obtain a causal relationship, we need to specify the dependent (i.e., output) and independent (i.e., input) variables from the physical quantities describing the system. To simplify the demonstration, we assume the form of the output variable as the Nusselt number $\mathrm{Nu} = \frac{qh}{\lambda \Delta T}$ (a more general case using q as the output will be presented later) and a list of physical quantities $\boldsymbol{p}$ as input variables. The causal relationship to be determined can be represented as

$$\mathrm{Nu} = \frac{qh}{\lambda \Delta T} = f(h, \Delta T, \lambda, g, \alpha, \nu, \kappa) = f(\boldsymbol{p}). \tag{1}$$

This is a high-dimensional parameter space. To explore it, we collect an experimental dataset of turbulent Rayleigh-Bénard convection from two different articles [18, 19], including 182 experiments with various input variables and corresponding output measurements (Fig. 1b). Many machine learning models can fit the data. However, most of those are black box models, such as neural networks, without good interpretability and physical insights. Alternatively, we aim to identify a low-dimensional scale-invariant scaling law that best represent the dataset. In the scaling law, the products of powers of the input variables $\boldsymbol{p}$ form a dimensionless number Π. Thus, the causal relationship can be rewritten as

$$\mathrm{Nu} = f_1(\Pi), \tag{2}$$

$$\Pi = h^{w_1} \Delta T^{w_2} \lambda^{w_3} g^{w_4} \alpha^{w_5} \nu^{w_6} \kappa^{w_7}, \tag{3}$$

where $\boldsymbol{w} = [w_1, ..., w_7]^T$ are the powers that produce the dimensionless number and are to be determined. In this example, we assume that there is only one input dimensionless number governing the process. An algorithm for determining the number of the dimensionless numbers required from the data is provided in Supplementary Information Section 1.3.

To embed the physical constraint of dimensional invariance, we conduct the dimensional analysis, i.e., the powers $\boldsymbol{w} = [w_1, ..., w_7]^T$ need to satisfy a linear system of equations

$$\boldsymbol{D}\boldsymbol{w} = 0, \tag{4}$$

where $\boldsymbol{D}$ is the dimension matrix of the input variables (Fig. 1c). Each column of the dimension matrix is the dimension vector of the corresponding variable. The dimension vector represents the exponents of the physical quantity with respect to the fundamental dimensions. It is worth noting that there are only seven fundamental dimensions in nature: mass [M], length [L], time [T], temperature [Θ], electric current [I], luminous intensity [J], and amount of substance [N] [20]. All the other dimensions are power-law monomials of the fundamental dimensions [1]. In this example, we use four fundamental dimensions, i.e., [M], [L], [T], and [Θ] (Fig. 1c). The dimension matrix includes the physical dimensions of the input variables. The linear system of equations $\boldsymbol{D}\boldsymbol{w} = 0$ guarantees that the power-law monomial of the input variables (Eqn. 3) is dimensionless [21]. Since the linear system is underdetermined (i.e., the number of unknown variables is more than the number of equations), there are infinitely many solutions, indicating that infinitely many forms of dimensionless numbers can be obtained from the dimensional analysis. Furthermore, we can represent the solutions of the linear system (Eqn.4) as linear combinations of three basis vectors $\boldsymbol{w_{b1}}$, $\boldsymbol{w_{b2}}$, and $\boldsymbol{w_{b3}}$

$$\boldsymbol{w} = \gamma_1 \boldsymbol{w_{b1}} + \gamma_2 \boldsymbol{w_{b2}} + \gamma_3 \boldsymbol{w_{b3}}, \tag{5}$$

where $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \gamma_2]^T$ are the coefficients with respect to the three basis vectors in this case. The number of basis vectors is equal to the number of input variables (seven in this case) minus the rank of the dimension matrix (four in this case). This formula aligns with the Buckingham π theorem [9]. Since the basis vectors can be computed based on Eqn. 4 (an algorithm for computing basis vectors is provided in Supplementary Information Section 1.2), the basis

