Learning approach to the detection of gravitational wave transients

E. Chassande-Mottin*
Equipe ILGA-Virgo (CNRS-EP 2122), Observatoire de la Côte d'Azur, BP 4229, F-06304 Nice Cedex 4, France
(Received 3 October 2002; published 7 May 2003)

We investigate the class of quadratic detectors (i.e., the statistic is a bilinear function of the data) for the detection of poorly modeled gravitational transients of short duration. We point out that all such detection methods are equivalent to passing the signal through a filter bank and linearly combining the output energy. Existing methods for the choice of the filter bank and of the weight parameters (to be multiplied by the output energy of each filter before summation) rely essentially on the two following ideas: (i) the use of the likelihood function based on a (possibly noninformative) statistical model of the signal and the noise; (ii) the use of Monte Carlo simulations for the tuning of parametric filters to get the best detection probability while keeping the false alarm rate fixed. We propose a third approach according to which the filter bank is ''learned'' from a set of training data. By-products of this viewpoint are that, contrarily to previous methods, (i) there is no requirement of an explicit description of the probability density function of the data when the signal is present and (ii) the filters we use are nonparametric. The learning procedure may be described as a two step process: first, estimate the mean and covariance of the signal with the training data; second, find the filters which maximize a contrast criterion referred to as the *deflection* between the ''noise only'' and ''signal + noise'' hypotheses. The deflection is homogeneous to the signal-to-noise ratio and it uses the quantities estimated at the first step. We apply this original method to the problem of the detection of supernovae core collapses. We use the catalog of waveforms provided recently by Dimmelmeier *et al.* to train our algorithm. We expect such a detector to have better performances in this particular problem provided that the reference signals are reliable.

DOI: 10.1103/PhysRevD.67.102001 PACS number(s): 04.80.Nn, 07.05.Kf, 95.55.Ym

I. MOTIVATIONS

A number of large scale gravitational wave interferometric detectors such as the Laser Interferometric Gravitational Wave Observatory (LIGO), TAMA, GEO600, and VIRGO [1] are taking scientific data or will reach this goal soon. The objective of this paper is to contribute to the arsenal of detection algorithms able to locate the weak and short signature of a gravitational wave of astrophysical origin in the long data stream produced by the detector.

In the list of candidates having a good chance for first detection, there are sources for which we can only make a rough guess or simulate the highly nonlinear physics which is involved. This causes the expected gravitational waveforms to be poorly modeled. Most of these are collapses of very massive astrophysical objects in the final stage, and this causes the resulting gravitational wave to be a burst.

In such cases, computer simulations may give good indications of what the waveform can be for some choices of the parameter values describing the physical phenomenon. However, it is generally not possible to have a tight sampling of the parameter space, i.e., to scan a large range of physical configurations. We only have at our disposal a catalog of waveforms, whose members are selected representatives of the large set of possibilities.

Two examples of such sources are supernovae core collapse and binary black hole merger. Despite some recent progress, work is still needed in the latter case to produce reliable waveforms in a realistic setup (including spins, for instance). Concerning the former case, hydrodynamic simulations of relativistic supernovae have recently been computed [2,3] and the expected waveforms for different parameter configurations were made available.

In this paper, we propose a method for systematically designing a decision statistic for the detection of gravitational transients by extracting the necessary information from a catalog of test waveforms emitted from a targeted source. We use the supernovae waveforms as one possible application. Although not considered in this paper, the presented approach may also apply to other problems encountered when analyzing the output of a gravitational wave interferometer, such as the classification and characterization of the noisy transients. (Because they worsen the detector sensitivity, such interferences deserve the development of algorithms to determine their actual origin.) In this context, the initial database could be a collection of characteristic individuals extracted ''by hand'' or with some other simple algorithm. In any case, we refer to the (gravitational wave or noise) transient(s) we want to detect as the *signal*.

Some attention has been paid to the choice of an adequate vector formalism to treat the problem and make the implementation on computers easier. The resulting notation is defined in Sec. II. This section also includes the formulation, within the chosen framework, of classical results such as the Plancherel formula which will be of use further on.

In Sec. III, we describe the detection problem we consider with the accompanying hypotheses. We assume the signal to be random and of unknown probability density function (PDF). This assumption explicitly translates the lack of knowledge about the signal. The only pieces of information at our disposal are its first and second order statistical moments (i.e., its mean and covariance). In the situation of interest here, these two quantities are not known, but they can

*Electronic address: ecm@obs-nice.fr

be estimated with sufficient accuracy from available sample sets.

We consider that the noise is Gaussian and stationary and that we know its correlation function (or equivalently its power spectral density). Analogously to the signal, an extension to the case where there is no reasonable noise model is possible through the use of estimates done with ''noise only'' data streams.

Since we do not have the signal PDF, it is impossible to write the exact form of the likelihood ratio, and thus to obtain the optimal statistic. However, a satisfactory solution can be obtained by first imposing the mathematical structure of the statistic and secondly looking at the selected set of functions for the best element by maximizing a contrast criterion.

The difficulty then lies in making the choice of a sufficiently general class of statistics and a sensible criterion for the problem considered. In Sec. IV, it is explained why the family of *quadratic detectors* (i.e., the detection statistic is a quadratic form of the observed data) and a measurement of the *deflection* (a quantity homogeneous to the signal-to-noise ratio) are good candidates. Coming back to our initial detection problem, we individuate in the selected class the statistic which performs best according to the chosen criterion and show that it can be expressed easily with the signal and noise covariance. Our proof was inspired by the work presented in [4] and [5] which we adapt to the case of interest (finite vector spaces and noncentered signals). We finish Sec. IV by showing that the proposed approach is not unfamiliar since it can be related to the well-known method of matched filtering.

In Sec. V, we put the quadratic detector of best deflection into practice with the problem of detecting gravitational wave transients from supernovae core collapse. In this specific case, we show how the signal covariance matrix can be estimated from the catalog of simulated waveforms. This yields a simplification of the detector and an efficient implementation for online detection. We also give some results about the determination of the decision threshold required to get a chosen false alarm rate.

The vector formalism introduced here is a general framework in which all quadratic detectors can be easily related and compared. In Sec. VI, we do this comparative study between the solution with optimal deflection and other detection techniques [6,7] proposed in the literature that also belong to the quadratic class.

II. NOTATION AND BASIC ALGEBRA

We will denote scalar quantities and scalar-valued functions with plain italics, e.g., x; vectors by boldface letters, e.g., $\boldsymbol{x}$; and matrices by boldface capitals, e.g., $\boldsymbol{X}$. We will represent the components of vectors and matrices with superscripts within parentheses, e.g. $\boldsymbol{X}^{(m,n)}$ designates the element located in the mth row and nth column of the matrix $\boldsymbol{X}$. Finally, $\boldsymbol{x}^t$ denotes the transpose of the vector $\boldsymbol{x}$. The symbol $\equiv$ will be used in the following to define our variables and therefore stands for ''equal by definition.''

We use parentheses for denoting continuous (random and/or deterministic) time (or frequency) series [e.g., $x(t)$], whereas square brackets are employed for discrete time (sampled) processes (e.g., $x[k]$). The samples of a discrete time signal are collected in a single column vector of $\mathbb{R}^N$, e.g.,

$$\boldsymbol{x} \equiv (x[k] = x(t_s k), k = 0, \ldots, N-1)^t \tag{1}$$

where $t_s \equiv 1/f_s$ is the sampling period and f_s the sampling rate.

We define the Fourier transform of $x[k]$ by

$$X(f) \equiv t_s \sum_{k=-\infty}^{+\infty} x[k] e^{-2i\pi k f/f_s}. \tag{2}$$

As a general rule, a Fourier transform is denoted by the same (capital) letter used for its associated time sequence. The function $X(f)$ is f_s-periodic [i.e., $X(f) = X(f+f_s)$ for all f] and its inverse may be calculated with the following inversion equation:

$$x[k] = \int_{-f_s/2}^{f_s/2} X(f) e^{2i\pi k f/f_s}\, df. \tag{3}$$

We recall that the Plancherel formula relates scalar products expressed in the time and frequency domains:

$$t_s \sum_{j=-\infty}^{+\infty} x[j] y[j] = \int_{-f_s/2}^{+f_s/2} X(f) \overline{Y(f)}\, df. \tag{4}$$

This equation may also be expressed using a vector scalar product, provided we assume that one of the two signals $x[k]$ or $y[k]$ has a finite support (denoted with $\mathrm{supp}\{\cdot\}$), as specified in the following lemma.

Lemma 1. Assuming that $\mathrm{supp}\{y\} \subset \{0, \ldots, N-1\}$, the Plancherel formula becomes

$$t_s \boldsymbol{x}^t \boldsymbol{y} = \int_{-f_s/2}^{+f_s/2} X(f) \overline{Y(f)}\, df. \tag{5}$$

Continuing with the same idea, the convolution of two signals $x[k]$ and $y[k]$ defined by

$$z[k] = t_s \sum_{j=-\infty}^{+\infty} x[k-j] y[j] \tag{6}$$

or equivalently by $Z(f) = X(f)Y(f)$ may also be rewritten using vectors under some support constraints as in the next lemma.

Lemma 2. Let $N_y < (N-1)/2$ be a positive integer and suppose that $\mathrm{supp}\{y\} = \{-N_y, \ldots, N_y\}$, then the collection of samples $\boldsymbol{z} = (z[N_y], \ldots, z[N-1-N_y])^t$ where $z[k]$ is the convolution of $x[k]$ and $y[k]$ as defined in Eq. (6) can be expressed as

$$\boldsymbol{z} = t_s \boldsymbol{Y} \boldsymbol{x} \tag{7}$$

where $\boldsymbol{Y} \in \mathbb{R}^{N-2N_y \times N}$ is a matrix of the form

$$\boldsymbol{Y}=\begin{pmatrix} y[N_y] & \cdots & y[-N_y] & 0 & \cdots & \cdots & 0 \\ 0 & y[N_y] & \cdots & y[-N_y] & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & \cdots & 0 & y[N_y] & \cdots & y[-N_y] \end{pmatrix}. \tag{8}$$

Proofs of both Lemmas 1 and 2 are simple and left to the reader.

