Heavy-tailed likelihoods for robustness against data outliers: Applications to the analysis of gravitational wave data

Argyro Sasli[*], Nikolaos Karnesis, and Nikolaos Stergioulas

Department of Physics, Aristotle University of Thessaloniki, Thessaloniki 54124, Greece

In recent years, the field of Gravitational Wave Astronomy has flourished. With the advent of more sophisticated ground-based detectors and space-based observatories, it is anticipated that Gravitational Wave events will be detected at a much higher rate in the near future. One of the future data analysis challenges is performing robust statistical inference in the presence of detector noise transients or nonstationarities, as well as in the presence of stochastic Gravitational Wave signals of possible astrophysical and/or cosmological origin. The incomplete knowledge of the total noise of the observatory can introduce challenges in parameter estimation of detected sources. In this work, we propose a heavy-tailed, Hyperbolic likelihood, based on the Generalized Hyperbolic distribution. With the Hyperbolic likelihood we obtain a robust data analysis framework against data outliers, noise nonstationarities, and possible inaccurate modeling of the noise power spectral density. We apply this methodology to examples drawn from gravitational wave astronomy, and in particular to synthetic data sets from the planned LISA mission.

I. INTRODUCTION

In recent years and since the first detection of a Gravitational Wave (GW) signal [1], more than 90 GW signals from the inspiral and merger of compact astrophysical objects have been included in published catalogues of GW detections achieved by the LIGO [2] and Virgo [3] ground-based detectors [4–10]. In the near future, the existing detectors (including KAGRA [11]) will operate at improved sensitivity, while the construction of LIGOIndia [12,13] is expected to start. In the meantime, a new generation of detectors (Einstein Telescope (ET) [14], Cosmic Explorer (CE) [15] and NEMO [16]) is in the planning phase, and the space-borne Laser Interferometer Space Antenna (LISA) [17] and TianQin and Taiji [18] are expected to operate after the mid-2030s. Each one of these observatories is based on different designs, which translates to different sensitivity curves aiming at different frequency ranges for the detected GW signals. One of their similarities, however, is that all 3rd-generation detectors will be characterized by higher rates of GW signals, with a significant fraction overlapping in time and frequency [19]. At the same time, instrumental noise knowledge is crucial to successfully detect and characterize GW signals [20–24]. Detector noise nonstationarities, such as slow noise Power Spectral Density (PSD) variations [25], instrumental transients (glitches) [26,27], data gaps [28,29], or noise bursts [30], may lead to biased results or even to wrong false alarm

rates. Ultimately, we expect noise nonstationarities to play a more important role in analyzing future science-rich data sets.

In this work, we focus on the case of the LISA mission and the mHz range of the GW spectrum. We have made this choice because it is certain that resolving the LISA noise budget will be a challenging task (e.g., see [20,31] and references therein). In addition, for the case of ground-based detectors, dedicated experiments can be performed on-site to calibrate the relevant noise models with high accuracy [32]. This practice is inaccessible in space.

LISA is currently in development and is expected to start collecting data after the mid-2020s. This means we currently do not have sufficient prior information about the possible outliers of the LISA noise. However, we can extrapolate from the LISA Pathfinder (LPF) mission data [33–35]. During the LPF mission, the differential acceleration noise level between 1 and 100 mHz was measured to be constantly reducing with time due to the constantly varying vacuum conditions inside the test-masses caging, throughout the duration of the mission [33,36]. Such an effect is also expected in the LISA noise data. Another example of a nonstationary stochastic GW signal in the LISA band will be generated by the ensemble signal of the Ultra Compact Binaries (UCBs) in the vicinity of our Galaxy [17,20,31,37,38]. This confusion signal will have cyclostationary properties due to the orbit of the LISA constellation facing in and out of the Galactic center, where most of these objects are located. More details for this type of signal are mentioned in Sec. III. Besides the UCBs, we

---

[*]asasli@auth.gr

also expect to detect stochastic signals originating from other types of astrophysical sources [39,40].

Another noise source that appears as a nonstationary feature in the overall PSD of the noise is the so-called tilt-to-length coupling effect (TTL). The TTL was first measured during the LPF operations, and it corresponds to the effect of space-craft jitter motion picked up by the sensitive interferometer measurement [34], generating a "bump"-like noise between 10–100 mHz [34,36] in the differential acceleration spectrum. This effect depends on the geometrical layouts of the optical components of the instruments [41,42], and can vary with time depending on environmental factors [34]. A TTL noise contribution is predicted for the LISA mission as well. Currently, the characterization plan is based on partially subtracting the effect of TTL in the postprocessing of the data [43–45]. Finally, the experience with LPF data also provided us with some insight into noisy glitches [26,27].

So far, several techniques have been developed to address each of the issues caused by noise nonstationarities. For example, one way to tackle the problems induced by rapid noisy transients is to model them and fit them simultaneously with the waveform model from the data [46]. This type of analysis is based on transdimensional methods, such as the Reversible-Jump Markov Chain Monte Carlo (MCMC) [47,48]. Following this methodology, a signal waveform is being searched in the data together with the unknown number of models representing the glitches. Then, the parameter space becomes dynamic since the number of such events is unknown. In the context of LISA data analysis, there have been quite a few applications of this technique [49–52]. Another approach was introduced in [53], which focused on the idea of testing the stationarity of the time series. Based on a surrogate data approach, this framework begins with detecting the data segments with different statistical noise properties. Then, a flexible PSD model is used to fit the different sections of the data [54].

Finally, there is the approach of tackling this kind of problem at the level of the adopted likelihood function, where models with wider tail properties than the usual Gaussian distribution can be used. Such a solution was proposed with Student's t-distribution [55–57], which has been applied to GW data by the LVK Collaboration. In [55], one begins by adopting the $\text{Inv} - \chi^2$ prior distribution for each of the PSD coefficients of the noise, which is conjugate with respect to the Gaussian likelihood. Then, the resulting joint conditional density takes the form of Student's t-distribution with $\nu > 0$ degrees of freedom, which, depending on the tuning parameter ν, has the desired heavier tails compared to the Gaussian distribution. This feature can accommodate any deviations of the data with respect to our modeled PSD of the noise, whether these originate from fast noisy transients or other types of noise nonstationarities. In addition, in [57], it was found that the uncertainties in the phase evolution of the binary

neutron star signal could be accommodated in addition to the uncertainty in the noise. However, following this approach, one should take into account that the choice of prior can have a non-negligible impact on the final posterior.

Besides the above examples, one can also follow the recipe of [58], where the non-Gaussianity of the data is modeled using a higher-order Edgeworth expansion for the unknown signal distribution. Other works [59–61], have adopted a linear combination of the Gaussian and the Log-Normal likelihoods. This was mostly done in order to accommodate the systematic bias due to data averaging. Older works also proposed using different linear combinations of distributions for the likelihood function [62,63].

Here, we introduce the Hyperbolic likelihood $\Lambda_{\mathcal{H}}$, which is based on a special case of the Generalized Hyperbolic ($\mathcal{GH}$) distribution, for the analysis of GW data. The $\mathcal{GH}$ distribution has been used in the statistical description of relatively short time series in the field of Finance and Econometrics [64–69]. The $\mathcal{GH}$ distribution has the advantage of being able to arrive to many known distributions from the exponential family [68], by simply tuning its overall shape to adjust to the given statistical properties of the input data. Thus, for example, while recovering the parameters of a transient signal, we can simultaneously recover the statistical properties of the underlying noise.

In practice, we increase the dimensionality of the problem by two or three parameters [see Eq. (21)], which control the shape of the distribution. Those are estimated from the residual data and are then used to infer their actual distribution, simply because the Hyperbolic distribution has a greater flexibility and can adjust to different distributions with perfect or acceptable accuracy, as we demonstrate. For example, if the residual data have a Gaussian distribution, then the parameters of the Hyperbolic distribution will converge to values that agree with a Gaussian distribution asymptotically (see Sec. II B for details).

Consequently, the main and most important advantage of this framework is that it comprises a robust approach for both parameter estimation for the transient signals and the characterization of the underlying noise. This feature is crucial in cases where the detector noise is "polluted" with noise transients, or even when the noise model is not completely known. As already mentioned, such will be the case of the signal-dominated LISA data, where the total noise will be the sum of the instrumental noise plus the given unresolvable stochastic GW signal [17,31,70].

In Sec. II we introduce the basic theory behind the $\mathcal{GH}$ distribution and propose the Hyperbolic likelihood, placing it in the context of detecting signals in noisy data. Some basic test cases are discussed and we apply this framework to real data from the LISA Pathfinder mission [33,34]. In Sec. III we apply this formulation to examples in GW astronomy, using synthetic data. We use different scenarios on our assumptions on the noise knowledge of the LISA

data channels and then apply this formulation to the case of a single Verification Binary taken from the LISA Data Challenge [71] catalogues. This application assumes cases where the PSD of the noise is considered either unknown or intentionally mischaracterized. In the case of unknown noise, adopting the Hyperbolic likelihood allows us to recover both the waveform and noise parameters. When the noise is assumed to be at the wrong level, the performance analysis based on the Hyperbolic likelihood is identical to the analysis based on the Whittle likelihood, which we use as a benchmark for our analyses. Finally, in Sec. IV we present our conclusions and discuss our findings.

II. THEORETICAL BACKGROUND

A. The Gaussian distribution case

We begin by assuming that a measurement $y(t)$ is the sum of a signal h that might depend on a parameter set $\boldsymbol{\theta}$ and a noise component as

$$y = h(\boldsymbol{\theta}) + n, \tag{1}$$

where we have omitted the dependence on time t for the sake of clarity. Then, assuming Gaussian properties of the noise, the likelihood of the measurement y given a parameter set $\boldsymbol{\theta}$, takes the form of

$$p(y|\boldsymbol{\theta}) = C \times e^{-\frac{1}{2}\langle y-h(\boldsymbol{\theta})|y-h(\boldsymbol{\theta})\rangle}, \tag{2}$$

where $\langle\cdot|\cdot\rangle$ denotes the noise-weighted inner product between two real time series. In the general multidimensional case of m data channels

$$\mathbf{a}(t) = \begin{pmatrix} a_1(t) \\ a_2(t) \\ \vdots \\ a_m(t) \end{pmatrix}, \quad \text{and} \quad \mathbf{b}(t) = \begin{pmatrix} b_1(t) \\ b_2(t) \\ \vdots \\ b_m(t) \end{pmatrix}, \tag{3}$$

we write the $\langle\cdot|\cdot\rangle$ in matrix form as

$$\langle\mathbf{a}|\mathbf{b}\rangle = 4\mathrm{Re}\int_0^\infty df\left[\tilde{\mathbf{a}}^\dagger(f)\mathbf{S}_n^{-1}(f)\tilde{\mathbf{b}}(f)\right], \tag{4}$$

where $\mathbf{S}_n$ becomes the one-sided cross-spectral matrix of the noise for the given arrays of time series measurements. The tilde ($\tilde{}$) denotes the Fourier transform, and the (†) represents the conjugate transpose operation. In the end, we conveniently write the log-likelihood as

$$\Lambda_{\mathcal{N}}(\boldsymbol{\theta}) \propto -\frac{1}{2}\langle y-h(\boldsymbol{\theta})|y-h(\boldsymbol{\theta})\rangle. \tag{5}$$

In Eq. (2), one assumes that the noise is Gaussian and thus one can calculate its PSD accurately, but this will not

be always the case. In case of strong glitches or other non-Gaussian features, the PSD calculation will be inaccurate, as it is based on the hypothesis of a Gaussian distribution. Also, in the case of future signal-dominated detectors, the instrumental noise and hence its PSD will be unknown for large parts of the spectrum.

