A unified approach to χ^2 discriminators for searches of gravitational waves from compact binary coalescences

Sanjeev Dhurandhar,[1,*] Anuradha Gupta,[2,1,†] Bhooshan Gadre,[1,‡] and Sukanta Bose[1,3,§]

[1]*Inter-University Centre for Astronomy and Astrophysics, Post Bag 4, Ganeshkhind, Pune 411 007, India*
[2]*Institute for Gravitation and Cosmos, Physics Department, Pennsylvania State University,
University Park, Pennsylvania 16802, USA*
[3]*Department of Physics and Astronomy, Washington State University,
1245 Webster, Pullman, Washington 99164-2814, USA*

We describe a general mathematical framework for χ^2 discriminators in the context of the compact binary coalescence (CBC) search. We show that with any χ^2 is associated a vector bundle over the signal manifold, that is, the manifold traced out by the signal waveforms in the function space of data segments. The χ^2 is then defined as the square of the L_2 norm of the data vector projected onto a finite-dimensional subspace (the fibre) of the Hilbert space of data trains and orthogonal to the signal waveform. Any such fibre leads to a χ^2 discriminator, and the full vector bundle comprising the subspaces and the base manifold constitute the χ^2 discriminator. We show that the χ^2 discriminators used so far in the CBC searches correspond to different fibre structures constituting different vector bundles on the same base manifold, namely, the parameter space. Several benefits accrue from this general formulation. It most importantly shows that there are a plethora of χ^2's available and further gives useful insights into the vetoing procedure. It indicates procedures to formulate new χ^2's that could be more effective in discriminating against commonly occurring glitches in the data. It also shows that no χ^2 with a reasonable number of degrees of freedom is foolproof. It could also shed light on understanding why the traditional χ^2 works so well. We show how to construct a generic χ^2 given an arbitrary set of vectors in the function space of data segments. These vectors could be chosen such that glitches have maximum projection on them. Further, for glitches that can be modeled, we are able to quantify the efficiency of a given χ^2 discriminator by a probability. Second, we propose a family of ambiguity χ^2 discriminators that is an alternative to the traditional one [B. Allen, Phys. Rev. D 71, 062001 (2005), B. Allen *et al.*, Phys. Rev. D 85, 122006 (2012).]. Any such ambiguity χ^2 makes use of the filtered output of the template bank, thus adding negligible cost to the overall search. It is termed so because it makes significant use of the ambiguity function. We first describe the formulation with the help of the Newtonian waveform, apply the ambiguity χ^2 to the spinless TaylorF2 waveforms, and test it on simulated data. We show that the ambiguity χ^2 essentially gives a clean separation between glitches and signals. We indicate how the ambiguity χ^2 can be generalized to detector networks for coherent observations. The effects of mismatch between signal and templates on a χ^2 discriminator using general arguments and the geometrical framework are also investigated.

I. INTRODUCTION

The detection of gravitational wave (GW) signals [1,2] intricately depends on comprehensively addressing the non-Gaussianity and nonstationarity of detector noise [3] and the implementation of effective measures for discriminating noise artifacts from true signals (see, e.g., Ref. [4]). Non-Gaussianity and nonstationarity can arise from various components of the detector itself or the environment (see, e.g., Ref. [5] and references therein). Matched filtering [6], the commonly employed technique for hunting for signals in noisy data, involves cross-correlating the data with a set of templates. These templates are based on gravitational waveforms that we expect the astrophysical sources of interest to emit. But this kind of filtering, by itself, is not always sufficient to distinguish a signal from noise with high confidence. This is because even if a noisy feature—or "glitch"—only slightly resembles a signal, it can give significantly high values of the matched-filter output when it has large power.