III. PROBLEM STATEMENT

The question we consider here is the detection of a (possibly nonstationary) random signal in a stationary Gaussian noise (where the signal and noise covariance function is known or could be estimated with accuracy). Using the notation defined in the previous section, the problem is to distinguish between two statistical hypotheses (H_0) and (H_1):

$$(H_0):\ \boldsymbol{x}=\boldsymbol{n}, \tag{9a}$$

$$(H_1):\ \boldsymbol{x}=\boldsymbol{s}+\boldsymbol{n}, \tag{9b}$$

with the following assumptions.

(1) The signal $\boldsymbol{s}$ is a random vector of mean $\boldsymbol{s}_m \equiv \mathsf{E}[\boldsymbol{s}]$ (where $\mathsf{E}[\cdot]$ denotes the expectation operator) and correlation matrix $\boldsymbol{R}_s \equiv \mathsf{E}[(\boldsymbol{s}-\boldsymbol{s}_m)(\boldsymbol{s}-\boldsymbol{s}_m)^t]$.

(2) The noise $\boldsymbol{n}$ is a zero-mean, stationary Gaussian vector with correlation matrix $\boldsymbol{R}_n$. Note that, since $\boldsymbol{n}$ is stationary, $\boldsymbol{R}_n$ is a Toeplitz symmetric matrix (the terms of $\boldsymbol{R}_n$ are given by the autocorrelation function $\boldsymbol{R}_n^{(j,j+k)} \equiv \mathsf{E}[n[j]n[j+k]]$).

(3) The signal and the noise are decorrelated, meaning that $\mathsf{E}[\boldsymbol{n}^t(\boldsymbol{s}-\boldsymbol{s}_m)]=0$,

This set of assumptions correspond to several different practical situations. A first one is when the signal or noise model is good enough to get reliable closed form expressions of $\boldsymbol{s}_m$, $\boldsymbol{R}_s$, and $\boldsymbol{R}_n$. The second situation is when a sufficiently large set of ''signal only'' and/or ''noise only'' realizations is available and can be used to obtain a good estimate of the first and second order moments of the signal and the noise. Note that, except for its first and second order moments, we made no hypothesis about the PDF of $\boldsymbol{s}$.

IV. QUADRATIC DETECTORS

Deciding (H_1) or (H_0) is classically done by finding a partition function $\Lambda(\cdot)$ dividing the observation space (here, $\mathbb{R}^N$) into two disjoint subsets:

$$\Lambda(\boldsymbol{x}) \geqslant \eta, \quad \text{decide } (H_1), \tag{10a}$$

$$\Lambda(\boldsymbol{x}) < \eta, \quad \text{decide } (H_0), \tag{10b}$$

where the detection threshold η defines the border between the two decision areas. Its value is given by fixing to a reasonable value the probability of deciding upon hypothesis (H_1) although no signal $\boldsymbol{s}$ is present, which we refer to as the ''false alarm probability.'' The function $\Lambda(\cdot)$ is referred to as the *detection statistic* or simply the *detector*.

A. Intuitive background

It is intuitive to search for some unknown signal by looking for an abnormal excess of power in one or several frequency bands of the observed signal spectrum. To implement this idea, we define the power of a signal x using an l^2 measure:

$$E_x \equiv 1/N \sum_{k=0}^{N-1} x[k]^2 = \boldsymbol{x}^t\boldsymbol{x}/N, \tag{11}$$

and a bank of filters which adequately select the signal in the frequency bands of interest. Let $\{g_m[k], k=0,\ldots,N-1 \text{ and } m=0,\ldots,M-1\}$ be the impulse responses (assumed to be of finite support) of the chosen bandpass filters. We get the signal $y_m[k]$ at the output of each filter by convolving the observed signal $x[k]$ with the corresponding impulse response. With the constraint that $\mathrm{supp}\{g_m\} \subset \{-N_g,\ldots,N_g\}$ for all m where $N_g < (N-1)/2$, we can apply Lemma 2 and express the output signal in vector form as

$$\boldsymbol{y}_m = t_s \boldsymbol{G}_m \boldsymbol{x}, \tag{12}$$

where $\boldsymbol{y}_m$ and $\boldsymbol{G}_m$ are as defined in Lemma 2.

Hence, we can write down the detection statistic corresponding to the basic idea mentioned above by summing up the power in all M bands which yields

$$\Lambda_{intuitive}(\boldsymbol{x}) \equiv \sum_{m=0}^{M-1} E_{y_m}^2 = \boldsymbol{x}^t \left(\frac{t_s^2}{N-2N_g} \sum_{m=0}^{M-1} \boldsymbol{G}_m^t \boldsymbol{G}_m \right) \boldsymbol{x}. \tag{13}$$

We conclude that the heuristic principle we chose is implemented in practice with a detection statistic which is a quadratic form of the data. Extending this result to cases where the kernel of the form is an arbitrary symmetric matrix, this leads us to define the following family of detectors.

Definition 1. Let $\boldsymbol{A} \in \mathbb{R}^{N\times N}$ be a symmetric real matrix; the function

$$\Lambda_{\boldsymbol{A}}(\boldsymbol{x}) = \boldsymbol{x}^t \boldsymbol{A} \boldsymbol{x} \tag{14}$$

defines a quadratic detector of kernel $\boldsymbol{A}$.

Quadratic detectors will be a central ingredient in this paper. It is worth noting that they have been extensively used in many applications (see, e.g., [8] for a list of examples). In the case of gravitational wave detection, the specific area of

interest here, several works [6,7,9–11] are based on this detector structure.

Beyond qualitative arguments, the following theoretical result is an important motivation for the use of quadratic detectors [12]: with the additional assumption of a zero-mean Gaussian signal (i.e., both signal and noise are Gaussian), the optimal solution in the Neymann-Pearson sense of the problem (9) belongs to the family defined in Definition 1. Although the problem considered here excludes a Gaussian signal, we expect quadratic detectors to retain their good performance when the signal PDF is close to Gaussian.

B. Optimal quadratic detectors

For our problem (9), we do not have a complete knowledge of the statistics of the input signal (the PDF of $\boldsymbol{s}$ is unknown). In consequence, we cannot write out the likelihood ratio which gives the best (in the Neymann-Pearson sense) detector among all possibilities.

We propose to overcome this difficulty by first reducing the class of possible solutions to the family of quadratic detectors defined in Definition 1, and, second, extracting from this smaller set the statistic which will best perform for our problem. More precisely, our objective is to get the quadratic detector (i.e., get the kernel matrix $\boldsymbol{A}$ which identifies this quadratic detector in the whole family) that maximizes the following contrast criterion based on the first and second order moments:

$$d^2(\Lambda_{\boldsymbol{A}}) = \frac{(\mathsf{E}[\Lambda_{\boldsymbol{A}}(\boldsymbol{x})|H_1] - \mathsf{E}[\Lambda_{\boldsymbol{A}}(\boldsymbol{x})|H_0])^2}{\mathrm{var}\{\Lambda_{\boldsymbol{A}}(\boldsymbol{x})|H_0\}}. \tag{15}$$

This criterion is generally referred to as the *signal-to-noise ratio* (statistics) or *deflection* (signal processing). The terminology ''signal-to-noise ratio'' is generally associated in most of the literature about gravitational wave data analysis with the quantity in Eq. (15) where $\Lambda(\cdot)$ is set to the matched filter statistic. In consequence, we adopt the term ''deflection'' to avoid confusion. The deflection may be viewed as a contrast measurement between the two statistical hypotheses H_0 and H_1 in the sense that it measures the distance between the centers of the PDFs of the statistic in the two hypotheses relative to the PDF width in the null hypothesis H_0.

In the context of the problem (9), we now apply this approach for selecting the best statistic among all quadratic detectors.

Lemma 3. In the situation described in Eq. (9), the deflection of a quadratic detector as defined in Definition 1 is

$$d^2(\Lambda_{\boldsymbol{A}}) = \frac{\mathrm{tr}^2\{\boldsymbol{A}\boldsymbol{C}_s\}}{2\,\mathrm{tr}\{(\boldsymbol{A}\boldsymbol{R}_n)^2\}}, \tag{16}$$

where $\boldsymbol{C}_s \equiv \mathsf{E}[\boldsymbol{s}^t\boldsymbol{s}]$ defines the (noncentral) covariance matrix and $\mathrm{tr}\{\cdot\}$ is the trace[1] operator.

[1]Let $\boldsymbol{A} \in \mathbb{R}^{N\times N}$, then the operator $\mathrm{tr}\{\boldsymbol{A}\} \equiv \Sigma_{n=0}^{N} \boldsymbol{A}^{(n,n)}$ defines the trace of $\boldsymbol{A}$.

The proof of this lemma is detailed in Appendix A. In order to find the best detector in the quadratic family, we need now to find which kernel matrix maximizes the deflection. This is stated by the following proposition.