Thus, a model of the spectrum of the noise can be adopted and fitted together with the parameters of the signal. Then, the logarithm of the likelihood can be written as

$$\Lambda_{\mathcal{W}}(\boldsymbol{\theta}) \propto -\frac{1}{2}\left(\sum_f[\ln(S_n(\boldsymbol{\theta}_n))] + \langle y-h(\boldsymbol{\theta}_h)|y-h(\boldsymbol{\theta}_h)\rangle\right), \tag{6}$$

where $\boldsymbol{\theta} = \boldsymbol{\theta}_n \sqcup \boldsymbol{\theta}_h$, and therefore $\boldsymbol{\theta}_n \subseteq \boldsymbol{\theta}$ the parameters of the model of the noise. The above expression constitutes the *Whittle likelihood* and is approximate for Gaussian and stationary time series [56,72,73]. In [74], the authors extend the posterior consistency result of [75] to non-Gaussian time series, providing a theoretical justification of posterior consistency under mild assumptions on the time series without having to assume Gaussianity. They suggest that this approach can be applied to non-Gaussian time series and may provide accurate spectral density estimates even if the data do not come from a normal distribution, provided that a large sample size is available. One can also go a step further and assume the uninformative and improper Jeffreys prior for the noise variance, and marginalize the noise spectrum out of the expression of Eq. (6) [55,76].

Shortcomings of the Whittle model have been extensively studied in the literature (e.g., see [56,73,77] and references therein), but one very relevant situation in GW astronomy is the measurement of time series with high autocorrelation, which could potentially reduce the efficiency of the likelihood model of Eq. (6).

As already mentioned in the introduction, a proposed solution to counterbalance those shortcomings would be to follow the strategy of [55,56], where a filter was introduced based on *Student's t-distribution*, which can be proven to be robust against data nonstationarities. One can start from Eq. (2) and adopt a prior for the variance of the noise that follows the $\mathrm{Inv}-\chi^2$ distribution. Then, following [55], we can compute the marginal posterior by integrating out the noise variance, and arrive at the desired probability density function of Student's t-distribution that is more heavy-tailed than the Gaussian case. Student's t-distribution can be tuned with the degrees-of-freedom parameter ν, which can either be chosen *a priori* or estimated directly from the data. A similar strategy was followed in other works, where the problem of data outliers was tackled by adopting a composite model for the total likelihood. For example, in [62] a Gaussian distribution was used for the noise and a

uniform model for the data bursts, while in [63], the total likelihood was a Gaussian mixture.

In the following subsection, we will discuss the generalized hyperbolic model, which, at the cost of adding extra dimensionality to the problem, offers a more generic framework to handle a variety of data-irregularity situations.

B. The Generalized Hyperbolic distribution

We can now attempt to negate the possible shortcomings of the Whittle approximation by using the family of $\mathcal{GH}$ distributions [64–68]. One of the advantages of this practice, is that we can arrive at virtually any distribution of the exponential family, simply by tuning the parameters of the $\mathcal{GH}$ function. In fact, the Student's t-distribution mentioned above is a special case of the $\mathcal{GH}$ function, which, apart from the Student's t-distribution, leads to a large number of limiting distributions. The $\mathcal{GH}$ family of distributions for a variable x can be expressed as

$$\mathcal{GH}(x|\lambda,\alpha,\beta,\delta,\mu) = a(\lambda,\alpha,\beta,\delta,\mu)(\delta^2 + (x-\mu)^2)^{(\lambda-\frac{1}{2})/2}$$
$$\times K_{\lambda-1/2}\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right)$$
$$\times \exp\left[\beta(x-\mu)\right], \quad (7)$$

where

$$a(\lambda,\alpha,\beta,\delta,\mu) = \frac{(\alpha^2 - \beta^2)^{\lambda/2}}{\sqrt{2\pi}\alpha^{\lambda-1/2}\delta^\lambda K_\lambda(\delta\sqrt{\alpha^2-\beta^2})}, \quad (8)$$

and K_λ is the modified Bessel function of the third kind. The domain of variation of the parameters $(\lambda,\alpha,\beta,\delta,\mu)$ is $\mu \in \mathbb{R}$ and

$$\delta \geq 0, |\beta| < \alpha, \quad \text{if } \lambda > 0,$$
$$\delta > 0, |\beta| < \alpha, \quad \text{if } \lambda = 0,$$
$$\delta > 0, |\beta| \leq \alpha, \quad \text{if } \lambda < 0. \quad (9)$$

Equation (7) describes a skewed distribution when $\beta \neq 0$ and a symmetric distribution when $\beta = 0$. The parameter μ tunes the position of the distribution along the x-axis. The analytic expression for the variance is [68]

$$\sigma^2 = \frac{\delta K_{\lambda+1}(\delta\gamma)}{\gamma K_\lambda(\delta\gamma)} + \frac{\beta^2\delta^2}{\gamma^2}\left[\frac{K_{\lambda+2}(\delta\gamma)}{K_\lambda(\delta\gamma)} - \frac{K_{\lambda+1}^2(\delta\gamma)}{K_\lambda^2(\delta\gamma)}\right], \quad (10)$$

with $\gamma \equiv \sqrt{\alpha^2 - \beta^2}$. The general expression for the mean of the $\mathcal{GH}$ distribution is [68]

$$\text{mean}(\mathcal{GH}) = \mu + \frac{\beta\delta}{\gamma}\frac{K_{\lambda+1}(\delta\gamma)}{K_\lambda(\delta\gamma)}. \quad (11)$$

The $\mathcal{GH}$ distributions have semiheavy tails. In particular, for $\mu = 0$, the asymptotic behavior for x $\rightarrow \pm\infty$ is

$$\mathcal{GH}(x|\lambda,\alpha,\beta,\delta) \sim |x|^{\lambda-1}\exp((\mp\alpha+\beta)x), \quad (12)$$

up to a multiplicative constant [78].

We can now define, as in [68], the *multivariate, d-dimensional* $\mathcal{GH}$ distribution for $x \in \mathbb{R}^d$, with d being the given dimensionality. In the case of a GW detector network, the dimensionality corresponds to the number of detectors. Then,

$$\mathcal{GH}_d(x|\lambda,\alpha,\beta,\delta,\mu) = A\frac{K_{\lambda-d/2}(\alpha\sqrt{\delta^2+r})}{(\alpha^{-1}\sqrt{\delta^2+r})^{d/2-\lambda}}\exp\left[\beta^{\text{T}}(x-\mu)\right], \quad (13)$$

with $\mu,\beta \in \mathbb{R}^d$,

$$A \equiv A(\lambda,\alpha,\beta,\delta,\mu) = \frac{(\sqrt{\alpha^2-B}/\delta)^d}{(2\pi)^{d/2}K_\lambda(\delta\sqrt{\alpha^2-B})}, \quad (14)$$

where we set

$$r = (x-\mu)^{\text{T}}\Delta^{-1}(x-\mu), \quad (15)$$

and $B = \beta^{\text{T}}\Delta^{-1}\beta$. In (15), Δ is a positive definite matrix $\in \mathbb{R}^{d\times d}$ with $|\Delta| = 1$.

The $\mathcal{GH}$ distribution can be quite flexible in the description of data, albeit at the cost of a larger number of parameters. For a given problem with d dimensions, the number of parameters is $2d+3$.

C. Limiting cases and subclasses of the Generalized Hyperbolic distribution

Several well-known distributions are limiting cases or subclasses of the $\mathcal{GH}$ distribution. Below, we list some cases that we are going to use to set up test cases in Sec. II E 1..

(i) The *Normal Gaussian* distribution $\mathcal{N}(\mu,\sigma^2)$ is a limiting case of the $\mathcal{GH}$ distribution, since

$$\mathcal{N} \leftarrow \mathcal{GH}(\lambda = 1, \beta = 0, \alpha \rightarrow \infty, \delta \rightarrow \infty, \mu), \quad (16)$$

provided that $\alpha \rightarrow \infty$ and $\delta \rightarrow \infty$ in such a way that $\delta/\alpha \rightarrow \sigma^2$.

(ii) The *Normal Inverse Gaussian* distribution $\mathcal{NIG}(\alpha,\beta,\delta,\mu)$ has the explicit form [79]

$$\mathcal{NIG}(x|\alpha,\beta,\delta,\mu) = \frac{\alpha\delta K_1(\alpha\sqrt{\delta^2+(x-\mu)^2})}{\pi\sqrt{\delta^2+(x-\mu)^2}}$$
$$\times \exp\left[\delta\gamma + \beta(x-\mu)\right], \quad (17)$$

which is a subclass of the $\mathcal{GH}$ distribution for $\lambda = -1/2$

$$\mathcal{NIG}(\alpha,\beta,\delta,\mu) = \mathcal{GH}(\lambda = -1/2, \alpha,\beta,\delta,\mu). \quad (18)$$

(iii) The *Hyperbolic* distribution $\mathcal{H}$ is a subclass of the $\mathcal{GH}$ distribution for $\lambda = 1$

$$\mathcal{H}(\alpha, \beta, \delta, \mu) = \mathcal{GH}(\lambda = 1, \alpha, \beta, \delta, \mu), \quad (19)$$

with $\delta \geq 0$ and $|\beta| < \alpha$. For $\beta = 0$ it becomes the *Symmetric Hyperbolic distribution* ($\mathcal{SH}$). More generally, in the multivariate, d-dimensional case, the *multivariate Hyperbolic* ($\mathcal{H}_d$) distribution is as subclass of the $\mathcal{GH}_d$ distribution for $\lambda = (d+1)/2$

$$\mathcal{H}_d(\alpha, \beta, \delta, \mu) = \mathcal{GH}_d\left(\lambda = \frac{d+1}{2}, \alpha, \beta, \delta, \mu\right). \quad (20)$$

(iv) The *Student's t-distribution* with ν degrees of freedom is a limiting case of the $\mathcal{GH}$ distribution when $\lambda = -\nu/2, \alpha = 0, \beta = 0, \delta = \sqrt{\nu}$ and $\mu = 0$.

In the Appendix, we demonstrate that one can recover the theoretical distributions of the above subclasses and limiting cases for test data, using Bayesian inference with appropriate likelihood functions.

We refer the reader to [64–68,80] and references therein for a more extensive list of limiting distributions and subclasses of the $\mathcal{GH}$ distribution.

D. The Hyperbolic likelihood function

Our main aim is to use a likelihood distribution that is sufficiently flexible, as to be able to reconstruct a large variety of distributions of data. As we demonstrate in the different examples in the next sections, we find that a likelihood based on the multivariate Hyperbolic ($\mathcal{H}_d$) distribution (i.e., the subclass of the $\mathcal{GH}_d$ distribution when $\lambda = (d+1)/2$) serves our purpose well.

For a distribution of data $x_i \in \mathbb{R}^d, 1 \leq i \leq n$, we derive the corresponding Hyperbolic likelihood function for inferring the parameters (α, δ, β) of their distribution as[1]

$$\Lambda_{\mathcal{H}}(\alpha, \delta, \beta) = n\left[\frac{d+1}{2}\ln\left(\frac{\gamma}{\delta}\right) + \frac{1-d}{2}\ln(2\pi)\right.$$
$$\left. - \ln(2\alpha) - \ln\left[K_{(d+1)/2}(\delta\gamma)\right]\right]$$
$$- \alpha\sum_{i=1}^{n}\sqrt{\delta^2 + r_i} + \beta^{\mathrm{T}}\sum_{i=1}^{n}x_i, \quad (21)$$

where

$$r_i = x_i^{\mathrm{T}}\hat{\Delta}^{-1}x_i. \quad (22)$$

We are going to use this form of the likelihood in test cases in Sec. II E, where, for $d = 1$, r_i reduces to $r_i = x_i^2$.

----

[1]Our result reduces to the hyperbolic distribution Λ_{hyp} in [66,68] for $\beta = 0$.

