Extending the Fisher Information Matrix in Gravitational-wave Data Analysis

Ziming Wang\textsuperscript{1}, Chang Liu\textsuperscript{1,2}, Junjie Zhao\textsuperscript{3}, and Lijing Shao\textsuperscript{2,4}

\textsuperscript{1}Department of Astronomy, School of Physics, Peking University, Beijing 100871, People’s Republic of China
\textsuperscript{2}Kavli Institute for Astronomy and Astrophysics, Peking University, Beijing 100871, People’s Republic of China; lshao@pku.edu.cn
\textsuperscript{3}Department of Astronomy, Beijing Normal University, Beijing 100875, People’s Republic of China
\textsuperscript{4}National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, People’s Republic of China

Received 2022 March 9; revised 2022 April 10; accepted 2022 April 27; published 2022 June 20

Abstract

The Fisher information matrix (FM) plays an important role in forecasts and inferences in many areas of physics. While giving fast parameter estimation with Gaussian likelihood approximation in the parameter space, the FM can only give the ellipsoidal posterior contours of the parameters and it loses the higher-order information beyond Gaussianity. We extend the FM in gravitational-wave (GW) data analysis by using the Derivative Approximation for Likelihoods (DALI), a method to expand the likelihood, while keeping it positive definite and normalizable at every order, for more accurate forecasts and inferences. When applied to two real GW events, GW150914 and GW170817, DALI can reduce the difference between the FM approximation and the real posterior by 5 times in the best case. The calculation times of DALI and the FM are at the same order of magnitude, while obtaining the real full posterior will take several orders of magnitude longer. Besides more accurate approximations, higher-order correction from DALI provides a fast assessment of the FM analysis and gives suggestions for complex sampling techniques that are computationally intensive. We recommend using the DALI method as an extension to the FM method in GW data analysis to pursue better accuracy while still keeping the speed.

Unified Astronomy Thesaurus concepts: Gravitational waves (678); Astronomy data analysis (1858); Bayesian statistics (1900); Astrostatistics (1882)

1. Introduction

The Fisher information matrix (FM) has been widely used to forecast and estimate parameters in many fields of physics. In gravitational-wave (GW) data analysis, the FM is a useful tool to characterize the parameter estimation performance of GW measurements (Finn 1992; Vallisneri 2008). Using the FM, one can estimate the measurement abilities of different GW detectors, with a view to providing guidance on the construction of GW detectors and obtaining a glimpse of the scientific outputs (Cutler & Flanagan 1994). As the next-generation GW detectors—including ground-based and space-borne ones—come into use, discussion about what and how accurately they are able to reach is critical and inevitable. Based on the FM, many works have investigated in detail the detection abilities of these detectors, and provided forecasts on their scientific significance (see, e.g., Cutler 1998; Berti et al. 2005; Isoyama et al. 2018; Liu et al. 2020; Ruan et al. 2020; Zhao et al. 2021; Liu & Shao 2022; Shuman & Cornish 2022).

The FM is not the only means of aiding GW parameter estimation, but it is the fastest. In a widely used approach, the parameters of interest are obtained via the Bayes’ theorem, and their posterior distributions can be expressed as the product of a subjectively selected prior and a model-dependent likelihood. One can evaluate the posterior directly, using sampling techniques such as Monte Carlo Markov Chain approaches and Nested Sampling (Skilling 2004, 2006). These methods can give the real full posterior, at the desired precision, in principle, but can also take considerable computing time and resources. This problem is particularly sharp in GW analysis, where one must deal with complicated waveform templates. For the dozens of real GW events detected so far (Abbott et al. 2019, 2021a, 2021b), people have been willing to investigate their parameters as accurately as possible, in spite of the computational costs. However, in the exploration of the characteristics of new detectors, in the light of rapidly increasing event numbers, numerous simulations and forecasts are required, and they cannot all rely on these delicate sampling techniques, given the limited computing resources. To some extent, forecasting prefers speed to accuracy. Therefore, the fast FM approximation comes into consideration. It only needs the first-order derivatives of the parameters of the waveform at one particular point, the truth point, which is known in forecasting studies. In mathematics, the FM gives the Gaussian part of the likelihood, which can be expressed easily and analytically, and the covariance matrix of the parameters is just the inverse of the FM (Cutler & Flanagan 1994). Given these properties, the FM has been used routinely in numerous studies of GWs. Although it benefits from its advantages, the FM also pays for its approximation. With the assumption of the Gaussian posterior (or the Gaussian likelihood, in the parameter space), the FM will by definition lead to ellipsoidal confidence level contours, whose principal axes represent the parameter degeneracy, with the areas of the contours reflecting the detection ability. However, the Gaussian likelihood in the parameter space is exact only when the noises are Gaussian and the model is linear in its parameters. In GW data analysis, we usually assume that the noise is stationary and Gaussian, while the GW templates are not linear in their parameters. In this case, the FM comes from the linear signal approximation of the waveform or, equivalently, the high signal-to-noise ratio (S/N) approximation (Finn 1992; Vallisneri 2008), which is not always satisfied in reality. Besides, one may encounter a (quasi-)singular FM in some cases, which hardly provides useful information. In any GW event, only taking out the Gaussian part is bound to deviate from the real likelihood, so it is worth...
investigating how to check whether the FM is applicable in the forecasting process (Vallisneri 2008; Shuman & Cornish 2022).

There are several ways of dealing with this problem without directly calculating the costly real likelihood. Selecting the appropriate parameters to obtain an approximately multivariate Gaussian likelihood in the transformed parameter space is one good choice (Joachimi & Taylor 2011). One can also consider the higher orders of the FM and extend them. When adding the higher-order terms, an approximation closer to the original likelihood can be obtained, and the accuracy of the FM can also be tested by evaluating the higher-order corrections. As mentioned above, the FM works well in the high-S/N limit (Finn 1992), so one can expand the likelihood in the small parameter “1/(S/N)” (see Appendix A.5 in Cutler & Flanagan 1994; Vallisneri 2008). An efficient semianalytical technique has been developed to acquire the exact sampling distribution of the maximum-likelihood (ML) estimator across noise realizations for any S/N (Vallisneri 2011). Here, we will focus on another path to higher orders, considering directly the Taylor expansion of the likelihood, which is also the most straightforward path.

