Extragalactic test of general relativity from strong gravitational lensing by using artificial neural networks

Jing-Yu Ran and Jun-Jie Wei[*]

*Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing 210023, China
and School of Astronomy and Space Sciences, University of Science and Technology of China,
Hefei 230026, China*

This study aims to test the validity of general relativity (GR) on kiloparsec scales by employing a newly compiled galaxy-scale strong gravitational lensing (SGL) sample. We utilize the distance sum rule within the Friedmann-Lemaître-Robertson-Walker metric to obtain cosmology-independent constraints on both the parametrized post-Newtonian parameter $\gamma_{\rm PPN}$ and the spatial curvature Ω_k, which overcomes the circularity problem induced by the presumption of a cosmological model grounded in GR. To calibrate the distances in the SGL systems, we introduce a novel nonparametric approach, artificial neural network (ANN), to reconstruct a smooth distance-redshift relation from the Pantheon+ sample of type Ia supernovae. Our results show that $\gamma_{\rm PPN} = 1.16^{+0.15}_{-0.12}$ and $\Omega_k = 0.89^{+1.97}_{-1.00}$, indicating a spatially flat universe with the conservation of GR (i.e., $\Omega_k = 0$ and $\gamma_{\rm PPN} = 1$) is basically supported within 1σ confidence level. Assuming a zero spatial curvature, we find $\gamma_{\rm PPN} = 1.09^{+0.11}_{-0.10}$, representing an agreement with the prediction of 1 from GR to a 9.6% precision. If we instead assume GR holds (i.e., $\gamma_{\rm PPN} = 1$), the curvature parameter constraint can be further improved to be $\Omega_k = 0.11^{+0.78}_{-0.47}$. These resulting constraints demonstrate the effectiveness of our method in testing GR on galactic scales by combining observations of strong lensing and the distance-redshift relation reconstructed by ANN.

I. INTRODUCTION

As an important cornerstone of modern physics, Einstein's theory of general relativity (GR) has withstood very strict tests (e.g., [1–4]). But testing GR at a much higher precision is still a vital task, because any possible violation of GR would have profound effects on our understanding of fundamental physics. Within the parametrized post-Newtonian (PPN) formalism, GR predicts that the PPN parameter $\gamma_{\rm PPN}$ which describes the amount of space curvature produced per unit rest mass should be exactly 1 [5]. Measuring $\gamma_{\rm PPN}$ therefore serves as a test of the validity of GR on large scales. That is, any deviation from $\gamma_{\rm PPN} = 1$ implies a possible violation of GR.

On solar system scales, the GR prediction for $\gamma_{\rm PPN}$ has been confirmed with high accuracy. By measuring the round-trip travel time of radar signals passing near the Sun, the Cassini spacecraft yielded $\gamma_{\rm PPN} = 1 + (2.1 \pm 2.3) \times 10^{-5}$ [6]. However, the extragalactic tests of GR are still insufficient and much less precise. On galactic scales, strong gravitational lensing (SGL), combined with stellar kinematics in the lensing galaxy, provides an effective way to test the validity of GR by constraining the PPN

parameter $\gamma_{\rm PPN}$. The pioneering work by the authors of Ref. [7] first utilized this approach and reported a result of $\gamma_{\rm PPN} = 0.98 \pm 0.07$ based on observations of 15 elliptical lensing galaxies from the Sloan Lens ACS (SLACS) Survey. Since then, numerous studies have been conducted to test GR using different SGL samples [8–14]. In this paper, we further explore the validity of GR by employing a newly compiled SGL sample [15], which consists of 161 galaxy-scale strong lensing systems. This larger SGL sample allows us to perform a more comprehensive analysis and obtain further insights into the behavior of gravity on galactic scales.

In practice, in order to constrain the PPN parameter $\gamma_{\rm PPN}$ using SGL systems, one has to know a ratio of three angular diameter distances (i.e., the distances from the observer to the lens, D_l, the observer to the source, D_s, and the lens to the source, D_{ls}). In most previous works, the required distance ratio is calculated within the context of the standard ΛCDM cosmological model. However, ΛCDM itself is established on the framework of GR, which leads to a circularity problem in testing GR [13,14]. To circumvent this problem, we will introduce the distance sum rule (DSR) in the Friedmann-Lemaître-Robertson-Walker (FLRW) metric. The two distances D_l and D_s can be directly determined from observations of type Ia supernovae (SNe Ia), but not the distance D_{ls}. The DSR enables

[*]jjwei@pmo.ac.cn

us to convert D_{ls} into a relationship with D_l, D_s, and the spatial curvature Ω_k. Based on the DSR in the FLRW metric, cosmology-independent constraints on both γ_{PPN} and Ω_k can thus be obtained by combing observations of strong lensing and SNe Ia [10,14].

Very recently, by employing the Gaussian process (GP) method, Liu *et al.* [13] reconstructed a smooth distance-redshift relation directly from SN Ia observation to calibrate the distances in the SGL sample, however, with a grossly underestimated error. GP allows for the reconstruction of a function from a dataset without assuming a specific model or parametrization, and it has been widely used in cosmological researches [16–22]. In the GP analysis, the errors in the observational data are assumed to follow a Gaussian distribution [23]. However, the actual observations might not follow Gaussian distributions. This may thus be a strong assumption for reconstructing a function from observational data. Moreover, due to the sparsity and scatter of data points at high redshifts, the GP reconstructed function from SN Ia data exhibits strange oscillations with large uncertainties. To address these concerns and ensure the reliability of the reconstructed function, we employ the artificial neural network (ANN) method, which is a machine learning technique and has been proven to be a "universal approximator" that can reconstruct a great variety of functions [24,25]. Thanks to the powerful property of neural networks, methods based on ANNs have been widely used in regression and estimation tasks. In this work, we will reconstruct the distance-redshift relation from SN Ia data using the ANN method, utilizing a code developed in Ref. [26].

This paper is organized as follows: in Sec. II, we introduce the methodology and observations used for testing GR on galactic scales. Cosmology-independent constraints on γ_{PPN} and Ω_k are shown in Sec. III. In Sec. IV, we make a summary and end with some discussions.

II. METHODOLOGY AND DATA

In the weak-field limit, the metric of space-time can be characterized as

$$ds^2 = c^2 dt^2 \left(1 - \frac{2GM}{c^2 r}\right) - dr^2 \left(1 + \frac{2\gamma_{PPN} GM}{c^2 r}\right) - r^2 d\Omega^2,$$
(1)

where γ_{PPN} is the PPN parameter, M is the mass of the central object, and Ω is the angle in the invariant orbital plane. In the framework of GR, γ_{PPN} is equal to unity.

