Learning orbital dynamics of binary black hole systems from gravitational wave measurements

Brendan Keith [*]
Center for Applied Scientific Computing, Lawrence Livermore National Laboratory, Livermore, California 94550, USA

Akshay Khadse
Department of Physics and Astronomy, The University of Mississippi, University, Mississippi 38677, USA

Scott E. Field
Department of Mathematics, University of Massachusetts, Dartmouth, Massachusetts 02747, USA
and Center for Scientific Computing and Visualization Research, University of Massachusetts, Dartmouth, Massachusetts 02747, USA

(Received 17 March 2021; accepted 4 October 2021; published 9 November 2021)

We introduce a gravitational waveform inversion strategy that discovers mechanical models of binary black hole (BBH) systems. We show that only a single time series of (possibly noisy) waveform data is necessary to construct the equations of motion for a BBH system. Starting with a class of universal differential equations parameterized by feed-forward neural networks, our strategy involves the construction of a space of plausible mechanical models and a physics-informed constrained optimization within that space to minimize the waveform error. We apply our method to various BBH systems including extreme and comparable mass ratio systems in eccentric and noneccentric orbits. We show that the resulting differential equations apply to time durations longer than the training interval, and relativistic effects, such as perihelion precession, radiation reaction, and orbital plunge, are automatically accounted for. The methods outlined here provide a data-driven approach to studying the dynamics of binary black hole systems.

DOI: 10.1103/PhysRevResearch.3.043101

I. INTRODUCTION

Classical physical theories begin with scientific laws as ansätze, which are validated by repeated scientific experiment. From these laws, one derives a set of equations (usually differential equations) that can be solved, either completely or partially, to deduce various conclusions about the physical system under consideration. In this paper, we follow a different approach to learning physical equations: we solve an optimization problem that isolates the most likely physical model (differential equations) that would deliver certain physical measurements (data). This approach is aligned with a growing trend in data-driven science; see, e.g., [1–12] and references therein.

The focus of this paper is on learning mechanical models for binary black hole (BBH) systems through gravitational wave measurements. As the black holes orbit one another, the motion of these massive objects generate gravitational waves that radiate away to the far-field where they can be observed by an international-network of detectors [13]. Complicated partial differential equations (PDEs) govern the entire process, and in particular connect the near-field dynamics to the far-field gravitational radiation. Traditionally, black hole orbital dynamics and gravitational waves have been computed by expensive simulation codes [14] or approximations to general relativity such as the post-Newtonian formalism [15].

Our principal contribution is to show that two-body relativistic orbital models can be deduced from gravitational wave measurements by solving an inverse problem [16,17] where the control variable is the vector of weights and biases in a neural network. Our numerical examples use gravitational waveform measurements from both a noisy detector and "clean" measurements from numerical relativity (NR) simulations. As such, the techniques described here may apply to both traditional modeling endeavors that require NR waveform data for calibration and gravitational-wave (GW) astronomy. In the latter case, where the training data is comprised of GW observations, our inversion strategy avoids the need to solve or analyze Einstein's equation of general relativity to learn the orbital model.

A key goal of our paper is to develop the computational framework for learning binary black hole dynamical models from gravitational waves, which is a new approach to the modeling problem. As such, we focus on simple modeling choices and apply them to illustrative examples. We will show that simple ansatz models parameterized by feed-forward neural networks can be used to discover complicated dynamics. Indeed, despite starting with an essentially Newtonian ansatz model [cf. Eq. (5)], our trained models accurately capture both the relativistic dynamics and the

[*]Corresponding author: keith10@llnl.gov

waveform (cf. Sec. III). Potential applications are considered in Sec. IV.

II. METHODOLOGY

A. Universal differential equations

In this paper, we rely on the following general class of dynamical system models referred to in [8] as (autonomous) universal differential equations (UDEs),

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathcal{F}(\mathbf{x})), \quad \mathbf{x}(0) = \mathbf{x}_0, \tag{1}$$

where $\mathbf{x}(t)$ is the solution vector and $\mathbf{x}_0$ specifies the initial conditions. Here, $\mathcal{F}(\mathbf{x})$ is neural network and the overdot symbol "˙" denotes differentiation with respect to time $t \in [0, T]$. For example, if $\mathcal{F}(\mathbf{x}) = \mathcal{F}(\mathbf{x}; \boldsymbol{\xi})$ is a feed-forward deep neural network with two hidden layers, it can be written as

$$\mathcal{F}(\mathbf{x}; \boldsymbol{\xi}) = W_3\, \sigma_2(W_2\, \sigma_1(W_1(\mathbf{x}) + b_1) + b_2) + b_3, \tag{2}$$

with $\boldsymbol{\xi} = (W_1, W_2, W_3, b_1, b_2, b_3)$. Here, W_i are matrices (weights), b_i are vectors (biases), and σ_i are the chosen activation functions.

We note that the UDE paradigm permits an immense variety of different parameterized functions $\mathcal{F}(\mathbf{x})$, not only neural network-based parametrizations. This flexibility provides numerous advantages, for example, prior scientific knowledge of the solution may be incorporated into the choice of function parametrization. We choose to use feed-forward neural networks, such as Eq. (2), because of the ease with which their parametrization can be determined using existing software [8].

B. BBH modeling

Our task is to define a family of physical models that can be used to describe relativistic orbital dynamics of two spherical objects of mass m_1 and m_2. That is, we ask that our model provides the position of object 1, $\mathbf{r}_1(t)$, and object 2, $\mathbf{r}_2(t)$, which are the solutions to the dynamical system model. In Newtonian physics, this is the familiar two-body problem of Kepler whose solution has been known since the earliest days of classical mechanics. In relativistic physics, the field of computational relativity is largely devoted to providing numerical solutions to this problem by solving the equations of general relativity (a nonlinear, coupled, hyperbolic-elliptic PDE system) on large supercomputers [14]. In order to motivate a dynamical system model, we consider special cases whereby the dynamical motion described by PDEs can be approximately reduced to ordinary differential equations (ODEs). Such approximations have been well developed over many decades, and we refer the reader to [15,18,19] and references therein. Throughout this paper, we use geometric units where both the speed of light c, and the gravitational constant G, are set to unity.

First, for slow moving objects ($v \ll c$), a powerful formalism known as the post-Newtonian approximation provides a systematic framework for adding relativistic corrections in powers of v/c to the Keplerian equations of motion. The post-Newtonian framework provides us with justification for treating the two-body problem as an effective one-body problem. Here, the relevant equations that govern the separation vector, $\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$, can be used to reconstruct the two-body motion through the relations,

$$\mathbf{r}_1 = (m_2/M)\mathbf{r}, \quad \mathbf{r}_1 = -(m_1/M)\mathbf{r}, \tag{3}$$

where $M = m_1 + m_2$. We call this an effective one-body problem as one often views there to be an "effective" body located at $\mathbf{r}$ relative to the system's center-of-mass. A different technique, known as the post-Minkowskian approximation, informs us that the dominant contribution to the gravitational radiation field can be computed according to the quadrupole formula, first derived by Einstein. Collectively, the two-body-to-one-body map and the quadrupole formula neglect a substantial amount of physics including terms proportional to v/c, as well as many more higher-order terms, some of which have been computed and others that remain unknown.

Second, in the limit of $m_1 \gg m_2$, the two-body problem reduces to a simpler setup whereby the larger object is fixed at the coordinate system's origin. The smaller orbiting object's motion is then described by a geodesic path in the Schwarzschild geometry set by the larger black hole. Recent numerical evidence suggests that the geodesic equations of motion (11) with self-force corrections may work unreasonably well even for near-equal mass systems [20–28]. More importantly for our purposes, we know from black-hole perturbation theory results that using geodesic equations of motion to describe the two-body problem neglects a substantial amount of physics including terms proportional to m_2/m_1.

Having motivated some of the physics behind the problem, we now outline our strategy to write down a model inspired by a combination of Newtonian and relativistic physics. Our orbital model will omit a significant amount of important physics that will be accounted for by deep neural networks trained on gravitational waveform data.

We write the two-body problem as an effective one-body one, and associate the orbital separation vector $\mathbf{r}$ with the location of the fictitious effective body orbiting a fixed, spherically-symmetric central object. Owing to the spherical symmetry of the central object, we may assume, without loss of generality, that the effective object's trajectory lies in the equatorial plane, which we take to be the plane perpendicular to the z axis and where the *angle* ϕ is between $\mathbf{r}$ and the x axis. Orbits are specified by the orbital parameters *eccentricity* $e(t)$ and *semilatus rectum* $p(t)$; in the Newtonian case these are constants while in the general relativistic case they are often interpreted as time-dependent functions. Finally, we use a well-known parametrization for the Euclidean norm of $\mathbf{r}$,

$$r(t) = p(t)M/(1 + e(t)\cos\chi(t)), \tag{4}$$

and evolve the *anomaly* $\chi(t)$ instead of $r(t)$ because the anomaly increases monotonically through radial turning points. To summarize, we assume the equations of motion for the effective object can be described by four time-dependent variables ϕ, χ, p, and e. The effective object's trajectory is provided ϕ, χ, whereas p, e parametrize the orbital configuration.

Upon denoting $\mathbf{x} = (\phi, \chi, p, e)$, we propose the following family of UDEs to describe the two-body relativistic dynamics:

$$\dot{\phi} = \frac{(1 + e\cos(\chi))^2}{Mp^{3/2}}(1 + \mathcal{F}_1(\cos(\chi), p, e)), \tag{5a}$$

$$\dot{\chi} = \frac{(1 + e\cos(\chi))^2}{Mp^{3/2}}(1 + \mathcal{F}_2(\cos(\chi), p, e)), \tag{5b}$$

$$\dot{p} = \mathcal{F}_3(p, e), \tag{5c}$$

$$\dot{e} = \mathcal{F}_4(p, e), \tag{5d}$$

with $\mathbf{x}(0) = (\phi_0, \chi_0, p_0, e_0)$. Note that the functional form of Eqs. (5) have been inspired by Eq. (11), which are the geodesic equations of motion for an infinitesimally small "particle" orbiting a supermassive black hole. In particular, Eq. (5) are rotationally invariant because the right-hand side omits the ϕ variable. Moreover, when $\mathcal{F}_3 = \mathcal{F}_4 = 0$, orbital energy $E(p, e)$ and orbital angular momentum $L(p, e)$ are conserved:

$$\dot{E} = \frac{\partial E}{\partial p}\dot{p} + \frac{\partial E}{\partial e}\dot{e}, \quad \dot{L} = \frac{\partial L}{\partial p}\dot{p} + \frac{\partial L}{\partial e}\dot{e}. \tag{6}$$

Due to the emission of gravitational waves (so-called radiation-reaction), we have that both $\dot{E}, \dot{L} < 0$ for all time. When each $\mathcal{F}_j = 0$, we recover Newtonian orbits.