When expanding the likelihood at the ML point—or the truth point in forecasting—the first nontrivial order is exactly the FM approximation. Therefore, the easiest way to expand the FM is to add the higher-order terms in the Taylor series. However, if truncated at a certain order, in general this expansion is not guaranteed to be a true mathematically valid probability distribution, i.e., one that is positive definite and normalizable. Sellentin et al. (2014) found that a simple rearrangement of the terms in the Taylor series can guarantee that the expansion remains a true probability distribution at each order of the derivatives of the model. Based on this, they proposed a new method, named Derivative Approximation for Likelihoods (DALI), to approximate the likelihood. At the leading order, the DALI series gives the FM approximation, while at any higher order the DALI approximation is positive definite and normalizable, keeping the properties of a probability density for the approximated likelihood. With DALI, one could obtain the well-defined non-Gaussian parts of the likelihood beyond the FM. As further described in Sellentin (2015), the DALI algorithm is independent of its physical application, meaning that the method can be applied in any field of physics, such as cosmology (Sellentin et al. 2014) and gravitational lensing (Sellentin and Schäfer 2016). However, the DALI methods in early works are approximations of the likelihood in finite-dimensional data spaces with finite random variables, while in GW analysis in theory we must deal with the infinite-dimensional data space, namely the GW strain $g(t)$. Therefore, the original DALI method, reviewed in the Appendix, needs to be changed slightly to fit the requirements of GW astrophysics.

In this work, we generalize the DALI method to fit the requirements of GW data analysis, and show how it works in reality. Similar to Sellentin et al. (2014), this method can show the degeneracy directions and regions of the parameter space beyond the FM. Meanwhile, it still only needs the derivatives at one point, meaning that it is still much faster than complex sampling techniques. To show that DALI is a useful and powerful method for forecasts and inferences, we apply it to two real GW events, GW150914 (Abbott et al. 2016) and GW170817 (Abbott et al. 2017). We find that compared to the FM, the posterior distribution obtained by the DALI approximation is about 1–5 times closer to the posterior from a sophisticated sampling method, measured by the Wasserstein distance (Vaserstein 1969; Peyré and Cuturi 2019). Meanwhile, the time costs of DALI and the FM are still at the same order of magnitude. We therefore establish DALI as a useful extension of the FM.

This paper is organized as follows. In Section 2, we introduce the original DALI algorithm in Sellentin et al. (2014), and derive the corresponding formulae in GW data analysis. In Section 3, we explain how to use the DALI approximation in forecasts and real data analysis. Section 4 gives two examples and compares the DALI approximation with the FM and sophisticated sampling methods. Section 5 presents a summary and discusses the application prospects of DALI.

# 2. DALI Algorithm in GW Data Analysis

As described in the introduction, the DALI algorithm is a powerful and fast method to obtain an approximated likelihood once the model has been given (Sellentin et al. 2014). The detailed expansion formulae of DALI for finite random variables can be found in the Appendix. To obtain this algorithm, we will start from the classical multidimensional Gaussian likelihood in the data space,

$$P(x|\Theta) = \exp\left\{ -\frac{1}{2} [x - \mu(\Theta)] M [x - \mu(\Theta)] \right\},$$

and its corresponding log-likelihood $L(\Theta) \equiv -\log P(x|\Theta)$ for a set of fixed data $x$. Here, $x \equiv (x_i) (i = 1, 2, \cdots, N_d)$ denotes a realization of $N_d$ random variables, $\Theta \equiv (\Theta_\alpha) (\alpha = 1, 2, \cdots, N_p)$ denotes $N_p$ parameters of interest, $\mu \equiv (\mu_\alpha(\Theta)) (i = 1, 2, \cdots, N_d)$ is the model that gives the prediction of $x$, $C$ is the parameter-independent covariance matrix in the data space with dimensions $N_d \times N_d$, and $M \equiv C^{-1}$ is the inverse of the covariance matrix. It is worth noting that the likelihood (1) is strictly Gaussian in the parameter space only when the model $\mu$ is linear in its parameters. With likelihood (1) and a prior $p(\Theta)$, one can obtain the posterior distribution of the parameters using the Bayes’ theorem:

$$P(\Theta|x) \propto p(\Theta) \cdot P(x|\Theta).$$

One can then evaluate the likelihood at any parameters analytically, or use sampling techniques to obtain the global distribution of the parameters numerically (Veitch et al. 2015).

However, when the model becomes complicated, evaluating this multidimensional likelihood can be a computationally costly procedure, and a good approximation of the likelihood will help to reduce the cost. Sellentin et al. (2014) found a systematic method, DALI, to obtain a well-defined approximation of the likelihood at any order. The principle of DALI is to rearrange the Taylor expansion of the likelihood in order of the derivatives of the model $\mu$. The rearrangement can be completed. When taking the ensemble average, denoted as $\langle \cdots \rangle$, the derivatives of $L$ at the ML point can be written as (Sellentin et al. 2014):

$$\langle L, \alpha \beta \gamma \delta \rangle = \mu_{\alpha} M_{\beta} \delta_{\gamma},$$

$$\langle L, \alpha \beta \gamma \delta \delta \delta \rangle = \mu_{\alpha} M_{\beta} \mu_{\delta} + \mu_{\gamma} \delta \delta \delta \mu_{\beta} + \mu_{\alpha} \beta \delta \delta \delta \mu_{\gamma},$$

$$\langle L, \alpha \beta \gamma \delta \gamma \delta \rangle = \mu_{\alpha} \beta \gamma \delta \delta \delta \mu + \mu_{\alpha} \gamma \delta \delta \delta \mu_{\beta} + \mu_{\gamma} \alpha \beta \delta \delta \delta \mu + \mu_{\alpha} \beta \delta \delta \delta \mu_{\gamma},$$

