

Machine learning forecasts of the cosmic distance duality relation with strongly lensed gravitational wave events

Rubén Arjona^{1,*}, Hai-Nan Lin^{2,†}, Savvas Nesseris^{1,‡} and Li Tang^{3,2,§}

¹*Instituto de Física Teórica UAM-CSIC, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain*

²*Department of Physics, Chongqing University, Chongqing 401331, China*

³*Department of Math and Physics, Mianyang Normal University, Mianyang 621000, China*



(Received 9 November 2020; accepted 14 April 2021; published 11 May 2021)

We use simulated strongly lensed gravitational wave events from the Einstein telescope to demonstrate how the luminosity and angular diameter distances, $d_L(z)$ and $d_A(z)$, respectively, can be combined to test in a model independent manner for deviations from the cosmic distance duality relation and the standard cosmological model. In particular, we use two machine learning approaches, the genetic algorithms and Gaussian processes, to reconstruct the mock data and we show that both approaches are capable of correctly recovering the underlying fiducial model and can provide percent-level constraints at intermediate redshifts when applied to future Einstein telescope data.

DOI: [10.1103/PhysRevD.103.103513](https://doi.org/10.1103/PhysRevD.103.103513)

I. INTRODUCTION

The first detection of gravitational waves (GWs) by the LIGO/Virgo Collaboration was not just limited to the discovery of new astrophysical objects, but it was also instrumental in furthering our understanding of the fundamental properties of gravity and cosmology, by providing tests of gravity in the strong field regime. These observations from black hole and neutron star mergers have, figuratively speaking, shone a light on the population of compact objects in the Universe and the mechanism by which they are formed [1], given some of the most rigorous direct tests to date of general relativity (GR) [2,3], and provided the first measurement of a cosmological parameter, the Hubble constant H_0 , using GW sources [4].

Moreover, the observation of the binary neutron star merger GW170817 availed us of the opportunity to test gravity both in the strong regime and at large scales, as it was followed by the nearly simultaneous detection of its optical counterpart and allowed us to strongly constrain the GW propagation speed to $|c_g - c| \lesssim \mathcal{O}(10^{-15})$ [5], thus challenging a wide range of modified gravity scenarios which are candidates to explain the current acceleration of the Universe [6–11]. The impact of GW observations will be further extended by third generation ground based detectors like the Einstein telescope (ET) [12] and the space-based interferometer LISA [13].

Similarly to photons, GWs can be gravitationally lensed by the presence of galaxies and clusters of galaxies, producing a deflection in their trajectories, thus generating multiple detection events. This phenomenon is quite intriguing because the clustered matter that lies in between the GW source and the observer can enhance the observed signal [14]. This in turn can cause the luminosity distance to the source and therefore its redshift, if combined with the Hubble parameter H_0 constraints [15], to be underestimated. Sequentially, this would lead to an overestimation of the chirp mass [16].

With the upgraded sensitivity of the third generation of GW detectors, such as the Einstein telescope (ET), the detection sensitivity of the GW events would be accordingly improved. Thus, with the sufficiently large number of detectable events foreseen [17,18], it is expected that some of these events could be gravitationally lensed, thus allowing the creation of a considerably large catalogue of strongly lensed GWs event within a few years of operation. For an extensive analysis on how GW lensing is enriched with concrete signatures and features and can be used to search for deviations of GR, see Ref. [19].

As the GW passes through near massive astrophysical objects, its path would be modified producing gravitational lensing [20–22]. Since its first proposal [23], efforts have been placed to search for signatures of gravitational lensing in binary black hole events from current detectors such as LIGO and Virgo [24] but with no strong evidence of this effect [25]. However, as the sensitivity of the detectors improve further, it is plausible to observe lensing effects with future detectors such as aLIGO, the ET [26], and the space-based detector LISA [27]. In Ref. [28] the authors

*ruben.arjona@uam.es

†linhn@cqu.edu.cn

‡savvas.nesseris@csic.es

§tang@cqu.edu.cn

improved previous analysis of GW lensing events by including effects created by the ellipticity of lensing galaxies, lens environments, and magnification bias. Indeed, these observations could lead to new applications in astrophysics, cosmology, and fundamental physics [29–35].

One of the advantages of strongly lensed GW events comes from their ability to provide simultaneous measurements of both the luminosity and angular diameter distance, i.e., $d_L(z)$ and $d_A(z)$, respectively, which in turn could be used to probe fundamental properties of the standard cosmological model. One such example of a possible probe is the cosmic distance duality relation (DDR), also known as the Etherington relation, which relates the luminosity distance to the angular diameter distance at any redshift z via [36]

$$d_L(z) = (1+z)^2 d_A(z), \quad (1)$$

which is valid for any metric theory of gravity like GR and under the condition that the number of gravitons or photons, depending in which context it is applied, is conserved and that they travel along null geodesics in a pseudo-Riemannian spacetime [37]. At this point we can introduce the duality parameter

$$\begin{aligned} \eta(z) &\equiv \frac{d_L(z)}{(1+z)^2 d_A(z)} \\ &\equiv (1+z)^{\epsilon(z)}, \end{aligned} \quad (2)$$

where $\eta(z)$ is a function that accounts for possible deviations from unity and is equal to unity when the DDR holds, while in the last line we have introduced a phenomenological parameter $\epsilon(z)$, usually assumed to be constant, i.e., $\epsilon(z) \simeq \epsilon_0 = \text{constant}$.

Hence, any violation of the DDR relation at any redshift, i.e., $\eta(z) \neq 1$ or $\epsilon_0 \neq 0$, would be a hint of new physics, which in the case of photons could be caused by different mechanisms, such as the annihilation of photons by the intergalactic dust [38], the coupling of photons with other particles like axions [39], and the variation of fundamental constants [40]. In fact several works have been devoted to test the DDR relation [37,41–55].

As mentioned earlier, in our analysis ϵ_0 is a phenomenological parameter that parametrizes deviations from the standard DDR, i.e., any values that are different from zero imply a deviation from the standard model. Clearly, any such deviations will be small, as otherwise they would be immediately obvious in a plethora of observations, including strong and weak lensing. Indeed, in Ref. [52] it was shown that current Type Ia supernovae (SnIa) and baryon acoustic oscillations (BAO) data constrain the parameter to be $\epsilon_0 = 0.013 \pm 0.029$ and that there is no evidence for a redshift evolution of $\epsilon(z)$. The latter was shown by splitting the data in two bins, in $0 < z < 0.9$ and $z \geq 0.9$, and testing if both bins give consistent results for the reconstruction of

the parameter ϵ_0 in each bin. In particular, it was found that the values of ϵ_0 in both bins were the same, note however that this analysis was made using data in small redshifts ($z < 1.5$), so deviations might be present at higher redshifts.

Furthermore, in Ref. [52] it was found that future large scale structure surveys like Euclid will be able to improve upon the constraints on ϵ_0 from currently available BAO and type Ia supernovae (SnIa) by a factor of six. In particular, current BAO and SnIa data provide a constraint of $\epsilon_0 = 0.013 \pm 0.029$, while Euclid will improve this to $\epsilon_0 = -0.0008 \pm 0.0049$, which is a tighter constraint on ϵ_0 by a factor of six [52].

Since $\epsilon(z)$ relates two geometric variables, i.e., the luminosity and angular diameter distances, then strong lensing (either with light or GWs) is ideally suited to constrain it, while in the case of weak lensing the effect is less clear, but as was shown in Ref. [52], the bulk of the constraining power will come from improving the bounds on Ω_m , thus breaking the degeneracies between Ω_m and ϵ .

In this paper we show how to reconstruct the DDR relation using mock datasets of strongly lensed GWs, as they allow us to measure both the angular diameter and luminosity distance is complementary to the approach of Ref. [56], where $\eta(z)$ was constrained using mocks of strongly lensed SnIa, based on the Large Synoptic Survey Telescope survey. Both methods have the advantage of allowing for measurements of the duality parameter without relying on multiple datasets, hence it is competitive with other more traditional tests of the DDR relation where the latter is constrained through the combination of SnIa and BAO observations, as for example it has been forecast for future surveys [52].

In Ref. [57] the authors proposed a novel method to test the cosmic distance duality relation using the strongly lensed GWs from the Einstein telescope and in Ref. [58] mock data points were generated for this ground-based detector, while a parametrized approach was used to constrain the DDR relation. Here we present a broader analysis by presenting a slightly different methodology which allows us to directly make robust $\eta(z)$ mocks, based on the mocks of d_L and d_A and then we use genetic algorithms (GA) and Gaussian processes (GP), two non-parametric and symbolic regression subclasses of machine learning methods, to reconstruct $\eta(z)$ directly without any underlying model.

The parametric and nonparametric methods, like the GA, were extensively compared using mock data in Ref. [52], where it was shown that the two approaches are consistent with each other, albeit the errors in the reconstructed quantities are slightly larger for the GA due to its non-parametric nature. On the other hand, a model for the duality parameter η , based on axion physics was studied in Ref. [53]. Since axions couple to the standard model and photons, it is expected that some of the axions will be converted to photons and vice versa, thus leading to a

surplus or deficit of photons. Since the DDR assumes the photon number conservation, this implies axions lead to a violation of the DDR and an duality parameter which is different from unity. Using mock data it was shown in Ref. [53], that both the GA and the GP can consistently reconstruct the cosmological distances and the duality parameter η in agreement with each other and the fiducial model, within the errors, thus we are confident for our reconstruction methods.

Here we follow the approach of Ref. [53], especially as we also use mock GW events, albeit we assume that they can be lensed so that we also extract the angular diameter distance. Using the mock data, where we know the fiducial cosmology, allows us to assess the quality of the fit and determine whether the GA and the GP can successfully determine the underlying model.

Our paper is organized as follows: In Sec. II we describe the methodology to generate the ET mock data points and the machine learning (ML) implementation. In Sec. III we present our reconstructions for the GP and GA, while in Sec. IV we summarize our conclusions.

II. METHODOLOGY

A. Strongly lensed GW events

1. The angular diameter distance d_A from strong lensing

Here we will now consider the case when a GW emission is strongly lensed by a foreground galaxy, whose mass profile can be modeled by the singular isothermal sphere (SIS) model [59]. We will assume however that GWs propagate following geometric optics,¹ i.e., we neglect wave effects; see Ref. [22] for more details. With this setup then we assume the two images will appear at angular positions θ_1 and θ_2 with respect to the lensing galaxy. See Fig. 1 the geometrical illustration of the lensing system. Thus, the Einstein radius $\theta_E = |\theta_1 - \theta_2|/2$ will be given by [59]

$$\theta_E = \frac{4\pi\sigma_{\text{SIS}}^2 d_A(z_l, z_s)}{c^2 d_A(z_s)}, \quad (3)$$

where the velocity dispersion of the lens galaxy is given by σ_{SIS} , the angular diameter distances from the observer to the source and from the lens to the source are given by $d_A(z_s)$ and $d_A(z_l, z_s)$, respectively, z_l and z_s are the redshifts of the lens and source, respectively. We can rearrange Eq. (3) to obtain the distance ratio, which will be given by

¹The frequency of a gravitational wave produced by the merger of neutron star-black hole (NS-BH) or neutron star-neutron star (NS-NS) binary, is about several hundred Hertz and the corresponding wavelength ($\sim 10^6$ m) is much smaller than the scale of lens galaxy (\sim kpc). Hence, it is unnecessary to consider the wave effect.

