Machine Learning Hidden Symmetries

Ziming Liu and Max Tegmark

Phys. Rev. Lett. **128**, 180201 — Published 6 May 2022

DOI: 10.1103/PhysRevLett.128.180201

Machine-learning hidden symmetries

Ziming Liu and Max Tegmark
Department of Physics, Massachusetts Institute of Technology, Cambridge, USA
(Dated: February 11, 2022)

We present an automated method for finding hidden symmetries, defined as symmetries that become manifest only in a new coordinate system that must be discovered. Its core idea is to quantify asymmetry as violation of certain partial differential equations, and to numerically minimize such violation over the space of all invertible transformations, parametrized as invertible neural networks. For example, our method rediscovers the famous Gullstrand-Painlevé metric that manifests hidden translational symmetry in the Schwarzschild metric of non-rotating black holes, as well as Hamiltonicity, modularity and other simplifying traits not traditionally viewed as symmetries.

INTRODUCTION

Philip Anderson famously argued that "It is only slightly overstating the case to say that physics is the study of symmetry" [1], and discovering symmetries has proven enormously useful both for deepening understanding and for solving problems more efficiently, in physics [1–3] as well as machine learning [4–10].

Discovering symmetries is useful but hard, because they are often not *manifest* but *hidden*, becoming manifest only after an appropriate coordinate transformation. For example, after Schwarzschild discovered his eponymous black hole metric, it took 17 years until Painlevé, Gullstrand and Lemaître showed that it had hidden translational symmetry: they found that the spatial sections could be made translationally invariant with a clever coordinate transformation, thereby deepening our understanding of black holes [11]. As a simpler example, Fig. 1 shows the same vector field in two coordinates systems where rotational symmetry is manifest and hidden, respectively.

Our results below are broadly applicable because they apply to a very broad definition of symmetry, including not only *invariance* and *equivariance* with respect to arbitrary Lie groups, but also *modularity* and *Hamiltonicity*. If a coordinate transformation is discovered that makes such simplifying properties manifest, this can not only deepen our understanding of the system in question, but also enable an arsenal of more efficient numerical methods for studying it.

Discovering hidden symmetries is unfortunately highly non-trivial, because it involves a search over all smooth invertible coordinate transformations, and has traditionally been accomplished by scientists making inspired guesses. The goal of this *Letter* is to present a machine learning method for automating hidden symmetry discovery. Its core idea is to quantify asymmetry as violation of certain partial differential equations, and to numerically minimize such violation over the space of all invertible transformations, parametrized as invertible neural networks. For example, the neural network automatically learns to transform Fig. 1(b) into Fig. 1(c), thereby making the hidden rotational symmetry manifest. Our method differs from previous work [9, 10, 12–15] that exploits manifest symmetries, partial differential equations or other physical properties to facilitate machine learning, but not the other way around to discover hidden symmetries with machine learning as a tool.

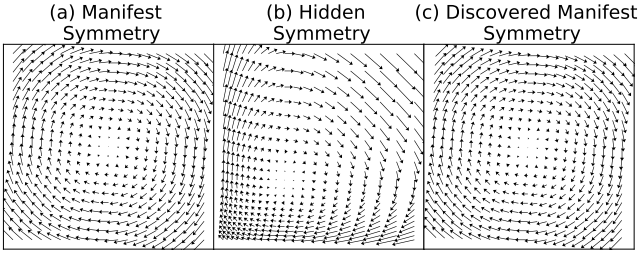

FIG. 1: 1D harmonic oscillator phase space flow vector field $\mathbf{f}(x,p)=(p,-x)$. The rotational symmetry of $\mathbf{f}$ is manifest in (a) and hidden in (b). Our algorithm can reveal the hidden symmetry by auto-discovering the transformation from (b) to (c).

TABLE I: PDE and Losses for Generalized Symmetries

Generalized symmetry	Linear operator $\hat{L}$	Loss ℓ	Examples
Translation invariance	$\hat{L}_j=\partial_j$	$\ell_{\rm TI}$	A,E,F
Lie invariance	$\hat{L}_j=K_j\mathbf{z}\cdot\nabla$	$\ell_{\rm INV}$	E,F
Lie equivariance	$\hat{L}_j=K_j\mathbf{z}\cdot\nabla\pm K_j$	$\ell_{\rm EQV}$	B
Canonical eqvariance	$\hat{L}_j^{\mathbf{x}}=K_j\mathbf{x}\cdot\nabla_{\mathbf{x}}-K_j^t\mathbf{p}\cdot\nabla_{\mathbf{p}}+K_j^t$ $\hat{L}_j^{\mathbf{p}}=K_j\mathbf{x}\cdot\nabla_{\mathbf{x}}-K_j^t\mathbf{p}\cdot\nabla_{\mathbf{p}}-K_j$	$\ell_{\rm CAN}$	C
Hamiltonicity	$\hat{L}_{ij}=-\mathbf{m}_i^t\partial_j+\mathbf{m}_j^t\partial_i$	$\ell_{\rm H}$	A,B,C,D
Modularity	$\hat{L}_{ij}=\mathbf{A}_{ij}\hat{\mathbf{z}}_i^t\partial_j$	$\ell_{\rm M}$	D

In the Method section, we introduce our notation and symmetry definition and present our method for hidden symmetry discovery. In the Results section, we apply our method to classical mechanics and general relativity examples to test its ability to auto-discover hidden symmetries.