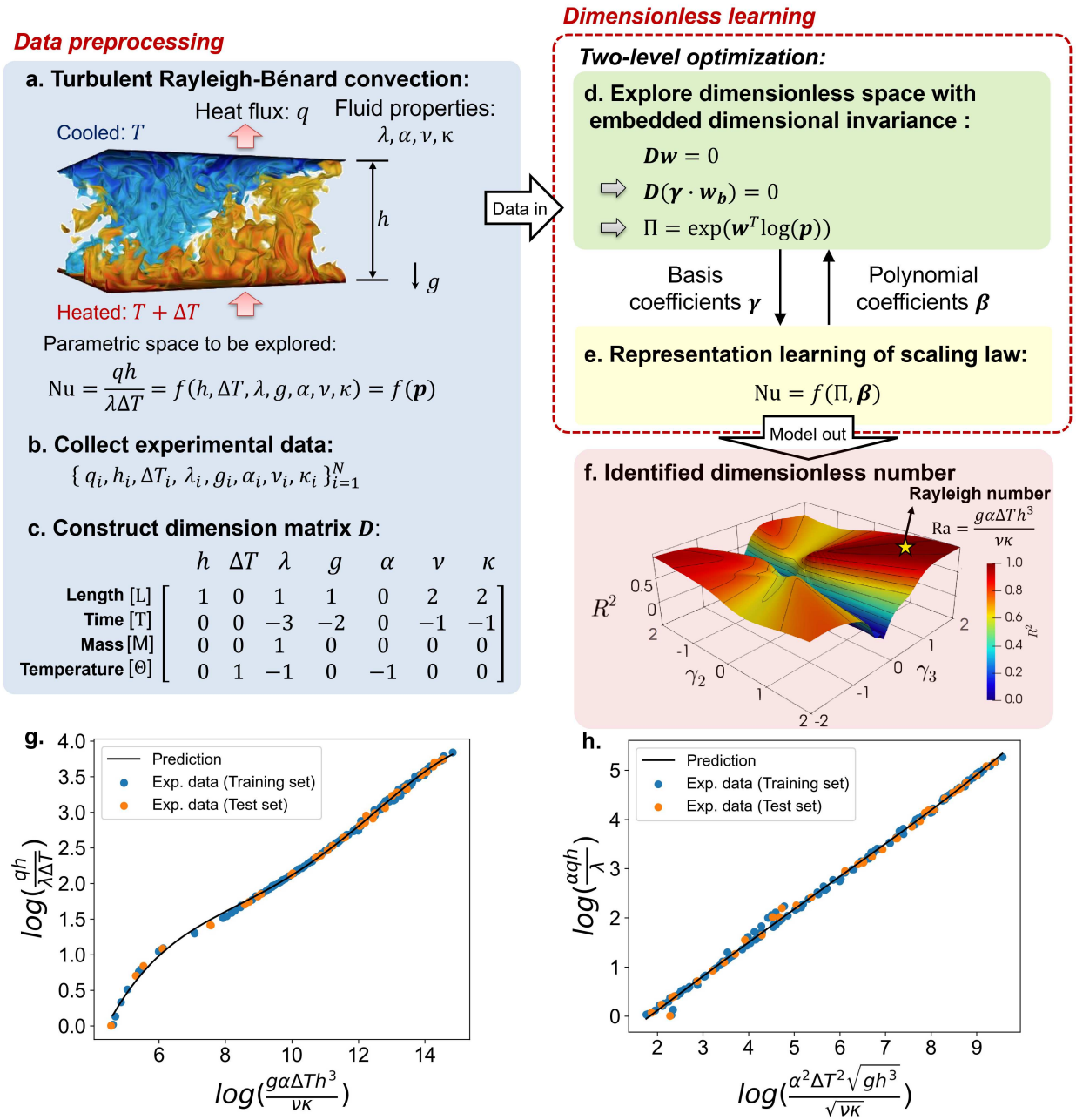

Figure 1: **The proposed dimensionless learning demonstrated on turbulent Rayleigh-Benard convection.** **(a)** A schematic of Rayleigh-Benard convection [17] with related physical quantities. **(b)** Collected experimental measurements. **(c)** Constructed dimension matrix $\boldsymbol{D}$ of the input variables. **(d)** First-level of the two-level optimization scheme for training the coefficients $\boldsymbol{\gamma}$ with respect to the computed basis vectors. **(e)** Second-level of the two-level optimization scheme for optimizing the unknown coefficients $\boldsymbol{\beta}$ in the representation learning. **(f)** Explored dimensionless space with a measure of R^2. The location with the maximum R^2 marked as a yellow star is related to the classical Rayleigh number. **(g)** Identified one-dimensional scaling law between Nu and Ra. **(h)** Discovered linear scaling law between two identified dimensionless numbers.

vectors' coefficients (or just "basis coefficients") are the unknowns to be determined. A set of computed basis vectors for this case are

$$\boldsymbol{w_{b1}} = [0, 0, 0, 0, 0, 1, -1]^T, \tag{6}$$

$$\boldsymbol{w_{b2}} = [0, 1, 0, 0, 1, 0, 0]^T, \tag{7}$$

$$\boldsymbol{w_{b3}} = [3, 0, 0, 1, 0, -2, 0]^T. \tag{8}$$

Once the basis coefficients γ_1, γ_2, and γ_3 are obtained, the form of the dimensionless number Π can be determined by Eqns. 3 and 5 (Fig. 1d).

To determine the values of the basis coefficients using the collected dataset, a model representing the scaling relation between the input and output dimensionless numbers is required, which introduces another set of unknown parameters $\boldsymbol{\beta}$ (i.e., the representation learning shown in Fig. 1e). We use a 5-ordered polynomial model in this case (more advanced models, such as tree-based models and deep neural networks, are optional depending on the complexity of the problem to be solved; see Section 2.3 and Supplementary Information Section Section 3 for more demonstrations). The polynomial model can be expressed as

$$\mathrm{Nu} = \beta_0 + \beta_1\Pi + \beta_2\Pi^2 + ... + \beta_5\Pi^5, \tag{9}$$

where $\boldsymbol{\beta} = [\beta_0, \beta_1, ..., \beta_5]^T$ are polynomial coefficients that represent the scaling relation.

We design an iterative two-level optimization scheme to determine the two sets of unknown parameters in the regression problem, i.e., the basis coefficients $\boldsymbol{\gamma}$ and polynomial coefficients $\boldsymbol{\beta}$. The optimization scheme includes multiple interactive steps. At each step, we adjust the first-level basis coefficients $\boldsymbol{\gamma}$ while fixing the second-level polynomial coefficients $\boldsymbol{\beta}$ and then optimize the second-level polynomial coefficients $\boldsymbol{\beta}$ while fixing the first-level basis coefficients $\boldsymbol{\gamma}$. This process is repeated until the result is converged, i.e., the values of $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are unchanged. There are several advantages of the proposed two-level approach over a single-level approach combining the two sets of unknowns together during the optimization. We can use different optimization methods and parameters (such as the learning rate) for these two-level models to significantly improve the efficiency of the optimization. More importantly, we can utilize the physical insights to inform the learning process. The first-level basis coefficients $\boldsymbol{\gamma}$ have a clear physical meaning, which is related to the powers that produce the dimensionless number. Thus, those values have to be rational numbers to maintain the dimensional invariance. Moreover, their range is typically not large. Note that absolute values of coefficients in most of the dimensionless numbers and scaling laws are less than four [1]. To leverage those physical insights or constraints, we design several methods for optimizing the first-level basis coefficients, including a simple grid search (used in this section) and a pattern search that is much more efficient (Supplementary Information Section 3.2). For the second-level coefficients, we conduct multiple standard representation learning methods, including the polynomial regression used in this section, tree-based extreme gradient boosting (XGBoost [22]) used in Section 2.3, and general gradient descent method (Supplementary Information Section 3.1). Details of the two-level optimization framework are provided in Supplementary Information Section 3.