Proposition 1. There exists a unique symmetric matrix $\boldsymbol{H}$ such that $d^2(\Lambda_{\boldsymbol{H}})$ obtained in Lemma 3 is maximum, and this matrix is

$$\boldsymbol{H} = \mathrm{argmax}_{\boldsymbol{A}}\{d^2(\Lambda_{\boldsymbol{A}})\} = \boldsymbol{R}_n^{-1}\boldsymbol{C}_s\boldsymbol{R}_n^{-1}. \tag{17}$$

Proof. We first note that the noise autocorrelation $\boldsymbol{R}_n$ is a symmetric positive definite matrix. Therefore, there exists a triangular matrix $\boldsymbol{T}_n$ with positive diagonal such that $\boldsymbol{R}_n = \boldsymbol{T}_n\boldsymbol{T}_n^t$. This factorization method is referred to as Cholesky factorization [13].

It is useful to introduce the two matrices $\boldsymbol{G} \equiv \boldsymbol{T}_n\boldsymbol{A}\boldsymbol{T}_n^t$ and $\boldsymbol{C} \equiv (\boldsymbol{T}_n^{-1})^t\boldsymbol{C}_s\boldsymbol{T}_n^{-1}$, which we substitute in the expression of the deflection we got in Lemma 3, thus reducing it to

$$d^2(\Lambda_{\boldsymbol{A}}) = \frac{\mathrm{tr}^2\{\boldsymbol{G}\boldsymbol{C}\}}{2\,\mathrm{tr}\{\boldsymbol{G}^2\}}. \tag{18}$$

Let $\mathcal{S}_N(\mathbb{R})$ be the vector space of real symmetric matrices of $\mathbb{R}^{N\times N}$. It is easily shown that $\langle \boldsymbol{A},\boldsymbol{B}\rangle_{\mathcal{S}_N(\mathbb{R})} \equiv \mathrm{tr}\{\boldsymbol{A}\boldsymbol{B}\}$ defines a scalar product on $\mathcal{S}_N(\mathbb{R})$. Since the matrices $\boldsymbol{G}$ and $\boldsymbol{C}$ belong to $\mathcal{S}_N(\mathbb{R})$, we can rewrite the deflection as a ratio of scalar products, also referred to as a Rayleigh quotient [13], namely,

$$d^2(\Lambda_{\boldsymbol{A}}) = \frac{\langle \boldsymbol{G},\boldsymbol{C}\rangle^2_{\mathcal{S}_N(\mathbb{R})}}{2\langle \boldsymbol{G},\boldsymbol{G}\rangle_{\mathcal{S}_N(\mathbb{R})}}. \tag{19}$$

Using the Cauchy-Schwarz inequality, we conclude that $d^2(\Lambda_{\boldsymbol{A}})$ is maximum if and only if $\boldsymbol{G} \propto \boldsymbol{C}$. Setting $\boldsymbol{G} = \boldsymbol{C}$ without loss of generality and replacing $\boldsymbol{G}$ and $\boldsymbol{C}$ by their definition directly yields Eq. (17). ■

We now have the final expression for the quadratic detector reaching the deflection optimum, namely,

$$\Lambda_{\boldsymbol{H}}(\boldsymbol{x}) = \boldsymbol{x}^t\boldsymbol{H}\boldsymbol{x} \quad \text{given} \quad \boldsymbol{H} = \boldsymbol{R}_n^{-1}\boldsymbol{C}_s\boldsymbol{R}_n^{-1}. \tag{20}$$

Before looking at how the approach proposed here may be practically implemented, we first give some interpretations of the detection statistic we obtained.

C. Interpretation and relation to matched filtering

With a direct calculation from its definition, we can separate $\boldsymbol{C}_s \equiv \mathsf{E}[\boldsymbol{s}^t\boldsymbol{s}]$ into two terms $\boldsymbol{C}_s = \boldsymbol{s}_m\boldsymbol{s}_m^t + \boldsymbol{R}_s$. The mean $\boldsymbol{s}_m$ can be viewed as a trend of the signal which happens systematically. In this sense, we refer to the first term as ''deterministic.'' The correlation matrix is related to the typical amplitude of the random fluctuations superimposed to the mean. For this reason, we refer to the second term as ''random.''

Injecting this expansion in Eq. (20), we obtain a similar separation of $\Lambda_{\boldsymbol{H}}(\boldsymbol{x})$:

$$\Lambda_{\boldsymbol{H}}(\boldsymbol{x}) = \Lambda_{\boldsymbol{H}}^{det}(\boldsymbol{x}) + \Lambda_{\boldsymbol{H}}^{rand}(\boldsymbol{x}), \tag{21}$$

where $\Lambda_{\boldsymbol{H}}^{det}(\boldsymbol{x}) = (\boldsymbol{s}_m^t \boldsymbol{R}_n^{-1} \boldsymbol{x})^2$ is related to the deterministic part of the signal model and $\Lambda_{\boldsymbol{H}}^{rand}(\boldsymbol{x}) = \boldsymbol{x}^t \boldsymbol{R}_n^{-1} \boldsymbol{R}_s \boldsymbol{R}_n^{-1} \boldsymbol{x}$ to its random part.

The two contributions of the detection statistic are worth further investigation. An interesting interpretation and a link to matched filtering [12] results from the reformulation in the frequency domain of $\boldsymbol{y}^t \boldsymbol{R}_n^{-1} \boldsymbol{x}$ where $\boldsymbol{x}$ and $\boldsymbol{y} \in \mathbb{R}^N$. This is the objective of the proposition whose proof is detailed in the next section.

1. Toward matched filtering

As a preamble, we define formally the whitening operation (i.e., filtering the signal with the inverse of the square root of the noise power spectral density) in the vector formalism used here. Let $\Gamma_n(f) \equiv \mathsf{E}[|N(f)|^2]$ be the power spectral density of the noise $n[k]$ and $\check{x}[k]$ be the result of the whitening of a given signal $x[k]$. We define

$$\check{X}(f) \equiv \frac{X(f)}{\sqrt{\Gamma_n(f)}}. \tag{22}$$

Clearly, $\check{x}[k]$ is the result of the convolution of the signal $x[k]$ with the whitening filter of the impulse response $w[k]$,

$$w[k] = \int_{-f_s/2}^{+f_s/2} \frac{1}{\sqrt{\Gamma_n(f)}} e^{2\pi i k f/f_s} \, df. \tag{23}$$

Using Lemma 2, we deduce that the whitening filter defined in Eq. (22) can be written in the vector formalism by

$$\check{\boldsymbol{x}} = t_s \boldsymbol{W} \boldsymbol{x}, \tag{24}$$

where $\check{\boldsymbol{x}}$ and $\boldsymbol{W}$ are as given in Lemma 2.

This expression for the whitening operation is needed in the proof of the following proposition, in which we get the vector form of the matched filtering statistic provided we take some care with the cancellation of finite size effects.

Proposition 2. Let $N_w < (N-1)/2$ be the half size of $\mathrm{supp}\{w\} = \{-N_w, \ldots, N_w\}$, i.e., the support of the impulse response of the whitening filter defined in Eq. (23), and let $y[k]$ be a signal whose whitened version has a finite size support, $\mathrm{supp}\{\check{y}\} \subset \{0, \ldots, N-1\}$, then

$$\boldsymbol{y}^t \boldsymbol{R}_n^{-1} \boldsymbol{x} = \int_{-f_s/2}^{+f_s/2} \frac{X(f)\overline{Y(f)}}{\Gamma_n(f)} \, df. \tag{25}$$

The proof of this proposition can be found in Appendix B.

2. Deterministic and random components

If we choose N large enough so that no finite size effect appears (i.e., the supports of all required signals respect the condition of Proposition 2), we can rewrite the deterministic term of the detection statistic as

$$\Lambda_{\boldsymbol{H}}^{det}(\boldsymbol{x}) = \left(\int_{-f_s/2}^{+f_s/2} \frac{X(f)\overline{S_m(f)}}{\Gamma_n(f)} \, df \right)^2. \tag{26}$$

From the following eigenexpansion of the signal correlation matrix

$$\boldsymbol{R}_s = \sum_{k=0}^{N-1} \sigma_k \boldsymbol{v}_k \boldsymbol{v}_k^t, \tag{27}$$

the random component may be expressed as

$$\Lambda_{\boldsymbol{H}}^{rand}(\boldsymbol{x}) = \sum_{k=0}^{N-1} \sigma_k (\boldsymbol{v}_k^t \boldsymbol{R}_n^{-1} \boldsymbol{x})^2. \tag{28}$$

Similarly to the deterministic component, assuming again that N is sufficiently large, it follows that

$$\Lambda_{\boldsymbol{H}}^{rand}(\boldsymbol{x}) = \sum_{k=0}^{N-1} \sigma_k \left(\int_{-f_s/2}^{+f_s/2} \frac{X(f)\overline{V_k(f)}}{\Gamma_n(f)} \, df \right)^2. \tag{29}$$

Equations (26) and (29) show that the quadratic detector with optimal deflection $\Lambda_{\boldsymbol{H}}(\boldsymbol{x})$ is closely related to the well-known technique of matched filtering [12]. The complete statistic can be equivalently implemented as a bank of $N+1$ matched filters (using the templates given by $\boldsymbol{s}_m$ and $\boldsymbol{v}_k$, $k = 0, \ldots, N-1$) whose output energies are combined with a weighted sum. Being a covariance matrix, $\boldsymbol{R}_s$ is positive definite. All eigenvalues σ_k are then real and positive numbers. In consequence, the weights (equaled to the eigenvalues) favor or attenuate the contribution of a corresponding term in the summation.

V. APPLICATION TO THE DETECTION OF GRAVITATIONAL WAVE TRANSIENTS

Because the physics driving supernovae explosions is highly nonlinear, the expected gravitational radiation is difficult to obtain through analytical means. However, numerical simulations are available [2,3,14] and a catalog of the reference waveforms associated with typical situations is accessible on the internet. The waveforms of this catalog, which we refer to as DFM (Dimmelmeier, Font, and Müller [5]), present an intrinsic diversity which motivates us to look at them as if they were produced by a single random mechanism.

