Testing general relativity with compact coalescing binaries: comparing exact and predictive methods to compute the Bayes factor

Walter Del Pozzo, Katherine Grover, Ilya Mandel and Alberto Vecchio

School of Physics and Astronomy, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK
E-mail: wdp@star.sr.bham.ac.uk

Received 10 June 2014
Accepted for publication 31 July 2014
Published 3 October 2014

Abstract
The second generation of gravitational-wave detectors is scheduled to start operations in 2015. Gravitational-wave signatures of compact binary coalescences could be used to accurately test the strong-field dynamical predictions of general relativity (GR). Computationally expensive data analysis pipelines, including TIGER (test infrastructure for general relativity), have been developed to carry out such tests. As a means to cheaply assess whether a particular deviation from GR can be detected, Cornish et al (2011 Phys. Rev. D 84 062003) and Vallisneri (2012 Phys. Rev. D 86 082001) recently proposed an approximate scheme to compute the Bayes factor between a GR gravitational-wave model and a model representing a class of alternative theories of gravity parametrized by one additional parameter. This approximate scheme is based on only two easy-to-compute quantities: the signal-to-noise ratio (SNR) of the signal and the fitting factor (FF) between the signal and the manifold of possible waveforms within GR. In this work, we compare the prediction from the approximate formula against an exact numerical calculation of the Bayes factor using the lalinference library. We find that, using frequency-domain waveforms, the approximate scheme predicts exact results with good accuracy, providing the correct scaling with the SNR at a FF value of 0.992 and the correct scaling with the FF at a SNR of 20, down to a FF of ∼0.9. We extend the framework for the approximate calculation of the Bayes factor, which significantly increases its range of validity, at least to FFs of ∼0.7 or higher.
1. Introduction

The upgraded versions of the ground-based gravitational wave detectors LIGO [1, 2] and Virgo [3–6] are expected to detect gravitational-wave signals from the coalescence of compact binary systems. The prospect of frequent detections, with expected rates between one per few years and a few hundred per year [7], promises to yield a variety of scientific discoveries. Among these, the possibility of testing the strong field dynamics of general relativity (GR) has received increasing attention (e.g., [9–14]). In fact, during the latest phase of the inspiral, typical orbital velocities are an appreciable fraction of the speed of light (v/c ~ 0.4); following merger, the compactness $G M / R c^2$ of the newly formed black hole that is undergoing quasinormal ringing is close to one. By comparison, the orbital velocity of the double pulsar J0737-3039 is $O(10^{-3})$ and its compactness is $\sim 10^{-6}$ [8]. Consequently, efforts have concentrated on the development of robust frameworks to reliably detect deviations from GR using gravitational-wave signatures of compact-binary mergers.

One of these frameworks is the so-called test infrastructure for general relativity (TIGER) [9, 13, 14]. TIGER operates by computing the odds ratio between GR and a test model in which one or more of the post-Newtonian coefficients are allowed to deviate from the value predicted by GR. The interested reader is referred to [9, 13, 14] for the details of the method and for analysis of its robustness against various potential systematic effects. To account for unmodelled effects, TIGER constructs a ‘background’ distribution of odds ratios between GR and the test hypothesis by analyzing $O(10^3)$ simulated GR signals. The background distribution defines the null hypothesis against which any particular observation (or catalog of observations) is tested. For validation purposes, the sensitivity of the algorithm to a specific deviation from GR is currently assessed by comparing it with a ‘foreground’ odds ratio distribution. The foreground distribution is constructed by simulating a variety of signals in which the chosen deviation from GR is introduced. If the integrated overlap between the foreground and background distributions is smaller than a given false alarm probability, sensitivity to that particular deviation can be claimed.

The process described in the previous paragraph is extremely computationally expensive. If arbitrary combinations of $k$ post-Newtonian coefficients are allowed to deviate from GR values, the total number of simulations that is necessary to construct the background is $2^k$ for each synthetic source.

