



Thermodynamics of Learning Physical Phenomena

Elias Cueto¹ · Francisco Chinesta^{2,3}

Received: 16 March 2023 / Accepted: 25 May 2023 / Published online: 14 June 2023
© The Author(s) 2023

Abstract

Thermodynamics could be seen as an expression of physics at a high epistemic level. As such, its potential as an inductive bias to help machine learning procedures attain accurate and credible predictions has been recently realized in many fields. We review how thermodynamics provides helpful insights in the learning process. At the same time, we study the influence of aspects such as the scale at which a given phenomenon is to be described, the choice of relevant variables for this description or the different techniques available for the learning process.

1 Introduction

In a 2009 compilation of essays, the fourth paradigm of science was first described [1]. After centuries of science based on observation—the empirical period of the first paradigm—came a period based on the establishment of scientific laws—the second paradigm, think of Newton—and much more recently a period in which simulation took an important role—the third paradigm. Very recently, the authors of this essay argue that we have entered a period in which data plays a prominent role in scientific discovery and where theory and experiments, symbiotically, help data to achieve higher goals. Scientific Machine Learning is precisely a new field in which data coming from scientific experiments is used massively to unveil new, still unknown scientific laws. Some authors have begun to think about an even more recent fifth paradigm of science, in which data is obtained not from experiments, but from simulations [2]. This approach is helping scientists to look for the origin of dark matter [3] or the structure of protein folding [4], to

name but two of the most relevant examples. Other data-driven approaches do not look for a closed-form scientific expression describing a particular phenomenon, but substitute phenomenological laws of low epistemic value (typically, constitutive or closure laws) by data [5–10]. Of course, this approach is somewhat more limited in terms of extrapolation capabilities. Although it largely exceeds the scope of this paper, the interested reader could find excellent reviews on machine learning in [11–13], for instance.

No doubt that this data-intensive approach is revolutionizing science. But, on a much more applied context, scientific machine learning is also revolutionizing industry. In the same way that these techniques can look for general laws of physics, they can also look for rigorous descriptions of the functioning of technical apparatus, thus helping us to develop digital twins in a very efficient manner [14–16].

However, distilling scientific laws about any physical phenomena, regardless of its practical importance, has profound implications. Galileo, for instance, in his book on two new sciences—one of which was mechanics of materials—failed completely in the description of beam bending phenomena. He assumed that rectangular sections of a beam rotate due to bending around an axis passing through the lower surface of the beam, and not through the center of mass, as it is well known today [17]. However, he was intelligent enough to think of rotating *sections* of the beam, which is actually the cornerstone of the celebrated Euler-Bernoulli-Navier bending theory. Thus, even if his theory was not entirely true, his choice of variables was correct. Or at least it was at that macroscopic scale of description. Of course, at that time the atomic structure of matter was not known, but this extremely fine scale of description is equally valid, albeit much less

✉ Elias Cueto
ecueto@unizar.es

Francisco Chinesta
Francisco.Chinesta@ensam.eu

¹ ESI Group Chair, Aragon Institute of Engineering Research, Universidad de Zaragoza, Maria de Luna s/n, 50018 Zaragoza, Spain

² ESI Group Chair, PIMM Lab, ENSAM Institute of Technology, 151 rue de l'Hôpital, 75013 Paris, France

³ CNRS@CREATE, 1 Create Way, #08-01 Create Tower, Singapore 138602, Singapore

efficient. Or that of the theory of elasticity within a continuum mechanics framework. Or that of the Cosserat brothers, that includes rotations of the material point—the beam bending theory could be considered itself as a Cosserat theory on a one-dimensional continuum [18]. Any of these scales is in principle valid, yet Galileo chose the right one, as did Euler, Bernoulli and Navier. Here, by “right” we mean the most useful one in terms of engineering practice. In Sect. 2 we briefly review the statistical mechanics aspects related to the choice of an appropriate level of description for a given physical phenomenon.

But the employ of machine learning is by no means restricted to unveil unknown physical laws. Still today, and despite the impressive interest on big data approaches in social sciences, the biggest supercomputers in the world continue to be devoted to the simulation of complex phenomena [19]. Full-field and high-resolution simulations continues to be challenging, and the possibility of developing a promising family of *learned simulators* is appealing [20–22]. These learned simulators share the characteristics of being considerably faster than traditional simulation techniques such as finite elements or finite volumes (once trained, of course, this does not take into account the considerable effort invested in training these networks). While traditional techniques employ a non-negligible effort to construct each model, learned simulators, in general, employ reusable architectures that can be employed for different, but related, problems. Additionally, the results of these learned simulators are, in general, as accurate as the data employed to train the network, and do not depend on models whose validity could be compromised by extreme parameter values, for instance. Another advantage includes the possibility of employ these simulators for inverse problems and optimization procedures, since they are based on differentiable networks [23–26].

Despite of the above-mentioned advantages, neural networks are not very popular among the scientific community in general, and particularly in the computational mechanics one [27, 28]. Common pitfalls include large deviations from the expected result arising as a consequence of small perturbations in data, or the well-known phenomenon of overfitting the data (basically, learning the noise in the data), for instance. This has motivated a strong effort of research towards the inclusion of previous knowledge about the physics taking place in the learning procedure. This knowledge could be included in the form of inductive biases, for instance [29] or the particular form of the differential equation governing the physics.

Recently, very popular methods based on neural networks have been developed for the solution of Partial Differential Equations (PDEs) from data [30] that try to overcome these limitations. These methods, coined globally as Physics-informed Neural Networks (PINNs), allow one to solve PDEs

from data by following a scheme that resembles a collocation method in some sense. Thus, if the PDE governing the problem is known—hence the “physics-informed” character—, its particular form can be unveiled from the available data. We devote Sect. 3 to this family of techniques.

There is still another family of machine learning techniques that consider the learning process as a sort of regression procedure operating on an Ordinary Differential Equation (ODE) governing the dynamics of the system. These techniques can be seen as a particular instance of the proposal by Weinan [31]. These techniques operate when the particular form of the PDE governing the physics is not known, nor sought. Instead, as will be seen, this particular form of seeing the learning problem leverages the vast corps of knowledge around dynamical systems to enforce the right structure of the system (conserved quantities, symmetries, etc.). To this family of techniques we devote Sect. 4.

As will be seen throughout this paper, the employ of neural networks for learning physical phenomena is an active field of research, that has produced hundreds of papers in a very limited period of time. Recently, methods have been developed that are able to unveil constitutive laws from displacement data alone—with global force but, notably, no stress data—which constitute a great breakthrough in this discipline, see [32–34]. Nevertheless, the topic is still in its infancy, and much more research is expected before we will be able to fully understand and predict the behavior of machine learning techniques, particularly neural networks, in this amazing discipline. The paper ends with some conclusions about these and other considerations in Sect. 5.

2 Statistical Mechanics of Coarse Graining

The simplest approach to learn physics, at least conceptually, consists in going down to the molecular dynamics scale, where Newton laws apply, label every molecule and follow them across their travel. This is not a useful approach, obviously, but conceptually is the simplest one. At this scale, everything is reversible. So if we denote by $\mathbf{z} = \{\mathbf{q}_1, \dots, \mathbf{q}_N, \mathbf{p}_1, \dots, \mathbf{p}_N\}$, where N is the number of molecules (on the order of 10^{23} to have meaningful results) and $\mathbf{q}_i, \mathbf{p}_i, i = 1, \dots, N$, represent the position and momentum of every molecule, the evolution in time of the system is well-known to obey a Hamiltonian evolution, i.e.,

$$\dot{\mathbf{z}} = \frac{\partial \mathbf{z}}{\partial t} = \mathbf{L} \nabla \mathcal{H} = \mathbf{L} \nabla E, \quad (1)$$

with \mathbf{L} the so-called symplectic (skew-symmetric) matrix and \mathcal{H} the Hamiltonian of the system, the total energy, E . So, learning the system at this scale means learning the particular form of \mathbf{L} and E by regression (a form of supervised learning). Distilling these from data will ensure that

predictions obtained through Eq. (1) will conserve energy, thanks to the symplectic structure of the learned description.

Since this approach is by no means practical, we will need to coarse-grain the description of the system. In other words, we will need to represent it with less degrees of freedom than it actually has. This process is described with particular elegance in a paper by Pep Español, whose title we borrow for this section [35]. Alternative approaches can be found, for instance, in [36]. In essence, the process consists in eliminating those degrees of freedom whose evolution in time is faster, keeping those whose evolution takes place over longer periods of time, thus allowing us to describe the system over longer intervals with less computational effort. At the other side of the spectrum, (equilibrium) thermodynamics is a description of the system that takes into account only invariants, magnitudes that do not evolve in time (typically, mass, momentum, energy).

Each possible level of description involves a set of variables, which we will denote by \mathbf{z} , regardless of the particular scale, if there is no risk of confusion. Coarser levels of description will provide us with less information, but will involve smaller sets \mathbf{z} . The molecular dynamics scale given by Eq. (1) has a typical time scale on the order of the collision time (again, unpractical for our purposes). In any case, every description will be governed by some equation that, very much like Eq. (1) predicts the evolution of the variables governing the system. It is also important to realize that, if a clear separation of scales exists between conserved and eliminated degrees of freedom, the resulting description will be Markovian. In other words, the coarse-grained model will not depend on history, but only in the present value of its variables.