Equations (5a) through (5d) define a family of trajectories $\mathbf{x}(t) = \mathbf{x}(t; \boldsymbol{\xi})$. Through Eq. (3), these trajectories determine the black hole orbits,

$$\mathbf{r}_1(t) = \frac{r(t)m_2}{M}(\cos(\phi(t)), \sin(\phi(t)), 0), \tag{7a}$$

$$\mathbf{r}_2(t) = -\frac{r(t)m_1}{M}(\cos(\phi(t)), \sin(\phi(t)), 0). \tag{7b}$$

Gravitational waves are generated by orbiting black holes, and so the waves encode detailed information about the dynamical variables $\mathbf{r}_1(t)$ and $\mathbf{r}_2(t)$. General relativity tells us that the dynamics and waves are connected through PDEs, which is a familiar scenario in the modeling of waves.

In the next section we summarize how to learn $\mathcal{F}_j$ from gravitational wave measurements. Despite the simplicity of Eqs. (5), we show that the learned ODEs can describe dynamics beyond the base mechanical model (the base model corresponds to setting the neural network parameters $\boldsymbol{\xi}$ to zero and, thus, each $\mathcal{F}_j = 0$). In our first numerical experiment III A, for example, we show that bound orbits of a test particle following geodesic motion on a Schwarzschild geometry can be accounted for. In our final set of numerical experiments we show that the dissipative dynamics can also be accounted for. Due to the flexible framework of our waveform inversion technique, one can easily swap out our base model (5) for others; e.g., the EOB model [29]. This suggests many possible future applications of gravitational waveform inversion.

C. Quadrupole formula, the loss function, and model discovery

Very far from a BBH system, where gravitational wave detectors are located, the gravitational radiation field is an outgoing spherical wavefront. On a sufficiently large sphere we can expand the radiation field into a complete basis of (tensorial) spherical harmonics labeled by (ℓ, m) harmonic indices.

In this paper, we consider only the dominant $(\ell, m) = (2, 2)$-mode gravitational waveforms (cf. Appendix A 2), however, our waveform inversion technique could be modified to easily include subdominant modes. Accordingly, we denote all waveforms by the variable $w = (r/M)\,\mathrm{Re}\{h^{22}\}$, where

$$h^{22}(t) = \frac{1}{r}\sqrt{\frac{4\pi}{5}}(\ddot{\mathcal{I}}_{xx} - 2i\ddot{\mathcal{I}}_{xy} - \ddot{\mathcal{I}}_{yy}) \tag{8}$$

and the trace-free mass quadrupole tensors $\mathcal{I}_{xx}$, $\mathcal{I}_{xy}$, and $\mathcal{I}_{yy}$ are defined in Eq. (A10) (see also [30, Eqs. 54–56]). Eq. (8) is the well-known quadrupole formula, which expresses the measurable waveform h^{22} in terms of the orbits $\mathbf{r}_1 = (x_1, y_1, 0)$ and $\mathbf{r}_2 = (x_2, y_2, 0)$. The quadrupole formula is a very simple approximation that will necessarily introduce systematic error when learning $\mathcal{F}_j$, but is sufficient for our purpose.

We assume that our waveform measurements appear as ordered pairs (t_k, w_k), where w_k denotes the value of the waveform data at time $t_k \in [0, T]$. In this setting, we define the mean-squared waveform error

$$\mathcal{J}(\mathbf{x}) = \langle J(\mathbf{x}, \cdot)\rangle := \frac{1}{T}\int_0^T J(\mathbf{x}, t)\,dt, \tag{9}$$

where $J(\mathbf{x}, t) = \sum_k (w_k - w(t))^2\delta(t - t_k)$ and bracket notation, $\langle\cdot\rangle$, denotes denotes averaging over the time interval. Accordingly, we choose to solve the inverse problem:

$$\min_{\boldsymbol{\xi}} \mathcal{J}(\mathbf{x}) \quad \text{subject to (5a) through (5d).} \tag{10}$$

In some situations, convergence to the solution of Eq. (10) can be improved by adding well-chosen, physics-informed penalty and regularization to Eq. (9); cf. Sec. III B.

We note that the exclusive use of gravitational-wave data in the loss function is motivated by the consideration that in experimental settings only gravitational-wave observations will be available and never a direct view of black hole orbits. Even in computational relativity simulations, the numerical measurement of black hole trajectories are complicated by coordinate ambiguities of general relativity that make it difficult to assign physical significance to their values. Waveforms computed from computational relativity simulations, on the other hand, are well-defined and physically meaningful.

The ODE-constrained optimization problem Eq. (10) delivers the calibrated dynamical system model $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathcal{F}(\mathbf{x}; \boldsymbol{\xi}^\star))$, where $\boldsymbol{\xi}^\star$ denotes the optimizer found by solving Eq. (10). This inverse problem can be solved with a number of standard methods. We choose to use a BFGS algorithm with backtracking line search [31] and an adjoint-based (implicit differentiation/adjoint sensitivity method) calculation of gradients [32] implemented with the Julia [33] software package DiffEqFlux [8]. The algorithm is described by the flowchart in Fig. 1. Our code is available for download at [34].

III. RESULTS

In this section, we present results with three different examples. The first demonstrates the ability of Eq. (10) to

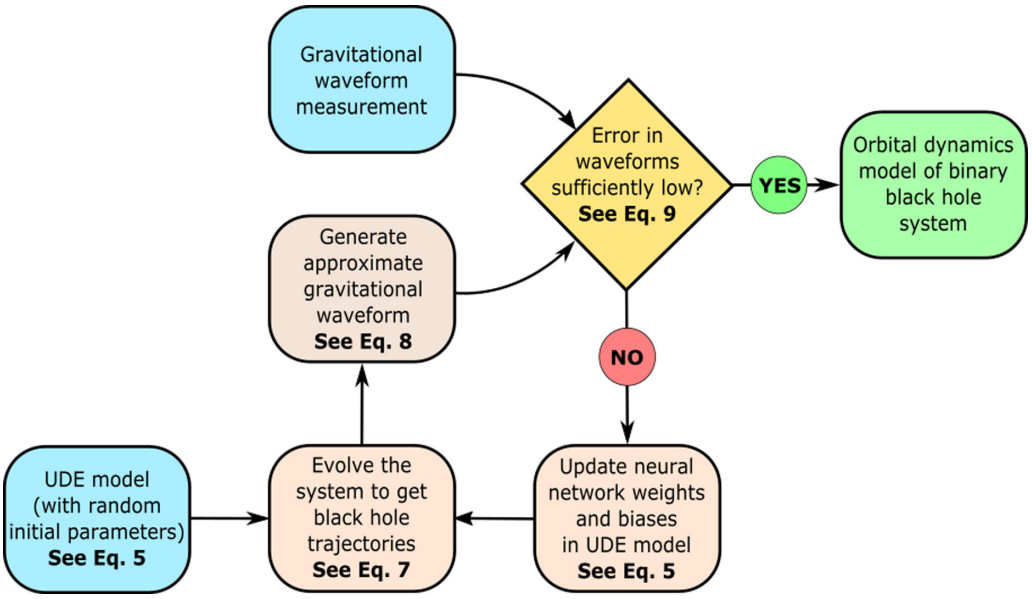

FIG. 1. Flowchart of algorithm to solve the inverse problem Eq. (10).

recover known orbital equations. The second two showcase the discovery of equations of motion for equal mass binary black hole mergers.

A. Extreme mass ratio systems

As our first motivating example, we consider a special case of the relativistic two-body problem where the exact solution is known. We show that from short-duration gravitational wave observations we are able to discover differential equations that are valid over much longer time scales.

In the regime of $m_1 \gg m_2$, formally the limit $m_1 \to M$, m_2 is a "test particle" whose motion obeys [19,35–37]

$$\dot{\phi} = \frac{(p-2-2e\cos\chi)(1+e\cos\chi)^2}{Mp^{3/2}[(p-2)^2-4e^2]^{1/2}}, \tag{11a}$$

$$\dot{\chi} = \frac{(p-2-2e\cos\chi)(1+e\cos\chi)^2[p-6-2e\cos\chi]^{1/2}}{Mp^{2}[(p-2)^2-4e^2]^{1/2}}, \tag{11b}$$

while $\mathbf{r}_1 = (0,0,0)$ and $\dot{e} = \dot{p} = 0$. We shall be interested in the parameter restriction $0 \leqslant e < 1$, for which the radial motion occurs between two turning points, $pM/(1+e)$ and $pM/(1-e)$ and the orbit is bounded. When $e = 0$, the orbit is circular. We let $\mathcal{F}_3 = \mathcal{F}_4 = 0$ and provide values for the initial conditions $\phi_0 = 0$ and $\chi_0 = \pi$. In our example, we set $e = 0.5$, $p = 100$, and $m_1 = 1$, although the results we show remain largely the same for other parameter values we have tested. For simplicity, in this first example we provide known values for e_0, p_0 while in Sec. III B we show how our approach performs when these parameters are also learned.

To prepare our ground-truth data, we numerically solve Eq. (11) on a dense time grid, thereby generating the black hole trajectory $\mathbf{r}_2(t)$. We then apply the quadropole formula Eq. (8) to generate a gravitational waveform sampled at 250 equally-spaced points spanning the time interval $[0, 0.6 \times 10^5]$, and shown in Fig. 2 (bottom panel; black dots). Note that in the extreme mass ratio limit, $m_2 \to 0$, and the waveform $h^{22} \propto m_2/m_1$ goes to zero. Therefore, in this example, we use $w = (m_1/m_2)\,\mathrm{Re}\{rh^{22}\}$ as gravitational-wave data; w is now independent of m_2.

Using the procedure summarized in Sec. II C, we recover the governing equations by optimizing for $\mathcal{F}_1$ and $\mathcal{F}_2$. In this setting, both abstract functions only depend on $\cos\chi$. We exploit this periodic structure by defining $\mathcal{F}_1$ and $\mathcal{F}_2$ with cosine activation functions, $\sigma_j = \cos$. We then construct $\mathcal{F}_1$ and $\mathcal{F}_2$ as feed-forward neural networks with two hidden layers each; see, e.g., Eq. (2). The exact network architecture and numerical discretization can be found in the file `EMR.jl` in [34]. Finally, we learn the corresponding neural network weights and biases by optimizing Eq. (10).