For more realistic applications, we will adopt the above Eq. (21) in the frequency domain. In particular, in Sec. III we adopt the *symmetric* Hyperbolic likelihood (i.e., $\beta = 0$). For multiple data channels—see Eq. (4)—and assuming that (α, δ) are the same for all frequencies, we then write

$$\Lambda_{\mathcal{H}}(\theta, \alpha, \delta) = n\left[\frac{d+1}{2}\ln\left(\frac{\alpha}{\delta}\right) + \frac{1-d}{2}\ln(2\pi)\right.$$
$$\left. - \ln(2\alpha) - \ln\left[K_{(d+1)/2}(\delta\alpha)\right]\right]$$
$$- \alpha\sum_{i}\sqrt{\delta^2 + \mathrm{Re}\{r_i(\theta)\}}, \quad (23)$$

where

$$r_i(\theta) = \tilde{\chi}_i^{\dagger}\mathbf{S}_n^{-1}\tilde{\chi}_i. \quad (24)$$

In the above relation, we have used

$$\tilde{\chi}_i = \sqrt{2\mathrm{d}f}\left(\tilde{y}_i - \tilde{h}_i(\theta)\right), \quad (25)$$

where $\mathrm{d}f$ is the given frequency resolution. The summation in Eq. (23) is over elements that correspond to a chosen frequency range $(f_{\min}, f_{\max})$.

Equation (24) implies that we have chosen a given estimate (or model) for the PSD of the noise, through $\mathbf{S}_n$. One of the great advantages of this likelihood formulation, as demonstrated in the following sections, is that a mis-modeling of the noise PSD could be compensated by the joint estimation of the (α, δ) parameters. These two parameters will also indicate possible departures from a Gaussian distribution, as described in Sec. II C. In the rather exceptional case where one does not want to assume any PSD model, but instead prefers to infer the noise properties by using the (α, δ) parameters of the Hyperbolic likelihood, then one can use

$$r_i(\theta) = \tilde{\chi}_i^{\dagger}\hat{\Delta}^{-1}\tilde{\chi}_i, \quad (26)$$

where now, since we operate in the frequency domain, and in order to arrive at a dimensionless r_i, the $\hat{\Delta}$ has units of $1/\mathrm{Hz}$ (alternatively, one could make $\tilde{\chi}_i$ dimensionless, by scaling it with an appropriate reference value with units of $1/\sqrt{\mathrm{Hz}}$).

After defining the likelihood function, and since we operate within a Bayesian framework, we can write the posterior of the parameters as

$$p(\theta|y) \propto p(y|\theta)p(\theta), \quad (27)$$

with $p(\theta)$ being the prior of the parameters. The marginal likelihood $p(y)$ usually acts as a normalization constant and for that reason it is omitted from Eq. (27).

E. Test cases

Before applying the hyperbolic likelihood $\Lambda_{\mathcal{H}}$ to examples in GW astronomy, we first demonstrate its ability to reconstruct non-Gaussian test distributions (several analytic distributions, as well as real data from the LISA Pathfinder [33,34] mission).

1. Reconstruction of one-dimensional test distributions

In the following, we will reconstruct different one-dimensional test distributions (corresponding to the liming cases and subclasses of the $\mathcal{GH}$ distributions discussed in Sec. II C), assuming that the $\mathcal{H}$ distribution can approximately describe the difference cases (even those with $\lambda \neq 1$). For each case, we infer the $(\alpha, \beta, \delta, \mu)$ parameters of $\mathcal{H}$ using a Metropolis-Hastings MCMC algorithm and $\Lambda_{\mathcal{H}}$ (with $d = 1$) as the likelihood function. The purpose of this first simple investigation is to demonstrate the flexibility of the $\Lambda_{\mathcal{H}}$ likelihood.

For the sake of simplicity, we generate[2] all data (5×10^5 samples for each test distribution) with zero-mean ($\mu = 0$). The detailed description of each test distribution is as follows:

(1) Normal Gaussian distribution $\mathcal{N}(\mu = 0, \sigma^2 = 1)$.
(2) Normal Inverse Gaussian distribution $\mathcal{NIG}(\alpha = 1, \beta = 0, \delta = 0.8, \mu = 0)$.
(3) Hyperbolic distribution, $\mathcal{H}(\alpha = 1.5, \beta = 0.75, \delta = 2, \mu = 0)$.
(4) Student's t-distribution with $\nu = 4$.

The generated test distributions described above are shown in Fig. 1. The solid lines represent the theoretical distribution in each case. The dashed lines represent our reconstruction using the $\Lambda_{\mathcal{H}}$ likelihood for all cases. We see that the two cases for which $\lambda = 1$ ($\mathcal{N}$ and $\mathcal{H}$) are reconstructed with high accuracy (the Jensen-Shannon divergence[3] is smaller than 10^{-5}), but, also, the two cases with $\lambda \neq 1$ ($\mathcal{NIG}$ and Student's t) are still reconstructed fairly accurately (the Jensen-Shannon divergence is 2.5×10^{-4} and 5.1×10^{-4}, respectively) when using the same $\Lambda_{\mathcal{H}}$ likelihood. This demonstrates the flexibility of the $\Lambda_{\mathcal{H}}$ likelihood to reconstruct different distributions of data and to compensate for different underlying values of λ through its other parameters.

Finally, one can use the $\{\alpha, \beta, \delta\}$ parameters to probe the statistical properties of the residuals. For example, in many cases, it is necessary to test for any departures from Gaussianity for a given data set. A very useful graphical tool for such applications is the so-called *shape triangle*, where, based on the recovered $\{\alpha, \beta, \delta\}$ coefficients, we can

FIG. 1.   Randomly generated data with Gaussian $\mathcal{N}$ (yellow), hyperbolic $\mathcal{H}$ (purple), $\mathcal{NIG}$ (red), and Student's t (light blue) distributions. The reconstructed distributions, using Bayesian inference with $\Lambda_{\mathcal{H}}$ as the likelihood function (solid lines), agree very well with the theoretical distributions (dashed lines).

qualitatively characterize the yielding distribution of the residuals [68,69]. We present more details about this methodology in the Appendix.

2. Application to real data from the LISA Pathfinder mission

LISA Pathfinder (LPF) was an European Space Agency mission launched in late 2015 and remained in operation until 2017 [33,34]. The primary goal of LPF was to test technologies for future GW observatories in space, such as the LISA mission. LPF was essentially a laboratory in space, which contained two cubic test masses maintained in free-fall conditions. The differential acceleration $\Delta g(t)$ between the two test masses was being monitored by means of laser interferometry. During the mission, the differential acceleration noise budget was studied and modeled, with the aim of building solid grounds for the development of the LISA mission. One of the most important measurements of the LPF mission was the pure acceleration noise measurements, where no experiments (excitations of the three-body system) were performed. During those measurements, a great variety of spurious transient signals (glitches) were recorded. While the physical origin of those glitches is not entirely known [26], their statistical properties were studied in detail [27]. For more studies on the LPF acceleration noise data, we refer the reader to [35,36,84–86].

For our example here, we take the $\Delta g(t)$ segment that was recorded between 2017-02-13 14:30:00.000 and 2017-03-02 21:50:19.000 UTC. This particular segment contained a series of glitches, with the loudest of them (shown in the top panel of Fig. 2) distorting the numerical estimations of the PSD of the data. In fact, such strong spurious signals were straightforwardly detected and

---

[2]The random values are generated with the SCIPY.STATS package [81].

[3]The Jensen-Shannon divergence [82,83] is a special case of the Kullback-Leibler divergence, which can be used as a metric for the similarity of two distributions. It is symmetric, and takes values close to zero if two distributions are similar.
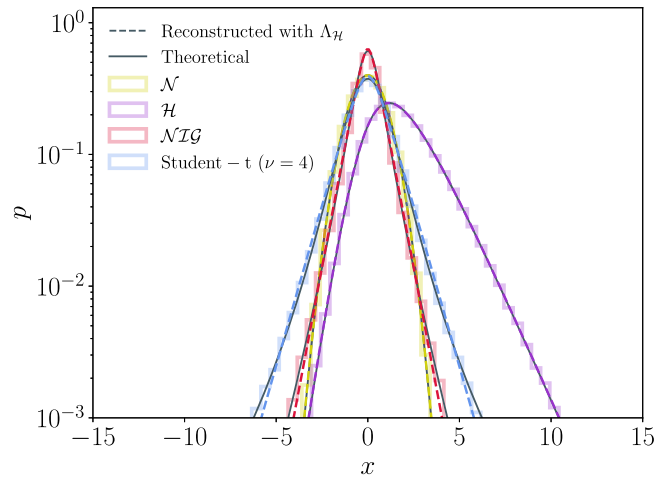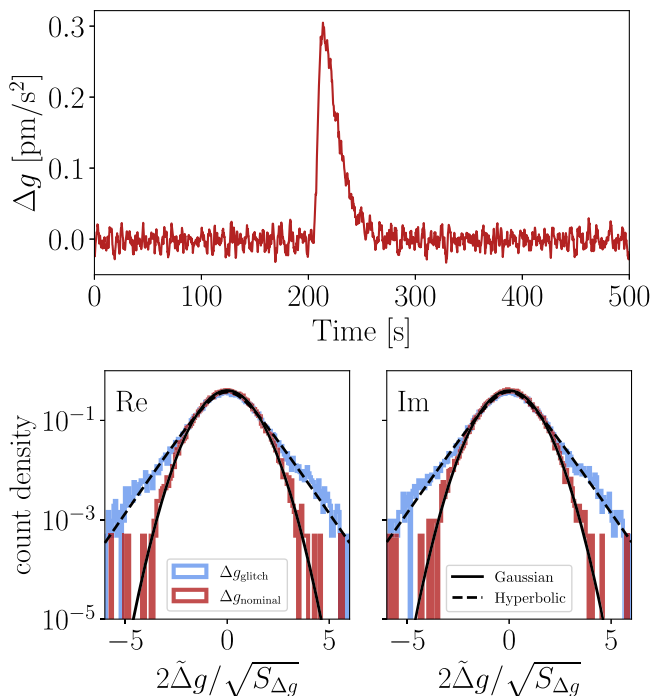
FIG. 2. *Top panel:* A loud glitch in the $\Delta g(t)$ time series data from the LPF mission (see text for details). The data presented here are detrended by subtracting their mean and low-passed for visualization purposes. *Bottom panel:* Histograms of the distributions $\tilde{\Delta} g$ in the frequency domain, normalized with the PSD $S_{\Delta g}$ of a nearby-in-time, outlier-free segment. *Blue:* segment that contains the loud glitch. *Red:* reference data set. The dashed and dotted lines represent Gaussian and Hyperbolic distributions, respectively.

subtracted from the data. However, this situation of data outliers is ideal for testing the heavy-tailed likelihood framework we introduced in Sec. II D.

We first crop[4] the $\Delta g(t)$ time series and determine the time of occurrence of the loudest glitch signals. We also mark the $\Delta g(t)$ segments when no glitches are present, with the aim of using them as benchmarks for the PSD of $\Delta g(t)$. Then, we proceed and apply the Hyperbolic likelihood function Eq. (21) for the distribution of the $\tilde{\Delta} g(f)$ data in the frequency domain and infer the free parameters α, β, and δ. In order to avoid getting trapped in local maxima on the posterior surface, we have used MCMC methods [48].

We plot the results in the bottom panel of Fig. 2, where the histograms of the distribution of the noise data points are plotted. The distribution of the Δg noise for the particular segment that contains the loud glitch is represented in red color. For the sake of comparison, we have also chosen a neighboring-in-time segment, where no data outliers are present (blue data). The hyperbolic density which is computed at the estimated α, β, and δ is shown

---

[4]We remove the first and last few data points to avoid distortions due to the application of filters.

with the dotted line, while the theoretical Gaussian distribution is represented with the dashed line. The heavy-tailed hyperbolic distribution, Eq. (21), indeed manages to correctly capture the tails of the distribution in the case where a loud glitch exists in the data, whereas the Gaussian distribution is suitable only for the segment where no large glitch was detected. We note that, due to the nearly symmetric distribution of the data, the parameter β was inferred to be very close to zero and a symmetric hyperbolic likelihood would still be a good choice.