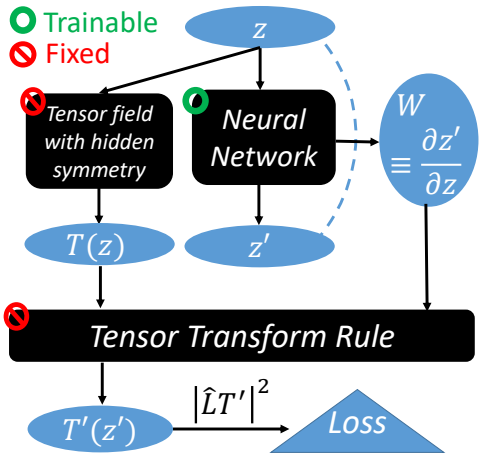

FIG. 2: Schematic workflow of our algorithm for discovering hidden symmetry

METHOD

PDEs encoding generalized symmetries

We seek to discover symmetries in various tensor fields $T(\mathbf{z})$ for $\mathbf{z} \in \mathbb{R}^n$, for example the vector field $\mathbf{f}(\mathbf{z})$ (a rank-1 tensor) defining a dynamical system $\mathbf{z}(t)$ through a vector differential equation $\dot{\mathbf{z}} = \mathbf{f}(\mathbf{z})$, or the metric $g(\mathbf{z})$ (a rank-2 tensor) quantifying spacetime geometry in general relativity. We say that a tensor field T has a *generalized symmetry* if it obeys a linear partial differential equation (PDE) $\hat{L}T = 0$, where $\hat{L}$ is a linear operator that encodes the symmetry generators. This definition covers a broad range of interesting situations, as illustrated by the examples below (see Table I for a summary).

Translational Invariance: A tensor field T is invariant under translation in the $\hat{j^{\text{th}}}$ coordinate direction $\hat{\mathbf{z}}_j$ if $T(\mathbf{z} + a\hat{\mathbf{z}}_i) = T(\mathbf{z})$ for all $\mathbf{z}$ and a, which is equivalent to satisfying the PDE $\partial T/\partial z_j = 0$, corresponding to the linear operator $\hat{L} = \partial_j$.

Lie invariance & equivariance: If $T(\mathbf{z})$ satisfies $T(g\mathbf{z}) = g^n T(\mathbf{z})$ for all elements g of some Lie group $\mathcal{G}$ and an integer $n = -1$, 0 or 1, then we say that T is *invariant* if $n = 0$, and *equivariant* otherwise ($n = 1$ corresponds to a *covariant* (1,0) vector field and $n = -1$ corresponds to a *contravariant* (0,1) vector field) [1]. Taking the derivative on the both sides of the identity $T(e^{K_j a}\mathbf{z}) = e^{nK_j a}T(\mathbf{z})$ with respect to a at $a = 0$ gives the PDEs $\hat{L}_j\,\mathbf{f} = 0$ with $\hat{L}_j \equiv K_j\mathbf{z}\cdot\nabla - nK_j$. Figure 1 (a) and (c) show examples of rotational equivariance.

[1] An (m,n)-tensor means the tensor has m covariant and n contravariant indices, which is a convenient notations especially in general relativity when dealing with metrics. We call a (1,0)-tensor a covariant vector, and a (0,1)-tensor a contravariant vector.

Hamiltonicity (a.k.a. *symplecticity*): A dynamical system $\mathbf{z}(t) \in \mathbb{R}^{2d}$ obeying a vector differential equation $\dot{\mathbf{z}} = \mathbf{f}(\mathbf{z})$ is called *Hamiltonian* or *symplectic* if $\mathbf{f} = \mathbf{M}\nabla H$ for a scalar function H, where

$$\mathbf{M} \equiv \begin{pmatrix} 0 & \mathbf{I} \\ -\mathbf{I} & 0 \end{pmatrix}, \tag{1}$$