This process delivers the red trajectory and waveform presented in Fig. 2. Not only do both waveforms and trajectories match over the training interval consisting of about about 6 orbits, they continue to agree when the learned dynamics are extrapolated to about 31 orbits, after which the orbit's perihelion precession has undergone a full cycle. In Fig. 2 we compare the true waveforms/trajectories to the learned waveforms/trajectories over the extended time interval $[0, 3 \times 10^5]$. To compare the learned model Eq. (5) to the exact model Eq. (11), we also compute the error in $\dot{\phi}$ and $\dot{\chi}$ over the extended time interval. Evidently, not only do the waveforms and trajectories match upon visual inspection, the learned model matches the true mechanical model to about two orders of magnitude. The learned model also recovers important relativistic effects, notably perihelion precession, from just a few gravitational-wave cycles. Finally, we note that once the dynamical model is known, it can be used to generate very long orbits and gravitational wave signals by integrating the ODE Eq. (5) and postprocessing the solution with the quadrupole formula Eq. (8).

This experiment demonstrates the potential power of waveform inversion in a simple setting with a known solution. It also demonstrates how the information content in the

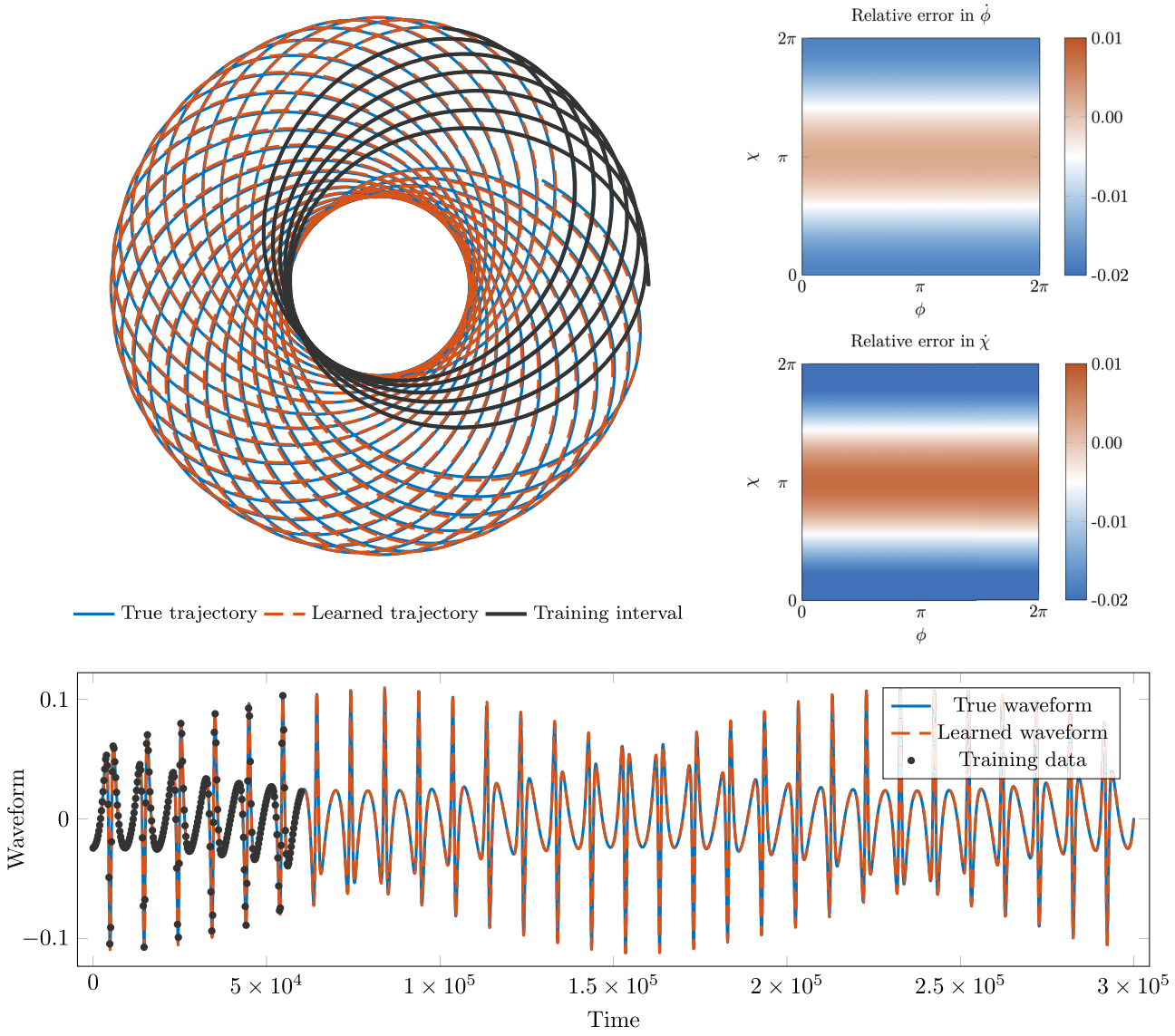

FIG. 2. Summary of our first experiment, where we have used gravitational-wave observations (black dots; bottom panel) to learn the underlying two-dimensional dynamical system model governing the relativistic two-body problem in the extreme mass ratio limit with orbital parameters $p = 100$, $e = 0.5$. Top left: Learned (dashed red) and exact (solid blue) trajectories extrapolated 4× the training interval. We also show the portion of the orbit (black) corresponding to the gravitational-wave training window, although no orbital data was used to learn the dynamics. Top right: Relative error between the learned model Eq. (5) and the exact model Eq. (11). Bottom: Learned (dashed red) and exact (solid blue) waveforms extrapolated 4× the training interval.

original waveform can be used to infer UDEs. Nevertheless, this system is conservative ($\dot{e} = \dot{p} = 0$), the quadrupole formula is prescribed exactly, and the learnable dynamics depend only on the χ variable. The following examples are more challenging because none of the aforementioned simplifications hold.

B. General relativistic orbital dynamics of binary black holes

In this pair of examples, we consider numerically generated waveform measurements from equal mass $m_1 = m_2 = 0.5$ binary black hole systems. Unlike the previous experiment, the orbital dynamics for these systems is much more complicated; The exact equations of motion are unknown and the dynamics include time-dependent values of the eccentricity and semilatus rectum. Although an extensive body of literature exists for deriving these equations from approximations of general relativity [15,29,39], we are unaware of any data-driven approaches focused on discovering orbital dynamics from waveform measurements.

Computational relativity codes provide exact (up to numerical discretization error) solutions to the general relativistic two-body problem, including both the corresponding trajectories of the center of the black holes and gravitational-wave data. Although the location of the black holes are coordinate-dependent they can still be used to compare with the trajectories obtained from our model. However, such comparisons should no longer be understood as model error since the coordinate system used for the data and model are necessarily different.

The simulations for this work were performed using the Spectral Einstein Code (SpEC) [38,40–45] developed by the Simulating eXterme Spacetimes (SXS) collaboration [40] and made publicly available through the Gravitational Waveform Database [38].

From now on, we augment the loss function $\mathcal{J}$ in Eq. (9) with non-negative penalty and regularization terms, motivated below:

$$\mathcal{J}(\mathbf{x}, \boldsymbol{\xi}) = \langle J(\mathbf{x}, \cdot)\rangle + \mathcal{P}_1(\mathbf{x}) + \mathcal{P}_2(\mathbf{x}) + \mathcal{R}(\boldsymbol{\xi}). \tag{12}$$

In this new expression, we define

$$\mathcal{P}_1(\mathbf{x}) = \gamma_1\langle(\dot{p})_+^2\rangle + \gamma_2\langle(\ddot{p})_+^2\rangle, \tag{13}$$

where $(f(t))_+ = \max\{f(t), 0\}$,

$$\mathcal{P}_2(\mathbf{x}) = \gamma_3\langle(-e)_+^2\rangle + \gamma_4\langle(e-e_0)_+^2\,\mathbf{1}_{\{p>6+2e_0\}}\rangle, \tag{14}$$

where $\mathbf{1}_\Omega$ denotes the indicator function on the set $\Omega \subset [0, T]$, and finally

$$\mathcal{R}(\boldsymbol{\xi}) = \gamma_5\|\boldsymbol{\xi}\|^2, \tag{15}$$

where $\|\boldsymbol{\xi}\|$ denotes the ℓ^2-norm of the expanded parameter vector $\boldsymbol{\xi}$. It is standard practice to use large coefficients for penalty terms and small coefficients for regularization terms. However, as explained in the paragraphs below, our penalty terms are present to help avoid nonphysical local minima and are not active in the optimized model. For this reason, we do not make a concerted attempt to tune these coefficients. In both of the coming experiments, we somewhat arbitrarily fix $\gamma_1 = 10^3$, $\gamma_2 = 10^2$, $\gamma_3 = 10^1$, and $\gamma_5 = 10^{-1}$. In the first experiment, we take $\gamma_4 = 1$, while in the second experiment we use $\gamma_4 = 0$.

The physical motivation for the terms in Eq. (13) relies on Eqs. (7). From these equations, we have that the distance between the two black holes r is proportional to p. Due to energy loss from the emitted gravitational waves, $r(t)$ converges to zero at a rate that increases throughout the system's evolution. The penalty terms $\langle(\dot{p})_+^2\rangle$ and $\langle(\ddot{p})_+^2\rangle$ have been chosen to encourage the selection of solutions with this physical behavior. The first term in Eq. (14) encourages the selection of a positive eccentricity function $e(t)$ for all time t. On the other hand, the final term in this definition is motivated by the stability condition $p \geqslant 6 + 2e$ for bound orbits. It is widely accepted that e decays in this range [35], and this term helps to direct the solution toward models with this property.

Clearly, if $\dot{p}$, $\ddot{p}$, $-e$, $(e-e_0)^2\,\mathbf{1}_{\{p>6+2e_0\}} \leqslant 0$, then $\mathcal{P}_1(\mathbf{x}) = \mathcal{P}_2(\mathbf{x}) = 0$. Our experiments appear to indicate that optimal solutions $p(t)$ and $e(t)$ satisfy each of these bounds, therefore, the penalty terms only act as guardrails throughout the optimization process. The Tikhonov–Phillips regularization term Eq. (15) helps convergence by ensuring continuous dependence between the data and the solution [46]. The Tikhonov regularization term $\|\boldsymbol{\xi}\|^2$ can also help avoid model degeneracies and overfitting in the presence of noisy data. For example, when the orbit is circular ($e = 0$) the model is degenerate in χ and, therefore, it is also degenerated in the weights and biases defining $\mathcal{F}_2$. Other penalty and regularization terms could be considered in future studies.