Even in the case in which the description is Markovian, it is worth noting that many microscopic states could lead to the same macroscopic state. This produces uncertainty in the future evolution of the coarse-grained description of the system, see [35, 37, 38]. This uncertainty is equivalent to fluctuation in the evolution, and fluctuation is equivalent to dissipation by the celebrated fluctuation-dissipation theorem [39, 40].

Español takes a system with two clearly separated scales to elaborate around this theory: a colloidal suspension [35]. The finest description is obtained, as discussed before, by taking position and momenta of both colloidal particles, $\mathbf{Q}_i, \mathbf{P}_i$ and solvent molecules, $\mathbf{q}_j, \mathbf{p}_j$, with $i = 1, \dots, N_{\text{col}}$, $j = 1, \dots, N_{\text{sol}}$ the number of colloidal particles and solvent molecules, respectively. At this scale, the energy of the system is composed by the potential and kinetic energy of the particles, and the typical time scale is on the order of picoseconds.

If we are not interest in describing the movement of every solvent molecule, we can substitute them by a continuous hydrodynamic field, in which extensive magnitudes as mass

density, momentum density and energy density fields substitute them. In this framework we loose information about the precise location of each molecule, but we have instead information about how many of them are located within a given region of the fluid. Particles entering and leaving the considered region in the fluid produce the fluctuations mentioned before [41]. These fluctuations are in turn responsible of the Brownian motion of the colloidal particles. Dissipative Particle Dynamics, for instance, could be employed for the description of the solvent at this scale [42–44].

If the colloidal particles are massive, if compared to those of the solvent, their change in position and momentum will be slow, compared to the scale of hydrodynamic interactions. Therefore, one could envisage the elimination of the hydrodynamic field in order to keep position and momenta of the colloidal particles as the sole variables of our system, $\mathbf{z} = \{\mathbf{P}_i, \mathbf{Q}_i\}$, $i = 1, \dots, N_{\text{col}}$. This description gives rise to a Fokker-Planck (FP) equation. The general FP equation works with a probability density function $\psi(\mathbf{z}, t)$ that reflects the probability of finding a particle at a given position at a given time instant. Its evolution in time is given by

$$\frac{\partial \psi(\mathbf{z}, t)}{\partial t} = -\frac{\partial}{\partial \mathbf{z}} \cdot (\mathbf{A}(\mathbf{z}, t)\psi(\mathbf{z}, t)) + \frac{1}{2} \frac{\partial}{\partial \mathbf{z}} \frac{\partial}{\partial \mathbf{z}} : (\mathbf{D}(\mathbf{z}, t)\psi(\mathbf{z}, t)). \tag{2}$$

\mathbf{A} is a term that takes into account deterministic effects related to the macroscopic velocity drift, while \mathbf{D} is a diffusion tensor related to Brownian effects [45, 46]. The FP equation has an equivalent Itô stochastic differential equation of the form

$$d\mathbf{z} = \mathbf{A}(\mathbf{z}, t)dt + \mathbf{B}(\mathbf{z}, t) \cdot d\mathbf{W}, \tag{3}$$

where $\mathbf{D} = \mathbf{B} \cdot \mathbf{B}^T$ and \mathbf{W} is a multidimensional Wiener process. In the case of our colloidal suspension, the influence of the hydrodynamic field is taken into account by introducing a friction tensor ζ_{ij} that depends on the relative position of colloidal particles i and j . Its equivalent Itô stochastic differential equation has the form

$$d\mathbf{Q}_i = \frac{\mathbf{P}_i}{M_i} dt, \quad d\mathbf{P}_i = \mathbf{F}_i^{CC} dt - \sum_j \zeta_{ij} \cdot \mathbf{V}_j dt + d\tilde{\mathbf{F}}_i,$$

with $\mathbf{V}_i = \mathbf{P}_i/M_i$, \mathbf{F}_i^{CC} a force of interaction among colloidal particles and $\tilde{\mathbf{F}}_i$ a stochastic force described by a Wiener process that takes the form $d\tilde{\mathbf{F}}_i \cdot d\tilde{\mathbf{F}}_j = 2k_B T \zeta_{ij} dt$ [47].

Still a coarser description can be obtained if the colloidal particles are placed at distant locations from each other. In that case, we can assume that $\zeta_{ij} = \delta_{ij} \mathbf{I} \zeta$, where ζ is a friction coefficient. This gives rise to the set of so-called Langevin equations. All this works well only if the density of colloidal particles is much bigger than that of the solvent. Otherwise, we can not isolate position and momenta of the colloidal particles as the sole variables in our model.

Coarse-graining this model ever further, one could think that the position of the colloidal particles changes much more slowly than momentum, and try to keep \mathbf{Q}_i alone as governing variables of our model. This gives rise to the so-called Smoluchowsky equation [48]. One can even think that we are not interested in knowing the position of each and every colloidal particle in our suspension, but on their density over a particular region. This is possible by introducing a concentration field as the sole variable of the system. This gives rise, in turn, to the well-known Fick equation.

All this process ends by considering that we are only interested in the system at equilibrium. The description that takes into account invariants of the system only is the scale of thermodynamics. For such a simple system we have mentioned six different possible scales at which we can describe its dynamics. As mentioned before, at each of these scales a given degree of uncertainty will appear, as a consequence of our lack of knowledge about the precise microscopic state (at the molecular dynamics scale). This uncertainty appears as stochasticity. In general, at each level of description, the least biased distribution is the one that maximizes the entropy functional at that scale.

It is therefore evident that by performing experiments at a given scale—something that depends usually on the available instruments and not on our choice—we are choosing a particular description of the phenomenon under scrutiny that may heavily influence the way in which the system shows to us. In the absence of previous knowledge on the system, we are not even in the position of ensuring the non-Markovian character of the description. Recently, however, data analysis methods have been devised that help in determining the precise number of internal, phenomenological variables for an accurate description of the history of the system [38].

Under this prism, we could ask ourselves if a suitable description of our system is already available in the literature, or if the phenomenon is completely unknown to us. In the former case, this description will be given frequently in the form of a partial differential equation. The general form of this PDE could be known, but not the precise values of its coefficients nor the boundary conditions applying under the laboratory conditions. Should this be the case, the formalism of physics-informed neural networks (PINNs) becomes the natural way to tackle the problem, see Sect. 3. If, on the other hand, we do not have any information on the particular form of the law being sought—think of the equations governing dark matter, for instance—a supervised learning approach based on the dynamical systems equivalence should be preferred. These will be deeply analyzed in Sect. 4 below.

3 Physics-Informed Neural Networks

Physics-informed neural networks (PINNs) have constituted a great success in computational mechanics and mathematics and are, perhaps, the responsible of a very active research activity in the field [30, 49–56]. PINNs assume that a general, nonlinear PDE of the form

$$\dot{\mathbf{z}} + \mathcal{N}[\mathbf{z}; \boldsymbol{\mu}] = \mathbf{0}, \quad (4)$$

is known to govern the physics taking place. Here, \mathcal{N} is a nonlinear differential operator and $\boldsymbol{\mu}$ represents a set of parameters. Of course, $\mathbf{z}(\mathbf{x}, t)$, $\mathbf{x} \in \Omega$ represent the governing variables of the problem, defined in some open set $\Omega \subset \mathbb{R}^n$. The problem is then established so as to find, given some measurement on the system, the value of the variables governing it, $\mathbf{z}(\mathbf{x}, t)$, and the fittest parameters $\boldsymbol{\mu}$ that produce these measurements.

PINNs somehow resemble a collocation method. Given some measurements $\mathbf{z}^i(t_z^i, \mathbf{x}_z^i)$, the residual of Eq. (4) is defined as

$$\mathcal{R} = \dot{\mathbf{z}} + \mathcal{N}[\mathbf{z}; \boldsymbol{\mu}].$$

By approximating \mathbf{z} through a deep neural network, and applying automatic differentiation, one arrives at a neural network approximation to \mathcal{R} . By minimizing the mean squared error defined as

$$MSE = MSE_z + MSE_{\mathcal{R}}, \quad (5)$$

with

$$MSE_z = \frac{1}{N_z} \sum_{i=1}^{N_z} \left| \mathbf{z}(t_z^i, \mathbf{x}_z^i) - \mathbf{z}^i \right|^2,$$

and

$$MSE_{\mathcal{R}} = \frac{1}{N_{\mathcal{R}}} \sum_{i=1}^{N_{\mathcal{R}}} \left| \mathcal{R}(t_{\mathcal{R}}^i, \mathbf{x}_{\mathcal{R}}^i) \right|^2,$$

with $\{t_z^i, \mathbf{x}_z^i, \mathbf{z}^i\}_{i=1}^{N_z}$ denote measurements on the initial and boundary values and $\{t_{\mathcal{R}}^i, \mathbf{x}_{\mathcal{R}}^i\}_{i=1}^{N_{\mathcal{R}}}$ represent the collocation points for \mathcal{R} .

When learning the physics governing some given phenomenon, knowing in advance the PDE best describing it may seem as a too stringent condition, but the fact is that we already know many details about much of the physics surrounding us. In fact, in 1929, Paul Dirac said that [57],

The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble.