and $\mathbf{I}$ is the $d \times d$ identity matrix. Such systems are of great importance in physics, where it is customary to write $\mathbf{z} = (\mathbf{x}, \mathbf{p})$, because the Hamiltonian function $H(\mathbf{z})$ (interpreted as energy) is a conserved quantity under the system evolution $\dot{\mathbf{z}} = \mathbf{f}(\mathbf{z}) = (\dot{\mathbf{x}}, \dot{\mathbf{p}}) = \mathbf{M}\nabla H = (\partial_{\mathbf{p}} H, -\partial_{\mathbf{x}} H)$. Hamiltonicity thus corresponds to $\mathbf{M}^{-1}\mathbf{f}$ being a gradient, *i.e.*, to its generalized curl (the antisymmetric parts of its Jacobian matrix) vanishing. Letting $\mathbf{J} \equiv \nabla\mathbf{f}$ denote the Jacobian of $\mathbf{f}$ ($J_{ij} \equiv f_{i,j}$) and using the fact that $\mathbf{M}^{-1} = \mathbf{M}^t = -\mathbf{M}$ (superscript t denotes transpose), Hamiltonicity is thus equivalent to satisfying the PDEs $\hat{L}_{ij}\mathbf{f} = 0$ where $\hat{L}_{ij}\mathbf{f} = (\mathbf{MJ} + \mathbf{J}^t\mathbf{M})_{ij}$ for all i and j ($\frac{n(n-1)}{2}$ independent ODE's in all), corresponding to $\hat{L}_{ij} = -\mathbf{m}_i^t\partial_j + \mathbf{m}_j^t\partial_i$, where $\mathbf{m}_i$ are the column vectors of $\mathbf{M}$. In other words, although Hamiltonicity is not traditionally thought of as a symmetry, it conveniently meets our generalized symmetry definition and can thus be auto-discovered with our method.

Canonical equivariance: We define a Hamiltonian system as canonically equivariant if $\mathbf{z} = (\mathbf{x}, \mathbf{p})$ and the vector field $\mathbf{f} \equiv (\mathbf{f}_{\mathbf{x}}, \mathbf{f}_{\mathbf{p}})$ satisfies $\mathbf{f}_{\mathbf{x}}(g\mathbf{x}, g^{-1}\mathbf{p}) = g^{-t}\mathbf{f}_{\mathbf{x}}$ and $\mathbf{f}_{\mathbf{p}}(g\mathbf{x}, g^{-1}\mathbf{p}) = g\mathbf{f}_{\mathbf{p}}$ for all $g \in \mathcal{G}$. These two equations are equivalent to the PDEs $\hat{L}_j^{\mathbf{x}}\mathbf{f}_{\mathbf{x}} = 0$ and $\hat{L}_j^{\mathbf{p}}\mathbf{f}_{\mathbf{p}} = 0$ with $\hat{L}_j^{\mathbf{x}} = K_j\mathbf{x}\cdot\nabla_{\mathbf{x}} - K_j^t\mathbf{p}\cdot\nabla_{\mathbf{p}} + K_j^t$ and $\hat{L}_j^{\mathbf{p}} = K_j\mathbf{x}\cdot\nabla_{\mathbf{x}} - K_j^t\mathbf{p}\cdot\nabla_{\mathbf{p}} - K_j$. In special cases when the generator K_j is anti-symmetric (*e.g.*, for the rotation group), $\hat{L}_j^{\mathbf{x}} = \hat{L}_j^{\mathbf{p}}$.

Modularity: A dynamical system $\mathbf{z}(t)$ obeying $\dot{\mathbf{z}} = f(\mathbf{z})$ is *modular* if the Jacobian $\mathbf{J} = \nabla\mathbf{f}$ is block-diagonal, which implies that the components of $\mathbf{z}$ corresponding to different blocks evolve independently of each other. More generally, we say that a system is $(n_1 + \cdots + n_k)$-modular if $\mathbf{J}$ vanishes except for blocks of size $n_1, \ldots n_k$, which we can write as $\mathbf{A} \circ \mathbf{J} = 0$ where $\circ$ denotes element-wise multiplication ($([\mathbf{A} \circ \mathbf{J}])_{ij} = \mathbf{A}_{ij}\mathbf{J}_{ij}$) and the elements of the mask matrix $\mathbf{A}$ equal 1 inside the blocks, vanishing otherwise. Although modularity is typically not viewed as a symmetry, it too thus meets our generalized symmetry definition and can be auto-discovered with our method using the matrix PDE $\mathbf{A} \circ \nabla\mathbf{f} = 0$, corresponding to the linear operators $\hat{L}_{ij} \equiv \mathbf{A}_{ij}\hat{\mathbf{z}}_i^t\partial_j$.

Our algorithm for discovering hidden symmetries

We now describe our algorithm of discovering hidden symmetries. Since $\hat{L}T = 0$ implies manifest symmetry, $|\hat{L}T|^2$ is a natural measure of manifest symmetry viola-

tion. We therefore define the *symmetry loss* as

$$\ell \equiv \frac{\langle|\hat{L}T(\mathbf{z})|^2\rangle}{\langle|\mathbf{z}|^2\rangle^\alpha}, \tag{2}$$

where angle brackets denote averaging over some set of points $\mathbf{z}_i$, and α is chosen so that ℓ is scale-invariant *i.e.*, invariant under a scale transformation $\mathbf{z} \to a\mathbf{z}$, $T \to a^{m-n}T$, $\hat{L} \to a^s\hat{L}$, $\ell \to a^{2(m-n+s-\alpha)}$ if T has m contravariate indices and n covariate indices. Hence we choose $\alpha = m - n + s$. We jointly search for multiple hidden symmetries by using the loss function $\ell = \sum_i \ell_i$ where each i corresponds to one symmetry, denoted by subscripts as in Tab. I.