In the following pair of examples, we construct feed-forward neural network parametrizations of $\mathcal{F}_j$, $j = 1, \ldots, 4$, with tanh activation functions. The exact network architecture we use can be found in files `SXS1.jl` and `SXS2.jl` [34].

1. Near-circular orbits from clean GW observations

For this experiment, we consider a binary black hole system with negligible eccentricity during the initial inspiral. For inspection, the center of mass-corrected trajectories of the binary black hole system are depicted in the top left-hand corner of Fig. 3 (solid-black lines), with the associated 1000 equally-spaced waveform data points in the bottom panel (black dots). From now on, we let $[0, T]$ denote the time interval between the first ($t = 0$) and final ($t = T$) measurement, where the final measurement occurs shortly *before* merger.

As in the previous experiment, we adopt the initial conditions $\phi_0 = 0$ and $\chi_0 = \pi$ and assume that r_0 is known. Using Eq. (4), these assumptions provide us with an explicit expression for p_0, $Mp_0 = r_0\,(1 + e_0\cos(\chi_0))$. Due to the nearly-zero eccentricity of the initial trajectories, we opt for the simple initial condition $e_0 = 0$. In the next and final experiment, we treat the more realistic case where both e_0 and χ_0 are unknown.

In order to avoid local minima, we solve Eq. (10) on a sequence of increasing time intervals $[0, T_0] \subsetneq [0, T_1] \subsetneq \cdots \subsetneq [0, T]$, using the optimal parameters $\boldsymbol{\xi}^\star$ from each preceding optimization problem (plus a small amount of Gaussian noise) as initial data for the subsequent problem. Using this incremental procedure, we are able to recover the overwhelming majority of the black hole trajectories, as indicated by the visual agreement between the learned (red and blue) and NR (black) trajectories shown in the top left-hand panel of Fig. 3. We also depict the relative disagreement between the NR trajectory of the first black hole $\hat{\mathbf{r}}_1$ and the learned trajectory $\mathbf{r}_1$. The model also recovers important general relativistic effects, notably the learned functions $\mathcal{F}_3$ and $\mathcal{F}_4$ cause a runaway inspiral process that drives the black holes to merge. This process is seen most clearly by monitoring the behavior of $p(t)$ in the upper right-hand panel of Fig. 3. We also observe that our model is able to naturally include both the inspiral ($p > 6$) and plunge ($p < 6$) orbital regimes. Note the upper right-hand panel of Fig. 3 shows that near this transition region the eccentricity quickly grows, which is at odds with our physical expectation for stable orbits and indicates a very different dynamical regime.

One complication in validating our learned dynamical system is how to perform meaningful comparisons with other models. Indeed, besides the waveform, the other three subpanels shown in Fig. 3 depict gauge-dependent quantities; that is their value depends on the coordinate system being used. In particular, our trajectories are not expressed in the same damped harmonic gauge coordinates used by SpEC simulations [47]. Nevertheless, recent studies have noted surprisingly good agreement between NR trajectories and those computed with post-Newtonian (PN) models [48] and PN-augmented dynamical models [49–51]. In particular, Ref. [49] conjectures that the main source of disagreement is due to the PN formula being expressed in the harmonic gauge [15]. The close agreement between NR and UDE trajectories shown in Fig. 3 is another example of surprisingly good agreement [48,49].

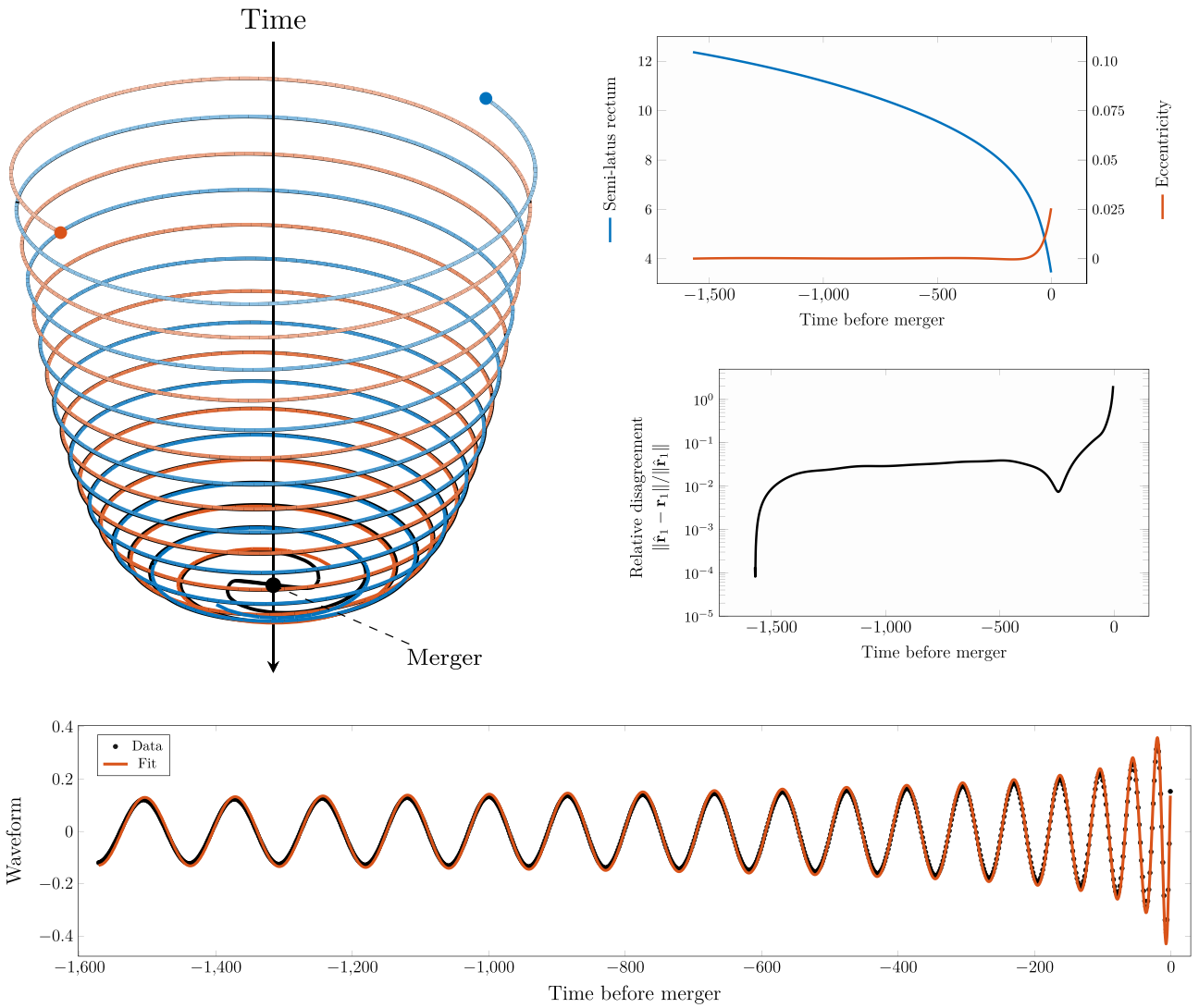

FIG. 3. Summary of our second experiment, where have used gravitational-wave observations (black dots; bottom figure) to learn the underlying dynamical system model governing the relativistic two-body problem for two equal mass black holes in quasi-circular orbit. Top left: Trajectories of the centers of mass of the black hole system `SXS:BBH:0217` (black lines), taken from the SXS Gravitational Waveform Database [38]. We also show the orbit computed from our learned dynamical system (blue and red lines). In the upper-right panel, we show the evolution of the eccentricity and semilatus rectum from our learned-dynamical system. The middle-right panel shows the disagreement between the NR trajectories and the ones computed from the learned dynamical system. We caution the reader that this figure should not be understood as a relative error because the numerical relativity black hole trajectories and our learned model are expressed in different coordinate systems that are impossible to relate. Bottom: Learned (red line) and computational relativity (black dots) waveform data corresponding to the real part of the h^{22} mode.

To avoid gauge ambiguities, comparisons of BBH dynamics focus on comparing gauge-invariants that are computable within different frameworks. For example, Refs. [52,53] explore the conservative dynamics by comparing the relationship between the total energy and total angular momentum from NR data to the corresponding analytical predictions from PN and EOB theory. However, the identification of a conserved energy or angular momentum within our setup is not obvious as our equations are not derived from a Hamiltonian; we will return this issue in Sec. IV. Instead, we follow Ref. [28] and compute the accumulated orbital phase as a function of the orbital frequency. This quantity includes both dissipative and conservative effects, and can be computed within different modeling frameworks.

The NR orbital phase is defined in terms of the waveform data as follows:

$$\phi^{\rm orb}_{\rm NR}(t) = \tfrac{1}{2} \arg h_{22}(t). \tag{16}$$

We also compute the orbital phase from a recently developed precessing EOB model (`SEOBNRv4PHM`) [54], a numerical relativity surrogate model (`NRSur7dq4`) [55], and our UDE model. Both `NRSur7dq4` and `SEOBNRv4PHM` are considered state-of-the-art and have been used by the LIGO-Virgo Collaboration to analyze recent gravitational-wave observations [13,56,57]. We represent each orbital phase by a degree 3 spline (using a smoothing factor of 0.0002), from which we

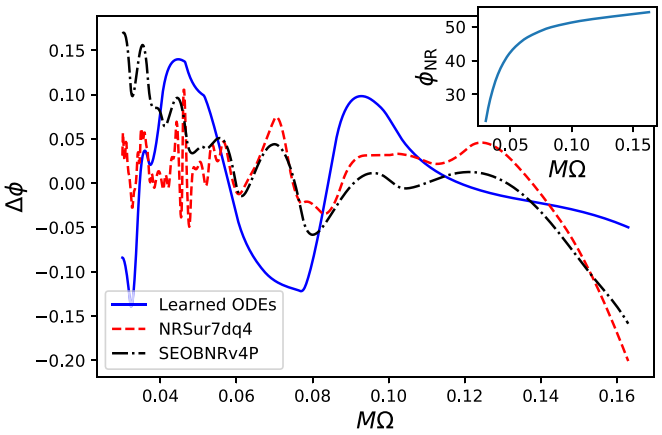

FIG. 4. Comparison of the accumulated orbital phase as a function of orbital frequency for the second experiment III B 1, an equal mass quasicircular inspiral of nonspinning black holes. The orbital phasing obtained from an NR simulation is shown in the inset figure. We show the difference between the NR data and (i) our UDE orbital model (solid blue line), (ii) `NRSur7dq4`, a numerical relativity surrogate model (dashed-red line), and (iii) `SEOBNRv4PHM`, a recently developed precessing EOB model (dash-dot black line). All three models show good agreement with the NR orbital phasing. The UDE (learned ODE) model is competitive with these two state-of-the-art models despite being built in a very different way.