$$\langle L, \alpha \beta \gamma \delta \gamma \delta \gamma \delta \rangle = \mu_{\alpha} \beta \gamma \delta \gamma \delta \delta \delta \mu + \mu_{\alpha} \beta \gamma \delta \gamma \delta \delta \delta \mu + \mu_{\gamma} \beta \gamma \delta \delta \delta \delta \delta \mu + \mu_{\alpha} \beta \gamma \delta \gamma \delta \delta \delta \mu + \mu_{\alpha} \beta \gamma \delta \gamma \delta \delta \delta \mu + \mu_{\gamma} \beta \gamma \delta \delta \delta \delta \delta \mu + \mu_{\alpha} \beta \gamma \delta \gamma \delta \delta \delta \mu.$$
and so on. Then, rearranging these terms in order of the derivatives of the model \( \mu \), one can obtain the DALI expansion in Equation (A4).

When calculating the likelihood in the GW data analysis, we need to calculate the probability of the GW signal \( g(t) \), when the physical parameters \( \Theta \) and waveform template \( h(t; \Theta) \) are known. In principle, the strain contains infinite random variables, essentially in a random process. Assuming that the noise \( n(t) \) is stationary and Gaussian, with a zero mean, the likelihood is (Finn 1992)

\[
P(g(t)|h(t; \Theta)) \propto e^{-\frac{1}{2}E(n-h,g-h)},
\]
where \((u, v)\) is the symmetric inner product of two data streams \( u(t) \) and \( v(t) \):

\[
(u, v) \equiv 2 \int_{-\infty}^{\infty} \frac{\tilde{u}(f) \tilde{v}(f)}{S_n(f)} df
\]
\[
= 4\Re \int_0^{\infty} \frac{\tilde{u}(f) \tilde{v}(f)}{S_n(f)} df.
\]

Here, \( S_n(f) \) is the one-sided power spectral density of the noise, and \( \tilde{u}(f) \) and \( \tilde{v}(f) \) are the Fourier transforms of \( u(t) \) and \( v(t) \), defined by

\[
\tilde{u}(f) \equiv \mathcal{F}(u(t)) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} u(t) e^{-2\pi if} dt.
\]

Now we use the DALI method to deal with the likelihood in Equation (4), which requires that the derivatives of the log-likelihood \( \mathcal{L} = -\log P(g(t)|h(t; \Theta)) \) have a similar form to Equation (3). Indeed, they do have the required form:

\[
\mathcal{L}_{\alpha} = (h - g, h_\alpha),
\]
\[
\mathcal{L}_{\alpha\beta} = ((h_{\beta}, h_\alpha) + (h - g, h_{\alpha\beta}))
\]
\[
= (h_{\beta}, h_\alpha) + (n, h_{\alpha\beta})
\]
\[
= (h_{\beta}, h_\alpha),
\]
\[
\mathcal{L}_{\alpha\beta\gamma} = ((h_{\beta\gamma}, h_\alpha) + (h_{\beta}, h_{\alpha\gamma}) + (h_{\gamma}, h_{\alpha\beta})
\]
\[
+ (h - g, h_{\alpha\beta\gamma}))
\]
\[
= (h_{\beta\gamma}, h_\alpha) + (h_{\beta}, h_{\alpha\gamma}) + (h_{\gamma}, h_{\alpha\beta}),
\]
\[
\mathcal{L}_{\alpha\beta\gamma\delta} = \ldots .
\]

Through the strict calculation above, we find that the likelihood (4) has the same properties as Equation (1). This is no coincidence. In fact, if one defines the inner product of two random vectors \( u \) and \( v \) in the data space as \( (u, v) \equiv uMv = uM^b_\lambda j \) in Equation (1), then the likelihood can be rewritten as

\[
P(x|\Theta) \propto e^{-\frac{1}{2}x - \mu(\Theta), x - \mu(\Theta)},
\]
which has the same form as Equation (4). The inner product in Equation (7) is a summation, while in Equation (4) the inner product becomes an integral.

We can also make other comparisons of the two likelihoods and inner products. For the two components of the inner product, both \( x - \mu \) and \( g(t) - h(t) \) represent the noise that leads to uncertainty and the probability of the inferred parameters. The covariance matrix \( C \) describes the correlation among the random variables, while in the continuous case it is the autocorrelation function \( R(\tau) \equiv \langle n(t)n(t + \tau) \rangle \) that takes this role. On the other hand, one can see that in Equation (7) the inner product is controlled by the inverse of the covariance matrix \( M = C^{-1} \), so in Equation (4) it is expected be controlled by \( R(\tau)^{-1} = 1/R(\tau) \). Note that Equation (4) is defined in the frequency domain, and \( F[R(\tau)] = S_n(f)/2 \) with \( f \geq 0 \) (Finn 1992). Therefore, we expect that the inner product will be controlled by \( 2/S_n(f) \). This is indeed what Equation (5) gives. Although Equations (1) and (4) can be uniformly described using the inner product, it is still necessary to derive the DALI algorithm, because the GW signals are continuous, which is different from the discrete case in Sellentin et al. (2014).

In real GW data analysis, we have finite data points due to the limited timespans and sampling rates of the data, and this is a discrete version of the continuous signal. We use discrete summation to approximate the integral in Equation (5), but one must be aware that it is not the summation in Equation (1).

Certainly, one can use the autocorrelation function to describe the correlations among the sampling points in the time domain, to obtain the covariance matrix, and to write a likelihood like Equation (1). However, this form is complicated due to the nondiagonal covariance matrix. When the sampling rate tends to infinity, Equation (1) will become the concise form in Equation (5), after some calculation (Finn 1992).