Consequently, the detection problem we face is similar to the one exposed in Eq. (9) provided that the second order statistics of both signal $s(t)$ and noise $n(t)$ are known. Strictly speaking, the covariance matrix $\boldsymbol{C}_s$ of the signal is not available but if we assume that the DFM gravitational waveforms are noise-free and independent realizations of the random process $s(t)$, they can be used to get a sufficiently accurate estimate.

We can then apply the method proposed in Sec. IV B to optimally detect $s(t)$. From the signal covariance estimate

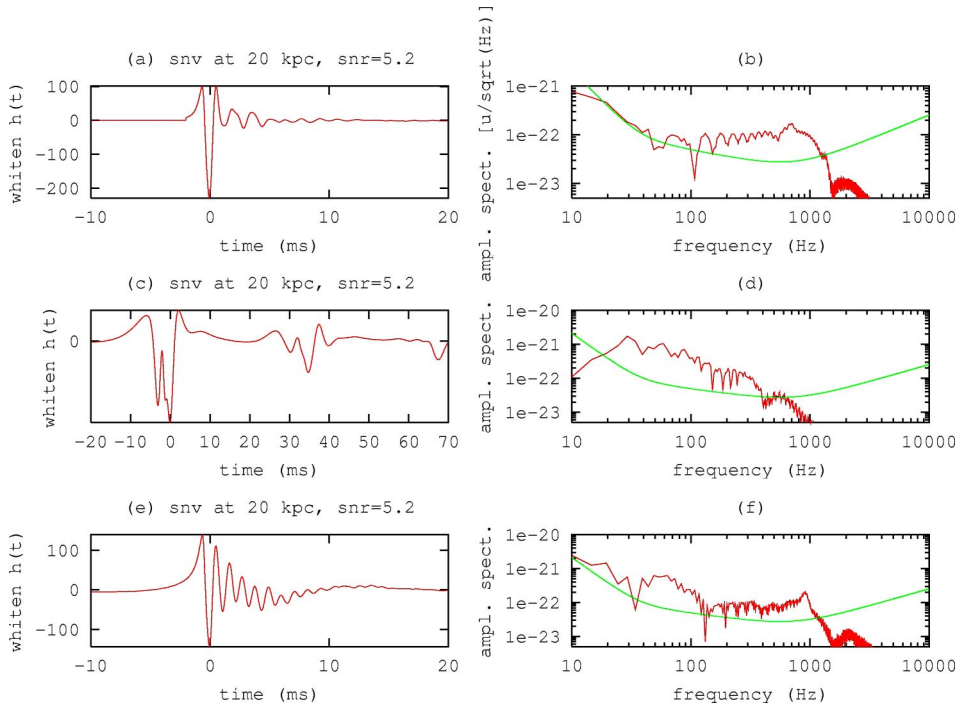

FIG. 1. (Color online) Examples of simulated gravitational transient emitted by a supernova core collapse taken from the DFM catalog. The DFM catalog of supernova gravitational transients can be separated into three types [2] which correspond to different collapse scenarios. In each of these cases, we present the waveform (column on the left-hand-side) which has been filtered by the whitening filter [in Eq. (22)] and the Fourier transform of the corresponding (nonwhitened) waveform (column on the right-hand-side in solid/red) superimposed on the objective spectral density of VIRGO noise (dashed/green). Each supernova has been placed at a distance of $d=20$ kpc from the earth, which corresponds to a signal-to-noise ratio (averaged value obtained with all waveforms in the catalog) of about $\mathcal{R}=5.2$. (a) and (b) "Regular collapse" (model reference A1B3G3). (c) and (d) "Multiple bounce collapse" (model reference A2B4G1). (e) and (f) "Rapid collapse" (model reference A1B3G5). The waveforms have clearly different shapes and characteristics (time duration and frequency bandwidth).

and a realistic noise model, we can calculate the quadratic detector with best deflection.

A. Finding the best quadratic detector

From the database publicly available on the internet [2], we have extracted the $N_z=25$ waveforms, which we have resampled at the constant rate of $f_s=20$ kHz. Actually two sets of waveforms can be used: the one drawn from a Newtonian simulation setup and the other from a fully relativistic code. We preferred to use the latter since the result is likely to be closer to reality. The waveforms are stored in column vectors $\{\boldsymbol{z}_k\}_{k=0,\ldots,N_z-1}$ of size $N=2251$ rows. This corresponds to a maximum burst duration of about 112.5 ms.

For each signal, we fixed the time axis origin to be located at the minimum of the largest negative bump of the whitened signal (in most cases, this is synchronized with the core bounce [2]); precisely, $t_0=\mathrm{argmin}_j(\check{z}_k[j]),\forall k$. Figure 1 presents three individuals of each type of supernova taken from the processed DFM catalog.

Assuming that the DFM gravitational waveforms are noise-free and independent realizations of the random process $s(t)$, we use these waveforms to estimate the covariance matrix of $s(t)$. This is done with the following empirical unbiased estimator:

$$\hat{\boldsymbol{C}}_s=\frac{1}{N_z}\sum_{k=0}^{N_z-1}\boldsymbol{z}_k\boldsymbol{z}_k^t. \tag{30}$$

For simplicity, we consider that the noise power spectrum is known *a priori* and is given by the expected sensitivity curve for the planned detectors. (Note that the noise correlation matrix may also be estimated from "noise only" data streams.) We restrict our study to the case of the VIRGO detector using the noise model available at the address [15]. Extensions to other large scale interferometers are straightforward. We can get the noise correlation matrix from the power spectrum by applying an inverse Fourier transformation. From the obtained values of $\hat{\boldsymbol{C}}_s$ and $\boldsymbol{R}_n$, we deduce the kernel of the best quadratic detector as given in Eq. (17).

This computation requires $O(N^3)$ operations to get the inverse of $\boldsymbol{R}_n$ and we need roughly the same number of operations to make the two matrix multiplications in Eq. (17). A more computationally efficient algorithm may be used. With this aim in view, we process the whole catalog of waveforms with the operation $\check{\boldsymbol{z}}_k\equiv\boldsymbol{R}_n^{-1}\boldsymbol{z}_k$. Roughly speaking, the multiplication by $\boldsymbol{R}_n^{-1}$ is equivalent to whitening the signal twice as suggested by the factorization of $\boldsymbol{R}_n^{-1}$ in Eq. (B10). The more precise relation $\check{Z}_k(f)=t_sZ_k(f)/\Gamma_n(f)$ can be stated by applying the result in Proposition 2 with $y[j]=\delta_{jm}$ for any $m\in\{0,\ldots,N-1\}$ and $x[j]=z_k[j]$. This

double whitening operation amounts to filtering the signal in the frequency bandwidth of interest where the noise is low and removing the remaining part where the noise is large. From Eqs. (20) and (30), it is easily shown that the objective kernel can be computed directly using the modified catalog with the relation

$$\hat{\boldsymbol{H}}=\frac{1}{N_z}\sum_{k=0}^{N_z-1}\check{\boldsymbol{z}}_k\check{\boldsymbol{z}}_k^t. \tag{31}$$

Since the correlation matrix $\boldsymbol{R}_n$ is Toeplitz symmetric, the computation of $\check{\boldsymbol{z}}_k$ is equivalent to solving a $N_z\times N_z$ Toeplitz linear system. This can be done efficiently with a variety of fast $O(N^2)$ algorithms. We selected and applied the Levinson algorithm [13].

The total gravitational energy radiated during the collapse varies according to the selected models. The peak amplitudes of the waveforms of the DFM catalog have values ranging in an interval as large as one order of magnitude. To ensure that all types of supernovae are treated equitably in the sum of Eq. (31), we scale all $\check{\boldsymbol{z}}_k$ by dividing by the expected signal-to-noise ratio defined as

$$\mathcal{R}_k\equiv\left(\int_{-f_s/2}^{f_s/2}|Z_k(f)|^2/\Gamma_n(f)df\right)^{1/2}. \tag{32}$$

A practical expression of this quantity can be obtained by first deducing from Proposition 2 that $\mathcal{R}_k=(\boldsymbol{z}_k^t\boldsymbol{R}_n^{-1}\boldsymbol{z}_k)^{1/2}$ and using the definition of the whitening operator in Eq. (24) and its relation to $\boldsymbol{R}_n$ in Eq. (B10), thus leading to $\mathcal{R}_k=\|\check{\boldsymbol{z}}_k\|_2/\sqrt{f_s}$ where $\|\boldsymbol{x}\|_2^2\equiv\boldsymbol{x}^t\boldsymbol{x}$ defines the l_2 norm.

At this point, it is worth noting that Eq. (31) implements a learning scheme which extracts systematically the necessary information from a (possibly large) database of reference waveforms and combines it in order to find (within a fixed family) the detector which maximizes the signal-to-noise ratio. Although in a different context, a similar idea has been used in [16] for the adaptive tuning of a network of gravitational wave interferometers.

B. Approximated detector

A close look to the detector kernel $\hat{\boldsymbol{H}}$ indicates that it is degenerate (its rank is much smaller than N). There are two reasons for that. First, as a result of the linear combination of $N_z\ll N$ rank-1 matrices [see Eq. (31)], the rank of the kernel cannot exceed N_z. The second reason is the fact that the waveforms of the DFM catalog have common features in their shapes (e.g., fundamental oscillation frequency, time duration, etc.). This causes the matrices $\check{\boldsymbol{z}}_k\check{\boldsymbol{z}}_k^t$ to share some linear dependency.