Addressing this problem is troublesome because there is a wide variety of noise artifacts that can produce large matched-filter output values. In spite of this difficulty, a few methods have been proposed to better discriminate signals from noise. In this work we focus on signals in ground-based detectors arising from compact binary coalescences

*[*]sanjeev@iucaa.in
[†]axg645@psu.edu
[‡]bug@iucaa.in
[§]sukanta@iucaa.in

(CBCs) involving black holes or neutron stars. These signals are transient, lasting between a fraction of a second to several minutes (see, e.g., Ref. [7] for their matched filter). The χ^2 test described in Refs. [8,9], termed here as the *traditional* χ^2 test (which we henceforth denote by χ_t^2) is one such discriminator. This test splits the broadband data into several smaller sub-bands and checks for consistency between the power expected from the signal in each sub-band with the observed power in that sub-band. This, however, is not the only type of discriminator that can be constructed; a few other tests have been proposed [10–12], not all of which follow the χ^2 distribution [6]. Moreover, while a set of χ^2 tests were introduced in Ref. [11], it remained to be explored whether they can be unified in a way that can facilitate the construction of other χ^2 tests and provide deeper insights into why and when some test is more successful than others.

In this paper, we lay down the foundations of such a formalism. We also propose a single detector and a multi-detector χ^2 test and show how they and some of the tests proposed in the past fit into this unified formalism. Let $x(t)$ describe the single-detector strain snippet providing a strong match with one of the templates of interest. When the match is above a chosen threshold value, we term the data snippet as a *trigger*. The appropriate setting for this formalism is the space of functions that describes single-detector strain of the same duration as $x(t)$, which itself belongs to the same space. This space includes detector strains that are pure noise as well as the GW signal $h(t)$ that corresponds to the template that was triggered by $x(t)$. Detector strains which have no overlap with $h(t)$ belong to a subspace of this function space orthogonal to $h(t)$. Ideally, noise artifacts should be detector strains which have a substantial projection on this orthogonal subspace. A χ^2 discriminator is a construct that allows one to quantify how large that projection is and, thereby, helps assess how different $x(t)$ is from a GW signal.

Ingenuity and knowledge about the characteristics of glitches affecting the search are required in identifying the orthogonal subspace that has large overlaps with the glitches. On the average, this will yield large values of χ^2 for the glitches, zero for the signal, and an expected moderate value for Gaussian noise that is equal to the degrees of freedom of the χ^2—the dimension of the subspace. Thus, a χ^2 discriminator defined on a given subspace complements another χ^2 discriminator defined on a different subspace; the two together would be more effective in distinguishing, or even separating, the signals from the noise artifacts. In fact, as will be clear from the general framework for χ^2 presented here, any two χ^2 discriminators can be realized as a single χ^2 whose orthogonal subspace is just the sum of the subspaces (in the vector-space sense) of the two individual χ^2's.

This paper accomplishes two tasks. First, a general framework for χ^2 discriminators is presented, and we show

that for each such discriminator there is an associated vector bundle. This is in the context of a parameter space or a family of waveforms that depend on a number of parameters. It was shown in Ref. [13] that the family of signals has the structure of a manifold, which was called the signal manifold, with the signal parameters playing the role of coordinates. The signal manifold is in turn a subset of the function space of data trains that is in fact a Hilbert space equipped with a scalar product. A given χ^2 discriminator involves a choice of a family of finite-dimensional subspaces of the function space of data trains such that any such subspace is orthogonal to the signal vector belonging to the signal manifold. Therefore, the χ^2 discriminator just turns out to be the L_2 norm squared of the data vector projected onto this subspace. Thus, any given χ^2 involves the selection of subspaces, one at each point of the signal manifold. If the subspaces are chosen in a smooth manner, the underlying structure of the χ^2 discriminator over the parameter space constitutes a smooth manifold which is a smooth vector bundle. With each χ^2 discriminator there is an associated fibre structure, while the base manifold of waveforms remains the same. This is a special subcategory of vector bundles in which the base manifold remains fixed. In fibre bundle language, the χ^2 discriminator can be described as a non-negative real-valued function on a section of the fibre bundle—the section being defined by the vector field of projections of a given data vector onto the finite-dimensional subspaces mentioned above. A framework for coherent χ^2 tests has been discussed in Ref. [11] which has some common features with the framework presented in this paper, but here we present a more mathematically rigorous framework which has several benefits, as will be discussed in the subsequent text.