Discovering hidden symmetry is equivalent to minimizing ℓ over all diffeomorphisms (everywhere differentiable and invertible coordinate transformations), which we parametrize with an invertible neural network. Figure 2 shows the workflow of our algorithm: (1) a neural network transforms $\mathbf{z} \mapsto \mathbf{z}'$ and obtains the transformation's Jacobian $\mathbf{W} \equiv d\mathbf{z}'/d\mathbf{z}$; (2) in parallel with (1), we evaluate the known tensor field T at $\mathbf{z}$; (3) we jointly feed $\mathbf{W}(\mathbf{z})$ and $T(\mathbf{z})$ into a module which implements the tensor transformation rule and gives $T'(\mathbf{z}')$; (4) we compute the symmetry loss of $\ell(T')$. Note that only the neural network is trainable, while both the tensor field with hidden symmetry $T(\mathbf{z})$ and tensor transformation rule are hard-coded in the workflow. We update the neural network with back-propagation to find the coordinate transformation $\mathbf{z} \mapsto \mathbf{z}'$ that minimizes ℓ. If the resulting ℓ is effectively zero, a hidden symmetry has been discovered.

Neural network training and symbolic regression

We parametrize the coordinate transformation $\mathbf{z} \mapsto \mathbf{z}'$ as $\mathbf{z}' = \mathbf{z} + \mathbf{f}_{\rm NN}(\mathbf{z})$, where $\mathbf{f}_{\rm NN}$ is a fully connected neural network with two hidden layers containing 400 neurons each. We use a silu activation function [16] rather than the popular ReLU alternative, because our method requires activation functions to be twice differentiable (since the loss function involves first derivatives of output with respect to input via the Jacobian $\mathbf{W}$). Derivatives of PDE losses and Jacobians are calculated with automatic differentiation and backpropagation. The invertibility of the mapping $\mathbf{z} \to \mathbf{z}'$ is guaranteed by the fact that $\det \mathbf{W} \to 0$ and the loss function $\ell \to \infty$ if $\mathbf{z} \to \mathbf{z}'$ approaches non-invertibility, as seen in equations (**??**)-(**??**) in the supplementary material. The supplementary material also provides further technical details on the selection of data points $\mathbf{z}$, neural network initialization and training. If multiple symmetries are tested, the training process is performed in multiple stages: at each stage, we add one more symmetry to the loss function and re-train to convergence.

We then apply AI Feynman, a physics-inspired symbolic regression module, to interpret what the neural network has learned; for details, see Appendix **??** and [17, 18].

RESULTS

We will now test our algorithm on 6 physics examples, ranging from classical mechanics to general relativity. Table II summarizes these examples, labeled A, B,...,F for easy reference, listing their manifestly non-symmetric equations, their simplifying coordinate transformations, their transformed and manifestly symmetric equations, and their discovered hidden symmetries. As we will see, all the symmetries we had hidden in our test examples were rediscovered by our algorithm. The only example is the transformation for A, where the problem is so simple that an infinite family of transformations give equal symmetry.

Warmup examples

To build intuition for how our method works, we first apply it to the simple warmup examples A, B and C, corresponding to systems involving free motion, harmonic oscillation or Kepler problem whose simplicity has been obfuscated by various coordinate transformations. For our examples, we consider a hidden symmetry to have been tentatively discovered if its corresponding loss drops below a threshold $\epsilon = 10^{-3}$ [2]. If that happens, we apply the AI Feynman symbolic regression package [17, 18] to try to discover a symbolic approximation of the learned transformation $\mathbf{f}_{\rm NN}$ that makes the symmetry loss zero. As can be seen in Tab. II and Fig. 3, all hidden symmetries are successfully discovered together with the coordinate transformations that reveal them. This includes not only traditional hidden symmetries such as translational invariance (example A) and rotational equivariance (example B), but also Hamiltonicity and modularity.

A-C were toy examples in the sense that we had hidden the symmetries by deliberate obfuscation. In contrast, the value of our algorithm lies in its ability to discover symmetries hidden not by people but by nature, as in example D (the linearized double pendulum). We see that our method auto-discovers both hamiltonicity (by finding the correct conjugate momentum variables) and

[2] How low the loss should be to warrant interpretation as a symmetry discovery depends on both training accuracy and data noise; see appendix G for details. For our examples, we found $\epsilon = 10^{-3}$ to be an small enough for symbolic regression to be able to discover the exact formula, after which the symmetry loss drops to exactly zero.