Precisely, it means that the kernel may be decomposed along a small number of preferred directions of the measurement space. The most adequate basis to check this is formed by the generalized eigenvectors of $\hat{\boldsymbol{C}}_s$ and $\boldsymbol{R}_n$ as explained in the following section. We show that the kernel degeneracy may be used to simplify the detector and reduce its computational complexity.

1. Truncating to principal directions

The vector $\boldsymbol{u}$ and scalar γ are, respectively, the generalized eigenvector and value of $\hat{\boldsymbol{C}}_s$ and $\boldsymbol{R}_n$ if the following equation is satisfied [13]:

$$\hat{\boldsymbol{C}}_s\boldsymbol{u}=\gamma\boldsymbol{R}_n\boldsymbol{u}. \tag{33}$$

Since $\boldsymbol{R}_n$ is a definite positive matrix, it can be decomposed using Cholesky factorization [13] as a product of invertible and triangular matrices, namely, $\boldsymbol{R}_n=\boldsymbol{T}_n\boldsymbol{T}_n^t$. Multiplying to the left both sides of Eq. (33) by $\boldsymbol{T}_n^{-1}$, the generalized eigenproblem above turns out to be equivalent to the standard one given by

$$\boldsymbol{\Gamma}\boldsymbol{v}=\gamma\boldsymbol{v}, \tag{34}$$

provided that $\boldsymbol{\Gamma}=\boldsymbol{T}_n^{-1}\hat{\boldsymbol{C}}_s\boldsymbol{T}_n^{-t}$ and $\boldsymbol{v}=\boldsymbol{T}_n^t\boldsymbol{u}$. Consequently, the matrix $\boldsymbol{\Gamma}$ may be expanded along its eigendirections $\{\boldsymbol{v}_k\}_{k=0,\ldots,N-1}$, namely,

$$\boldsymbol{\Gamma}=\sum_{k=0}^{N-1}\gamma_k\boldsymbol{v}_k\boldsymbol{v}_k^t. \tag{35}$$

Since we have $\hat{\boldsymbol{H}}=\boldsymbol{T}_n^{-1}\boldsymbol{\Gamma}\boldsymbol{T}_n^{-t}$, the previous expansion yields that of the detector kernel along the generalized eigenbasis defined in Eq. (33):

$$\hat{\boldsymbol{H}}=\sum_{k=0}^{N-1}\gamma_k\boldsymbol{u}_k\boldsymbol{u}_k^t. \tag{36}$$

Combining adequately a Cholesky and a Schur decomposition [13,17], we computed the solutions of Eq. (33). The eigenvalues are sorted in decreasing order $\gamma_0>\gamma_1>\cdots>\gamma_{N-1}$ and presented in Fig. 2. It appears clear that the resulting spectrum is essentially dominated by the first few eigenvalues.

A consequence is that the sum in Eq. (36) can be fairly approximated by the summation truncated to the first terms. Let $n<N$ be the truncation limit; we get the kernel

$$\tilde{\boldsymbol{H}}_n\equiv\sum_{k=0}^{n-1}\gamma_k\boldsymbol{u}_k\boldsymbol{u}_k^t, \tag{37}$$

which we use to compute the approximated detection statistic

$$\Lambda_{\tilde{\boldsymbol{H}}_n}(\boldsymbol{x})=\sum_{k=0}^{n-1}\gamma_k(\hat{\boldsymbol{u}}_k^t\boldsymbol{x})^2\approx\Lambda_{\hat{\boldsymbol{H}}}(\boldsymbol{x}). \tag{38}$$

The value of the truncation index n is essentially related to the intrinsic complexity of the initial waveform database. In the case of interest, n is much smaller than N by several orders of magnitude (a nonempirical choice of n is described in the next section), and its value remains stable when N increases. In consequence, the approximated statistic (38) is computed with $O(N^2)$ floating point operations versus a total cost of $O(N^3)$ in the nonapproximated case.

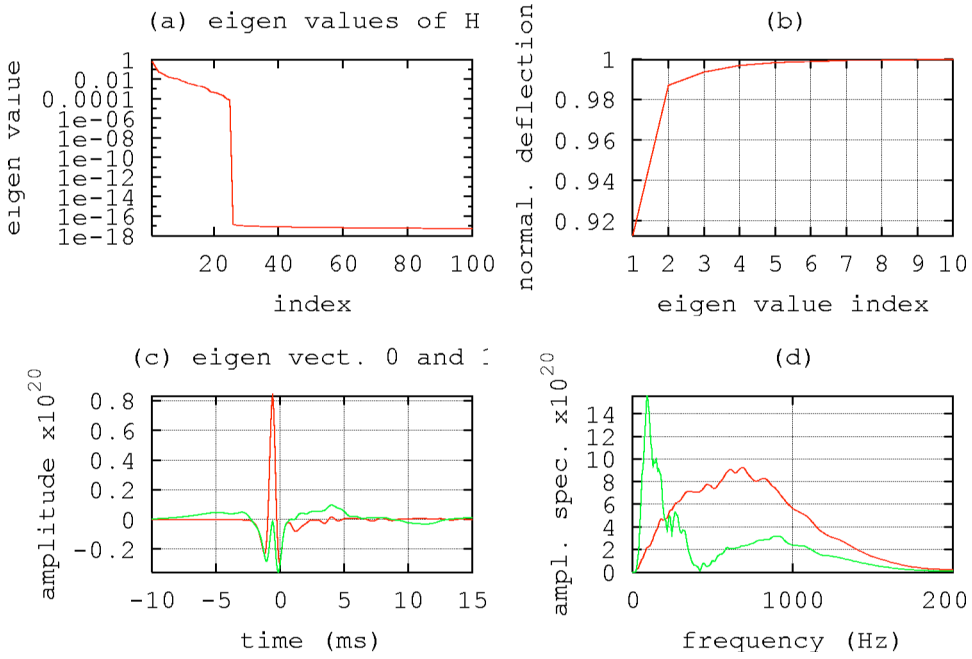

FIG. 2. (Color online) Generalized eigenvalues and eigenvectors of $\hat{\boldsymbol{H}}$. With these plots, we summarize the information carried by the generalized eigenvalues γ_k and vectors $\boldsymbol{u}_k$ defined by Eq. (33). The generalized spectrum of $\hat{\boldsymbol{H}}$ is largely dominated by the first few eigenvalues [the first 100 are shown in (a)]. This degeneracy can be used to simplify the statistic by truncating the eigenexpansion (36) to the first terms [see Eq. (37)]. The number of terms to keep is given by the amount of deflection we tolerate losing due to this truncation. This is indicated in (b) where we see that keeping the $n=2$ dominating eigenvectors is sufficient to reach $\approx 99\%$ of the optimal deflection. These two eigenvectors $u_0[k]$ (solid/red) and $u_1[k]$ (dashed/green) are presented in (c) with their corresponding Fourier transform in (d). From its shape, it appears that $u_0[k]$ grabs most of the peak occurring in the bounce phase of the supernova (this represents about 91% of the total deflection) and $u_1[k]$ the few oscillations of the ringdown phase (which are the 8% remaining). It is worth noting that both $U_0(f)$ and $U_1(f)$ are nonzero in frequency bands ranging from 200 Hz to 1 kHz and from 50 Hz to 100 Hz.

2. Loss in deflection due to approximation

The truncation to a few eigendirections causes $\Lambda_{\tilde{\boldsymbol{H}}_n}(\boldsymbol{x})$ to be suboptimal, i.e., the resulting deflection is smaller than the one obtained with $\Lambda_{\hat{\boldsymbol{H}}}(\boldsymbol{x})$. Two interesting questions are (i) how much deflection do we lose? and (ii) can we adjust n so that the loss is acceptable? To address these questions, it is convenient to define the *loss in deflection*: $l_n \equiv d^2(\Lambda_{\tilde{\boldsymbol{H}}_n})/d^2(\Lambda_{\hat{\boldsymbol{H}}})$. This index whose values are between 0 and 1 measures the degree of "suboptimality" of the truncated detector.

Replacing $\boldsymbol{A}$ in the expression of the deflection obtained in Lemma 3 with the truncated sum in Eq. (37), and using the fact that $\{\boldsymbol{u}_k\}_{k=0,\ldots,N-1}$ form a basis which diagonalizes simultaneously $\boldsymbol{R}_n$ and $\hat{\boldsymbol{C}}_s$, i.e., more precisely $\boldsymbol{u}_k^t \boldsymbol{R}_n \boldsymbol{u}_j = \delta_{jk}$ and $\boldsymbol{u}_k^t \hat{\boldsymbol{C}}_s \boldsymbol{u}_j = \gamma_k \delta_{jk}$ (see [13] for details), a straightforward calculation leads to

$$d^2(\Lambda_{\tilde{\boldsymbol{H}}_n}) = \sum_{k=0}^{n-1} \gamma_k^2 . \tag{39}$$

This results holds also for $n=N$, yielding the maximum value of the deflection, which we denote by $d_{max} \equiv d^2(\Lambda_{\hat{\boldsymbol{H}}}) = \Sigma_{k=0}^{N-1} \gamma_k^2$. The loss in deflection can then be expressed as

$$l_n = \sum_{k=0}^{n-1} \frac{\gamma_k^2}{d_{max}} \tag{40}$$

and is presented in Fig. 2(b). We conclude that, with $n=2$, i.e., keeping the first two leading eigendirections, the truncated detector has a performance index of about 99% (1% from optimum). Figure 2(c) details the waveforms of the two leading eigenvectors.