We illustrate the first-level grid search for γ_2 and γ_3 ranging from -2 to 2 with a interval of 1/100 (Fig. 1f). We fix γ_1 as one to avoid too large or too small powers of the dimensionless number Π. For each $\boldsymbol{\gamma}$ in the dimensionless space, the polynomial coefficients $\boldsymbol{\beta}$ are trained based on the collected data. The dataset is separated into 80% training set and 20% test set. The coefficient of determination (R^2) of the test set is shown in Fig. 1f as a measure of learning performance. We can identify a unique point with the maximum R^2 (0.999) from Fig. 1f (marked as a yellow star), where γ_1=γ_2=γ_3 = 1. Using these optimized basis coefficients, the expression of the dominant dimensionless number can be identified as

$$\Pi = \frac{g\alpha\Delta T h^3}{\nu\kappa}. \tag{10}$$

This form is identical to the classical Rayleigh number, indicating that the proposed dimensionless learning can rediscover the well-known dimensionless number directly from data. Moreover, we demonstrate that for the given parameter list, the Rayleigh number is the unique dimensionless number to best fit the dataset because there is only one global maximum of R^2 within the dimensionless space (Fig. 1f). The log-log scaling relation between Ra and Nu is a simple one-dimensional pattern where all the datapoints collapse onto a single curve (Fig. 1g).

The proposed dimensionless learning can deal with dimensional output variables as well. A combination of input variables with the same dimension as the output variable can be searched to non-dimensionalize the output variable (the detailed algorithm for output non-dimensionalization is provided in Supplementary Information Section 1.3). Using the heat flux q as the output variable (instead of Nu, which was used in the previous example), the dimensionless space is expanded, and thus more dominant dimensionless numbers and scaling laws can be found. We discover a new set of dimensionless numbers to best represent (R^2 = 0.999) the collected experimental measurements. More interestingly, the identified log-log scaling relation between the dimensionless numbers is almost linear (Fig. 1h). This finding could lead to new physical insights into the complex turbulent Rayleigh-Bénard dynamics.

2.2 Vapor depression dynamics in laser-metal interaction

Another challenging problem in the application of dimensionless learning is laser-metal interaction dynamics. People have been curious about the physical responses of a metallic material due to a high-power laser irradiation since 1964, when Patel invited an electric discharge CO_2 laser [23] that was dramatically scaled up in power shortly after. During the laser-metal interaction, a vapor-filled depression (termed a keyhole) frequently forms on a puddle of liquid metal melted by the laser. The keyhole is caused by vaporization-induced recoil pressure, and its dynamics is inherently difficult to understand because of its complex dependence upon many physical mechanisms but important to be able to quantify because it is closely related to energy absorption and defect formation in many industrial and military applications, such as laser-based materials processing and manufacturing [24], high-energy laser weapons [25], and aerospace laser-propulsion engines [26].

Recently, high-quality in situ experimental data on keyhole dynamics became available via high-speed X-ray imaging [27]. Using X-ray pulses, images of the keyhole region inside the metals can be recorded with micrometer spatial resolution [28]. The keyhole depth e can be measured from the X-ray images (Fig. 2a), and it depends on different materials and several process parameters, such as the effective laser power ηP, the laser scan speed V_s, and the laser beam radius r_0. We collect a dataset of keyhole X-ray images from the literature, including ninety-one experiments with various process parameters and three different materials: a titanium alloy (Ti6Al4V), an aluminium alloy (Al6061), and a stainless steel (SS316) [29, 24]. We represent a material using a set of material properties: the thermal diffusivity α, the material density ρ, the heat capacity C_p, and the difference between melting temperature and ambient temperature $T_l - T_0$. Therefore, the casual relationship can be expressed as

$$e = f(\eta P, V_s, r_0, \alpha, \rho, C_p, T_l - T_0). \tag{11}$$

We can use the dimensionless learning described in the previous section to extract a low-dimensional scale-free relation from the parameter list. The dimension matrix $\boldsymbol{D}$ and computed basis vectors $\boldsymbol{w_{b1}}$, $\boldsymbol{w_{b2}}$, and $\boldsymbol{w_{b3}}$ of this example are provided in Supplementary Information Section 2. We first demonstrate the grid search ranging from -2 to 2 with a interval of 1/100 for the first-level optimization and 5-ordered polynomial regression for the second-level optimization. We fix γ_1 as 0.5 and normalize the output variable as the keyhole aspect ratio $e^* = \frac{e}{r_0}$, which is a widely used dimensionless parameter to represent the keyhole characteristic [30]. By searching the dimensionless parameter space, we can find one local optimum in term of the R^2 criteria marked as a blue star ($R^2 = 0.64$) in Fig. 2d. The expression of the dimensionless number $\Pi = \frac{\rho C_p (T_l - T_0) V_s^{1.5} r_0^{2.5}}{\sqrt{\alpha} \eta P}$ is computed based on the basis coefficients $\gamma_2 = \gamma_3 = -1$. However, the datapoints are scattered as shown in Fig. 2c, indicating that the dimensionless number located at the local maximum of the dimensionless space is not a good scaling parameter for this problem. The global optimum of the dimensionless space, where $\gamma_2 = \gamma_3 = 1$, provides a much better scaling behavior, with a 0.98 R^2 score (Fig. 2b). The dominant dimensionless number that emerged in the keyhole dynamics is

$$\Pi = \frac{\eta P}{(T_l - T_0)\pi \rho C_p \sqrt{\alpha V_s r_0^3}}. \tag{12}$$