A. Gravitational lensing theory

The main idea of testing the validity of GR via SGL systems is that the mass enclosed within the Einstein radius derived separately from the gravitational theory and the dynamical theory should be equivalent, i.e., $M_E^{grl} = M_E^{dyn}$. From the theory of gravitational lensing [27], the Einstein angle θ_E reflecting the angular separations between multiple images is related to the gravitational mass M_E^{grl},

$$\theta_E = \sqrt{\frac{1 + \gamma_{PPN}}{2} \left(\frac{4GM_E^{grl}}{c^2} \frac{D_{ls}}{D_l D_s}\right)^{1/2}},$$
(2)

where D_l, D_s, and D_{ls} are, respectively, the angular diameter distances from the observer to the lens, the observer to the source, and the lens to the source. By introducing the Einstein radius $R_E = D_l \theta_E$, Eq. (2) can be rearranged as

$$\frac{GM_E^{grl}}{R_E} = \frac{2}{1 + \gamma_{PPN}} \frac{c^2}{4} \frac{D_s}{D_{ls}} \theta_E.$$
(3)

To estimate the dynamical mass M_E^{dyn} from the spectroscopic measurement of the lens velocity dispersion, one must first set a mass distribution model for the lensing galaxy. Here we use the common mass model with power-law density profiles [15,28]:

$$\rho(r) = \rho_0 \left(\frac{r}{r_0}\right)^{-\alpha}$$

$$\nu(r) = \nu_0 \left(\frac{r}{r_0}\right)^{-\delta}$$

$$\beta(r) = 1 - \sigma_t^2/\sigma_r^2,$$
(4)

where r is defined as the spherical radial coordinate from the lens center, $\rho(r)$ is the total (including luminous and dark matter) mass density distribution, and $\nu(r)$ represents the distribution of luminous density. The parameter $\beta(r)$ describes the anisotropy of the stellar velocity dispersion, where σ_t and σ_r are the velocity dispersions in the tangential and radial directions, respectively. In the literature, β is always assumed to be independent of r (e.g., [28,29]). Following previous studies [7,9,10,13–15], we set a Gaussian prior $\beta = 0.18 \pm 0.13$, informed by the constraint from a well-studied sample of elliptical galaxies [30]. That is, β will be marginalized using a Gaussian prior of $\beta = 0.18 \pm 0.13$ over the 2σ range of $[-0.08, 0.44]$. Also, α and δ are the power-law indices of the total mass density profile and the luminosity density profile, respectively. It has been confirmed in previous works [15,31] that α is significantly related with the lens redshift z_l and the surface mass density of the lensing galaxy. Therefore, we treat the parametrized model of α as [15]

$$\alpha = \alpha_0 + \alpha_z z_l + \alpha_s \log_{10} \tilde{\Sigma},$$
(5)

where α_0, α_z, and α_s are arbitrary constants. Here $\tilde{\Sigma}$ stands for the normalized surface mass density, and is expressed as $\tilde{\Sigma} = \frac{(\sigma_0/100 \text{ km s}^{-1})^2}{R_{eff}/10 \ h^{-1} \text{ kpc}}$, where σ_0 is the observed velocity dispersion, R_{eff} is the lensing galaxy's half-light radius, and $h = H_0/(100 \text{ km s}^{-1} \text{ Mpc}^{-1})$ is the reduced Hubble constant.

Following the well-known radial Jeans equation in spherical coordinate [32], the radial velocity dispersion of the luminous matter σ_r in early-type lens galaxies takes the form

$$\sigma_r^2(r) = \frac{G \int_r^\infty dr' r'^{2\beta-2} \nu(r') M(r')}{r^{2\beta} \nu(r)}, \quad (6)$$

where $M(r)$ is the total mass included within a sphere with radius r,

$$M(r) = \int_0^r dr' 4\pi r'^2 \rho(r') = 4\pi \frac{\rho_0}{r_0^{-\alpha}} \frac{r^{3-\alpha}}{3-\alpha}. \quad (7)$$

The dynamical mass M_E^{dyn} enclosed within a cylinder of radius equal to the Einstein radius R_E can be written as [15]

$$M_E^{dyn} = 2\pi^{3/2} \frac{R_E^{3-\alpha}}{3-\alpha} \frac{\Gamma(\frac{\alpha-1}{2})}{\Gamma(\frac{\alpha}{2})} \frac{\rho_0}{r_0^{-\alpha}}, \quad (8)$$

where $\Gamma(x)$ is Euler's Gamma function. By combining Eqs. (7) and (8), we get the relation between $M(r)$ and M_E^{dyn}:

$$M(r) = \frac{2}{\sqrt{\pi}} \frac{1}{\lambda(\alpha)} \left(\frac{r}{R_E}\right)^{3-\alpha} M_E^{dyn}, \quad (9)$$

where $\lambda(\alpha) = \Gamma(\frac{\alpha-1}{2})/\Gamma(\frac{\alpha}{2})$. By substituting Eqs. (4) and (9) into Eq. (6), we obtain

$$\sigma_r^2(r) = \frac{2}{\sqrt{\pi}} \frac{G M_E^{dyn}}{R_E} \frac{1}{\xi - 2\beta} \frac{1}{\lambda(\alpha)} \left(\frac{r}{R_E}\right)^{2-\alpha}, \quad (10)$$

where $\xi = \alpha + \delta - 2$.

The actual velocity dispersion of the lensing galaxy is the component of luminosity-weighted average along the line of sight and measured over the effective spectroscopic aperture R_A, which can be expressed as (see Ref. [15] for more details)

$$\sigma_0^2(\leq R_A) = \frac{c^2}{2\sqrt{\pi}} \frac{2}{1+\gamma_{PPN}} \frac{D_s}{D_{ls}} \theta_E F(\alpha, \delta, \beta) \left(\frac{R_A}{R_E}\right)^{2-\alpha}, \quad (11)$$

where

$$F(\alpha, \delta, \beta) = \frac{3-\delta}{(\xi - 2\beta)(3-\xi)} \frac{\lambda(\xi) - \beta\lambda(\xi+2)}{\lambda(\alpha)\lambda(\delta)}. \quad (12)$$

The theoretical value of the velocity dispersion inside the radius $R_{eff}/2$ can then be calculated by [28]

$$\sigma_0^{th} = \sqrt{\frac{c^2}{2\sqrt{\pi}} \frac{2}{1+\gamma_{PPN}} \frac{D_s}{D_{ls}} \theta_E F(\alpha, \delta, \beta) \left(\frac{\theta_{eff}}{2\theta_E}\right)^{2-\alpha}}, \quad (13)$$

where $\theta_{eff} = R_{eff}/D_l$ denotes the effective angular radius of the lensing galaxy.

Based on the spectroscopic data, one can measure the luminosity-weighted average of the line-of-sight velocity

dispersion σ_{ap} within the circular aperture with the angular radius θ_{ap}. In practice, σ_{ap} should be normalized to the velocity dispersion within the typical physical aperture with a radius $R_{eff}/2$,

$$\sigma_0^{obs} = \sigma_{ap}[\theta_{eff}/(2\theta_{ap})]^\eta, \quad (14)$$

where the value of the correction factor is taken as $\eta = -0.066 \pm 0.035$ [33]. Then, the total uncertainty of σ_0^{obs} can be obtained by

$$(\Delta\sigma_0^{SGL})^2 = (\Delta\sigma_0^{stat})^2 + (\Delta\sigma_0^{AC})^2 + (\Delta\sigma_0^{sys})^2, \quad (15)$$

where $\Delta\sigma_0^{stat}$ is the statistical error propagated from the measurement error of σ_{ap}, and $\Delta\sigma_0^{AC}$ is the aperture-correction-induced error propagated from the uncertainty of η. The systematic error due to the extra mass contribution from the outer matters of the lensing galaxy along the line of sight, $\Delta\sigma_0^{sys}$, is taken as an uncertainty of $\sim 3\%$ to the velocity dispersion [34].

