Application of asymptotic expansions for maximum likelihood estimators errors to gravitational waves from binary mergers: The single interferometer case

M. Zanolin,[1] S. Vitale,[1,3] and N. Makris[2]

[1]*Embry-Riddle Aeronautical University, 3700 Willow Creek Road, Prescott, Arizona, 86301, USA*
[2]*Massachusetts Institute of Technology, 77 Mass Ave, Cambridge, Massachusetts, 02139, USA*
[3]*LPTMC–Université Pierre-et-Marie-Curie, 4 Place Jussieu, 75005 Paris, France*
(Received 23 February 2010; published 25 June 2010)

In this paper we apply to gravitational waves (GW) from the inspiral phase of binary systems a recently derived frequentist methodology to calculate analytically the error for a maximum likelihood estimate of physical parameters. We use expansions of the covariance and the bias of a maximum likelihood estimate in terms of inverse powers of the signal-to-noise ration (SNR)s where the square root of the first order in the covariance expansion is the Cramer Rao lower bound (CRLB). We evaluate the expansions, for the first time, for GW signals in noises of GW interferometers. The examples are limited to a single, optimally oriented, interferometer. We also compare the error estimates using the first two orders of the expansions with existing numerical Monte Carlo simulations. The first two orders of the covariance allow us to get error predictions closer to what is observed in numerical simulations than the CRLB. The methodology also predicts a necessary SNR to approximate the error with the CRLB and provides new insight on the relationship between waveform properties, SNR, dimension of the parameter space and estimation errors. For example the timing match filtering can achieve the CRLB only if the SNR is larger than the Kurtosis of the gravitational wave spectrum and the necessary SNR is much larger if other physical parameters are also unknown.

I. INTRODUCTION

The ground-based gravitational waves detectors LIGO, Virgo, and GEO 600 [1–3] are rapidly improving in sensitivity. By 2015, advanced versions of these detectors should be taking data with a design sensitivity approximately 10 times greater than the previous generation, and the probed volume will grow by a factor of about a thousand. Such improvements in detector sensitivity mean that the first gravitational-wave signature of a compact-binary coalescence (CBC) could be detected in the next few years (see, for example, [4]).

Among the expected signals, a special role is covered by inspiralling compact binaries. This follows from the ability to model the phase and amplitude of the signals quite accurately and consequently to maximize the signal-to-noise ratio (SNR) by using matched filtering techniques. Matched filters also provide a maximum likelihood estimation (MLE) of the waveform parameters such as component masses or time of coalescence. The choice of the MLEs as reference estimators is also motivated by the fact that if an unbiased estimator that attains the CRLB exists, is the MLE [5].

The existing GW frequentist literature [5–18] evaluates the MLE accuracy in two ways: (a) analytically by calculating the so-called Fisher information matrix (FIM), or equivalently the Cramer Rao lower bound (CRLB), which is the square root of the diagonal elements of the inverse FIM, and (b) numerically by performing Monte Carlo simulations. The FIM was derived analytically in [10,11,13] using Newtonian waveforms, extended to

second-order post-Newtonian [14,15] and recently revisited up to 3.5 PN [6,7].

In [6,7] the authors calculate the CRLB for the three *standard* binary systems (NNS, NBH, BBH), and show how the errors change when the successive different PN orders are taken into account. They consider initial LIGO, advanced LIGO and Virgo noises. They also considers PN corrections to the *amplitude*. Monte Carlo simulations were performed, for example, in [8,9], for the lowest PN orders, where it is also suggested that the inclusion of the higher PN orders would be computationally expensive. More recent Monte Carlo simulations with 3.5 PN waveforms are described in [18]. We did not try to compare the uncertainties derived here to other existing papers (especially those from the 1990s) since different parameter sets, noise spectra (even for the same antenna) and PN terms were used. For example, in [12] a comparison between the CRLB and other bounds is done for a waveform at the 0th PN order. This work also uses different conventions on the waveform spectrum than more recent literature. In [13] phasing is extended to the 1.5 PN order. The spin parameters are taken in account but the noise spectrum for LIGO is different than the currently used design noise curves. In [8,14,15] the 2 PN wave was used. In the work [14] interesting observations are made about the fluctuation of the parameters variance with the PN orders, analyzing both the case of spin and spinless systems. The fluctuations of the variance in the spinless case is also stressed in [6].

The CRLB is a convenient tool to approximate the accuracies in large SNRs and to obtain error bounds for unbiased estimators. Unfortunately, for low SNRs (below

20) where the first detections might emerge, the CRLB can grossly underestimate the errors [8,12,19,20]. The reason is that with nonlinear data models and (or) with non-Gaussian noises, the CRLB depends only on the curvature of the likelihood function around the true value of the parameter.

In this paper we apply a recently derived analytical tool [21–24]to better predict a MLE accuracy and to establish necessary conditions on the SNR for the MLE error to attain the CRLB. Explicitly, within the frequentist framework, for arbitrary probability distribution of the noise, expansions of the bias and the covariance of a MLE in inverse powers of the SNR are discussed. The first order of the expansion of the variance is the inverse of the FIM. By requiring that the second-order covariance is smaller, or much smaller, than the first order, this approach predicts a necessary SNR to approximate the error with the CRLB. The higher order of the expansions are determinant in the low SNR regime where the inverse FIM underestimates the error.

We compared the errors computed using the first two orders of the expansions to the Monte Carlo simulations in [18]. We observed the first two orders of the expansions provide error predictions closer than the CRLB to what is observed in the numerical simulations. In [18] the simulations are related to the FIM to establish ranges of SNR where the CRLB describes the error. Our expansions predict the same SNR range of validity for the CRLB. The expansions are sensitive to the side lobes of the likelihood function because they make use of higher order derivatives than the second one (which is only sensitive to the curvature of the main lobe). The methodology also provides new insight on the relationship between waveform properties, SNR, dimension of the parameter space, and estimation errors. For example, as noted in [22] for an arrival-time matched-filter estimate in white noise, the accuracy achieves the CRLB only if the SNR is larger than the Kurtosis of the gravitational wave spectrum. Here we also notice that the necessary SNR is much larger if the other physical parameters are unknown. More specifically, the MLE of the arrival time for NS-NS binary signals might require a SNR equal to 2 with the time as the only parameter or 15 when all the other parameters are unknown. These results are important to understand the domain of validity of recent papers like [16] that defines 90% confidence regions in direction reconstruction with time triangulation. The regions discussed in [16] for SNR smaller than 10 are based on timing MLEs, with the arrival time being the only unknown parameter, and the time uncertainties quantified by the CRLB.

We also note that [25], using a formalism introduced in [19,20], describes a consistency criterion, different from the condition derived in this paper, for the validity of the CRLB that, if applied to a 4 pp compact-binary signal computed with a 2 PN expansion and $m_1 = m_2 = 10 M_\odot$,

requires a SNR of at least 10. At the time of the writing of this paper, we established with M. Vallisneri that the Eq. (3.19) of this paper becomes, in the one parameter case and colored Gaussian noise, equivalent to Eq. (60) in [25] or (A35) in [19]. A comparison for the Gaussian noise and multiparameter case is the object of current work, while a comparison for arbitrary noise is not possible because [19,25] use Gaussian noises from the beginning of their derivations. The explicit calculations shown here for different GWs are also not performed in [19,25].

In Sec. II we present the explicit expressions of the expansions of the bias and the covariance matrix for arbitrary noise and size of the parameter space. In Sec. III we explain how the expansion can be evaluated for signals in additive colored Gaussian noise. In Sec. IV we describe the post-Newtonian inspiral waveform used for the examples, the parameter space and the initial and advanced LIGO noises. In Sec. V we study the one-dimensional parameter space results when only one parameter at a time is considered unknown. In Sec. VI we present the results for full parameter space with the initial and advanced LIGO noises. We also compare our results with published results from Monte Carlo simulations. In Sec. VII we present some conclusions, and in the appendix we describe the derivation of the expansions as well as the relationship of this method with the statistics literature.

II. EXPANSIONS FOR THE BIAS AND COVARIANCE MATRIX OF A FREQUENTIST MLE IN ARBITRARY NOISE

In this section we present the first two orders of the expansions in inverse powers of the SNR for the bias and the covariance matrix. The details of the derivation are provided in Appendix A. Given an experimental data vector $\underline{x} = \{x_1, \ldots, x_N\}$, where N is the dimension, we assume that the data are described by a probability density $P(\underline{x}, \underline{\vartheta})$ that depends on a D-dimensional parameter vector $\underline{\vartheta} = \{\vartheta_1, \ldots, \vartheta_D\}$. According to [26], we suppose that the MLE $\hat{\vartheta} = \{\hat{\vartheta}_1, \ldots, \hat{\vartheta}_D\}$ of $\underline{\vartheta}$ is given by a stationary point of the likelihood function $l(\underline{x}, \underline{\vartheta}) = \ln(P(\underline{x}, \underline{\vartheta}))$ with respect to the components of $\underline{\vartheta}$

$$l_r(\underline{x}, \hat{\vartheta}) = \frac{\partial l(\underline{x}, \underline{\vartheta})}{\partial \vartheta_r}\bigg|_{\underline{\vartheta} = \hat{\vartheta}} = 0 \quad r = 1, \ldots, D. \quad (2.1)$$

If we introduce the notations

$$l_{a_1 a_2 \ldots a_s} = l_{a_1 a_2 \ldots a_s}(\underline{x}, \underline{\vartheta}) = \frac{\partial^s l(\underline{x}, \underline{\vartheta})}{\partial \vartheta_{a_1} \partial \vartheta_{a_2} \ldots \partial \vartheta_{a_s}}$$

$$v_{a_1 a_2 \ldots a_s, \ldots, b_1 b_2 \ldots b_s} = E[l_{a_1 a_2 \ldots a_s} \ldots l_{b_1 b_2 \ldots b_s}]$$

where $-v_{ab}$ is the Fisher information matrix $i_{ab} = -v_{ab} = -E[l_{ab}] = E[l_a l_b]$ ($E[.]$ is the expected value), the first two orders of the bias for the MLE of the r component of the parameter vector $\underline{\vartheta}$ are given by

$$b_1(\hat{\vartheta}^m) = \tfrac{1}{2} i^{ma} i^{bc} (v_{abc} + 2v_{c,ab}) \quad (2.2)$$

$$b_2(\hat{\vartheta}^m) = -\frac{i^{ma}i^{bc}}{2}[v_{abc} + 2v_{ab,c}] + \frac{i^{ma}i^{bd}i^{ce}}{8}[v_{abcde} + 4v_{ac,bde} + 8v_{de,abc} + 4v_{abce,d} + 4v_{abc,d,e} + 8v_{ab,cd,e}]$$

$$+ \frac{i^{ma}i^{bc}i^{df}i^{eg}}{4}[(2v_{afed}v_{gb,c} + 2v_{bedf}v_{ac,g} + 4v_{abed}v_{gf,c}) + (v_{afed}v_{gcb} + +2v_{abed}v_{gcf} + 2v_{dbeg}v_{acf})$$

$$+ (2v_{aed}v_{gb,fc} + 4v_{acf}v_{dg,eb} + 4v_{bed}v_{ac,gf} + 2v_{fcb}v_{ag,ed}) + (4v_{afe,g}v_{db,c} + 4v_{afe,c}v_{db,g} + 4v_{dbe,g}v_{af,c})$$

$$+ (2v_{abe,g}v_{cdf} + 4v_{dbe,g}v_{acf} + 4v_{abe,f}v_{cfg} + 2v_{dge,b}v_{acf}) + (4v_{ag,fc}v_{ed,b} + 4v_{ed,fc}v_{ag,b} + 4v_{ag,ed}v_{fc,b})$$

$$+ (4v_{acg}v_{ef,b,d} + 2v_{cde}v_{ab,f,g}) + \frac{2}{3}v_{abde}v_{c,f,g}] + \frac{i^{ma}i^{bc}i^{de}i^{fg}i^{ti}}{8}[v_{adf}(v_{ebc}v_{gti} + 2v_{etc}v_{gbi} + 4v_{gbe}v_{tci}$$

$$+ 8v_{gbt}v_{eci} + 2v_{ebc}v_{gt,i} + 4v_{etc}v_{gb,i} + 2v_{gti}v_{eb,c} + 4v_{gtc}v_{eb,i} + 8v_{gbt}v_{ce,i} + 8v_{gbt}v_{ci,e} + 8v_{gbe}v_{ct,i}$$

$$+ 8v_{cte}v_{gb,i} + 4v_{cti}v_{gb,e} + 4v_{gt,i}v_{eb,c} + 4v_{eb,i}v_{gt,c} + 8v_{gt,b}v_{ic,e} + 8v_{gt,e}v_{ic,b} + 4v_{bet}v_{g,c,i})$$

$$+ v_{dci}(8v_{bgt}v_{ae,f} + 4v_{bgf}v_{ae,t} + 8v_{ae,t}v_{bg,f} + 8v_{ae,f}v_{bg,t} + 8v_{af,b}v_{ge,t})] \tag{2.3}$$

were we assumed the Einstein convention to sum over repeated indices. For the covariance matrix the first order is the inverse of the Fisher information matrix while the second order is given in by (for simplicity we provide the diagonal terms):