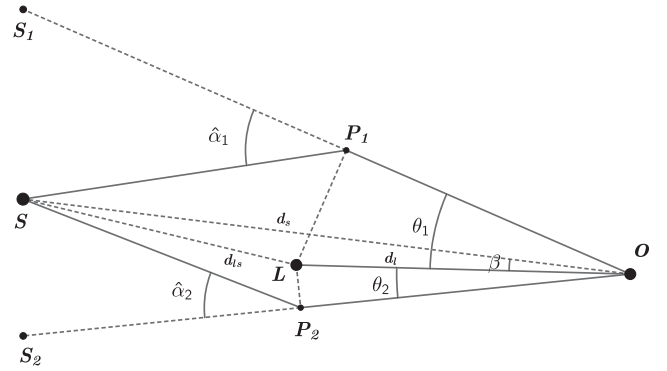


FIG. 1. The geometry of gravitational lensing. The source (S) at redshift z_s is strongly lensed by a foreground galaxy (L) at redshift z_l . The observer (O) sees two images (S_1 and S_2) at angular positions θ_1 and θ_2 , respectively. The actual angular position of the source with respect to the line-of-sight from observer to lens is β .

$$R_A \equiv \frac{d_A(z_l, z_s)}{d_A(z_s)} = \frac{c^2 \theta_E}{4\pi\sigma_{\text{SIS}}^2}. \quad (4)$$

If the angular positions and the velocity dispersion are well measured, which would require a precise localization of the GW sources that should be achievable with a network of interferometers,² then we can obtain the distance ratio R_A from Eq. (4).

As the two images propagating along different paths will take different amounts of time to reach Earth, then the time delay between the images will be given by [59]

$$\Delta t = \frac{(1+z_l) d_A(z_l) d_A(z_s)}{c d_A(z_l, z_s)} \Delta\phi, \quad (5)$$

where

$$\Delta\phi = \frac{(\theta_1 - \beta)^2}{2} - \Psi(\theta_1) - \frac{(\theta_2 - \beta)^2}{2} + \Psi(\theta_2), \quad (6)$$

is the Fermat potential difference between two paths, β is the actual angular position of the source, and $\Psi(\theta)$ is the rescaled projected gravitational potential of the lens galaxy. From equation (5) we can thus obtain the time-delay distance

$$D_{\Delta t} \equiv \frac{d_A(z_l) d_A(z_s)}{d_A(z_l, z_s)} = \frac{c}{1+z_l} \frac{\Delta t}{\Delta\phi}. \quad (7)$$

If the gravitational potential of the lens galaxy can be measured from photometric and spectroscopic observations,

²The angular separation between the images in a typical strongly lensing system is about several arcseconds [60]. In order to identify the images, the angular resolution of GW detectors should be better than arcseconds. A larger network of interferometers is necessary in order to reach such a high accuracy.

and if the time delay between two images can be well measured, we can obtain the time-delay distance $D_{\Delta t}$ according to Eq. (7).

In a spatially flat universe, the angular diameter distance from lens to source, $d_A(z_l, z_s)$, can be expressed in terms of $d_A(z_s)$ and $d_A(z_l)$ as [57]

$$d_A(z_l, z_s) = d_A(z_s) - \frac{1 + z_l}{1 + z_s} d_A(z_l). \quad (8)$$

Equation (8), together with equations (4) and (7), allow us to uniquely solve for $d_A(z_s)$, which reads

$$d_A(z_s) = \frac{1 + z_l R_A D_{\Delta t}}{1 + z_s (1 - R_A)}, \quad (9)$$

where R_A and $D_{\Delta t}$ are given by equations (4) and (7), respectively. The error on $d_A(z_s)$ propagates from the errors on R_A and $D_{\Delta t}$. Using the standard error propagating formula, and assuming that R_A and $D_{\Delta t}$ are uncorrelated, we obtain,

$$\frac{\delta d_A(z_s)}{d_A(z_s)} = \sqrt{\left(\frac{\delta R_A}{R_A(1 - R_A)}\right)^2 + \left(\frac{\delta D_{\Delta t}}{D_{\Delta t}}\right)^2}, \quad (10)$$

where

$$\frac{\delta R_A}{R_A} = \sqrt{\left(\frac{\delta \theta_E}{\theta_E}\right)^2 + 4\left(\frac{\delta \sigma_{\text{SIS}}}{\sigma_{\text{SIS}}}\right)^2}, \quad (11)$$

and

$$\frac{\delta D_{\Delta t}}{D_{\Delta t}} = \sqrt{\left(\frac{\delta \Delta t}{\Delta t}\right)^2 + \left(\frac{\delta \Delta \phi}{\Delta \phi}\right)^2}. \quad (12)$$

In order for our method to work, we must independently measure the following observables: (z_l , z_s , Δt , $\Delta \phi$, θ_E , σ_{SIS}). If a GW event is accompanied by electromagnetic counterparts, the redshifts of the lens and source can be measured spectrometrically with negligible uncertainty, just as in the strongly lensed quasar or galaxy case [61,62]. The time delay between two images can be measured by comparing the light curves of two images at percentage level [63]. Especially, in the strongly lensed GW case, due to the transient property of GW signal, the time delay can be measured with negligible uncertainty. The difference of Fermat potentials $\Delta \phi$, and the velocity dispersion of lens galaxy σ_{SIS} , can be measured through photometric and spectroscopic observations of the lens galaxy [62,64]. The Einstein radius, $\theta_E = |\theta_1 - \theta_2|/2$, can be obtained by measuring the image positions θ_1 and θ_2 . The measurements of the latter three observables, however, may be uncertain. Following Ref. [65], we assume 0.6%, 1.0%, and 5.0% uncertainties on $\Delta \phi$, θ_E , σ_{SIS} , respectively.

Having obtained all the observables, $d_A(z_s)$ and its uncertainty can be derived from Eqs. (9)–(12). We see that the uncertainty on $d_A(z_s)$ mainly comes from the uncertainty on σ_{SIS} . The uncertainties on the rest observables will not strongly affect our results. To improve the accuracy, more accurate determinations of the velocity dispersion are required.

2. The luminosity distance d_L from GW signals

We now consider an unlensed GW source. In this case the luminosity distance to the source can be directly obtained by matching the GW signals to the GW templates. GW detectors based on the interferometers, such as ET, measure the change of difference of two optical paths caused by the pass of GW signals. In general, the response of a GW detector on GW signals will depend on the spacetime strain, which is the linear combination of the two polarization states $h_+(t)$ and $h_\times(t)$,

$$h(t) = F_+(\theta, \varphi, \psi)h_+(t) + F_\times(\theta, \varphi, \psi)h_\times(t), \quad (13)$$

where the beam-pattern functions $F_+(\theta, \varphi, \psi)$ and $F_\times(\theta, \varphi, \psi)$ do not only depend on the configuration of the detector, but they also depend on the position of the GW source (θ, φ) and the polarization angle ψ . For example, in the case of the ET, the beam-pattern functions can be found in Ref. [66].

In the post-Newtonian and stationary phase approximation, the strain $h(t)$ can be written in the Fourier space by [66,67]

$$\mathcal{H}(f) = \mathcal{A}f^{-7/6} \exp[i(2\pi f t_0 - \pi/4 + 2\psi(f/2) - \varphi_{(2,0)})], \quad (14)$$

where f is the GW frequency, t_0 is the time of merger. The explicit expressions of the phase terms $\psi(f/2)$ and $\varphi_{(2,0)}$ can be found in Ref. [68], but they are unimportant in our study here. The Fourier amplitude in Eq. (14) is given by

$$\begin{aligned} \mathcal{A} &= \frac{1}{d_L} \sqrt{F_+^2(1 + \cos^2 \iota)^2 + 4F_\times^2 \cos^2 \iota} \\ &\times \sqrt{\frac{5\pi}{96}} \pi^{-7/6} \mathcal{M}_c^{5/6}, \end{aligned} \quad (15)$$

where ι is the inclination angle, d_L is the luminosity distance, $\mathcal{M}_c = M\eta^{3/5}$ is the chirp mass, $M = m_1 + m_2$ is the total mass, $\eta = m_1 m_2 / M^2$ is the symmetric mass ratio, m_1 and m_2 are the component masses of the binary in the comoving frame. Here and after, we work in the natural units, so $c = G = 1$. In the case of a GW source at redshift z , \mathcal{M}_c in equation (15) should be interpreted as the chirp mass in the observer frame, which can be related to that of the comoving frame via $\mathcal{M}_{c,\text{obs}} = (1 + z)\mathcal{M}_{c,\text{com}}$. Finally, it should be noted that the exponential term on the

right-hand side of Eq. (14) is just a phase term, which is unimportant in our analysis.

The signal-to-noise ratio (SNR) of the detector's response to a GW signal is given by [67]

$$\rho_i = \sqrt{\langle \mathcal{H}, \mathcal{H} \rangle}, \quad (16)$$

where the inner product may be defined as

$$\langle a, b \rangle = 4 \int_{f_{\text{lower}}}^{f_{\text{upper}}} \frac{\tilde{a}(f)\tilde{b}^*(f) + \tilde{a}^*(f)\tilde{b}(f)}{2} \frac{df}{S_h(f)}, \quad (17)$$

and in the latter equation, \tilde{a} and a^* stand for the Fourier transformation and complex conjugation of a , respectively, while $S_h(f)$ is the one-side noise power spectral density (PSD) of ET, $f_{\text{lower}} = 1$ Hz and $f_{\text{upper}} = 2f_{\text{LSO}}$ are the lower and upper cutoffs of the frequency, $f_{\text{LSO}} = 1/(6^{3/2}2\pi M_{\text{obs}})$ is the orbit frequency at the last stable orbit, $M_{\text{obs}} = (1+z)(m_1 + m_2)$ is the total mass in the observer frame. Finally, the PSD for ET is given by [69]

$$S_h(f) = 10^{-50} (2.39 \times 10^{-27} x^{-15.64} + 0.349 x^{-2.145} + 1.76 x^{-0.12} + 0.409 x^{1.1})^2 \text{ Hz}^{-1}, \quad (18)$$

where x is the GW frequency in unit of 100 Hz, i.e., $x = f/100$ Hz. For the ET, three arms interfere with each other in pairs, hence the combined SNR is given by

$$\rho = \left[\sum_{i=1}^3 \rho_i^2 \right]^{1/2}. \quad (19)$$

In general, there is degeneracy between the luminosity distance d_L and inclination angle ι , so the uncertainty on d_L may be large. However, if the GW event is accompanied by a short gamma-ray burst (GRB), we can assume that the inclination angle is small, since GRB is expected to be produced in a narrow beam. In this case the degeneracy between d_L and ι breaks, and the uncertainty on d_L can be estimated as

$$\delta d_L = \sqrt{\left(\frac{2d_L}{\rho} \right)^2 + (0.05z d_L)^2}, \quad (20)$$

where the $0.05z$ term represents the uncertainty arising from weak lensing effect caused by the intergalactic medium along the line-of-sight.