TABLE II: Physical Systems studied

ID	Name	Original dynamics $\dot{\mathbf{z}}=\mathbf{f}(\mathbf{z})$ or metric $g(\mathbf{z})$	Transformation $\mathbf{z}\mapsto\mathbf{z}'$	Symmetric dynamics $\dot{\mathbf{z}}'=\mathbf{f}'(\mathbf{z}')$ or metric $g'(\mathbf{z}')$	Manifest Symmetries
A	1D Uniform Motion	$\frac{d}{dt}\begin{pmatrix}a\\b\end{pmatrix}=\begin{pmatrix}\frac{1}{2}(a+b)\ln(\frac{a-b}{2})\\\frac{1}{2}(a+b)\ln(\frac{a-b}{2})\end{pmatrix}$	$\begin{pmatrix}a\\b\end{pmatrix}=\begin{pmatrix}e^{\frac{1}{2}x}+e^{\frac{1}{2}p}\\e^{\frac{1}{2}x}-e^{\frac{1}{2}p}\end{pmatrix}$	$\frac{d}{dt}\begin{pmatrix}x\\p\end{pmatrix}=\begin{pmatrix}p\\0\end{pmatrix}$	Hamiltonicity 1D Translational invariance
B	1D Harmonic Oscillator	$\frac{d}{dt}\begin{pmatrix}a\\b\end{pmatrix}=\begin{pmatrix}(1+a)\ln(1+b)\\-(1+b)\ln(1+a)\end{pmatrix}$	$\frac{d}{dt}\begin{pmatrix}a\\b\end{pmatrix}=\begin{pmatrix}e^{\frac{1}{2}x}-1\\e^{\frac{1}{2}p}-1\end{pmatrix}$	$\frac{d}{dt}\begin{pmatrix}x\\p\end{pmatrix}=\begin{pmatrix}p\\-x\end{pmatrix}$	Hamiltonicity SO(2)-equivariance
C	2D Kepler	$\frac{d}{dt}\begin{pmatrix}a\\b\\c\\d\end{pmatrix}=\begin{pmatrix}(1+a)\ln(1+b)\\-\frac{(1+b)\ln(1+a)}{8(\ln^2(1+a)+\ln^2(1+c))^{3/2}}\\(1+c)\ln(1+d)\\-\frac{(1+d)\ln(1+c)}{8(\ln^2(1+a)+\ln^2(1+c))^{3/2}}\end{pmatrix}$	$\begin{pmatrix}a\\b\\c\\d\end{pmatrix}=\begin{pmatrix}e^{\frac{1}{2}x}-1\\e^{\frac{1}{2}p_x}-1\\e^{\frac{1}{2}y}-1\\e^{\frac{1}{2}p_y}-1\end{pmatrix}$	$\frac{d}{dt}\begin{pmatrix}x\\p_x\\y\\p_y\end{pmatrix}=\begin{pmatrix}p_x\\-x/r^3\\p_y\\-y/r^3\end{pmatrix}$	Hamiltonicity Can SO(2)-equivariance
D	Double Pendulum	$\frac{d}{dt}\begin{pmatrix}\theta_1\\\dot{\theta}_1\\\theta_2\\\dot{\theta}_2\end{pmatrix}=\begin{pmatrix}\dot{\theta}_1\\-\frac{(m_1+m_2)g}{m_1 l}\theta_1+\frac{m_2 g}{m_1 l}\theta_2\\\dot{\theta}_2\\\frac{(m_1+m_2)g}{m_1 l}\theta_1-\frac{(m_1+m_2)g}{m_1 l}\theta_2\end{pmatrix}$	$\begin{pmatrix}\theta_1\\\dot{\theta}_1\\\theta_2\\\dot{\theta}_2\end{pmatrix}=\begin{pmatrix}-1&1&0&0\\a&a&0&0\\0&0&-1&1\\0&0&a&a\end{pmatrix}\begin{pmatrix}\theta_+\\\dot{\theta}_+\\\theta_-\\\dot{\theta}_-\end{pmatrix}$ $a=\sqrt{\frac{m_1+m_2}{m_2}}$	$\frac{d}{dt}\begin{pmatrix}\theta_+\\\dot{\theta}_+\\\theta_-\\\dot{\theta}_-\end{pmatrix}=\begin{pmatrix}\dot{\theta}_+\\-\omega_+^2\theta_+\\\dot{\theta}_-\\-\omega_-^2\theta_-\end{pmatrix}$ $\omega_\pm^2=\frac{(m_1+m_2)g}{m_1 l}(1\pm\sqrt{\frac{m_2}{m_1+m_2}})$	Hamiltonicity (2 + 2)-Modularity
E	Expanding universe & empty space	$\mathbf{g}=\begin{pmatrix}1&0&0&0\\0&-(r^2+\frac{kx^2}{1-kr^2})\frac{t^2}{r^2}&-\frac{kxyt^2}{1-kr^2}&-\frac{kxzt^2}{1-kr^2}\\0&-\frac{kxyt^2}{1-kr^2}&-(r^2+\frac{ky^2}{1-kr^2})\frac{t^2}{r^2}&-\frac{kyzt^2}{1-kr^2}\\0&-\frac{kxzt^2}{1-kr^2}&-\frac{kyzt^2}{1-kr^2}&-(r^2+\frac{kz^2}{1-kr^2})\frac{t^2}{r^2}\end{pmatrix}$	$\begin{pmatrix}t'\\x'\\y'\\z'\end{pmatrix}=\begin{pmatrix}t\sqrt{1+r^2}\\tx\\ty\\tz\end{pmatrix}$	$\mathbf{g}=\begin{pmatrix}1&0&0&0\\0&-1&0&0\\0&0&-1&0\\0&0&0&-1\end{pmatrix}$	SO(3,1)-Invariance 4D Translational Invariance
F	Schwarzchild black hole & GP metric	$\mathbf{g}=\begin{pmatrix}1-\frac{2M}{r}&0&0&0\\0&-1-\frac{2Mx^2}{(r-2M)r^2}&-\frac{2Mxy}{(r-2M)r^2}&-\frac{2Mxz}{(r-2M)r^2}\\0&-\frac{2Mxy}{(r-2M)r^2}&-1-\frac{2My^2}{(r-2M)r^2}&-\frac{2Myz}{(r-2M)r^2}\\0&-\frac{2Mxz}{(r-2M)r^2}&-\frac{2Myz}{(r-2M)r^2}&-1-\frac{2Mz^2}{(r-2M)r^2}\end{pmatrix}$	$\begin{pmatrix}t'\\x'\\y'\\z'\end{pmatrix}=\begin{pmatrix}t+2M\left[2u+\ln\frac{u-1}{u+1}\right]\\x\\y\\z\end{pmatrix}$ $u\equiv\sqrt{\frac{r}{2M}}$	$\mathbf{g}=\begin{pmatrix}1-\frac{2M}{r}&-\sqrt{\frac{2M}{r}}\frac{x}{r}&-\sqrt{\frac{2M}{r}}\frac{y}{r}&-\sqrt{\frac{2M}{r}}\frac{z}{r}\\-\sqrt{\frac{2M}{r}}\frac{x}{r}&-1&0&0\\-\sqrt{\frac{2M}{r}}\frac{y}{r}&0&-1&0\\-\sqrt{\frac{2M}{r}}\frac{z}{r}&0&0&-1\end{pmatrix}$	SO(3)-Invariance 3D Translational Invariance