These first results are very encouraging, thus we can now proceed and apply this framework to more complicated investigations. Such cases, as already discussed, are situations of having to search for signals in noisy data, where it is crucial to be able to correctly model the statistical properties of the residuals and the underlying noise. In the following sections, we will apply the hyperbolic likelihood function in more realistic applications drawn from challenges in Gravitational Wave data analysis.

In this first study, we are going to focus on symmetric hyperbolic distributions (i.e., with $\beta = 0$) for two main reasons. Firstly, we have found that in the simple applications presented in this work, the symmetric Hyperbolic distribution is more than sufficient, and, secondly, because it is lighter in terms of computational requirements. Indeed, for a measurement of n data points, the number of Bessel function calculations is reduced to one, instead of $n + 1$ for the case of asymmetry, where $\beta \neq 0$.

III. APPLICATIONS TO THE ANALYSIS OF GRAVITATIONAL WAVE DATA

The detection and characterization of any given signal requires a good level of knowledge of the instrumental noise. Concerning the GW measurements from ground-based detectors, there has been extensive work on the statistical properties of the instrumental noise (e.g., [87–89] and references therein). This, together with monitoring the status of the instruments through a large number of auxiliary channels, contributes to the accurate representation of the instrument capabilities, virtually at all times.

However, future detectors, such as LISA, are going to be signal-dominated, and the noise of the observatory will not be completely known. In particular, LISA is going to measure the complete population of Ultra Compact Binaries emitting from within the Milky Way [17]. These objects are mostly Double White Dwarfs, and their total number is estimated to be of order $\mathcal{O}(10^6)$, emitting nearly monochromatic GW radiation. Depending on the given population model, current estimates place the total number of sources resolvable by LISA at $\mathcal{O}(10^4)$, while the rest will generate a nonstationary confusion signal between about 0.01 and 0.2 mHz [31,37,38]. At the same time, the number of calibration instruments available on board a space-borne observatory (thermometers, magnetometers, etc.) will be fairly restricted, due to space and power supply limitations.

The above, in combination with the limited data transmission rate, might contribute to having less information about the instrumental noise, when compared to ground-based observatories. We will utilize examples drawn from LISA data scenarios, focusing mostly on analyzing signals from UCBs. In the following section, we will compare the performance of the different likelihood formulations described in Sec. II, under different noise assumptions.

A. Methodology

In order to test the performance of the Hyperbolic likelihood, we designed a series of experiments with synthetic data. We chose to work only with the signals from UCBs, mostly due to their monochromatic nature, and also for practical reasons, such as the very small computational time. However, it should be mentioned here, that our formulation applies straightforwardly to the chirping signals of supermassive black hole binaries. For those signals, and the simplest noise case of a varying PSD level across the complete frequency band (same spectral shape, different overall amplitude), we obtain very similar results as with the UCBs investigations presented in this section. The situation becomes more complicated, however, when the noise PSD level assumed in the likelihood function has a different spectral shape than the true noise PSD. We leave this kind of investigation to future work.

For our first experiment, we simulate data given a particular instrument sensitivity and then perform the analysis assuming fixed PSD of the noise. We use three likelihood formulations. The first type refers to the standard Gaussian likelihood of Eq. (5), but with an assumed noise PSD which is chosen to be different than the one that was used to generate the synthetic data, simulating situations of instrumental noise mismodeling. The second type refers to the Whittle likelihood of Eq. (6), where the model for the noise PSD is being fitted simultaneously with the GW waveform parameters. Finally, the third refers to the hyperbolic likelihood of Eq. (23), with fixed PSD for the noise as in the case of the Gaussian likelihood above, but with the added flexibility of extra parameters.

It is worth noting here, that for our first experiments, we have assumed ideal synthetic data. This means that we simulate perfectly Gaussian and uninterrupted data, directly from the LISA sensitivity curves [90]. At the same time, we assume a rigid LISA constellation, while the test-masses acceleration noise and spacecraft interferometric noises are assumed to be equal for all three spacecraft [91]. For this special case, we can confidently utilize the noise-orthogonal[5] Time Delay Interferometry (TDI) A, E, and T channels [90,92], which are retrieved from the detector outputs X, Y, Z as

---

[5]Under the assumptions of [90,92], the A and E channels are uncorrelated and only the diagonal components of their covariance matrix survive.

$$A = \frac{1}{\sqrt{2}}(Z - X), \qquad E = \frac{1}{\sqrt{6}}(X - 2Y + Z),$$
$$T = \frac{1}{\sqrt{3}}(X + Y + Z). \tag{28}$$

The T channel is the so-called *null* channel, which greatly suppresses the GW signal. Although the T TDI channel seems useful when considering the complete frequency band, it is neglected in our analysis because, as described below, we will focus on a narrow frequency band. Under our assumptions, the A and E channels used in the analysis will have the same spectral shape, which under given assumptions about the future instrument, is given by [90,93]

$$S_n^{\mathrm{AE}}(f) = 8\sin^2(f^*)\left[2S_{\delta\nu/\nu}^{\mathrm{acc}}(A;f)(3 + 2\cos(f^*) + \cos(2f^*))\right.$$
$$\left. + S_{\delta\nu/\nu}^{\mathrm{OMS}}(P;f)(2 + \cos(f^*))\right], \tag{29}$$

where $f^* = 2\pi L f/c$, L is the LISA arm-length, while the noise components $S_{\delta\nu/\nu}^{\mathrm{acc}}(A;f)$ and $S_{\delta\nu/\nu}^{\mathrm{OMS}}(P;f)$ can be found in [90], expressed in relative frequency units (thus the index $\delta\nu/\nu$). These functions depend on parameters A and P, respectively, for which a conservative estimate consistent with current requirement levels for the instrument is [61,90,93]

$$\sqrt{A} = 3 \ \mathrm{fm/sec^2/Hz}, \quad \mathrm{and}$$
$$\sqrt{P} = 15 \ \mathrm{pm/Hz}. \tag{30}$$

Considering the above, the likelihood computation of Eqs. (5), (6), and (23) reduces to just the sum of the likelihoods for the two uncorrelated A and E channels.

Finally, in order to perform our analysis, we sample the posterior of the parameters with Markov Chain Monte Carlo algorithms, enhanced with *parallel tempering* [48]. For the applications in the sections below, we have set up our sampler with an adjusting temperature ladder of 30 temperatures, each running with 50 independent walkers. We nominally allow for 5×10^4 samples for the burn-in phase, and 10^5 samples per walker for the parameter estimation phase.

B. Ultra compact galactic binaries

As already discussed, the UCBs emitting from the vicinity of our Galaxy are mostly Double White Dwarfs (DWDs) [38,94,95]. Those will be the most numerous types of sources in the LISA band, with millions of them emitting in the mHz range. For our study here, we will perform parameter estimation on a single binary embedded in Gaussian noise, with a PSD given by Eq. (29). For the actual waveforms, we utilize the fast frequency-domain implementation that was first presented in [96]. The two polarizations are then written as

TABLE I. The parameter values of the injected UCB signal in the data. This injection was used for the investigations in both Secs. III B 2 and III B 3.

Parameters	Value
Amplitude, $\log_{10}(\mathcal{A}[\text{Strain}])$	-21.37729133
Initial frequency, f_0 [mHz]	2.61301
Frequency derivative, $\log_{10} \dot{f}_0$	-16.53675992
Inclination, $\cos(\iota\ [\text{rad}])$	0.05407993
Ecliptic Latitude, $\sin(\beta_{\text{E}}\ [\text{rad}])$	0.1012303
Ecliptic Longitude, λ_{E} [rad]	4.052962883
Polarization, ψ [rad]	0.80372815
Phase, ϕ_0 [rad]	3.12184095

$$h_+(t) = \frac{2\mathcal{M}}{D_L}(\pi f_0(t))^{2/3}(1 + \cos^2 \iota)\cos \psi(t),$$

$$h_\times(t) = -\frac{4\mathcal{M}}{D_L}(\pi f_0(t))^{2/3}\cos \iota \sin \psi(t), \qquad (31)$$

with $\mathcal{M}$ being the chirp mass, f_0 the instantaneous gravitational wave frequency, D_L the luminosity distance, ι the inclination of the binary orbit and ψ the gravitational wave phase over time. For more details about the waveform model, we refer the reader to [96–98]. The final inferred parameters are $\boldsymbol{\theta} = \{\log_{10}\mathcal{A}, f_0 \log_{10}\dot{f}_{\text{gw}}, \phi_0, \cos \iota, \psi, \lambda_{\text{E}}, \sin \beta_{\text{E}}\}$, where $\mathcal{A}$ is the overall amplitude, which, in relation to Eqs. (31), can be expressed as

$$\mathcal{A} = D_L^{-1}(2\mathcal{M}^{5/3}\pi^{2/3}f_0^{2/3}). \qquad (32)$$

Above, λ_{E} and β_{E} are the given ecliptic latitude and longitude, respectively. The graphics processing unit-accelerated software we have used can be found in [99], which is essentially an adaptation of the waveform implemented for the LISA Data Challenges (LDC) software [100] and in earlier works. The injected parameters for the single binary are presented in Table I.

1. Unknown noise spectrum

We first investigate the case, where the true PSD of the noise is completely unknown. A partial solution for this problem would be to assume a parametrized model of the noise PSD and infer its parameters simultaneously with the signal, using the Whittle likelihood of Eq. (6). Here, instead, we will use the Hyperbolic likelihood $\Lambda_{\mathcal{H}}$ of Eq. (23), with $\boldsymbol{r}_i(\boldsymbol{\theta})$ given by Eq. (26), without making an explicit assumption about the noise PSD.

By adopting the noise-orthogonal A and E channels, Eq. (26) reduces to

$$r_i(\boldsymbol{\theta}) = 2\text{d}f \sum_c^{\{A,E\}} |\tilde{y}_c - \tilde{h}_c(\boldsymbol{\theta})|^2. \qquad (33)$$

Essentially, this corresponds to using a unitary diagonal $\hat{\Delta}$ matrix.

Given the simulated data, we should recover a set of α and δ parameters that would reflect the characteristics of the noise. Since, in our example, we generated Gaussian noise, the posterior distribution of α and δ should reflect the characteristics of the Gaussian normal distribution that we have adopted for the simulations.

For this investigation, we can treat the instrumental noise as white around the injection frequency, simply because this type of signal is almost monochromatic and, thus, the variation of the noise PSD in a narrow frequency band around the signal is negligible [see Fig. 3(a)]. This further simplifies our analysis because it allows us to use the common α and δ parameters for all frequencies considered in the analysis. In other cases, such as when one simulates multiple injections at different frequencies, different α, δ parameters will need to describe different parts of the spectrum as demonstrated in an example using Student's t-distribution in [56]. Another simplification is that we use the same α and δ coefficients for both the A and E TDI variables, assuming that the two data channels have the same noise levels.[6]

Considering the above, we should expect to recover parameters that while $\alpha \to \infty$ and $\delta \to \infty$, their ratio remains constant and converges to the variance of the noise around the source initial emission frequency f_0. Concerning the signal, we draw a relatively "loud" source from the list of Verification Binaries [71,100], which we further tune to get a slightly higher SNR. The injected waveform parameters are listed in Table I, while our synthetic data set has a duration of $T_{\text{obs}} = 1$ year, which gives a frequency resolution of $\sim 10^{-8}$ Hz.

We then perform parameter estimation simultaneously for the waveform parameters listed in Table I and the parameters α, δ of the $\Lambda_{\mathcal{H}}$ likelihood. Figure 3(b) displays the 2D posterior distribution of the parameters α, δ. For a symmetric distribution, Eq. (10) reduces to $\sigma^2 \to \delta/\alpha$ in the limit of a Gaussian distribution,[7] i.e., when $\alpha \to \infty$ and $\delta \to \infty$. The dashed red line in Fig. 3(b) represents $\log_{10}\delta - \log_{10}\alpha = \log_{10}\sigma^2$, where σ^2 here corresponds to the "correct" value of the power spectral density $S_n^{\text{AE}}(f_0)$ in the same dimensionless units we chose for the data. From the alignment of the 2D posterior distribution with this line, it follows that the inferred parameters α, δ adjust the Hyperbolic distribution to the true Gaussian distribution of the data residuals.