compute the orbital frequency,

$$\Omega = \frac{d\phi^{\rm orb}}{dt}, \tag{17}$$

by taking a derivative of the spline, and finally forming the function $\phi^{\rm orb}(\Omega) = \phi^{\rm orb}(t(\Omega))$. We then compare $\phi^{\rm orb}_{\rm NRSur7dq4}$, $\phi^{\rm orb}_{\rm SEOBNRv4P}$, and $\phi^{\rm orb}_{\rm UDE}$ to $\phi^{\rm orb}_{\rm NR}$. For each comparison, we form the difference, $\Delta\phi = \phi^{\rm orb}_{\rm model} - \phi^{\rm orb}_{\rm NR}$ after phase alignment [58]. Figure 4 shows $\Delta\phi$ for each model. Over the range of orbital frequencies shown, the L_2-error in the phase is 2.5×10^{-2} (UDE model), 2.0×10^{-2} (`NRSur7dq4`), and 2.2×10^{-2} (`SEOBNRv4PHM`). All three models do an excellent job at tracking the NR orbital phase throughout the late inspiral phase to $\Omega \approx 0.16$, which is when a common apparent horizon appears in the NR simulation.

One of the most important practical uses of a dynamical model is as an intermediate step towards generating gravitational waveforms. While a full study is outside the scope of this paper, we provide a preliminary look at this here. We compute an L_2-type error measurement (see Eq. 21 from Ref. [50]) between the complexified NR waveform and each model's prediction of the waveform after optimizing for phase and time alignments. We find the errors to be 3.1×10^{-3} (UDE model), 1.2×10^{-3} (`SEOBNRv4PHM`), and 1.1×10^{-5} (`NRSur7dq4`). All three models have been calibrated to $q = 1$ NR waveform data, so this comparison is only meant to be suggestive of how well the modeling techniques can perform on the training set and not its generalization error.

We note that all three models used in our comparisons have been built in very different ways. The `SEOBNRv4PHM` model is a highly sophisticated analytical model thats been under investigation for two decades [29] while the `NRSur7dq4` model was trained against 1528 NR simulations using numerical techniques that have been in development for nearly a decade [59,60]. By comparison, our UDE model is new and our modelization choices are simple. Given that our UDE model is able to perform comparably well against state-of-the-art models demonstrates the potential of waveform inversion as a new tool for model builders to consider in future work.

2. *Eccentric orbits from noisy GW measurements*

This experiment proceeds in much the same way as the previous one. Here, however, we learn the dynamics of an eccentric binary black hole system whose trajectories are depicted in the top left-hand corner of Fig. 5 (solid-black lines) with the associated waveform data in the bottom panel (black dots). Unlike the previous experiments, we do not assume known values for the initial conditions e_0, p_0, or χ_0 but instead make these part of the learning process. We continue to adopt the initial conditions $\phi_0 = 0$. As can be seen in Fig. 5, we introduce additive Gaussian noise to the waveform data of the form $w(t_i) + n(t_i)$, where $n(t_i)$ is draw from a normal distribution of mean 0 and standard deviation of $\sigma = 10^{-2}$. As the typical waveform amplitude is $\sim$0.1, this corresponds to a coefficient of variation of around $\sigma/0.1 = 0.1$.

In spite of these challenges, we are still able to recover the original trajectories as indicated by the visual agreement between the learned (red and blue) and NR (black) trajectories shown in the top left-hand panel of Fig. 5. This is achieved by simultaneously optimizing for both e_0 or χ_0 in Eq. (10), in addition to the neural network parameters $\boldsymbol{\xi}$, and deducing the associated value of p_0 directly from Eq. (4). The model also recovers important general relativistic effects, notably, as in the previous example, radiation-reaction effects. In this case, the tendency for the orbit to circularize, $e(t) \to 0$, is seen in the upper right-hand panel of Fig. 5. As before, the eccentricity has an inflection point around $p \approx 6$ and quickly grows thereafter.

As follow-on to this experiment, we test the stability of the learned solution to different signal-to-noise ratios (SNRs). Specifically, we use the learned solution parameters $\boldsymbol{\xi}$, e_0, and χ_0 from Fig. 5 as initial guesses in a set of model discovery problems that fit the original `SXS:BBH:1356` waveform data, but have white Gaussian noise of different variance superimposed. After solving each of these model discovery problems, we measure how much the learned waveform differs from the true waveform. The curve in Fig. 6 shows the relative error in the learned waveform as a function of the SNR in the waveform measurement. The relative error grows as the SNR decreases, however, the results are surprisingly accurate down to SNR $\approx$32. We note that while all BBH gravitational-wave detections to-date have had SNRs below 32 [13], future detectors such as LISA [61], Cosmic Explorer [62], or Einstein Telescope [63] will routinely detect events with an SNR greater than 32.

IV. DISCUSSION

We have presented a data-driven gravitational waveform inversion strategy, which generates mechanical models of binary black hole systems. We start with a structurally very simple set of universal differential equations and parametrize

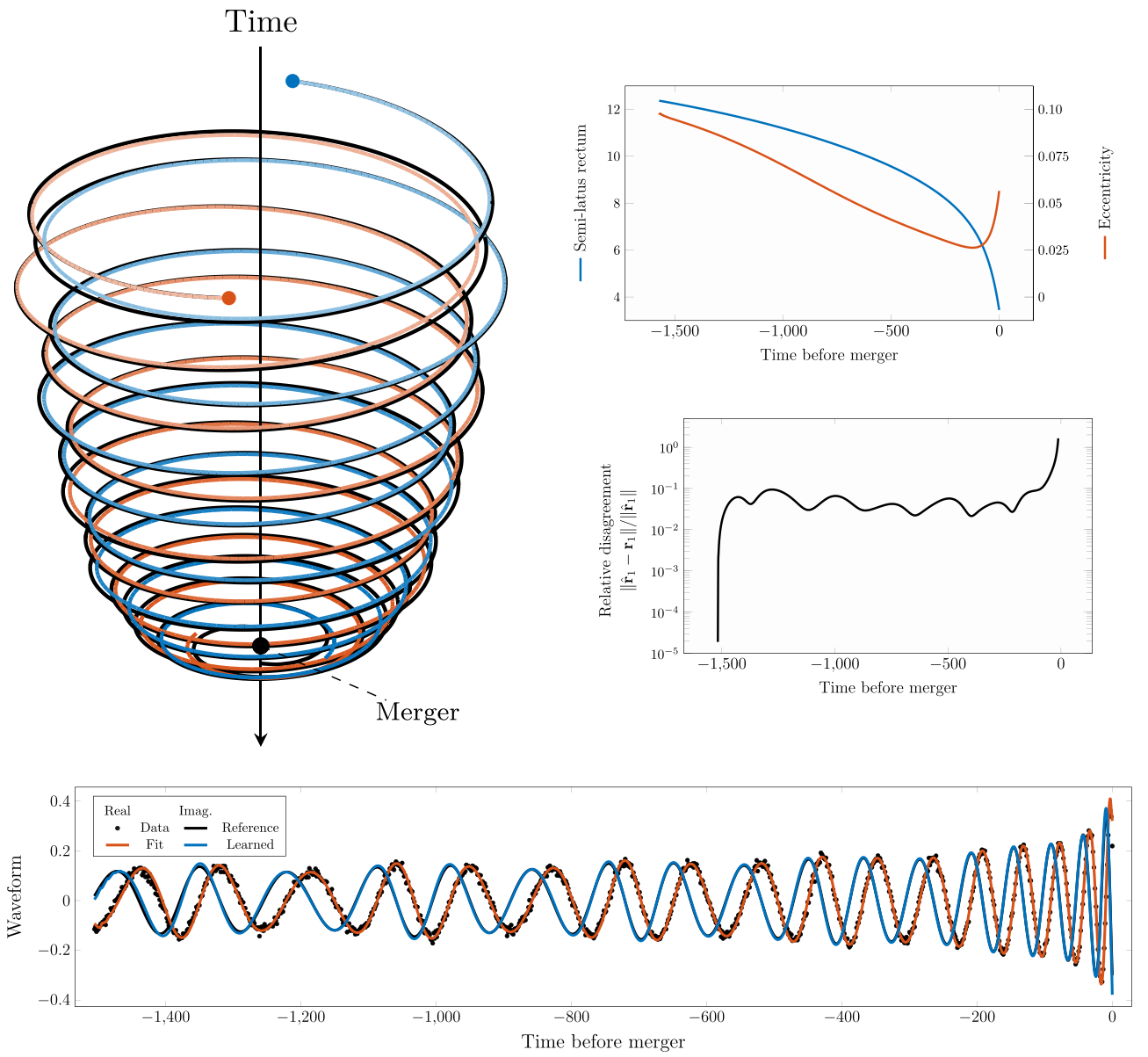

FIG. 5. Summary of our third experiment, where have used noisy gravitational-wave observations (black dots; bottom figure) to learn the underlying dynamical system model governing the relativistic two-body problem for two equal mass black holes in an eccentric orbit. Top left: Trajectories of the centers of mass of a black hole system `SXS:BBH:1356` (black lines), taken from the SXS Gravitational Waveform Database [38]. We also show the orbit computed from our learned dynamical system (blue and red lines). In the upper-right panel, we show the evolution of the eccentricity and semilatus rectum from our learned-dynamical system. The middle right panel shows the disagreement between the NR trajectories and the ones computed from the learned dynamical system. We caution the reader that this figure should not be understood as a relative error because the numerical relativity black hole trajectories and our learned model are expressed in different coordinate systems that are impossible to relate. Bottom: Learned (red line) and computational relativity (black dots) waveform data corresponding to the real part of gravitational waveforms. Here, we also include the learned imaginary part of the waveform (blue line), reconstructed with the quadrupole formula, and compare it with the reference imaginary part taken from the SXS database (black line).

the space of models with feed-forward neural networks. Our differential equations are trained by solving a physics-informed constrained optimization problem that seeks to minimize the waveform error. We tested our method on various BBH systems including extreme and comparable mass ratio systems in eccentric and noneccentric orbits, and train on portions of the waveform corresponding to orbital plunge right up to the time of merger. We find that the resulting differential equations agree remarkably well with the black hole trajectories computed through purely numerical means. Our models can be extrapolated in time and recover various known relativistic effects despite these being previously unknown to the universal differential equations. The main contribution of our paper is to show that two-body relativistic models can be deduced from gravitational wave measurements.