This dimensionless number is identified directly from data, and it has the same form as the newly discovered keyhole number Ke [29] (sometimes called normalized enthalpy [31]), which can be derived from heat transfer theory. Even if we use dimensional variable e as the output, the dimensionless learning algorithm still confirms that the form of the keyhole number (i.e., Eqn. 12) is unique and dominant for controlling the value of the keyhole aspect ratio. Details of the procedure and results are provided in Supplementary Information Section 4.1. Using the identified dimensionless number, a simple scaling law emerges to control the keyhole aspect ratio, which simplifies the original high-dimensional problem into a univariate scaling law as

$$e^* = 0.12\mathrm{Ke} - 0.30. \tag{13}$$

Providing a sufficient parameter list is critical for dimensionless learning. If one or more important quantities are omitted, it is impossible to achieve a high R^2 for the learning and identify the right form of the dimensionless number(s). We demonstrate in Supplementary Information Section 5.1 that if we assume a parameter list omitting the thermal diffusivity α, the maximum R^2 is below 0.80 over the dimensionless space, which is much less than the value for the sufficient parameter list (i.e., Eqn. 11). Another scenario that often happens in the applications is that we consider more quantities than are sufficient, including some irrelevant or unimportant quantities. We demonstrate this scenario in Supplementary Information Section 5.2 by considering one more quantity, e.g., the latent heat of melting L_m or the difference between boiling temperature and ambient temperature $T_l - T_0$, in the parameter list. The form of the keyhole number can still be identified in the scenario. Moreover, there are a few more dimensionless numbers, which consist of

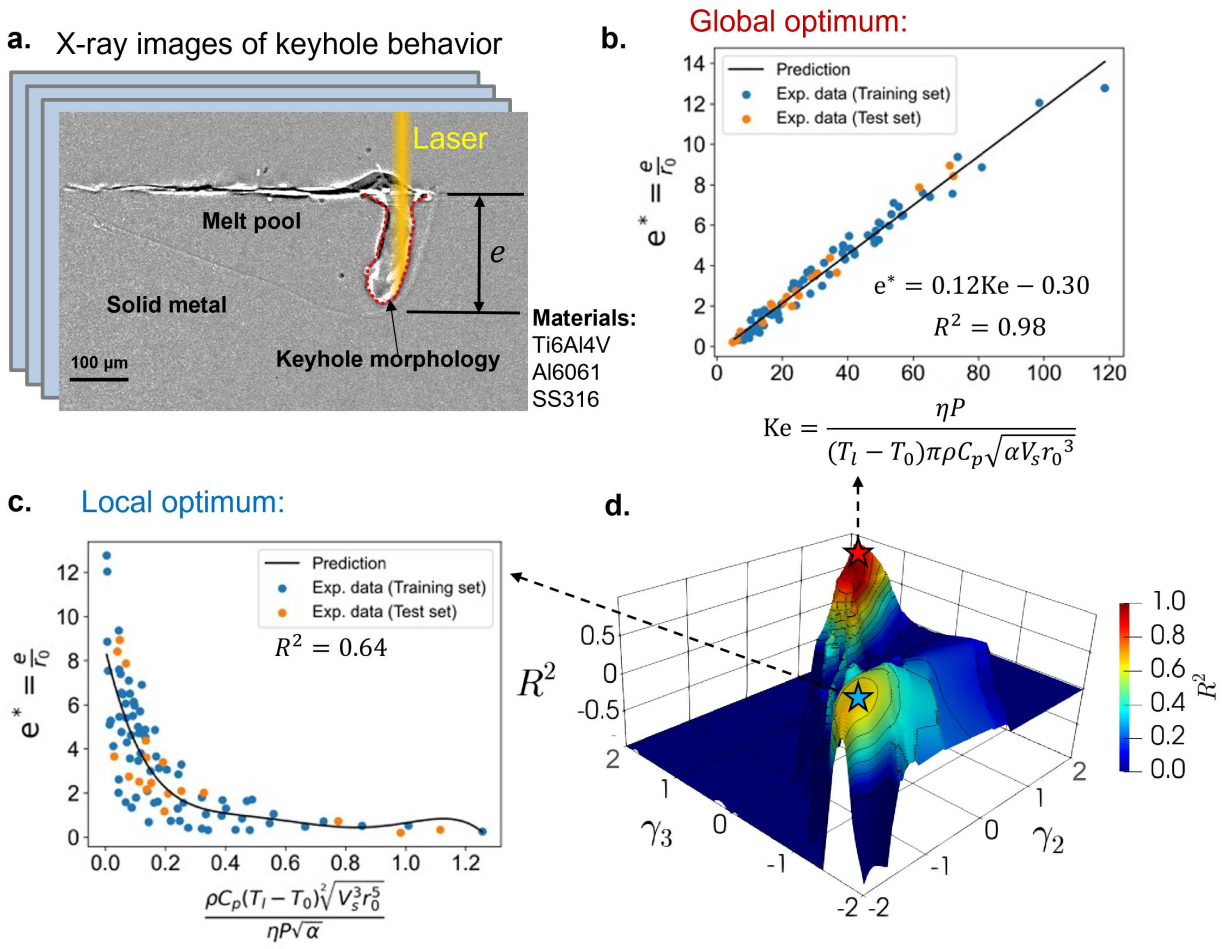

Figure 2: **Discover dimensionless numbers governing keyhole dynamics in laser-metal interaction.** **(a)** An illustrative X-ray image of keyhole morphology [29]. The dataset including X-ray imaging experiments on three different materials. **(b)** Global optimum of the dimensionless space, which represents the scaling law between the keyhole aspect ratio and the identified keyhole number using dimensionless learning. **(c)** Local optimum of the dimensionless space. **(d)** Dimensionless space using γ_2 and γ_3 as coordinates. The values of R^2 indicate the learning performance for the corresponding dimensionless number in the dimensionless space. The values of R^2 less than -1 are shown as -1.

the added quantity, providing a high R^2 as the keyhole number. This implies that more experiments are needed to select the distinguished one from the identified dimensionless numbers.

We provide two efficient algorithms, i.e., gradient-based and pattern search-based two-level optimization schemes, in Supplementary Information Section 3 to improve the efficiency of the optimization used in this section. These algorithms are especially helpful for exploring a high-dimensional parameter space, including many parameters to describe the physical system and several dimensionless numbers to construct the low-dimensional pattern.

2.3 Porosity formation in 3D printing of metals.

Three-dimensional (3D) printing or additive manufacturing is a disruptive technology of making three-dimensional solid objects from a digital file, which provides a new paradigm to change manufacturing [32]. In metal 3D printing, metallic parts are built layer by layer via local melting and (re)solidification of metallic powders by a laser or electron beam. 3D printing enables remarkable freedom for designing local geometrical and compositional features. However, this process has large numbers of parameters to be considered when making a part, and tends to produce defects, such as internal porosity, during the process if inappropriate process parameters are used (Fig. 3a).