---

[6]For future applications, it would be interesting to generalize the $\mathcal{GH}$ distribution to $\alpha, \delta \in \mathbb{R}^d$.

[7]Notice that, as evident from Fig. 3(a), $S_n^{\text{AE}}(f_0) \sim 10^{-42}$ Hz^{-1}, which explains the particular numerical values of α, δ in the chosen system of units, where $\hat{\Delta}$ is assumed to have dimensions of $1/$Hz without any further scaling of the $\tilde{\chi}_i$ data. With appropriate rescaling, one can bring both α, δ in the regime of $\to \infty$.
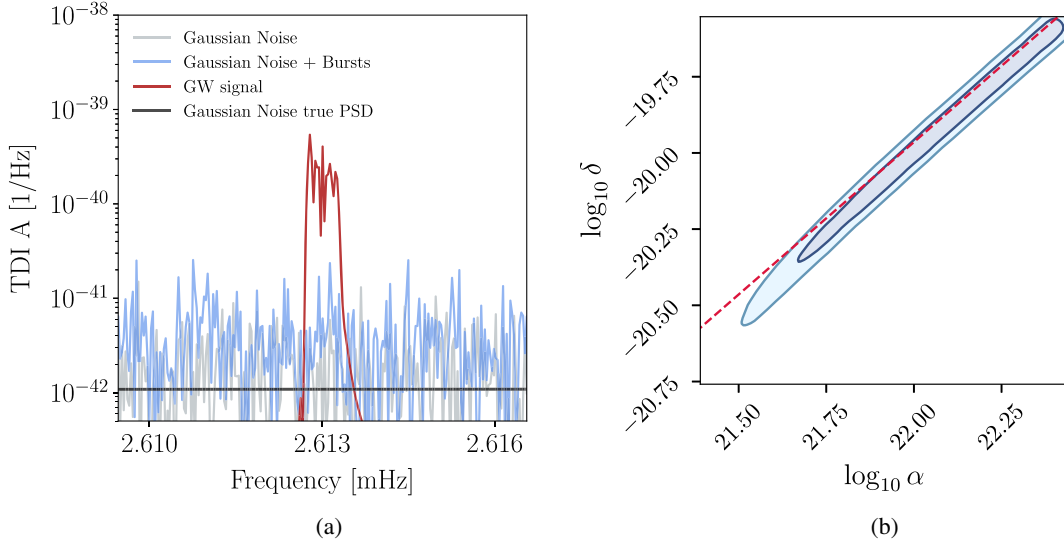
FIG. 3. *(a)* The different sets of data used in the investigations of Sec. III. In the given frequency range, the PSD of the instrumental noise is shown with the solid black line [see Eqs. (29) and (30)]. This model, with added Gaussian noise was used in order to generate the data shown in gray color. For the investigation in Sec. III B 3, we additionally injected a series of Gaussian bursts, which resulted in the noise data shown in blue color. The injected UCB signal is shown in dark red (see Table I). *(b)* The 2D posterior slice between the parameters α and δ of the Hyperbolic likelihood defined in Eqs. (23) and (33), for the first investigation of Sec. III B 2. The contours represent the sampled posterior surface at 67% and 90% C.I., while the dashed red line corresponds to the logarithm of the true injected PSD levels of the instrumental noise (see text for details).

Finally, the waveform parameters are also recovered with identical posteriors as the baseline Gaussian likelihood case, where the noise in Eq. (5) is set to the true PSD value. This first result demonstrates the capabilities and robustness of the Hyperbolic likelihood in recovering both the correct parameters of the noise and the statistical properties of the residuals.

2. Noise spectral model mismodeling

Our next step is to test this framework in a different setting, which focuses on mismodeling the PSD of the noise. This situation is a bit closer to reality, where we start making some assumptions about the underlying PSD of the noise. This will be the case for LISA around the mHz level, due to the unknown contribution of the stochastic signal generated by astrophysical populations [31,38–40,70].

In practice, we simulate a mismodeling situation by plugging a wrong PSD model of the noise to the likelihood function of Eqs. (5), (6), and (23). Since our signal is almost monochromatic, this translates to an overall PSD level difference between the true and adopted models for the noise. We use the same source signal as before (see Table I), and simulate data for an observation duration of $T_{\rm obs} = 1$ year. Using a PSD based on Eq. (29) and assuming that the values of Eq. (30) correspond to the "correct" noise levels, we obtain a signal-to-noise ratio (SNR) of $\rho = \sqrt{\langle h|h\rangle} = 122.7$.

Next, we assume that we do not know the "correct" noise levels and are forced to guess, adopting a different,

"wrong" PSD, using Eq. (29), but with the "wrong" values,

$$\sqrt{A} = 2.5 \text{ fm/sec}^2/\text{Hz} \quad \text{and}$$
$$\sqrt{P} = 11 \text{ pm/Hz}, \tag{34}$$

which means that by adopting the above model, we greatly underestimate the level of the noise by a factor of ~1.57 at the particular emission frequency of the binary. We then sample the parameter space with parallel tempering MCMC [48] and investigate three different cases.

(1) The *first* parameter estimation analysis was performed by adopting the standard Gaussian likelihood ($\Lambda_{\mathcal{N}}$) of Eq. (5), and assuming the "wrong" PSD levels of Eq. (34) for the noise.

(2) The *second* investigation was analyzed using the Whittle likelihood ($\Lambda_{\mathcal{W}}$) of Eq. (6), with a model of the PSD that represents a flat PSD spectrum around the frequencies of interest ($f_{\rm min} = f_0 - 1$ [mHz], $f_{\rm max} = f_0 + 1$ [mHz]) multiplied by a single parameter $n_{\rm level}$. For $n_{\rm level} = 1$, we take the PSD to be the value given by Eq. (29) at $f = f_0$, but in which the "wrong" parameters of Eq. (34) were used. By fitting for the free parameter $n_{\rm level}$, we essentially search for a correction to the adopted "wrong" PSD.

(3) Finally, a *third* case that we consider, is to use the Hyperbolic likelihood formulation ($\Lambda_{\mathcal{H}}$) of Eqs. (23) and (24), again assuming the "wrong" PSD parameters of Eq. (34) for the noise. Here, we allow the
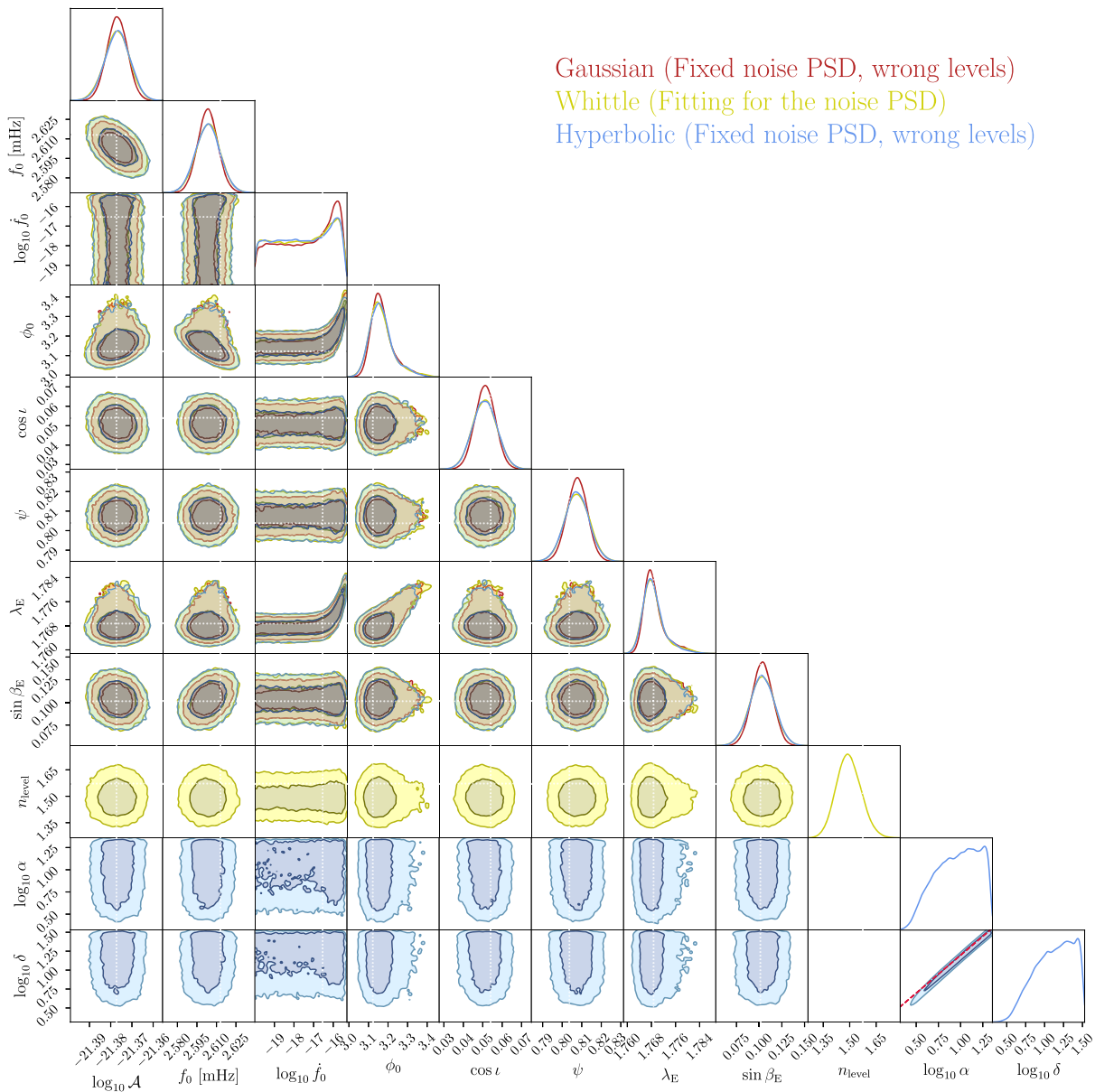
FIG. 4. Corner plot showing the 2D posterior slices for all the parameters that characterize the injected waveform and the properties of the noise, for the case of noise mismodeling of Sec. III B 2. The optimal signal-to-noise ratio is 122.7. In particular, with dark red we plot the posteriors for the case of the Gaussian likelihood $\Lambda_{\mathcal{N}}$ with PSD at a wrong level, whereas with yellow we plot the posteriors of the Whittle likelihood ($\Lambda_{\mathcal{W}}$), which allows for fitting of the noise level. Finally, with light blue, we plot the posteriors retrieved with the Hyperbolic likelihood, with the wrong PSD levels in the whitening. The true injected values for all parameters (with the exception of α and δ for the Hyperbolic likelihood case) are shown with white dashed lines. The red dashed curve represents the $\log_{10} \delta - \log_{10} \alpha = \log_{10} n_{\text{level}}^{(\text{correct})}$ line.

free parameters α, δ of ($\Lambda_{\mathcal{H}}$) to offset the wrong assumption of the PSD. This application highlights one of the advantages of the Hyperbolic likelihood, which is the detection of any deviations from a given assumed noise model, or any kind of departures from the assumed Gaussianity of the residuals.

To validate the performance of each parameter estimation run, we compare them to the baseline case, which is the Gaussian likelihood with the correct noise model described by Eq. (29) and the parameters of Eqs. (30).