$$C_2(\vartheta^j) = -i^{jj} + i^{jm}i^{jn}i^{pq}(2v_{nq,m,p} + v_{nmpq} + 3v_{nq,pm}$$

$$+ 2v_{nmp,q} + v_{mpq,n}) + i^{jm}i^{jn}i^{pz}i^{qt}[(v_{npm}$$

$$+ v_{n,mp})(v_{qzt} + 2v_{t,zq}) + v_{npq}(\tfrac{5}{2}v_{mzt} + 2v_{m,tz}$$

$$+ v_{m,t,z}) + v_{nq,z}(6v_{mpt} + 2v_{pt,m} + v_{mp,t})]. \tag{2.4}$$

III. EXPANSIONS FOR SIGNALS IN ADDITIVE COLORED GAUSSIAN NOISE

For this analysis we assume that the output of an optimally oriented GW interferometer has the form:

$$x(t) = h(t, \underline{\theta}) + w(t) \tag{3.1}$$

where $h(t, \underline{\theta})$ is the signal, which depends on the parameters vector $\underline{\theta}$, and $w(t)$ a stationary Gaussian noise with zero mean. The probability distribution can be written as

$$p(x) \propto \exp\left\{-\frac{1}{2}\int [x(t) - h(t, \underline{\theta})]\Omega(t - t_1)[x(t_1)\right.$$

$$\left. - h(t_1, \underline{\theta})]dtdt_1\right\}. \tag{3.2}$$

The first and second derivative of the log-likelihood give

$$l_a \equiv \frac{\partial \log p(x)}{\partial \theta_a}$$

$$= \int h_a(t, \underline{\theta})\Omega(t - t_1)[x(t_1) - h(t_1, \underline{\theta})]dtdt_1, \tag{3.3}$$

$$l_{ab} \equiv \frac{\partial l_a}{\partial \theta_b}$$

$$= \int [h_{ab}(t, \underline{\theta})\Omega(t - t_1)[x(t_1) - h(t_1, \underline{\theta})]$$

$$- h_a(t, \underline{\theta})\Omega(t - t_1)h_b(t_1, \underline{\theta})]dtdt_1, \tag{3.4}$$

$$i_{ab} = E[l_a l_b] = -E[l_{ab}]$$

$$= \int h_a(t, \underline{\theta})\Omega(t - t_1)h_b(t_1, \underline{\theta})dtdt_1 = \tag{3.5}$$

$$= \int df df' h_a(f)h_b(f')\Omega(-f, -f'). \tag{3.6}$$

It is easy to verify that

$$\Omega(f, f') = \frac{\delta(f + f')}{S_h(f)}, \tag{3.7}$$

where we have introduced, $S_h(f)$, the *one sided power spectral density* defined as the Fourier transform of the noise autocorrelation:

$$R(t) = E[n(t + \tau)n(\tau)]. \tag{3.8}$$

$$S_h(f) \equiv \int dt e^{-2\pi i f t} R(t). \tag{3.9}$$

Notice that the sign convention in (3.7) follows from the implicit assumption that $R(t)$ is the Fourier transform of $E[n(f)n(f')]$. In the literature, another convention with the minus sign is also found corresponding to defining $R(t)$ as the Fourier transform of $E[n(f)n^*(f')]$. Using the relation $h(-f) = h(f)^*$, we can finally write the FIM:

$$i_{ab} = E[l_a l_b] = \langle h_a(f), h_b(f)\rangle, \tag{3.10}$$

where $h_a(f)$ are the derivatives of the Fourier transform of the signal with respect to the ath parameter. We have introduced a mean in the frequency space:

$$\langle u(f), v(f)\rangle \equiv 4\mathcal{R}\left[\int_{f_{\text{low}}}^{f_{\text{cut}}} df \frac{u(f)v(f)*}{S_h(f)}\right], \tag{3.11}$$

where the range of integration depends on the antenna properties and the theoretical model for the binary system.

The SNR corresponds to the optimal filter:

$$\rho^2 \equiv \langle h(f), h(f) \rangle = 4 \int_{f_{low}}^{f_{cut}} df \frac{|h(f)|^2}{S_h(f)}.$$

We can express in the same way all the quantities we need in order to calculate the second-order variance, like scalar products of $h(f)$ derivatives.

$$v_{a,b} = -v_{ab} = i_{ab} = \langle h_a, h_b \rangle \qquad (3.12)$$

$$v_{ab,c} = \langle h_{ab}, h_c \rangle \qquad (3.13)$$

$$v_{abc} = -\langle h_{ab}, h_c \rangle - \langle h_{ac}, h_b \rangle - \langle h_{bc}, h_a \rangle \qquad (3.14)$$

$$v_{ab,cd} = \langle h_{ab}, h_{cd} \rangle + \langle h_a, h_b \rangle \langle h_c, h_d \rangle \qquad (3.15)$$

$$v_{abc,d} = \langle h_{abc}, h_d \rangle \qquad (3.16)$$

$$v_{abcd} = -\langle h_{ab}, h_{cd} \rangle - \langle h_{ac}, h_{bd} \rangle - \langle h_{ad}, h_{bc} \rangle - \langle h_{abc}, h_d \rangle$$
$$- \langle h_{abd}, h_c \rangle - \langle h_{acd}, h_b \rangle - \langle h_{bcd}, h_a \rangle \qquad (3.17)$$

$$v_{ab,c,d} = \langle h_a, h_b \rangle \langle h_c, h_d \rangle = i_{ab} i_{cd}. \qquad (3.18)$$

If one uses these functions, the form for the second-order variance, Eq. (2.4) can be further simplified:

$$C_2(\vartheta^j) = i^{jm} i^{jn} i^{pq} (v_{nmpq} + 3\langle h_{nq}, h_{pm} \rangle + 2v_{nmp,q}$$
$$+ v_{mpq,n}) + i^{jm} i^{jn} i^{pz} i^{qt} (v_{npm} v_{qzt} + \tfrac{5}{2} v_{npq} v_{mzt}$$
$$+ 2v_{qz,n} v_{mtp} + 2v_{qp,z} v_{nmt} + 6v_{mqp} v_{nt,z}$$
$$+ v_{pqz} v_{nt,m} + 2v_{mq,z} v_{pt,n} + 2v_{pt,z} v_{mq,n}$$
$$+ v_{mz,t} v_{nq,p}). \qquad (3.19)$$

IV. INSPIRAL PHASE WAVEFORM FOR BINARY SYSTEMS

We apply the general results of the theory to the case of a binary system in the inspiral phase. Starting from the 3.5 PN phasing formula [27] we write the Fourier transform of the chirp signal:

$$h(t) = a(t)[e^{i\phi(t)} + e^{-i\phi(t)}], \qquad (4.1)$$

where $\phi(f)$ is the implicit solution of the 3.5 PN phasing formula, using the *stationary phase approximation* (SPA) [28,29]. The final result is

$$h(f) = \mathcal{A} f^{-(7/6)} e^{i\psi(f)}. \qquad (4.2)$$

The phase is given by

$$\psi(f) = 2\pi f t - \phi - \frac{\pi}{4} + \frac{3}{128 \eta v^5} \sum_{k=0}^{N} \alpha_k v^k, \qquad (4.3)$$

where t and ϕ are the arrival time and the arrival phase. The function v can be defined either in terms of the total mass of the binary system, $M = m_1 + m_2$, or in terms of the chirp mass, $\mathcal{M} = \eta^{3/5} M$:

$$v = (\pi M f)^{1/3} = (\pi f \mathcal{M})^{1/3} \eta^{-(1/5)},$$

where η is the symmetrized mass ratio

$$\eta = \frac{m_1 m_2}{M^2}.$$

The amplitude is a function of the chirp mass, the effective distance, and the orientation of the source:

$$\mathcal{A} \propto \mathcal{M}^{5/6} Q(\text{angles})/D.$$

The coefficients α_k's with $k = 0 \dots N$ (the meaning of each terms being the $\frac{k}{2}$ PN contribution) are given by

$$\alpha_0 = 1, \qquad \alpha_1 = 0, \qquad \alpha_2 = \frac{20}{9}\left(\frac{743}{336} + \frac{11}{4}\eta\right), \qquad \alpha_3 = -16\pi, \qquad \alpha_4 = 10\left(\frac{3\,058\,673}{1\,016\,064} + \frac{5429}{1008}\eta + \frac{617}{144}\eta^2\right)$$

$$\alpha_5 = \pi\left(\frac{38\,645}{756} + \frac{38\,645}{252}\log\frac{v}{v_{lso}} - \frac{65}{9}\eta\left[1 + 3\log\frac{v}{v_{lso}}\right]\right)$$

$$\alpha_6 = \left(\frac{11\,583\,231\,236\,531}{4\,694\,215\,680} - \frac{640}{3}\pi^2 - \frac{6848}{21}\gamma\right) + \eta\left(-\frac{15\,335\,597\,827}{3\,048\,192} + \frac{2255}{12}\pi^2 - \frac{1760}{3}\theta + \frac{12\,320}{9}\lambda\right)$$
$$+ \frac{76\,055}{1728}\eta^2 - \frac{127\,825}{1296}\eta^3 - \frac{6858}{21}\log 4v,$$

$$\alpha_7 = \pi\left(\frac{77\,096\,675}{254\,016} + \frac{378\,515}{1512}\eta - \frac{74\,045}{756}\eta^2\right),$$

where $\lambda \simeq -0.6451$, $\theta \simeq -1.28$. γ is the Euler constant, and $v_{lso} = (\pi M f_{lso})^{1/3}$, with f_{lso} the *last stable orbit* frequency for a test mass in a Scharzschild space-time of mass M:

$$f_{lso} = (6^{3/2}\pi M)^{-1} \qquad (4.4)$$

which will be also used as the upper cutoff for the integrals (3.11), (3.12), (3.13), (3.14), (3.15), (3.16), (3.17), and (3.18).

Given the waveform (4.2), one can easily calculate the Fisher information matrix, and its inverse, the CRLB. Equation (4.2) contains five unknown parameters,

$(\mathcal{A}, t, \phi, \mathcal{M}, \eta)$ (the total mass M could be used instead of the chirp mass), so that one should calculate a five dimensional square matrix.

It was already observed by [13] that the errors in the distance, and consequently the amplitude $\mathcal{A}$, are uncorrelated with the errors in other parameters, i.e. that the Fisher information is block diagonal. We observed that this is also the case for the waveform (4.2) we use here for both the FIM and the second-order covariance matrix. We can therefore work in a four-dimensional parameter space $(t, \phi, \mathcal{M}, \eta)$. However, it is worth observing that this is not an obvious point since in general the amplitude estimation can be coupled with other parameters if they enter in the signal in specific ways (see Ch. 3 of [5]).

For the analysis of this paper, the SNR and the amplitude $\mathcal{A}$ are related like follows:

$$\rho^2 \equiv \langle h(f), \qquad h(f) \rangle = 4\mathcal{A}^2 \int_{f_{\text{low}}}^{f_{\text{cut}}} df \frac{f^{-(7/3)}}{S_h(f)}. \quad (4.5)$$

We perform the calculations using analytical forms of the design initial and advanced LIGO noise spectrum (4.6). The initial one-sided power spectral density of LIGO can be written for $f \geq f_{\text{low}}$ ($S_h(f) = \infty$, $f \leq f_{\text{low}}$):

$$S_h(f)) = S_0[(4.49x)^{-56} + 0.16x^{-4.52} + 0.52 + 0.32x^2], \quad (4.6)$$

where the lower frequency cutoff value is $f_{\text{low}} = 40$ Hz, $x \equiv \frac{f}{f_0}$, $f_0 = 500$ Hz, and $S_0 = 9 \times 10^{-46}$ Hz^{-1}.

The Advanced LIGO one sided psd has the following expression, for $f \geq f_{\text{low}}$ ($S_h(f) = \infty$, $f \leq f_{\text{low}}$):

$$S_h(f) = S_0\left[x^{-4.14} - 5x^{-2} + 111\frac{1 - x^2 + x^4/2}{1 + x^2/2}\right], \quad (4.7)$$

where the lower frequency cutoff value is $f_{\text{low}} = 20$ Hz, $x \equiv \frac{f}{f_0}$, $f_0 = 215$ Hz, and $S_0 = 10^{-49}$ Hz^{-1}. They are plotted in Fig. 1.
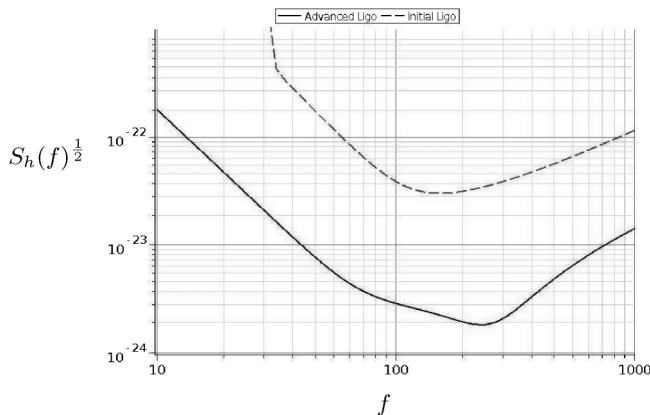

FIG. 1. The Initial (dashed line) and Advanced (solid line) LIGO noise spectrum.