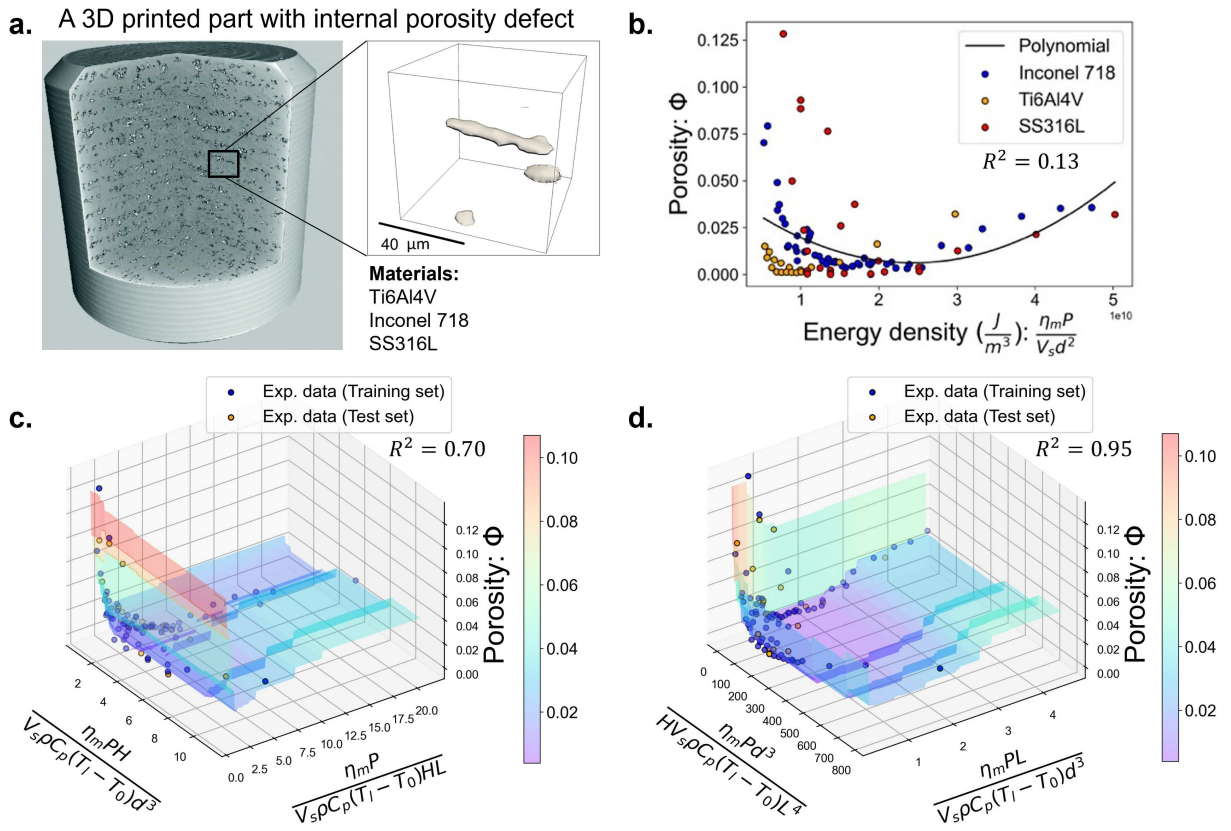

Figure 3: **Discover dimensionless numbers governing porosity formation in 3D printing.** **(a)** A schematic of a 3D printed metal part with internal porosity defect [39]. The dataset, including X-ray micro-computed tomography (micro-CT) measurements on three different materials. **(b)** Porosity measurements with various energy density values, a traditional combined parameter to correlate porosity with process parameters. **(c)** Identified 2D scaling relation combining both lack of fusion and keyhole porosity with two discovered dimensionless numbers. **(d)** Another identified 2D scaling relation with a higher R^2 score. The reduced parameter space can be easily visualized and interpreted, while the original high-dimensional problem is hard since the porosity is governed by nine parameters.

To extract elegant insights into the complex behavior of porosity formation in 3D printing, we collect an experimental dataset collected from six independent studies [33, 34, 35, 36, 37, 38], including ninety-three 3D printed parts with measured porosity volume fraction and various process parameters. Three kinds of materials were used: a titanium alloy (Ti6Al4V), a nickel-based alloy (Inconel 718), and a stainless steel (SS316L). The porosity volume fraction Φ depends on many process parameters and materials used in the experiments, which can be expressed as

$$\Phi = f(\eta_m P, V_s, d, \rho, C_p, \alpha, T_l - T_0, H, L), \tag{14}$$

where $\eta_m P$ is the effective laser power input, V_s is the laser scan speed, d is the laser beam diameter, ρ is the material density, C_p is the material heat capacity, α is the thermal diffusivity, $T_l - T_0$ is the difference between melting temperature and ambient temperature, H is the hatch spacing between the two adjacent laser scans, and L is the layer thickness of the metallic powders. It is a high-dimensional relation, and thus hard to understand and visualize. Traditionally, some combined parameters, such as energy density $\frac{\eta_m P}{V_s d^2}$, are used to simplify this relation. However, the R^2 score of a polynomial model using energy density as input is very low (0.13), as shown in Fig. 3b, indicating that a universal physical relation, which is valid for different materials and processing conditions, cannot be built by using the energy density alone because it is not a scale-free parameter. The form of the relation has to be modified when the energy scale is changed in the experiments with varying process parameters or materials.

We apply dimensionless learning to this challenging engineering problem and discover some dominant dimensionless numbers that provide a universal physical relation remained accurate for all the experimental conditions. The dimension

matrix and computed basis vectors of this example are provided in Supplementary Information Section 2. The two-level optimization applied in this example includes a pattern search for the first-level and a XGBoost method to capture the second-level relationships (Supplementary Information Section 3.2). We find that two dimensionless numbers are necessary to represent the dataset since no high value of the R^2 score (e.g., greater than 0.5) can be achieved if we set only one dimensionless number in the training. A systematic algorithm that determines the number of dimensionless numbers required to govern a physical system is provided in Supplementary Information Section 1.3.