As a means to cheaply evaluate the detectability of particular deviations from GR, Cornish et al [11] proposed an approximate formula to calculate the odds ratio between GR and an alternative model for gravity (AG). Subsequently, Vallisneri [15] proposed a similar approximation derived from the Fisher matrix formalism. Vallisneri’s approximation considers the distribution of the odds ratio in the presence of noise and characterizes the efficiency and false alarm of a Bayesian detection scheme for alternative theories of gravity. Whilst neither of these approaches can replace the necessary analysis for real data, the possibility of having a quick and easily understandable formalism to check the performance of complex pipelines such as TIGER and assess whether a specific type of deviation is detectable without having to run thousands of simulations seems quite attractive.
In this work we investigate, in an idealized and controlled scenario, whether the predictions from [11, 15] are in agreement with the output of a numerical Bayesian odds-ratio calculation. We find in particular that the analytical prescription of [15] is in reasonable agreement with the numerical result when the fitting factor (FF) between AG and GR waveforms is $\geq 0.9$, and that for FF $\leq 0.8$, both analytical prescriptions overestimate the exact odds ratio.

Nevertheless, when the analytical odds ratio is regarded as an upper limit, useful indications of the detectability of a given deviation from GR can be drawn.

We analytically correct the approximate framework for computing the Bayes factor by introducing terms that are negligible at $\sim FF^{-1}$, reproducing the proposed analytical expressions given in [11, 15], but become significant at lower values of the FF. We show that these corrections extend the range of validity of the approximate expressions at least down to FF values of $\sim 0.7$.

The rest of the paper is organized as follows: in section 2 we briefly review the Bayesian definition of the odds ratio; in section 2.1 we introduce the formula from [15]. In section 3 we present our findings, and finally we discuss them in section 4.

2. Bayesian inference for gravitational wave signals

In a Bayesian context, the relative probability of two or more alternative hypotheses given observed data $d$ is described by the odds ratio (see, e.g., [10]). If GR is the general relativity hypothesis and AG is the hypothesis corresponding to some alternative theory of gravity, the odds ratio is given by

$$O_{AG,GR} = \frac{p(AG | d)}{p(GR | d)} = \frac{p(AG) p(d | AG)}{p(GR) p(d | GR)} \equiv \frac{p(AG)}{p(GR)} B_{AG,GR},$$

where we introduced the Bayes factor $B_{AG,GR}$, which is the ratio of the marginalized likelihoods (or evidences). The marginal likelihood is the expectation value of the likelihood of observing the data given the specific model $H$ under consideration of the prior probability distribution for all the model parameters $\theta$

$$p(d | H) \equiv Z = \int d\theta p(d | \theta, H)p(\theta | H).$$

With the exception of a few idealized cases, the integral (2) is, in general, not tractable analytically. In gravitational-wave data analysis, the parameter space is at least nine-dimensional (for binaries with components that are assumed to have zero spin), and up to 15-dimensional for binaries with arbitrary component spins, and the integrand is a complex function of the data and the waveform model. For stationary Gaussian noise

$$p(d | \theta, H) \propto \exp \left[-(d - h(\theta)\bar{d} - h(\theta))/2\right];$$

where $h(\theta) \equiv h(\theta | H)$ is the model waveform given parameters $\theta$ and we introduced the scalar product

$$(a | b) \equiv 2 \int_0^\infty df \frac{a(f)b(f) + a(f)^*b(f)}{S(f)}$$

with the one-sided noise power spectral density $S(f)$. We analyzed data from a single detector with a noise spectral density corresponding to the zero-detuning, high-power advanced LIGO design configuration [25].
2.1. Analytical approximation

Vallisneri [15] proposed an analytical approximation to the integral (2). He considered the following assumptions:

- linear signal approximation leading to a quadratic approximation of the log likelihood;
- only one additional dimension is necessary to describe the AG model;
- uniform prior distributions for all parameters describing both GR and AG models;
- the distance between the AG waveform and the manifold of GR waveforms is small so that the FF between the two, defined as

\[
\text{FF} = \left( \frac{\left(h_{AG} | h_{GR}(\theta)\right)}{\sqrt{(h_{AG} | h_{AG}(\theta) | h_{GR}(\theta))}} \right)_{\max \over \theta},
\]

is close to unity.