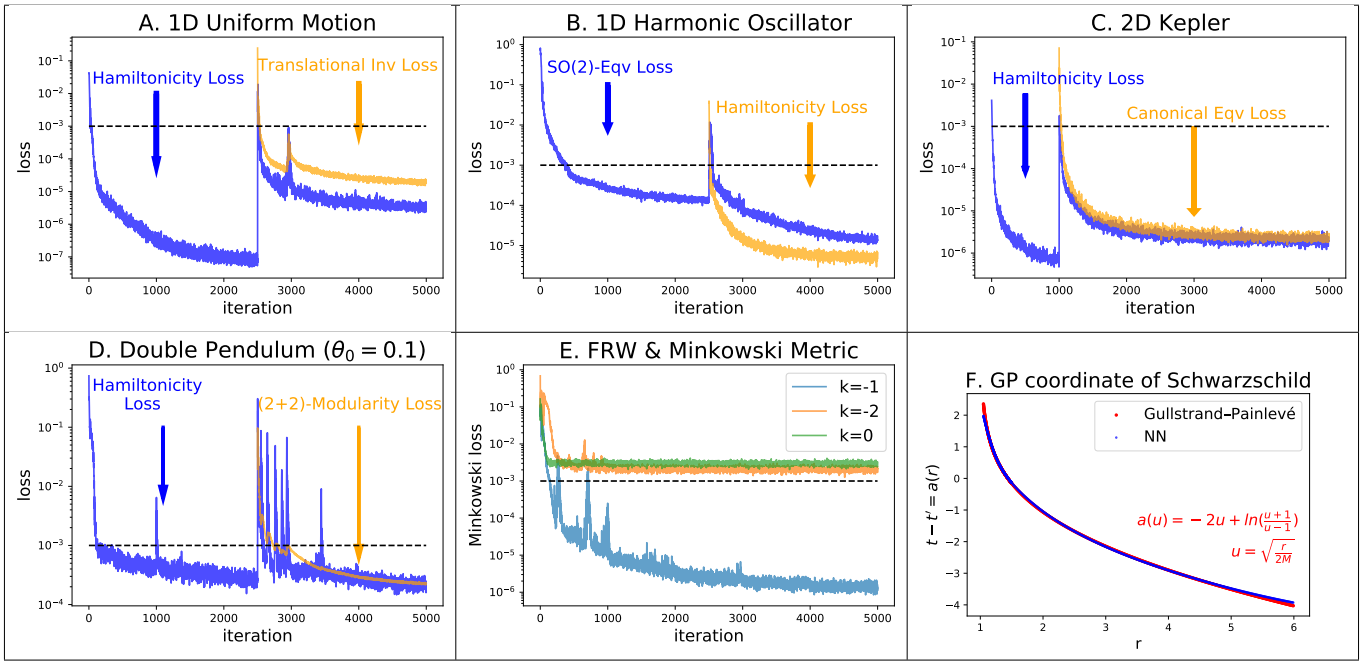

FIG. 3: All hidden symmetries in six tested systems are discovered by our algorithm. The last figure shows that the neural network accurately learns the Gullstrand-Painlevé transformation.