It seems therefore reasonable to think that some form of PDE could be envisageable for the phenomena under scrutiny. Even if this is not completely true—think again of an engineer trying to model a beam from observation: should he or she consider an Euler-Bernoulli-Navier model or a Timoshenko one?—PINNs constitute nowadays a very popular method for solving PDEs from data.

As discussed in Sect. 2 above, learning physics from data is even more difficult than that. The just presented approach does no guarantee, of course, that for any reason, the resulting prediction made by such a learned PDE will be consistent with the principles of thermodynamics that, as justified before, act as a restriction of a very high epistemic level.

If some symmetries of system are known in advance (remember that, thanks to the Noether’s theorem, for each symmetry of the system there is a conserved quantity, see [58] or its recent translation to English [59]) these can be imposed beforehand. This is precisely the path followed in [60] for Hamiltonian systems of the form given by Eq. (1). The PDE (4) is then restricted to an ordinary differential equation (ODE). We will come back to this approach in Sect. 4 below.

3.1 Thermodynamics-Based Artificial Neural Networks

In a series of papers, I. Stefanou and coworkers presented recently a technique that employs thermodynamics as a restriction for learning constitutive laws [61–64]. They coined the technique as Thermodynamics-based Artificial Neural Networks (TANNs). TANNs encode the two laws of thermodynamics by making use of automatic differentiation, thus ensuring by constructions the fulfillment of these laws, without the need to learn them from data. Thus, it also avoids the lack of fulfillment of these laws for previously unseen data.

Conservation of energy can be expressed as a PDE of the form

$$\rho \dot{e} = \sigma \cdot \nabla^s \mathbf{v} - \nabla \cdot \mathbf{q} + \rho h, \tag{6}$$

with ρ the density, σ the stress tensor, \mathbf{v} the velocity field, \mathbf{q} the heat flux and h the energy supply per unit mass. In turn, the second principle of thermodynamics can be expressed as

$$\rho(\theta \dot{s} - \dot{e}) + \sigma \cdot \nabla^s \mathbf{v} - \frac{\mathbf{q} \cdot \nabla \theta}{\theta} \geq 0, \tag{7}$$

with s the specific entropy and θ the temperature.

TANNs try to avoid black-box ANNs by imposing the fulfillment of Eqs. (6) and (7). When fed with the current state of the material, $\mathbf{z}_t = \{\epsilon_t, \Delta \epsilon, \sigma_t, \theta_t, \zeta_t, \Delta t\}$, with ζ a set of internal variables and t the time, they produce an output composed by $\{\Delta \zeta, \Delta \theta, F_{t+\Delta t}, \Delta \sigma\}$, where $F = E - S\theta$ is the

Helmholtz free energy, $S = \rho s$ and E is an energy potential (assumed to be rate-independent such that $\sigma = \frac{\partial E}{\partial \epsilon}$ and $\theta = \frac{\partial E}{\partial s}$).

TANNs actually learn from data the values of two scalars, the Helmholtz free energy F and the dissipation $D = \rho(\theta \dot{s} - \dot{e}) + \sigma \cdot \nabla^s \mathbf{v}$. For this to be possible, the increment in internal variables and temperature are also learned. The rest of the ingredients of Eqs. (6) and (7) are computed by automatic differentiation.

A somewhat related approach has recently been presented in [65]. An alternative approach based on exterior calculus is also developed in [66].

3.2 Variational Onsager Neural Networks

Huang and coworkers developed recently a technique based on the application of the Onsager variational principle [67, 68] which they coined as Variational Onsager Neural Networks, VONNs [69]. Again, their technique is based on the learning of two potentials: the free energy F and the dissipation potential D . The resulting NN enforces strongly the fulfillment of the second law of thermodynamics.

For systems at constant temperature and free of inertial effects, the Onsager principle looks like

$$\min_{\zeta} \mathcal{R}[\mathbf{z}, \zeta], \tag{8}$$

where \mathcal{R} is the so-called Rayleighian and \mathbf{z} and ζ represent the state variables and the process (internal) variables, respectively. The Rayleighian takes the form

$$\mathcal{R}[\mathbf{z}, \zeta] = \dot{F}[\mathbf{z}, \zeta] + D[\mathbf{z}, \zeta] + P[\mathbf{z}, \zeta], \tag{9}$$

with $P[\mathbf{z}, \zeta]$ the power supplied by external forces.

From the variational principle, Eq. (8), we arrive at

$$\frac{\delta \dot{F}}{\delta \zeta} + \frac{\delta D}{\delta \zeta} + \frac{\delta P}{\delta \zeta} = 0.$$

Both F and D are learned through a different NN, by minimizing the loss function *à la* PINN, see Eq. (5). The convexity of the dissipation potential D must be imposed explicitly, by making use of the input convex neural network paradigm [70]. This may be also the case for the free energy potential F .

VONNs work after data for \mathbf{z} and ζ . This means that, very much like TANNs, this technique needs for data on internal variables that are often impossible to measure. If these networks work is by the availability of synthetic data, coming from simulations. We will discuss further about this limitation in Sect. 4. In any case, both techniques are able to impose in a soft manner the fulfillment of the second law of thermodynamics.

In the next section we analyze a different family of techniques based on dynamical systems, ODEs, instead of PDEs, and discuss their relative advantages and disadvantages.

4 Learning Based on a Dynamical Systems Analogy

On a somewhat different setting, Weinan showed that the learning problem has the same structure of a dynamical system [31, 37, 71]. In these works, it is suggested that supervised learning has the same structure of a dynamical system of the form

$$\frac{dz}{dt} = f(z, t), \quad z(0) = z_0, \tag{10}$$

so that the flow map

$$z_0 \rightarrow z(T, z_0),$$

is produced by a non-linear function f whose precise form is sought. This function can be found by employing classical regression methodologies such as linear regression, support vector machines [72], or others. Of course, given the availability of the universal approximation theorem for neural networks, these appear as an appealing choice [73–75]. But the most interesting part of this approach is to recognize that, given the form of a dynamical system, all our previous knowledge in the field can be advantageously exploited to impose known structures in the system. The simplest structures one can think of are, of course, the Hamiltonian structure given by Eq. (1) if the system is known to be conservative, or the gradient flow structure, for instance [76]. For non-conservative variables, their evolution can be established after some potential \mathcal{R} in the form [77]

$$\frac{dz}{dt} = -\frac{\partial \mathcal{R}}{\partial z}.$$

As can be readily noticed, this framework assumes no known form for the problem at hand. In sharp contrast to the framework of PINNs, whose governing PDE is assumed to be known, this dynamical systems equivalence assumes no previous knowledge on the physics taking place from which data are obtained. We explore this formalism more in detail in the following sections.

4.1 Neural Networks for Conservative Systems

This framework has attracted the interest of many researchers in recent times. For instance, in [60] a method is developed for systems of the form given by Eq. (1). To better explain this method, let us introduce some notation first.

A differentiable map $\phi : U \subset \mathbb{R}^{2N} \rightarrow \mathbb{R}^{2N}$ is said to be symplectic if

$$\left(\frac{\partial \phi}{\partial x}\right)^\top L^\top \left(\frac{\partial \phi}{\partial x}\right) = L^\top.$$

In particular, if $\phi_t(z_0)$ is the flow map of a Hamiltonian system, it is a symplectic map,

$$\left(\frac{\partial \phi_t}{\partial z_0}\right)^\top L^\top \left(\frac{\partial \phi_t}{\partial z_0}\right) = L^\top.$$

Assume that we are in the position of obtaining experimental data about the system at hand in the form

$$\mathcal{D} = \{z_i, y_i = \phi_h(z_i)\}_{i=1}^{n_{\text{samples}}},$$

with h the time step size for an adequate numerical integration scheme on the problem. With these data we can think of constructing a feedforward neural network whose loss is of the form

$$MSE = MSE_d + \lambda \cdot MSE_s,$$

with

$$MSE_d = \frac{1}{n_{\text{samples}}} \sum_{i=1}^{n_{\text{samples}}} \|\phi_h(z_i) - y_i\|^2$$

the typical loss on the accuracy of the predictions and

$$MSE_s = \frac{1}{n_{\text{samples}}} \sum_{i=1}^{n_{\text{samples}}} \left\| \left[\left(\frac{\partial \phi_t}{\partial z}\right)^\top L^\top \left(\frac{\partial \phi_t}{\partial z}\right) \right] (z_i) - L^\top \right\|^2. \tag{11}$$

λ is a parameter to take into account the different relative size of both losses. This loss (11) enforces the symplectic structure of the sought matrix L , thus enforcing in turn the Hamiltonian character of the resulting approximation and its inherent conservative character. The resulting network works actually as a time integrator with time step h .

Many researchers have followed similar strategies. For instance, a year or so before Jin et al. and Greydanus et al. introduced the so-called Hamiltonian Neural Networks for canonical, discrete systems, in which the loss term is of the form [78]

$$MSE = \left\| \frac{\partial \mathcal{H}}{\partial p} - \frac{\partial q}{\partial t} \right\|_2 + \left\| \frac{\partial \mathcal{H}}{\partial q} + \frac{\partial p}{\partial t} \right\|_2, \tag{12}$$

where q and p represent, respectively, position and momenta of discrete particles. A similar approach is followed in [79]. The same loss function in Eq. (12) is employed in [80]. However, in this work the authors solve parametric problems, so the input data set is collected for different values of

these parameters, thus generalizing this type of networks, while keeping their conservative character.