Provided that $u_0[k]$ and $u_1[k]$ have support in $\{0,\ldots,N-1\}$ [this is the case in the example presented here; see Fig. 2(c)], we can apply Lemma 1 and get the truncated detector (38) expressed in the frequency domain:

$$\Lambda_{\tilde{\boldsymbol{H}}}(\boldsymbol{x}) = \gamma_0 f_s^2 \left| \int_{-f_s/2}^{f_s/2} X(f) \overline{U_0(f)}\, df \right|^2 + \gamma_1 f_s^2 \left| \int_{-f_s/2}^{f_s/2} X(f) \overline{U_1(f)}\, df \right|^2 . \tag{41}$$

We conclude that the detection statistic is computed by first selecting the interesting frequency content of the spectrum of the observed data with the two [bandpass, as shown by Fig. 2(d)] filters $U_0(f)$ and $U_1(f)$ and then combining the energy of the filter outputs with a weighted sum. The weight

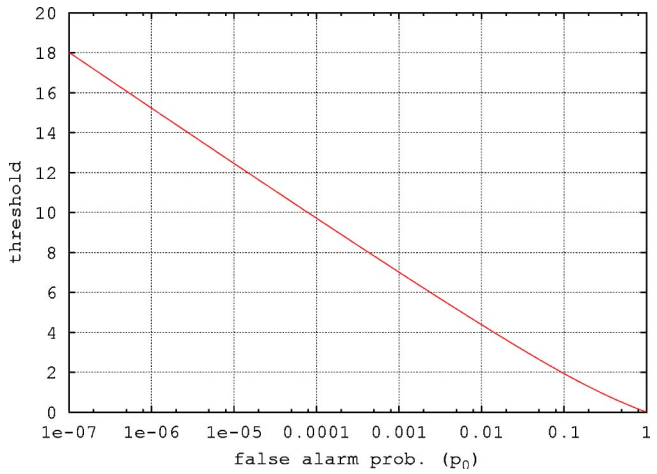

FIG. 3. (Color online) Detection threshold satisfying a given requirement on the false alarm rate. This plot is the diagram (solid/red) of the function $\eta = F^{-1}(1-p_0)$ relating the detection threshold η to apply in order to get a fixed false alarm rate p_0. The function $F(\cdot)$ is the cumulative probability function of the statistic $\Lambda_{\tilde{H}}$ in the H_0 null hypothesis ("noise only"). It is the integral of the PDF in Eq. (43) where we fixed $\gamma_0 \approx 0.627$ and $\gamma_1 \approx 0.181$. Typical values for p_0 are ranging in the interval between 10^{-7} and 10^{-5} (this roughly gives a false alarm rate of a few per 10 min to a few per day) which correspond to values of the threshold between 11 and 18. In this range of interest, the fit $\eta \approx -11/4 \log_{10}(p_0) - 5/4$ (dashed/green) gives a satisfactory approximation (in this region, the two curves superimpose).

parameters can be interpreted as "confidence coefficients" in finding a (supernova) signal in the corresponding frequency bands.

In other words, the proposed method extracts systematically from a database of reference signals the frequency bands which need to be considered in order to maximize the deflection.

3. Detection threshold and false alarm probability

Under the noise only (H_0) assumption, the detector (38) is a finite sum of the squares of the random variables defined by $n_k \equiv \boldsymbol{u}_k^t \boldsymbol{n}$:

$$\Lambda_{\tilde{\boldsymbol{H}}}(\boldsymbol{n}) = \sum_{k=0}^{n-1} \gamma_k n_k^2 . \tag{42}$$

These variables can be easily shown to be Gaussian and zero mean. Furthermore, since $\{\boldsymbol{u}_k\}_{k=0,\ldots,N-1}$ diagonalizes the noise correlation $\boldsymbol{R}_n$, we have $\mathsf{E}[n_j n_k] = \delta_{jk}$, from which we conclude that $\{n_k\}_{k=0,\ldots,n-1}$ is a sequence of independent and identically distributed Gaussian variables of PDF $\mathcal{N}(0,1)$.

Let $f(\lambda)$ be the PDF of $\Lambda_{\tilde{\boldsymbol{H}}}(\boldsymbol{n})$ when there is only noise. In the case where $n=2$ eigenvectors are sufficient to get a good approximation of the optimal detector, we have for $\lambda > 0$ [18]

$$f(\lambda) = \frac{1}{2\sqrt{\gamma_0 \gamma_1}} \exp\left[-\frac{1}{4}\left(\frac{1}{\gamma_1} + \frac{1}{\gamma_0}\right)\lambda \right] I_0\left[\frac{1}{4}\left(\frac{1}{\gamma_1} - \frac{1}{\gamma_0}\right)\lambda\right] \tag{43}$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind [19] and $f(\lambda) = 0$ if $\lambda \leqslant 0$.

Integrating the PDF in Eq. (43), we obtained the cumulative probability function $F(\lambda) \equiv \int_0^\lambda f(\nu) d\nu = \mathbb{P}(\Lambda_{\tilde{\boldsymbol{H}}}(\boldsymbol{x}) < \lambda | H_0)$. The threshold ensuring a given false alarm probability p_0 is thus given by $\eta = F^{-1}(1-p_0)$. This function is presented in Fig. 3 in the range of useful values of p_0 and using the first two leading eigenvalues of $\hat{\boldsymbol{H}}$, namely, $\gamma_0 \approx 0.627$ and $\gamma_1 \approx 0.181$.

4. Basic check of the detection procedure

With the expression of the statistic in Eq. (41) and the value of the threshold (see the previous section), we have now in hand all the ingredients required to implement the decision test. Checks can be made to verify whether the procedure obtained effectively detects supernovae when placed at realistic distances.

To do so, let us first define $\lambda_k \equiv \Lambda_{\tilde{\boldsymbol{H}}}(\boldsymbol{z}_k)$ for $k = 0, \ldots, N_z - 1$. The waveforms $\boldsymbol{z}_k$ from the DFM catalog are all scaled to the same signal-to-noise ratio $\mathcal{R}_k = 5.2$ as defined in Eq. (32). This means that the supernovae are located at different distances from the earth, the average value being of about the largest galactic distance $d \approx 20$ kpc. We set the false alarm probability to $p_0 = 10^{-5}$, so that we are led to a threshold $\eta \approx 12$ (see Fig. 3). The average contribution of the noise to the statistic can be calculated to be $\mathsf{E}[\Lambda_{\tilde{\boldsymbol{H}}}(\boldsymbol{n})] = 1$ [see Eq. (A2) and the normalization explained in Sec. V A] so that we conclude that a supernova is detected if $\lambda_k \gtrsim 11$.

In the present setup (i.e., with the $N_z = 25$ available waveforms), 16 waveforms (of type I or I/II) pass the test (with $\lambda_k \approx 15$) and 9 waveforms fail: two of them with $\lambda_k \approx 9$ are type III supernovae, the rest, $\lambda_k \approx 4$, are type II supernovae. This result is expected in the sense that the method relies on the assumption that all waveforms in the learning database are from a single random process. This assumption is not realistic for the very different physical scenarios that lead to type I or type II/III supernovae.

This indicates that at least two statistics have to be used to take the large shape differences between type I and type II/III waveforms into account. To ensure reliability, this has to be done with a finer sampling of the parameter space (i.e., with a larger number of waveforms, especially of type II and III). This approach is currently under investigation [20].

5. Time running implementation

Until now, we have only considered the statistical test of the presence of a supernova transient at a given time. The date of arrival of the gravitational wave being unknown, we must apply the detection procedure at any given time instants. To do this, we select the N data samples starting from a given time index m, namely, $\boldsymbol{x}_m \equiv (x[m+k], k = 0, \ldots, N-1)^t$. We compute the detection statistic $\Lambda_{\boldsymbol{H}}(\boldsymbol{x}_m) = \boldsymbol{x}_m^t \boldsymbol{H} \boldsymbol{x}_m$