With the above assumptions, the integral equation (2) can be approximately computed analytically and the Bayes factor (1) is then given by

\[B_{AG, GR} \approx \sqrt{2\pi \frac{\Delta \theta_{a \text{ prior}}}{\Delta \theta_{a \text{ est}}}} e^{\rho (1 - \text{FF})},\]

where \(\rho\) denotes the optimal signal-to-noise ratio (SNR)\(^1\)

\[\rho \equiv 2 \sqrt{\int_0^\infty df \frac{|h(\theta_{\text{true}})|^2}{S(f)}}.\]

The terms \(\Delta \theta_{a \text{ prior}}\) and \(\Delta \theta_{a \text{ est}}\) are the width of the prior distribution and of the Fisher matrix 1-\(\sigma\) uncertainty estimate for the additional AG parameter, respectively. Equation (6) given here is valid for the case in which a zero realization of the noise is present in the data. In [15] noise is considered and the appropriate formulae for the distribution of the Bayes factor over noise realizations can be found there. We opted for a zero-noise case for ease of comparison.

3. Comparison between the exact calculation and the analytical approximation

We compare the prediction from equation (6) with the evidence calculated by the Nested sampling algorithm [18] as implemented in lalinfer\(\text{ence}\) [17] in a simple experiment. Using the test waveform model presented in [14], we generate inspiral signals which would span a range of FFs. The testing waveform is a frequency-domain stationary phase approximation waveform, based on the TaylorF2 approximant [16], that has been modified in such a way that the post-Newtonian coefficients are allowed to vary around the GR values within a given range. The TaylorF2 waveform for a face-on, overhead binary is given by

\[h(f) = \frac{1}{D} \sqrt{\frac{5}{24}} \pi^{-2/3} M^{3/2} \lambda^{-7/6} e^{\Psi(f)} h,\]

\(^1\) Note that the definition by Vallisneri of the signal-to-noise ratio (SNR) is different from ours. In [15] the SNR quantity that appears in equation (6) is the norm of an hypothetical GR signal whose parameters are exactly the same as the ‘true’ AG waveform, but with the extra AG parameter set to zero. In our case, the SNR corresponds to the power in the AG signal (in the ideal case when it is filtered with AG templates). However, when the AG parameter is present only in the phase of the gravitational wave these two SNRs coincide.
where $D$ is the luminosity distance, $\mathcal{M}$ is the chirp mass and the phase $\Psi(f)$ is

$$
\Psi(f) = \frac{2\pi f t_\text{coa} - q_0 - \pi}{4} + \sum_{i=0}^{7} \left[ \psi_i + \psi_i^{(I)} \ln f \right] f^{(i-5)/3}.
$$

The explicit forms of the coefficients $\psi_i$ and $\psi_i^{(I)}$ in $(\mathcal{M}, \eta)$, where $\eta$ is the symmetric mass ratio, can be found in [23]. In all our experiments we kept the parameters of the simulated sources fixed, with the exception of the 1.5 post-Newtonian coefficient $\psi_3$ which we varied between $[0.5, 1.5]$ times its GR value by adding an arbitrary shift $d\chi_3$ between $[-0.5, 0.5]$.

$$
\psi_3 \rightarrow \psi_3 \left( 1 + d\chi_3 \right).
$$

The Nested sampling algorithm was set up to sample from the following prior:

- the component masses were allowed to vary uniformly $\in [1, 7]M_\odot$ with the total mass constrained to the range $\in [2, 8]M_\odot$. This choice results in an allowed region of triangular shape in the $\mathcal{M}, \eta$ plane, see figure 1;
- uniform on the two-sphere for sky position and orientation parameters;
- uniform in Euclidean volume for the luminosity distance;
- for recovery with AG templates, we used only one free testing parameter ($d\chi_3$) which was allowed to vary uniformly between $[-0.5, 0.5]$ times its GR value.

The FFs were computed from the maximum likelihood values obtained from the lalinf simulations, see appendix A. The parameter uncertainty for equation (6) was computed using a five-dimensional Fisher matrix calculation in which we varied the two mass parameters, the time of coalescence, the phase at coalescence and the deviation parameter $d\chi_3$.

Our experiments were performed analyzing simulated signals from a system whose component masses were chosen to be $1.4M_\odot + 4.5M_\odot$. We chose this system because it lies in the centre of our prior probability distribution over the masses, far away from prior boundaries. This minimizes the impact of the prior on the FF and Bayes factor computations, which
ensures that we can make a fair comparison with equation (6) derived under the assumption of a uniform prior\(^2\).