The above discussion is applicable for unlensed GW events. However, the situation is subtle for strongly lensed GW events. Due to the magnification effect of lensing, the luminosity distance determined from the strongly lensed GW signals is not the true distance. Since the amplitude of GW signal \mathcal{A} is magnified by the lensing effect by a factor of $\sqrt{\mu_{\pm}}$ [70], and the luminosity distance d_L is inversely

proportional to \mathcal{A} , the true luminosity distance should be $d_L^{\text{true}} = \sqrt{\mu_{\pm}} d_L^{\text{obs}}$. If the magnification factor $\sqrt{\mu_{\pm}}$ can be independently determined through photometric observations, we can obtain the true distance d_L^{true} . The uncertainty of μ_{\pm} will also propagate to d_L . Therefore, the total uncertainty on $d_L(z_s)$ is given by

$$\frac{\delta d_L^{\text{total}}}{d_L} = \sqrt{\left(\frac{2}{\rho} \right)^2 + (0.05z_s)^2 + \frac{1}{4} \left(\frac{\delta \mu_{\pm}}{\mu_{\pm}} \right)^2}. \quad (21)$$

Due to the contamination of the image flux by the foreground lensing galaxy, the magnification factor is highly uncertain. Here we follow Ref. [65] and assume a 20% uncertainty on μ_{\pm} .

Theoretically, only the merger of NS-NS or NS-BH binaries may be accompanied by a short GRB, while the merger of BH-BH binary is expected to have no electromagnetic counterpart. Our method requires the direct measurement of source redshift, which is achievable only for NS-NS or NS-BH mergers. Unfortunately, according to numerical simulations [26], most of the lensed GW events are produced by the BH-BH merger. Without the redshift for the BH-BH events, they cannot be directly used to test DDR. If, however, the GW event can be precisely localized, it is possible to infer the redshift of GW source statistically [71], but will introduce additional uncertainty.

B. The mock DDR data points

1. The fiducial cosmological distances

In order to forecast direct measurements of the duality parameter $\eta(z)$ from the ET, we use mock distance data points based on individual measurements of the luminosity and angular diameter distances $d_L(z)$ and $d_A(z)$, respectively, as described previously. To join the two measurements and derive the $\eta(z)$ data points, we follow the pioneering work of Ref. [56] using a Markov Chain-Monte Carlo (MCMC) approach to create mock samples.

In a nutshell we can summarize this approach as follows. First, we assume a fiducial cosmology based on the cosmological constant Λ and cold dark matter (Λ CDM) model with a Hubble constant $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, a matter density parameter $\Omega_{m,0} = 0.3$, and assuming flatness ($\Omega_k = 0$). Then, based on the redshift distribution of sources, see Fig. 1 in Ref. [58] and Fig. 2 in Ref. [26] for either NS-NS or NS-BH, we calculate at every point in redshift the corresponding angular diameter distance $d_A(z)$ and the luminosity distance $d_L(z)$ via the methodology of Ref. [58] as described earlier. At this point, we also introduce a modification of the luminosity distance so that the observed luminosity distance would be proportional to the ‘‘bare’’ one as:

$$d_{L,\text{obs}}(z) = (1+z)^{\epsilon(z)} d_{L,\text{bare}}(z), \quad (22)$$

such that it corresponds to a duality parameter $\eta(z) = (1+z)^{\epsilon(z)}$, which should be equal to unity if no deviations are present, i.e., $\epsilon(z) \rightarrow 0$ in the Λ CDM model and $\eta(z) = 1$. In particular, we assume that to lowest order, any deviations are small enough that we can assume a constant $\epsilon(z) = \epsilon_0$. In general, any such deviations on the GW sector could be due to modifications of gravity, see for example Ref. [53]. Specifically, in what follows we will assume four specific scenarios: the vanilla Λ CDM case for $\epsilon_0 = 0$ and three more cases with one mild and two stronger deviations of the duality relation with $\epsilon_0 = (0.01, 0.05, 0.1)$.

As mentioned earlier, current supernovae and BAO data constrain the parameter to be $\epsilon_0 = 0.013 \pm 0.029$, and there is no evidence for a redshift evolution of $\epsilon(z)$, while that Euclid will be able to improve the constraints by a factor of six [52]. Thus, our choices of the ϵ_0 are realistic at the lower end ($\epsilon_0 \sim 0.01$) and high enough to sufficiently test our methodology at the higher end ($\epsilon_0 \sim 0.1$). In any case the constant $\epsilon(z) = \epsilon_0$ is the simplest ansatz used to test for deviations from the duality relation, and it does not really affect our analysis or our conclusions.

2. The mock samples of $\eta(z)$

After we have calculated the fiducial values of the cosmological distances, we can then make mock samples of the duality parameter $\eta(z)$ directly via the following procedure: First, at each redshift we create mock distances $d_A(z_s)$ and $d_L(z_s)$ based on a Gaussian distribution using the fiducial values and 1σ errors based on the methodology of Ref. [56], such that for the mock we have

$$(D_{i,\text{mock}}, \sigma_{i,\text{mock}}) \rightarrow \mathcal{N}(D_{i,\text{fid}}, \sigma_{i,\text{fid}}), \quad (23)$$

where $i = 1 \dots N_{\text{lens}}$, D_i represents either d_A or d_L , while $\sigma_{i,\text{fid}}$ are the errors, and $\mathcal{N}(\mu, \sigma)$ stands for a normal distribution with mean μ and standard deviation σ . Then, to make a mock sample of $\eta(z_i)$ values we can use Eq. (2) and finally, we employ an MCMC-like approach to obtain the mean values and the errors of the data points as follows:

- (1) Using the mock distances at each redshift $D_{i,\text{mock}}$ we draw 10,000 random samples from the assumed distribution for $D_{i,\text{mock}}$.
- (2) We then estimate $\eta(z_i)$ at each redshift z_i for each of the 10,000 random points using Eq. (2) to obtain 10,000 realizations of the distribution of $\eta(z_i)$.
- (3) We estimate the mean and standard deviation of $\log_{10} \eta(z_i)$ at each redshift point to create our final mock sample.

The main advantage of this MCMC-like approach is that it does not depend on error propagation for the various quantities, which could be highly nontrivial for complicated modified gravity models, but it also preserves the statistical properties of the samples. This allows us to obtain our results without any further dependence on the

cosmological model, and in Ref. [53] it was shown that this approach allows for the creation of mocks that have minimal external biases, theoretical, statistical, or otherwise. For example, this means that we no longer have to assume that the distributions of the $\log_{10} \eta(z_i)$ data points are sufficiently Gaussian, as implied by standard error propagation and which may bias the results by introducing artificial deviations from the fiducial model.

Finally, we should note that we chose to make mocks of $\log_{10} \eta(z_i)$ instead of simply $\eta(z_i)$, as we found that the distribution of the latter is somewhat non-Gaussian, while on the other hand, $\log_{10} \eta(z_i)$ is very close to being normally distributed around zero, i.e., $\log_{10} \eta(z_i) \sim \mathcal{N}(0, \sigma_{\log_{10} \eta(z_i)})$. Then, having created the $\log_{10} \eta(z_i)$ samples, we consider a likelihood \mathcal{L} of the form [56]:

$$-2 \ln \mathcal{L} = \sum_{i=1}^{N_{\text{lens}}} \left(\frac{\log_{10} \eta(z_i) - \log_{10} \eta^{\text{th}}(z_i)}{\sigma_{\log_{10} \eta(z_i)}} \right)^2, \quad (24)$$

where $\eta^{\text{th}}(z_i)$ is the theoretical value of $\eta(z_i)$. In our actual analysis we will consider the somewhat optimistic case of $N_{\text{lens}} = 100$ as an optimistic case for the possible number of events that could be detected in the coming decades.

C. Machine learning

Machine learning (ML) is a subset of artificial intelligence designed to model a given dataset. ML approaches have been proven to be successful at processing and extracting essential information from large amounts of data and can get rid of the problem of model bias [72], while also being very effective in testing the consistency of the dataset model independently and also for searching tensions or systematics.

In the context of GW physics, ML can be a good ally since it can tackle challenges that upcoming GW astronomy is fronting [73,74] such as a fast and systematic method to characterize properly the signal and the detector, accurate reconstructions of GW signals and a correct estimate of their statistical and systematic errors, and can help to improve and be more sensitive to different searching techniques such as matched filtering [75], cross-correlation methods [76], and time-coincident detection of coherent excess power between several detectors [77].

ML algorithms have been used to improve the sensitivity of ground-based GW detectors, to reduce and characterize nonastrophysical detector noise, and also it has been applied for fast determinations of parameter estimation [78,79]. See Ref. [80] for a review on several ML methods. ML is able to process GW signals on real-time, for example the algorithm named Deep Filtering [81] based on neural networks has been created for parameter estimation, reaching a similar performance compared to matched filtering but faster. In the last few years ML techniques have also been applied with success for glitch classification [82],

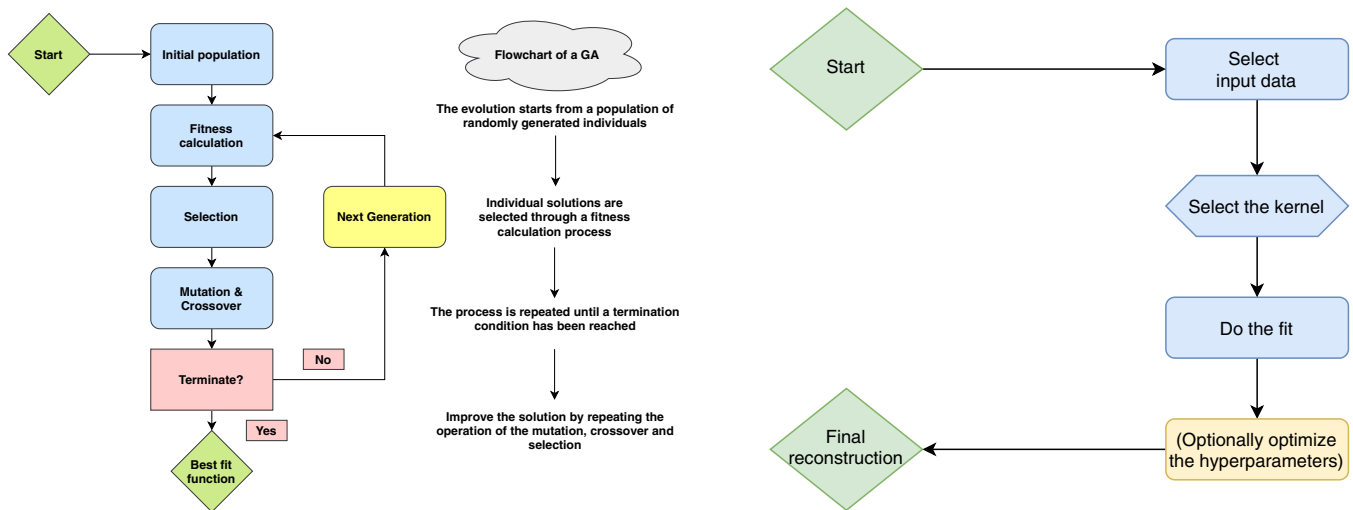


FIG. 2. Flowcharts of the list of steps of a usual genetic algorithm (left) and Gaussian process (right).

earthquake prediction [83] and to supplement existing Bayesian methods [84]. It is worth also mentioning the recent progress of neural networks in producing in a fast manner one and two dimensional marginalized Bayesian posteriors [85] for GW parameter estimation, showing how ML can give results very similar to Bayesian statistics [86].