Once we know the ratio of the angular diameter distances D_s/D_{ls}, the constraints on the PPN parameter γ_{PPN} can be derived by comparing the observational and theoretical values of the velocity dispersions [see Eqs. (13) and (14)]. Conventionally, the distance ratio D_s/D_{ls} is calculated within the standard ΛCDM cosmological model [9,10]. However, ΛCDM itself is built on the framework of GR and this leads to a circularity problem [13,14]. To avoid this problem, we will use a cosmological-model-independent method which is based upon the sum rule of distances in the FLRW metric to constrain γ_{PPN}.

B. Distance sum rule

In a homogeneous and isotropic space, the dimensionless comoving distance $d(z_l, z_s) \equiv (H_0/c)(1+z_s)D_A(z_l, z_s)$ can be written as

$$d(z_l, z_s) = \frac{1}{\sqrt{|\Omega_k|}} sinn\left(\sqrt{|\Omega_k|} \int_{z_l}^{z_s} \frac{dz'}{E(z')}\right), \quad (16)$$

where Ω_k denotes the spatial curvature density parameter at the present time and $E(z) = H(z)/H_0$ is the dimensionless expansion rate. Also, $sinn(x)$ is $sinh(x)$ when $\Omega_k > 0$, x when $\Omega_k = 0$, and $sin(x)$ when $\Omega_k < 0$. By applying the notations $d(z) \equiv d(0, z)$, $d_{ls} \equiv d(z_l, z_s)$, $d_l \equiv d(0, z_l)$, and $d_s \equiv d(0, z_s)$, one can derive a sum rule of distances along the null geodesics of the FLRW metric as [35–37]

$$\frac{d_{ls}}{d_s} = \sqrt{1 + \Omega_k d_l^2} - \frac{d_l}{d_s}\sqrt{1 + \Omega_k d_s^2}. \quad (17)$$

This relation provides a cosmology-independent probe to test both the spatial curvature and the FLRW metric. The validity of the FLRW metric can be tested by comparing the derived Ω_k from the three distances ($d_l, d_s,$ and d_{ls}) for any two pairs of (z_l, z_s).

With Eq. (17), the distance ratio D_s/D_{ls}[1] in Eq. (13) is only related to the curvature parameter Ω_k and the dimensionless distances d_l and d_s. If independent measurements of d_l and d_s are given, we can put constraints on both γ_{PPN} and Ω_k from Eqs. (13) and (17) without assuming any specific cosmological model.

C. Artificial neural network

To calibrate the distances d_l and d_s of the SGL systems [i.e., the distances d_l and d_s on the right side of Eq. (17)], we use a new nonparametric approach, ANN, to reconstruct a smooth distance-redshift relation $d(z)$ from SN Ia observation.

ANNs possess several desirable properties, including high-level abstraction of neural input-output transformation, the ability to generalize from learned instances to new unseen data, adaptability, self-learning, fault tolerance, and nonlinearity [38]. According to the universal approximation theorem [24,39], ANNs can function as universal function approximators to simulate arbitrary input-output relationships using multilayer feedforward networks with a sufficient number of hidden units. Therefore, we can input the redshift z into the neural network, with the corresponding comoving distance $d(z)$ and its associated error $\sigma_{d(z)}$ as the desired outputs. Once the network has been trained using the Pantheon+ sample, we will obtain an approximate function capable of predicting both $d(z)$ and its error $\sigma_{d(z)}$ at any given redshift z.

The authors of Ref. [26] have developed a Python code for the reconstruction of functions from observational data employing an ANN. They have substantiated the reliability of these reconstructed functions by estimating cosmological parameters through the utilization of the reconstructed Hubble parameter $H(z)$ and the luminosity distance $D_L(z)$, in direct comparison with observational data. In our study, we will employ this code to reconstruct the distance-redshift relation.

The general structure of an ANN consists of an input layer, one or more hidden layers, and an output layer. The basic units of these layers are referred to as neurons, which serve as both linear transformation units and nonlinear activation functions for the input vector. In accordance with Ref. [26], we employ the exponential linear unit as our chosen activation function, as defined by its form in [40]:

$$f(x) = \begin{cases} x & x > 0 \\ \alpha(e^x - 1) & x \le 0, \end{cases} \qquad (18)$$

where the hyperparameter α is set to 1.

The network is trained by minimizing a loss function, which quantitatively measures the discrepancy between the

ground truth and predicted values. In this analysis, we adopt the mean absolute error (MAE), also known as the L1 loss function, as our choice of loss function. The linear weights and biases within the network are optimized using the backpropagation algorithm. We employ the Adam optimizer [41], a gradient-based optimization technique, to iteratively update the network parameters during training. This choice of optimizer also contributes to faster convergence. After multiple iterations, the network parameters are adjusted to minimize the loss. We have determined that a sufficient number of iterations for training convergence is reached when the loss no longer decreases, which we set to be 3×10^5. Batch normalization [42] is a technique designed to stabilize the distribution of inputs within each layer, allowing for higher learning rates and reduced sensitivity to initialization.

To determine the optimal network model, we train the network using 1701 SNe Ia from the Pantheon+ sample (more on this below) and assess the fitting effect through K-fold cross-validation [43]. In K-fold cross-validation, the training set is divided into k smaller sets, with k-1 folds used as training data for model training, and the remaining fold used for validation. This process is repeated k times, with each fold serving as the validation data once. The final performance of the model is determined by averaging the performance across these k iterations. This approach is particularly useful when the number of samples available for learning is insufficient to split into traditional train, validation, and test sets, as is the case in our analysis. Additionally, it helps mitigate issues arising from the randomness in data partitioning. As a general guideline, we have selected $k = 10$ for our cross-validation procedure and have utilized the mean squared error (MSE) as the metric for validating the performance of the model.

Through our experimentation, we have found that the network model with a single hidden layer comprising 4096 neurons and without batch normalization yields the best results. We conduct comparisons with models having varying numbers of hidden layers, and we observe diminished performance as the number of hidden layers increased, accompanied by increased computational resource consumption. Regarding the number of neurons in the hidden layer, we observe negligible impact on the results, as reflected by the final MSE values consistently hovering around 0.0042, regardless of whether the number of neurons was set to 1024, 2048, 4096, or 8192. Importantly, the final MSE value with 4096 neurons was slightly smaller compared to the other three configurations, and as a result, we select this configuration. The validation values for implementing batch normalization or not implementing it is 0.0049 or 0.0042, respectively.

Subsequently, we will employ the optimal network model, as described above, to reconstruct our distance-redshift curve.

---

[1]Note that D_s/D_{ls} is actually equal to the dimensionless distance ratio d_s/d_{ls}.