Even more sophisticate loss functions could be considered. For instance, Bertalan et al. assume that their data includes not only position and momenta, but also their derivatives. Their loss term is therefore of the form [81]

$$\text{Loss} = \sum_{k=1}^4 c_k f_k,$$

where each one of the four terms look like

$$\begin{aligned} f_1 &= \left\| \frac{\partial \mathcal{H}}{\partial \mathbf{p}} - \frac{\partial \mathbf{q}}{\partial t} \right\|_2, \\ f_2 &= \left\| \frac{\partial \mathcal{H}}{\partial \mathbf{q}} + \frac{\partial \mathbf{p}}{\partial t} \right\|_2, \\ f_3 &= (\mathcal{H}(\mathbf{q}_0, \mathbf{p}_0) - \mathcal{H}_0)^2, \end{aligned}$$

(this term serves for disambiguation only, since the Hamiltonian can be determined up to a constant. It is therefore assume to be known at a point $(\mathbf{q}_0, \mathbf{p}_0)$), and

$$f_4 = \left\| \frac{\partial \mathcal{H}}{\partial \mathbf{q}} \dot{\mathbf{q}} + \frac{\partial \mathcal{H}}{\partial \mathbf{p}} \dot{\mathbf{p}} \right\|_2.$$

The number of works that employs related approaches is huge and dates back to at least 1993 [82]. In [83], for instance, an improved training method is developed for this type of networks that is based on the symplectic character of the equations. Finzi et al. suggest a modification of the above techniques by working on cartesian coordinates and not in the phase space [84]. To improve the expressiveness of these networks, Tong et al. introduce a method that adds Taylor series expansions designed with symmetric structure [85].

Always with the same Hamiltonian structure in mind, Chen et al. introduce them in the realm of recurrent neural networks, RNNs [86]. Even Generative Neural Networks exist under this Hamiltonian prism [87]. Again following a similar rationale, Di Pietro et al. improve the results by introducing a fourth-order time integration scheme in the learning scheme [88]. Choudhary et al. focus their attention in the transition from ordered to chaotic systems, and show that Hamiltonian Neural Networks (HNNs) perform much better than classical, black-box NNs [89]. Other works also confirm the superiority of HNNs over classical NNs [90, 91].

But these architectures are interesting not only by their inductive biases, that force them to follow conservative dynamics. Galimberti and coworkers have also demonstrated that HNNs eliminate by construction the problem of vanishing gradients [92], present in many NN architectures [93]. Even a very recent survey paper has been written on this particular family of techniques [94].

An alternative route to follow in this vast family of conservative phenomena is to employ Lagrangian, instead of Hamiltonian, formalisms. One advantage of doing so is the possibility of employing arbitrary coordinates instead of canonical coordinates. Although we will see later on that this does not constitute a true limitation for many systems of practical interest, canonical coordinates satisfy a set of rules in terms of the so-called Poisson bracket. The Lagrangian formalism assumes a data set composed by $\mathbf{z} = \{\mathbf{q}, \dot{\mathbf{q}}\}_{i=1}^{n_{\text{samples}}}$ (position and velocities of the particles). As it is well known, the Lagrangian formalism defines the so-called action functional as

$$S = \int_{t_0}^{t_1} (T(\mathbf{q}, \dot{\mathbf{q}}) - V(\mathbf{q})) dt,$$

with T the kinetic energy and V the potential energy, so that a dynamical system will follow a path given by the minimum value of S . This value is obtained through the restriction to the so-called Euler-Lagrange equations,

$$\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{\mathbf{q}}} = \frac{\partial \mathcal{L}}{\partial \mathbf{q}},$$

with $\mathcal{L} = T - V$ the Lagrangian of the system. This Lagrangian is precisely the objective to reconstruct from data. By standard algebraic manipulations, we arrive at

$$\ddot{\mathbf{q}} = (\nabla_{\dot{\mathbf{q}}} \nabla_{\dot{\mathbf{q}}}^T \mathcal{L})^{-1} \left[\nabla_{\mathbf{q}} \mathcal{L} - (\nabla_{\mathbf{q}} \nabla_{\dot{\mathbf{q}}}^T \mathcal{L}) \dot{\mathbf{q}} \right].$$

From this expression it is therefore straightforward to define a loss function

$$\text{Loss} = \|\dot{\mathbf{q}}^{\mathcal{L}} - \ddot{\mathbf{q}}^{\text{true}}\|_2,$$

so as to obtain an approximation to \mathcal{L} . This is the approach followed in [95–101], among others. Again, the amount of works devoted to this approach and their recent dates prove the interest of the community in these approaches.

4.2 Neural Networks for Dissipative Phenomena

Despite the success in the development of network architectures that impose conservation of energy (achieved mainly in the last two years), many researchers have realized that, following the reasoning in Sect. 2, dissipation is present in nearly every phenomenon of interest. Therefore, it is of utmost importance to develop techniques able to satisfy the principles of thermodynamics in the presence of dissipation. We make an overview of these techniques in this section.

4.2.1 Deconstructing Inductive Biases

Some authors, aware of the need to include dissipation in the formulation, have opted for its direct inclusion by just relaxing

the fulfillment of the restrictions (inductive biases) introduced in the last section. This is the approach followed in the Symplectic ODE nets (symODEN) approach [102, 103], inspired by the literature on controlling dynamical systems. It is also the approach followed in a very recent approach [104]. Actually, both approaches are heavily influenced by the port-Hamiltonian approach to dynamical systems, which has a strong tradition in the introduction of dissipation and control in the formulation of dynamical systems [105–107]. Essentially, port-Hamiltonian systems consider an evolution of the system in the form

$$\begin{bmatrix} \dot{q} \\ \dot{p} \end{bmatrix} = \left(\begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{I} & \mathbf{0} \end{bmatrix} - \mathbf{D}(q) \right) \begin{bmatrix} \frac{\partial \mathcal{H}}{\partial q} \\ \frac{\partial \mathcal{H}}{\partial p} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{g}(q) \end{bmatrix} \mathbf{u}, \quad (13)$$

where, as can be noticed, dissipation is included through a symmetric, positive semi-definite matrix \mathbf{D} , and control is considered through an actuation \mathbf{u} . The formulation in Eq. (13) recovers the Hamiltonian structure if no dissipation nor control are present. However, as will be demonstrated later, this formulation does not guarantee the fulfillment of the principles of thermodynamics and can be considered, in some sense, phenomenological.

Of course, neural networks based upon this formulation have been developed in the last years [108–113]. They all share the more or less the same ingredients and are based upon the formulation just discussed.

An alternative route is the one followed by Wang and coworkers [114]. Given the equivalence of symmetries and conservation laws for different magnitudes, the authors choose to begin by employing equivariant neural networks [115–117]. Then, these requirements on invariance are relaxed. But again, the lack of thermodynamic foundations of this method does not guarantee, in principle, the fulfillment of the principles of thermodynamics.

4.2.2 Metriplectic Neural Networks

As mentioned above, the port-Hamiltonian or relaxed equivariance formalisms do not guarantee but a phenomenological fulfillment of the laws of thermodynamics. There is no guarantee that, once faced to previously unseen data, or trained with noisy data, these networks will produce an output with the right amount of dissipation.

In order to develop a consistent formulation, let us assume that, at a given level of description, as discussed in Sect. 2, the invariants of the system, $I(\mathbf{z})$ can be expressed as a function of the resolved variables,

$$I(\mathbf{z}) = \mathcal{I}(\mathbf{z}),$$

for some suitable function \mathcal{I} . Let us also assume that the Hamiltonian of the system can be expressed as a function of these variables,

$$\mathcal{H}(\mathbf{z}) = E(\mathbf{z}),$$

with E the actual energy of the system. In these circumstances, the Fokker-Planck equation (2) and, more particularly, its equivalent Itô stochastic differential equation (3) takes the form [42]

$$d\mathbf{z} = \left[\mathbf{L}(\mathbf{z}) \frac{\partial E}{\partial \mathbf{z}} + \mathbf{M}(\mathbf{z}) \frac{\partial S}{\partial \mathbf{z}} + k_B \nabla \mathbf{M}(\mathbf{z}) \right] dt + d\tilde{\mathbf{z}}, \quad (14)$$

where k_B is the Boltzmann constant, $\mathbf{M}(\mathbf{z})$ is a symmetric, positive semi-definite dissipation matrix, S is a second potential (the so-called Massieu potential, entropy at this level of description) and $d\tilde{\mathbf{z}}$ is a Wiener process that satisfies

$$d\tilde{\mathbf{z}} = \mathbf{B}(\mathbf{z}) d\mathbf{W}(t),$$

with \mathbf{B} a non-square matrix satisfying

$$\mathbf{B}(\mathbf{z})\mathbf{B}(\mathbf{z})^\top = 2k_B \mathbf{M}(\mathbf{z}).$$