modularity (by auto-discovering the two normal modes), even though neither of these symmetries were manifest in the most obvious physical coordinates (the pendulum angles and angular velocities) [19].

General relativity examples

As a first general relativity (GR) application (example E), we consider the Friedmann-Robertson-Walker (FRW) metric[3] describing a homogeneous and isotropic expanding universe with negative spatial curvature ($k = 1$) and cosmic scale factor evolution $a(t) = t$. A GR expert will realize that its Riemann tensor vanishes everywhere, so that there must exist a coordinate transformation revealing this to be simply empty space in disguise, with Poincaré symmetry (Lorenz symmetry and 4D translational symmetry). Discovering this transformation is

[3] We use $r \equiv \sqrt{x^2+y^2+z^2}$ for brevity in Tab. II, but not in the neural network, which actually takes (t, x, y, z) as inputs.

non-trivial, and is sometimes assigned as a homework problem in graduate courses.

It is easy to show that any metric with Poincaré symmetry must be a multiple of the Minkowski metric η, so we define our Poincaré symmetry loss as $\ell \equiv \langle ||T(\mathbf{z}) - \eta||^2 \rangle / \langle ||T(\mathbf{z})||^2 \rangle$. Figure 3 (E) shows that the Minkowski loss drops below 10^{-3}, indicating that the $k = -1$ FRW metric is indeed homomorphic to Minkowski space, while the loss gets stuck above 10^{-3} for $k = -2$ and $k = 0$. Applying the AI Feynman symbolic regression package [17] to the learned transformation $\mathbf{f}_{\rm NN} = (t, x, y, z)$ reveals the exact formula $(x', y', z', t') = (tx, ty, tz, t\sqrt{1 + x^2 + y^2 + z^2})$, which gives vanishing loss.

We now turn to studying the spacetime of a non-rotating black hole described by the Schwarzschild metric (without loss of generality, we set $2M = 1$). This problem proved so difficult that it took physicists 17 years to clear up the misconception that something singular occurs at the event horizon, until it was finally revealed that the apparent singularity at $r = 2M$ was merely caused by a poor choice of coordinates [20–23], just as the z-axis is merely a coordinate singularity in spherical coordinates. Our method auto-discovers hidden translational symmetry in the spatial coordinates (x, y, z), revealed by the coordinate transformation $t' = t + 2M\left[2u + \ln\frac{u-1}{u+1}\right]$, where $u \equiv \sqrt{r/2M}$, which is auto-discovered by applying AI Feynman [17] to the learned transformation $\mathbf{f}_{\rm NN}$ (see Fig. 3, panel F). Since both the original and target metrics have the SO(3) (rotational) spatial symmetry, our neural network parametrizes the coordinate transformation $(x, y, z, t) \to (x', y', z', t')$ via a two-dimensional transformation $(r, t) \to (r', t')$, where $r \equiv \sqrt{x^2 + y^2 + z^2}$. This transforms the Schwarzschild metric into the famous Gullstrand-Painlevé metric [22, 23], which is seen to be perfectly regular at the event horizon and can be intuitively interpreted simply as flat space flowing inward with the classical escape velocity [11, 24].

CONCLUSIONS

We have presented a machine-learning algorithm for auto-discovering hidden symmetries, and shown it to be effective for a series of examples from classical mechanics and general relativity. Our symmetry definition is very broad, corresponding to the data satisfying a differential equation, which encompasses both traditional invariance and equivariance as well as Hamiltonicity and modularity.

Our work is linked to Noether's theorem [25], which states that a continuous symmetry leaving the Lagrangian invariant corresponds to a conservation law. The Lagrangian is a scalar, a special case of the tensors of this paper. If we rewrite the dynamical equations in the form of Euler-Lagrange equations, then the invariance of the Lagrangian under a symmetry group is equivalent to the equivariance of the dynamical equation under the same symmetry group, both of which imply the same conservation laws.

In future work, it will be interesting to seek hidden symmetries in data from both experiments and numerical simulations. Although our examples involved no more than two symmetries at once, it is straightforward to auto-search for a whole library of common symmetries, adopting the best-fitting one and recursively searching for more hidden symmetries until all are found.