The approximate formula in equation (6) depends essentially on two quantities: the SNR \(\rho\) and the FF. Below, we describe our investigations of the dependence of the Bayes factor on these two quantities.

For Nested sampling calculations, the uncertainty on the calculated value of the evidence \(Z\) is evaluated as \[^{18}\Delta \approx Z H n \log , (11)\]

where \(H\) is the Kullback–Leibler divergence (or relative entropy) between the posterior distribution and the prior distribution, and \(n\) is the number of live points used for Nested sampling. \(H\) is computed by the Nested sampling alongside the evidence \(Z\). Typical values for \(\Delta Z \log \) are \(-O(10^{-1})\).

Finally, it was recently pointed out that when a signal terminates abruptly in the detector band, measurement uncertainty may be significantly smaller than predicted by the Fisher matrix calculation \[^{22}\]. To avoid these complications, we limited our analysis to frequencies between 30 and 512 Hz.

### 3.1. Scaling with the SNR

We investigated the dependence of the Bayes factor on SNR by comparing the output of lalinference and equation (6) for a \(1.4 M_\odot + 4.5 M_\odot\) system at SNRs of 10, 20, 30 and 40 at a fixed value of the FF, 0.992. Figure 2 shows the Bayes factors from the two calculations.

The quadratic dependence of the Bayes factor on the SNR was verified by means of a simple chi-squared fit to an expression of the form

\[
\ln B_{AG,GR} = \alpha \text{SNR}^\beta + \gamma.
\]  

\(^2\) For example, an equal mass system would lie exactly on the prior boundary at \(\eta = 0.25\). For this reason, the GR model has very little room in the \(\eta\) direction to accommodate the additional phase shift due to a non-zero \(d_\chi\). The net result is a very rapid drop in FF towards negative \(d_\chi\).
The scaling of $\log (B_{AG,GR})$ with the SNR appears to be consistent with the expected quadratic dependence: we find $\beta = 1.95 \pm 0.4$.

### 3.2. Scaling with the FF

We evaluate the dependence of the Bayes factor on FF by again injecting a signal from a $1.4M_\odot + 4.5M_\odot$ binary, now at a fixed SNR of 20 but with varying FF. As in the previous section, we vary the FF by adding arbitrary deviations from the GR value to the 1.5 post-Newtonian phase coefficient. In particular, we varied $d\chi^3$ between $-0.5$ and 0.5, leading to FF $\in [0.7, 1.0]$. We verified that our injection was sufficiently far from prior boundaries by confirming that the Bayes factor is the same for positive and negative values of $d\chi^3$ that yield the same FF.

Figure 3 shows the logarithmic Bayes factor computed by lalinference and from equation (6). The two methods agree for FF $\sim 1$. At FF $\leq 0.9$, the analytical approximation overestimates the value of the Bayes factor compared to lalinference. Moreover, the disagreement gets worse with decreasing FF, suggesting a nonlinear dependence on the FF. In the next section, we investigate the approximate analytical expression in greater detail and derive additional corrections that extend its validity to lower FFs.

### 3.3. Correcting the analytical expression for lower FFs

Under the assumption that the region of likelihood support on the parameter space is small, and that over this region the prior does not vary significantly, the evidence for any of the models $H_i$, depending on parameters $\theta$, under consideration can be approximated as (e.g., [24], correcting for a typo in the exponent of $(2\pi)$)

$$Z(H_i) \propto \left[ L_{H_i} \right]_{\text{max over } \theta} (2\pi)^{N/2} \prod_i \frac{\Delta \theta_{\text{est}}^i}{\Delta \theta_{\text{prior}}^i} ,$$  

where $N$ is the number of parameters. Strictly speaking, the equation above is only valid when parameters are uncorrelated. In the general case of correlated parameters, $\prod_i \Delta \theta_{\text{est}}^i$ should be replaced with the uncertainty volume in which the likelihood has support, while $\prod_i \Delta \theta_{\text{prior}}^i$ is
shorthand for the total prior volume. However, such correlations do not affect the scaling of the uncertainty with the SNR, and do not impact our conclusions.