The importance of thermal fluctuations is controlled by the relative value of the Boltzmann constant k_B with respect to the average value of entropy. Given that E , S , \mathbf{L} and \mathbf{M} do not depend on k_B , if these effects are of low importance, we can take the limit $k_B \rightarrow 0$, resulting in

$$d\mathbf{z} = \mathbf{L}(\mathbf{z}) \frac{\partial E}{\partial \mathbf{z}} + \mathbf{M}(\mathbf{z}) \frac{\partial S}{\partial \mathbf{z}}. \quad (15)$$

This same assumption, $k_B \rightarrow 0$, induces two additional consequences (we omit the proof, the interested reader can consult [42])

$$\mathbf{L}(\mathbf{z}) \frac{\partial S}{\partial \mathbf{z}} = \mathbf{0}, \quad (16)$$

and

$$\mathbf{M}(\mathbf{z}) \frac{\partial E}{\partial \mathbf{z}} = \mathbf{0}, \quad (17)$$

which constitute the ingredients of the celebrated General Equation for the non-Equilibrium Reversible-Irreversible Coupling, GENERIC, equations [118–122]. This type of formulations are also known as metriplectic formulations, since they combine metric and symplectic terms [123, 124]. However, in GENERIC Eqs. (16) and (17), known as degeneracy conditions, play a fundamental role. They are key ingredients in the demonstration of the a priori satisfaction of the two laws of thermodynamics: conservation of energy in closed systems and non-negative entropy production. Indeed, given Eqs. (16) and (17), it is straightforward to prove that, given the anti-symmetry of \mathbf{L} ,

$$\dot{E}(z) = \frac{\partial E}{\partial z} \dot{z} = 0,$$

and

$$\dot{S} = \frac{\partial S}{\partial z} \dot{z} = \frac{\partial S}{\partial z} M(z) \frac{\partial S}{\partial z} \geq 0,$$

given the positive semi-definiteness of M . Therefore, the GENERIC structure appears to be much more interesting than port-Hamiltonian ones. It consistently guarantees the satisfaction of the laws of thermodynamics by construction. This makes GENERIC a very appealing choice for the construction of inductive biases in the learning of physical phenomena.

We assume that data sets \mathcal{D}_i contain labelled pairs of a single-step state vector z_t and its evolution in time z_{t+1} ,

$$\mathcal{D} = \{\mathcal{D}_i\}_{i=1}^{N_{\text{sim}}}, \quad \mathcal{D}_i = \{(z_t, z_{t+1})\}_{t=0}^T, \tag{18}$$

so that a neural network can be constructed by means of two loss terms, a data loss term that takes into account the correct prediction of the state vector time evolution using the GENERIC integrator, defined as

$$\mathcal{L}_n^{\text{data}} = \left\| \frac{dz^{\text{GT}}}{dt} - \frac{dz^{\text{net}}}{dt} \right\|_2^2,$$

where $\|\cdot\|_2$ denotes the L2-norm. The choice of the time derivative instead of the state vector itself is to regularize the global loss function to a uniform order of magnitude with respect to the degeneracy terms.

A second loss term takes into account the fulfillment of the degeneracy equations,

$$\mathcal{L}_n^{\text{deg}} = \left\| L \frac{\partial S}{\partial z_n} \right\|_2^2 + \left\| M \frac{\partial E}{\partial z_n} \right\|_2^2.$$

This formulation gave rise to the so-called structure-preserving neural networks [125] and thermodynamics-informed neural networks [126–128]. These networks have been employed recently in the development of physics perception with the help of computer vision techniques [129, 130].

The global loss term is a weighted mean of the two terms over the shuffled N_{batch} batched snapshots.

$$\mathcal{L} = \frac{1}{N_{\text{batch}}} \sum_{n=0}^{N_{\text{batch}}} (\lambda \mathcal{L}_n^{\text{data}} + \mathcal{L}_n^{\text{deg}}). \tag{19}$$

Note that the energy and entropy are learned through data about their gradients, so they are learnt up to an integration constant value. Note also that the activation functions must have a sufficient degree of continuity to allow for this.

Recently, alternative approaches have been developed to impose these degeneracy restrictions in hard form, instead

of soft form. For instance, [131] employ a particular parametrization of the L and M matrices. Zhang and coworkers [132] employ skew-symmetric matrices for forcing orthogonality. Notably, in this last work a universal approximation theorem is provided for this class of GENERIC-based networks, thus proving their expressivity.

An approach has been developed in which both inductive biases for the thermodynamic structure of the problem and a graph structure in the network are used in conjunction [128]. This approach has demonstrated to be very convenient for the development of learned simulators trained from finite element data. The result is a network for the problem of interest in which previously unseen geometric modifications can be introduced in the domain, as well as remeshings associated to these changes, without any decrease in accuracy.

Alternative versions based on GENERIC, but employing classical, piece-wise linear regressions instead of neural networks also exist, see [133–137]. More details on the geometric and thermodynamic structure of information can be found at [138].

Something remarkable from this GENERIC approach is that it works whenever the set of state variables z is able to adequately represent the energy of the system [42]. This opens the possibility of constructing reduced-order models and apply TINNs to learn the evolution equations in this reduced-order manifold. This is the approach followed in [126]. In Fig. 1 a sketch of this architecture is depicted. At a first instance, a sparse autoencoder is trained [139]. This autoencoder is trained with the full-field variables z . The autoencoder, that employs $L1$ -norm enforcing sparsity, then finds a latent representation of the phenomenon x . It is precisely in this latent space in which the method seeks for a GENERIC representation.

It is worth highlighting that the reduced-order modeling must follow this path: first reduce, then learn the GENERIC representation. The opposite (first learn the GENERIC model, then reduce) will make the ingredients of this formalism, L , M , E and S to loose their required properties.

4.2.3 Generalized GENERIC Structures

One of the most frequent criticisms to the employ of GENERIC as an inductive bias when learning physical phenomena is that metriplectic formalisms are only able to consider quadratic potentials. However, this is not entirely true. While the early descriptions of metriplectic formalisms and of GENERIC itself included only quadratic dissipation potentials, see [118, 119, 123, 124], this is no longer the case for modern descriptions of the GENERIC formalism [121, 140].

To develop such a theory, we first introduce the conjugate variables

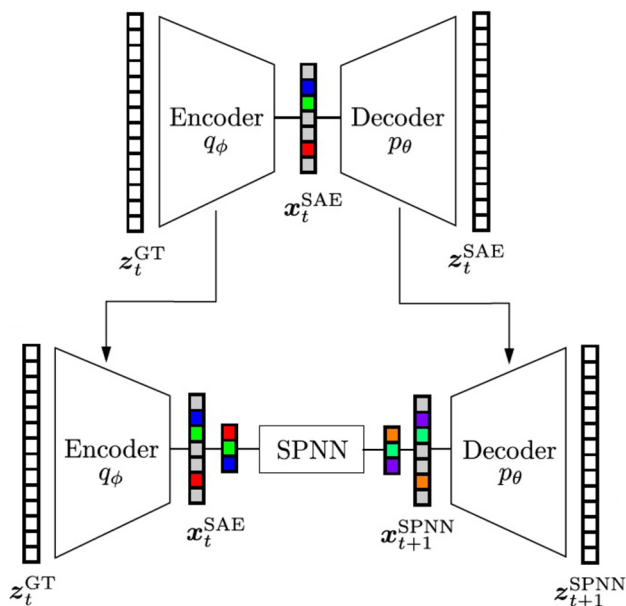


Fig. 1 Sketch of the architecture employed in [126] to learn reduced-order models of a given phenomenon

$$z^* = \frac{\partial S}{\partial z} = S_z.$$

We also introduce the so-called dissipative thermodynamic forces $\mathcal{X}(z^*)$. The third ingredient of the model is a dissipation potential

$$\Xi = \Xi(z, \mathcal{X}),$$

with the following properties:

1. Ξ is a real-valued and regular function of (z, \mathcal{X}) .
2. $\Xi(z, 0) = 0$.
3. $\Xi(z, \mathcal{X})$ reaches its minimum at $\mathcal{X} = \mathbf{0}$.
4. $\Xi(z, \mathcal{X})$ is convex in a neighborhood of $\mathcal{X} = \mathbf{0}$.
5. $z^* \frac{\partial \Xi}{\partial z^*} = K(\mathcal{X}, \frac{\partial \Xi}{\partial \mathcal{X}}) > 0$.

With these conditions, we have that

$$\dot{z} = L(z) \frac{\partial E}{\partial z} + \frac{\partial \Xi}{\partial z^*}. \tag{20}$$

Solutions to Eq. (20) satisfy the conservation of energy and non-negative entropy production,

$$\dot{S} = \frac{\partial S}{\partial z} \frac{\partial \Xi}{\partial z^*} = K\left(\mathcal{X}, \frac{\partial \Xi}{\partial \mathcal{X}}\right) > 0.$$

Equation (20) reduces to the standard GENERIC equation if $\mathcal{X}(z^*) = z^*$ and $\Xi(z, \mathcal{X}) = \frac{1}{2} z^* M z^*$, with M the usual dissipation matrix, symmetric and positive semi-definite.