In all, the DALI algorithm still works in GW data analysis, as guaranteed by the same algebraic structure underpinning Equations (1) and (4). Therefore, defining the ML \( \hat{\Theta} \) and the measurement error \( \theta \equiv \Theta - \hat{\Theta} \), the DALI expansion here gives

\[
P(\Theta) \propto \exp \left[ -\frac{1}{2} (h_{\alpha}, h_{\beta}) \theta_{\alpha} \theta_{\beta} - \frac{1}{8} (h_{\alpha\beta}, h_{\gamma\delta}) \theta_{\alpha} \theta_{\beta} \theta_{\gamma} \theta_{\delta} + 1 \theta_{\alpha} \theta_{\beta} \theta_{\gamma} \theta_{\delta} \theta_{\alpha} \theta_{\beta} \theta_{\gamma} \theta_{\delta} + \frac{1}{72} (h_{\alpha\beta\gamma\delta})(\theta_{\alpha\beta\gamma\delta}) \theta_{\alpha} \theta_{\beta} \theta_{\gamma} \theta_{\delta} \right] + O(4).
\]

Note that \( O(4) \) here means terms that contain fourth or higher orders of the derivatives of the model. The highest-order term in Equation (8) is proportional to \( (h_{\alpha\beta\gamma}, h_{\delta\tau\sigma})(\theta_{\alpha\beta\gamma\delta})(\theta_{\delta\tau\sigma}) = (\delta_{\alpha\beta\gamma\delta})(\theta_{\alpha\beta\gamma\delta}) \), which is positive definite and makes the exponential part of Equation (8) tend to negative infinity for large \( \theta \). Therefore, the DALI approximation is normalizable, keeping the properties of a probability density.

Once the ML parameters, the sensitivities of the detectors, and the derivatives of the waveform are given, one can immediately obtain the approximate likelihood for the parameters of interest. The logarithm of Equation (8) is polynomial, making it easy to calculate. Meanwhile, the terms in the two large pairs of round brackets in Equation (8) are called “double-DALI” and “triplet-DALI”, respectively, giving information beyond the Gaussian ellipsoids (Sellentin et al. 2014). It is worth noting that this formula is applicable to any random process with likelihoods and inner products similar to Equations (4) and (5).
3. Considerations in Applying DALI to GW Events

To obtain a better approximation of the likelihood, we introduce the following specifics, before using the DALI method in GW data analysis.

As an extension of the FM, DALI still maintains some properties of the FM (Sellentin et al. 2014), and the most important one is that DALI represents the “average” detection ability of detectors, which is also why we take the data average (\cdots) in Equations (3) and (6).\(^5\) When doing forecasts, one usually chooses a set of typical parameters as the true parameters, and expands the likelihood at the truth point, after an ensemble average of noises. In this case, the results of DALI are obtained after “averaging” over the ensemble, and this cancels out all the effects of particular noise realizations that will occur in a set of real data (see, e.g., Zanolin et al. 2010). We are more interested in the “average” detection capability of the detector, because there is no real signal and a specific noise realization. But for a specific event (corresponding to a specific noise realization), we need to obtain the ML point from the data of the real GW signal at first. Due to the noise realization, the ML point will deviate slightly from the truth, and the likelihood contours will also change accordingly. Therefore, when doing parameter estimation, the DALI series is not expected to be exact with the real posterior in a strict sense. DALI here is an approximation, with errors in both noise and truncations of series. However, as we will show in Section 4, if one can find a parameter point close enough to the truth, DALI can also give an approximation close to the real likelihood. In principle, when it is applied to forecasting, DALI is exact in the meaning of average, but when it is applied to measuring, DALI is an approximation. Nonetheless, in both cases, DALI will improve the results given by the FM.

However, not all of the useful properties of the FM are shared by DALI. The biggest disadvantage of DALI is that the likelihood shapes of DALI must be obtained numerically, while in the FM, because of the good properties of the Gaussian function, we can easily evaluate the \(n-\sigma\) \((n = 1, 2, 3, \cdots)\) confidence level contours analytically, using the inverse of the FM when a uniform prior is chosen (Sellentin et al. 2014).

Based on the rearrangements of the Taylor series, it is very important to choose which parameters to expand, because different parameter expansions have different convergence behaviors. This matter may not be so important in the FM, where one only considers the first derivatives of the model, but in DALI we must investigate the convergence very carefully, according to the model. For example, there are some mathematically equivalent parameter sets controlling the binary masses: (i) the two component masses \(\{m_1, m_2\}\); (ii) the chirp mass and the mass ratio \(\{M_\text{ch}, q\}\); and (iii) the chirp mass and the symmetric mass ratio \(\{M_\text{ch}, \eta\}\). But note that the waveform \(h\) is roughly the polynomial of the chirp mass and symmetric mass ratio. Therefore, expanding \(M_\text{ch}\) and \(\eta\) is expected to have better convergence. Another example is the luminosity distance \(d_L\), which, in parameter estimation, could be replaced by (i) the comoving volume \(V_c\); (ii) the comoving distance \(d_c\); or even (iii) the reciprocal of \(d_c\), namely \(d_c^{-1}\). Because the waveform is proportional to \(d_c^{-1}\), expanding by \(d_c\) will encounter the jumping of the derivative signs, while the higher-order derivatives of the waveform to \(d_c^{-1}\) are zero. Likewise, the convergence of the Taylor series of functions in the form of “1/ \(x^n\)” at the expanding point \(x_0\) is not fine, especially for small \(x_0\). Therefore, in practice, we choose parameters to obtain DALI series with fast convergence, just as Joachimi & Taylor (2011) used the Box–Cox transformations to obtain an approximately Gaussian likelihood in the transformed parameter space.

The selection of the parameters in the likelihood changes the specific functional form of the prior. Some parameters can provide convenience for derivation, but at the same time lead to complex and extreme priors. In GW analysis, we usually choose the prior of the comoving volume \(V_c\) to be uniform, meaning that GW events are uniform in the comoving space. If one chooses the equivalent parameter \(\beta = d_c^{-1}\) (approximately equal to \(d_c^{-1}\) when the redshift \(z\) is small) for the convenience of derivation, the prior will not stay uniform and will become \((dV_c/d\beta) \propto \beta^{-4}\). When \(\beta\) is small, this prior dominates the posterior, and the shape of likelihood will be difficult to depict. The influences of the prior also exist in the FM, and a detailed discussion can be found in Vallisneri (2008).