It is of course not necessary to choose the same dimension for the subspace at each point of the signal manifold, although it seems sensible to split the signal manifold into regions and fix the dimension of the subspace over each region, with each region having subspaces of possibly different dimensions. In this case we have a union of vector bundles instead of a single vector bundle, where the base manifold is split up into several disjoint manifolds. Since there are various ways to select such subspaces, in principle, a plethora of χ^2 discriminators are possible. Thus, the important observation of this paper is that one may be able to *tune* the χ^2 discriminator to a given family of glitches. In general, one could use different families of subspaces (χ^2's) for different families of glitches.

The general framework has many benefits. It shows that there is an exceedingly rich structure yielding in principle a large number of χ^2 discriminators. It indicates how new χ^2's could be formulated and be more effective in discriminating real signals with commonly occurring glitches: the probing directions for the χ^2 could be taken along the glitches. It gives valuable insights into the currently used χ^2 discriminators and simplifies previous proofs. It also shows

that a χ^2 with reasonable degrees of freedom, say tens or a few hundreds, is not foolproof. The effects of mismatch between a signal and template can be computed, in general, without reference to any specific χ^2 by using general arguments. We show how to construct a generic χ^2 from an arbitrary set of vectors in the function space of data trains. These vectors could be chosen along glitches for maximum effectiveness of the χ^2. Further, we are able to quantify the efficiency of a given χ^2 for glitches that can be modeled as a conditional probability; we use sine-Gaussian glitches as an example.

Second, a family of discriminators is proposed for the CBC searches. We call it the *ambiguity* χ^2; the reason for this terminology will become clear from the text that follows. The ambiguity χ^2 has very little overhead and uses essentially precomputed results, namely, the filtered output from a set of templates. In this regard it coincides very much in spirit with Hanna's construction [10]; however, in many other respects it differs from it. The first advantage is that it is inexpensive because the additional computations are few compared to those required for searching through the data with a bank of templates. Recently, Nitz [14] has shown that the traditional χ^2 can be computed in such a way that it does not add too much to the overall cost; it nonetheless involves a fair amount of computation compared to the method we propose. Our new χ^2 discriminator will supplement the traditional one and, thus, help to better discriminate against noise.

The χ^2 we construct is based on the following observation: if it is a signal that has triggered a given template, then the matched filter output in the vicinity of the trigger must essentially follow the ambiguity function consistent with the parameters of the trigger, such as the signal-to-noise ratio (SNR) and other signal parameters. On the other hand, if it is due to a noise glitch, it will not follow the ambiguity function, in general. A χ^2 statistic thus can be constructed to distinguish between these two situations. The principal idea is to sample the region in the parameter space in the vicinity of the trigger template and compare the "observed" with the "expected" values. It is desirable to remain in the neighborhood of the trigger because the effect of the signal (more precisely the ambiguity function) would not have died out. In order for the χ^2 discriminator's degrees of freedom to be used effectively, either the signal should have significant projection on the selected template vectors of the χ^2, or the glitch should have significant projection on the selected subspaces. Being in the neighborhood of the trigger makes sure that at least one condition is satisfied. Further, to save on computational costs, we make use of the readily available filtered output from the search pipeline which yields the observed values on a subset of the neighborhood of the trigger template. The expected values are obtained by scaling the ambiguity function by the observed SNR of the trigger template at the corresponding

points of the subset. Then the differences in the expected and observed values can be used to construct a χ^2 statistic.

For simplicity, we consider the spinless TaylorF2 signal waveform [15] described by the two masses (or equivalently two mass parameters) and the kinematic parameters, the time of coalescence t_c, and the phase at coalescence ϕ_c. Since our method is more generally applicable (e.g., for IMRPhenomD [16]), many of the formulas given here hold in general. For a demonstration of how our χ^2 works, we restrict ourselves to the TaylorF2 spinless waveform. Templates are constructed at $\phi_c = 0$, $\pi/2$ and at those discretely located mass parameters which are obtained according to a predecided minimal match, say 0.97, after maximizing over t_c and ϕ_c; this is conveniently done by defining a metric on the parameter space [13,17]. Since a fast Fourier transform (FFT) is used in the search algorithm to compute the match over a data segment to search for t_c, the filtered output obtained is (almost) continuous in t_c. Further, the filtered output is obtained at $\phi_c = 0$ and $\pi/2$ and at the discretely placed templates in the mass parameters. The ambiguity χ^2 makes use of tens of points in the parameter space at which the filtered output is available and which lie in the neighborhood of the trigger waveform vector. Here we present a general analysis and formulas involving both phases. For our numerical simulations and results, we consider both phases $\phi_c = 0$ and $\pi/2$. In Sec. V B we make an exception, where for simplicity of demonstration we consider the Newtonian waveform and a single dominant phase. Without loss of generality we take the dominant phase to be $\phi_c = 0$ (if the dominant phase turns out to be $\pi/2$, the formulas can easily be adapted by interchanging the roles of 0 and $\pi/2$ in the formulas).