We identify several low-dimensional patterns with the propriety of the scale-free from data. They can achieve a high R^2 for both the training and test sets (a table summarizing the identified dimensionless numbers is provided in Supplementary Information Section 4.2). Interestingly, we identify another dimensionless number (besides the keyhole number), which has been discovered by the theory-driven approach [33, 29]: the normalized energy density (NED) (Fig. 3c). It can be expressed as

$$\mathrm{NED} = \frac{\eta_m P}{V_s \rho C_p (T_l - T_0) H L}. \tag{15}$$

The NED represents the ratio of laser energy input within the powder layer to the sensible heat of melting. This dimensionless number governs the lack of fusion porosity in metal 3D printing, which is a well-known porosity mechanism due to insufficient laser energy input to fully melt the powder material [40]. The other dimensionless number $\frac{\eta_m P H}{V_s \rho C_p (T_l - T_0) d^3}$ in Fig. 3c is related to another porosity mechanism, i.e., keyhole porosity, caused by bubbles of gas trapped underneath the surface during the fluctuation of an unstable keyhole [28]. This dimensionless number is a modified normalized enthalpy product, i.e., $\mathrm{NEP} \cdot \frac{H}{d}$, where the normalized enthalpy product NEP is proven to be related to the keyhole instability, and a unstable keyhole with a high NEP could lead to keyhole pores [31]. The NEP can be expressed as

$$\mathrm{NEP} = \frac{\eta_m P}{V_s \rho C_p (T_l - T_0) d^2}. \tag{16}$$

Since the NEP is derived from the single-track laser scan condition [31], the modified term $\frac{H}{d}$ emerges to consider the effect of multiple-track scanning. Another identified low-dimensional pattern that can achieve and even higher R^2 (0.95) is shown in Fig. 3d. It is also based on NED and NEP, but with different modification terms. By reducing high-dimensional parameter space, many fewer experiments would be required to determine optimal processing conditions and parameters for new materials and thus ease the Edisonian burden endemic among current metal 3D printing practitioners.

3 Discussion

The two-level optimization scheme makes dimensionless learning very flexible. The first-level scheme guarantees the dimensional invariance (or dimensional homogeneity), and thus many representation learning methods can be used for the second-level scheme to capture scale-free relationships. We demonstrate polynomial and tree-based XGBoost [22] methods in the previous sections. However, the capability of the dimensionless learning can be improved by leveraging more methods, including deep neural networks [41], symbolic regression [42], and Bayesian machine learning [43]. Another highly promising candidate is the sparse identification of nonlinear dynamics (SINDy) [15], which can identify ordinary differential equations (ODEs) or partial differential equations (PDEs) from data by selecting dominant terms in a library including all the possible candidates. Dimensionless learning can significantly improve the capability of the data-driven discovery methods, such as SINDy, in which the identified governing equations are not necessarily dimensionally homogeneous because the candidate terms in the library could have different physical dimensions. By integrating the proposed dimensionless learning with SINDy, the forms of dimensionally homogeneous differential equations can be discovered directly from data, which provides even more insights and interpretations of the physical system. We demonstrate this using a simple example of a spring-mass-damper system in Supplementary Information Section 6. Using the proposed two-level optimization with SINDy to analyze the time series of the spring-mass-damper system, we can obtain not only the governing equation, but several key physical parameters, such as the natural time scale (i.e., the inverse of natural frequency), the natural length scale and the dimensionless damping coefficient (Fig. 4). The dimensionless form of the governing equation involves a reduced number of variables (from six to three), which is the minimalistic representation of the system.

The proposed approach reduces the learning space and eliminates the strong dependence between different variables by embedding a physical constraint (i.e., dimensional invariance) into machine learning algorithms. It is a physics-based