Therefore, the elements in the GENERIC description of complex solids is composed by

1. the state variables z .
2. the kinematics expressed by the Poisson bracket $\{a, b\}$,
3. the dissipative forces \mathcal{X} and
4. three potentials, $E(z)$, $S(z)$ and $\Xi(z, \mathcal{X})$.

To the best of the author’s knowledge, there is no learning strategy developed on top of this last theory that, nevertheless, is worth exploring to extend the capabilities of metriplectic formalisms.

4.3 Strategies for Open Systems

The variational Onsager strategy arising from Eq. (9) presents one important advantage over strategies based upon metriplectic (GENERIC) formalisms: they are valid for externally-driven systems. On the contrary, GENERIC assumes by construction a closed system. Very recently, however, the metriplectic approach inherent in GENERIC has been extended to open systems in a port-Hamiltonian-like style [141]. This new formalism can be seen as a sort of port-metriplectic approach in which dissipative, open systems can be tackled within the metriplectic framework by simply adding ports to the formulation, through which energy is exchanged with the environment. This opens the possibility to learn systems of systems, i.e., systems composed by sub-system, possibly of different nature, while ensuring the fulfillment of the right thermodynamic structure of the problem. Under the umbrella of the port-metriplectic formalism, the evolution of the open system can be analyzed as

$$\begin{aligned} \dot{z} &= \{z, E\}_{\text{bulk}} + [z, S]_{\text{bulk}} \\ &= \{z, E\} + [z, S] - \{z, E\}_{\text{boun}} - [z, S]_{\text{boun}}, \end{aligned} \tag{21}$$

where $\{\cdot, \cdot\}$ is the Poisson bracket and $[\cdot, \cdot]$ represents the dissipative bracket that constitute the conservative and dissipative, respectively, terms of GENERIC. In fact, the reader can notice that both brackets are decomposed additively into bulk and boundary contributions. The degeneracy conditions, Eqs. (16) and (17) are satisfied by the bulk operators only.

5 Conclusions

We have explored the recent interest in developing learning strategies for physical phenomena that take into account inductive biases originated from the non-equilibrium thermodynamic theory. The growing interest in this field is motivated by the interest of the mechanics community on the

development of strategies that avoid as much as possible black-box approaches to the problem. For our community, but also for industrialists, credible methods are necessary. These days somewhat resemble the early days of finite elements. By that times, industrialists did not understand how finite elements worked. This motivated a huge effort of research within the applied maths and engineering communities that soon dissipated any doubt. If we want these data driven techniques to be adopted by the industry in a near future, we should provide responses, very much like the ones given for finite elements. This will make it necessary a joint effort from the applied maths community, from the engineering community but also from the physics and thermodynamics community, as we have tried to demonstrate in this paper.

In any case, so far these techniques have already demonstrated a substantial increase in robustness and credibility with respect to black-box approaches. It is also worth noting that something that we have already learnt is that the more physics knowledge we add to the learning process as an inductive bias, the less data will be necessary for the same level of accuracy and, conversely, the lower the error on the approximation will be.

In general, we strongly believe that these techniques, and those that will be developed in the future, will change the way we think of simulation, paving the way for instantaneous responses for parametric problems to the designers and analysts.

Acknowledgements This work has been partially funded by the Spanish Ministry of Science and Innovation, AEI/10.13039/501100011033, through Grant number PID2020-113463RB-C31 and the Regional Government of Aragon and the European Social Fund, group T24-20R. The authors also acknowledge the support of ESI Group through the project UZ-2019-0060.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Hey T, Tansley S, Tolle KM (2009) The fourth paradigm: data-intensive scientific discovery. Microsoft Research
- Bishop C (2022) Ai4science to empower the fifth paradigm of scientific discovery. <https://www.microsoft.com/en-us/research/blog/ai4science-to-empower-the-fifth-paradigm-of-scientific-discovery/>
- Nishimichi T, Takada M, Takahashi R, Osato K, Shirasaki M, Oogi T, Miyatake H, Oguri M, Murata R, Kobayashi Y et al (2019) Dark quest. I. Fast and accurate emulation of halo clustering statistics and its application to galaxy clustering. *Astrophys J* 884(1):29
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A et al (2021) Highly accurate protein structure prediction with alpha-fold. *Nature* 596(7873):583–589
- Kirchdoerfer T, Ortiz M (2016) Data-driven computational mechanics. *Comput Methods Appl Mech Eng* 304:81–101
- Conti S, Müller S, Ortiz M (2020) Data-driven finite elasticity. *Arch Ration Mech Anal* 237(1):1–33
- Conti S, Müller S, Ortiz M (2018) Data-driven problems in elasticity. *Arch Ration Mech Anal* 229(1):79–123
- Ibañez R, Borzacchiello D, Aguado JV, Abisset-Chavanne E, Cueto E, Ladevèze P, Chinesta F (2017) Data-driven non-linear elasticity: constitutive manifold construction and problem discretization. *Comput Mech* 60(5):813–826
- Ibanez R, Abisset-Chavanne E, Aguado JV, Gonzalez D, Cueto E, Chinesta F (2018) A manifold learning approach to data-driven computational elasticity and inelasticity. *Arch Comput Methods Eng* 25(1):47–57
- Carrara P, Ortiz M, De Lorenzis L (2022) Model-free fracture mechanics and fatigue. Current trends and open problems in computational mechanics. Springer, New York, pp 75–82
- Chinesta F, Cueto E (2022) Empowering engineering with data, machine learning and artificial intelligence: a short introductory review. *Adv Model Simul Appl Sci* 9:21
- Mahesh B (2020) Hine learning algorithms-a review. *Int J Sci Res (IJSR)* 9:381–386
- Mitchell T, Buchanan B, DeJong G, Dietterich T, Rosenbloom P, Waibel A (1990) Machine learning. *Annu Rev Comput Sci* 4(1):417–433
- Chinesta F, Cueto E, Abisset-Chavanne E, Duval JL, Khaldi FE (2020) Tual, digital and hybrid twins: a new paradigm in data-based engineering and engineered data. *Arch Comput Methods Eng* 27(1):105–134
- Sancarlos A, Cameron M, Abel A, Cueto E, Duval J-L, Chinesta F (2021) From rom of electrochemistry to AI-based battery digital and hybrid twin. *Arch Comput Methods Eng* 28(3):979–1015
- Moya B, Badías A, Alfaro I, Chinesta F, Cueto E (2020) Digital twins that learn and correct themselves. *Int J Numer Methods Eng* 123:3034
- Galilei G (1914) Two new sciences. Dover, Mineola
- Cosserrat E (1909) Théorie des corps déformables. Librairie Scientifique A, Hermann et Fils
- BBVA OpenMind (2019) The top 10 supercomputers, the new scientific giants. Web page
- Stachenfeld K, Fielding DB, Kochkov D, Cranmer M, Pfaff T, Godwin J, Cui C, Ho S, Battaglia P, Sanchez-Gonzalez A (2021) Learned simulators for turbulence. In: International conference on learning representations
- Allen KR, Lopez-Guevara T, Stachenfeld K, Sanchez-Gonzalez A, Battaglia P, Hamrick J, Pfaff T (2022) Physical design using differentiable learned simulators. <http://arxiv.org/abs/2202.00728>
- Klimesch J, Holl P, Thuerey N (2022) Simulating liquids with graph networks. <http://arxiv.org/abs/2203.07895>
- Um K, Brand R, Fei YR, Holl P, Thuerey N (2020) Solver-in-the-loop: learning from differentiable physics to interact with iterative pde-solvers. *Adv Neural Inf Process Syst* 33:6111–6122