The results of this application are summarized in Fig. 4 and in Table II. It is quite evident from the corner plot of the 2D posterior distributions in Fig. 4, that the naive approach of using the Gaussian likelihood $\Lambda_{\mathcal{N}}$ with a PSD based on the "wrong" noise parameters of Eq. (34), yields underestimated posterior widths. On the other hand, when

TABLE II. Jensen-Shannon (JS) divergence of the marginal posterior probability density functions of all parameters, between the three different likelihood choices with wrong PSD levels in Sec. III B 2 and the baseline model of a Gaussian distribution with the correct PSD (see text for details).

θ	Jensen-Shannon divergence ($\times 10^{-3}$)		
	$\Lambda_{\mathcal{N}}$	$\Lambda_{\mathcal{W}}$	$\Lambda_{\mathcal{H}}$
$\log_{10} \mathcal{A}$	11	0.4	0.3
f_0	12	0.7	0.4
$\log_{10}(\dot{f}_0)$	1	0.4	0.9
$\cos \iota$	11	0.4	0.4
$\sin \beta_{\mathrm{E}}$	11	0.4	0.5
λ_{E}	7	1.1	1.7
ψ	12	0.4	0.6
ϕ_0	5	0.9	2.3

correcting the PSD, through the free n_{level} parameter of $\Lambda_{\mathcal{W}}$, we obtain results that agree with the baseline case. The n_{level} parameter is recovered as 1.5 ± 0.12 (90% CI), which includes the correct value of $n_{\mathrm{level}}^{(\mathrm{correct})} \sim 1.57$. Very similar 2D posterior distributions are also obtained with the Hyperbolic likelihood $\Lambda_{\mathcal{H}}$, but without assuming any parametrized model for the PSD. Instead, the correction of the wrong PSD levels is achieved through the free parameters α, δ of $\Lambda_{\mathcal{H}}$. In Fig. 4 we also plot the $\log_{10} \delta - \log_{10} \alpha = \log_{10} n_{\mathrm{level}}^{(\mathrm{correct})}$ line,[8] which aligns with the 2D posterior distributions of α, δ as $\alpha \to \infty$ and $\delta \to \infty$, confirming that using $\Lambda_{\mathcal{H}}$ even with a wrong PSD in the inner product, one obtains posteriors equivalent to the baseline with the correct PSD.

In Table II we present the Jensen Shannon divergence of the marginal posterior probability density functions of all parameters, between the three different likelihood choices with wrong PSD levels and the baseline model of a Gaussian distribution with the correct PSD. It is quite evident, that the Hyperbolic distribution agrees really well with the baseline case and corrects the wrong PSD assumption with comparable accuracy as the Whittle likelihood.

This result demonstrates the robustness of the Hyperbolic likelihood formulation in situations where the noise spectral shape is not entirely known, or when has features that are not modeled properly. As already discussed, such is the case of the future LISA data, where the confusion noise of the ensemble signal of all the UCBs in our Galaxy will generate a confusion stochastic foreground. Thus, a possible application would be to adopt a heavy-tailed likelihood, such as the one presented here, in order to perform parameter estimation and search under a robust statistical framework.

---

[8]This equation is expected to hold, because the data are whitened with the PSD of the noise, see Eq. (24).

As a final note, we should mention that we retrieved very similar results when inverting the noise assumptions for this exercise, i.e., the true noise being lower than the one assumed in the likelihood functions. In this case, the posterior spread was overestimated with the $\Lambda_{\mathcal{N}}$ likelihood and correctly estimated with the other two choices.

3. Gaussian bursts

For our third application, we focus on a somewhat more realistic scenario. In particular, we inject the data generated in Sec. III B 2 with bursts of Gaussian noise, placed randomly in the time series of the A and E TDI channels. The noisy bursts are generated by the same model of Eq. (29) and with parameters drawn uniformly from $\log_{10} A \sim \mathcal{U}[-32, -28]$, and $\log_{10} P \sim \mathcal{U}[-23.6, -19.6]$ [in the units of Eq. (30)]. Their duration d_{burst} is also sampled uniformly, as $d_{\mathrm{burst}} \sim \mathcal{U}[0.01, 10]$ days. The resulting time series thus consists of the waveform model injected in Gaussian noise with the addition of the Gaussian bursts and is shown in Fig. 5. Naturally, the presence of such noise nonstationarities is expected to distort the spectral shape of the overall noise PSD. However, in our application, we focus on a single UCB, which is very well localized in frequency. Thus the effect of the bursts is similar to the mismodeling of the PSD in Sec. III B 2 above. This is illustrated in Fig. 3(a), where the individual components of the signal, the simulated Gaussian instrumental noise, and the Gaussian noise bursts are shown in the frequency domain.

Hence, we can adopt the same analysis methodology, which uses the same list of likelihood formulations that we used in previous sections. We first sample the posterior of the parameters given the data with a Gaussian likelihood, which assumes the artifacts-free noise model ($\Lambda_{\mathcal{N}}$). This assumption about the data is expected to yield a biased estimate of the waveform parameters posterior widths. We then repeat the analysis using the Whittle likelihood $\Lambda_{\mathcal{W}}$ and fitting for the noise level n_{level}. Finally, we use the Hyperbolic distribution ($\Lambda_{\mathcal{H}}$), where we fit for the α and δ parameters of Eq. (23). Notice that in this application, we do not have a baseline run, because we do not have access to the "true" noise PSD due to the randomly injected Gaussian bursts.

The results can be summarized in Fig. 6, where we plot the 2D posterior slices of the sampled parameters for each case. It is fairly obvious that the Hyperbolic and the noise-fitting Gaussian likelihoods present almost identical performance, with the corresponding marginal posterior distributions agreeing very well with each other. This is verified when we compute the Jensen-Shannon divergence, which yields figures of order $\sim 10^{-4}$ for all parameters. In addition, as expected, the n_{level} noise parameter estimated with the Whittle likelihood (the maximum posterior value for n_{level} is represented by the black dashed line in Fig. 6), agrees with the δ/α ratio recovered by the analysis using
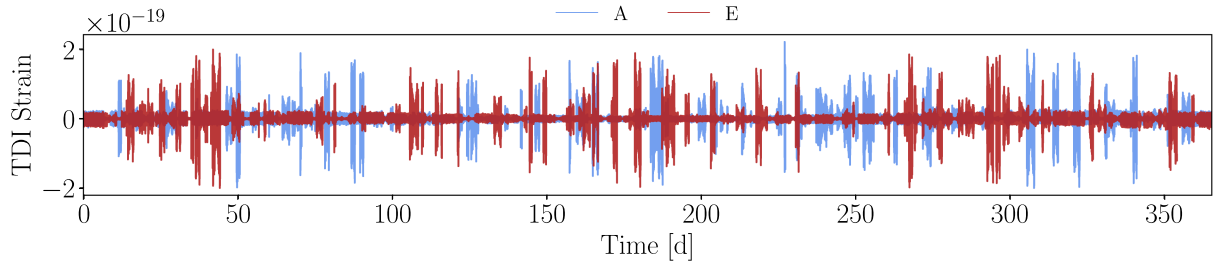
FIG. 5. The time series of the *A* (light blue) and *E* (dark red) TDI channels include the injection binary signal, the nominal instrumental noise, and the noise bursts. The noise bursts are randomly placed in the synthetic time series, while their duration and spectral shapes are also generated by sampling from predefined distributions (see text for details).
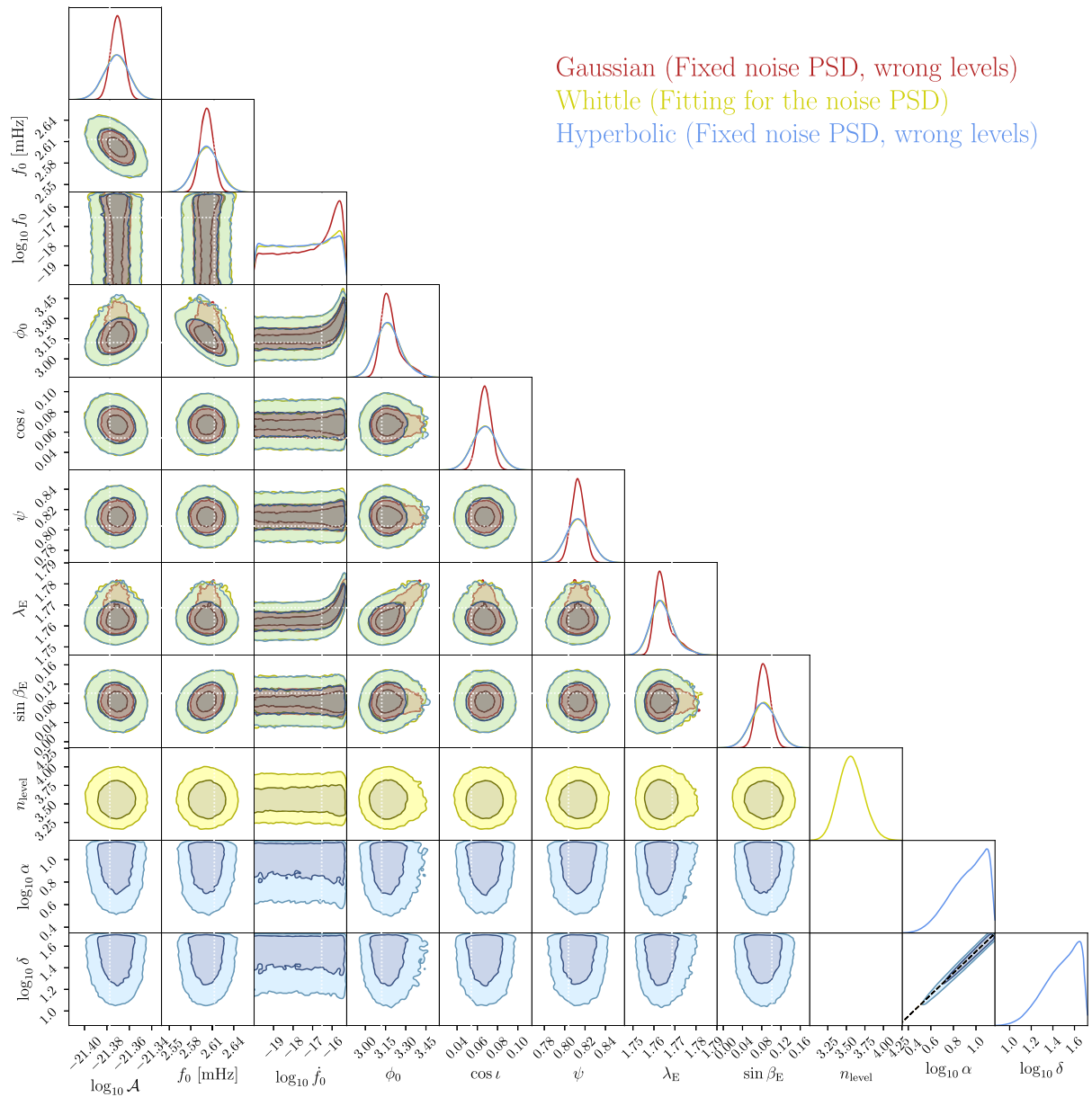

FIG. 6. Same as Fig. 4, but for the case of noise nonstationarities of Sec. III B 3. The dashed black here represents the n_{level} noise parameter as estimated with the Whittle likelihood analysis (see text for details).

the hyperbolic likelihood. This result confirms, once again, the versatility and robustness of the Hyperbolic likelihood $\Lambda_{\mathcal{H}}$ in parameter estimation situations, where the instrumental noise properties are not completely known.

IV. CONCLUSIONS AND DISCUSSION

We have introduced a heavy-tailed likelihood framework for robust inference and demonstrated its applications in Gravitational Wave data analysis. In particular, we have adopted the Generalized Hyperbolic ($\mathcal{GH}$) distribution, which has a number of widely used distributions of the exponential family as limiting distributions. This built-in flexibility allows us to model the residual noise with more accuracy, which is very useful when dealing with real data. Such cases are commonly encountered in GW Astronomy, where the GW signals need to be searched for in the noise of a given detector or of a network of detectors. This task can be challenging due to the limited knowledge of the statistical properties of the detector noise, which might depart from the ideal Gaussian properties. This usually happens due to noise transients and bursts, data gaps, spectral lines, or even the presence of foreground stochastic GW signals.