We now calculate the second-order covariance for the full four-dimensional parameter space $(t, \phi, \mathcal{M}, \eta)$.

To compare our results with the literature more easily, we study a binary neutron star system (NS-NS), a neutron star—black hole system (NS-BH), and a black hole system (BH-BH). A neutron star is assumed to have a mass of $1.4M_\odot$ and a black hole of $10M_\odot$.

We performed our calculations following these steps: (I) We give to the total mass M and the mass ratio η a numerical value depending on the binary system we consider. This makes the upper cutoff (4.4) to have a numerical value. (II) We compute analytically the derivatives of the wave function $h(f)$ and use it to compute the Fisher information (3.10) and his inverse, the CRLB. (III) We calculate the v's functions (3.12), (3.13), (3.14), (3.15), (3.16), (3.17), and (3.18) and use them to compute the second-order covariance (3.19). (IV) We plot for the four parameters the CRLB, the second-order variance and the total error (without bias, see end of Sec. V).

V. ONE-DIMENSIONAL PARAMETER SPACE

In this section we describe the results for the instance where only one of the parameters (we call it θ) is considered unknown, while the others have known values. It can be shown that in this case the second-order variance (3.19) can be written as

$$C_2(\theta) = (i^{\theta\theta})^3(8i^{\theta\theta}\langle h_{\theta\theta}, h_\theta\rangle^2 - \langle h_{\theta\theta\theta}, h_\theta\rangle) \quad (5.1)$$

Let us consider the case where θ is the arrival time t in the waveform (4.2). The ratio between the second order and the first order variance turns out to be

$$\frac{C_2(t)}{\sigma_t^2} = \frac{1}{4\rho^2}\frac{\frac{K_4}{K_0}}{(\frac{K_2}{K_0})^2} \quad (5.2)$$

where $K_\alpha \equiv \int df \frac{f^\alpha}{S_h(f)}|h(f)|^2$ is the αth moment of the signal's frequency spectrum. The second order is negligible with respect to the first if

$$\rho^2 \gg \frac{\frac{K_4}{K_0}}{(\frac{K_2}{K_0})^2}, \quad (5.3)$$

that is, if the SNR is much larger than the *Kurtosis* of the signal's spectrum. This means that for two signals with the same bandwidth, the one with a more *peaked* spectrum requires higher SNR to attain the CRLB. It must be noted that the K_α are functions of the total mass via the upper limit of the integral [see (4.4)]. This kind of observation was also introduced in [22] for an arrival-time matched-filter estimate from data disturbed by additive white noise.

It can be shown that for a BH-BH (NS-NS) system the second order becomes equal to the first for $\rho = 1.32$ ($\rho = 1.97$). These values of SNR are smaller than those we will derive with the full parameters space. This indicates that it is much harder to estimate all the parameters at the same

time. Also notice that if someone requires the second order to be much smaller than the first, for example, 10%, the conditions become more stringent.

A similar calculation can be done when one of the other parameters is considered as unknown. For M and η the same analysis does not give a result that is equally compact and we only show a plot of the ratio between the second order and the first order variance, for a fixed value of SNR, $\rho = 10$; These values, are presented in Fig. 2, for different values of η and M and 1 PN waveforms. When η is the unknown parameter (upper plot), the ratio becomes smaller when the total mass of the system increases. This is in agreement with the Monte Carlo simulations (see e.g. [18], and Sec. 8.3) performed in the full four parameter space. The necessary SNR is not very sensitive to the value of η. Similar considerations can be drawn when the total mass M is the unknown parameter (bottom plot), except that now the necessary SNR seems to be slightly more sensitive to the value of η. In both panels of Fig. 2, the second order is always about 30% of the first order. If one works with a one-dimensional parameter space more compact expressions can also be given for the first and second-order bias,

Eqs. (2.2) and (2.3):

$$b[1] = -\tfrac{1}{2}(i^{\theta\theta})^2 \langle h_{\theta\theta}, h_\theta \rangle \tag{5.4}$$

$$\begin{aligned}
b[2] = -(i^{\theta\theta})^3 \big(&\tfrac{1}{8}\langle h_{\theta\theta\theta\theta}, h_\theta \rangle + \tfrac{5}{4}\langle h_{\theta\theta\theta}, h_{\theta\theta} \rangle - \tfrac{3}{2}\langle h_{\theta\theta\theta}, h_\theta \rangle \\
& - i^{\theta\theta}\langle h_{\theta\theta\theta}, h_\theta \rangle \langle h_{\theta\theta}, h_\theta \rangle - \tfrac{9}{2}i^{\theta\theta}\langle h_{\theta\theta}, h_{\theta\theta} \rangle \langle h_{\theta\theta}, h_\theta \rangle \\
& + \tfrac{9}{8}(i^{\theta\theta})^2 \langle h_{\theta\theta}, h_\theta \rangle^3 \big)
\end{aligned} \tag{5.5}$$

We observed that the first two orders of the bias are, for all of the examples in this paper, few order of magnitude smaller than the variance contributions. Therefore, we do not include them in the presented results. Ongoing research on waveforms including the merger and ringdown phases show that the bias can also become important for systems with masses beyond the range considered here.

VI. FULL PARAMETER SPACE

We present results for the full parameter space beginning with the examples obtained using the initial LIGO noise (4.6), see Figs. 3–5 and then show the same results for the advanced LIGO noise (4.7), see Figs. 6–8. In each plot we show three curves as a function of the SNR. The dotted one is the CRLB (square root of the diagonal elements of the inverse of the FIM). The dashed one is the square root of the corresponding diagonal elements in the second-order covariance matrix, and the continuous one is the square root of the sum of the diagonal elements of the of the FIM and of the second-order covariance matrix. For all the cases analyzed in this paper, the bias resulted in a negligible contribution and, as a result, is not included in the plots.

For the bottom two of the four plots panels, the curves are divided by the actual value of the parameter in order to express the relative uncertainty. The general trend is that the CRLB fails to describe correctly the error for SNRs lower than 20. For the t and the M, this regime is of physical interest while for the symmetrized mass ratio and the phase the CRLB already predicts very large uncertainties. It is also worth noticing that the SNR at which the second-order covariance matrix contribution becomes comparable to the CRLB is much larger when the full parameter space is involved. For t for example in the NS-NS case the two are the same at $\rho = 2$ while for the full parameter space they equate around $\rho = 15$, see the first panel of Fig. 3. These results appear to indicate that also the timing based directional reconstruction accuracy is worse when the other physical parameters of the waveform are unknown.

It is interesting to compare our results with Monte Carlo simulations for 3.5 PN waveforms, [18].

In Fig. 9 we reproduce the plots of [18] obtained for a NS-NS system (first panel from top) and for a low mass BH-BH system—$10M_\odot$— (third panel from top) when the initial LIGO noise is used.
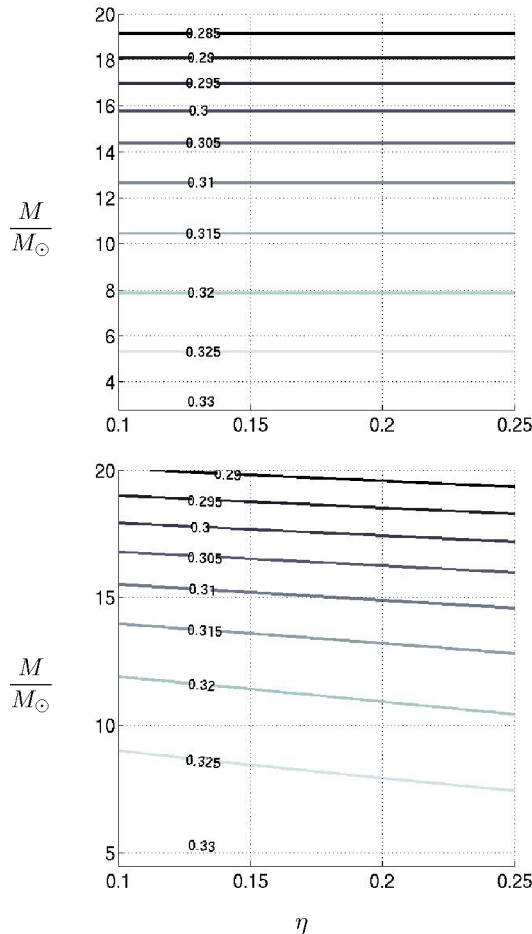

FIG. 2 (color online). The ratio $C_2(\eta)/i^{\eta\eta}$ (top) and $C_2(M)/i^{MM}$ (bottom).
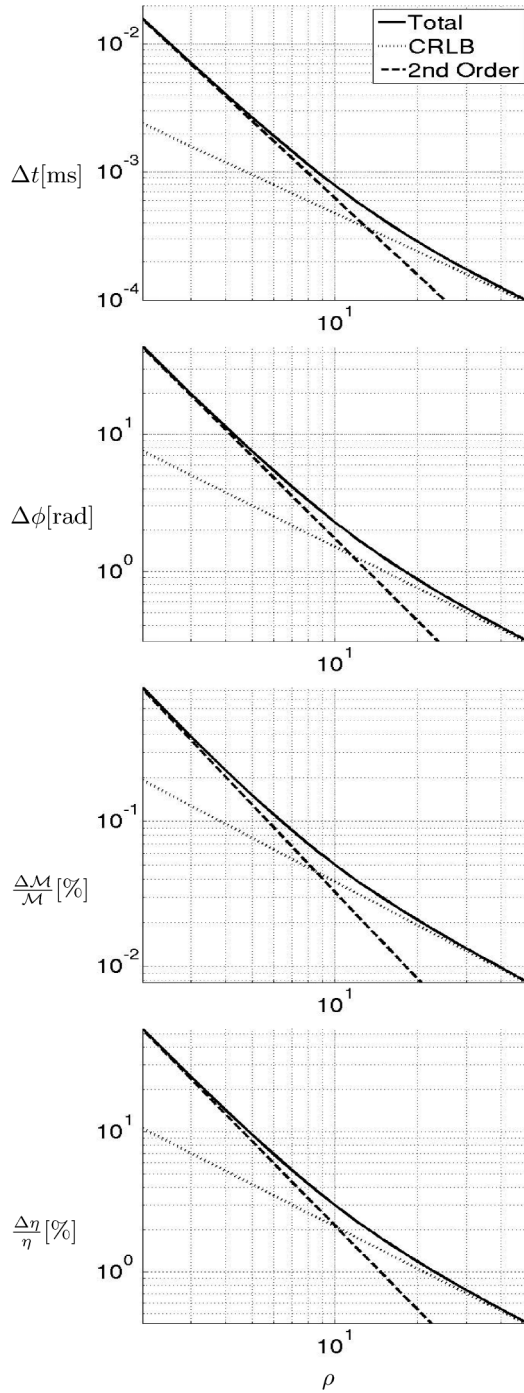
FIG. 3. NS-NS signal in initial LIGO noise. The dotted line is the CRLB. The dashed line is the square root of the second-order covariance matrix, and the continuous line is the square root of the sum of the diagonal elements of the FIM and of the second-order covariance matrix. In the last two panels the errors are divided by the value of the parameter.

FIG. 4. NS-BH signal in initial LIGO noise. The dotted line is the CRLB. The dashed line is the square root of the second-order covariance matrix, and the continuous line is the square root of the sum of the diagonal elements of the FIM and of the second-order covariance matrix. In the last two panels the errors are divided by the value of the parameter.