24. Coros S, Macklin M, Thomaszewski B, Thürey N (2021) Differentiable simulation. In SIGGRAPH Asia 2021 Courses, pp 1–142
25. Heiden E, Millard D, Coumans E, Sheng Y, Sukhatme GS (2021) Neursim: augmenting differentiable simulators with neural networks. In: 2021 IEEE international conference on robotics and automation (ICRA), IEEE, pp 9474–9481
26. Schenck C, Fox D (2018) Spnets: differentiable fluid dynamics for deep neural networks. In: Conference on robot learning, PMLR, pp 317–335
27. Zhang GP (2006) Avoiding pitfalls in neural network research. *IEEE Trans Syst Man Cybern Part C* 37(1):3–16
28. Eitel F, Schulz M-A, Seiler M, Walter H, Ritter K (2021) Promises and pitfalls of deep neural networks in neuroimaging-based psychiatric research. *Exp Neurol* 339:113608
29. Battaglia PW, Hamrick JB, Bapst V, Sanchez-Gonzalez A, Zambaldi V, Malinowski M, Tacchetti A, Raposo D, Santoro A, Faulkner R, et al (2018) Relational inductive biases, deep learning, and graph networks. <http://arxiv.org/abs/1806.01261>
30. Raissi M, Perdikaris P, Karniadakis GE (2019) Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J Comput Phys* 378:686–707
31. Weinan E (2017) A proposal on machine learning via dynamical systems. *Commun Math Stat* 5(1):1–11
32. Thakolkaran P, Joshi A, Zheng Y, Flaschel M, De Lorenzis L, Kumar S (2022) Nn-euclid: deep-learning hyperelasticity without stress data. *J Mech Phys Solids* 169:105076
33. Flaschel M, Kumar S, De Lorenzis L (2023) Automated discovery of generalized standard material models with euclid. *Comput Methods Appl Mech Eng* 405:115867
34. Marino E, Flaschel M, Kumar S, De Lorenzis L (2023) Automated identification of linear viscoelastic constitutive laws with euclid. *Mech Mater* 181:104643
35. Español P (2004) Statistical mechanics of coarse-graining. Springer, Berlin Heidelberg, pp 69–115
36. Turkington B, Chen Q-Y, Thalabard S (2016) Coarse-graining two-dimensional turbulence via dynamical optimization. *Nonlinearity* 29(10):2961
37. Ma C, Wang J, Weinan E (2018) Model reduction with memory and the machine learning of dynamical systems
38. González D, Chinesta F, Cueto E (2021) Learning non-Markovian physics from data. *J Comput Phys* 428:109982
39. Weber J (1956) Fluctuation dissipation theorem. *Phys Rev* 101(6):1620
40. Kubo R (1966) The fluctuation-dissipation theorem. *Rep Prog Phys* 29(1):255
41. Landau LD, Lifshitz EM, Beyer RT et al (1992) Hydrodynamic fluctuations. Perspectives in theoretical physics. Elsevier, Amsterdam, pp 359–361
42. Espanol P, Warren P (1995) Statistical mechanics of dissipative particle dynamics. *EPL (Europhys Lett)* 30(4):191
43. Warren PB (1998) Dissipative particle dynamics. *Curr Opin Colloid Interface Sci* 6(3):620–624
44. Espanol P, Warren PB (2017) Perspective: dissipative particle dynamics. *J Chem Phys* 146(15):150901
45. Keunings R (2000) Advances in the computer modeling of the flow of polymeric liquids. *Comput Fluid Dyn J* 9(1):449–458
46. Keunings R (2004) Micro-macro methods for the multiscale simulation of viscoelastic flow using molecular models of kinetic theory. *Rheol Rev* 67–98:2004
47. Owens RG, Phillips TN (2002) Computational rheology. World Scientific, Singapore
48. Wilemski G (1976) On the derivation of smoluchowski equations with corrections in the classical theory of brownian motion. *J Stat Phys* 14(2):153–169
49. Mao Z, Jagtap AD, Karniadakis GE (2020) Physics-informed neural networks for high-speed flows. *Comput Methods Appl Mech Eng* 360:112789
50. Pang G, Lu L, Karniadakis GE (2019) fpinns: fractional physics-informed neural networks. *SIAM J Sci Comput* 41(4):A2603–A2626
51. Misyris GS, Venzke A, Chatzivasileiadis S (2020) Physics-informed neural networks for power systems. In: 2020 IEEE power & energy society general meeting (PESGM), IEEE, pp 1–5
52. Jagtap AD, Kawaguchi K, Karniadakis GE (2020) Adaptive activation functions accelerate convergence in deep and physics-informed neural networks. *J Comput Phys* 404:109136
53. Cai S, Wang Z, Wang S, Perdikaris P, Karniadakis GE (2021) Physics-informed neural networks for heat transfer problems. *J Heat Transf* 143(6):060801
54. Krishnapriyan A, Gholami A, Zhe S, Kirby R, Mahoney MW (2021) Characterizing possible failure modes in physics-informed neural networks. *Adv Neural Inf Process Syst* 34:26548–26560
55. Cai S, Mao Z, Wang Z, Yin M, Karniadakis GE (2022) Physics-informed neural networks (PINNs) for fluid mechanics: a review. *Acta Mech Sin* 37:1–12
56. Cuomo S, Di Cola VS, Giampaolo F, Rozza G, Raissi M, Piccialli F (2022) Scientific machine learning through physics-informed neural networks: where we are and what's next. <http://arxiv.org/abs/2201.05624>
57. Dirac PAM (1929) Quantum mechanics of many-electron systems. *Proc R Soc Lond Ser A* 123(792):714–733
58. Noether AE (1918) Nachr kgl ges wiss göttingen. *Math Phys KI II*:235
59. Noether E (1971) Invariant variation problems. *Transp Theory Stat Phys* 1(3):186–207
60. Jin P, Zhang Z, Zhu A, Tang Y, Karniadakis GE (2020) Symnets: intrinsic structure-preserving symplectic networks for identifying Hamiltonian systems. *Neural Netw* 132:166–179
61. Masi F, Stefanou I, Vannucci P, Maffi-Berthier V (2021) Thermodynamics-based artificial neural networks for constitutive modeling. *J Mech Phys Solids* 147:104277
62. Masi F, Stefanou I, Vannucci P, Maffi-Berthier V (2020) Material modeling via thermodynamics-based artificial neural networks. In: Workshop on joint structures and common foundations of statistical physics, information geometry and inference for learning. Springer, pp 308–329
63. Masi F, Stefanou I (2021) Thermodynamics-based artificial neural networks (tann) for multiscale modeling of materials with inelastic microstructure. <http://arxiv.org/abs/2108.13137>
64. Masi F, Stefanou I (2022) Multiscale modeling of inelastic materials with thermodynamics-based artificial neural networks (tann). *Comput Methods Appl Mech Eng* 398:115190
65. Patel RG, Manickam I, Trask NA, Wood MA, Lee M, Tomas I, Cyr EC (2022) Thermodynamically consistent physics-informed neural networks for hyperbolic systems. *J Comput Phys* 449:110754
66. Trask N, Huang A, Xiaozhe H (2022) Enforcing exact physics in scientific machine learning: a data-driven exterior calculus on graphs. *J Comput Phys* 456:110969
67. Onsager L (1931) Reciprocal relations in irreversible processes I. *Phys Rev* 37(4):405
68. Arroyo M, Walani N, Torres-Sanchez A, Kaurin D (2018) Onsager's variational principle in soft matter: introduction and application to the dynamics of adsorption of proteins onto fluid membranes. The role of mechanics in the study of lipid bilayers. pp 287–332
69. Huang S, He Z, Reina C (2022) Variational onsager neural networks (vonns): a thermodynamics-based variational