D. Supernova data

In order to reconstruct the distance function $d(z)$, we choose the latest combined sample of SNe Ia called Pantheon+ [44], which consists of 1701 light curves of 1550 SNe Ia, covering the redshift range $0.001 < z < 2.3$. For each SN Ia, the distance modulus μ is related to the luminosity distance D_L by

$$\mu(z) = 5\log_{10}\left[\frac{D_L(z)}{\text{Mpc}}\right] + 25, \tag{19}$$

and the observed distance modulus is

$$\mu_{\text{obs}}(z) = m_B(z) + \kappa \cdot X_1 - \omega \cdot \mathcal{C} - M_B, \tag{20}$$

where m_B is the rest-frame B band peak magnitude, X_1 and $\mathcal{C}$, respectively, represent the time stretch of light curve and the SN color at maximum brightness, and M_B is the absolute B-band magnitude. Through the "BEAMS with Bias Corrections" method [45], the two nuisance parameters κ and ω can be calibrated to zero. Then, the observed distance modulus can be simplified as

$$\mu_{\text{obs}}(z) = m_{\text{corr}}(z) - M_B, \tag{21}$$

where m_{corr} is the corrected apparent magnitude. The absolute magnitude M_B is exactly degenerate with the Hubble constant H_0. Once the value of M_B or H_0 is known, the luminosity distances $D_L(z)$ can be obtained from SNe Ia.

In this work, we adopt $H_0 = 70 \text{ km s}^{-1}\text{ Mpc}^{-1}$ to normalize the SN Ia $D_L(z)$ data as the observational $d(z)$. That is, $d(z) = (H_0/c)D_L(z)/(1 + z)$. Note that the choice of H_0 has no impact on our results, since the required distance ratio D_s/D_{ls} [see Eq. (13)] is completely independent of H_0. Having obtained the dataset of $d(z)$, we adopt ANN to reconstruct the distance function $d(z)$, and the results are shown in Fig. 1. The black line represents the reconstructed function of $d(z)$, and the shaded region is the corresponding 1σ confidence level.

E. Strong-lensing data

Recently, the authors of Ref. [15] compiled a galaxy-scale SGL sample including 161 systems with stellar velocity dispersion measurements, which is assembled with strict selection criteria to meet the assumption of spherical symmetry on the lens mass model. The observational information for each SGL system are listed in the Appendix of Ref. [15], including the lens redshift z_l, the source redshift z_s, the Einstein angle θ_E, the central velocity dispersion of the lensing galaxy σ_{ap}, the spectroscopic aperture angular radius θ_{ap}, and the half-light angular radius of the lensing galaxy θ_{eff}.

By fitting the two-dimensional power-law luminosity profile convolved with the instrumental point spread function to the high-resolution Hubble Space Telescope imaging data over a circle of radius $\theta_{\text{eff}}/2$ centered on the lensing galaxies, the authors of Ref. [15] measured the
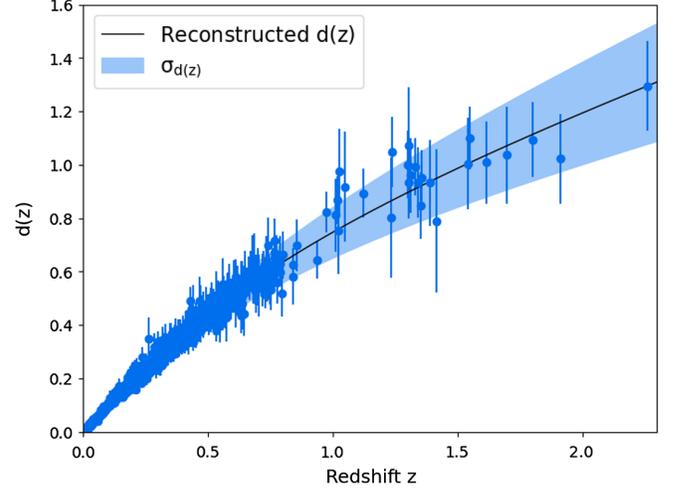

FIG. 1. Reconstruction of the dimensionless comoving distance $d(z)$ from Pantheon+ SNe Ia using ANN. The shaded area is the 1σ confidence level of the reconstruction. The blue dots with error bars represent the observational data.

slopes of the luminosity density profile δ for the 130 lensing galaxies in the full sample. They showed that δ should be treated as an observable for each lens in order to get an unbiased estimate of the cosmological parameter Ω_m. Therefore, the SGL sample we adopt here is the truncated sample of 130 SGL systems with δ measurements, for which the redshift ranges of lenses and sources are $0.0624 \leq z_l \leq 0.7224$ and $0.1970 \leq z_s \leq 2.8324$, respectively. In this work, we use the reconstructed distance function $d(z)$ from Pantheon + SNe Ia to calibrate the distances d_l and d_s of the SGL systems. However, the SN Ia catalog extends only to $z = 2.3$. As such, we shall employ only a subset of the SGL sample that overlaps with the SN Ia data for the calibration. Thus, only 120 SGL systems with $z_s \leq 2.3$ are available in our analysis.

F. The likelihood function

By using the Python Markov chain Monte Carlo module EMCEE [46] to maximize the likelihood function $\mathcal{L}$, we simultaneously place limits on the PPN parameter γ_{PPN}, the curvature parameter Ω_k, and the lens model parameters (α_0, α_z, and α_s). The likelihood function is defined as

$$\mathcal{L} = \prod_{i=1}^{120} \frac{1}{\sqrt{2\pi}\Delta\sigma_{0,i}^{\text{tot}}} \exp\left[-\frac{1}{2}\left(\frac{\sigma_{0,i}^{\text{th}} - \sigma_{0,i}^{\text{obs}}}{\Delta\sigma_{0,i}^{\text{tot}}}\right)^2\right], \tag{22}$$

where the variance

$$(\Delta\sigma_0^{\text{tot}})^2 = (\Delta\sigma_0^{\text{SGL}})^2 + (\Delta\sigma_0^{\text{SN}})^2 \tag{23}$$

is given in terms of the total uncertainty $\Delta\sigma_0^{\text{SGL}}$ derived from the SGL observation [Eq. (15)] and the propagated uncertainty $\Delta\sigma_0^{\text{SN}}$ derived from the distance calibration by SNe Ia. With Eq. (13), the propagated uncertainty $\Delta\sigma_0^{\text{SN}}$ can be estimated as

$$\Delta\sigma_0^{\rm SN} = \sigma_0^{\rm th}\frac{\Delta D_r}{2D_r}, \tag{24}$$

where D_r is a convenient notation for the distance ratio in Eq. (13), i.e., $D_r \equiv D_s/D_{ls} = d_s/d_{ls}$, and its uncertainty is ΔD_r. With the reconstructed distance function $d(z)$, as well as its 1σ uncertainty $\Delta d(z)$, from the SN Ia data, we can calibrate the distances (d_l and d_s) and their corresponding uncertainties (Δd_l and Δd_s) for each SGL system. Thus, the

uncertainty ΔD_r of the distance ratio can be easily derived from Eq. (17), i.e.,