To describe the computational framework, we have focused on simple choices such as our ansatz neural ODE model Eq. (5) and the quadrupole formula for computing the

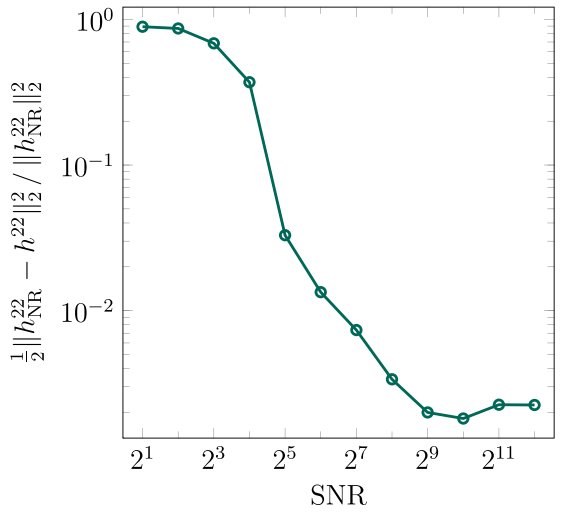

FIG. 6. We quantify the method's robustness to increasing amounts of Gaussian noise by comparing the true NR waveform, $h^{22}_{\rm NR}$, to the learned waveform, h^{22}. The amount of noise is quantified by the signal's SNR. We compute the waveform error according to $\frac{1}{2}\|h^{22}_{\rm NR} - h^{22}\|^2_2 \,/\, \|h^{22}_{\rm NR}\|^2_2$, which is related to the overlap error (cf. Ref. [50]) commonly used in gravitational-wave data analysis.

harmonic modes. These modeling choices are distinct from the overall computational framework and can be easily swapped out for different choices. For example, future high-accuracy studies should seek better orbit-to-waveform maps as the quadrupole formula is likely a large source of systematic error.

Our framework for learning the dynamics of binary black holes is quite general, and we expect that it can be applied to a variety of cases we have not considered including unequal masses, aligned-spins systems, and precessing systems. Our method should be especially useful for discovering equations of motion for systems where traditional approaches are less well-developed including eccentric binaries, the highly relativistic late-inspiral and plunge dynamical regimes, and beyond-GR theories. Given the close agreement with NR trajectories, other possible applications could include setting NR initial data whereby the neural ODEs could be used to predict an NR trajectory before the simulation is performed.

One of the most important applications of our technique may involve calibrating existing orbital dynamics models (and high-accuracy gravitational-wave models) by using a base model different from Eq. (5). Given that all modern inspiral-merger-ringdown waveform models require calibration of unknown parameters to numerical relativity simulations, the waveform inversion technique described here could benefit these efforts. For example, if implemented within the effective-one-body (EOB) approach, a suitably modified version of our methodology could be used to calibrate for missing terms in the EOB Hamiltonian. Precessing NR surrogate models also require a dynamical model, which is found through a direct fitting for the right-hand-side of the relevant ODEs. This computationally costly step might be simplified and accelerated with our techniques.

Another potential use of our methodology is training dynamical models entirely from gravitational-wave data sets instead of solving or analyzing Einstein's equation of general relativity as is traditionally done. In Fig. 6 we explore how well the algorithm performs as the signal's SNR is systematically varied. We find that, at least for the examples considered here, the method continues to work down to an SNR of about 32. Consequently, our method is most applicable for future GW detections, including those made with LISA [61], Cosmic Explorer [62], Einstein Telescope [63], or perhaps the upcoming LIGO-Virgo-Kagra science run. Our method will need to be modified to achieve SNR levels of approximately 10 (which would cover most GW detections to date [13]) without compromising accuracy. Some approaches could include: (i) comparing waveforms with the Wasserstein metric, which is known to be more robust to phase trappings, which shows up in higher noise, (ii) detecting an ensemble of noisy signals, and training on this observation set, (iii) using Bayesian networks where the network parameters are probability functions, or (iv) to apply our technique to filtered waveforms using a model-agnostic approach such as wavelet methods or denoising methods. We leave such extensions for future work.

ACKNOWLEDGMENTS

We thank Alvin Chua, Sebastian Khan, Gaurav Khanna, Tom O'Leary-Roseberry, and Harald Pfeiffer for helpful discussions. S.E.F. is supported by NSF Grants No. PHY-1806665 and No. DMS-1912716. This paper was written while the authors were in residence at the Institute for Computational and Experimental Research in Mathematics (ICERM) in Providence, RI, during the Advances in Computational Relativity program, supported by the National Science Foundation under Grant No. DMS-1439786. This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344, LLNL-JRNL-819108.

APPENDIX: METHODS

In this section, we elaborate on the technical elements of our work, which are necessary for replication of the results. We open with a brief overview of the adjoint sensitivity method [32], which we used to compute derivatives of $\mathcal{J}$ [see Eqs. 9 and 12] with respect to $\boldsymbol{\xi}$ and, in turn, facilitate solving problem Eq. (10). The section then closes with an overview of the quadrupole formula we have used to model the gravitational waveform Eq. (8). Because both of these topics are well known in specific (but mainly disjoint) scientific communities, we keep the exposition brief but include numerous references to the literature.

1. Calculation of derivatives

The ODE-constrained optimization problem Eq. (10) delivers a calibrated dynamical system model

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x};\boldsymbol{\xi}), \quad \mathbf{x}(0) = \mathbf{x}_0, \tag{A1}$$

where $\mathbf{F}(\mathbf{x};\boldsymbol{\xi}) = \mathbf{f}(\mathbf{x}, \mathcal{F}(\mathbf{x};\boldsymbol{\xi}))$. Because $\mathbf{x} = \mathbf{x}(\boldsymbol{\xi})$ depends implicitly on $\boldsymbol{\xi}$ through Eq. (A1), the main technical difficulty with solving such optimization problems lies in computing

total derivatives of the functional

$$\mathcal{J}(\mathbf{x},\boldsymbol{\xi}) = \frac{1}{T}\int_0^T J(\mathbf{x},\boldsymbol{\xi})\,dt, \tag{A2}$$

with respect to $\boldsymbol{\xi}$.

Indeed, assume that we are working with the definition of $\mathcal{J}$ given in Eq. (9). Here, any gradient-based optimization algorithm requires the calculation of the total derivative

$$d_{\boldsymbol{\xi}}\mathcal{J} = \frac{1}{T}\int_0^T \partial_{\mathbf{x}} J d_{\boldsymbol{\xi}}\mathbf{x} + \partial_{\boldsymbol{\xi}} J\,dt. \tag{A3}$$

One may easily note that, in the specific setting given to us through Eq. (9), we have $\mathcal{J} = \mathcal{J}(\mathbf{x})$ and, therefore, the partial derivative $\partial_{\boldsymbol{\xi}} J$ vanishes. In more general situations, the term $\partial_{\boldsymbol{\xi}} J$ is routine to derive. On the other hand, the calculation of $d_{\boldsymbol{\xi}}\mathbf{x}$ is problematic due to the fact that $\mathbf{x}(\boldsymbol{\xi})$ is not available in closed form. One approach to computing $d_{\boldsymbol{\xi}}\mathcal{J}$ involves directly estimating $d_{\boldsymbol{\xi}}\mathbf{x}$ via finite differences, however, the cost of this approach scales linearly with the size of $\boldsymbol{\xi}$. This makes it prohibitive for most practical problems. We choose to compute these gradients using in an alternative way often referred to as the adjoint sensitivity method [32].

The adjoint sensitivity method has been used extensively in engineering design [64,65] and, more recently, in machine learning research [4,8,66]. It involves the integration of two ODEs over the time interval $[0, T]$: the original governing ODE Eq. (A1) and an adjoint ODE (integrated backwards in time).

The derivation of this method usually begins with the Lagrangian

$$\mathcal{L} = \frac{1}{T}\int_0^T J(\mathbf{x}(t),\boldsymbol{\xi}) - \boldsymbol{\lambda}(t)^\top(\dot{\mathbf{x}}(t) - \mathbf{F}(\mathbf{x}(t);\boldsymbol{\xi}))\,dt - \boldsymbol{\mu}^\top(\mathbf{x}(0) - \mathbf{x}_0), \tag{A4}$$

which comes from writing the $\boldsymbol{\xi}$ minimization of Eq. (A2), constrained by solutions of Eq. (A1), as a saddle-point problem [67]. This functional is clearly designed such that $d_{\boldsymbol{\xi}}\mathcal{J} = d_{\boldsymbol{\xi}}\mathcal{L}$ for any $\mathbf{x}$ satisfying Eq. (A1). In addition, one observes that

$$d_{\boldsymbol{\xi}}\mathcal{L} = \partial_{\mathbf{x}}\mathcal{L}\,d_{\boldsymbol{\xi}}\mathbf{x} + \partial_{\boldsymbol{\lambda}}\mathcal{L}\,d_{\boldsymbol{\xi}}\boldsymbol{\lambda} + \partial_{\boldsymbol{\mu}}\mathcal{L}\,d_{\boldsymbol{\xi}}\boldsymbol{\mu} + \partial_{\boldsymbol{\xi}}\mathcal{L} \tag{A5}$$

$$= \frac{1}{T}\int_0^T \partial_{\boldsymbol{\xi}} J(\mathbf{x}) - \boldsymbol{\lambda}^\top \partial_{\boldsymbol{\xi}}\mathbf{F}(\mathbf{x};\boldsymbol{\xi})\,dt, \tag{A6}$$

if $\partial_{\boldsymbol{\lambda}}\mathcal{L} = \partial_{\boldsymbol{\mu}}\mathcal{L} = \partial_{\mathbf{x}}\mathcal{L} = 0$. A straightforward calculation shows that the first two of these equations are equivalent to the original dynamical system Eq. (A1). On the other hand, $\partial_{\mathbf{x}}\mathcal{L} = 0$ is equivalent to the *adjoint equation*

$$-\dot{\boldsymbol{\lambda}} = [\partial_{\mathbf{x}}\mathbf{F}(\mathbf{x};\boldsymbol{\xi})]^\top\boldsymbol{\lambda} + \partial_{\mathbf{x}} J(\mathbf{x},\boldsymbol{\xi}), \quad \boldsymbol{\lambda}(T) = \mathbf{0}. \tag{A7}$$

Evidently, the ODE system Eq. (A7) depends on the solution to Eq. (A1), $\mathbf{x}(t)$. Therefore, the algorithm for computing $d_{\boldsymbol{\xi}}\mathcal{J}$ must follow a specific order:

(1) Integrate $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x};\boldsymbol{\xi})$ from $t = 0$ to T with the initial condition $\mathbf{x}(0) = \mathbf{x}_0$.