Currently, our method can only search for symmetries from a list of candidates pre-specified by the user, and cannot search for unknown symmetries. In future work, it will be interesting to enable search also for unknown symmetries, *e.g.*, by making the Lie generators trainable. In other words, if there is *any* differential equation that a suitably transformed dataset satisfies, one would seek to auto-discover both the transformation and the differential equation.

Acknowledgement We thank the Center for Brains, Minds, and Machines (CBMM) for hospitality. This work was supported by The Casey and Family Foundation, the Foundational Questions Institute, the Rothberg Family Fund for Cognitive Science and IAIFI through NSF grant PHY-2019786.

[1] P. W. Anderson, More is different, Science **177**, 393 (1972), https://science.sciencemag.org/content/177/4047/393.full.pdf.
[2] D. J. Gross, Symmetry in physics: Wigner's legacy, Physics Today **48**, 46 (1995).
[3] D. J. Gross and F. Wilczek, Asymptotically free gauge theories. i, Physical Review D **8**, 3633 (1973).
[4] T. Cohen and M. Welling, Group equivariant convolutional networks, in *International conference on machine learning* (PMLR, 2016) pp. 2990–2999.
[5] N. Thomas, T. Smidt, S. Kearnes, L. Yang, L. Li, K. Kohlhoff, and P. Riley, Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds, arXiv preprint arXiv:1802.08219 (2018).
[6] F. B. Fuchs, D. E. Worrall, V. Fischer, and M. Welling, Se (3)-transformers: 3d roto-translation equivariant attention networks, arXiv preprint arXiv:2006.10503 (2020).
[7] R. Kondor, Z. Lin, and S. Trivedi, Clebsch-gordan nets: a fully fourier space spherical convolutional neural network, arXiv preprint arXiv:1806.09231 (2018).
[8] V. G. Satorras, E. Hoogeboom, and M. Welling, E (n) equivariant graph neural networks, arXiv preprint arXiv:2102.09844 (2021).
[9] G. Kanwar, M. S. Albergo, D. Boyda, K. Cranmer, D. C. Hackett, S. Racanière, D. J. Rezende, and P. E. Shanahan, Equivariant flow-based sampling for lattice gauge theory, Phys. Rev. Lett. **125**, 121601 (2020).
[10] D. Boyda, G. Kanwar, S. Racanière, D. J. Rezende, M. S. Albergo, K. Cranmer, D. C. Hackett, and P. E. Shana-

han, Sampling using SU(n) gauge equivariant flows, Phys. Rev. D **103**, 074504 (2021).
[11] C. Misner, K. Thorne, and J. Wheeler, *Gravitation* (Princeton University Press, 2017).
[12] M. Raissi, P. Perdikaris, and G. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, Journal of Computational Physics **378**, 686 (2019).
[13] T. Beucler, M. Pritchard, S. Rasp, J. Ott, P. Baldi, and P. Gentine, Enforcing analytic constraints in neural networks emulating physical systems, Phys. Rev. Lett. **126**, 098302 (2021).
[14] M. Cranmer, S. Greydanus, S. Hoyer, P. Battaglia, D. Spergel, and S. Ho, Lagrangian neural networks, arXiv preprint arXiv:2003.04630 (2020).
[15] Z. Liu, B. Wang, Q. Meng, W. Chen, M. Tegmark, and T.-Y. Liu, Machine-learning nonconservative dynamics for new-physics detection, Phys. Rev. E **104**, 055302 (2021).
[16] S. Elfwing, E. Uchibe, and K. Doya, Sigmoid-weighted linear units for neural network function approximation in reinforcement learning, Neural Networks **107**, 3 (2018), special issue on deep reinforcement learning.
[17] S.-M. Udrescu, A. Tan, J. Feng, O. Neto, T. Wu, and M. Tegmark, Ai feynman 2.0: Pareto-optimal symbolic regression exploiting graph modularity (2020), arXiv:2006.10782 [cs.LG].
[18] S.-M. Udrescu and M. Tegmark, Ai feynman: A physics-inspired method for symbolic regression, Science Advances **6**, 10.1126/sciadv.aay2631 (2020).
[19] H. Goldstein, C. Poole, and J. Safko, Classical mechanics (2002).
[20] R. Penrose, Gravitational collapse and space-time singularities, Phys. Rev. Lett. **14**, 57 (1965).
[21] M. D. Kruskal, Maximal extension of schwarzschild metric, Phys. Rev. **119**, 1743 (1960).
[22] P. Painlevé, La mécanique classique et la théorie de la relativité, C. R. Acad. Sci. (Paris) **173**, 677 (1921).
[23] A. Gullstrand, Allgemeine lösung des statischen einkörrperproblems in der einsteinschen gravitationstheorie, Arkiv för Matematik, Astronomi och Fysik **16 (8)**, 1 (1922).
[24] A. J. S. Hamilton and J. P. Lisle, The river model of black holes, American Journal of Physics **76**, 519–532 (2008).
[25] E. Noether, Invariante variationsprobleme, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse **1918**, 235 (1918).