$$(\Delta D_r)^2 = D_r^4 \left(\frac{\Omega_k d_l}{\sqrt{1 + \Omega_k d_l^2}} - \frac{\sqrt{1 + \Omega_k d_s^2}}{d_s} \right)^2 (\Delta d_l)^2$$
$$+ D_r^4 \left(\frac{d_l}{d_s^2\sqrt{1 + \Omega_k d_s^2}} \right)^2 (\Delta d_s)^2. \tag{25}$$
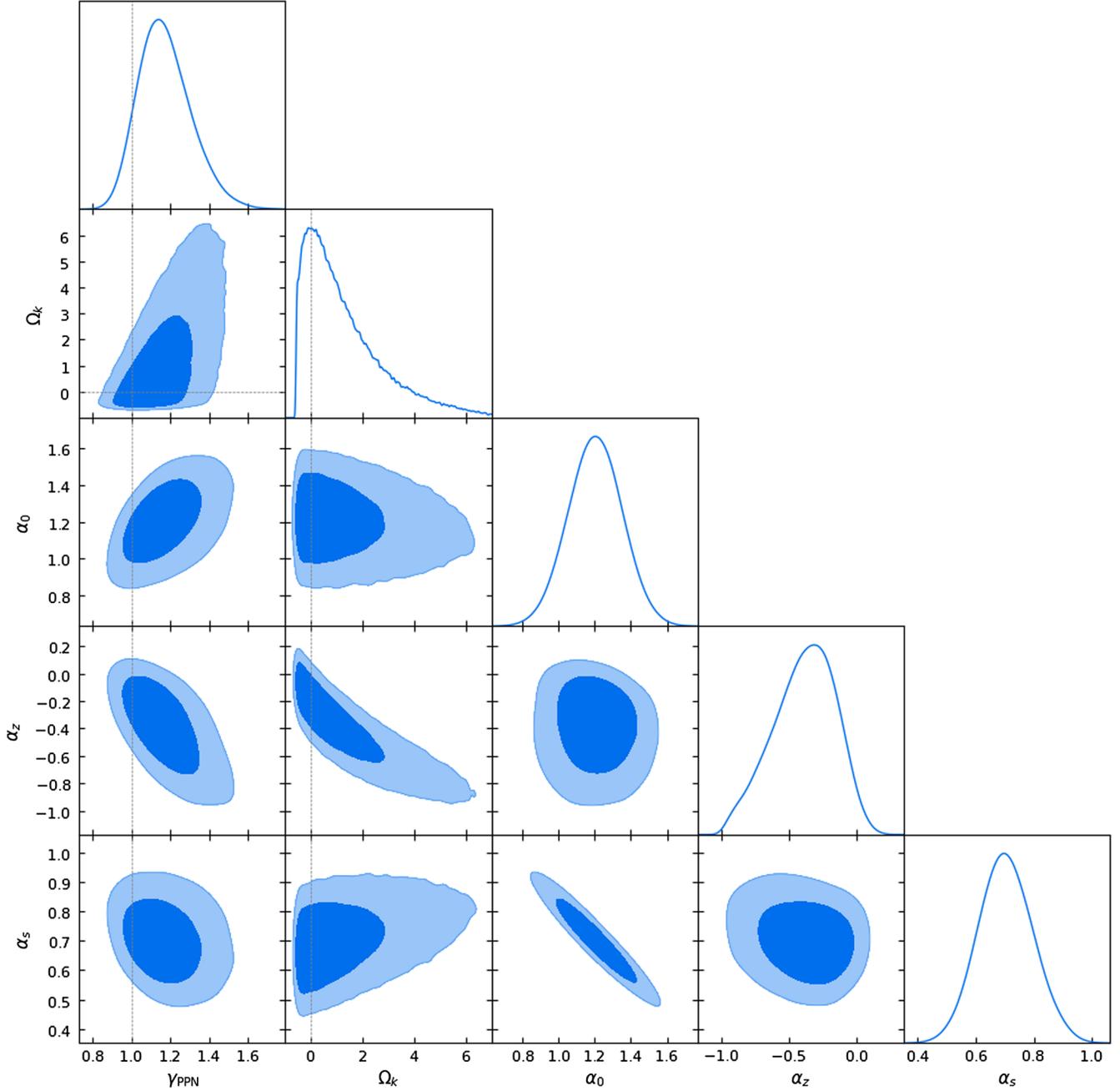

FIG. 2. 1D marginalized probability distributions and 2D $1-2\sigma$ confidence contours for the PPN parameter $\gamma_{\rm PPN}$, the cosmic curvature Ω_k, and the lens model parameters (α_0, α_z, and α_s). The dashed lines represent $\gamma_{\rm PPN} = 1$ and $\Omega_k = 0$, corresponding to a flat universe with the validity of GR.

TABLE I. Constraint results for all parameters with different priors.

Priors	γ_{PPN}	Ω_k	α_0	α_z	α_s
None	$1.16^{+0.15}_{-0.12}$	$0.89^{+1.97}_{-1.00}$	$1.20^{+0.15}_{-0.15}$	$-0.37^{+0.22}_{-0.26}$	$0.70^{+0.10}_{-0.09}$
$\Omega_k = 0$	$1.09^{+0.11}_{-0.10}$		$1.22^{+0.14}_{-0.14}$	$-0.20^{+0.11}_{-0.11}$	$0.67^{+0.09}_{-0.09}$
$\gamma_{PPN} = 1$		$0.12^{+0.78}_{-0.47}$	$1.10^{+0.11}_{-0.12}$	$-0.20^{+0.15}_{-0.16}$	$0.74^{+0.08}_{-0.08}$

III. RESULTS

The 1D marginalized probability distributions and 2D plots of the $1-2\sigma$ confidence regions for the PPN parameter γ_{PPN}, the cosmic curvature Ω_k, and the lens model parameters (α_0, α_z, and α_s), constrained by 120 SGL systems, are presented in Fig. 2, and the best-fitting results

are listed in Table I. These contours show that at the 1σ confidence level, the inferred parameter values are $\gamma_{PPN} = 1.16^{+0.15}_{-0.12}$, $\Omega_k = 0.89^{+1.97}_{-1.00}$, $\alpha_0 = 1.2^{+0.15}_{-0.15}$, $\alpha_z = -0.37^{+0.22}_{-0.26}$, and $\alpha_s = 0.70^{+0.10}_{-0.09}$. We find that the measured γ_{PPN} is consistent with the prediction of $\gamma_{PPN} = 1$ from GR, and its constraint accuracy is about 11.6%. While Ω_k is weakly constrained, it is still compatible with zero spatial curvature within the 1σ confidence level. We also find that the inferred α_z and α_s separately deviate from zero at $\sim 2\sigma$ and $\sim 8\sigma$ levels, confirming the previous finding that the total mass density slope α strongly depends on both the lens redshift and the surface mass density [15].