Neural networks (NN) have been also tested on open data; for example in Ref. [87] the authors searched for a gravitational wave signal from an isolated neutron star from a remnant of GW170817. NN were applied as well in continuous gravitational waves from unknown spinning neutron stars [88] and for gravitational-wave transients associated with gamma-ray bursts [89]. Other applications, general reports, and reviews for the use of GW data analysis with ML can be found in Refs. [90–96].

In what follows we will describe two particular classes of ML methods, the genetic algorithms (GA) and the Gaussian processes (GP) which we use to perform our analysis. One of the advantages of the GA against other symbolic regression methods, such as neural networks, is that the GA provides analytical functions that describe the data provided. In our paper we have also used the GP to compare our results with the GA, but it is beyond the scope of our paper to compare all of the different symbolic regression ML approaches.

1. The genetic algorithms

Here we introduce the theoretical background of the implementation of the GA in our analysis. For a similar discussion on the GA and several applications to cosmology see Refs. [51,97–103]. The GA have been also applied in a wide range of areas such as particle physics [104–106], astronomy and astrophysics [107–109], and other fields like computational science, economics, medicine, and engineering [110,111]. For other symbolic regression methods applied in physics and cosmology see [112–119].

The GA is a particular class of ML methods mostly developed to perform unsupervised regression of data, which implies that the GA can be used for nonparametric reconstructions finding an analytic function that describes the data, using one or more variables. The unsupervised regression nature of the GA means that the algorithm tries to extract patterns and features on its own and without external help by using features and patterns on its own using unlabeled datasets. This is in contrast to supervised machine-learning where the algorithm is trained on a model to create good predictions based on the response to labeled datasets. On the other hand, the main goal of a classification problem is to identify to which class a new data point will fall under, given a training sample. In the left panel of Fig. 2 we present the flowchart of the GA and we will now proceed to discuss the implementation in detail.

The GA works by mimicking the notion of biological evolution via the concept of natural selection, as conveyed by the genetic operations of mutation and crossover. In essence, a set of test functions evolves over time through the effect of the stochastic operators of crossover, i.e., the joining of two or more candidate functions to form another one, and mutation, i.e., a random alteration of a candidate function. This process is consequently repeated thousands of times so as to ensure convergence, while different random seeds can be tested to further explore the functional space.

In what follows, we will briefly describe an example that will help illustrate these two operators. Given two functions $f_1(x) = 1 + x + x^2$ and $f_2(x) = \sin(x) + \cos(x)$, the mutation operation will randomly alter the various terms in the previous expressions, e.g., the outcome might be that the two functions have become $f_1(x) = 1 + 2x + x^2$ and $f_2(x) = \sin(x^2) + \cos(x)$, where in the former the coefficient of the second term was modified from one to two while in the latter $x = x^1$ was modified to x^2 . The crossover

TABLE I. The grammars used in the GA analysis. Other complicated forms are automatically produced by the mutation and crossover operations as described in the text.

Grammar type	Functions
Polynomials	$c, x, 1 + x$
Fractions	$\frac{x}{1+x}$
Trigonometric	$\sin(x), \cos(x), \tan(x)$
Exponentials	$e^x, x^x, (1+x)^{1+x}$
Logarithms	$\log(x), \log(1+x)$

operation, on the other hand, randomly merges the two functions to create two more, e.g., the outcome might be that the terms $1 + 2x$ from f_1 and $\cos(x)$ combined to $\tilde{f}_1(x) = 1 + 2x + \cos(x)$, while the rest of the terms combined to $\tilde{f}_2(x) = x^2 + \sin(x^2)$. For further details on the GA and some applications to cosmology see Refs. [97,99].

Since by construction the GA is a stochastic approach, the probability that a set or a population of functions will produce offspring is normally assumed to be proportional to its fitness to the data, which in our analysis is a χ^2 statistic and indicates how good every individual agrees with the data. For the data we use in our analysis, indeed we assume that the likelihoods are sufficiently Gaussian that we use the χ^2 in our GA approach. Note that in the GA both the probability to have offspring and the fitness of each individual is proportional to the likelihood. This then causes evolutionary pressure that favors the best-fitting functions in every population, thus driving the fit towards the minimum in a few generations.

In this analysis we reconstruct the DDR parameter $\log_{10} \eta(z_i)$ directly from the data, and the procedure to its reconstruction proceeded as follows. First, our predefined grammar consisted on the following orthogonal basis of functions: exp, log, polynomials, etc., and a set of operations $+, -, \times, \div$; see Table I for a complete list.

We also imposed a prior motivated by physical reasons. The only assumption made is that $\eta(z)$ is equal to 1 at the present time $z = 0$. This is natural to expect since mechanisms where the DDR is violated are cumulative, as photons interact with interceding constituents along the line of sight. Hence, such an event does not have time to occur at vanishing redshifts. This can also be seen by taking the limit for $z = 0$ at Eq. (2) and assuming the Hubble law, i.e., $\lim_{z \rightarrow 0} \eta(z) = 1$. Finally, we make no assumptions on the curvature of the Universe or any modified gravity or dark energy model.

We also required that all functions reconstructed by the GA are continuous and differentiable, without any singularities in the redshift probed by the data, avoiding in this manner overfitting or any spurious reconstructions. We emphasize that the choice of the grammar and the

population size has already been tested in Ref. [97].³ In the same manner, the seed numbers play also a big role since they are used to create the initial population of functions used later on by the GA.

Once the initial population has been constructed, the fitness of each member is evaluated by a χ^2 statistic, using the $\eta(z)$ data points directly as input. Afterwards, using a tournament selection, see Ref. [97] for more details, the best-fitting functions in each generation are chosen and the two stochastic operations (crossover and mutation) are used. In order to assure convergence, the GA process is then repeated thousands of times and with various random seeds, so as to properly explore the functional space. Then the final output of the code is a function of $\eta(z)$ that describes the evolution of the DDR.

Finally, the error estimates of the reconstructed function are determined via the path integral approach, which was originally implemented in Refs. [99,100]. This approach consists of having an analytical estimate of the error of the reconstructed quantity by calculating a path integral over all possible functions around the best fit GA that may contribute to the likelihood; and it has been shown that this can be performed whether the data points are correlated or uncorrelated. This error reconstruction method has been exhaustively examined and compared against a bootstrap Monte Carlo by Ref. [99]. Thus, given a reconstructed function $f(x)$ from the GA, the path integral approach of Ref. [99] provides us with the 1σ error $\delta f(x)$. This can be also compared to error propagation if one assumes that the error in a quantity is taken as $\sigma_f = f(p)\delta p$, where p would represent a parameter. We have extensively tested our approach finding that this assumption is appropriate for the data set used here.

2. The Gaussian processes

We also use the Gaussian processes approach in order to provide an alternative to the GA reconstruction and minimize any potential biases due to the reconstruction approach. Traditionally, a Gaussian process (GP) is defined as an ensemble of random variables that have a joint Gaussian distribution [120]. The GP in general is determined by the mean, usually assumed to be zero or some fiducial model, and the covariance. In our case the GP random variables stand for the duality parameter $\log_{10} \eta(z_i)$.

On the other hand, the covariance function, also known as a kernel, is denoted by $k(x, \tilde{x})$ and encodes the correlations of two different GP random variables denoted by x and \tilde{x} , which in our case correspond to the values of the duality parameter $\eta(z)$ at different values of z , i.e., two different data points of the data set. In summary then, the kernel is used to join up the data points in order to build a function. In practice, the kernel is related to the input data

³See for example Fig. 2 of Ref. [97].

as it is used as a measure of the similarity between points, i.e., a covariance function, and is used to predict the value for an unseen point from training data.

Lately the GPs have been used in the reconstruction of a plethora of cosmological datasets, see e.g., [121–130], while the proper choice of the kernel remains a hotly debated issue in the literature, as it can strongly affect the GP reconstruction. In Ref. [131] it was found that a kernel that works quite well for cosmological datasets is the so-called Matérn class of kernels, given by [120]:

$$k(x, \tilde{x}) = \sigma_M^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}(x - \tilde{x})}{\ell} \right)^\nu \times K_\nu \left(\frac{\sqrt{2\nu}(x - \tilde{x})}{\ell} \right), \quad (25)$$

where K_ν is a modified Bessel function, ν determines the shape of the covariance function, which asymptotes to the Gaussian limit as $\nu \rightarrow \infty$, while $\Gamma(\nu)$ is the gamma function.

Furthermore, the parameter ℓ describes the length scales over which the function varies, while the parameter σ_M corresponds to the magnitude of these variations. The parameter ν is further chosen to be a half integer to minimize the dependence on the Bessel function [131]. High values of ν make the GP smoother, but for $\nu \geq 7/2$ the results are practically indistinguishable from each other, so we make the choice $\nu = 5/2$. Overall, we find that altering either the GP kernels or ν does not impact the performance of the GP. In our analysis we use the GP PYTHON package GEORGE [132] to reconstruct $\log_{10} \eta(z)$ with the kernel as described above. Also note that in the GP the log-likelihood given by Eq. (24) is used, by maximizing it, to optimize the value of any hyperparameters in the kernel.

Finally, similarly to the GA case described in the previous section, we have also imposed a prior on the GP reconstructions which is motivated by physical reasons. Specifically, we again demand that $\eta(z)$ is equal to unity at out present time $z = 0$, i.e., $\eta(z = 0) = 1$. Again, this is necessary to ensure our reconstructions are physical, while at the same time keeping our analysis general enough.

III. RESULTS

We will now present the results of the reconstruction for both the GA and GP approaches.⁴ Note that in both cases the input data are the values of $\log_{10} \eta_i$ in the form of sets of points given by $(z, \log_{10} \eta_i, \sigma_{\log_{10} \eta_i})$, which are the inputs to the GP and GA and can be used to create a likelihood, as discussed in the previous point. In the case of the GA the data enter only via the likelihood of Eq. (24), while on the other hand the GP builds a function that essentially joins up

⁴Both codes are very efficient and it takes a few seconds for the GA and less than a second for the GP to converge, which is comparable to other traditional parametric approaches.

the data points according to the GP kernel, but also uses the data (via the likelihood) to optimize the value of any hyperparameters.