The paper is organized as follows. In Sec. II we set up the preliminaries required for our purpose and describe the matched filtering paradigm. Section III presents a general framework for the χ^2 discriminators in terms of vector bundles. In Sec. IV we show how a generic χ^2 can be constructed from an arbitrary set of vectors by projecting out components parallel to the trigger templates and using the principal axes transformation. In Sec. V we construct the ambiguity χ^2 where we choose the test vectors from the template bank and show how the ambiguity functions come into play. The filtered output of the search bank forms the input to the ambiguity χ^2, thus adding very little to the computational cost of the search. Here we also present numerical results where the ambiguity χ^2 is tested on simulated data in Sec. V. We indicate how this framework can be generalized in a straightforward manner to the coherent multidetector case [18,19] in Sec. VI. The effect of mismatch between the signal and the templates on the χ^2 is evaluated in Sec. VII. In Sec. VIII we conclude.

II. THE MATCHED FILTERING PARADIGM

Since the matched filter [7] is central to the construction of the CBC detection statistic [7,9,20], we briefly describe

it here. Let us consider two functions, $x(t)$ and $y(t)$, defined over a data segment $[0, T]$ of duration T. When using them as vectors in the space of detector data we denote them in boldface. We define their scalar product as given below, in terms of their respective Fourier transforms $\tilde{x}(f)$ and $\tilde{y}(f)$:

$$(\mathbf{x}, \mathbf{y}) = 4\Re \int_{f_{\text{lower}}}^{f_{\text{upper}}} df \frac{\tilde{x}^*(f)\tilde{y}(f)}{S_h(f)}, \qquad (2.1)$$

where $S_h(f)$ is the one-sided power spectral density (PSD) of the noise, which we denote by $n(t)$. The bandwidth of the detector is $[f_{\text{lower}}, f_{\text{upper}}]$. Often we will write $f_{\text{lower}} = f_s$, where f_s is the seismic cutoff frequency. The noise $n(t)$ is a stochastic process defined over the data segment has an ensemble mean of zero, and is stationary in the wide sense; in the Fourier domain these properties imply that

$$\langle \tilde{n}(f) \rangle = 0, \qquad \langle \tilde{n}^*(f)\tilde{n}(f') \rangle = \frac{1}{2} S_h(f)\delta(f - f'), \quad (2.2)$$

where angular brackets denote ensemble averages while round brackets denote scalar products. This construction makes the space of data segments a Hilbert space—a L_2 space with measure $d\mu \equiv df/S_h(f)$. We denote this space by $\mathcal{D} = L_2([0, T], \mu)$, which is a function space.

The most commonly used post-Newtonian (PN) approximant is TaylorF2, which is computed in the Fourier domain using the stationary phase approximation. We use this waveform for the signal in our discussion on the ambiguity χ^2. The general form of the signal, denoted by h, is

$$\tilde{h}(f) = \mathcal{A}f^{-7/6}e^{-i\psi(f)}, \qquad (2.3)$$

where the overall amplitude $\mathcal{A}$ depends on the binary component masses, the source distance, sky position, and the relative orientation of the binary orbit to the detector. The phase $\psi(f)$ is computed to 3.5PN order explicitly [15], and depends on t_c, ϕ_c, and the mass parameters. (The sign convention for the Fourier transform in Ref. [15] is opposite to ours.) We will view these waveforms as vectors in $\mathcal{D}$ and denote them by the boldfaced letter $\mathbf{h}$.