We further explore the scenario of adopting a of flatness, i.e., $\Omega_k = 0$. For this scenario, as shown in Fig. 3 and Table I, the marginalized distribution gives
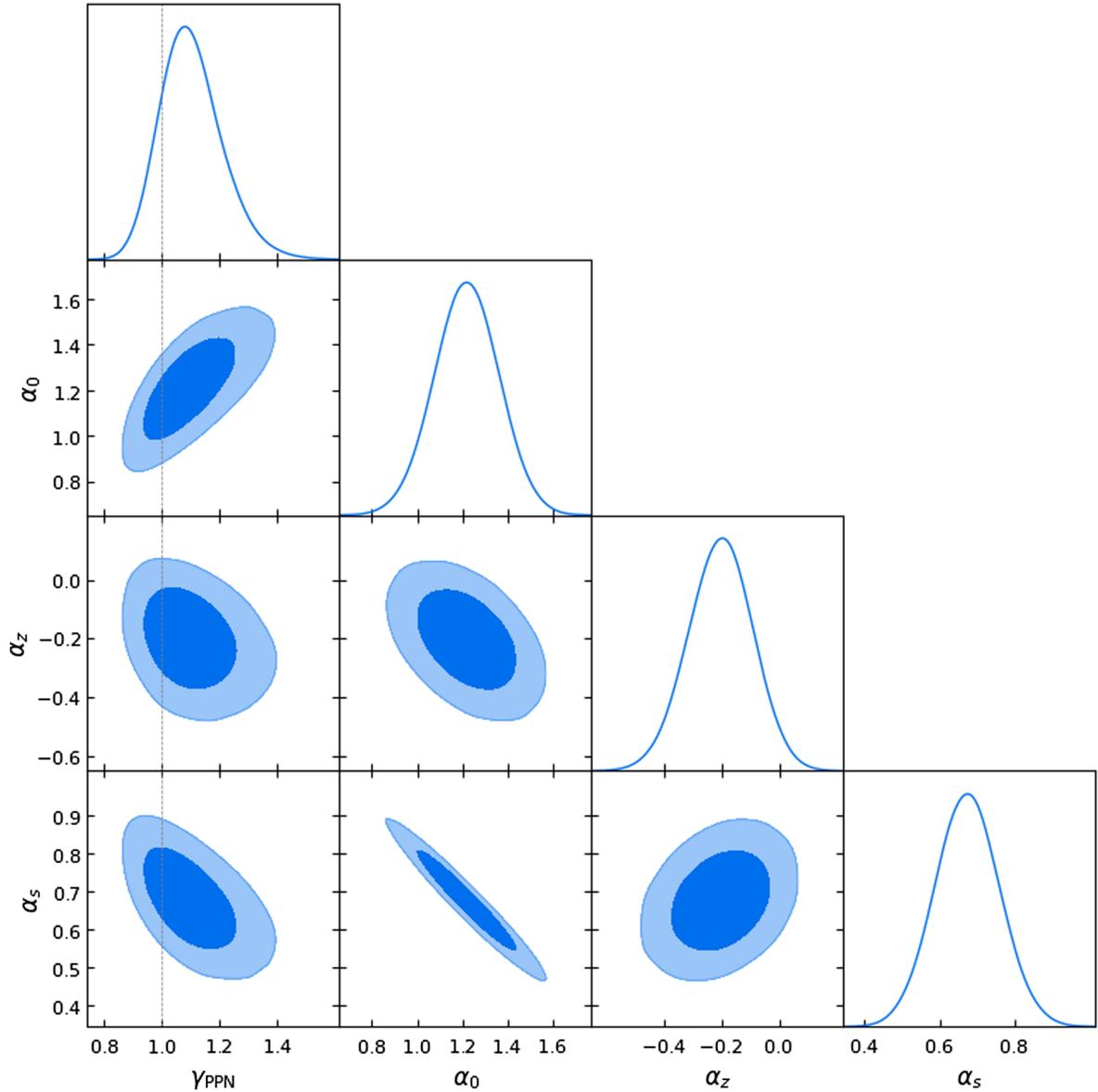

FIG. 3.   Same as Fig. 2, except now for the scenario with a prior of $\Omega_k = 0$. The dashed line represents $\gamma_{PPN} = 1$ predicted by GR.
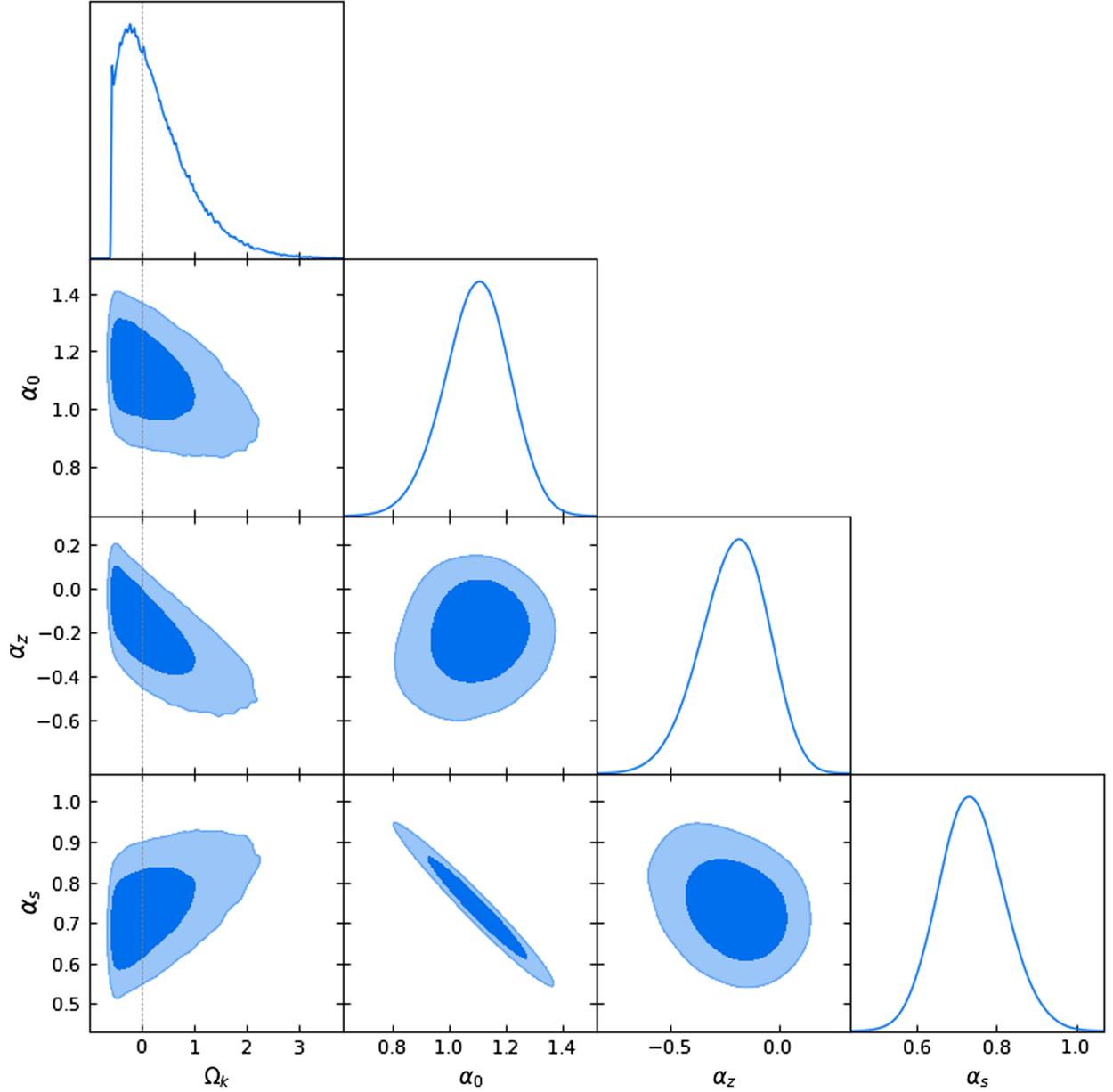
FIG. 4.   Same as Fig. 2, except now for the scenario with a prior of $\gamma_{\rm PPN} = 1$. The dashed line represents a spatially flat universe.