Based on a subclass of the ($\mathcal{GH}$) distribution (the Hyperbolic distribution, $\mathcal{H}$), we derived the Hyperbolic likelihood function $\Lambda_{\mathcal{H}}$ of Eq. (21), which depends on parameters α, δ and $\boldsymbol{\beta}$ and demonstrated its performance and flexibility in simple test cases. We simulated random variables from distributions that are known to be limiting cases of $\mathcal{GH}$. Those are the Gaussian ($\mathcal{N}$), $\mathcal{NIG}$, and Student's t-distributions. In each case, the inferred parameters α, δ and $\boldsymbol{\beta}$ of $\Lambda_{\mathcal{H}}$ were demonstrated (as evidenced by the Jensen-Shannon divergence) to lead to a close match between the target and posterior distributions, even for cases where one of the $\mathcal{GH}$ parameters (λ) of the target distribution was different than the $\lambda = 1$ value assumed in the Hyperbolic distribution. This is a nice demonstration that the Hyperbolic distribution has sufficient degrees of freedom to adjust (with acceptable accuracy) to different distributions.

As as next step, we applied this methodology to more realistic applications, drawn from problems encountered in GW Astronomy. For the case of the current ground-based detectors, the spectral shape of the noise can be understood fairly well to allow inference with relative confidence, due to the accessibility of the instrument and the numerous auxiliary channels available. Even in that case, there is always the risk of transient events, such as noisy glitches, which can happen randomly in time. The situation is quite different for future space-based observatories, such as LISA, where the spectral shape of the noise will not be completely known. This is due to the limited number of available auxiliary channels, as well as due to the "confusion noise" generated by the ensemble of the vast number of signals detectable with LISA.

For demonstration purposes, we limited our analyses to a single signal from a Double White Dwarf binary, injected in colored Gaussian noise representative of the LISA detector with relatively high SNR ($\rho = 122.7$, see Table I for its waveform parameters) and investigated three different applications. In the first investigation, we performed parameter estimation using the Hyperbolic likelihood $\Lambda_{\mathcal{H}}$ we propose in Eqs. (23) and (26), where the level of the noise was assumed completely unknown (no PSD was assumed and no whitening was performed in the frequency domain). From the inferred parameters $\{\alpha, \delta\}$ of the symmetric Hyperbolic distribution, we were able to characterize the noise. In particular, as theory predicts, we found that the distribution of the residuals asymptotically (as $\alpha, \delta \rightarrow \infty$) tends to the assumed Gaussian distribution, with $\log \delta - \log \alpha = \log S_n^{\mathrm{AE}}(f_0)$, where $S_n^{\mathrm{AE}}(f_0)$ is the spectral noise density at the emission frequency f_0 of the binary.

Then, we investigated a case of mismodeling the PSD levels of the noise. Essentially, we assume a wrong noise level for the parameter estimation process, which is then plugged into the Gaussian likelihood of Eq. (5). For this investigation, we assumed that the noise PSD was lower than the true value that was used to simulate the data. As before, we also performed the parameter estimation analysis of a single Galactic Binary using the Hyperbolic likelihood $\Lambda_{\mathcal{H}}$, with the expectation that the $\{\alpha, \delta\}$ coefficients would "tune" the likelihood to mitigate for this mismodeling of the noise PSD. This was verified with the results presented in Fig. 3 and Table II. The estimated waveform posteriors for the parameters, when using $\Lambda_{\mathcal{H}}$, are almost identical to those obtained with the baseline Gaussian likelihood using the correct noise levels. In addition, the ratio δ/α obtained from posteriors of the parameters of the hyperbolic likelihood converged to the ratio $n_{\mathrm{level}}^{(\mathrm{true})}$ between the true noise PSD level and the mismodeled case. The true PSD was thus fully characterized using the $\Lambda_{\mathcal{H}}$ likelihood.

Finally, we tested this framework in a somewhat more realistic scenario, where the instrumental noise was "polluted" with Gaussian bursts, randomly placed in time for both TDI channels considered in the analysis. Using the same approach as before, we obtained results that verify the functionality of the heavy-tailed framework introduced in this work. In particular, we demonstrated that with the hyperbolic likelihood, we were able to recover the correct posterior widths for the waveform parameters, which were larger than predicted by the simple Gaussian likelihood, due to the additional uncertainty induced by the noisy bursts (see Sec. III B 3 for details).

We expect that the new framework we introduced in this work will be very useful for data from future observatories, where the noise model will not be completely accessible. The Hyperbolic likelihood is quite versatile, as the shape of the Hyperbolic distribution can adjust to fit the residual

data. In that sense, an advantage to the Whittle approximation of the likelihood, is that the Hyperbolic likelihood can converge to a different distribution than the Gaussian. Thus, there are many applications where this framework can be utilized, a primary example being the modeling of noise artifacts, such as bursts, data gaps, glitches, or other types of nonstationarities. At the same time, even in the absence of noise nonstationarities, the Hyperbolic likelihood can be employed in order to probe any mismodeling of our noise models. Finally, another application would be to construct Gaussianity tests on different segments of data, based on the values of the parameters of the Hyperbolic likelihood and tools such as the shape triangle.

In our computations, we used Parallel Tempering MCMC methods [48], and graphics processing unit accelerated waveforms [99] and we make our codes available as open source software [101].

ACKNOWLEDGMENTS

APPENDIX: THE SHAPE TRIANGLE

The shape triangle is a very useful graphical tool for visualizing the degree to which a distribution belonging to the exponential family is heavy-tailed or skewed. One can begin by using the Generalized Hyperbolic distribution to characterize a new distribution. Then, based on the recovered $\{\alpha, \beta, \delta\}$ coefficients, one can qualitatively categorize the distribution of interest by placing it into a shape triangle [68,69]. To do that, we use a different parametrization from the usual $\{\alpha, \beta, \delta\}$ set. Following [68,69], we can compute

$$\zeta = \delta\sqrt{\alpha^2 - \beta^2}, \qquad \varrho = \beta/\alpha,$$
$$\xi = (1 + \zeta)^{-1/2}, \qquad \chi = \xi\varrho, \qquad (A1)$$

and then use the scale and location-invariant parameters χ and ξ, which form a triangle in the $\chi - \xi$ plane, since $0 \le |\chi| < \xi < 1$. The position inside the shape triangle gives us visual information about the distribution of heavy-tailedness and skewness. This visual tool can be helpful in practical applications, such as monitoring data quality in real time or classifying different data segments, depending on their statistical properties [102].

Here, we investigate the same test distributions as in Sec. II E, but for the two cases, for which the parameter λ is negative, we adopt the $\mathcal{GH}$ distribution for $\lambda < 0$ as in [68]

$$
\begin{aligned}
\Lambda_{\mathcal{GH}_d}(\alpha, \delta) = {}& n\left[-\lambda\log(\delta) + \frac{d}{2}\log(2\pi) - \log\left(K_\lambda(\delta\alpha)\right)\right] \\
& - \frac{1}{2}\left(\lambda - \frac{d}{2}\right)\sum_{j=1}^{n}(\delta^2 + x_j^2) \\
& + \sum_{j=1}^{n} K_{\lambda-\frac{d}{2}}\left(\alpha\sqrt{\delta^2 + x_j^2}\right). \qquad (A2)
\end{aligned}
$$

The reason that we assume two different likelihoods, one of the positive values of λ [$\Lambda_{\mathcal{H}}$ of Eq. (21)] and another one for the negative values [$\Lambda_{\mathcal{GH}_d}$ of Eq. (A2)], arises from the definition of the domain in Eq. (9). Thus, the difference here compared to Sec. II E is that the posteriors for the $\{\alpha, \beta, \delta\}$ are estimated using the $\Lambda_{\mathcal{GH}_d}$ Eq. (A2) when the test-case distribution is either the Student's t-distribution or $\mathcal{NIG}$.

After we obtain the posteriors, we calculate the $\{\chi, \xi\}$ parameters from Eqs. (A1) and place them onto the shape triangle, which is shown in Fig. 7. The shape triangle, gives us a graphical representation (on the $\chi - \xi$ plane) of the characteristics of the resulting distribution. We have marked the injected values with crosses, while the recovered parameters are marked with colored dot markers ($\mathcal{N}$ with yellow, $\mathcal{H}$ with purple, $\mathcal{NIG}$ with red and, and finally the Student's t-distribution case with light blue). From the results displayed in Fig. 7, it is evident that we were able to recover the underlying distributions as special or limiting cases of the $\mathcal{GH}$ distribution.
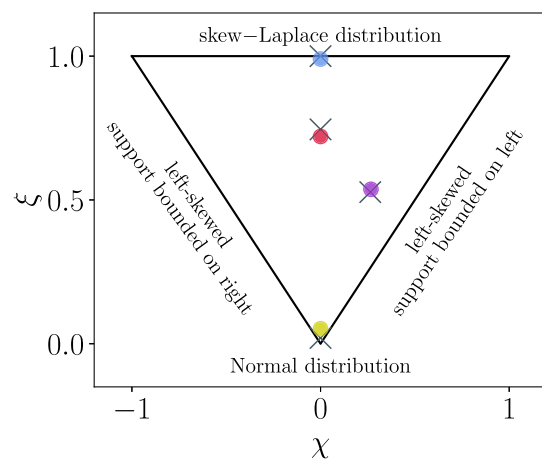

FIG. 7. Triangle shape for the injected values of χ, ξ (crosses) and the recovered values (filled circles) for the data generated assuming Gaussian $\mathcal{N}$ (yellow), Hyperbolic $\mathcal{H}$ (purple), Normal Inverse Gaussian $\mathcal{NIG}$ (red) and Student's t-distributions (light blue).

[1] B. P. Abbott *et al.*, Phys. Rev. Lett. **116**, 061102 (2016).
[2] J. Aasi *et al.*, Classical Quantum Gravity **32**, 074001 (2015).
[3] F. Acernese *et al.*, Classical Quantum Gravity **32**, 024001 (2014).
[4] B. P Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. X **6**, 041015 (2016).
[5] B. P Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. X **9**, 031040 (2019).
[6] R. Abbott *et al.* (LIGO Scientific Collaboration and Virgo Collaboration), Phys. Rev. X **11**, 021053 (2021).
[7] R. Abbott *et al.*, GWTC-2.1: Deep extended catalog of compact binary coalescences observed by LIGO and Virgo during the first half of the third observing run (2021).
[8] R. Abbott *et al.*, GWTC-3: Compact binary coalescences observed by LIGO and Virgo during the second part of the third observing run (2021).
[9] R. Abbott *et al.*, The population of merging compact binaries inferred using gravitational waves through GWTC-3 (2021).
[10] A. H. Nitz, S. Kumar, Y.-F. Wang, S. Kastha, S. Wu, M. Schäfer, R. Dhurkunde, and C. D. Capano, Astrophys. J. **946**, 59 (2023).
[11] Y. Aso, Y. Michimura, K. Somiya, M. Ando, O. Miyakawa, T. Sekiguchi, D. Tatsumi, and H. Yamamoto, Phys. Rev. D **88**, 043007 (2013).
[12] C. S. Unnikrishnan, Int. J. Mod. Phys. D **22**, 1341010 (2013).
[13] M. Saleem, J. Rana, V. Gayathri, A. Vijaykumar, S. Goyal, S. Sachdev, J. Suresh, S. Sudhagar, A. Mukherjee, G. Gaur, B. Sathyaprakash, A. Pai, R. X. Adhikari, P. Ajith, and S. Bose, Classical Quantum Gravity **39**, 025004 (2021).
[14] M. Punturo *et al.*, Classical Quantum Gravity **27**, 194002 (2010).
[15] V. Srivastava, D. Davis, K. Kuns, P. Landry, S. Ballmer, M. Evans, E. D. Hall, J. Read, and B. S. Sathyaprakash, Astrophys. J. **931**, 22 (2022).
[16] K. Ackley *et al.*, Publ. Astron. Soc. Pac. **37**, e047 (2020).
[17] P. Amaro-Seoane *et al.*, Laser interferometer space antenna (2017).
[18] Y. Gong, J. Luo, and B. Wang, Nat. Astron. **5**, 881 (2021).
[19] S. Wu and A. H. Nitz, Phys. Rev. D **107**, 063022 (2023).
[20] T. B. Littenberg and N. J. Cornish, Phys. Rev. D **107**, 063004 (2023).
[21] N. Christensen and R. Meyer, Rev. Mod. Phys. **94**, 025001 (2022).
[22] T. B. Littenberg and N. J. Cornish, Phys. Rev. D **91**, 084034 (2015).
[23] Q. Baghi, N. Karnesis, J.-B. Bayle, M. Besançon, and H. Inchauspé, J. Cosmol. Astropart. Phys. 04 (2023) 066.
[24] L. Speri, N. Karnesis, A. I. Renzini, and J. R. Gair, Nat. Astron. **6**, 1356 (2022).
[25] M. C. Edwards, P. Maturana-Russel, R. Meyer, J. Gair, N. Korsakova, and N. Christensen, Phys. Rev. D **102**, 084062 (2020).
[26] M. Armano *et al.* (LISA Pathfinder Collaboration), Phys. Rev. D **106**, 062001 (2022).
[27] Q. Baghi, N. Korsakova, J. Slutsky, E. Castelli, N. Karnesis, and J.-B. Bayle, Phys. Rev. D **105**, 042002 (2022).