Therefore, the Bayes factor between the AG and GR model can be approximated as the ratio of the maximum likelihoods times the product of the ratios of posterior widths to prior supports

\[
B_{\text{AG,GR}} \approx \frac{[L_{\text{AG}}]_{\text{max over } \theta'}}{[L_{\text{GR}}]_{\text{max over } \theta'}} \sqrt{2\pi} \prod_{i=0}^{N-1} \frac{\Delta \theta_i^\prime / \Delta \theta_i^{\prime \text{prior}}}{\Delta \theta_i^{\prime \text{est}} / \Delta \theta_i^{\prime \text{prior}}},
\]

where \( \theta' \) and \( \theta \) are parameter vectors within the AG and GR models, respectively, and \( N \) is the dimensionality of the AG parameter space.

We begin by considering just the first term in equation (14), which scales exponentially with the SNR in contrast to the components of the second term, which scale inversely with the SNR. Neglecting the second term, we find

\[
\log \left( B_{\text{AG,GR}} \right) \propto \log \left( [L_{\text{AG}}]_{\text{max over } \theta'} \right) - \log \left( [L_{\text{GR}}]_{\text{max over } \theta} \right).
\]

Using equation (A.9), we find

\[
\log \left( B_{\text{AG,GR}} \right) \propto \frac{\rho^2}{2} \left( 1 - FF^2 \right),
\]

which is the expression originally proposed in [11]. At FF close to unity, \( (1 - FF^2) \approx 2(1 - FF) \), the approximation implicitly made in [15], and we recover equation (6). However, we expect (16) to lead to a better fit at low FFs. The filled (red) dots in figure 3 show the Bayes factor computed via equation (16), with the proportionality constant fixed to be the same as in equation (6). Indeed, equation (16) predicts Bayes factors that are in closer agreement with the exact ones than equation (6). In this case, disagreements with the exact result can be seen for FF \( \sim 0.75 \), when the differences in the local shapes of the GR and AG manifolds can become significant.

Vallisneri [15] further assumed that the priors and measurement uncertainties on all parameters except the one describing the deviation from GR, \( \theta_i^a \), are the same for the AG and GR models (which, in turn, is a statement about the similarity in the shape of the two waveform manifolds near the maximum likelihood locations). In this case, the Bayes factor between the two models is [see (6)]

\[
B_{\text{AG,GR}} \propto \frac{[L_{\text{AG}}]_{\text{max over } \theta'}}{[L_{\text{GR}}]_{\text{max over } \theta'}} \sqrt{2\pi} \frac{\Delta \theta_i^{\prime \text{est}}}{\Delta \theta_i^{\prime \text{prior}}},
\]

where \( \theta_i^a \) again refers to the one additional AG parameter which describes the deviation from GR.

However, we should not expect that the posterior widths will be identical in the AG and GR models, for all parameters except the additional AG parameter are the same in the AG and GR models. At high SNRs where the log likelihood can be approximated by a quadratic, posterior widths should scale inverse with the SNR \( \rho \). While \( \rho \) is the optimal SNR recovered when AG templates are used within the AG model, the maximal SNR recoverable when using GR templates within the GR model is lower. By definition, this GR SNR is
\[ \rho_{\text{GR}} \equiv \left( \frac{h_{\text{AG}}(\theta)}{\sqrt{h_{\text{GR}}(\theta) h_{\text{GR}}(\theta)}} \right)_{\text{max over } \theta} \equiv \text{FF } \rho . \] (18)

Assuming the inverse SNR scaling of the posteriors, and using identical priors on common parameters in the AG and GR models, equation (14) reduces to

\[ B_{\text{AG,GR}} \approx \frac{L_{\text{AG}}}{L_{\text{GR}}} \text{max over } \theta' \text{FF}^{N-1} \sqrt{2\pi} \frac{\Delta \theta_{\text{est}}^2}{\Delta \theta_{\text{prior}}^2}. \] (19)

Taking a logarithm of this equation and again using \( (\rho^2/2)(1 - \text{FF}^2) \) for the difference between maximum likelihoods (A.9), we find

\[ \log \left( B_{\text{AG,GR}} \right) \approx \frac{\rho^2}{2} \left( 1 - \text{FF}^2 \right) + (N - 1) \log (\text{FF}) \]

\[ + \log \left( \sqrt{2\pi} \frac{\Delta \theta_{\text{est}}^2}{\Delta \theta_{\text{prior}}^2} \right) . \] (20)

Equation (20) reduces to equation (6) for \( \text{FF} \sim 1 \). However, it is accurate for a much wider range of FFs. Figure 4 shows the comparison between the log-Bayes factors from \text{lalinference} (error bars), the ones from equation (6) (circles) and finally the ones from equation (20) (red dots). Indeed, the log-Bayes factors from equation (20) show a very close agreement with the numerical values. Thus, equation (20) provides a good approximation to the exact values of the log-Bayes factors.