The results we obtain for the same systems are shown in the second and fourth panels from the top of Fig. 9, where the CRLB and the square root of the sum of the inverse 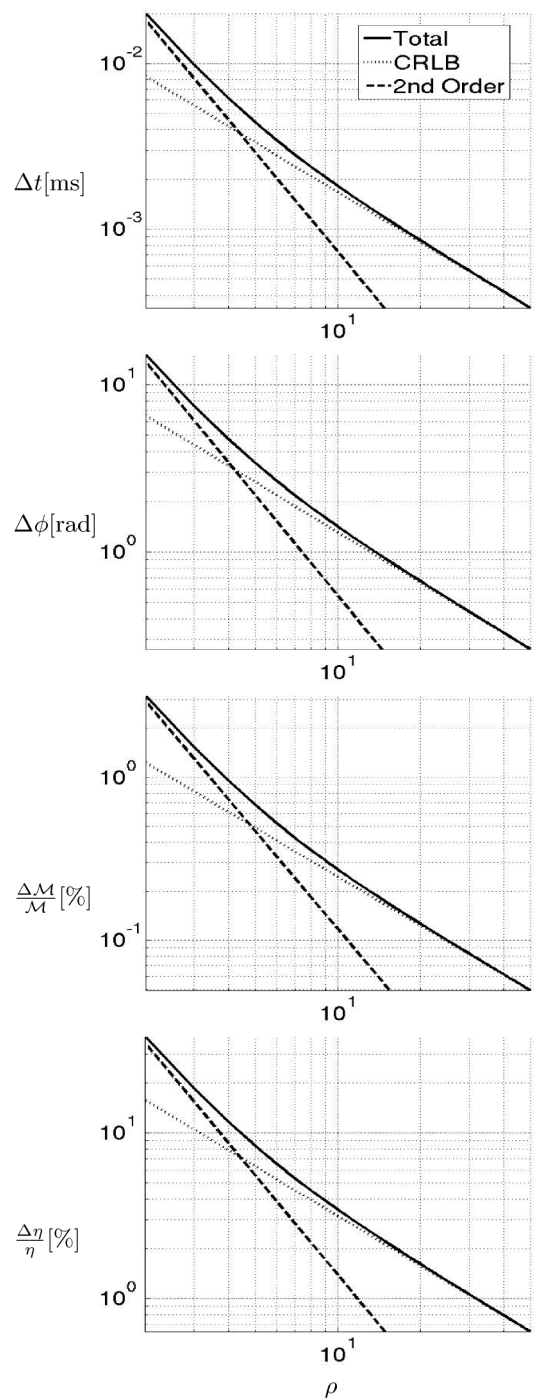FIM and the second-order covariance are plotted. The plots show that the use of the second-order covariance predicts correctly the SNR regime where the CRLB fails to describe the MLE accuracy. The explicit percentage discrepancies
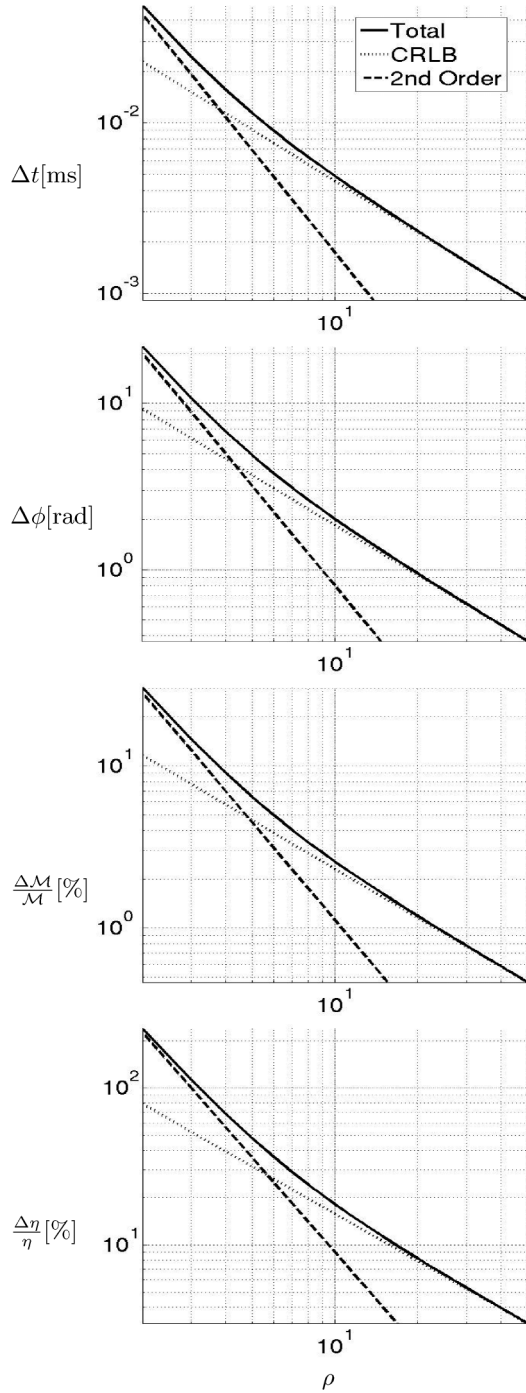
FIG. 5.   BH-BH signal in initial LIGO noise. The dotted line is the CRLB. The dashed line is the square root of the second-order covariance matrix, and the continuous line is the square root of the sum of the diagonal elements of the FIM and of the second-order covariance matrix. In the last two panels the errors are divided by the value of the parameter.

FIG. 6.   NS-NS signal with Advanced LIGO noise. The dotted line is the CRLB. The dashed line is the square root of the second-order covariance matrix, and the continuous line is the square root of the sum of the diagonal elements of the FIM and of the second-order covariance matrix. In the last two panels the errors are divided by the value of the parameter.

presented for SNR equal to 20, 10, and 6 in Table I seem to indicate that higher orders of the expansions might be necessary to fully reproduce the error derived with the Monte Carlo simulations.
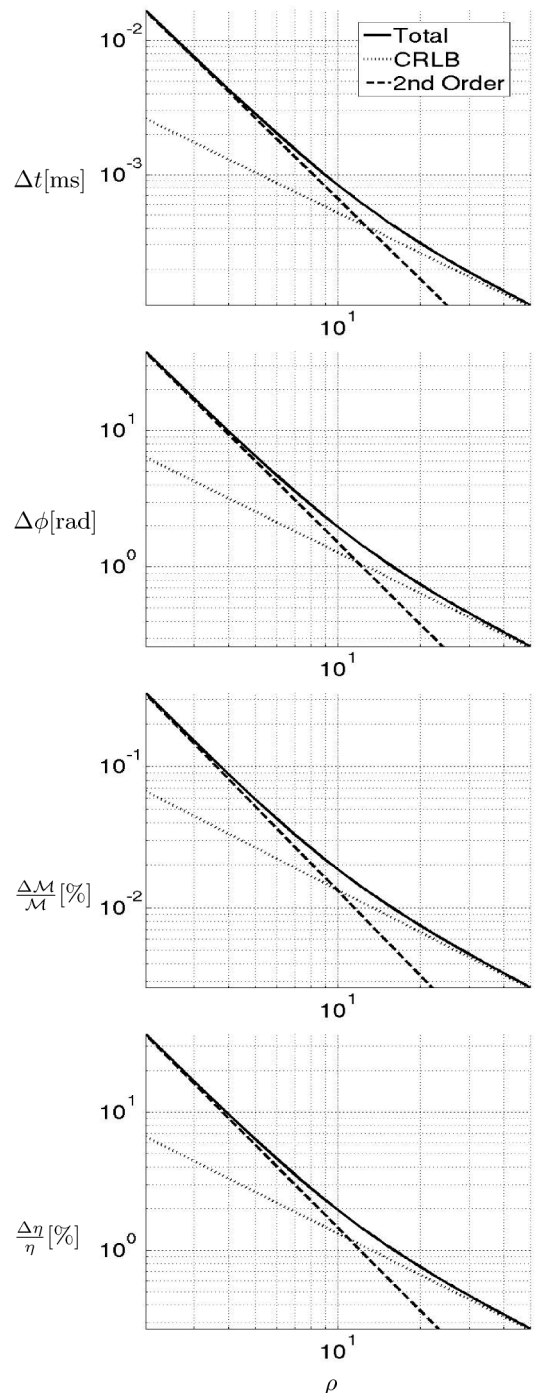
VII. CONCLUSIONS

In this paper we applied a recently derived statistical methodology to gravitational waves generated by the in-

FIG. 7.   NS-BH signal with advanced LIGO noise. The dotted line is the CRLB. The dashed line is the square root of the second-order covariance matrix, and the continuous line is the square root of the sum of the diagonal elements of the FIM and of the second-order covariance matrix. In the last two panels the errors are divided by the value of the parameter.
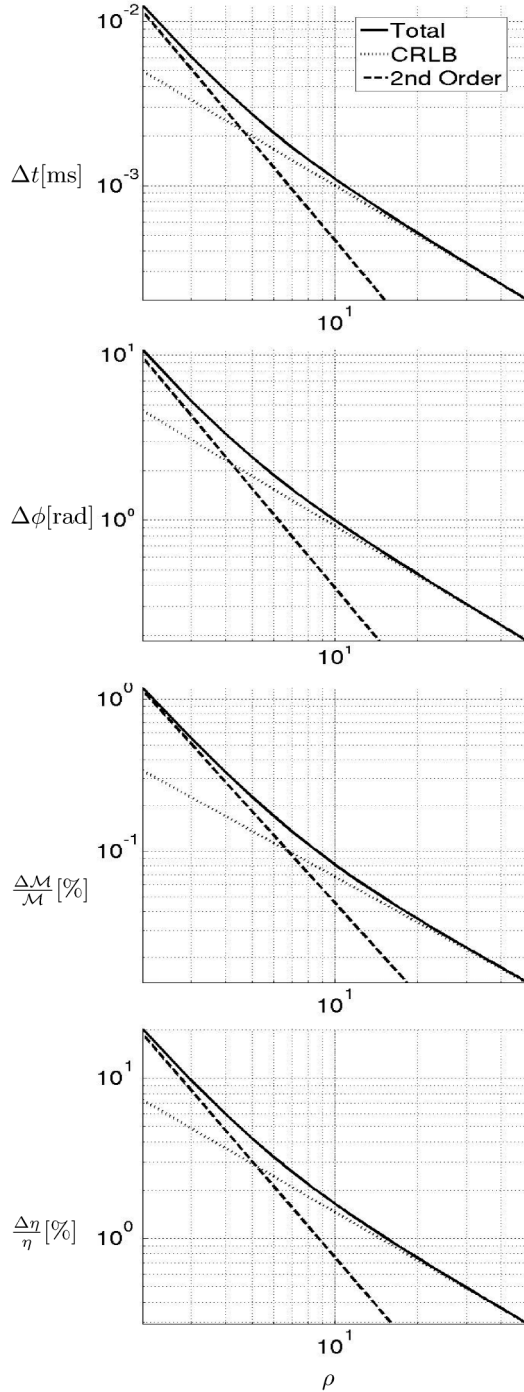

FIG. 8.   BH-BH signal with advanced LIGO noise. The dotted line is the CRLB. The dashed line is the square root of the second-order covariance matrix, and the continuous line is the square root of the sum of the diagonal elements of the FIM and of the second-order covariance matrix. In the last two panels the errors are divided by the value of the parameter.

spiral phase of binary mergers and for noise spectral densities of gravitational wave interferometers. Explicitly, we computed the first two orders of MLE expansions of bias and covariance matrix to evaluate MLE uncertainties. We
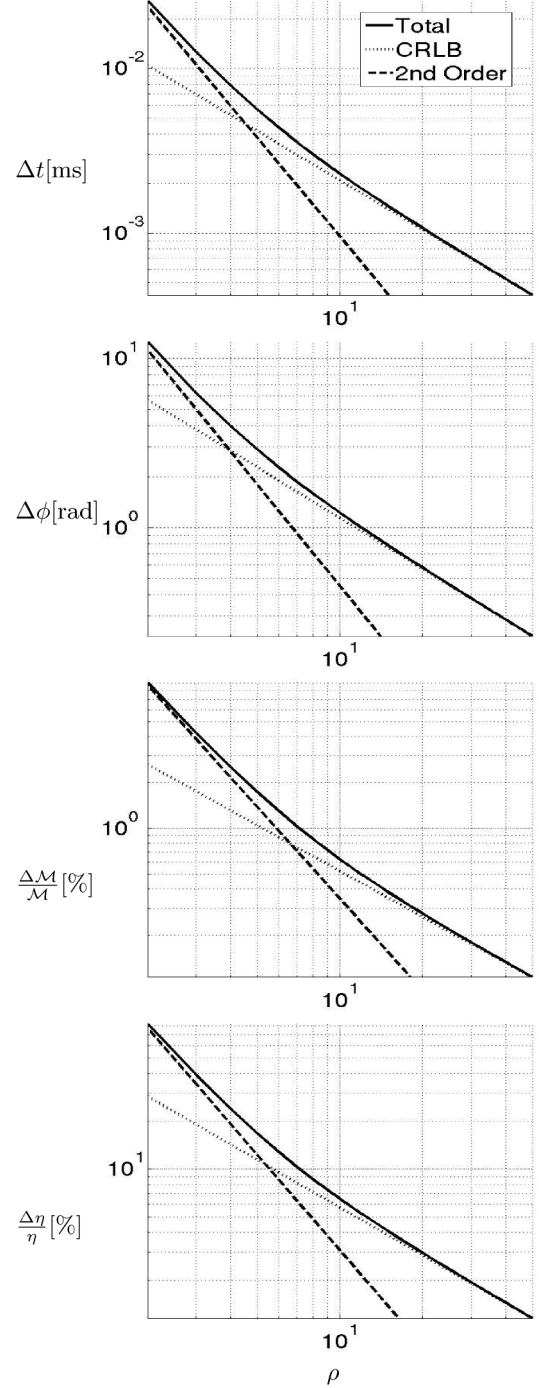
also compared the improved error estimate with existing numerical estimates. The value of the second order of the variance expansions allows us to get error predictions closer to what is observed in numerical simulations than
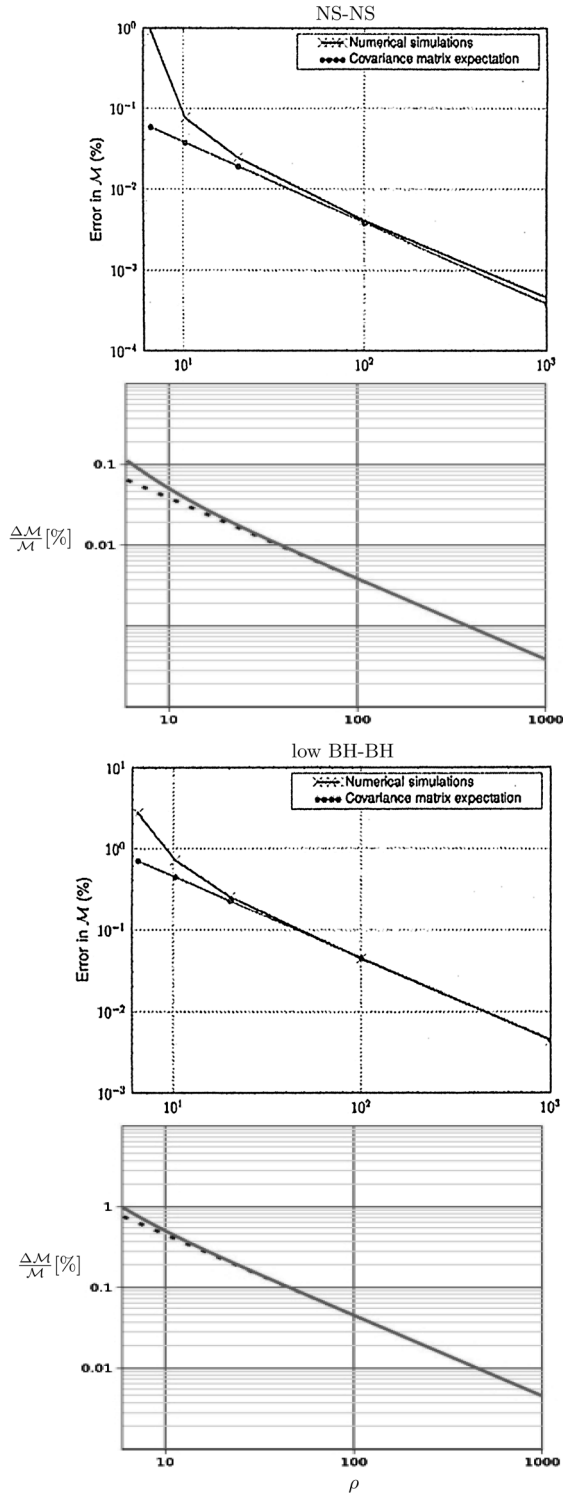
FIG. 9.   The simulations of T. Cokelaer (first and thrid panel from top), compared with our result (see the text for a discussion).