$\gamma_{\rm PPN} = 1.09^{+0.11}_{-0.10}$, representing a precision of 9.6%, in good agreement with the prediction of GR. If instead we adopt a prior of $\gamma_{\rm PPN} = 1$ (i.e., assuming GR holds) and allow Ω_k to be a free parameter, the resulting constraints on Ω_k and the lens model parameters are displayed in Fig. 4 and Table I. The marginalized Ω_k constraint is $\Omega_k = 0.12^{+0.78}_{-0.47}$, consistent with a spatially flat universe. The comparison among lines 1–3 of Table I suggests that different choices of priors have little effect on the lens model parameters (α_0, α_z, and α_s).

From Figs. 2–4, we can see that there is a strongly squeezed relation between α_0 and α_s. This seems to indicate that to some extent one can replace α_s by mass density independent α_0, but not entirely. This may also imply that

the adopted parametrization for the total mass density slope α [Eq. (5)] is not the most compatible lens model. Nevertheless, it is worth pointing out that besides the dependence of α on lens redshift and surface mass density, other unexplored but important dependencies may also be found in the future that can result in a more accurate phenomenological model for lensing galaxies [15], thereby providing more reliable constraints on the PPN parameter $\gamma_{\rm PPN}$.

IV. CONCLUSION AND DISCUSSIONS

Galaxy-scale SGL systems, combined with stellar velocity dispersion measurements of lensing galaxies, provide a powerful probe to test the validity of GR by constraining the PPN parameter $\gamma_{\rm PPN}$ on kiloparsec scales. Testing GR in

this manner, however, it is necessary to know the angular diameter distances between the observer, lens, and source. Conventionally, the required distances are calculated within the standard ΛCDM cosmological model. Such distance calculations would involve a circularity problem in testing GR, since ΛCDM itself is established on the framework of GR. In this paper, in order to address the circularity problem, we have employed the DSR in the FLRW metric to estimate not only γ_{PPN} but also the spatial curvature Ω_k independently of any specific cosmological model. To calibrate the distances of the SGL systems, we have introduced a new nonparametric approach for reconstructing the distance-redshift relation from the Pantheon+ SN Ia sample using an ANN, which has no assumptions about the observational data and is a completely data-driven approach.

By combining 120 well-selected SGL systems with the reconstructed distance function from 1701 data points of SNe Ia, we have obtained simultaneous estimates of γ_{PPN} and Ω_k without any specific assumptions about the contents of the universe or the theory of gravity. Our results show that $\gamma_{PPN} = 1.16^{+0.15}_{-0.12}$ and $\Omega_k = 0.89^{+1.97}_{-1.00}$. The measured γ_{PPN} is in good agreement with the prediction of GR with 11.6% accuracy. If we use flatness as a prior (i.e., $\Omega_k = 0$), we infer that $\gamma_{PPN} = 1.09^{+0.11}_{-0.10}$, representing a precision of 9.6%. If we instead assume the conservation of GR (i.e., $\gamma_{PPN} = 1$) and allow Ω_k to be a free parameter, we find $\Omega_k = 0.12^{+0.78}_{-0.47}$. The measured Ω_k is consistent with zero spatial curvature, suggesting that there is no significant deviation from a flat universe.

In the literature, the authors of Ref. [7] used 15 SLACS lenses to give a result of $\gamma_{PPN} = 0.98 \pm 0.07$, based on priors on galaxy structure from local observations and the assumption of a ΛCDM model with density parameters $(\Omega_m, \Omega_\Lambda) = (0.3, 0.7)$. Utilizing a sample of 80 SGL systems, Ref. [10] obtained the constraint accuracy of the PPN parameter γ_{PPN} to be 25% under the assumption of ΛCDM with parameters taken from Planck observations. Within the same context of ΛCDM, Ref. [11] concluded that $\gamma_{PPN} = 0.97 \pm 0.09$ (representing a precision of 9.3%) by analyzing the nearby lens ESO 325-G004. Through the reanalysis of four time-delay lenses, Ref. [12] obtained simultaneous constraints of γ_{PPN} and the Hubble constant H_0 for flat ΛCDM, yielding $\gamma_{PPN} = 0.87^{+0.19}_{-0.17}$ (representing a precision of 21%) and $H_0 = 73.65^{+1.95}_{-2.26}$ km s^{-1} Mpc^{-1}. Within a flat FLRW metric, Ref. [13] used 120 lenses to achieve a model-independent estimate of $\gamma_{PPN} = 1.065^{+0.064}_{-0.074}$ (representing a precision of 6.5%) by employing the GP method to reconstruct the SN distances. As a further refinement, Ref. [14] removed the flatness assumption and implemented the DSR to obtain model-independent constraints of $\gamma_{PPN} = 1.11^{+0.11}_{-0.09}$ (representing a precision of 9.0%) and $\Omega_k = 0.48^{+1.09}_{-0.71}$. Note that in Ref. [14] the distances of the SGL systems were determined by fitting a third-order polynomial to the SN Ia data. Unlike the polynomial fit that relies on the assumed

parametrization, the ANN used in this work is a completely data-driven approach that could reconstruct a function from various data without assuming a parametrization of the function. Moreover, unlike the GP method that relies on the assumption of Gaussian distributions for the observational random variables, the ANN method has no assumptions about the data. More importantly, compared to previous results, our work yielded comparable resulting constraints on γ_{PPN}, which indicates the effectiveness of data-driven modeling based on the ANN.

Note that the dataset used here is roughly 10 times larger than that used by the authors of Ref. [7]. Yet the results obtained are almost the same level of accuracy. We attribute the absence of significantly enhanced accuracy to three primary reasons. (i) Reference [7] modeled the probability distribution of the total mass density slope α as a Gaussian prior of $\alpha = 1.93 \pm 0.08$. But, Ref. [10] highlighted that the most substantial source of systematic error on γ_{PPN} is the scatter of α. That is, without the Gaussian prior on α, the systematic error on γ_{PPN} would be much larger. Our worse γ_{PPN} accuracy appears to be due to the fact that Ref. [7] set a very narrow Gaussian prior on α, while we consider the dependence of α on lens redshift and surface mass density [see Eq. (5)] and adopt wide flat priors on the free parameters α_0, α_z and α_s. (ii) Compared to our sample [28,47,48], the sample employed in Ref. [7] exhibits notably higher accuracy in velocity dispersion measurements. (iii) Reference [7] assumed ΛCDM with specific cosmological parameters to compute distances D_s and D_{ls}, thereby lacking the uncertainty derived from distance calibration.