In Figs. 3 and 4 we show a particular realization of the $\log_{10} \eta(z_i)$ mocks for $N_{\text{lens}} = 100$, along with the fiducial model (dashed black line), the case of $\log_{10} \eta = 0$ (dashed green line), and the corresponding best-fit (solid colored line) for the GA and the GP for $\epsilon = (0, 0.01, 0.05, 0.10)$ in the top left, top right, bottom left, and bottom right panels, respectively. The data points are shown in the background as gray points with their 1σ error bars and the shaded band corresponds to the 1σ confidence region for the GA (red band) and the GP (magenta band) in the two plots.

Note that for the different values of ϵ for the mocks, we keep the same random seed number so that our analysis is not complicated by statistical fluctuations and the interpretation of our results is more straightforward. The apparent lack of events at high redshifts is due to the expected redshift distribution of the BH-NS and NS-NS events; see Fig. 1 in Ref. [58] and Fig. 2 in Ref. [26], which also has the side effect of increasing the errors of the reconstructions and making the fits increase with redshift at high z . The expected probability density of events as a function of redshift, i.e., the redshift distribution of the BH-NS and NS-NS events for the ET, was determined in Ref. [26] by using the intrinsic merger rates of these objects with the help of a population synthesis evolutionary code.

The only physical prior used in the reconstruction was the assumption that $\lim_{z \rightarrow 0} \eta(z) \rightarrow 1$, which follows naturally from the definition of $\eta(z)$ via Eq. (2) and the fact that, at zero redshift the causes of any deviation (for example either axions for light or modified gravity for the GW) have had no time to yet act, which is necessary to ensure that our results are physical.

As can be seen in Figs. 3 and 4, in all cases both the GA and the GP capture the behavior of the data points accurately and remain close to the fiducial model, well within the 1σ region. In particular, on average the difference between the GA or GP best-fit and the fiducial model remains close to a percent level in all reconstructions. Furthermore, for both ML approaches we find that the reconstructed errors are consistent with each other, thus we are confident in our reconstruction as the GA and the GP are in principle rather different reconstruction methods. In particular, we see in both the GA and GP cases that when $\epsilon = (0.05, 0.10)$ both ML approaches find a clear deviation from the null hypothesis, i.e., $\log_{10} \eta = 0$ (dashed green line) for $0 \leq z \leq 3.5$. For higher redshifts, due to the lack of points the errors of the reconstruction become larger and the statistical significance of the detection diminishes.⁵

⁵As a consistency test of our approach, we also fit the parametrization $\eta(z) = (1+z)^{\epsilon_0}$, with $\epsilon_0 = \text{constant}$, to the mock data and we discuss our results in the Appendix.

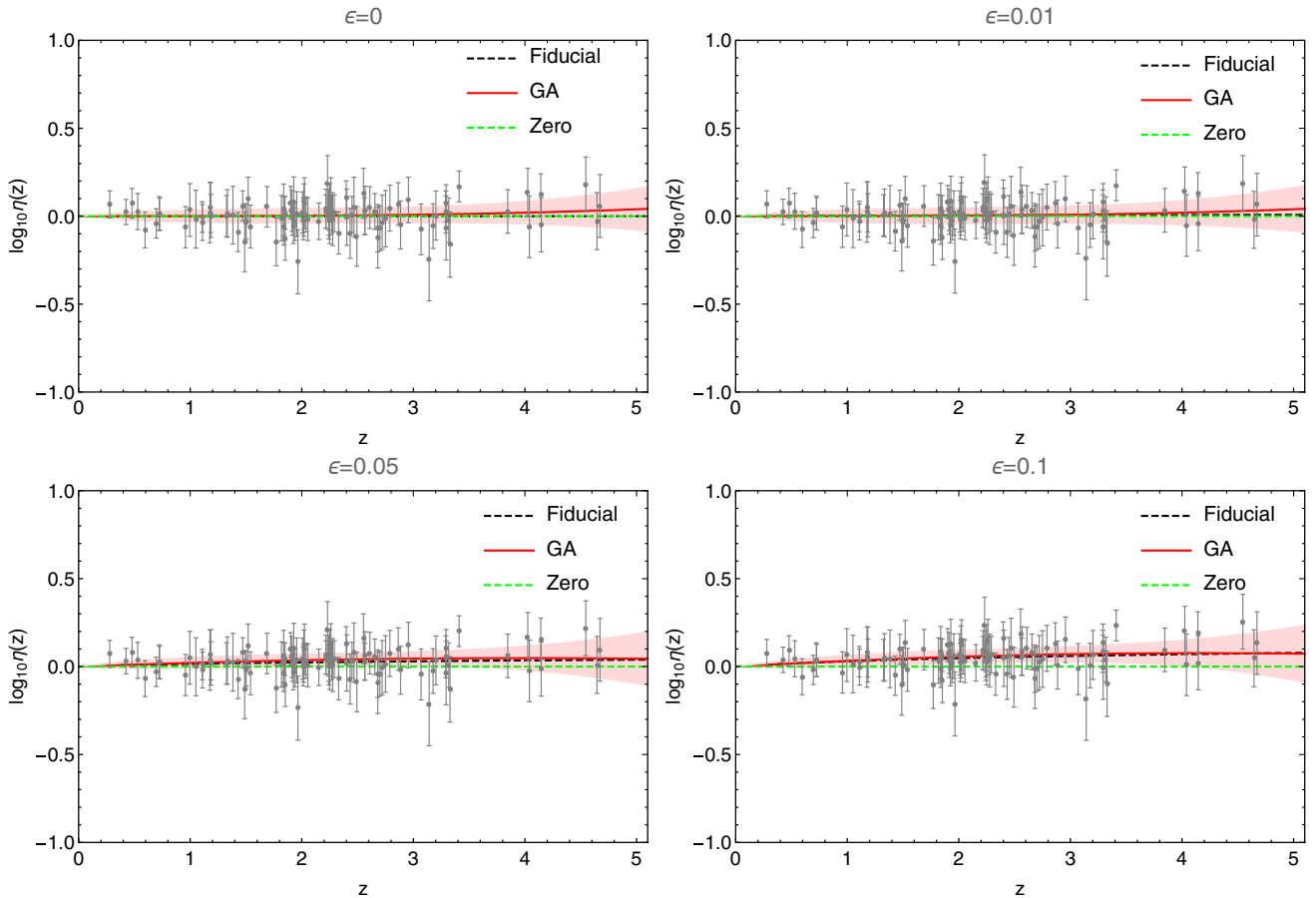


FIG. 3. The $\eta(z)$ mocks along with the fiducial model given by Eq. (2) (black dashed line), the case of $\log_{10} \eta = 0$ (dashed green line), the corresponding best-fit (solid colored line) for the GA for $\epsilon = (0, 0.01, 0.05, 0.10)$ in the top left, top right, bottom left, and bottom right panels, respectively. In all cases the $\eta(z)$ data points are shown in the background as gray points with their 1σ error bars and the shaded band corresponds to the 1σ confidence region for the GA (red band).

Finally, we also vary two key parameters of our analysis, the amplification error $\delta\mu/\mu$ and the number of events N . Using the GA, we reconstruct $\eta(z)$ with the amplification error $\delta\mu/\mu$ taking the values $[0.1, 0.2, 0.5]$ and then we do the same with the number of lenses with $N = [50, 100, 200]$ events for the case of $\epsilon = 0.1$. We show the results in Fig. 5 and we find that as expected, increasing the amplification error (top row of plots in Fig. 5) has no obvious effect when $\delta\mu/\mu$ changes from 0.1 to 0.2. As can be seen in Eq. (21) when $\rho = 16$, which is the critical value for which we assume to claim the detection of GW signal, then the $2/\rho$ term is about 0.1, the $0.05z$ term is also about 0.1 for $z \sim 2$ and is even larger for $z > 2$. Since there is a factor $1/4$ before the term $\delta\mu/\mu$, changing $\delta\mu/\mu$ from 0.1 to 0.2 does not affect significantly the total error on the luminosity distance $d_L(z)$. On the other hand, when $\delta\mu/\mu = 0.5$, then the last term in Eq. (21) dominates and this results in larger errors, by roughly $\sim 20\%$, for the GA reconstruction of $\eta(z)$, compared to when $\delta\mu/\mu$ is 0.1 or 0.2.

On the other hand, the effect of varying the number of lenses is more subtle. As can be seen in the bottom row of Fig. 5, for 50 lenses we have a deviation at $\sim 1\sigma$ below $z \sim 4$, while in the case of 100 lenses the errors of the GA reconstruction become smaller for $z < 4$ raising the deviation to slightly more than 1σ but at $z > 4$ surprisingly become larger again. This tightening of the errors at intermediate redshifts ($z < 4$) and enlarging at $z > 4$ as we increase the points from 50 to 100 is due to the redshift distribution of the points (for 100 points more events are located at $z < 3.5$). When we increase the number of events to 200, we see the error at high redshifts is now more uniform, even though the reconstructions start to get dominated by the systematic errors. In summary, the number of events necessary to obtain a statistically significant deviation depends both on the value of ϵ and the number of events, so for example with $N = 50$ we can have a 1σ deviation up to $z \sim 4$, while to go to higher redshifts ($z \sim 5$) we require $N = 200$.