We will describe our method in terms of the Newtonian waveform, which is simple (even if somewhat inaccurate) and thus allows us to draw the reader's attention more toward the intricacies of the method. However, when applying our method on simulated injections and glitches, we use PN approximants. The normalized Newtonian inspiral binary waveform in the Fourier domain is given by

$$\tilde{h}(f; t_c, \tau_0, \phi_c) = \mathcal{N}f^{-\frac{7}{6}}e^{-i\psi_N(f; t_c, \tau_0, \phi_c)}, \qquad (2.4)$$

where $\mathcal{N}$ is the normalization constant, which is determined by setting $(\mathbf{h}, \mathbf{h}) = 1$. Thus,

$$\mathcal{N} = \frac{1}{2}\left[\int_{f_{\text{lower}}}^{f_{\text{upper}}} \frac{df}{f^{7/3}S_h(f)}\right]^{-\frac{1}{2}}. \qquad (2.5)$$

The phase $\psi_N(f)$ is given by

$$\psi_N(f; t_c, \tau_0, \phi_c) = 2\pi f t_c + \frac{6\pi f_s \tau_0}{5}\left(\frac{f}{f_s}\right)^{-5/3} - \phi_c - \frac{\pi}{4}, \qquad (2.6)$$

where t_c and ϕ_c are the coalescence time and phase, respectively. Furthermore, we have expressed the phase in terms of a parameter more suited to this work than the chirp mass [7], namely, the chirp time τ_0 [7,21]. The chirp times are used to construct template banks because the metric components are nearly constant in these parameters, so that the templates with a fixed minimal match cover the parameter space uniformly. Physically, τ_0 is the time taken for the binary to coalesce starting from some fiducial frequency f_a. We take this fiducial frequency to be the seismic cutoff frequency $f_s = 10$ Hz for Advanced LIGO (aLIGO):

$$\tau_0 = \frac{5}{256\pi f_s}(\pi \mathcal{M}f_s)^{-5/3}$$
$$\simeq 1393\left(\frac{f_s}{10 \text{ Hz}}\right)^{-8/3}\left(\frac{\mathcal{M}}{M_\odot}\right)^{-5/3} \text{ sec}. \qquad (2.7)$$

Here $\mathcal{M} = \mu^{3/5}M^{2/5}$ is the chirp mass, with μ and M being the reduced and the total mass, respectively. $M_\odot$ denotes the mass of the Sun. In terms of this normalized waveform we now define the signal and the templates.

The signal $\mathbf{s}$ in the data is just an amplitude A multiplying the normalized waveform $\mathbf{h}$; thus, $\mathbf{s} = A\mathbf{h}$. The data vector, which we denote by $\mathbf{x}$, is then $\mathbf{x} = \mathbf{s} + \mathbf{n}$, when a signal is present; otherwise it is just noise, i.e., $\mathbf{x} = \mathbf{n}$, when a signal is absent. The match c (correlation) is the scalar product between the data $\mathbf{x}$ and a (normalized) template $\mathbf{h}$, that is, $c = (\mathbf{x}, \mathbf{h})$, which is then a function of the signal parameters. In the analysis of the data for searching signals the match is maximized over signal parameters and compared with a preset threshold. In practice, for the parameters t_c, ϕ_c, the templates only need to be defined at $\phi_c = 0$ and $\phi_c = \pi/2$, and for $t_c = 0$. This is because the search over these parameters can be done efficiently using quadratures for ϕ_c and the FFT algorithm for t_c in a continuous fashion. The search over the mass parameters denoted by the vector parameter ϑ needs to be carried out with a finely sampled discrete bank of templates so that signals are not missed. Thus, the search pipeline outputs the correlations c_0 and $c_{\pi/2}$ continuously in the parameter t_c[1] and at the template locations in ϑ.

---

[1]In practice, data are sampled at a finite rate in time. Hence, the correlations are also computed at discrete values of t_c. However, the sampling rate is very high, and is at least twice the Nyquist frequency of the sought signal.