Another merit of equation (20) is that it can be generalized to an arbitrary number of extra non-GR parameters. If we have \( k \) non-GR parameters, equation (20) becomes

\[ \log \left( B_{\text{AG,GR}} \right) \approx \frac{\rho^2}{2} \left( 1 - \text{FF}^2 \right) + (N - k) \log (\text{FF}) \]

\[ + \log \left( 2\pi^{k/2} \prod_{i=1}^{k} \frac{\Delta \theta_{\text{est}}^2}{\Delta \theta_{\text{prior}}^2} \right) . \] (21)
Throughout this work, we have restricted our attention to the zero-noise realization. Vallisneri analyzed the distribution of the Bayes factor under different noise realizations and showed (see equation (15) of [15]) that fluctuations in the logarithm of the Bayes factor have a standard deviation of $\sim \sqrt{2} \rho \sqrt{1 - \text{FF}}$. While our additional corrections to the Bayes factor also lead to corrections in this quantity, we neglect these second-order effects.

We can compare the two systematic corrections discussed above to the level of these statistical fluctuations due to noise. The difference in the log-Bayes factor between equation (6) and equation (16), i.e., the difference between the approximations of [15] and [11], is $(1/2)\rho^2(1 - \text{FF}^2) - \rho^2(1 - \text{FF}) = -(1/2)\rho^2(1 - \text{FF})^2$. This difference is approximately equal to the statistical fluctuation in the log-Bayes factor for $\rho = 20$ and FF $\sim 0.73$, corresponding to the rightmost points in figure 4. Meanwhile, the new correction to the log-Bayes factor which we introduced in equation (20) has a magnitude of $(N - 1) \log \text{FF}$; for $N = 9$ and other parameters as above, it is several times smaller than the noise-induced fluctuations.

Therefore, these corrections are unlikely to impact the detectability of a deviation from GR; in any case, in practice, the detectability of the deviation would be determined by an analysis of the data and a numerical computation of the Bayes factor, not approximate predictive techniques. However, these corrections are useful in explaining the apparent difference between numerical and analytical calculations, and therefore help validate both approaches by enabling a successful cross-check.

4. Discussion

We computed the Bayes factor between a GR model and an alternative gravity model for a gravitational-wave signature of an inspiraling compact binary. We compared two calculations of the Bayes factor: an exact numerical computation with lalinference and an approximate analytical prediction due to Vallisneri [15]. We verified that the analytical approximation yields the correct scaling of the logarithm of the Bayes factor with the square of the SNR at high FF values. However, the predicted scaling of the Bayes factor with the fitting factor is inaccurate for $\text{FF} \leq 0.9$.

We extended the regime of validity of the analytical approximation of [15] to lower FFs by including additional FF-dependent terms and by extending to multiple non-GR parameters. We confirmed that the more complete analytical prediction that we derived in this work, equation (20), remains valid down to FFs of $\leq 0.7$.

It is worth noting that equation (6) loses accuracy precisely in the regime where it becomes possible to differentiate GR and alternative gravity models. The FF is very close to unity in the regime in which the GR waveform can still match a signal which violates GR through different choices of the values of the binary’s parameters within the GR framework. The Bayes factor in this case is not significantly different from 1, thus no decision on the nature of the signal can be made at an acceptably low false alarm probability\(^3\). Therefore, our extension of the analytical expression for the Bayes factor to lower FFs provides a useful, easy-to-compute approximate technique precisely in the regime of interest in the case of a zero noise realization.