TABLE I.   The discrepancies between the CRLB error prediction and Monte Carlo simulations are presented above the discrepancies observed using the first and second-order covariance matrix. The discrepancies are presented for three values of the SNR: 20, 10, and 6.

NS-NS SNR	20	10	6
Cokelaer	25%	200%	700%
Z.V.M.	8.6%	131%	174%
BH-BH SNR	20	10	6
Cokelaer	10%	150%	230%
Z.V.M.	2.9%	111%	129%

error with the CRLB and provides new insight on the relationship between waveform properties SNRs and estimation errors. Future applications include IMR waveforms, a network of detectors, and source location estimation.

ACKNOWLEDGMENTS

APPENDIX A: DERIVATION OF THE EXPANSIONS

Analytic expressions for the moments of a MLE are often difficult to obtain given a nonlinear data model. However, it is known from [21] that likelihood expansions can be used to obtain approximate expressions for the moments of a MLE in terms of series expansions in inverse powers of the sample size (n). These expansions are valid in small samples or SNR where the MLE may have significant bias and may not attain minimum variance.

An expression for the second covariance equivalent to the one used here was first given in a longer form (about a factor 2), and with the prescription of isolating the different asymptotic orders inside the tensors, in [22]. The expressions presented here were first derived in an unpublished MIT technical report [30] by two of the authors of this paper, where (a) the Bartlett identities [31,23] were used to simplify the expression of the second-order variance, and derive the second-order bias. (b) the prescription on the tensors no longer needed to be implemented. The final form of the second-order covariance has already been published without proof, in two papers [24,32], where the first and third author of this paper are coauthors, involving MLE of source and environment parameters that use acoustic propagation within a shallow water ocean wave guide.

In this section, we derive the second-order bias for a multivariate MLE and we introduce a chain rule that allows

the inverse of the FIM. The condition where the second-order covariance is negligible with respect to the first order predicts correctly the necessary SNR to approximate the

the derivation of the second-order covariance matrix from the second-order bias. The explanation follows closely [30]. The derivation of the bias is performed in two steps: first we derive the expansions in the nonphysical scenario of n statistically independent identical measurements and then set $n = 1$ for the case of interest of this paper.

The derivation follows the approach of [26] for likelihood expansions, [33] for the large sample approximations and [34,35] for the notation of the observed likelihood expansions. For the small sample case of interest, *asymptotic* orders generated from the likelihood expansion may contain different powers of $\frac{1}{n}$, and the contributions to a single power of $\frac{1}{n}$ may have to be collected from different *asymptotic* orders. The authors of [36] extending the work of [21], confronted the equivalent problem of separating the powers in n within the expectations of products of linear forms and arrived at expressions for the second-order bias of the MLE by embedding the derivation in a discrete probability scheme. Some applications of their results are given in [37–41]. We present here an independent derivation for the second-order bias that is valid for general discrete or continuous random variables. Let us consider a set of n independent and identically distributed experimental data vectors $\underline{x}_i = \{x_{i1}, \ldots, x_{iN}\}$, where N is the dimension of every vector. We assume that these data are described by a probability density $P(\underline{x}, \underline{\vartheta}) = \Pi_i p(\underline{x}_i, \underline{\vartheta})$ that depends on a D-dimensional parameter vector $\underline{\vartheta} = \{\vartheta_1, \ldots, \vartheta_D\}$. According to [26], we suppose that the MLE $\hat{\underline{\vartheta}} = \{\hat{\vartheta}_1, \ldots, \hat{\vartheta}_D\}$ of $\underline{\vartheta}$ is given by a stationary point of the likelihood function $l(\underline{x}, \underline{\vartheta}) = \ln(P(\underline{x}, \underline{\vartheta})) = \sum_{i=1}^{n} \ln(p(\underline{x}_i, \underline{\vartheta}))$ with respect to the components of $\underline{\vartheta}$

$$l_r(\underline{x}, \hat{\underline{\vartheta}}) = \frac{\partial l(\underline{x}, \underline{\vartheta})}{\partial \vartheta_r} \bigg|_{\underline{\vartheta} = \hat{\underline{\vartheta}}} = 0 \quad r = 1, \ldots, D. \quad (A1)$$

The first step in deriving the likelihood expansion, if $l_r(\underline{x}, \hat{\underline{\vartheta}})$ can be expanded as a Taylor series in the components of the observed error $\hat{\underline{\vartheta}} - \underline{\vartheta}$, consists of writing $l_r(\underline{x}, \hat{\underline{\vartheta}})$ as

$$\begin{aligned} 0 = l_r(\underline{x}, \hat{\underline{\vartheta}}) &= l_r(\underline{x}, (\hat{\underline{\vartheta}} - \underline{\vartheta}) + \underline{\vartheta}) \\ &= l_r(\underline{x}, \underline{\vartheta}) + l_{rs}(\underline{x}, \underline{\vartheta})(\hat{\vartheta} - \vartheta)^s \\ &\quad + \tfrac{1}{2} l_{rtu}(\underline{x}, \underline{\vartheta})(\hat{\vartheta} - \vartheta)^t (\hat{\vartheta} - \vartheta)^u + \ldots, \quad (A2) \end{aligned}$$

where $(\hat{\vartheta} - \vartheta)^r$ for $r = 1, \ldots, D$ are the components of the observed error. We will use the notation

$$v_{\{a_1 a_2 \ldots a_s\} \ldots \{b_1 b_2 \ldots b_s\}} = E[H_{a_1 a_2 \ldots a_s} \ldots H_{b_1 b_2 \ldots b_s}],$$

where $H_{a_1 a_2 \ldots a_s} = l_{a_1 a_2 \ldots a_s} - v_{a_1 a_2 \ldots a_s}$ and $-v_{ab}$ is the information matrix $i_{ab} = -v_{ab} = -E[l_{ab}] = E[l_a l_b]$. Introducing j^{ab} as the inverse of the matrix whose elements are given by $j_{ab} = -l_{ab}$, Eq. (A2) can be rearranged to solve for $(\hat{\vartheta}_r - \vartheta_r) = (\hat{\vartheta} - \vartheta)^r$

$$\begin{aligned} (\hat{\vartheta} - \vartheta)^r &= j^{rs} l_s + \tfrac{1}{2} j^{rs} l_{stu} (\hat{\vartheta} - \vartheta)^t (\hat{\vartheta} - \vartheta)^u \\ &\quad + \tfrac{1}{6} j^{rs} l_{stuv} (\hat{\vartheta} - \vartheta)^t (\hat{\vartheta} - \vartheta)^u (\hat{\vartheta} - \vartheta)^v \\ &\quad + \tfrac{1}{24} j^{rs} l_{stuvw} (\hat{\vartheta} - \vartheta)^t (\hat{\vartheta} - \vartheta)^u (\hat{\vartheta} - \vartheta)^v \\ &\quad \times (\hat{\vartheta} - \vartheta)^w + \ldots. \quad (A3) \end{aligned}$$

Finally we can iterate Eq. (A3) with respect to the components of the observed error and expand j^{ab} in terms of the information matrix inverse $i^{ab} = (i^{-1})_{ab}$ ([42], page 149)

$$\begin{aligned} [j^{-1}]_{ab} = j^{ab} &= [(1 - i^{-1}(i - j))^{-1} i^{-1}]^{ab} \\ &= i^{ab} + i^{at} i^{bu} H_{tu} + i^{at} i^{bu} i^{vw} H_{tv} H_{uw} \\ &\quad + i^{at} i^{bu} i^{vw} i^{cd} H_{tv} H_{wc} H_{du} + \ldots. \quad (A4) \end{aligned}$$

From (A3) and (A4) the terms that contribute to each *asymptotic* order of the observed error can be obtained. However, in order to isolate the *asymptotic* orders necessary to compute the second-order bias we have chosen only a finite number of terms within Eqs. (A3) and (A4). This choice can be made by (a) observing that $v_{a_1 a_2 \ldots a_s}$ is proportional to n

$$v_{a_1 a_2 \ldots a_s} = E[l_{a_1 a_2 \ldots a_s}] = n E \left[\frac{\partial^s ln(p(\underline{x}_i, \underline{\vartheta}))}{\partial \vartheta_{a_1} \partial \vartheta_{a_2} \ldots \partial \vartheta_{a_s}} \right], \quad (A5)$$

where the value of i is irrelevant because all the vector data have the same distribution, and (b) using the large sample approximation expressed by ([33], page 221)

$$H_{a_1 a_2 \ldots a_s} = l_{a_1 a_2 \ldots a_s} - E[l_{a_1 a_2 \ldots a_s}] \sim n^{1/2}, \quad (A6)$$

and proved for discrete and continuous random variables in the next paragraph.

Equation (A6) indicates that the expectation of a product of c generic factors $H_{a_1 a_2 \ldots a_s}$ is a polynomial of integer powers of n, where the highest power of n is the largest integer less than or equal to $\frac{c}{2}$ (we use the notation int($\frac{c}{2}$)). For example, $v_{\{a_1 a_2 \ldots a_s\}\{b_1 b_2 \ldots b_p\}\{c_1 c_2 \ldots c_q\}}$ is proportional to n.

The proof of the large sample approximation (A6) is an extension to continuous random variables of the analysis performed in ([43], page 36), for discrete random variables. To prove Eq. (A6) we show that the quantities

$$P_m(n) = E[H_{\underline{\alpha}_1} H_{\underline{\alpha}_2} \ldots H_{\underline{\alpha}_m}] \quad (A7)$$

obtained as the expectation of products of m functions $H_{\underline{\alpha}_j} = \sum_{i=1}^{n} h_{i, \underline{\alpha}_j}$ are polynomials in integer powers of n of order less than or equal to $\frac{m}{2}$. Here, $h_{i, \underline{\alpha}_j} = l_{i, \underline{\alpha}_j} - \langle l_{i, \underline{\alpha}_j} \rangle$. The subscripts $\underline{\alpha}_i$ that appear in $H_{\underline{\alpha}_i}$ represent collections of indexes as introduced in Eq. (A6).

The factors $H_{\underline{\alpha}_1} H_{\underline{\alpha}_2} \ldots H_{\underline{\alpha}_m}$ appearing in the expectation can be seen as the first m terms of a succession $H_{\underline{\alpha}_i}$, where we choose arbitrary $H_{\underline{\alpha}_i}$ for $i > m$. Using such a succession we can define the quantity

$$F(y) = \sum_{s=0}^{\infty} y^s P_s(n)$$

$$= \sum_{s=0}^{\infty} y^s E\left[\left(\sum_{j_1=1}^{n} h_{j_1,\underline{\alpha}_1}\right) \cdots \left(\sum_{j_s=1}^{n} h_{j_s,\underline{\alpha}_s}\right)\right],$$

where we assume that all the expectations exist.