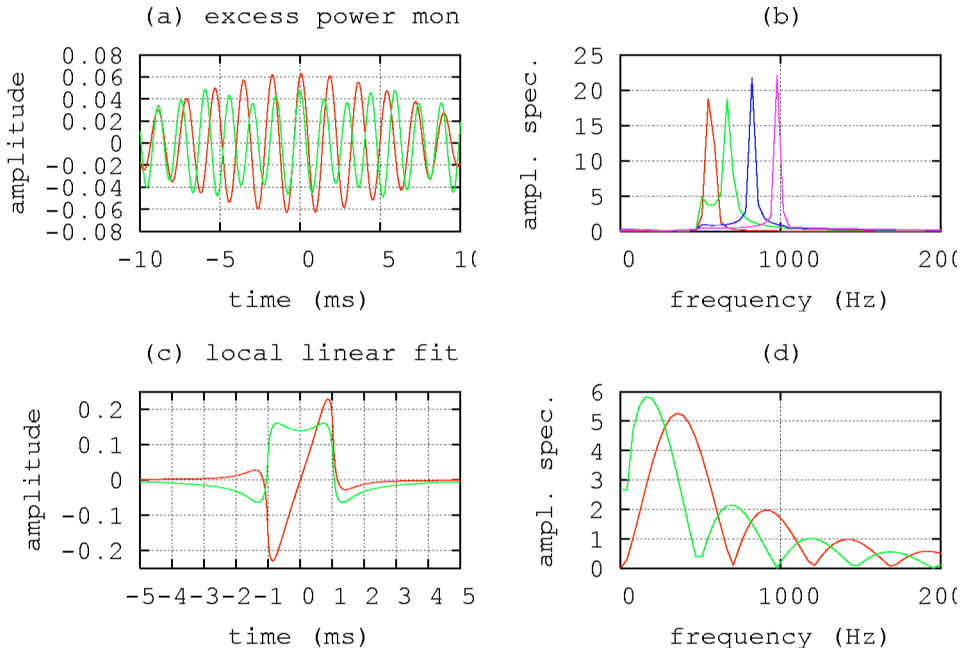

FIG. 4. (Color online) Generalized eigenvectors of $\boldsymbol{H}_{eps}$ and $\boldsymbol{H}_{alf}$. In this figure, we present the generalized eigenvectors (left-hand-side column) of the detection kernel used by the excess power (EPS) and the local linear fit (ALF) statistics and their respective Fourier transform (right-hand-side column). Concerning the EPS statistic (top row), we chose a time window with $N=512$ samples, corresponding to a duration $\delta t \approx 25$ ms provided a sample rate of $f_s=20$ kHz, and a frequency window of $\delta f=500$ Hz centered around $f_0=750$ Hz. This gives a time-frequency volume [6] of $V \approx 2\times 25$ ms$\times 500$ Hz$=25$. These parameters lead to the limit indices $j_-=13$ and $j_+=26$ in the sum (45). The detector kernel has about 14 large generalized eigenvalues which we sort in decreasing order. The corresponding eigenvectors form a set of bandpass filters (width ≈ 80 Hz) covering uniformly the selected frequency window. The waveforms of the first and the fourth eigenvectors are plotted in (a), and we show in (b) the spectra of the eigenvectors of rank 1, 4, 8, and 12. The linear fit done by the ALF statistic (bottom row) is computed using $N=40$ samples of data (i.e., in a time window of 2 ms) which is the best time window duration found in [7] for supernova transients. The two eigenvectors $\boldsymbol{W}^t\boldsymbol{t}$ (solid/red) and $\boldsymbol{W}^t\boldsymbol{1}$ (dashed/green) are shown in (c) and their respective Fourier transforms in (d). It appears that first filter selects frequencies in 350 ± 200 Hz, and the second in 155 ± 135 Hz. In a sense, although the bandwidths are not exactly the same, this filter bank is similar to the deflection optimal detector.

for increasing and equally spaced values of $m=0,\delta_m,\ldots$. This is similar to selecting the data with a time sliding window.

Using the approximated statistic expressed as in Eq. (41) and noting that $X_m(f)=X(f)e^{-2\pi imf/f_s}$, we get

$$\Lambda_{\tilde{\boldsymbol{H}}}(\boldsymbol{x}_m)=\gamma_0 f_s^2(y_0[m])^2+\gamma_1 f_s^2(y_1[m])^2, \tag{44}$$

where $y_{(0,1)}[m]=\int_{-f_s/2}^{+f_s/2}X(f)\overline{U_{(0,1)}(f)e^{-2\pi imf/f_s}}\,df$ are obtained by passing the signal through a time-invariant linear filter. Assuming $U_0(f)$ and $U_1(f)$ are stored in memory, the computation of $y_0[m]$ and $y_1[m]$ can be efficiently computed with the fast Fourier transform (and inverse) algorithm.

VI. RELATION TO OTHER DETECTION TECHNIQUES

We have shown in Sec. IV C that the quadratic detector with optimal deflection can be related to matched filtering. In fact, many of the methods for transient detection available in the literature, e.g., [6,7,9–11], belong to the class of quadratic detectors defined in Definition 1. The vector formalism used here constitutes a general framework in which all these methods can be reformulated and easily compared. Considering that the noise model remains the same as the one we used in the previous section, we get the shape of the kernel used by the two contributions described in [6] and [7] to which we limit the investigation. With this "back-engineering" approach, we can retrieve the *a priori* assumption on the signal covariance needed for the considered detector to have optimal deflection. We make this comparison by looking at the generalized eigenbasis of the kernel obtained.

A. Excess power statistic

The basic idea is to monitor the power in one (or several) given frequency band $f_0\pm\delta f/2$ (similarly to Sec. IV A). Let $X[j]\equiv\Sigma_{k=0}^{N-1}x[k]e^{-2\pi ijk/N}$ be the discrete Fourier transform (DFT). The excess power statistic presented in [6] reads

$$\Lambda_{eps}(\boldsymbol{x})=\sum_{j=j_-}^{j_+}\frac{|X[j]|^2}{\Gamma_n(jf_s/N)}, \tag{45}$$

where the limit indices are defined as $j_\pm=N(f_0\pm\delta f/2)/f_s$.

The DFT can be reexpressed as a scalar product $X[j]=\overline{\boldsymbol{f}_j^t}\boldsymbol{x}$ with the Fourier exponentials $\boldsymbol{f}_j\equiv(e^{2\pi ijk/N},k=0,\ldots,N-1)^t$. It is straightforward to show that the above statistic is a quadratic detector as in Definition 1 with the kernel

$$\boldsymbol{H}_{eps} = \sum_{j=j_-}^{j_+} \frac{\boldsymbol{f}_j \overline{\boldsymbol{f}_j^t}}{\Gamma_n(j f_s/N)}. \tag{46}$$

Roughly speaking, assuming that the noise power spectral density is ''flat'' in the selected frequency bandwidth, $\{\boldsymbol{f}_j\}_{j=j_-,\ldots,j_+}$ diagonalizes $\boldsymbol{R}_n$. In this case, the kernel of the excess power statistic has the number $(j_+ - j_- + 1) = V/2$ generalized eigenvectors. Figure 4 presents some of these eigenvectors and their Fourier transforms.

B. Linear fit filter

The detection statistic $\Lambda_{alf}(\boldsymbol{x})$ is obtained [7] from a local linear fit of the whitened signal $\check{x}$. The mean square rule yields the two parameters of the fit:

$$a = \frac{\langle t\check{x}\rangle - \langle t\rangle^2}{\langle t^2\rangle - \langle t\rangle^2}, \tag{47}$$

$$b = \langle \check{x}\rangle - a\langle t\rangle \tag{48}$$

which are orthonormalized, squared, and combined to get $\Lambda_{alf}(\boldsymbol{x})$.

It turns out to be convenient to set the time origin at the center of the data chunk, which we assume to have an odd number N of samples. We can do this with no loss of generality. In this setup, the fit parameters are given by the scalar products $a = \boldsymbol{t}^t\check{\boldsymbol{x}}/\|\boldsymbol{t}\|_2^2$ and $b = \boldsymbol{1}^t\check{\boldsymbol{x}}$ where we defined $\boldsymbol{t} = (-L, \ldots, L)^t$, $L \equiv (N-1)/2$ being the half size of the data chunk, and $\boldsymbol{1} = (1, \ldots, 1)^t$. After the orthonormalization and combination, the detection statistic appears to be a quadratic detector as in Definition 1, of kernel

$$\boldsymbol{H}_{alf} = t_s^2 \boldsymbol{W}^t \left(\frac{\boldsymbol{t}\boldsymbol{t}^t}{\|\boldsymbol{t}\|_2^2} + \frac{\boldsymbol{1}\boldsymbol{1}^t}{\|\boldsymbol{1}\|_2^2} \right) \boldsymbol{W} \tag{49}$$

where $\boldsymbol{W}$ is the whitening matrix defined in Eq. (24).

It can be easily shown that $\boldsymbol{W}^t\boldsymbol{1}$ and $\boldsymbol{W}^t\boldsymbol{t}$ are the two generalized eigenvectors of $\boldsymbol{H}_{alf}$ associated with the eigenvalue 1 (this is the only nonzero eigenvalue). They are presented in Fig. 4. In this degenerate case, similar calculations as the ones done in Sec. V B 3 yields the PDF of $\Lambda_{alf}(\boldsymbol{x})$ in the noise only case [18], namely,

$$f_{alf}(\lambda) = \frac{1}{2} e^{-\lambda/2}, \tag{50}$$

from which the threshold can be obtained for a given false alarm probability. This is a complementary contribution to the analysis made in [7] about the local linear fit method.

VII. CONCLUSIONS

Quadratic detectors (i.e., statistics that are bilinear functions of the data) can be essentially viewed as a filtering of the data through a selection of frequency bands, the power of the filtered data being further linearly combined. We introduced a method which systematically extracts from a complicated and possibly large database of target signals the important features that need to be considered in order to design this filter bank and choose the parameters for the energy combination. In the context of the detection of supernovae core collapses, we show that the method gives the intuitively appealing result of a filter bank composed of two elements (the one selecting the bounce pulse and the other the few oscillations of the ringdown phase) whose output powers are combined in such a way as to favor the bounce (which is the most energetic part of the signal). We checked that the procedure effectively detects type I supernovae whereas the method needs to be improved [20] for type II and III supernovae because of the large shape differences of their waveforms.

The scope of the approach presented here is general. The algorithm can be adapted to other problems of the same type (for instance, the detection of the final merger part of a binary black coalescence or of nonstationary noise interferences) provided that a sufficient number of training waveforms are available.

APPENDIX A: PROOF OF LEMMA 3 OF SEC. IV B

Proof. The proof of this lemma may be found in other articles [4,21] for zero mean signals and infinite vector spaces. Here, we give the proof for noncentral signals (i.e., $\boldsymbol{s}_m \neq \boldsymbol{0}$) and in the case of discrete signals forming vectors of finite size.