[28] K. Dey, N. Karnesis, A. Toubiana, E. Barausse, N. Korsakova, Q. Baghi, and S. Basak, Phys. Rev. D **104**, 044035 (2021).
[29] Q. Baghi, I. Thorpe, J. Slutsky, J. Baker, T. Dal Canton, N. Korsakova, and N. Karnesis, Phys. Rev. D **100**, 022003 (2019).
[30] S. Drasco and E. E. Flanagan, Phys. Rev. D **67**, 082003 (2003).
[31] N. Karnesis, S. Babak, M. Pieroni, N. Cornish, and T. Littenberg, Phys. Rev. D **104**, 043019 (2021).
[32] B. P. Abbott *et al.* (The LIGO Scientific Collaboration and the Virgo Collaboration), Classical Quantum Gravity **37**, 055002 (2020).
[33] M. Armano *et al.*, Phys. Rev. Lett. **120**, 061101 (2018).
[34] M. Armano *et al.*, Phys. Rev. Lett. **116**, 231101 (2016).
[35] G. Anderson *et al.*, Phys. Rev. D **98**, 102005 (2018).
[36] M. Armano *et al.*, Phys. Rev. D **97**, 122002 (2018).
[37] M. Georgousi, N. Karnesis, V. Korol, M. Pieroni, and N. Stergioulas, Mon. Not. R. Astron. Soc. **519**, 2552 (2022).
[38] V. Korol, N. Hallakoun, S. Toonen, and N. Karnesis, Mon. Not. R. Astron. Soc. **511**, 5936 (2022).
[39] S. Babak, C. Caprini, D. G. Figueroa, N. Karnesis, P. Marcoccia, G. Nardini, M. Pieroni, A. Ricciardone, A. Sesana, and J. Torrado, J. Cosmol. Astropart. Phys. 08 (2023) 034.
[40] F. Pozzoli, S. Babak, A. Sesana, M. Bonetti, and N. Karnesis, arXiv:2302.07043.
[41] M.-S. Hartig, S. Schuster, and G. Wanner, J. Opt. **24**, 065601 (2022).
[42] M. Armano *et al.*, arXiv:2308.02398.
[43] S. Paczkowski, R. Giusteri, M. Hewitson, N. Karnesis, E. D. Fitzsimons, G. Wanner, and G. Heinzel, Phys. Rev. D **106**, 042005 (2022).
[44] D. George, J. Sanjuan, P. Fulda, and G. Mueller, Phys. Rev. D **107**, 022005 (2023).
[45] N. Houba, S. Delchambre, T. Ziegler, G. Hechenblaikner, and W. Fichter, Phys. Rev. D **106**, 022004 (2022).
[46] N. J. Cornish and T. B. Littenberg, Classical Quantum Gravity **32**, 135012 (2015).
[47] P. J. Green, Biometrika **82**, 711 (1995).
[48] N. Karnesis, M. L. Katz, N. Korsakova, J. R. Gair, and N. Stergioulas, arXiv:2303.02164.
[49] R. Umstätter, N. Christensen, M. Hendry, R. Meyer, V. Simha, J. Veitch, S. Vigeland, and G. Woan, Phys. Rev. D **72**, 022001 (2005).
[50] T. B. Littenberg, N. J. Cornish, K. Lackeos, and T. Robson, Phys. Rev. D **101**, 123021 (2020).
[51] T. B. Littenberg and N. J. Cornish, Phys. Rev. D **107**, 063004 (2023).
[52] Q. Baghi, N. Karnesis, J.-B. Bayle, M. Besançon, and H. Inchauspé, J. Cosmol. Astropart. Phys. 04 (2023) 066.
[53] M. C. Edwards, P. Maturana-Russel, R. Meyer, J. Gair, N. Korsakova, and N. Christensen, Phys. Rev. D **102**, 084062 (2020).
[54] M. C. Edwards, R. Meyer, and N. Christensen, Phys. Rev. D **92**, 064011 (2015).
[55] C. Röver, Phys. Rev. D **84**, 122004 (2011).
[56] C. Röver, R. Meyer, and N. Christensen, Classical Quantum Gravity **28**, 015010 (2011).

[57] S. Banagiri, M. W. Coughlin, J. Clark, P. D. Lasky, M. A. Bizouard, C. Talbot, E. Thrane, and V. Mandic, Mon. Not. R. Astron. Soc. **492**, 4945 (2020).

[58] L. Martellini and T. Regimbau, Phys. Rev. D **89**, 124009 (2014).

[59] S. Hamimeche and A. Lewis, Phys. Rev. D **77**, 103013 (2008).

[60] L. Verde, H. V. Peiris, D. N. Spergel, M. R. Nolta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S. S. Meyer, L. Page, G. S. Tucker, E. Wollack, and E. L. Wright, Astrophys. J. Suppl. Ser. **148**, 195 (2003).

[61] R. Flauger, N. Karnesis, G. Nardini, M. Pieroni, A. Ricciardone, and J. Torrado, J. Cosmol. Astropart. Phys. 01 (2021) 059.

[62] J. D. E. Creighton, Phys. Rev. D **60**, 021101 (1999).

[63] B. Allen, J. D. E. Creighton, E. E. Flanagan, and J. D. Romano, Phys. Rev. D **65**, 122002 (2002).

[64] M. L. Bianchi, A. Hitaj, and G. L. Tassinari, arXiv:2005. 06390.

[65] S. Borak, A. Misiorek, and R. Weron, Models for heavy-tailed asset returns, in *Statistical Tools for Finance and Insurance*, edited by P. Cizek, W. K. Härdle, and R. Weron (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011), pp. 21–55.

[66] K. Prause, Albert-Ludwigs-Universität Freiburg fDM Report No. 48, 1997.

[67] E. Eberlein and K. Prause, in *Mathematical Finance–Bachelier Congress 2000* (Springer, New York, 2002), pp. 245–267.

[68] K. Prause (1999).

[69] U. Kächler, K. Neumann, M. Sørensen, and A. Streller, Math. Comput. Modell. **29**, 1 (1999).

[70] P. Auclair *et al.* (LISA Cosmology Working Group), Living Rev. Relativity **26**, 5 (2023).

[71] The LISA Data Challenges, https://lisa-ldc.lal.in2p3.fr/, accessed: 2023-01-01.

[72] P. Whittle, J. R. Stat. Soc. Ser. B **19**, 38 (1957).

[73] N. Choudhuri, S. Ghosal, and A. Roy, Biometrika **91**, 211 (2004).

[74] Y. Tang, C. Kirch, J. E. Lee, and R. Meyer, arXiv:2103. 01357.

[75] C. Kirch, M. C. Edwards, A. Meier, and R. Meyer, arXiv: 1701.04846.

[76] S. Vitale *et al.*, Phys. Rev. D **90**, 042003 (2014).

[77] A. Contreras-Cristán, E. Gutiérrez-Peña, and S. G. Walker, Commun. Stat. **35**, 857 (2006).

[78] O. E. Barndorff-Nielsen and P. Blaesild (1981).

[79] O. Barndorff-Nielsen, Proc. R. Soc. A **353**, 401 (1977).

[80] E. Eberlein and E. A. v. Hammerstein, in *Seminar on Stochastic Analysis, Random Fields and Applications IV*, edited by R. C. Dalang, M. Dozzi, and F. Russo (Birkhäuser Basel, Basel, 2004), pp. 221–264.

[81] P. Virtanen *et al.*, Nat. Methods **17**, 261 (2020).

[82] M. Menéndez, J. Pardo, L. Pardo, and M. Pardo, J. Franklin Inst. **334**, 307 (1997).

[83] F. Nielsen, Entropy **22**, 221 (2020).

[84] J. I. Thorpe *et al.*, Astrophys. J. **883**, 53 (2019).

[85] M. Armano *et al.*, Phys. Rev. D **99**, 082001 (2019).

[86] M. Armano *et al.*, Phys. Rev. Lett. **118**, 171101 (2017).

[87] D. V Martynov *et al.*, Phys. Rev. D **93**, 112004 (2016).

[88] R. Abbott *et al.*, Classical Quantum Gravity **37**, 055002 (2020).

[89] F. Acernese *et al.*, Classical Quantum Gravity **40**, 185006 (2023).

[90] S. Babak, A. Petiteau, and M. Hewitson, arXiv:2108. 01167.

[91] T. A. Prince, M. Tinto, S. L. Larson, and J. W. Armstrong, Phys. Rev. D **66**, 122002 (2002).

[92] M. Tinto and S. V. Dhurandhar, Living Rev. Relativity **8**, 4 (2005).

[93] LISA Science Study Team, LISA science requirements document, Technical Report No. ESA-L3-EST-SCI-RS-001, ESA, 2018.

[94] A. P. Stephan, S. Naoz, A. M. Ghez, M. R. Morris, A. Ciurlo, T. Do, K. Breivik, S. Coughlin, and C. L. Rodriguez, Astrophys. J. **878**, 58 (2019).

[95] V. Korol, E. M. Rossi, P. J. Groot, G. Nelemans, S. Toonen, and A. G. A. Brown, Mon. Not. R. Astron. Soc. **470**, 1894 (2017).

[96] N. J. Cornish and T. B. Littenberg, Phys. Rev. D **76**, 083006 (2007).

[97] T. Robson, N. J. Cornish, N. Tamanini, and S. Toonen, Phys. Rev. D **98**, 064012 (2018).

[98] M. L. Katz, C. Danielski, N. Karnesis, V. Korol, N. Tamanini, N. J. Cornish, and T. B. Littenberg, Mon. Not. R. Astron. Soc. **517**, 697 (2022).

[99] GBGPU, https://zenodo.org/record/6500434, accessed: 2023-01-01.

[100] The LISA Data Challenges Software, https://gitlab.in2p3 .fr/LISA/LDC, accessed: 2023-01-01.

[101] A. Sasli, N. Karnesis, and N. Stergioulas, hyperbolic_ likelihood_filter, version 0.1, 2023, https://github.com/ asasli/hyperbolic_likelihood_filter.

[102] D. Hartmann and C. Christiansen, Sedimentology **39**, 697 (1992).