Moving all the sums outside of the expectations and redefining the notation in the sums, we can obtain

$$F(y) = \sum_{s=0}^{\infty} y^s \sum_{\substack{a_{1,\underline{\alpha}_1}+\ldots+a_{n,\underline{\alpha}_1}+\ldots+a_{1,\underline{\alpha}_s}+\ldots+a_{n,\underline{\alpha}_s}=s}}^{a_{p,\underline{\alpha}_q}=0,1}$$
$$\times E[h_{1,\underline{\alpha}_1}{}^{a_{1,\underline{\alpha}_1}} \ldots h_{n,\underline{\alpha}_1}{}^{a_{n,\underline{\alpha}_1}} \ldots h_{1,\underline{\alpha}_s}{}^{a_{1,\underline{\alpha}_s}} \ldots h_{n,\underline{\alpha}_s}{}^{a_{n,\underline{\alpha}_s}}],$$
$$(A8)$$

where the quantity $a_{p,\underline{\alpha}_q}$ can be equal to 0 or to 1 to indicate that the quantity $h_{p,\underline{\alpha}_q}$ is present or not in the product inside the expectation. The summation over all $a_{i,\underline{\alpha}_j}$ is intended to account for all possible configurations of the indexes (choosing some equal to zero and the others equal to one) with the constraint that the number of factors within the expectation is equal to s. We can now group together terms $(h_{i,\underline{\alpha}_j})^{a_{i,\underline{\alpha}_j}}$ for different i and define $\sum_{j=1}^{s} a_{i,\underline{\alpha}_j} = s_i$, where $\sum_{i=1}^{n} s_i = s$. In this manner we can also factorize the expectations in (A8) as expectations for different data vectors. By making use of the statistical independence of the different data vectors, and defining $\lambda(a_{i,\underline{\alpha}_{j_1}}, \ldots, a_{i,\underline{\alpha}_{j_p}}) = E[h_{i,\underline{\alpha}_{j_1}} \ldots h_{i,\underline{\alpha}_{j_p}}]$, (A8) can be rewritten as follows:

$$F(y) = \sum_{s=0}^{\infty} y^s \sum_{s_1=0}^{s} \sum_{a_{1,\underline{\alpha}_1}+\ldots+a_{1,\underline{\alpha}_s}=s_1} \lambda(a_{1,\underline{\alpha}_1}, \ldots, a_{1,\underline{\alpha}_s})$$
$$\times \sum_{s_2=0}^{s-s_1} \sum_{a_{2,\underline{\alpha}_1}+\ldots+a_{2,\underline{\alpha}_s}=s_2} \lambda(a_{2,\underline{\alpha}_1}, \ldots, a_{2,\underline{\alpha}_s}) \ldots$$
$$\times \sum_{s_n=s-s_1-s_2-\ldots-s_{n-1}} \sum_{a_{n,\underline{\alpha}_1}+\ldots+a_{n,\underline{\alpha}_s}=s_n} \lambda(a_{n,\underline{\alpha}_1}, \ldots, a_{n,\underline{\alpha}_s})$$
$$= \left(\sum_{s_1=0}^{\infty} y^{s_1} \sum_{\sum_{j=1}^{s} a_{1,\underline{\alpha}_j}=s_1} \lambda(a_{1,\underline{\alpha}_1}, \ldots, a_{1,\underline{\alpha}_s})\right) \ldots$$
$$\times \left(\sum_{s_n=0}^{\infty} y^{s_n} \sum_{\sum_{j=1}^{s} a_{n,\underline{\alpha}_j}=s_n} \lambda(a_{n,\underline{\alpha}_1}, \ldots, a_{n,\underline{\alpha}_s})\right)$$
$$= \left(\sum_{s=0}^{\infty} y^s \sum_{\sum_{j=1}^{s} a_{1,\underline{\alpha}_j}=s_1} \lambda(a_{1,\underline{\alpha}_1}, \ldots, a_{1,\underline{\alpha}_s})\right)^n. \qquad (A9)$$

We observe that, when the expectations contain only one factor, for $s_1 = 1$, we have $E[h_{1,\underline{\alpha}_1}{}^{a_{1,\underline{\alpha}_1}} h_{1,\underline{\alpha}_2}{}^{a_{1,\underline{\alpha}_2}} \ldots] = E[h_{1,\underline{\alpha}_j}] = \lambda(a_{1,\underline{\alpha}_j}) = 0$ for any j. As a consequence, the Taylor expansion of $f(y)$ in y can be written as $f(y) =$

$c_0 + c_2 y^2 + c_3 y^3 + \ldots$. Eventually, it is the absence of $c_1 y$ in this expansion that allows us to explain the properties of the polynomials in Eq. (A7) and finally the large sample approximation (A6). To accomplish this last step, we explain how the polynomials $P_m(n)$ are related to the coefficients c_i. Let us consider the contribution to $P_m(n)$ that originates from the product of k factors $c_{i_1} y^{i_1}, \ldots, c_{i_k} y^{i_k}$ with the constraints $i_l \geq 1$ and $i_1 + \ldots + i_k = m$. The number of ways these products can be formed for an assigned set of $i_1, \ldots, i_k$ is proportional to $\binom{n}{k} \simeq n^k +$ lower powers in n. Moreover, the contributions to $P_m(n)$ are formed by an integer number of factors less than or equal to $\frac{m}{2}$ because $c_1 = 0$. This property limits the highest power in n contained in $P_m(n)$ with the largest integer number smaller than or equal to $\frac{m}{2}$. Equation (A7) is then proved as is the large sample approximation (A6).

Each *asymptotic* order m of the rth component of the observed error is denoted by $(\hat{\vartheta} - \vartheta)^r[m]$, where the index is given by $m = \frac{s+1}{2}$ and s is a natural number. For example, the *asymptotic* order for $m = 1$ is given by

$$(\hat{\vartheta} - \vartheta)^r[1] = \tfrac{1}{2} v^{rst} l_s l_t + i^{rs} i^{tu} H_{st} l_u, \qquad (A10)$$

where we have adopted the lifted indexes notation $v^{a\ldots z} = i^{a\alpha} \ldots i^{z\delta} v_{\alpha\ldots\delta}$. The *asymptotic* orders of the bias are then given as the expectation of the *asymptotic* orders of the observed error

$$\tilde{b}[\hat{\vartheta}^r, m] = E[(\hat{\vartheta} - \vartheta)^r[m]]. \qquad (A11)$$

The *asymptotic* order $\tilde{b}[\hat{\vartheta}^r, m]$ contains different powers of $\frac{1}{n}$ as we discuss in this paragraph. It follows from Eqs. (A3)–(A6) that $\tilde{b}[\hat{\vartheta}^r, m]$ is the sum of terms having the structure $\underbrace{i^{(.)} \ldots i^{(.)}}_{a} \underbrace{v_{(.)} \ldots v_{(.)}}_{b} \underbrace{E[H_{(.)} \ldots H_{(.)}]}_{c}$ where a, b, and c are the numbers of factors in the three products satisfying $a - b - \frac{c}{2} = m$. Different powers of $\frac{1}{n}$ can be generated because $\underbrace{E[H_{(.)} \ldots H_{(.)}]}_{c}$ can contain all the integer powers of n equal

to or less than $n^{\text{int}(c/2)}$. However, from the fact that no power higher than $n^{\text{int}(c/2)}$ can be generated follows that an asymptotic order m will never generate powers in the sample size $\frac{1}{n^p}$ with p smaller than m. It still remains to prove which is the largest power in $\frac{1}{n}$ contained in the asymptotic order m. We show that in the following.

Since the largest range of powers of $\frac{1}{n}$ is allowed when $b = 0$, we study these terms to evaluate the range of powers of $\frac{1}{n}$ contained in $\tilde{b}[\hat{\vartheta}^r, m]$. The structure of Eq. (A3) implies that its iteration with respect to the observed error components generates an equal number of j^{ab} and $H_{(.)}$ (we recall that $l_{(.)} = H_{(.)} + v_{(.)}$). Similarly, the number of $i^{(.)}$ generated from the expansion of j^{ab} given in Eq. (A4) is equal to the number of H_{ab} plus 1. These two observations imply that the terms where $b = 0$ also verify

$a = c = 2m$. As a consequence, the powers of $\frac{1}{n}$ that $\tilde{b}[\hat{\vartheta}^r, m]$ contains can range from *the smallest integer number larger than or equal to m*, up to $2m - 1$. The analysis above implies that in order to compute the contributions to $\frac{1}{n}$ and to $\frac{1}{n^2}$ in the bias expansion that we denote with $b[\hat{\vartheta}^r, 1]$ and $b[\hat{\vartheta}^r, 2]$, it is necessary to obtain the first three *asymptotic* orders of the bias, $\tilde{b}[\hat{\vartheta}^r, 1]$, $\tilde{b}[\hat{\vartheta}^r, \frac{3}{2}]$, and $\tilde{b}[\hat{\vartheta}^r, 2]$. In the explicit expressions below, to condense the notation, we introduce the quantities $I^{\alpha\beta,\ldots,\gamma\delta} = i^{\alpha\beta} \ldots i^{\gamma\delta}$, so that

$$\tilde{b}[\hat{\vartheta}^r, 1] = I^{ri,kl} v_{ik,l} + \frac{1}{2} I^{rj,lp} v_{jlp} \qquad (A12)$$

$$\begin{aligned}
\tilde{b}[\hat{\vartheta}^r, \tfrac{3}{2}] = \tfrac{1}{6}(&I^{r\alpha,s\beta,t\gamma,u\delta} v_{\alpha\beta\gamma\delta} \\
&+ 3I^{r\alpha,s\beta,\gamma\delta,tg,uv} v_{\alpha\beta\gamma} v_{\delta,g,v} v_{s,t,u} \\
&+ I^{r\alpha,s\beta,u\gamma,tv} v_{\alpha\beta\gamma} v_{uv,s,t} \\
&+ \tfrac{1}{2} I^{rs,t\alpha,u\beta,v\gamma} v_{\alpha\beta\gamma} v_{st,u,v} + \tfrac{1}{2} I^{rs,tu,vw} v_{stv,u,w} \\
&+ I^{rs,tu,vw} v_{st,vu,w} + 2I^{rs,tw} v_{st,w} + I^{r\alpha,\beta\gamma} v_{\alpha\beta\gamma},
\end{aligned}$$
$$\qquad (A13)$$

$$\begin{aligned}
\tilde{b}[\hat{\vartheta}^r, 2] = (&I^{ra,bq,cd,tp} + \tfrac{1}{2} I^{rd,ta,bq,cp}) v_{\{abc\}\{dt\}\{q\}\{p\}} \\
&+ \tfrac{1}{2}(I^{ra,bs} v^{cpq} + I^{bp,cq} v^{rsa}) v_{\{abc\}\{s\}\{p\}\{q\}} \\
&+ (\tfrac{1}{2} I^{ap,tq} v^{rbg} + I^{ra,tq} v^{bpg} + \tfrac{1}{2} I^{ra,bg} v^{tpq} \\
&+ I^{bg,tq} v^{rpa}) v_{\{ab\}\{gt\}\{p\}\{q\}} + \tfrac{1}{2}[v^{rsz} v^{pqt} \\
&+ v^{rzqs} i^{tp} + \tfrac{1}{3} v^{pzqt} i^{rs} + (I^{dz,eq,pt} v^{rcs} \\
&+ I^{rp,dz,et} v^{sqc} + 2I^{dq,es,pt} v^{rzc}) v_{cde}] v_{\{sp\}\{z\}\{q\}\{t\}} \\
&+ I^{ra,st,bc,de} v_{\{ab\}\{cd\}\{et\}\{s\}} \\
&+ \tfrac{1}{6} I^{rj,kl,mp,qz} v_{\{jlpz\}\{k\}\{m\}\{q\}} + [\tfrac{1}{24} v^{r\alpha\beta\gamma\delta} \\
&+ \tfrac{1}{6} v^{r\alpha a} I^{\beta b,g\delta,\gamma d} v_{abgd} + \tfrac{1}{4} v^{r\alpha\gamma\delta} I^{\beta c,b\alpha} v_{abc} \\
&+ \tfrac{1}{8} v^{rva} I^{w\alpha,z\beta,b\gamma,g\delta} v_{vwz} v_{abg} \\
&+ \tfrac{1}{2} v^{r\alpha v} i^{w\beta} v_{vwz} v^{z\gamma\delta}] v_{\{\alpha\}\{\beta\}\{\gamma\}\{\delta\}}.
\end{aligned}$$
$$\qquad (A14)$$

The first order of the bias $b[\hat{\vartheta}^r, 1]$ can then be obtained by substituting the explicit expressions for the tensors in $\tilde{b}[\hat{\vartheta}^r, 1]$. The second order $b[\hat{\vartheta}^r, 2]$ takes contributions from $\tilde{b}[\hat{\vartheta}^r, \frac{3}{2}]$ and $\tilde{b}[\hat{\vartheta}^r, 2]$. However, while $\tilde{b}[\hat{\vartheta}^r, \frac{3}{2}]$ generates only n^{-2} contributions, $\tilde{b}[\hat{\vartheta}^r, 2]$ generates n^{-2} and n^{-3} contributions. Consequently, to collect all and only the contributions to n^{-2}, it is necessary to extract the n^{-2} component from $\tilde{b}[\hat{\vartheta}^r, 2]$ and add it to $\tilde{b}[\hat{\vartheta}^r, \frac{3}{2}]$. The extraction can be done by introducing into $\tilde{b}[\hat{\vartheta}^r, 2]$ only the highest powers of n of the tensors.

The first two orders of the bias for the MLE of the r component of the parameter vector $\underline{\vartheta}$ then become (2.2) and (2.3).