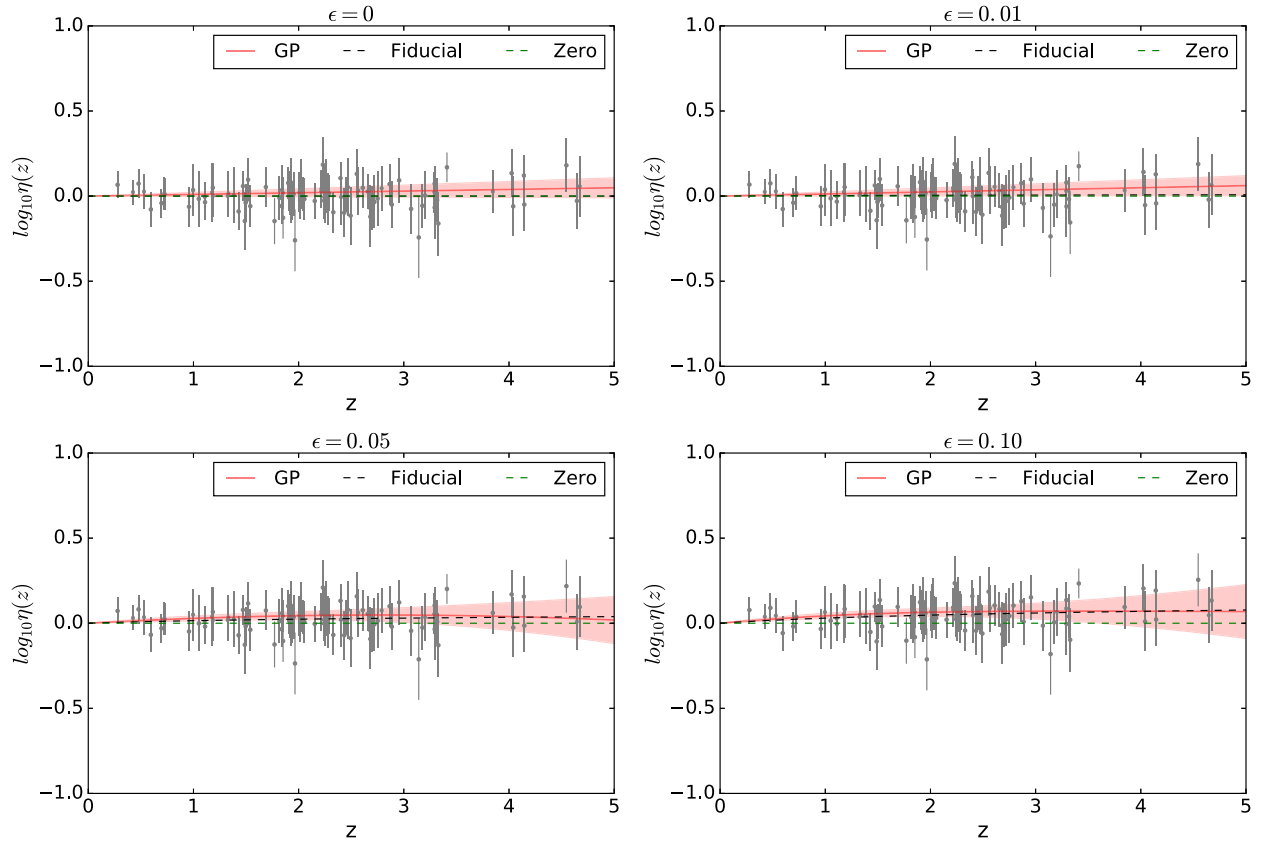


FIG. 4. The $\eta(z)$ mocks along with the fiducial model given by Eq. (2) (black dashed line), the case of $\log_{10}\eta = 0$ (dashed green line), the corresponding best-fit (solid colored line) for the GP for $\epsilon = (0, 0.01, 0.05, 0.10)$ in the top left, top right, bottom left, and bottom right panels, respectively. In all cases the $\eta(z)$ data points are shown in the background as gray points with their 1σ error bars and the shaded band corresponds to the 1σ confidence region for the GP (magenta band).

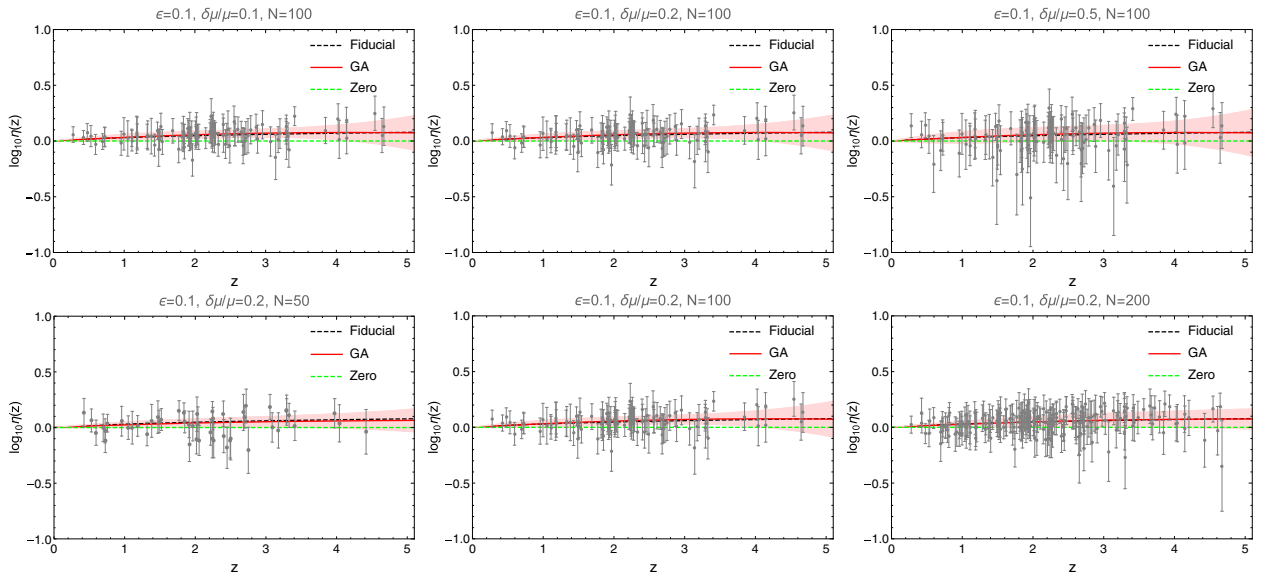


FIG. 5. The $\eta(z)$ mocks along with the fiducial model given by Eq. (2) (black dashed line), the case of $\log_{10}\eta = 0$ (dashed green line), the corresponding best-fit (solid colored line) for the GA for $(\epsilon = 0.10, \delta\mu/\mu = 0.1, N = 100)$, $(\epsilon = 0.10, \delta\mu/\mu = 0.2, N = 100)$, $(\epsilon = 0.10, \delta\mu/\mu = 0.5, N = 100)$, $(\epsilon = 0.10, \delta\mu/\mu = 0.2, N = 50)$, $(\epsilon = 0.10, \delta\mu/\mu = 0.2, N = 100)$, $(\epsilon = 0.10, \delta\mu/\mu = 0.2, N = 200)$ in the top left, top center, top right, bottom left, bottom center, and bottom right panels, respectively. In all cases the $\eta(z)$ data points are shown in the background as gray points with their 1σ error bars and the shaded band corresponds to the 1σ confidence region for the GA (red band).

IV. CONCLUSIONS

With the advent of GW observations an exciting new window has opened into the Universe. Moreover, a possible detection of strong GW lensing will allow for the testing of fundamental hypotheses of the standard cosmological model as it will provide a test of gravity in the strong field regime, but will also allow for tests of the DDR, similar to that proposed in Ref. [56] for strongly lensed SnIa systems. This exciting possibility was first proposed in Ref. [58], where parametric constraints of the DDR variable $\eta(z)$, given by Eq. (2) were presented.

Here, we extended the work of Ref. [58] in two crucial ways. First, we presented a methodology which allows for the direct creation of $\eta(z)$ mocks, similar to that of Ref. [56]. We showed an example of this approach using mock ET measurements of d_L and d_A from strongly lensed GW events, which were then combined to create mock $\eta(z)$ data points with an MCMC-like approach, as described in Sec. II B. It is important to stress that given the raw measurements of d_L and d_A , the measurements of $\eta(z)$

can be derived without using any dark energy model or beyond the standard model (BSM) theory.

Second, instead of using parametric models for $\eta(z)$, which may carry theoretical bias or miss important features in the data, here we used two specific ML approaches. In particular, we employed the genetic algorithms and the Gaussian processes, which are two nonparametric and symbolic regression subclasses of ML methods, to reconstruct $\eta(z)$ directly without using any underlying model.

Following our methodology, we created a realization of mock $\eta(z)$ data points for $\epsilon = (0, 0.01, 0.05, 0.10)$, assuming the ET specifications, and then used the GA and the GP to directly reconstruct $\eta(z)$. The reconstructions are shown in Figs. 3 and 4 where as can be seen, both the GA and the GP capture the behavior of the data points accurately and remain close to the fiducial model, well within the 1σ region for all values of the duality parameter ϵ . In particular, on average the difference between the GA or GP best-fit and the fiducial model remains close to a percent level in all reconstructions. Furthermore, in the two most extreme cases of $\epsilon = (0.05, 0.10)$ both the GA

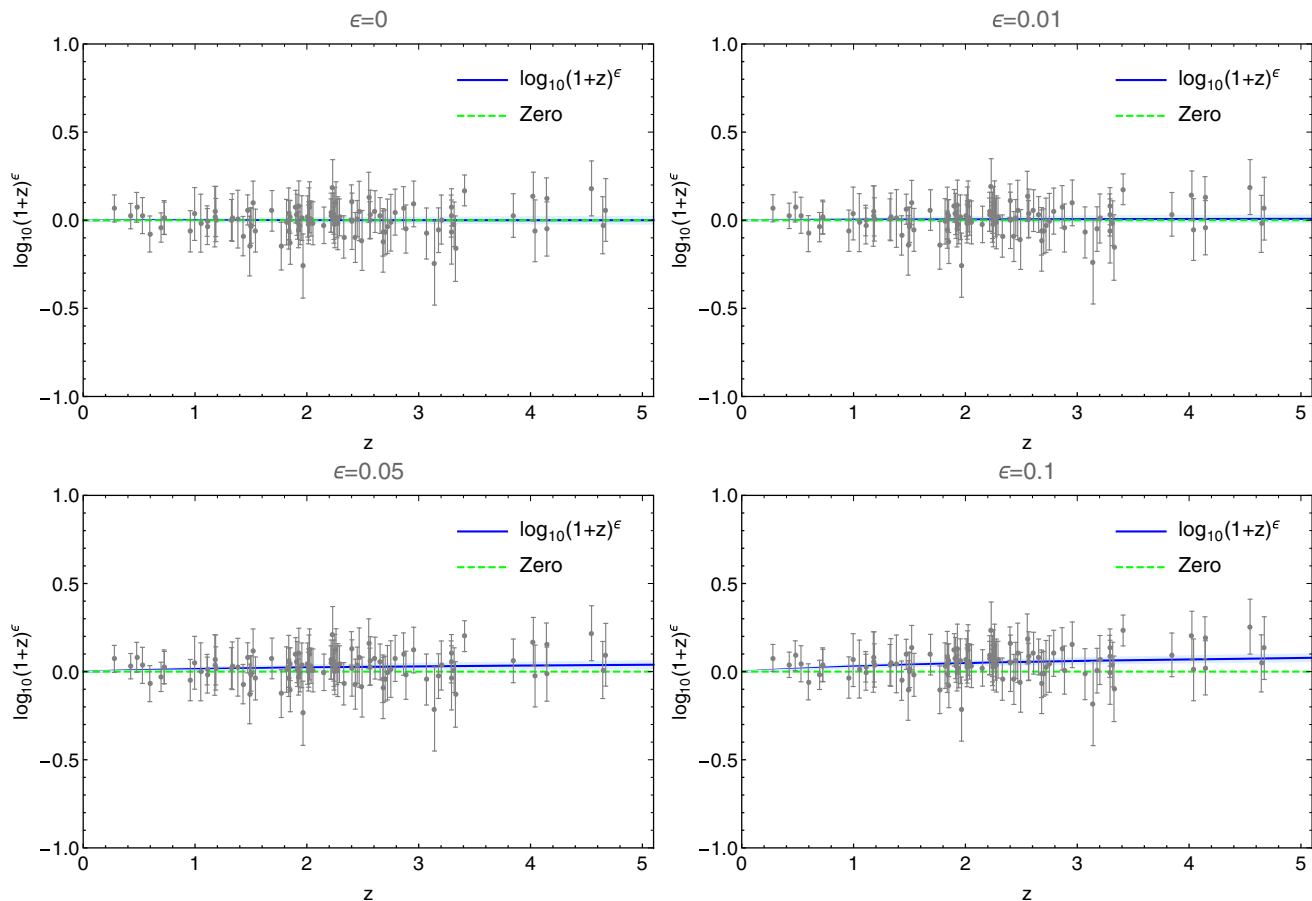


FIG. 6. The $\eta(z)$ mocks along with the fiducial model given by Eq. (2) (blue line) for $\epsilon = (0, 0.01, 0.05, 0.10)$ in the top left, top right, bottom left, and bottom right panels, respectively, and the case of $\log_{10} \eta = 0$ (dashed green line). In all cases the $\eta(z)$ data points are shown in the background as gray points with their 1σ error bars and the shaded band corresponds to the 1σ confidence region for the best-fit fiducial model (blue band).

and the GP find deviations from zero in the redshift range $0 \leq z \leq 3.5$.