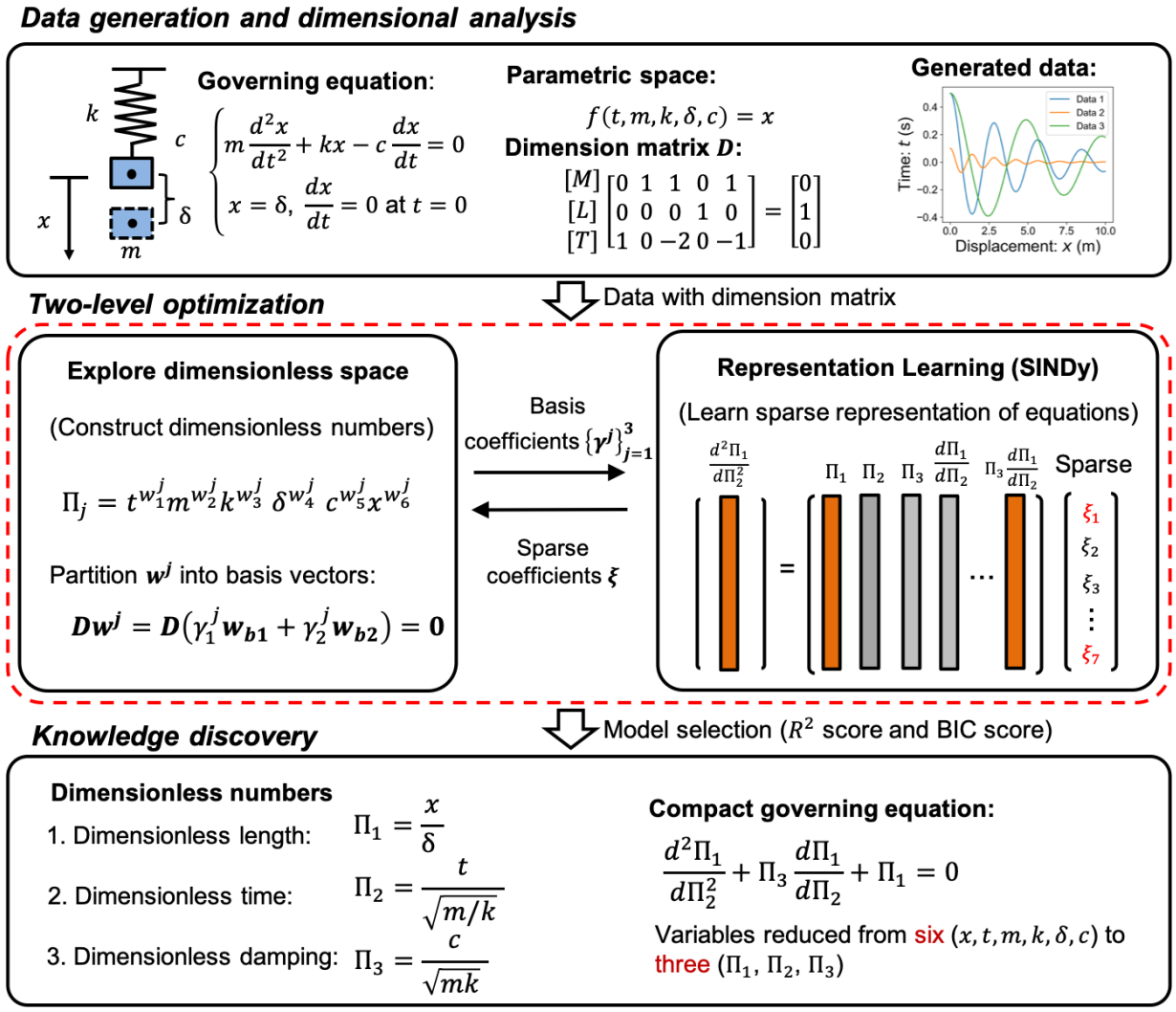

Figure 4: **Recover dimensionless numbers and the governing equation of a spring-mass-damping system.** In the first step, we generate three sets of data based on the governing equation with different combinations of parameters (spring constant k, mass m, initial displacement δ, and damping coefficient c). After analyzing the parameter space, we can obtain the dimension matrix $\boldsymbol{D}$. In the second step, a two-level optimization is developed to find the best basis coefficients and sparse coefficients sequentially. In third step, we can obtain the best dimensionless numbers and dimensionless governing equation after model selection using the R^2 score and Bayesian information criterion (BIC) score. Detailed information is shown in Supplementary Information Section 6.

dimension reduction approach, which can represent features as dimensionless numbers and transform the datapoints into a low-dimensional pattern that is insensitive to units and scales. It significantly improves the interpretability of the representation learning since the dimensionless numbers are physically interpretable, which enables qualitative and quantitative analysis of the systems of interest. It is worth noting that other kinds of invariances have been embedded into machine learning algorithms and achieved prominent successes, such as translation invariance in convolutional neural networks (CNNs), time translation invariance in recurrent neural networks (RNNs), and permutation invariance in graph neural networks (GNNs).

The proposed dimensionless learning enables a systematic and automatic learning of scale-free low-dimensional laws from a high-dimension parameter space, including many experimental conditions with different parameter settings. It can be applied to a vast number of physical, chemical, and biological systems to discover new or modify existing dimensionless numbers. Furthermore, it can be combined with other data-driven methods, such as SINDy, to discover dimensionless differential equations from high-resolution measurements. In material science, the identified compact mathematical expressions provide simple transition rules that translate optimal process parameters from one material (or existing materials) to another (or new materials). Dimensionless learning is able to reduce complex, highly multivariate problem spaces into descriptions involving just a few dimensionless parameters with clear physical meanings, which is particularly useful for engineering problems including many adjustable parameters with various dimensions or units, such as advanced materials processing and manufacturing, microfluidic flow control for precise drug delivery, solar energy systems design, and modern financial market analysis.

References

[1] Grigory Isaakovich Barenblatt. *Scaling*, volume 34. Cambridge University Press, 2003.

[2] Qing-Ming Tan. *Dimensional analysis: with case studies in mechanics*. Springer Science & Business Media, 2011.

[3] Josef Kunes. *Dimensionless physical quantities in science and engineering*. Elsevier, 2012.

[4] Osborne Reynolds. Xxix. an experimental investigation of the circumstances which determine whether the motion of water shall be direct or sinuous, and of the law of resistance in parallel channels. *Philosophical Transactions of the Royal society of London*, 174(174):935–982, 1883.

[5] Stephen J Kline. *Similitude and approximation theory*. Springer Science & Business Media, 2012.

[6] Kunal Ghosh and Binoy Krishna Mistry. Large incidence hypersonic similitude and oscillating nonplanar wedges. *AIAA Journal*, 18(8):1004–1006, 1980.

[7] Amir N Nahavandi, Frank S Castellana, and Edick N Moradkhanian. Scaling laws for modeling nuclear reactor systems. *Nuclear science and engineering*, 72(1):75–83, 1979.

[8] Dracos Vassalos. Physical modelling and similitude of marine structures. *Ocean engineering*, 26(2):111–123, 1998.

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

[10] Dale K Osborne. On dimensional invariance. *Quality and Quantity*, 12(1):75–89, 1978.

[11] Patricio F Mendez and Fernando Ordonez. Scaling laws from statistical data and dimensional analysis. *Journal of applied mechanics*, 72(5):648–657, 2005.

[12] A Murari, E Peluso, M Gelfusa, I Lupelli, and P Gaudio. A new approach to the formulation and validation of scaling expressions for plasma confinement in tokamaks. *Nuclear Fusion*, 55(7):073009, 2015.

[13] 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.

[14] Paul G Constantine, Zachary del Rosario, and Gianluca Iaccarino. Many physical laws are ridge functions. *arXiv e-prints*, pages arXiv–1605, 2016.

[15] 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.

[16] Lord Rayleigh. Lix. on convection currents in a horizontal layer of fluid, when the higher temperature is on the under side. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 32(192):529–546, 1916.

[17] Physics of Fluids Group University of Twente. https://www.youtube.com/watch?v=YkN6F6LJzGU, 2021. Accessed: 2021-10-26.