In all, an appropriate selection of the expansion parameters can help us to reach a better approximation when truncating the DALI series at a finite order, and both the uniformity of the prior and the convenience of derivation need to be considered carefully.

4. Results

In this section, we show the results of our DALI analysis. We choose two famous GW events as examples: the binary black hole event GW150914 (Abbott et al. 2016) and the binary neutron star (BNS) event GW170817 (Abbott et al. 2017). We use data obtained from the Gravitational Wave Open Science Center,\(^6\) a service of the LIGO Laboratory, the LIGO Scientific Collaboration, the Virgo Collaboration, and KAGRA.

To evaluate the full posterior, we use the open source software PARALLEL BILBY (Ashton et al. 2019; Romero-Shaw et al. 2020; Smith et al. 2020) and the nested sampling package DYNESTY (Skilling 2004, 2006; Speagle 2020) to implement parameter inference for the real strain data. For the waveform templates, we use IMRPHENOMD (Husa et al. 2016; Khan et al. 2016) for GW150914 and TAYLORF2 (Buonanno et al. 2009) for GW170817. The derivatives of these two waveforms are easy to obtain, and choosing other waveforms will not change our results significantly.

For the FM and DALI approximation, we expand the likelihood at the medians of the marginalized distributions from the nested sampling results, because these median values are expected to be close to the truth values. These examples have two meanings. On the one hand, one can take the real event as a “simulation” in the natural laboratory, and we are forecasting the results via the truth. This reflects the forecasting ability of DALI. On the other hand, the results show how good DALI can be, as long as we know where to expand the likelihood. In practice, the expansion point can be obtained independent of detailed sampling results, because one only needs a reliable ML point, which is still much easier and faster to obtain than the full posterior, for example, via a steepest descent method.

Both GW150914 and GW170817 have 10 variable parameters. In the parameter estimation of GW150914, we use the

---

\(^5\) In this work, we use the FM in the frequentist approach, which is defined as the data average of \(E_{\text{data}}\). The FM can also be defined in the Bayesian approach, and the differences between these two definitions are discussed in Sellentin et al. (2014).

\(^6\) https://www.gw-openscience.org
chirp mass $M_c$, mass ratio $q$, two aligned dimensionless spins $\chi_1$ and $\chi_2$, R.A. $\alpha$ and decl. $\delta$, polarization angle $\psi$, inclination angle $i$, luminosity distance $d_L$, and trigger time $t_c$. For GW170817, we set the spins of the BNS to be zero, as one would approximate at the leading order from astrophysics, but add the tidal deformability parameters $\Lambda_1$ and $\Lambda_2$. The priors of the above parameters are uniform, except $\delta$ and $i$; the priors of $\sin \delta$ and $\cos i$ are uniform, as in common practice. When expanding the likelihood, we choose the parameter set \{\(M_c\), $\eta$, $\chi_1$, $\chi_2$, $\alpha$, $\delta$, $\psi$, $i$, $t_c$, $\Delta t_c$\} for GW150914 and \{\(M_c\), $\eta$, $\Lambda_1$, $\Lambda_2$, $\alpha$, $\delta$, $\psi$, $i$, $V_c^{-1/3}$, $\Delta t_c$\} for GW170817, where $\eta$ is the symmetric mass ratio, and $\Delta t_c$ is just $t_c$ minus a constant for calculation convenience. As explained in Section 3, using $V_c$ instead of $d_L^{-1}$ to expand the likelihood for distant events can prevent the undesirable features resulting from a nonuniform prior, while for nearby events the other choice is better, because of the convergence of DALI series.

After calculating the derivatives at the expansion point, we use the EMCEE package (Foreman-Mackey et al. 2013) to obtain the approximate posteriors of the FM and DALI numerically. The final results are shown in Figures 1 and 2. For the convenience of the analysis and discussion, we only show the contours of three important parameters \{\(M_c\), $q$, $d_L$\}. The behaviors of the other parameters are similar.

Figure 1 shows the $q$–$M_c$ results of the FM, doublet-DALI, and triplet-DALI, compared with the accurate full posterior of GW150914. The nonelliptical contours in the $q$–$M_c$ distribution given by the FM are due to parameter transformation and the truncation of the prior on boundaries. Similar to Sellentin et al. (2014), DALI gradually improves the posterior given by the FM. One can clearly see that the FM gives contours much larger than the full posterior, which comes from the degeneracy between $m_1$ and $m_2$. If one had chosen the priors of the two component masses to be uniform, the degeneracy would become even more serious. Containing higher-order effects, doublet-DALI reduces the broad distribution of $q$ given by the FM, and keeps the shape of distribution of $M_c$. Triplet-DALI not only cuts down the peak of $q$ in doublet-DALI, but also gives a distribution closer to the full posterior of $M_c$.

Figure 2 shows the posteriors given by the FM, doublet-DALI, triplet-DALI, and the nested sampling of GW150914 (left panel) and GW170817 (right panel). In this figure, there are many interesting results that reflect the properties of DALI. The $d_L$ distribution of GW150914 given by doublet-DALI is worse than the FM, because in GW150914 we choose $V_c$ instead of $d_L^{-1}$ to expand the likelihood, and if we use $d_c^{-1}$ or $d_L^{-1}$ both the FM and DALI will be controlled by the prior. As explained in Section 3, the Taylor expansion of a “1/$x$” type function will suffer from jumps of the derivative signs. Indeed, doublet-DALI shifts the maximum of the $d_L$ marginalized distribution to a larger value, and when adding one more order triplet-DALI makes the maximum smaller, between the FM and doublet-DALI, which is very similar to the oscillation behavior for the Taylor expansion of 1/$x$. When applied to GW170817, DALI has little improvement on the FM in $M_c$ and $q$, compared to GW150914. This can be explained by the small $M_c$ of the BNS and the long chirp signal in the data, leading to a high S/N and a quasi-Gaussian likelihood in the mass parameters $M_c$ and $\eta$. In this case, the FM approximation can give almost the full posterior (expanded with appropriate parameters), and DALI only makes tiny corrections to it. But for $d_L$, DALI’s improvement on the FM is obvious, benefiting from choosing $V_c^{-1/3}$ to expand the likelihood. If one had chosen $V_c$ instead, a more serious shift in the doublet-DALI
would occur, like that of GW150914, because of the small $d_L$ here. These behaviors are all consistent with the discussions in Section 3.