We also determined the number of GW lensed events necessary to determine whether a deviation from the null hypothesis is present in the data. We find that the number of events necessary to find a deviation depends both on the value of ϵ and the number of events, so for example with $N = 50$ we can probe for deviations from the null hypothesis ($\epsilon = 0$) up to $z \sim 4$, while to go to higher redshifts we require $N = 200$.

We thus find that both machine learning approaches are capable of correctly recovering the underlying fiducial model and providing percent-level constraints when comparing the fiducial model and the reconstructions at intermediate redshifts, when applied to future Einstein telescope data, thus opening the door to direct tests of the fundamental principles of the standard cosmological model in the coming decades.

Numerical Analysis Files: The genetic algorithm codes used by the authors in the analysis of the paper can be found at <https://github.com/snesseris> and <https://github.com/RubenArjona>. For the Gaussian process analysis we use the publicly available PYTHON package GEORGE found at <https://github.com/dfm/george>.

ACKNOWLEDGMENTS

The authors thank N. Hogg and F. Renzi for useful discussions and comments on the draft. S.N. and R. A.

acknowledge support from the Research Project No. PGC2018-094773-B-C32 and the Centro de Excelencia Severo Ochoa Program No. SEV-2016-0597. S.N. also acknowledges support from the Ramón y Cajal program through Grant No. RYC-2014-15843. The work of H.-N. Lin, and L. Tang has been supported by the National Natural Science Fund of China under Grants No. 11603005, No. 11775038, and No. 11947406.

APPENDIX: COMPARATIVE ANALYSIS

Here we present a reconstruction of the duality parameter $\eta(z)$ using the mock data presented in Sec. II by fitting them to the parametrization of Eq. (2) with $\epsilon(z) = \epsilon_0 = \text{constant}$. This allows us to compare our GA and GP reconstructions to the standard parametric approach used widely in the literature.

In particular, in Fig. 6 we present the different fiducial mocks for $\epsilon = \{0, 0.01, 0.05, 0.1\}$ along with the best-fit parametrizations and their respective errors (blue line and blue shaded region). Comparing these against the reconstructions of Fig. 3 we find that they are in good agreement, albeit the parametric approach has somewhat smaller errors compared to the GA, due to its parametric nature, something which was also observed in Ref. [52]. As all three reconstructions, that of the GA, the GP, and the parametric ones are in good agreement; we are confident in our methodology.

-
- [1] B. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Astrophys. J. Lett.* **882**, L24 (2019).
 - [2] B. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. Lett.* **123**, 011102 (2019).
 - [3] B. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), *Phys. Rev. D* **100**, 104036 (2019).
 - [4] B. Abbott *et al.* (LIGO Scientific, Virgo, 1M2H, Dark Energy Camera GW-E, DES, DLT40, Las Cumbres Observatory, VINROUGE, and MASTER Collaborations), *Nature (London)* **551**, 85 (2017).
 - [5] B. P. Abbott, R. Abbott, T. Abbott, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. Adhikari, V. Adya *et al.*, *Astrophys. J. Lett.* **848**, L13 (2017).
 - [6] P. Creminelli and F. Vernizzi, *Phys. Rev. Lett.* **119**, 251302 (2017).
 - [7] J. M. Ezquiaga and M. Zumalacárregui, *Phys. Rev. Lett.* **119**, 251304 (2017).
 - [8] T. Baker, E. Bellini, P. Ferreira, M. Lagos, J. Noller, and I. Sawicki, *Phys. Rev. Lett.* **119**, 251301 (2017).
 - [9] J. Sakstein and B. Jain, *Phys. Rev. Lett.* **119**, 251303 (2017).
 - [10] L. Lombriser and A. Taylor, *J. Cosmol. Astropart. Phys.* **03** (2016) 031.
 - [11] L. Lombriser and N. A. Lima, *Phys. Lett. B* **765**, 382 (2017).
 - [12] M. Maggiore *et al.*, *J. Cosmol. Astropart. Phys.* **03** (2020) 050.
 - [13] P. Amaro-Seoane *et al.* (LISA Collaboration), [arXiv:1702.00786](https://arxiv.org/abs/1702.00786).
 - [14] G. Cusin, R. Durrer, and I. Dvorkin, [arXiv:1912.11916](https://arxiv.org/abs/1912.11916).
 - [15] J. M. Ezquiaga, D. E. Holz, W. Hu, M. Lagos, and R. M. Wald, *Phys. Rev. D* **103**, 064047 (2021).
 - [16] M. Oguri, *Mon. Not. R. Astron. Soc.* **480**, 3842 (2018).
 - [17] X.-L. Fan, K. Liao, M. Biesiada, A. Piórkowska-Kurpas, and Z.-H. Zhu, *Phys. Rev. Lett.* **118**, 091102 (2017).
 - [18] K. Liao, X.-L. Fan, X. Ding, M. Biesiada, and Z.-H. Zhu, *Nat. Commun.* **8**, 1 (2017).
 - [19] J. M. Ezquiaga and M. Zumalacárregui, *Phys. Rev. D* **102**, 124048 (2020).
 - [20] H. Ohanian, *Int. J. Theor. Phys.* **9**, 425 (1974).
 - [21] K. S. Thorne, [arXiv:gr-qc/9704042](https://arxiv.org/abs/gr-qc/9704042).
 - [22] R. Takahashi and T. Nakamura, *Astrophys. J.* **595**, 1039 (2003).
 - [23] Y. Wang, A. Stebbins, and E. L. Turner, *Phys. Rev. Lett.* **77**, 2875 (1996).