- learning strategy for non-equilibrium pdes. *J Mech Phys Solids* 163:104856
70. Amos B, Xu L, Kolter JZ (2017) Input convex neural networks. In: *International conference on machine learning*. PMLR, pp 146–155
 71. Barbaresco F, Nielsen F et al (2021) *Geometric structures of statistical physics, information geometry, and learning*. Springer, New York
 72. Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B (1998) Support vector machines. *IEEE Intell Syst Appl* 13(4):18–28
 73. Cybenko G (1989) Approximation by superpositions of a sigmoidal function. *Math Control Signals Syst* 2(4):303–314
 74. Hornik K, Stinchcombe M, White H (1989) Multilayer feed-forward networks are universal approximators. *Neural Netw* 2(5):359–366
 75. Scarselli F, Tsoi AC (1998) Universal approximation using feedforward neural networks: a survey of some existing methods, and some new results. *Neural Netw* 11(1):15–37
 76. Hohenberg PC, Halperin BI (1977) Theory of dynamic critical phenomena. *Rev Mod Phys* 49(3):435
 77. Weinan E (2020) Machine learning and computational mathematics. *Commun Comput Phys* 28(5):1639–1670
 78. Greydanus S, Dzamba M, Yosinski J (2019) Hamiltonian neural networks. In: *Advances in neural information processing systems*, vol 32
 79. Mattheakis M, Sondak D, Dogra AS, Protopapas P (2022) Hamiltonian neural networks for solving equations of motion. *Phys Rev E* 105:065305
 80. Han C-D, Glaz B, Haile M, Lai Y-C (2021) Adaptable Hamiltonian neural networks. *Phys Rev Res* 3(2):023156
 81. Bertalan T, Dietrich F, Mezić I, Kevrekidis IG (2019) On learning Hamiltonian systems from data. *Chaos Interdiscip J Nonlinear Sci* 29(12):121107
 82. De Wilde P (1993) Class of Hamiltonian neural networks. *Phys Rev E* 47(2):1392
 83. David M, Méhats F (2021) Symplectic learning for Hamiltonian neural networks. <http://arxiv.org/abs/2106.11753>
 84. Finzi M, Wang KA, Wilson AG (2020) Simplifying Hamiltonian and Lagrangian neural networks via explicit constraints. *Adv Neural Inf Process Syst* 33:13880–13889
 85. Tong Y, Xiong S, He X, Pan G, Zhu B (2021) Symplectic neural networks in Taylor series form for Hamiltonian systems. *J Comput Phys* 437:110325
 86. Chen Z, Zhang J, Arjovsky M, Bottou L (2019) Symplectic recurrent neural networks. <http://arxiv.org/abs/1909.13334>
 87. Toth P, Rezende DJ, Jaegle A, Racaniere S, Botev A, Higgins I (2019) Hamiltonian generative networks. <http://arxiv.org/abs/1909.13789>
 88. DiPietro D, Xiong S, Zhu B (2020) Sparse symplectically integrated neural networks. *Adv Neural Inf Process Syst* 33:6074–6085
 89. Choudhary A, Lindner JF, Holliday EG, Miller ST, Sinha S, Ditto WL (2020) Physics-enhanced neural networks learn order and chaos. *Phys Rev E* 101(6):062207
 90. Miller ST, Lindner JF, Choudhary A, Sinha S, Ditto WL (2020) Mastering high-dimensional dynamics with Hamiltonian neural networks. <http://arxiv.org/abs/2008.04214>
 91. Galimberti CL, Xu L, Trecate GF (2021) A unified framework for Hamiltonian deep neural networks. In: *Learning for dynamics and control*. PMLR, pp 275–286
 92. Hochreiter S (1998) The vanishing gradient problem during learning recurrent neural nets and problem solutions. *Int J Uncertain Fuzziness Knowl-Based Syst* 6(02):107–116
 93. Galimberti CL, Furieri L, Xu L, Ferrari-Trecate G (2021) Hamiltonian deep neural networks guaranteeing non-vanishing gradients by design. <http://arxiv.org/abs/2105.13205>
 94. Chen Z, Feng M, Yan J, Zha H (2022) Learning neural Hamiltonian dynamics: a methodological overview. <http://arxiv.org/abs/2203.00128>
 95. Cranmer M, Greydanus S, Hoyer S, Battaglia P, Spergel D, Ho S (2020) Lagrangian neural networks. <http://arxiv.org/abs/2003.04630>
 96. Roehrl MA, Runkler TA, Brandtstetter V, Tokic M, Obermayer S (2020) Modeling system dynamics with physics-informed neural networks based on Lagrangian mechanics. *IFAC-PapersOnLine* 53(2):9195–9200
 97. Allen-Blanchette C, Veer S, Majumdar A, Leonard NE (2020) LagNetvip: a Lagrangian neural network for video prediction. <http://arxiv.org/abs/2010.12932>
 98. Lee S-C, Kim Y-H (2002) An enhanced Lagrangian neural network for the eld problems with piecewise quadratic cost functions and nonlinear constraints. *Electr Power Syst Res* 60(3):167–177
 99. Zhong YD, Leonard N (2020) Unsupervised learning of Lagrangian dynamics from images for prediction and control. *Adv Neural Inf Process Syst* 33:10741–10752
 100. Lutter M, Ritter C, Peters J (2019) Deep Lagrangian networks: using physics as model prior for deep learning. <http://arxiv.org/abs/1907.04490>
 101. Bhattoo R, Ranu S, Krishnan NM (2021) Lagrangian neural network with differentiable symmetries and relational inductive bias. <http://arxiv.org/abs/2110.03266>
 102. Zhong YD, Dey B, Chakraborty A (2020) Dissipative symoden: encoding Hamiltonian dynamics with dissipation and control into deep learning. <http://arxiv.org/abs/2002.08860>
 103. Zhong YD, Dey B, Chakraborty A (2021) Benchmarking energy-conserving neural networks for learning dynamics from data. In: *Learning for dynamics and control*. PMLR, pp 1218–1229
 104. Gruver N, Finzi M, Stanton S, Wilson AG (2022) Deconstructing the inductive biases of Hamiltonian neural networks. <http://arxiv.org/abs/2202.04836>
 105. Van Der Schaft A, Jeltsema D et al (2014) An introductory overview. *Found Trends Syst Control* 1(2–3):173–378
 106. Rashad R, Califano F, van der Schaft AJ, Stramigioli S (2020) Twenty years of distributed port-Hamiltonian systems: a literature review. *IMA J Math Control Inf* 37(4):1400–1422
 107. Beattie CA, Mehrmann V, Van Dooren P (2019) Robust port-Hamiltonian representations of passive systems. *Automatica* 100:182–186
 108. Massaroli S, Poli M, Califano F, Faragasso A, Park J, Yamashita A, Asama H (2019) Port-Hamiltonian approach to neural network training. In: *2019 IEEE 58th conference on decision and control (CDC)*. IEEE, pp 6799–6806
 109. Cherifi K (2020) An overview on recent machine learning techniques for port Hamiltonian systems. *Physica D* 411:132620
 110. Desai SA, Mattheakis M, Sondak D, Protopapas P, Roberts SJ (2021) Port-Hamiltonian neural networks for learning explicit time-dependent dynamical systems. *Phys Rev E* 104(3):034312
 111. Poli M, Massaroli S, Yamashita A, Asama H, Park J (2020) Port-Hamiltonian gradient flows. In: *ICLR 2020 workshop on integration of deep neural models and differential equations*
 112. Furieri L, Galimberti CL, Zakwan M, Ferrari-Trecate G (2022) Distributed neural network control with dependability guarantees: a compositional port-Hamiltonian approach. In: *Learning for dynamics and control conference*. PMLR, pp 571–583
 113. Eidnes S, Stasik AJ, Sterud C, Bohn E, Riemer-Sørensen S (2022) Port-Hamiltonian neural networks with state dependent ports. <http://arxiv.org/abs/2206.02660>
 114. Wang R, Walters R, Yu R (2022) Approximately equivariant networks for imperfectly symmetric dynamics. <http://arxiv.org/abs/2201.11969>

115. Cohen T, Welling M (2016) Group equivariant convolutional networks. In: International conference on machine learning. PMLR, pp 2990–2999
116. Satorras VG, Hoogeboom E, Welling M (2021) E (n) equivariant graph neural networks. In: International conference on machine learning. PMLR, pp 9323–9332
117. Keriven N, Peyré G (2019) Universal invariant and equivariant graph neural networks. In: Advances in neural information processing systems, vol 32
118. Öttinger HC, Grmela M (1997) Dynamics and thermodynamics of complex fluids. II. Illustrations of a general formalism. *Phys Rev E* 56(6):6633
119. Öttinger HC (2005) Beyond equilibrium thermodynamics. Wiley, Hoboken
120. Grmela M (2018) Generic guide to the multiscale dynamics and thermodynamics. *J Phys Commun* 2(3):032001
121. Grmela M, Klika V, Pavelka M (2019) Gradient and generic evolution towards reduced dynamics. *Philos Trans R soc A* 378:20190472
122. Pavelka M, Klika V, Grmela M (2018) Multiscale thermodynamics. In: Multiscale thermo-dynamics. de Gruyter
123. Morrison PJ (1984) Bracket formulation for irreversible classical fields. *Phys Lett A* 100(8):423–427
124. Morrison PJ (1986) A paradigm for joined Hamiltonian and dissipative systems. *Physica D* 18(1–3):410–419
125. Hernández Q, Badiás A, González D, Chinesta F, Cueto E (2021) Structure-preserving neural networks. *J Comput Phys* 426:109950
126. Hernandez Q, Badiás A, Gonzalez D, Chinesta F, Cueto E (2021) Deep learning of thermodynamics-aware reduced-order models from data. *Comput Methods Appl Mech Eng* 379:113763
127. Chinesta F, Cueto E, Grmela M, Moya B, Pavelka M, Šípka M (2020) Learning physics from data: a thermodynamic interpretation. In: Workshop on joint structures and common foundations of statistical physics, information geometry and inference for learning. Springer, pp 276–297
128. Hernández Q, Badiás A, Chinesta F, Cueto F (2022) Thermodynamics-informed graph neural networks. <http://arxiv.org/abs/2203.01874>
129. Moya B, Badiás A, Gonzalez D, Chinesta F, Cueto E (2021) Physics perception in sloshing scenes with guaranteed thermodynamic consistency. <http://arxiv.org/abs/2106.13301>
130. Moya B, Badiás A, Gonzalez D, Chinesta F, Cueto E (2022) Physics-informed reinforcement learning for perception and reasoning about fluids. <http://arxiv.org/abs/2203.05775>
131. Lee K, Trask N, Stinis P (2021) Machine learning structure preserving brackets for forecasting irreversible processes. *Adv Neural Inf Process Syst* 34:5696–5707
132. Zhang Z, Shin Y, Em Karniadakis G (2022) Gfnnns: generic formalism informed neural networks for deterministic and stochastic dynamical systems. *Philos Trans R Soc A* 380(2229):20210207
133. González D, Chinesta F, Cueto E (2019) Thermodynamically consistent data-driven computational mechanics. *Continuum Mech Thermodyn* 31(1):239–253
134. González D, Chinesta F, Cueto E (2019) Learning corrections for hyperelastic models from data. *Front Mater* 6:14
135. Moya B, González D, Alfaro I, Chinesta F, Cueto E (2019) Learning slosh dynamics by means of data. *Comput Mech* 64(2):511–523
136. González D, García-González A, Chinesta F, Cueto E (2020) A data-driven learning method for constitutive modeling: application to vascular hyperelastic soft tissues. *Materials* 13(10):2319
137. Moya B, Alfaro I, Gonzalez D, Chinesta F, Cueto E (2020) Physically sound, self-learning digital twins for sloshing fluids. *PLoS ONE* 15(6):e0234569
138. Barbaresco F (2020) Souriau-casimir lie groups thermodynamics and machine learning. In: Workshop on joint structures and common foundations of statistical physics, information geometry and inference for learning. Springer, pp 53–83
139. Ng A et al (2011) Sparse autoencoder. *CS294A Lecture Notes* 72:1–19
140. Grmela M (2003) A framework for elasto-plastic hydrodynamics. *Phys Lett A* 312(3–4):136–146
141. Hernández Q, Badiás A, Chinesta F, Cueto E (2023) Port-metric neural networks: thermodynamics-informed machine learning of complex physical systems. *Comput Mech* 2023:1–9

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.