The analytical expressions presented in [11, 15] and in this work are predicated on the assumption that the $(N - 1)$-dimensional GR manifold and the $N$-dimensional AG manifold are sufficiently similar near the maximum-likelihood values that the parameter uncertainties

\(^3\) This regime is a case of the so-called ‘fundamental bias’ [19]. It is treated using the analytical approximation presented in [15] by [20]. A numerical study with the lalinference code can be found in [21].
can be assumed to be equal (up to scaling with the inverse SNR) on the two manifolds. Differences in the local curvature of the two manifolds could become significant when the distance between them is large, or the systematic bias between true and best-fit parameters is significant relative to statistical measurement uncertainty. Therefore, this assumption could (although need not) break down either at small FFs, or, somewhat paradoxically, at large SNR for a fixed FF. Specifically, when \( \rho^2 (1 - FF^2) \gg N \), the uncertainty region within a manifold is much smaller than the distance between manifolds or between true and best-fit parameters within a manifold, and the AG and GR manifolds may no longer yield similar parameter uncertainties.

Another possible cause of the breakdown of the analytical approximation is the impact of priors. If the prior distribution is very non-uniform within the region of likelihood support, particularly if a sharp prior boundary is present within this region, the analytical approach described above is no longer valid. A further limitation is the restriction to high SNRs. The widths of the posterior distributions are inversely proportional to the SNR only when the linearized-signal approximation is valid (i.e., when the covariance matrix is well approximated by the inverse of the Fisher matrix).

In summary, the analytical approximation presented by Cornish et al [11] and Vallisneri [15], and its extensions as given in equations (20) and (21), provide a computationally cheap way of predicting the detectability of a deviation from GR for a given AG theory without the need to run expensive numerical simulations, subject to the limitations outlined above. Hence, these analytical approximations can be a very useful tool to get quick indications of whether a particular class and magnitude of one-parameter deviations from GR are detectable. However, these methods are merely predictive, and inference on actual data must rely on parameter estimation and model comparison with complete data-analysis pipelines.

**Acknowledgments**

We thank Michele Vallisneri, Neil Cornish, John Veitch, Will Farr, Christopher Berry, Carl-Johan Haster and Zachary Hafen for useful comments and discussions. The work was funded in part by a Leverhulme Trust research project grant. The numerical simulations were performed on the Tsunami cluster of the University of Birmingham.

**Appendix A. Computing FFs from log likelihoods**

The FF equation (5) can be extracted directly from the Nested sampling runs without the need to search over a parameter grid. Begin by writing the logarithmic likelihood in a zero noise realization

\[
\log (L) = \text{const} + \langle h_{\text{true}} | h(\theta) \rangle - \frac{(h_{\text{true}}) h_{\text{true}}}{2} - \frac{(h(\theta)) h(\theta)}{2},
\]

where \( h_{\text{true}} \) is the gravitational wave signal in the data stream, \( h(\theta) \) is the search template, and const is a constant. Consider the difference \( \Delta \lambda \) between the maximum log likelihoods given for the AG and GR models given an AG signal

\[
\Delta \lambda = \log (L_{\text{GR}})_{\text{max over } \theta} - \log (L_{\text{AG}})_{\text{max over } \theta'},
\]
where
\[ \log(L_{GR}) = \text{const} + \left( h_{\text{true}} | h_{GR}(\theta) \right) - \frac{(h_{\text{true}}h_{\text{true}})}{2} - \frac{(h_{GR}(\theta)h_{GR}(\theta))}{2} \] (A.3)

and \( \log(L_{AG})_{\text{max over } \theta} = \text{const} \) since this likelihood is maximized for \( h_{AG}(\theta') = h_{\text{true}} \).

We can maximize the GR log-likelihood analytically over the amplitude of \( h_{GR}(\theta) \) by defining
\[ h_{GR}(\theta) = A\hat{h}_{GR}(\xi) \] (A.4)

with \( \xi \equiv \theta \setminus A \) being the set of parameters other than the amplitude. One can solve for the value of the amplitude that satisfies
\[ \frac{d \log(L_{AG})}{dA} = 0 \] (A.5)

and obtain
\[ A = \frac{(h_{\text{true}}h_{GR}(\xi))}{(\hat{h}_{GR}(\xi)\hat{h}_{GR}(\xi))} \] (A.6)