- [24] K. K. Ng, K. W. Wong, T. Broadhurst, and T. G. Li, *Phys. Rev. D* **97**, 023012 (2018).
- [25] O. Hannuksela, K. Haris, K. Ng, S. Kumar, A. Mehta, D. Keitel, T. Li, and P. Ajith, *Astrophys. J. Lett.* **874**, L2 (2019).
- [26] M. Biesiada, X. Ding, A. Piorkowska, and Z.-H. Zhu, *J. Cosmol. Astropart. Phys.* **10** (2014) 080.
- [27] M. Sereno, A. Sesana, A. Bleuler, P. Jetzer, M. Volonteri, and M. Begelman, *Phys. Rev. Lett.* **105**, 251101 (2010).
- [28] S.-S. Li, S. Mao, Y. Zhao, and Y. Lu, *Mon. Not. R. Astron. Soc.* **476**, 2220 (2018).
- [29] S. Jung and C. S. Shin, *Phys. Rev. Lett.* **122**, 041103 (2019).
- [30] K.-H. Lai, O. A. Hannuksela, A. Herrera-Martín, J. M. Diego, T. Broadhurst, and T. G. Li, *Phys. Rev. D* **98**, 083005 (2018).
- [31] L. Dai, S.-S. Li, B. Zackay, S. Mao, and Y. Lu, *Phys. Rev. D* **98**, 104029 (2018).
- [32] M. Sereno, P. Jetzer, A. Sesana, and M. Volonteri, *Mon. Not. R. Astron. Soc.* **415**, 2773 (2011).
- [33] S. Mukherjee, B. D. Wandelt, and J. Silk, *Phys. Rev. D* **101**, 103509 (2020).
- [34] S. Mukherjee, B. D. Wandelt, and J. Silk, *Mon. Not. R. Astron. Soc.* **494**, 1956 (2020).
- [35] S. Mukherjee, T. Broadhurst, J. M. Diego, J. Silk, and G. F. Smoot, *arXiv:2006.03064*.
- [36] I. Etherington, *Gen. Relativ. Gravit.* **39**, 1055 (2007).
- [37] B. A. Bassett and M. Kunz, *Phys. Rev. D* **69**, 101305 (2004).
- [38] P. S. Corasaniti, *Mon. Not. R. Astron. Soc.* **372**, 191 (2006).
- [39] P. Tiwari, *Phys. Rev. D* **95**, 023005 (2017).
- [40] G. F. R. Ellis, R. Poltis, J.-P. Uzan, and A. Weltman, *Phys. Rev. D* **87**, 103530 (2013).
- [41] S. Rasanen, J. Valiviita, and V. Kosonen, *J. Cosmol. Astropart. Phys.* **04** (2016) 050.
- [42] R. Lazkoz, S. Nesseris, and L. Perivolaropoulos, *J. Cosmol. Astropart. Phys.* **07** (2008) 012.
- [43] C. Ma and P.-S. Corasaniti, *Astrophys. J.* **861**, 124 (2018).
- [44] V. F. Cardone, S. Spiro, I. Hook, and R. Scaramella, *Phys. Rev. D* **85**, 123510 (2012).
- [45] S. Khedekar and S. Chakraborti, *Phys. Rev. Lett.* **106**, 221301 (2011).
- [46] R. Holanda and V. Busti, *Phys. Rev. D* **89**, 103517 (2014).
- [47] K. Liao, Z. Li, S. Cao, M. Biesiada, X. Zheng, and Z.-H. Zhu, *Astrophys. J.* **822**, 74 (2016).
- [48] X. Li and H. N. Lin, *Mon. Not. R. Astron. Soc.* **474**, 313 (2018).
- [49] H.-N. Lin, M.-H. Li, and X. Li, *Mon. Not. R. Astron. Soc.* **480**, 3117 (2018).
- [50] X. Zheng, K. Liao, M. Biesiada, S. Cao, T.-H. Liu, and Z.-H. Zhu, *arXiv:2002.09909*.
- [51] R. Arjona, *J. Cosmol. Astropart. Phys.* **08** (2020) 009.
- [52] M. Martinelli *et al.* (EUCLID Collaboration), *Astron. Astrophys.* **644**, A80 (2020).
- [53] N. B. Hogg, M. Martinelli, and S. Nesseris, *J. Cosmol. Astropart. Phys.* **12** (2020) 019.
- [54] R. Holanda, V. Busti, and J. Alcaniz, *J. Cosmol. Astropart. Phys.* **02** (2016) 054.
- [55] R. Holanda, V. Busti, F. Lima, and J. Alcaniz, *J. Cosmol. Astropart. Phys.* **09** (2017) 039.
- [56] F. Renzi, N. B. Hogg, M. Martinelli, and S. Nesseris, *arXiv:2010.04155*.
- [57] H.-N. Lin and X. Li, *Chin. Phys. C* **44**, 075101 (2020).
- [58] H.-N. Lin, X. Li, and L. Tang, *Chin. Phys. C* **45**, 015109 (2021).
- [59] S. Mollerach and E. Roulet *Gravitational Lensing and Microlensing* (World Scientific, 2002).
- [60] A. S. Bolton, S. Bures, L. V. E. Koopmans, T. Treu, R. Gavazzi, L. A. Moustakas, R. Wayth, and D. J. Schlegel, *Astrophys. J.* **682**, 964 (2008).
- [61] T. E. Collett, *Astrophys. J.* **811**, 20 (2015).
- [62] S. H. Suyu *et al.*, *Mon. Not. R. Astron. Soc.* **468**, 2590 (2017).
- [63] A. Aghamousa and A. Shafieloo, *Astrophys. J.* **834**, 31 (2017).
- [64] H.-W. Rix and S. D. M. White, *Mon. Not. R. Astron. Soc.* **254**, 389 (1992).
- [65] S. Cao, J. Qi, Z. Cao, M. Biesiada, J. Li, Y. Pan, and Z.-H. Zhu, *Sci. Rep.* **9**, 11608 (2019).
- [66] W. Zhao, C. Van Den Broeck, D. Baskaran, and T. G. F. Li, *Phys. Rev. D* **83**, 023005 (2011).
- [67] B. Sathyaprakash and B. Schutz, *Living Rev. Relativity* **12**, 2 (2009).
- [68] T. G. F. Li *Extracting Physics from Gravitational Waves: Testing the Strong-Field Dynamics of General Relativity and Inferring the Large-Scale Structure of the Universe* (Springer, 2015).
- [69] C. K. Mishra, K. G. Arun, B. R. Iyer, and B. S. Sathyaprakash, *Phys. Rev. D* **82**, 064010 (2010).
- [70] Y. Wang, A. Stebbins, and E. L. Turner, *Phys. Rev. Lett.* **77**, 2875 (1996).
- [71] H.-Y. Chen, M. Fishbach, and D. E. Holz, *Nature (London)* **562**, 545 (2018).
- [72] M. Ntampaka *et al.*, *arXiv:1902.10159*.
- [73] N. Aggarwal *et al.*, *arXiv:2011.12414*.
- [74] M. Hannam, *Gen. Relativ. Gravit.* **46**, 1767 (2014).
- [75] B. Sathyaprakash and S. Dhurandhar, *Phys. Rev. D* **44**, 3819 (1991).
- [76] S. Klimentenko *et al.*, *Phys. Rev. D* **93**, 042004 (2016).
- [77] B. Allen and J. D. Romano, *Phys. Rev. D* **59**, 102001 (1999).
- [78] D. George and E. Huerta, *Phys. Lett. B* **778**, 64 (2018).
- [79] H. Gabbard, M. Williams, F. Hayes, and C. Messenger, *Phys. Rev. Lett.* **120**, 141103 (2018).
- [80] E. Cuoco *et al.*, *Mach. Learn. Sci. Technol.* **2**, 011002 (2021).
- [81] D. George and E. Huerta, *Phys. Rev. D* **97**, 044039 (2018).
- [82] M. Zevin *et al.*, *Classical Quantum Gravity* **34**, 064003 (2017).
- [83] M. Coughlin *et al.*, *Classical Quantum Gravity* **34**, 044004 (2017).
- [84] P. Graff, F. Feroz, M. P. Hobson, and A. Lasenby, *Mon. Not. R. Astron. Soc.* **421**, 169 (2012).
- [85] A. J. Chua and M. Vallisneri, *Phys. Rev. Lett.* **124**, 041102 (2020).
- [86] H. Gabbard, C. Messenger, I. S. Heng, F. Tonolini, and R. Murray-Smith, *arXiv:1909.06296*.
- [87] A. L. Miller *et al.*, *Phys. Rev. D* **100**, 062005 (2019).
- [88] C. Dreissigacker, R. Sharma, C. Messenger, R. Zhao, and R. Prix, *Phys. Rev. D* **100**, 044009 (2019).

- [89] R. Abbott *et al.* (LIGO Scientific and Virgo Collaborations), [arXiv:2010.14550](https://arxiv.org/abs/2010.14550).
- [90] Y.-C. Lin and J.-H. P. Wu, *Phys. Rev. D* **103**, 063034 (2021).
- [91] E. Huerta *et al.*, [arXiv:2012.08545](https://arxiv.org/abs/2012.08545).
- [92] M. W. Coughlin, *Nat. Astron.* **4**, 550 (2020).
- [93] E. Huerta *et al.*, *Nat. Rev. Phys.* **1**, 600 (2019).
- [94] G. Allen *et al.*, [arXiv:1902.00522](https://arxiv.org/abs/1902.00522).
- [95] R. J. Foley *et al.*, [arXiv:1903.04553](https://arxiv.org/abs/1903.04553).
- [96] M. Lightman, J. Thurakal, J. Dwyer, R. Grossman, P. Kalmus, L. Matone, J. Rollins, S. Zairis, and S. Marka, *J. Phys. Conf. Ser.* **32**, 58 (2006).
- [97] C. Bogdanos and S. Nesseris, *J. Cosmol. Astropart. Phys.* **05** (2009) 006.
- [98] S. Nesseris and A. Shafieloo, *Mon. Not. R. Astron. Soc.* **408**, 1879 (2010).
- [99] S. Nesseris and J. Garcia-Bellido, *J. Cosmol. Astropart. Phys.* **11** (2012) 033.
- [100] S. Nesseris and J. Garca-Bellido, *Phys. Rev. D* **88**, 063521 (2013).
- [101] D. Sapone, E. Majerotto, and S. Nesseris, *Phys. Rev. D* **90**, 023012 (2014).
- [102] R. Arjona and S. Nesseris, *J. Cosmol. Astropart. Phys.* **11** (2020) 042.
- [103] R. Arjona and S. Nesseris, *Phys. Rev. D* **101**, 123525 (2020).
- [104] S. Abel, D. G. Cerdeo, and S. Robles, [arXiv:1805.03615](https://arxiv.org/abs/1805.03615).
- [105] B. C. Allanach, D. Grellscheid, and F. Quevedo, *J. High Energy Phys.* **07** (2004) 069.
- [106] Y. Akrami, P. Scott, J. Edsjo, J. Conrad, and L. Bergstrom, *J. High Energy Phys.* **04** (2010) 057.
- [107] M. Wahde and K. Donner, *Astron. Astrophys.* **379**, 115 (2001).
- [108] V. Rajpaul, in *Proceedings, 56th Annual Conference of the South African Institute of Physics (SAIP 2011): Gauteng, South Africa, 2011* (2012), pp. 519–524 [[arXiv:1202.1643](https://arxiv.org/abs/1202.1643)].
- [109] M. Ho, M. M. Rau, M. Ntampaka, A. Farahi, H. Trac, and B. Poczos, *Astrophys. J.* **887**, 25 (2019).
- [110] M. Affenzeller, S. Wagner, S. Winkler, and A. Beham, *Genetic Algorithms and Genetic Programming: Modern Concepts and Practical Applications* (Chapman and Hall/CRC, London, 2009).
- [111] S. Sivanandam and S. Deepa, in *Introduction to Genetic Algorithms* (Springer, New York, 2008), pp. 15–37.
- [112] S.-M. Udrescu and M. Tegmark, *Sci. Adv.* **6**, eaay2631 (2020).
- [113] Y. Setyawati, M. Prerer, and F. Ohme, *Classical Quantum Gravity* **37**, 075012 (2020).
- [114] H. Vaddireddy, A. Rasheed, A. E. Staples, and O. San, [arXiv:1911.05254](https://arxiv.org/abs/1911.05254).
- [115] K. Liao, A. Shafieloo, R. E. Keeley, and E. V. Linder, *Astrophys. J. Lett.* **886**, L23 (2019).
- [116] E. Belgacem, S. Foffa, M. Maggiore, and T. Yang, *Phys. Rev. D* **101**, 063505 (2020).
- [117] Y. Li, Y. Hao, W. Cheng, and R. Rainer, [arXiv:1904.05683](https://arxiv.org/abs/1904.05683).
- [118] M. Bernardini, L. Mayer, D. Reed, and R. Feldmann, *Mon. Not. R. Astron. Soc.* **496**, 5116 (2020).
- [119] A. Gmez-Valent and L. Amendola, in *15th Marcel Grossmann Meeting on Recent Developments in Theoretical and Experimental General Relativity, Astrophysics, and Relativistic Field Theories* (2019) [[arXiv:1905.04052](https://arxiv.org/abs/1905.04052)].
- [120] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning* (MIT Press, Cambridge, MA, 2006).
- [121] T. Holsclaw, U. Alam, B. Sansó, H. Lee, K. Heitmann, S. Habib, and D. Higdon, *Phys. Rev. D* **82**, 103502 (2010).
- [122] T. Holsclaw, U. Alam, B. Sansó, H. Lee, K. Heitmann, S. Habib, and D. Higdon, *Phys. Rev. Lett.* **105**, 241302 (2010).
- [123] T. Holsclaw, U. Alam, B. Sansó, H. Lee, K. Heitmann, S. Habib, and D. Higdon, *Phys. Rev. Lett.* **84**, 083501 (2011).
- [124] A. Shafieloo, A. G. Kim, and E. V. Linder, *Phys. Rev. D* **85**, 123530 (2012).
- [125] M. Seikel, C. Clarkson, and M. Smith, *J. Cosmol. Astropart. Phys.* **06** (2012) 036.
- [126] M.-J. Zhang and H. Li, *Eur. Phys. J. C* **78**, 460 (2018).
- [127] M. Martinelli, N. B. Hogg, S. Peirone, M. Bruni, and D. Wands, *Mon. Not. R. Astron. Soc.* **488**, 3423 (2019).
- [128] F. Gerardi, M. Martinelli, and A. Silvestri, *J. Cosmol. Astropart. Phys.* **07** (2019) 042.
- [129] N. B. Hogg, M. Bruni, R. Crittenden, M. Martinelli, and S. Peirone, *Phys. Dark Universe* **29**, 100583 (2020).
- [130] D. Benisty, [arXiv:2005.03751](https://arxiv.org/abs/2005.03751).
- [131] M. Seikel and C. Clarkson, [arXiv:1311.6678](https://arxiv.org/abs/1311.6678).
- [132] S. Ambikasaran, D. Foreman-Mackey, L. Greengard, D. W. Hogg, and M. O’Neil, *IEEE Trans. Pattern Anal. Mach. Intell.* **38**, 252 (2016).