(2) Integrate $-\dot{\boldsymbol{\lambda}} = [\partial_{\mathbf{x}}\mathbf{F}(\mathbf{x};\boldsymbol{\xi})]^\top\boldsymbol{\lambda} + \partial_{\mathbf{x}} J(\mathbf{x},\boldsymbol{\xi})$ from $t = T$ to 0 with the initial condition $\boldsymbol{\lambda}(T) = \mathbf{0}$.

(3) Compute $d_{\boldsymbol{\xi}}\mathcal{J} = \frac{1}{T}\int_0^T \partial_{\boldsymbol{\xi}} J(\mathbf{x}) - \boldsymbol{\lambda}^\top\partial_{\boldsymbol{\xi}}\mathbf{F}(\mathbf{x};\boldsymbol{\xi})\,dt$.

Extension of this algorithm to the scenario where the initial condition $\mathbf{x}_0$ is also optimized for (as considered in, e.g., Sec. III B 1) is straightforward. For thorough accounts of the numerical implementation of the adjoint sensitivity method for UDEs, we refer the interested reader to [8,68].

2. Gravitational waves from an orbit

In the context of general relativity, gravitational waves are associated with the outgoing, radiative parts of the spacetime metric and are solutions to Einstein field equation. The motion of massive objects produce gravitational waves, and our model outlined in Sec. II B provides the equations of motion for object 1 of mass m_1 and position $\mathbf{r}_1(t)$, and object 2 of mass m_2 and position $\mathbf{r}_2(t)$.

The quadrupole formula provides one simple method of obtaining the gravitational radiation from these orbital trajectories. In this framework, we assume both black holes to be "point sources" (i.e., Dirac delta functions). The Newtonian mass density of two objects orbiting in the x-y plane is

$$\rho(t,x,y,z) = m_1\delta(z)\delta(x - x_1(t))\delta(y - y_1(t)) + m_2\delta(z)\delta(x - x_2(t))\delta(y - y_2(t)), \tag{A8}$$

and note that in the special case $m_2 \gg m_1$ we have $\mathbf{r}_2(t) = (0,0,0)$ and $\mathbf{r}_1(t) = \mathbf{r}(t)$. Given the density, the quadrupole formula tells us that the dominant quadrupole mode of the gravitational radiation field tensor in the transverse-traceless gauge is

$$rH^{ab} = 2\frac{\partial^2}{\partial t^2}\mathcal{I}^{ab}, \tag{A9}$$

where the trace-free mass quadrupole tensor is

$$\mathcal{I}^{ab} = I^{ab} - \tfrac{1}{3}\delta^{ab}\delta_{cd}I^{cd}, \tag{A10}$$

δ^{ab} is the Kronecker delta. In Cartesian coordinates, the indices take on values of "x", "y", and "z". For example, we have H^{xx}, H^{yy}, H^{xy}, etc. The components of the mass quadrupole tensor, I^{ab}, that are relevant to Eq. (A9) (nonzero temporal derivatives) are

$$I^{xx} = \int d^3x\rho x^2 = m_1x_1(t)^2 + m_2x_2(t)^2, \tag{A11a}$$

$$I^{yy} = \int d^3x\rho y^2 = m_1y_1(t)^2 + m_2y_2(t)^2, \tag{A11b}$$

$$I^{xy} = \int d^3x\rho xy = m_1x_1(t)y_1(t) + m_2x_2(t)y_2(t), \tag{A11c}$$

and by symmetry $I^{xy} = I^{yx}$.

This framework computes the gravitational wave as perturbations, H^{ab}, of the background spacetime metric. However, numerical simulations instead provide the plus h_+ and cross $h_\times$, gravitational-wave polarizations defined on a "sphere at infinity", which are obtained after contracting H^{ab} with the polarization tensors [30]. On this sphere, it is common to define a complexified gravitational wave

$$h(t,\theta,\phi) = h_+(t,\theta,\phi) - \mathrm{i}h_\times(t,\theta,\phi) \tag{A12}$$

$$= \sum_{\ell=2}^{\infty}\sum_{m=-l}^{l} h^{\ell m}(t)\,{}_{-2}Y_{\ell m}(\theta,\phi), \tag{A13}$$

and subsequently expand h into a complete basis of spin $=-2$ weighted spherical harmonics labeled by (ℓ, m) harmonic indices, ${}_{-2}Y_{\ell m}$. Here θ and ϕ are the polar and azimuthal angles. For example, the SXS (2,2) mode gravitational waveform data, h^{22}, was used extensively in this paper. Given the orbital trajectories, the (2,2) mode,

$$h^{22}(t) = \frac{1}{r}\sqrt{\frac{4\pi}{5}}(\ddot{\mathcal{I}}_{xx} - 2i\ddot{\mathcal{I}}_{xy} - \ddot{\mathcal{I}}_{yy}), \tag{A14}$$

can be computed directly from the trace-free mass quadrupole tensor [30, Eqs. 54–56].

To summarize, from the orbital motion we compute the three components of the mass quadrupole tensor Eq. (A11), compute the trace-free mass quadrupole tensor, compute the time derivatives using finite differences, and then finally assemble the (2,2)-multipolar component of the outgoing gravitational waves, h^{22}, from Eq. (A14).

While a full discussion is outside the scope of this Appendix, we point out that the quadrupole formula is the simplest possible one and, consequently, ignores a lot of the relevant physics. Future work could consider incorporating more physics into the gravitational waveform model, including relativistic definitions of the density, higher-order post-Minkowskian corrections, subdominant harmonic modes, or near-field-to-far-field transport maps. Nevertheless, some of missing physics might already be accounted for through the orbital model, where the deep networks may try to account for missing physics in the waveform model by modifying the orbital dynamics model.

[1] J. P. Crutchfield and B. S. McNamara, Equations of motion from a data series, Complex Syst. **1**, 417 (1987).

[2] S. L. Brunton, J. L. Proctor, and J. N. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems, Proc. Natl. Acad. Sci. USA **113**, 3932 (2016).

[3] M. Raissi and G. E. Karniadakis, Hidden physics models: Machine learning of nonlinear partial differential equations, J. Comput. Phys. **357**, 125 (2018).

[4] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud, Neural ordinary differential equations, in *Advances in Neural Information Processing Systems*, Vol. 31, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Curran Associates, Inc., Red Hook, NY, 2018) pp. 6571–6583.

[5] Z. Long, Y. Lu, X. Ma, and B. Dong, PDE-net: Learning PDEs from data, in *International Conference on Machine Learning* (PMLR, 2018) pp. 3208–3216.

[6] J. Han, C. Ma, Z. Ma, and E. Weinan, Uniformly accurate machine learning-based hydrodynamic models for kinetic equations, Proc. Natl. Acad. Sci. USA **116**, 21983 (2019).

[7] K. Wu and D. Xiu, Numerical aspects for approximating governing equations using data, J. Comput. Phys. **384**, 200 (2019).

[8] C. Rackauckas, Y. Ma, J. Martensen, C. Warner, K. Zubov, R. Supekar, D. Skinner, and A. Ramadhan, Universal differential equations for scientific machine learning, arXiv:2001.04385.

[9] D. Z. Huang, K. Xu, C. Farhat, and E. Darve, Learning constitutive relations from indirect observations using deep neural networks, J. Comput. Phys. **416**, 109491 (2020).

[10] Y. Sun, L. Zhang, and H. Schaeffer, NeuPDE: Neural network based ordinary and partial differential equations for modeling time-dependent data, in *Proceedings of The First Mathematical and Scientific Machine Learning Conference*, Proceedings of Machine Learning Research, Vol. 107, edited by J. Lu and R. Ward (PMLR, Princeton University, Princeton, NJ, USA, 2020) pp. 352–372.

[11] A. Yazdani, L. Lu, M. Raissi, and G. E. Karniadakis, Systems biology informed deep learning for inferring parameters and hidden dynamics, PLoS Comput. Biol. **16**, e1007575 (2020).

[12] R. Dandekar, V. Dixit, M. Tarek, A. Garcia-Valadez, and C. Rackauckas, Bayesian neural ordinary differential equations, arXiv:2012.07244.

[13] R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, A. Adams, C. Adams, R. Adhikari, V. Adya, C. Affeldt *et al.*, GWTC-2: Compact Binary Coalescences Observed by LIGO and Virgo during the First Half of the Third Observing Run, Phys. Rev. X **11**, 021053 (2021).

[14] L. Lehner and F. Pretorius, Numerical relativity and astrophysics, Annu. Rev. Astron. Astrophys. **52**, 661 (2014).

[15] L. Blanchet, Gravitational radiation from post-Newtonian sources and inspiralling compact binaries, Living Rev. Relativity **17**, 2 (2014).

[16] C. R. Vogel, *Computational Methods for Inverse Problems*, Frontiers in applied mathematics (Society for Industrial and Applied Mathematics, Philadelphia, 2002).

[17] S. Arridge, P. Maass, O. Öktem, and C.-B. Schönlieb, Solving inverse problems using data-driven models, Acta Numerica **28**, 1 (2019).

[18] E. Poisson and C. M. Will, *Gravity: Newtonian, Post-Newtonian, Relativistic* (Cambridge University Press, Cambridge, 2014).

[19] S. Chandrasekhar, *The Mathematical Theory of Black Holes*, Vol. 69 (Oxford University Press, Oxford, 1998).

[20] A. G. Lewis, A. Zimmerman, and H. P. Pfeiffer, Fundamental frequencies and resonances from eccentric and precessing binary black hole inspirals, Class. Quantum Grav. **34**, 124001 (2017).

[21] A. Zimmerman, A. G. M. Lewis, and H. P. Pfeiffer, Redshift Factor and the First Law of Binary Black Hole Mechanics in Numerical Simulations, Phys. Rev. Lett. **117**, 191101 (2016).