To quantify DALI’s improvements on the FM, we calculate the Kullback–Leibler (KL) divergences (Kullback & Leibler 1951) and the Wasserstein distances (Vaserstein 1969; Peyré and Cuturi 2019) of the 10 one-dimensional marginalized distributions between the approximations and the full posterior. Both the KL divergence and the Wasserstein distance are distance functions defined between two probability distributions that measure their similarity. Compared with the more commonly used KL divergence, the Wasserstein distance is more universal and can reflect the similarity between two distributions that overlap very little. Intuitively, if one regards each probability distribution as an equal (normalized) pile of soil, then the Wasserstein distance can be interpreted as the minimum required “work” to deform one pile of soil into another, which is equal to the amount of soil to be moved multiplied by the average “distance” between the two piles (distributions). Similar distributions will have a small Wasserstein distance.

We show the Wasserstein distances in Table 1. The KL divergences between the parameter distributions of GW150914 and GW170817 are generally consistent with the Wasserstein distances. In these tables, we can clearly see DALI’s improvements to the FM. For example, triplet-DALI reduces the difference between the FM approximation and the real posterior in the chirp mass $M_\text{c}$ distribution of GW150914 by a factor of 5. As the first-order correction after the FM, doublet-DALI is usually radical, giving generally good but sometimes bad results. Triplet-DALI is relatively more conservative in some parameters, such as $\chi_1$ and $\chi_2$, but it improves the FM’s results more comprehensively. Among all the parameters, the sky location parameters $\alpha$ and $\delta$ are worth noting. For GW150914, DALI produces significantly better results than the FM, while for GW170817 DALI is not much different from the FM. This can be explained by the fact that we have data from three detectors for GW170817, but only from two detectors for GW150914. So the former already has almost a Gaussian posterior, while the latter has more room for improvement. Generally, the improvements of DALI for GW150914 are better than those for GW170817, given the lower S/N of GW150914, whose likelihood is more likely to deviate from Gaussian form.

| Parameter | Fisher | Doublet-DALI | Triplet-DALI |
|-----------|--------|--------------|--------------|
| $M_\text{c}$ | 0.776  | 0.767        | 0.157        |
| $q$      | 1.34   | 0.358        | 0.356        |
| $\chi_1$ | 8.07   | 3.79         | 4.61         |
| $\chi_2$ | 8.23   | 3.71         | 4.90         |
| $\alpha$ | 0.614  | 0.217        | 0.160        |
| $\beta$  | 1.09   | 0.677        | 0.438        |
| $\psi$   | 0.389  | 0.326        | 0.367        |
| $\iota$  | 0.22   | 0.915        | 0.241        |
| $d_L$    | 0.534  | 1.64         | 0.258        |
| $\Delta t_c$ | 0.452  | 0.587        | 0.611        |

Table 1. The Wasserstein Distances between FM/DALI and Nested Sampling

Figure 2. The same as Figure 1, but for the parameters $M_\text{c}$, $q$, $d_L$ of GW150914 (left panel) and GW170817 (right panel). Note that in the right panel, we have used $\Delta M_\text{c} = M_\text{c} - 1.19700 M_\odot$. 

Figure 2.
According to Section 3, the results of DALI must be obtained by numerical calculations, leading to longer time costs than the FM. However, the priors of the examples here are not uniform, so discarding the priors and only using the FM to calculate the contours directly leads to additional errors (Vallisneri 2008). Therefore, we evaluate the approximate posteriors for both the FM and DALI numerically, which mathematically equates to replacing complex likelihoods using waveforms and noise-weighted inner products with the trivial polynomial in Equation (8). In practice, DALI indeed takes more time than the FM, but the calculation times of DALI and the FM are still at the same order of magnitude, while obtaining the full posterior will take several orders of magnitude more time than the FM and DALI. Overall, DALI provides improved performance, obtaining a posterior closer to the real full posterior, while not costing much more time than the FM. In particular, the triplet-DALI algorithm performs very well in terms of extending the FM.

5. Conclusion

The FM is a useful tool for obtaining an approximation for complex models and likelihoods when making forecasts and inferences. Benefiting from its high speed and efficiency, the FM is widely used in many fields of physics, especially in the GW community. Lots of works have made use of the FM to estimate the detection ability of GW detectors for specific sources (e.g., Finn 1992; Cutler & Flanagan 1994; Poisson & Will 1995; Cutler 1998; Berti et al. 2005; Zhao & Wen 2018; Wang et al. 2019; Ruan et al. 2020; Kang et al. 2021; Shuman & Cornish 2022). Without doubt, the FM has greatly contributed to the development of GW studies, eventually leading to the first real GW event, GW150914 (Abbott et al. 2016), and the field of GW astrophysics (Abbott et al. 2019, 2021a, 2021b). At the same time, the disadvantages of the FM are obvious. The FM attempts to approximate any likelihood with a Gaussian distribution, which does not apply well in some cases (Vallisneri 2008; Shuman & Cornish 2022). Longing for a more accurate approximation of the likelihood, we have good reasons to extend the FM.