Looking forward, the forthcoming Large Synoptic Survey Telescope (LSST) survey, with its excellent operation performance, holds great promise for detecting a large number of lenses, potentially reaching up to 120,000 in the most optimistic scenario [49]. By setting a prior on the curvature parameter $-0.007 < \Omega_k < 0.006$, Ref. [10] showed that 53,000 simulated LSST strong lensing data would set a stringent constraint of $\gamma_{PPN} = 1.000^{+0.0009}_{-0.0011}$, reaching a precision of $10^{-3} \sim 10^{-4}$. Similarly, the authors of Ref. [50] performed a robust extragalactic test of GR using a well-defined sample of 5,000 simulated strong lenses from LSST, yielding an accuracy of 0.5%. In brief, much more severe constraints on both γ_{PPN} and Ω_k, as discussed in this work, can be expected with the help of future lens surveys.

ACKNOWLEDGMENTS

[1] F. W. Dyson, A. S. Eddington, and C. Davidson, Phil. Trans. R. Soc. A **220,** 291 (1920).

[2] R. V. Pound and G. A. Rebka, Phys. Rev. Lett. **4,** 337 (1960).

[3] I. I. Shapiro, Phys. Rev. Lett. **13,** 789 (1964).

[4] J. H. Taylor, L. A. Fowler, and P. M. McCulloch, Nature (London) **277,** 437 (1979).

[5] K. S. Thorne and C. M. Will, Astrophys. J. **163,** 595 (1971).

[6] B. Bertotti, L. Iess, and P. Tortora, Nature (London) **425,** 374 (2003).

[7] A. S. Bolton, S. Rappaport, and S. Burles, Phys. Rev. D **74,** 061501 (2006).

[8] T. L. Smith, arXiv:0907.4829.

[9] J. Schwab, A. S. Bolton, and S. A. Rappaport, Astrophys. J. **708,** 750 (2010).

[10] S. Cao, X. Li, M. Biesiada, T. Xu, Y. Cai, and Z.-H. Zhu, Astrophys. J. **835,** 92 (2017).

[11] T. E. Collett, L. J. Oldham, R. J. Smith, M. W. Auger, K. B. Westfall, D. Bacon, R. C. Nichol, K. L. Masters, K. Koyama, and R. van den Bosch, Science **360,** 1342 (2018).

[12] T. Yang, S. Birrer, and B. Hu, Mon. Not. R. Astron. Soc. **497,** L56 (2020).

[13] X.-H. Liu, Z.-H. Li, J.-Z. Qi, and X. Zhang, Astrophys. J. **927,** 28 (2022).

[14] J.-J. Wei, Y. Chen, S. Cao, and X.-F. Wu, Astrophys. J. Lett. **927,** L1 (2022).

[15] Y. Chen, R. Li, Y. Shu, and X. Cao, Mon. Not. R. Astron. Soc. **488,** 3745 (2019).

[16] M. Seikel, S. Yahya, R. Maartens, and C. Clarkson, Phys. Rev. D **86,** 083001 (2012).

[17] S. Yahya, M. Seikel, C. Clarkson, R. Maartens, and M. Smith, Phys. Rev. D **89,** 023503 (2014).

[18] A. Montiel, R. Lazkoz, I. Sendra, C. Escamilla-Rivera, and V. Salzano, Phys. Rev. D **89,** 043007 (2014).

[19] J.-J. Wei and X.-F. Wu, Astrophys. J. **838,** 160 (2017).

[20] F. Melia, Mon. Not. R. Astron. Soc. **481,** 4855 (2018).

[21] K. Liao, A. Shafieloo, R. E. Keeley, and E. V. Linder, Astrophys. J. Lett. **886,** L23 (2019).

[22] S. Dhawan, J. Alsing, and S. Vagnozzi, Mon. Not. R. Astron. Soc. **506,** L1 (2021).

[23] M. Seikel, C. Clarkson, and M. Smith, J. Cosmol. Astropart. Phys. 06 (2012) 036.

[24] G. V. Cybenko, Math. Control Signals Syst. **2,** 303 (1989).

[25] K. Hornik, Neural Netw. **4,** 251 (1991).

[26] G.-J. Wang, X.-J. Ma, S.-Y. Li, and J.-Q. Xia, Astrophys. J. Suppl. Ser. **246,** 13 (2020).

[27] P. Schneider, J. Ehlers, and E. E. Falco, Gravitational Lenses (Springer Berlin, Heidelberg, 1992), 10.1007/978-3-662-03758-4.

[28] L. V. E. Koopmans, T. Treu, A. S. Bolton, S. Burles, and L. A. Moustakas, Astrophys. J. **649,** 599 (2006).

[29] T. Treu, M. W. Auger, L. V. E. Koopmans, R. Gavazzi, P. J. Marshall, and A. S. Bolton, Astrophys. J. **709,** 1195 (2010).

[30] O. Gerhard, A. Kronawitter, R. P. Saglia, and R. Bender, Astron. J. **121,** 1936 (2001).

[31] A. Sonnenfeld, T. L. Treu, R. Gavazzi, S. H. Suyu, P. J. Marshall, M. W. Auger, and C. Nipoti, Astrophys. J. **777,** 98 (2013).

[32] J. Binney, Mon. Not. R. Astron. Soc. **190,** 873 (1980).

[33] M. Cappellari, R. Bacon, M. Bureau, M. C. Damen, R. L. Davies, P. T. de Zeeuw, E. Emsellem, J. Falcón-Barroso, D. Krajnović, H. Kuntschner et al., Mon. Not. R. Astron. Soc. **366,** 1126 (2006).

[34] G. Jiang and C. S. Kochanek, Astrophys. J. **671,** 1568 (2007).

[35] P. J. E. Peebles, Principles of Physical Cosmology (Princeton University Press, Princeton, 1993), 10.1515/9780691206721.

[36] G. Bernstein, Astrophys. J. **637,** 598 (2006).

[37] S. Räsänen, K. Bolejko, and A. Finoguenov, Phys. Rev. Lett. **115,** 101301 (2015).

[38] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. Mohamed, and H. Arshad, Heliyon **4,** e00938 (2018).

[39] K. Hornik, M. Stinchcombe, and H. White, Neural Netw. **2,** 359 (1989).

[40] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, arXiv:1511.07289.

[41] D. P. Kingma and J. Ba, arXiv:1412.6980.

[42] S. Ioffe and C. Szegedy, arXiv:1502.03167.

[43] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg et al., J. Mach. Learn. Res. **12,** 2825 (2011), http://jmlr.org/papers/v12/pedregosa11a.html.

[44] D. Scolnic, D. Brout, A. Carr, A. G. Riess, T. M. Davis, A. Dwomoh, D. O. Jones, N. Ali, P. Charvu, R. Chen et al., Astrophys. J. **938,** 113 (2022).

[45] R. Kessler and D. Scolnic, Astrophys. J. **836,** 56 (2017).

[46] D. Foreman-Mackey, D. W. Hogg, D. Lang, and J. Goodman, Publ. Astron. Soc. Pac. **125,** 306 (2013).

[47] A. S. Bolton, S. Burles, L. V. E. Koopmans, T. Treu, and L. A. Moustakas, Astrophys. J. **638,** 703 (2006).

[48] T. Treu, L. V. Koopmans, A. S. Bolton, S. Burles, and L. A. Moustakas, Astrophys. J. **640,** 662 (2006).

[49] T. E. Collett, Astrophys. J. **811,** 20 (2015).

[50] Y. Lian, S. Cao, T. Liu, M. Biesiada, and Z.-H. Zhu, Astrophys. J. **941,** 16 (2022).