We compute the first two statistical moments of $\Lambda_{\boldsymbol{A}}(\boldsymbol{x})$ under hypotheses H_0 and H_1. Starting conditionally at H_0, we have

$$\mathsf{E}[\Lambda_{\boldsymbol{A}}(\boldsymbol{x})|H_0] = \mathsf{E}[\boldsymbol{n}^t\boldsymbol{A}\boldsymbol{n}]. \tag{A1}$$

Using the identity $\boldsymbol{x}^t\boldsymbol{x} = \mathrm{tr}\{\boldsymbol{x}\boldsymbol{x}^t\}$, this can be reduced to

$$\mathsf{E}[\Lambda_{\boldsymbol{A}}(\boldsymbol{x})|H_0] = \mathrm{tr}\{\boldsymbol{A}\mathsf{E}[\boldsymbol{n}\boldsymbol{n}^t]\} = \mathrm{tr}\{\boldsymbol{A}\boldsymbol{R}_n\}. \tag{A2}$$

Under the ''signal + noise'' H_1 hypothesis, we expand the quadratic form

$$\mathsf{E}[\Lambda_{\boldsymbol{A}}(\boldsymbol{x})|H_1] = \mathsf{E}[\boldsymbol{s}^t\boldsymbol{A}\boldsymbol{s} + \boldsymbol{s}^t\boldsymbol{A}\boldsymbol{n} + \boldsymbol{n}^t\boldsymbol{A}\boldsymbol{s} + \boldsymbol{n}^t\boldsymbol{A}\boldsymbol{n}]. \tag{A3}$$

Then, with the identity mentioned previously, we can simplify the autoterms in the expansion $\mathsf{E}[\boldsymbol{s}^t\boldsymbol{A}\boldsymbol{s}] = \mathrm{tr}\{\boldsymbol{A}\boldsymbol{C}_s\}$ with $\boldsymbol{C}_s \equiv \mathsf{E}[\boldsymbol{s}^t\boldsymbol{s}]$ while the cross terms vanish: $\mathsf{E}[\boldsymbol{n}^t\boldsymbol{s}] = \mathsf{E}[\boldsymbol{n}^t]\boldsymbol{s}_m = 0$ and $\mathsf{E}[\boldsymbol{n}^t\boldsymbol{A}\boldsymbol{s}] = \mathsf{E}[\boldsymbol{s}^t\boldsymbol{A}\boldsymbol{n}] = 0$, which yields

$$\mathsf{E}[\Lambda_{\boldsymbol{A}}(\boldsymbol{x})|H_1] = \mathrm{tr}\{\boldsymbol{A}(\boldsymbol{C}_s + \boldsymbol{R}_n)\}. \tag{A4}$$

For the general expressions of higher order moments of Gaussian quadratic forms in [18], we get the variance under H_0:

$$\mathrm{var}[\Lambda_{\boldsymbol{A}}(\boldsymbol{x})|H_0] = \mathrm{var}(\boldsymbol{n}^t\boldsymbol{A}\boldsymbol{n}) = 2\,\mathrm{tr}\{(\boldsymbol{A}\boldsymbol{R}_n)^2\}. \tag{A5}$$

The combination of all these ingredients leads to the result. ∎

APPENDIX B: PROOF OF PROPOSITION 2 OF SEC. IV C 1

Proof. We first compute $\mathsf{E}[\check{\boldsymbol{n}}\check{\boldsymbol{n}}^t]$ and get a first expression using Eq. (24):

$$\mathsf{E}[\check{\boldsymbol{n}}\check{\boldsymbol{n}}^t]=t_s^2\boldsymbol{W}\boldsymbol{R}_n\boldsymbol{W}^t. \tag{B1}$$

A second expression is obtained by writing each term of the considered matrix in the Fourier domain. The component located at the jth row and kth column may be expressed as

$$\mathsf{E}[\check{n}[j]\check{n}[k]] = \int\int_{-f_s/2}^{+f_s/2}\frac{\mathsf{E}[N(f)\overline{N(f')}]}{\sqrt{\Gamma_n(f)\Gamma_n(f')}}e^{2\pi i(jf-kf')/f_s}\,df\,df'. \tag{B2}$$

The integrand can be worked out by applying the Wiener-Khinchine theorem [12]:

$$\mathsf{E}[N(f)\overline{N(f')}]=\delta(f-f')\Gamma_n(f) \tag{B3}$$

where $\delta(f)=t_s\Sigma_{k=-\infty}^{+\infty}e^{-2\pi ikf/f_s}$.

The function δ acts on the elements of the set $\mathcal{P}(f_s)$ (containing the periodic functions of period f_s) in the same way the Dirac operator acts on functions of $L^2(\mathbb{R})$. Let Φ be a test function of $\mathcal{P}(f_s)$ and ϕ its inverse Fourier transform. We have

$$\int_{-f_s/2}^{+f_s/2}\Phi(f)\,\delta(f)\,df=t_s\sum_{k=-\infty}^{+\infty}\phi[k]=\Phi(0). \tag{B4}$$

Note also that δ is f_s-periodic, i.e., $\delta(f+f_s)=\delta(f)$, for all f. The proofs of these two properties of δ are left to the reader.

With Eq. (B3), we have

$$\mathsf{E}[\check{n}[j]\check{n}[k]]=\int_{-f_s/2}^{+f_s/2}\frac{e^{2\pi ijf/f_s}}{\sqrt{\Gamma_n(f)}}\times\left(\int_{-f_s/2}^{+f_s/2}\sqrt{\Gamma_n(f')}\,\delta(f-f')\times e^{-2\pi ikf'/f_s}\,df'\right)df, \tag{B5}$$

which simplifies with the change of variable $u=f-f'$ and using the periodicity of the functions δ and Γ_n,

$$\mathsf{E}[\check{n}[j]\check{n}[k]]=\int_{-f_s/2}^{+f_s/2}\frac{e^{2\pi i(j-k)f/f_s}}{\sqrt{\Gamma_n(f)}}\times\left(\int_{-f_s/2}^{+f_s/2}\sqrt{\Gamma_n(f-u)}\,\delta(u)e^{2\pi iku/f_s}\,du\right)df, \tag{B6}$$

leading finally, using Eq. (B4), to

$$\mathsf{E}[\check{n}[j]\check{n}[k]]=\int_{-f_s/2}^{+f_s/2}e^{2\pi i(j-k)f/f_s}\,df, \tag{B7}$$

or equivalently to

$$\mathsf{E}[\check{n}[j]\check{n}[k]]=f_s\delta_{jk}, \tag{B8}$$

where δ_{jk} is the Kronecker symbol (by definition, $\delta_{jk}=1$ if $j=k$, and 0 otherwise).

We conclude that

$$\mathsf{E}[\check{\boldsymbol{n}}\check{\boldsymbol{n}}^t]=f_s\boldsymbol{I}. \tag{B9}$$

Assuming that $\boldsymbol{R}_n$ is invertible and combining both Eqs. (B1) and (B9), we get the relation between the whitening operator and the noise correlation matrix, namely,

$$\boldsymbol{R}_n^{-1}=t_s^3\;\boldsymbol{W}^t\boldsymbol{W}. \tag{B10}$$

With Eq. (24) and the Plancherel formula in Eq. (1),

$$\boldsymbol{y}^t\boldsymbol{R}_n^{-1}\boldsymbol{x}=t_s\check{\boldsymbol{y}}^t\check{\boldsymbol{x}}=\int_{-f_s/2}^{+f_s/2}\check{X}(f)\overline{\check{Y}(f)}\;df, \tag{B11}$$

provided that $\check{y}[k]$ has support in $\{0,\ldots,N-1\}$. Replacing $\check{Y}(f)$ and $\check{X}(f)$ by their definitions completes the proof. ■

[1] Here is a list of Internet sites where more information can be found on the various detectors: LIGO, http://www.ligo.caltech.edu; TAMA, http://tamago.mtk.nao.ac.jp; GEO600, http://www.geo600.uni-hannover.de; VIRGO, http://www.virgo.infn.it.

[2] H. Dimmelmeier, J.A. Font, and E. Müller, Astron. Astrophys. **393**, 523 (2002).

[3] H. Dimmelmeier, J.A. Font, and E. Müller, Astron. Astrophys. **388**, 917 (2002).

[4] G. Matz and F. Hlawatsch, Technical Report No. 96-05, Institut für Nachrichtentechnik und Hochfrequenztechnik, Vienna, 1996.

[5] B. Picinbono and P. Duvaut, IEEE Trans. Inf. Theory **34**, 304 (1988).

[6] W. Anderson, P. Brady, J. Creighton, and E. Flanagan, Phys. Rev. D **63**, 042003 (2001).

[7] T. Pradier, N. Arnaud, M.-A. Bizouard, F. Cavalier, M. Davier, and P. Hello, Phys. Rev. D **63**, 042002 (2001).

[8] Z. Wang and P. Willet, IEEE Trans. Signal Process. **48**, 2682 (2000).

[9] A. Vicerè, Phys. Rev. D **66**, 062002 (2002).
[10] N. Arnaud, F. Cavalier, M. Davier, and P. Hello, Phys. Rev. D **59**, 082002 (1999).
[11] S. Mohanty, Phys. Rev. D **61**, 122002 (2000).
[12] H.L. Van Trees, *Detection, Estimation and Modulation Theory* (Wiley, New York, 1968), Pt. I.
[13] G. Golub and C. Van Loan, *Matrix Computations*, 2nd ed. (John Hopkins University Press, Baltimore, 1989).
[14] T. Zwerger and E. Müller, Astron. Astrophys. **320**, 209 (1997).
[15] ''The Virgo sensitivity curve (2002),'' http://www.virgo.infn.it/senscurve
[16] S. Hughes, gr-qc/0209012.
[17] The figures of the simulations shown in this article were computed with GNU OCTAVE and GNUPLOT, 2002, http://www.octave.org
[18] N.L. Johnson and S. Kotz, *Distribution in Statistics. Continuous Univariate distributions—2* (Wiley, New York, 1970).
[19] I.S. Gradshteyn and I.M. Ryzhik, *Tables of Integrals, Series and Products* (Academic Press, New York, 1980).
[20] E. Chassande-Mottin (in preparation).
[21] F. Hlawatsch, *Time-Frequency Analysis and Synthesis of Linear Signal Spaces*: *Time-Frequency Filters*, *Signal Detection and Estimation*, *and Range-Doppler Estimation* (Kluwer, Boston, 1998).