Starting from (2.3), the form for the second-order covariance matrix can be obtained using the following theorem:

Theorem: Let $g(.)$ be an invertible and differentiable function and let $\underline{\xi} = g(\underline{\vartheta})$ be a set of parameters dependent on $\underline{\vartheta}$. Let $b(\hat{\vartheta}_r)$ and $b(\hat{\xi}_r)$ be the biases relative to the estimation of the components of $\underline{\vartheta}$ and $\underline{\xi}$, which can be computed as power series of $\frac{1}{n}$, as explained in Sec. II. The terms of the expansion in powers of $\frac{1}{n}$ for $b(\hat{\xi}_r)$ can be obtained from the terms of the expansion for $b(\hat{\vartheta}_r)$ by replacing the derivatives with respect to the components of $\underline{\vartheta}$, with derivatives with respect to the components of $\underline{\xi}$. The explicit dependence on $\underline{\xi}$ can be removed at the end of a derivation by means of the Jacobian matrix J defined by $\frac{\partial}{\partial \xi_m} = \frac{\partial \vartheta_s}{\partial \xi_m} \frac{\partial}{\partial \vartheta_s} = J_{ms} \frac{\partial}{\partial \vartheta_s}$.

To prove the chain-rule theorem given in Sec. II, we analyze how the derivation of the expansion in powers of $\frac{1}{n}$ for $b(\vartheta_r) r = 1, \ldots, D$, which is described there, changes if we are interested in the expansion for $b(\xi_r) r = 1, \ldots, D$, where $\underline{\xi}$ is a vector of parameters in an invertible relationship with $\underline{\vartheta}$: $\underline{\xi} = f(\underline{\vartheta})$; $\underline{\vartheta} = f^{-1}(\underline{\xi})$. The starting point for the derivation of the expansion of $b(\xi_r) r = 1, \ldots, D$ is the stationarity condition

$$l_r(\underline{x}, \hat{\xi}) = \frac{\partial l(\underline{x}, f^{-1}(\underline{\xi}))}{\partial \xi_r}\Bigg|_{\underline{\xi} = \hat{\xi}} = 0 \qquad r = 1, \ldots, D.$$
$$\qquad (A15)$$

that can be obtained from Eq. (A1) by replacing only the derivatives $\frac{\partial}{\partial \vartheta_r}$ for $r = 1, \ldots, D$ with $\frac{\partial}{\partial \xi_r}$ for $r = 1, \ldots, D$, because $\underline{\vartheta} = \hat{\vartheta}$ implies $\underline{\xi} = f(\hat{\vartheta}) = \hat{\xi}$. The subsequent steps can then be obtained from Eqs. (A2) up to (2.3) by replacing the derivatives in the same way. Since $\frac{\partial}{\partial \xi_m} = \frac{\partial \vartheta_s}{\partial \xi_m} \frac{\partial}{\partial \vartheta_s} = J_{ms} \frac{\partial}{\partial \vartheta_s}$, where the components of the Jacobian matrix J_{ms} behave like constants in the expectations, the substitution of the derivatives can also be done only in the final result [for example, in the orders given in Eqs. (A14) and (2.3)]. The expectations contained in the expansion of $b(\vartheta_r)$ in powers of $\frac{1}{n}$ can also be computed before the substitution of the derivatives if the likelihood function dependence on the parameters is expressed in terms of auxiliary functions. Examples of auxiliary functions are $g_r(\underline{\vartheta}) = \vartheta_r$ for a general parametric dependence and $g_1(\vartheta) = \mu(\vartheta)$ $g_2(\vartheta) = \sigma(\vartheta)$ for a scalar Gaussian distribution. By means of these auxiliary functions, the derivatives $\frac{\partial}{\partial \vartheta_m}$ and $\frac{\partial}{\partial \xi_m}$ become $\frac{\partial g_p}{\partial \vartheta_m} \frac{\partial}{\partial g_p}$ and $\frac{\partial g_p}{\partial \xi_m} \frac{\partial}{\partial g_p}$. As a consequence the orders of the expansion for $b(\xi_r)$ can be found from the orders for the expansion of $b(\vartheta_r)$ implementing, in the result of the expectations, the substitutions

$$\frac{\partial^{i_1 + i_2 + \ldots + i_p} g_m(\underline{\vartheta})}{(\partial \vartheta_1)^{i_1} \ldots (\partial \vartheta_D)^{i_D}} \rightarrow \frac{\partial^{i_1 + i_2 + \ldots + i_D} g_m(f^{-1}(\xi))}{(\partial \xi_1)^{i_1} \ldots (\partial \xi_D)^{i_p}.} \qquad (A16)$$

The converse of the chain rule, in which higher moments of the observed error are used to compute lower moments, is not possible. We can observe, for example, that the expansion of a general moment of a MLE does not always contain all the powers of $\frac{1}{n}$. The lowest order present in the expansion of the m order moment is given by the largest integer number smaller than or equal to $\frac{m}{2}$ [we use the notation int$(\frac{m}{2})$]:

$$E[(\hat{\vartheta} - \vartheta)^{i_1} \ldots (\hat{\vartheta} - \vartheta)^{i_m}] = \frac{1}{n^{\text{int}(m/2)}}$$

$$+ \text{ higher powers of } \frac{1}{n}.$$

The consequence of this observation is that only the bias and the error correlation matrix may contain first order terms. For this reason an inverse chain rule would have to generate nonzero orders of lower moments expansions from the corresponding orders of the higher moments expansions that are zero for powers of $\frac{1}{n}$ lower than $\frac{1}{n^{\text{int}(m/2)}}$.

Let us consider how the chain rule makes it possible to compute the expansion of the error correlation matrix $\Xi(\hat{\vartheta})$ defined by $\Xi(\hat{\vartheta}) = E[(\underline{\hat{\vartheta}} - \underline{\vartheta})(\underline{\hat{\vartheta}} - \underline{\vartheta})^T]$ and the covariance matrix. Using the invariance property of the MLE,

$$\Xi(\hat{\vartheta}) = b(\widehat{\underline{\vartheta}\,\underline{\vartheta}^T}) - \underline{\vartheta}(b(\hat{\vartheta}))^T - b(\hat{\vartheta})\underline{\vartheta}^T, \qquad (A17)$$

where $(.)^T$ is the transpose of a vector and $b(\widehat{\underline{\vartheta}\,\underline{\vartheta}^T})$ is a matrix whose components are the bias of the product of two components of $\underline{\vartheta}$. Once $\Xi(\hat{\vartheta})$ and the bias are known, then the covariance matrix $C(\hat{\vartheta})$ can also be computed by means of

$$C(\hat{\vartheta}) = \Xi(\hat{\vartheta}) - b(\hat{\vartheta})b(\hat{\vartheta})^T. \qquad (A18)$$

To compute the right-hand side of Eq. (A17), we express it in terms of the components, obtaining

$$\Xi_{ij}(\hat{\vartheta}) = b(\widehat{\vartheta_i \vartheta_j}) - \vartheta_i b(\hat{\vartheta}_j) - b(\hat{\vartheta}_i)\vartheta_j. \qquad (A19)$$

It is important to realize that knowledge of $b(\hat{\vartheta}_r)$ is sufficient because the expansion of $b(\widehat{\vartheta_i \vartheta_j})$ can be derived from it by means of the theorem given above. In fact, if we choose

$$\underline{\xi} = \{\vartheta_1, \ldots, \vartheta_{i-1}, \vartheta_i \vartheta_j, \vartheta_{i+1}, \ldots, \vartheta_D\} \qquad (A20)$$

as a new set of parameters, $b(\widehat{\vartheta_i \vartheta_j})$ becomes $b(\hat{\xi}_i)$. However, it is necessary to insure that the relationship between $\underline{\vartheta}$ and $\underline{\xi}$ is invertible. This condition holds if both $\vartheta_i \neq 0$, where $i = 1, \ldots, D$ and the sign of the components of $\underline{\vartheta}$ is known, for example, by means of constraints on the data model. In a scalar estimation scenario, Eq. (A17) becomes simply

$$\Xi(\hat{\vartheta}) = E[(\hat{\vartheta} - \vartheta)^2] = b(\hat{\vartheta}^2) - 2\vartheta b(\hat{\vartheta}) \qquad (A21)$$

and the variance $C(\hat{\vartheta}) = \Xi(\hat{\vartheta}) - b(\hat{\vartheta})^2$. In this case $b(\hat{\vartheta}^2)$

can be derived from $b(\hat{\vartheta})$ if we choose $\xi = \vartheta^2$ as the new parameter and the Jacobian matrix becomes $\frac{1}{2\vartheta}$ because $\frac{\partial}{\partial \xi} = \frac{\partial \vartheta}{\partial \xi}\frac{\partial}{\partial \vartheta} = \frac{1}{2\vartheta}\frac{\partial}{\partial \vartheta}$. A useful simplification of the algebra of the chain rule in the derivation of second-order moments is described in the following two paragraphs. The chain-rule and the subsequent conversion of the derivatives require the substitutions

$$\frac{\partial^m[.]}{\partial \vartheta_{i_1} \ldots \partial \vartheta_{i_m}} \rightarrow \frac{\partial^m[.]}{\partial \xi_{i_1} \ldots \partial \xi_{i_m}}$$

$$\rightarrow J_{i_1 j_1}(\vartheta)\frac{\partial}{\partial \vartheta_{j_1}} J_{i_2 j_2}(\vartheta)\frac{\partial}{\partial \vartheta_{j_2}} \ldots$$

$$\times J_{i_m j_m}(\vartheta)\frac{\partial}{\partial \vartheta_{j_m}}[.] \qquad (A22)$$

From the right-hand side of (A22) it is clear that the derivatives will also be applied to the Jacobian matrix, thereby generating $m!$ terms for every derivative of order m. However, it can be observed that the terms contributing to the final result are only a small subset. In fact, among all the terms generated from the conversion of the derivatives in the bias expansion, only those where a first derivative of the Jacobian matrix appears must be considered. For example, we can consider the term $\frac{1}{2}I^{ra,sb,vc}v_{abc}v_{v,w}I^{wd,te,uf}v_{def}v_{s,t,u}$ that comes from Eq. (2.3), which must be used to derive the second-order $\Xi(\hat{\vartheta})$. In this case, we need to consider only the 3 terms in which one of the derivatives represented by a, b, c operates on a Jacobian matrix and d, e, f operate on the likelihood function, plus the 3 terms where the role of a, b, c and d, e, f are inverted. In general, the total number of relevant terms among those generated in every derivative is equal to or less than the order of the derivative m. The detailed analysis of Eq. (A19) reveals that the terms generated in bias$(\widehat{\vartheta_i \vartheta_j})$ can be divided into three groups: (a) the terms where no derivative of the Jacobian matrix appears are equal to $((J^{-1})^T)_{is}$bias$(\hat{\vartheta}_s)$ [we show in Example 1 that $((J^{-1})^T)_{is}$bias$(\hat{\vartheta}_s)$ cancels with $-\vartheta_i b(\hat{\vartheta}_j) - b(\hat{\vartheta}_i)\vartheta_j$ after its introduction in Eq. (A19)]; (b) the terms where only one derivative of the Jacobian matrix appears give the error correlation matrix; and (c) the terms that contain higher derivatives of the Jacobian matrix or more than one first derivative of the Jacobian matrix in the same term, summed together, give zero. To clarify the use of the chain-rule and the algebraic issues discussed above, we present two examples. In Example 1 we use the first order term of the bias in a general vector estimate to derive the first order term of $\Xi(\hat{\vartheta})$. It is useful to recall that the expansion of $\Xi(\hat{\vartheta})$ and the expansion of $C(\hat{\vartheta})$ can be different only from the second order on, so this example also describes a derivation of the first order of the covariance matrix $C(\hat{\vartheta})$. Following the same approach, the second-order term of the error correlation matrix expansion can be derived from Eq. (2.3) and the second-order

covariance matrix can also be derived if we also use (2.2) and (A18). In Example 2 we illustrate the way the chain rule can still be used, if the available expression of the bias expansion is explicitly computed for a particular data model. In particular, we derive the second-order mean square error and the second-order variance from the second-order bias in two scalar Gaussian models. In Example 2 we also illustrate the simplification introduced above for the algebra involved in the conversion of the derivatives.