In our work, we extend the FM in GW analysis using the DALI method proposed by Sellentin et al. (2014). Note that the original DALI method deals with likelihoods in finite-dimensional data spaces, so in Section 2 we first extend the DALI method to the continuous case, to fit the requirements of GW analysis. The extended result in Equation (8), which uses the inner product to characterize the likelihood whether the data space is finite or not, is universally valid. Therefore, this formula can be applied to other random processes with stationary Gaussian noise and other types of likelihoods, as long as their algebraic structures are the same as Equations (4) and (7). To illustrate the usefulness of DALI, we choose two famous GW events, GW150914 and GW170817, as examples. The results in Section 4 can be regarded as applications of both forecasting and measuring. Compared to the FM, we find that both doublet-DALI and triplet-DALI bring obvious improvements on the posteriors for many parameters, and that triplet-DALI has overall better performance than doublet-DALI, according to Table 1. In the two examples, DALI can reduce the difference between the FM approximation and the real posterior by a factor of 5 in the best case, and the calculation times of DALI and the FM are still at the same order of magnitude. Therefore, we suggest using triplet-DALI with appropriate parameters to provide stable and reliable corrections to the FM. As in Sellentin et al. (2014) and Sellentin (2015), DALI can reduce the degeneracy that may exist in the FM (see Figure 1), giving a fast and one of the simplest validity checks on the FM analysis in forecasts. Recently, it has been possible to estimate the approximated full posteriors for real GW events in seconds (Chua & Vallisneri 2020; Dax et al. 2021) or minutes (Cornish 2021). In these cases, DALI can still help the sampling techniques by determining beforehand the high-probability regions to explore. In forecasts, the FM and DALI are still important, because the number of mock events is very large and there is no real GW signal. DALI has proved to be a successful extension of the FM and to have successful applications in GW data analysis, maintaining a good balance between the FM approximation and the full posterior in terms of fidelity and computational costs.

There are still a few possible areas where we can improve DALI’s performance. The most direct one will be to consider more higher-order terms, which will definitely lead to more accurate posteriors. However, the number of derivatives required increases exponentially as the order increases, as does the computational cost. Another way will be to choose more appropriate parameters. The parameter sets in Section 4 can provide good results, but there may be better ones. The third involves the measurements. In this work, we use the expansion points given by nested sampling, but for measurements using DALI, we need to find them independently. A fast and reliable method for finding the ML point will significantly improve the speed and fidelity of DALI.

As an extension of the FM, DALI is a useful approximation for likelihoods, and makes better forecasts and inferences beyond the FM. It can fill the gap between the simplest but low-fidelity FM method and the high-fidelity but costly sampling methods. DALI’s ability to break the parameter degeneracy in the FM is also helpful for some aspects of GW analysis.

We thank Zihe An and Yi-Fan Wang for discussions, and the anonymous referees for helpful comments that improved the paper. This work was supported by the National Natural Science Foundation of China (11975027, 12147177, 11991053, and 11721303), the National SKA Program of China (2020SKA0120300), the Young Elite Scientists Sponsorship Program by the China Association for Science and Technology (2018QNRC001), the Max Planck Partner Group Program funded by the Max Planck Society, and the High-Performance Computing Platform of Peking University. Z.W. was supported by the Hui-Chun Chin and Tsung-Dao Lee Chinese Undergraduate Research Endowment (Chun-Tsung Endowment) at Peking University, and J.Z. was supported by the “LiYun” postdoctoral fellowship of Beijing Normal University.

Facility: LIGO/Virgo/KAGRA.

Software: Bilby (Ashton et al. 2019), ChainConsumer (Hinton 2016), dynasty (Skilling 2004, 2006; Speagle 2020), emcee (Foreman-Mackey et al. 2013), Parallel Bilby (Smith et al. 2020), PyCBC (Nitz et al. 2022).
Appendix

DALI Algorithm

The Derivative Approximation for LIkelihoods (DALI) algorithm was developed by Sellentin et al. (2014). A brief discussion of the principles and implementation of the DALI algorithm is given below. It also sets the notations for the paper.

The ultimate goal of parameter estimation is to obtain a probability density function, $P(\Theta)$, where $\Theta \equiv \{ \Theta_\alpha \}$ ($\alpha = 1, 2, \ldots, N_\nu$) collectively denotes $N_\nu$ parameters of interest. We use Greek indices to run over the parameters, and Latin indices to run over the data. By performing the Taylor expansion on the log-likelihood, $L(\Theta) \equiv -\log P(\Theta)$, around the best-fitting parameters, $\hat{\Theta}$, one obtains (Vallisneri 2008; Sellentin et al. 2014):

$$P(\Theta) \propto \exp \left[ -\frac{1}{2} F^{\alpha \beta} \theta_\alpha \theta_\beta - \frac{1}{3!} S^{\alpha \beta \gamma} \theta_\alpha \theta_\beta \theta_\gamma + O(\theta^5) \right], \quad (A1)$$

where $\theta \equiv \Theta - \hat{\Theta}$, and we have defined

$$F^{\alpha \beta} \equiv \frac{\partial^2 L}{\partial \Theta_\alpha \partial \Theta_\beta} \bigg|_{\Theta = \hat{\Theta}},$$

$$S^{\alpha \beta \gamma} \equiv \frac{\partial^3 L}{\partial \Theta_\alpha \partial \Theta_\beta \partial \Theta_\gamma} \bigg|_{\Theta = \hat{\Theta}},$$

$$Q^{\alpha \beta \gamma \delta} \equiv \frac{\partial^4 L}{\partial \Theta_\alpha \partial \Theta_\beta \partial \Theta_\gamma \partial \Theta_\delta} \bigg|_{\Theta = \hat{\Theta}}. \quad (A2)$$

We will use shorthand notations for derivatives in the parameter space,

$$F^{\alpha \beta} = L_{\alpha \beta}, \quad S^{\alpha \beta \gamma} = L_{\alpha \beta \gamma}, \quad Q^{\alpha \beta \gamma \delta} = L_{\alpha \beta \gamma \delta}, \quad (A3)$$

where they are understood to be evaluated at $\Theta = \hat{\Theta}$. The expansion (A1) is only well defined as a normalizable probability density at its second order, characterized by $F^{\alpha \beta}$, and $F^{\alpha \beta}$ is the usual FM (see, e.g., Finn 1992; Cutler & Flanagan 1994). The terms characterized by $S^{\alpha \beta \gamma}$ and $Q^{\alpha \beta \gamma \delta}$ are not positively defined in general (Sellentin et al. 2014).