Substituting this into equation (A.3) and setting \( (h_{\text{true}}h_{\text{true}}) \equiv \rho^2 \) yields
\[ \Delta \lambda = \frac{1}{2} \left( \frac{(h_{\text{true}}h_{GR}(\xi))^2}{(\hat{h}_{GR}(\xi)\hat{h}_{GR}(\xi))} \right)_{\text{max over } \xi} - \frac{\rho^2}{2}. \] (A.7)

Meanwhile, the FF can be similarly written as
\[ \text{FF} = \left[ \frac{(h_{\text{true}}h_{GR}(\xi))}{\rho \sqrt{(\hat{h}_{GR}(\xi)\hat{h}_{GR}(\xi))}} \right]_{\text{max over } \xi}. \] (A.8)

Thus,
\[ \Delta \lambda = \frac{1}{2} \rho^2 \text{FF}^2 - \frac{\rho^2}{2} = -\frac{\rho^2}{2} \left( 1 - \text{FF}^2 \right) \] (A.9)

and
\[ \text{FF} = \sqrt{\frac{2\Delta \lambda}{\rho^2} + 1}. \] (A.10)

The numerically computed values of \( \Delta \lambda \) have an intrinsic variability due to the stochastic nature of the sampler. The standard deviation for \( \Delta \lambda \) derived from our simulations is \( \sigma_{\Delta \lambda} = 0.016 \). The corresponding uncertainty in our FF estimate is given by
\[ \sigma_{\text{FF}} = \frac{1}{\rho^2 \text{FF}} \sigma_{\Delta \lambda}. \] (A.11)

For an SNR of 20 and a FF of 1, \( \sigma_{\text{FF}} = 4 \times 10^{-5} \). For all practical purposes we can consider our estimated FFs to be exact.
References

[1] Abbott B et al 2009 Rep. Prog. Phys. 72 076901
[2] Harry G M 2010 (LIGO scientific collaboration) Class. Quantum Grav. 27 084006
[3] Acernese F et al 2008 Class. Quantum Grav. 25 184001
[4] Accadia T et al 2011 Class. Quantum Grav. 28 114002
[5] Accadia T et al 2012 J. Instrum. 7 P03012
[6] www.cascina.virgo.infn.it/advirgo
[7] Abadie J, Abbott B P, Abbott R et al 2010 Class. Quantum Grav. 27 173001
[8] Burgay M, D’Amico N, Possenti A et al 2003 Nature 426 531
[9] Agathos M, Del Pozzo W, Li T G F et al 2014 Phys. Rev. D 89 082001
[10] Del Pozzo W, Veitch J and Vecchio A 2011 Phys. Rev. D 83 082002
[11] Cornish N, Sampson L, Yunes N and Pretorius F 2011 Phys. Rev. D 84 062003
[12] Sampson L, Cornish N and Yunes N 2013 Phys. Rev. D 87 102001
[13] Li T G F, Del Pozzo W, Vitale S et al 2012 J. Phys. Conf. Ser. 363 012028
[14] Li T G F, Del Pozzo W, Vitale S et al 2012 Phys. Rev. D 85 082003
[15] Vallisneri M 2012 Phys. Rev. D 86 082001
[16] Buonanno A, Iyer B R, Ochsner E, Pan Y and Sathyaprakash B S 2009 Phys. Rev. D 80 084043
[17] Veitch J et al in preparation
[18] Skilling J 2004 Am. Inst. Phys. Conf. Ser. 735 395
[19] Yunes N and Pretorius F 2009 Phys. Rev. D 80 122003
[20] Vallisneri M and Yunes N 2013 Phys. Rev. D 87 102002
[21] Vitale S and Del Pozzo W 2014 Phys. Rev. D 89 022002
[22] Mandel I, Berry C, Ohme F, Fairhurst S and Farr W 2014 Class. Quantum Grav. 31 155005
[23] Mishra C K, Arun K G, Iyer B R and Sathyaprakash B S 2010 Phys. Rev. D 82 064010
[24] Veitch J, Mandel I, Aylott B, Farr B, Raymond V, Rodriguez C, van der Sluys M, Kalogera V and Vecchio A 2012 Phys. Rev. D 85 104045
[25] Advanced LIGO anticipated sensitivity curves https://dcc.ligo.org/cgi-bin/DocDB/ShowDocument?docid=2974