Example 1

Using the Bartlett identities [31,23], we rewrite the first order bias, given by Eq. (A12), as

$$b[\hat{\vartheta}_r, 1] = -\tfrac{1}{2} i^{rj} i^{lp} (v_{j,l,p} + v_{j,lp}). \quad \text{(A23)}$$

From Eq. (A23), $b(\widehat{\vartheta_i \vartheta_j}) = b(\xi_i)$ can be computed by means of the chain rule by replacing the derivatives with respect to the components of $\underline{\vartheta}$ with derivatives with respect to the components of $\underline{\xi}$, where $\underline{\xi}$ is given in Eq. (A20) and using the corresponding Jacobian matrix. The tensors in Eq. (A23) become

$$i_{lp}(\xi) = E\left[\frac{\partial^2 l}{\partial \xi_l \partial \xi_p}\right] = J_{lr} i_{rs}(\vartheta)(J^T)_{sp}$$

$$= J_{lr} E\left[\frac{\partial^2 l}{\partial \vartheta_i \partial \vartheta_j}\right](J^T)_{sp} i^{lp},$$

$$(\xi) = ((J^{-1})^T)_{lr} i^{rs}(\vartheta)(J^{-1})_{sp} v_{j,l,p},$$

$$(\xi) = J_{j\alpha} J_{l\beta} J_{p\gamma} v_{\alpha,\beta,\gamma}(\vartheta) v_{j,lp},$$

$$(\xi) = J_{j\alpha} J_{l\beta}\left(J_{p\gamma} v_{\alpha,\beta,\gamma}(\vartheta) + \frac{\partial J_{p\gamma}}{\partial \beta} v_{\alpha,\gamma}(\vartheta)\right),$$

where we have specified in the bracket beside the tensors the dependence on the parameter sets. Inserting these expressions in Eqs. (A23) and observing that Eq. (A19) can be expressed in the form

$$\Xi(\hat{\vartheta})_{ij} = b(\widehat{\vartheta_i \vartheta_j}) - ((J^{-1})^T)_{is} b(\hat{\vartheta}_s) \quad \text{(A24)}$$

the first order term of the error correlation matrix can then be obtained as

$$\Xi_{ij}[\hat{\vartheta}, 1] = \frac{1}{2} i^{\alpha\beta} \frac{\partial (J^{-1})_{\alpha i}}{\partial \vartheta_\beta} = \frac{1}{2} i^{\alpha\beta} \gamma_{\alpha\beta} = i^{ij}, \quad \text{(A25)}$$

where we have introduced the tensor

$$\gamma_{\alpha\beta} = \begin{cases} 1 & \text{if } \alpha = i\beta = j \text{ or } \alpha = j\beta = i \\ 0 & \text{otherwise.} \end{cases}$$

Example 2

In this example we determine the second-order variance and mean square error from the second-order bias for two cases of parameter dependence for the scalar Gaussian density

$$p(x, \vartheta) = \frac{1}{(2\pi c)^{n/2}} \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \frac{(x_i - \mu)^2}{c}\right) \quad \text{(A26)}$$

as a direct application of Eq. (A21). In the case where c does not depend on the parameters ($\frac{\partial c}{\partial \vartheta} = 0$) the second-order bias can be derived using the scalar version of Eq. (2.3) and of the tensors v defined in Eq. (A5). For this parameter dependence, the *asymptotic* order for $m = \frac{3}{2}$ [Eq. (A13)] is zero, and the second-order bias can be directly obtained also from Eq. (A14). The result is $b[\frac{1}{n^2}] = \frac{c^2}{(\mu)^8}[\frac{5}{4} \ddot{\mu}\, \ddot{\mu}(\dot{\mu})^2 - \frac{1}{8} \dddot{\mu}(\dot{\mu})^3 - \frac{15}{8} \dot{\mu}(\ddot{\mu})^3]$. Applying the chain-rule, the second-order mean square error for the estimation of ϑ becomes $\Xi(\vartheta)[\frac{1}{n^2}] = \frac{15\ddot{\mu}^2}{4\dot{\mu}^6} - \frac{\dddot{\mu}}{\dot{\mu}^5}$, where the full conversions of the derivatives are given by $\dot{\mu} \to \frac{\partial \mu}{\partial \vartheta^2} = \frac{\dot{\mu}}{2\vartheta}$; $\ddot{\mu} \to \frac{\partial^2 \mu}{(\partial \vartheta^2)^2} = \frac{1}{2\vartheta}(\frac{\ddot{\mu}}{2\vartheta} - \frac{\dot{\mu}}{2\vartheta^2})$; $\dddot{\mu} \to \frac{\partial^3 \mu}{(\partial \vartheta^2)^3} = \frac{1}{2\vartheta} \times (\frac{\dddot{\mu}}{4\vartheta^2} - \frac{3\ddot{\mu}}{4\vartheta^3} + \frac{3\dot{\mu}}{4\vartheta^4})$; $\ddddot{\mu} \to \frac{\partial^4 \mu}{(\partial \vartheta^2)^4} = \frac{\ddddot{\mu}}{16\vartheta^4} - \frac{3\dddot{\mu}}{8\vartheta^5} + \frac{15\ddot{\mu}}{16\vartheta^6} - \frac{15\dot{\mu}}{16\vartheta^7}$. By means of these conversions and the expression of the second-order bias, it can be observed that among the 18 terms that are in principle generated by the chain rule, only 6 contribute to the second-order mean square error.

In the following we show that the expansion in the inverse of the sample size is equivalent to an expansion in 1/SNR. The derivation of the asymptotic orders in $\frac{1}{n}$ would be the same for an expansion in any quantity if (A5) and (A6) can be derived for a certain quantity γ instead of the sample size n. In this section we illustrate indeed that this is the case where the signal-to-noise ratio for a set of scalar data distributed according to a deterministic Gaussian PDF takes the role of the sample size. The probability distribution and parameter dependent part of the likelihood function are given by

$$p(x, \vartheta) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(\sum_{i=1}^{n}(x_i - \mu(\vartheta))^2/\sigma^2)} \quad \text{(A27)}$$

$$l(\vartheta) = -\frac{\sum_{i=1}^{n}(x_i - \mu(\vartheta))^2}{\sigma^2} \quad \text{(A28)}$$

We also define the signal-to-noise ratio for this example as $\gamma = \mu^2/\sigma^2$ following standard practice for scalar deterministic Gaussian data. We can obtain

$$H_\alpha(\vartheta) = 2\mu_\alpha(\vartheta)\frac{\sum_{i=1}^{n}(x_i - \mu(\vartheta))}{\sigma^2} \quad \text{(A29)}$$

$$l_\alpha(\vartheta) = -n\sum_p c(p)\frac{\mu^p \mu^{\alpha-p}}{\sigma^2} + 2\mu_\alpha(\vartheta)\frac{\sum_{i=1}^{n}(x_i - \mu(\vartheta))}{\sigma^2} \quad \text{(A30)}$$

$$v_\alpha = \langle l_\alpha(\vartheta) \rangle = n\frac{\mu^2}{\sigma^2}\left[-\sum_p c(p)\left(\frac{\mu^p}{\mu}\right)\left(\frac{\mu^{\alpha-p}}{\mu}\right)\right] \quad \text{(A31)}$$

where α or p denote the order of the derivative or an arbitrary set of derivatives.

From (A31) it becomes obvious that v_α is proportional to not only the sample size n, but also to the γ. The term inside the square brackets is simply a sum over normalized

derivatives of the mean and contains information about the shape of the signal. In the above equations we use

$$E[(x - \mu(\vartheta))^{2n}] = \sigma^{2n} \frac{(2n)!}{2^n n!}. \qquad (A32)$$

The next step is to determine the dependence of $E[H_{\alpha_1} \ldots H_{\alpha_p}]$ on γ. This is shown below:

$$E[H_{\alpha_1} \ldots H_{\alpha_p}] = 2^p \mu_{\alpha_1} \ldots \mu_{\alpha_p} \frac{1}{\sigma^{2p}} E\left[\sum_{i=1}^n (x_i - \mu(\vartheta))^p \right]$$

$$= 2^p \left(\frac{\mu_{\alpha_1}}{\mu} \right) \ldots \left(\frac{\mu_{\alpha_p}}{\mu} \right) \frac{\mu^p}{\sigma^{2p}} \sigma^p \frac{(2 \frac{p}{2})!}{2^{p/2}(\frac{p}{2})!} n^{p/2}$$

$$= \left(\frac{\mu^2}{\sigma^2} \right)^{p/2} \ldots \left(\frac{\mu_{\alpha_1}}{\mu} \right) \ldots \left(\frac{\mu_{\alpha_p}}{\mu} \right) n^{p/2}$$

$$\Rightarrow E[H_{\alpha_1} \ldots H_{\alpha_p}] \propto (n\gamma)^{p/2} \qquad (A33)$$

It is therefore found that $E[H_{\alpha_1} \ldots H_{\alpha_p}]$ is proportional to $(\gamma)^{\mu/2}$ and we have proved the analogy between sample size and γ.

Note that for deterministic Gaussian data, the noninteger asymptotic orders of the bias are zero and the integer orders are equal to b_1, b_2, etc. This is sufficient to prove that the orders in $1/n$ of the bias expansion are also orders in $1/\gamma$. This also holds for the variance expansion. A similar, although longer proof can be written for the SNR definition provided in (4.5).

[1] B. P. Abbott *et al.*, Rep. Prog. Phys. **72**, 076901 (2009).
[2] H. Grote, Classical Quantum Gravity **25**, 114043 (2008).
[3] F. Acernese *et al.*, Classical Quantum Gravity **23**, 563 (2006).
[4] V. Kalogera, K. Belczynski, C. Kim Richard, W. O'Shaughnessy, and B. Willems, Phys. Rep. **442**, 75 (2007).
[5] A. Willsky and J. Shapiro, MIT Report No. EE6432, 1999.
[6] K. G. Arun, B. R. Iyer, B. S. Sathyaprakash, and P. A. Sundararajan, Phys. Rev. D **71**, 084008 (2005).
[7] K. G. Arun, B. R. Iyer, B. S. Sathyaprakash, and P. A. Sundararajan, Phys. Rev. D **72**, 069903(E) (2005).
[8] R. Balasubramanian, B. S. Sathyaprakash, and S. V. Dhurandhar, Phys. Rev. D **53**, 3033 (1996).
[9] R. Balasubramanian and S. V. Dhurandhar, Phys. Rev. D **57**, 3408 (1998).
[10] L. S. Finn and D. F. Chernoff, Phys. Rev. D **47**, 2198 (1993).
[11] P. Jaranowski and A. Krolak, Phys. Rev. D **49**, 1723 (1994).
[12] D. Nicholson and A. Vecchio, Phys. Rev. D **57**, 4588 (1998).
[13] C. Cutler and E. E. Flanagan, Phys. Rev. D **49**, 2658 (1994).
[14] E. Poisson and C. M. Will, Phys. Rev. D **52**, 848 (1995).
[15] A. Królak, K. D. Kokkotas, and G. Schäfer (unpublished).
[16] S. Fairhrst, New J. Phys. **11**, 123006 (2009); Phys. Rev. D **52**, 2089 (1995).
[17] P. Ajith and S. Bose, Phys. Rev. D **79**, 084032 (2009).
[18] T. Cokelaer, Classical Quantum Gravity **25**, 184007 (2008).
[19] C. Cutler, and E. Flanagan, Phys. Rev. D **49** (6) 2658 (1994).
[20] L. S. Finn, Phys. Rev. D **49** (6) 2658 (1992).
[21] Haldane and Smith, Biometrika **43**, 96 (1956).
[22] E. Naftali and N. C. Makris, J. Acoust. Soc. Am. **110**, 1917 (2001).
[23] M. S. Bartlett, Biometrika **40**, 306 (1953).
[24] A. Thode, M. Zanolin, E. Naftali, I. Ingram, P. Ratilal, and N. C. Makris, J. Acoust. Soc. Am. **112** (5) (2001).
[25] M. Vallisneri, Phys. Rev. D **77**, 042001 (2008).
[26] D. N. Lawley, Biometrika **43**, 295 (1956).
[27] L. Blanchet, G. Faye, B. R. Iyer, and B. Joguet, Phys. Rev. D **65**, 061501(R) (2002).
[28] T. Damour, B. R. Iyer, and B. S. Sathyaprakash, Phys. Rev. D **63**, 044023 (2001).
[29] T. Damour, B. R. Iyer, and B. S. Sathyaprakash, Phys. Rev. D **66**, 027502 (2002).
[30] M. Zanonlin and N. Makris, MIT Tech. Report 2001 (unpublished).
[31] M. S. Bartlett, Biometrika **40**, 12 (1953).
[32] M. Zanolin, I. Ingram, A. Thode, and N. C. Makris, J. Acoust. Soc. Am. **116**, 2031 (2004).
[33] P. McCullagh, *Tensor Methods in Statistics* (Chapman and Hall, London, 1987).
[34] O. E. Barndoff-Nielsen, D. R. Cox, and N. Reid, Int. Stat. Review **54**, 1 (1986).
[35] O. E. Barndoff-Nielsen, W. S. Kendall, and M. N. M. Lieshout, *Stocastic Geometry: Likelihood and Computation* (Chapman and Hall, London, 1999).
[36] L. R. Shenton and K. O. Bowman, J. R. Stat. Soc. 1963).
[37] K. O. Bowman and L. R. Shenton, Commun. Statist. Theory Methods **27** (11) 2743 (1998).
[38] K. O. Bowman, L. R. Shenton, Ann. Inst. Stat. Math. **44**, 781 (1992).
[39] K. O. Bowman and L. R. Shenton, Commun. Stat. Simul. Comput. **12**, 697 (1983).
[40] L. R. Shenton and K. O. Bowman, in Sankhia: The Indian Journal of Statistics **48**, 219 (1986), Series A, Part 2.
[41] L. R. Shenton and K. O. Bowman, Commun. Statist. Theory Methods , 1511 (1990).
[42] O. E. Barndoff-Nielsen and D. R. Cox, *Inference and Asymptotic* (Chapman-Hall, London, 1994).
[43] L. R. Shenton and K. O. Bowman, *Maximum Likelihood Estimation in Small Samples* (Griffin, London, 1977).