To cure the problem, instead of performing the Taylor expansion, the DALI algorithm expands $P(\Theta)$ in the order of the derivatives (Sellentin et al. 2014). Assume a theoretical model $\mu(\Theta)$, whose component $\mu_i(\Theta)$ ($i = 1, 2, \ldots, N_\nu$) predicts the $i$th data point, and a covariance matrix $C$ with dimensions $N_d \times N_d$ in the data space whose inverse is $M \equiv C^{-1}$. The DALI expansion gives

$$P(\Theta) \propto \exp \left[ -\frac{1}{2} \mu_\alpha M_{\alpha \beta} \mu_\beta + \frac{1}{6} \mu_\alpha \mu_\beta \mu_\gamma M_{\alpha \beta \gamma} + \frac{1}{12} \mu_\alpha \mu_\beta \mu_\gamma \mu_\delta M_{\alpha \beta \gamma \delta} + O(4) \right], \quad (A4)$$

where the first term in the exponent gives the usual FM, while the following two pairs of round brackets give the corrections at higher orders, and are named as the “doublet-DALI” and the “triplet-DALI”, respectively (Sellentin et al. 2014). Such an expansion guarantees the properties of probability densities like Equation (8).

ORCID iDs

Ziming Wang © https://orcid.org/0000-0002-8742-8397
Chang Liu © https://orcid.org/0000-0001-7649-6792
Junjie Zhao © https://orcid.org/0000-0002-9233-3683
Lijing Shao © https://orcid.org/0000-0002-1334-8853

References

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2016, PhRvL, 116, 061102
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, PhRvL, 119, 161101
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2019, PhRvX, 9, 031040
Abbott, R., Abbott, T. D., Abraham, S., et al. 2021a, PhRvX, 11, 021053
Abbott, R., Abbott, T. D., Acernese, F., et al. 2021b, arXiv:2111.03606
Ashton, G., Hübner, M., Lasky, P. D., et al. 2019, ApJS, 241, 27
Berti, E., Buonanno, A., & Will, C. M. 2005, PhRvD, 71, 084025
Buonanno, A., Iyer, B., Ochsner, E., Pan, Y., & Sathyaprakash, B. S. 2009, PhRvD, 80, 064043
Chua, A. J. K., & Vallisneri, M. 2020, PhRvL, 124, 041102
Cornish, N. J. 2021, PhRvD, 104, 104054
Cutler, C. 1998, PhRvD, 57, 7089
Cutler, C., & Flanagan, E. E. 1994, PhRvD, 49, 2658
Dax, M., Green, S. R., Gair, J., et al. 2021, PhRvL, 127, 241103
Finn, L. S. 1992, PhRvD, 46, 5236
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Hinton, S. R. 2016, JOSS, 1, 00045
Husa, S., Khan, S., Hannam, M., et al. 2016, PhRvD, 93, 044006
Isogawa, S., Nakano, H., & Nakamura, T. 2018, PTEP, 2018, 073E01
Joachimi, B., & Taylor, A. N. 2011, MNRAS, 416, 1010
Kang, Y., Liu, C., & Shao, L. 2021, AJ, 162, 247
Khan, S., Husa, S., Hannam, M., et al. 2016, PhRvD, 93, 044007
Kullback, S., & Leibler, R. A. 1951, Ann. Math. Stat., 22, 79
Liu, C., & Shao, L. 2022, ApJ, 926, 158
Liu, C., Shao, L., Zhao, J., & Gao, Y. 2020, MNRAS, 496, 182
Nitz, A., Harry, I., Brown, D., et al. 2022, gwaastro/pycbc: Release v2.0.1 of pyCBC, Zenodo, doi:10.5281/zenodo.5825666
Pepey, G., & Cuturi, M. 2019. Found. Trends Mach. Learn., 11, 355
Poisson, E., & Will, C. M. 1995, PhRvD, 52, 848
Poisson-Shaw, I. M., et al. 2020, MNRAS, 499, 3295
Ruan, W.-H., Liu, C., Guo, Z.-K., Wu, Y.-L., & Cai, R.-G. 2020, NatAs, 4, 108
Sellentin, E., Quartin, M., & Amendola, L. 2014, MNRAS, 441, 1831
Sellentin, E., & Schügerl, B. M. 2016, MNRAS, 456, 1645
Shuman, K. J., & Cornish, N. J. 2022, PhRvD, 105, 064055
Skilling, J. 2004, in AIP Conf. Ser. 735, Bayesian Inference and Maximum Entropy Methods in Science and Engineering, ed. R. Fischer, R. Preuss, & U. V. Toussaint (Melville, NY: AIP), 395
Skilling, J. 2006, BayeAn, 1, 833
Smith, R. J. E., Ashton, G., Vajpeyi, A., & Talbot, C. 2020, MNRAS, 498, 4492
Speagle, J. S. 2020, MNRAS, 493, 3122
Vallisneri, M. 2008, PhRvD, 77, 042001
Vallisneri, M. 2011, PhRvL, 107, 191104
Vaserstein, L. N. 1969, Probl. Peredachi Inf., 5, 64, http://www.ams.org/mathscinet-getitem?mr=314115
Veitch, J., Raymond, V., Farr, B., et al. 2015, PhRvD, 91, 042003
Wang, H.-T., Jiang, Z., Sesana, A., et al. 2019, PhRvD, 100, 043003
Zanolin, M., Vitale, S., & Makris, N. 2010, PhRvD, 81, 124048
Zhao, J., Shao, L., Guo, Y., et al. 2021, PhRvD, 104, 084008
Zhao, W., & Wen, L. 2018, PhRvD, 97, 064031

Wang et al.