[18] JJ Niemela and Katepalli R Sreenivasan. Turbulent convection at high rayleigh numbers and aspect ratio 4. *Journal of fluid mechanics*, 557:411–422, 2006.

[19] Xavier Chavanne, F Chilla, B Chabaud, B Castaing, and B Hebral. Turbulent rayleigh–bénard convection in gaseous and liquid he. *Physics of Fluids*, 13(5):1300–1320, 2001.

[20] Ernst Gőbel, Ian Mills, and Andrew Wallard. The international system of units (si). 2006.

[21] D Calvetti and E Somersalo. Dimensional analysis and scaling. *The Princeton Companion to Applied Mathematics, Princeton University Press, Princeton, NJ, USA*, pages 90–93, 2015.

[22] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[23] C Kumar N Patel. Continuous-wave laser action on vibrational-rotational transitions of c o 2. *Physical review*, 136(5A):A1187, 1964.

[24] Cang Zhao, Qilin Guo, Xuxiao Li, Niranjan Parab, Kamel Fezzaa, Wenda Tan, Lianyi Chen, and Tao Sun. Bulk-explosion-induced metal spattering during laser processing. *Physical Review X*, 9(2):021052, 2019.

[25] Joung R Cook. High-energy laser weapons since the early 1960s. *Optical Engineering*, 52(2):021007, 2012.

[26] A PIRRI and R WEISS. Laser propulsion. In *Society of Naval Architects and Marine Engineers, and US Navy, Advanced Marine Vehicles Meeting*, page 719, 1972.

[27] Cang Zhao, Kamel Fezzaa, Ross W Cunningham, Haidan Wen, Francesco De Carlo, Lianyi Chen, Anthony D Rollett, and Tao Sun. Real-time monitoring of laser powder bed fusion process using high-speed x-ray imaging and diffraction. *Scientific reports*, 7(1):1–11, 2017.

[28] Cang Zhao, Niranjan D Parab, Xuxiao Li, Kamel Fezzaa, Wenda Tan, Anthony D Rollett, and Tao Sun. Critical instability at moving keyhole tip generates porosity in laser melting. *Science*, 370(6520):1080–1086, 2020.

[29] Zhengtao Gan, Orion L Kafka, Niranjan Parab, Cang Zhao, Lichao Fang, Olle Heinonen, Tao Sun, and Wing Kam Liu. Universal scaling laws of keyhole stability and porosity in 3d printing of metals. *Nature communications*, 12(1):1–8, 2021.

[30] Remy Fabbro, Morgan Dal, Patrice Peyre, Frédéric Coste, Matthieu Schneider, and Valerie Gunenthiram. Analysis and possible estimation of keyhole depths evolution, using laser operating parameters and material properties. *Journal of Laser Applications*, 30(3):032410, 2018.

[31] Jianchao Ye, Saad A Khairallah, Alexander M Rubenchik, Michael F Crumb, Gabe Guss, Jim Belak, and Manyalibo J Matthews. Energy coupling mechanisms and scaling behavior associated with laser powder bed fusion additive manufacturing. *Advanced Engineering Materials*, 21(7):1900185, 2019.

[32] Andrew Dawood, B Marti Marti, V Sauret-Jackson, and Alastair Darwood. 3d printing in dentistry. *British dental journal*, 219(11):521–529, 2015.

[33] Zekun Wang and Moubin Liu. Dimensionless analysis on selective laser melting to predict porosity and track morphology. *Journal of Materials Processing Technology*, 273:116238, 2019.

[34] Galina Kasperovich, Jan Haubrich, Joachim Gussone, and Guillermo Requena. Correlation between porosity and processing parameters in tial6v4 produced by selective laser melting. *Materials & Design*, 105:160–170, 2016.

[35] Pankaj Kumar, Farah Jano, Akram Javed, Chong Teng, Jon Ginn, and Misra Mano. Influence of laser processing parameters on porosity in inconel 718 during additive manufacturing. *The International Journal of Advanced Manufacturing Technology*, 103(1-4):1497–1507, 2019.

[36] JA Cherry, HM Davies, S Mehmood, NP Lavery, SGR Brown, and J Sienz. Investigation into the effect of process parameters on microstructural and physical properties of 316l stainless steel parts by selective laser melting. *The International Journal of Advanced Manufacturing Technology*, 76(5-8):869–879, 2015.

[37] A Leicht, M Rashidi, U Klement, and E Hryha. Effect of process parameters on the microstructure, tensile strength and productivity of 316l parts produced by laser powder bed fusion. *Materials Characterization*, 159:110016, 2020.

[38] Jacob C Simmons, Xiaobo Chen, Arad Azizi, Matthias A Daeumer, Peter Y Zavalij, Guangwen Zhou, and Scott N Schiffres. Influence of processing and microstructure on the local and bulk thermal conductivity of selective laser melted 316l stainless steel. *Additive Manufacturing*, 32:100996, 2020.

[39] Anton Du Plessis, Igor Yadroitsev, Ina Yadroitsava, and Stephan G Le Roux. X-ray microcomputed tomography in additive manufacturing: a review of the current technology and applications. *3D Printing and Additive Manufacturing*, 5(3):227–247, 2018.

[40] Anton du Plessis. Effects of process parameters on porosity in laser powder bed fusion revealed by x-ray tomography. *Additive Manufacturing*, 30:100871, 2019.

[41] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.

[42] Michael Schmidt and Hod Lipson. Distilling free-form natural laws from experimental data. *science*, 324(5923):81–85, 2009.

[43] David Barber. *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.

4 Acknowledgements

This study was supported by the National Science Foundation (NSF) through grants CMMI-1934367.

5 Author contributions

Z.G. proposed the original ideas, supervised the project, and wrote the manuscript. X.X. performed the research and wrote the manuscript. W.K.L. contributed to discussions and supervised the project. All the authors reviewed and edited the manuscript.

6 Competing interests

The authors declare no competing interests.

Supplementary Files

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

- Supplementaryinformation.pdf