[22] A. Le Tiec, A. H. Mroue, L. Barack, A. Buonanno, H. P. Pfeiffer, N. Sago, and A. Taracchini, Periastron Advance in Black-Hole Binaries, Phys. Rev. Lett. **107**, 141101 (2011).

[23] A. Le Tiec, E. Barausse, and A. Buonanno, Gravitational Self-Force Correction to the Binding Energy of Compact Binary Systems, Phys. Rev. Lett. **108**, 131103 (2012).

[24] A. Le Tiec, The overlap of numerical relativity, perturbation theory and post-Newtonian theory in the binary black hole problem, Int. J. Mod. Phys. D **23**, 1430022 (2014).

[25] A. Le Tiec, A. Buonanno, A. H. Mroue, H. P. Pfeiffer, D. A. Hemberger, G. Lovelace, L. E. Kidder, M. A. Scheel, B. Szilágyi, N. W. Taylor, and S. A. Teukolsky, Periastron advance in spinning black hole binaries: Gravitational self-force from numerical relativity, Phys. Rev. D **88**, 124027 (2013).

[26] N. E. Rifat, S. E. Field, G. Khanna, and V. Varma, Surrogate model for gravitational wave signals from comparable and large-mass-ratio black hole binaries, Phys. Rev. D **101**, 081502(R) (2020).

[27] A. Pound, B. Wardell, N. Warburton, and J. Miller, Second-Order Self-Force Calculation of Gravitational Binding Energy, Phys. Rev. Lett. **124**, 021101 (2020).

[28] M. van de Meent and H. P. Pfeiffer, Intermediate Mass-Ratio Black Hole Binaries: Applicability of Small Mass-Ratio Perturbation Theory, Phys. Rev. Lett. **125**, 181101 (2020).

[29] A. Buonanno and T. Damour, Effective one-body approach to general relativistic two-body dynamics, Phys. Rev. D **59**, 084006 (1999).

[30] N. T. Bishop and L. Rezzolla, Extraction of gravitational waves in numerical relativity, Living Rev. Relativity **19**, 2 (2016).

[31] J. Nocedal and S. Wright, *Numerical optimization* (Springer Science & Business Media, Berlin, 2006).

[32] V. Boltyanskiy, R. V. Gamkrelidze, Y. Mishchenko, and L. Pontryagin, *Matematicheskaya teoriya optimal'nykh protsessov (Mathematical Theory of Optimal Processes)* (Nauka, Leningrad, 1961).

[33] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, Julia: A fresh approach to numerical computing, SIAM Rev. **59**, 65 (2017).

[34] B. Keith, A. Khadse, and S. E. Field, Three examples of learning orbital dynamics of binary black hole systems from gravitational wave measurements with Julia (2021), http://doi.org/10.5281/zenodo.4477649.

[35] C. Cutler, D. Kennefick, and E. Poisson, Gravitational radiation reaction for bound motion around a Schwarzschild black hole, Phys. Rev. D **50**, 3816 (1994).

[36] K. Martel, Gravitational waveforms from a point particle orbiting a Schwarzschild black hole, Phys. Rev. D **69**, 044025 (2004).

[37] S. E. Field, J. S. Hesthaven, and S. R. Lau, Discontinuous Galerkin method for computing gravitational waveforms from extreme mass ratio binaries, Class. Quantum Grav. **26**, 165010 (2009).

[38] M. Boyle, D. Hemberger, D. A. Iozzo, G. Lovelace, S. Ossokine, H. P. Pfeiffer, M. A. Scheel, L. C. Stein, C. J. Woodford, A. B. Zimmerman *et al.*, The SXS collaboration catalog of binary black hole simulations, Class. Quantum Grav. **36**, 195006 (2019).

[39] E. Poisson, A. Pound, and I. Vega, The motion of point particles in curved spacetime, Living Rev. Relativity **14**, 7 (2011).

[40] The Spectral Einstein Code, http://www.black-holes.org/SpEC.html.

[41] J. W. York, Jr., Conformal "Thin Sandwich" Data for the Initial-Value Problem, Phys. Rev. Lett. **82**, 1350 (1999).

[42] H. P. Pfeiffer and J. W. York, Jr., Extrinsic curvature and the Einstein constraints, Phys. Rev. D **67**, 044022 (2003).

[43] L. Lindblom, M. A. Scheel, L. E. Kidder, R. Owen, and O. Rinne, A New generalized harmonic evolution system, Class. Quantum Grav. **23**, S447 (2006).

[44] O. Rinne, L. T. Buchman, M. A. Scheel, and H. P. Pfeiffer, Implementation of higher-order absorbing boundary conditions for the Einstein equations, Class. Quantum Grav, **26**, 075009 (2009).

[45] SXS Collaboration, The SXS collaboration catalog of gravitational waveforms, http://www.black-holes.org/waveforms.

[46] M. Benning and M. Burger, Modern regularization methods for inverse problems, Acta Numerica **27**, 1 (2018).

[47] B. Szilagyi, L. Lindblom, and M. A. Scheel, Simulations of binary black hole mergers using spectral methods, Phys. Rev. D **80**, 124010 (2009).

[48] M. Boyle, L. E. Kidder, S. Ossokine, and H. P. Pfeiffer, Gravitational-wave modes from precessing black-hole binaries, arXiv:1409.4431.

[49] V. Varma, L. C. Stein, and D. Gerosa, The binary black hole explorer: On-the-fly visualizations of precessing binary black holes, Class. Quantum Grav. **36**, 095007 (2019).

[50] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, D. A. Hemberger, P. Schmidt, and R. Smith, A surrogate model of gravitational waveforms from numerical relativity simulations of precessing binary black hole mergers, Phys. Rev. D **95**, 104023 (2017).

[51] J. Blackman, S. E. Field, M. A. Scheel, C. R. Galley, C. D. Ott, M. Boyle, L. E. Kidder, H. P. Pfeiffer, and B. Szilágyi, Numerical relativity waveform surrogate model for generically precessing binary black hole mergers, Phys. Rev. D **96**, 024058 (2017).

[52] T. Damour, A. Nagar, D. Pollney, and C. Reisswig, Energy Versus Angular Momentum in Black Hole Binaries, Phys. Rev. Lett. **108**, 131101 (2012).

[53] A. Nagar, T. Damour, C. Reisswig, and D. Pollney, Energetics and phasing of nonprecessing spinning coalescing black hole binaries, Phys. Rev. D **93**, 044046 (2016).

[54] S. Ossokine, A. Buonanno, S. Marsat, R. Cotesta, S. Babak, T. Dietrich, R. Haas, I. Hinder, H. P. Pfeiffer, M. Pürrer, C. J. Woodford, M. Boyle, L. E. Kidder, M. A. Scheel, and B. Szilagyi, Multipolar effective-one-body waveforms for precessing binary black holes: Construction and validation, Phys. Rev. D **102**, 044055 (2020).

[55] V. Varma, S. E. Field, M. A. Scheel, J. Blackman, D. Gerosa, L. C. Stein, L. E. Kidder, and H. P. Pfeiffer, Surrogate models for precessing binary black hole simulations with unequal masses, Phys. Rev. Research **1**, 033015 (2019).

[56] R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. Adhikari, V. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, A. Aich, L. Aiello, A. Ain, P. Ajith, S. Akcay, G. Allen, A. Allocca *et al.*, GW190521: A Binary Black Hole Merger with a Total Mass of 150 m, Phys. Rev. Lett. **125**, 101102 (2020).

[57] R. Abbott, T. Abbott, S. Abraham, F. Acernese, K. Ackley, C. Adams, R. Adhikari, V. Adya, C. Affeldt, M. Agathos, K. Agatsuma, N. Aggarwal, O. D. Aguiar, A. Aich, L. Aiello, A. Ain, P. Ajith, S. Akcay, G. Allen, A. Allocca *et al.*, GW190412: Observation of a binary-black-hole coalescence with asymmetric masses, Phys. Rev. D **102**, 043015 (2020).

[58] We allow for $\phi_{\rm model} \to \phi_{\rm model} + c$, for some constant c, which is equivalent to a rotation in the orbital plane.

[59] S. E. Field, C. R. Galley, F. Herrmann, J. S. Hesthaven, E. Ochsner, and M. Tiglio, Reduced Basis Catalogs for Gravitational Wave Templates, Phys. Rev. Lett. **106**, 221102 (2011).

[60] S. E. Field, C. R. Galley, J. S. Hesthaven, J. Kaye, and M. Tiglio, Fast Prediction and Evaluation of Gravitational Waveforms Using Surrogate Models, Phys. Rev. X **4**, 031006 (2014).

[61] P. Amaro-Seoane, H. Audley, S. Babak, J. Baker, E. Barausse, P. Bender, E. Berti, P. Binetruy, M. Born, D. Bortoluzzi, J. Camp, C. Caprini, V. Cardoso, M. Colpi, J. Conklin, N. Cornish, C. Cutler, K. Danzmann, R. Dolesi, L. Ferraioli *et al.*, Laser interferometer space antenna, arXiv:1702.00786.

[62] B. P. Abbott *et al.* (LIGO Scientific), Exploring the sensitivity of next generation gravitational wave detectors, Class. Quantum Grav. **34**, 044001 (2017).

[63] S. Hild, M. Abernathy, F. Acernese, P. Amaro-Seoane, N. Andersson, K. Arun, F. Barone, B. Barr, M. Barsuglia, M. Beker *et al.*, Sensitivity studies for third-generation gravitational wave observatories, Class. Quantum Grav. **28**, 094013 (2011).

[64] A. Jameson, Aerodynamic design via control theory, J. Sci. Comput. **3**, 233 (1988).

[65] M. B. Giles and N. A. Pierce, An introduction to the adjoint approach to design, Flow, Turbul. Combust. **65**, 393 (2000).

[66] F. d. A. Belbute-Peres, T. Economon, and Z. Kolter, Combining differentiable PDE solvers and graph neural networks for fluid flow prediction, in *International Conference on Machine Learning* (PMLR, 2020) pp. 2402–2411.

[67] I. Ekeland and R. Temam, *Convex Analysis and Variational Problems*, Classics in applied mathematics (Society for Industrial and Applied Mathematics, Philadelphia, 1999).

[68] A. Gholami, K. Keutzer, and G. Biros, ANODE: Unconditionally accurate memory-efficient gradients for neural ODEs, in *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence* (IJCAI-19) (2019) pp. 730–736.