In the ambiguity χ^2, we propose to choose our test waveforms from the template bank so that the pipeline output provides a ready-made input for computing the ambiguity χ^2, which in turn means that negligible additional computational costs are involved. However, before going over to the ambiguity χ^2 we will first lay down a general framework of χ^2 discriminators and then show how to construct a generic χ^2. The construction of the ambiguity χ^2 then follows easily from the generic construction.

III. A GENERAL FRAMEWORK OF THE χ^2 DISCRIMINATORS: χ^2 AS A VECTOR BUNDLE

In this section we show that there is a natural vector bundle structure that can be associated with a given χ^2 discriminator. The χ^2 then turns out to be just the square of the L_2 norm of the data vector projected onto the fibre of the vector bundle.

A. The mathematical structure of a χ^2 discriminator

As mentioned above, the space of data trains over a time segment $[0, T]$ is the Hilbert space $\mathcal{D} = L_2([0, T], \mu)$, where μ is a measure defined conveniently in the Fourier domain by $d\mu = df/S_h(f)$, with $S_h(f)$ being the one-sided noise PSD. The scalar product on this space is given by Eq. (2.1). This is essentially a L_2 space with the norm suitably modified. The χ^2 discriminator is defined so that its value for the signal is zero and for Gaussian noise it has a χ^2 distribution with a certain number of degrees of freedom. In principle, $\mathcal{D}$ is infinite dimensional. In practice the data are sampled at a finite but large number of points. Therefore, although $\mathcal{D}$ is finite dimensional, its dimension is very large, typically of the order of $\sim 10^6$ or more; so, for all practical purposes $\mathcal{D}$ can be treated as infinite dimensional.

Let us first pick a single waveform $\mathbf{h}$. Now consider the space normal to $\mathbf{h}$, $\mathcal{N}_{\chi^2}(\mathbf{h})$, defined by

$$\mathcal{N}_{\chi^2}(\mathbf{h}) = \{\mathbf{x} \in \mathcal{D} | (\mathbf{x}, \mathbf{h}) = 0\}. \qquad (3.1)$$

$\mathcal{N}_{\chi^2}(\mathbf{h})$ is infinite dimensional and isomorphic to the quotient space $\mathcal{D}/\{\mathbf{h}\}$, where $\{\mathbf{h}\}$ is the one-dimensional subspace generated by $\mathbf{h}$. The χ^2 test for the waveform $\mathbf{h}$ is defined by choosing a finite-dimensional subspace of $\mathcal{N}_{\chi^2}(\mathbf{h})$, say, of dimension p, which we denote by $\mathcal{S}$. This is similar to the projection operator defined in Ref. [10]. Then we claim that the χ^2 pertaining to the waveform $\mathbf{h}$ is just the square of the L_2 norm of the data vector $\mathbf{x}$ projected onto $\mathcal{S}$. Specifically, we decompose the data vector $\mathbf{x} \in \mathcal{D}$ as,

$$\mathbf{x} = \mathbf{x}_\mathcal{S} + \mathbf{x}_{\mathcal{S}^\perp}, \qquad (3.2)$$

where $\mathcal{S}^\perp$ is the orthogonal complement of $\mathcal{S}$ in $\mathcal{D}$. $\mathbf{x}_\mathcal{S}$ and $\mathbf{x}_{\mathcal{S}^\perp}$ are projections of $\mathbf{x}$ into the subspaces $\mathcal{S}$ and $\mathcal{S}^\perp$, respectively. The orthogonal complement of $\mathcal{S}$ is defined as

$$\mathcal{S}^\perp = \{\mathbf{x} \in \mathcal{D} | (\mathbf{x}, \mathbf{y}) = 0, \quad \forall \, \mathbf{y} \in \mathcal{S}\}, \qquad (3.3)$$

and we may write $\mathcal{D}$ as a direct sum of $\mathcal{S}$ and $\mathcal{S}^\perp$, that is, $\mathcal{D} = \mathcal{S} \oplus \mathcal{S}^\perp$.

Then the statistic χ^2 is

$$\chi^2(\mathbf{x}) = \|\mathbf{x}_\mathcal{S}\|^2. \qquad (3.4)$$

Thus the χ^2 is defined with respect to some finite-dimensional subspace $\mathcal{S}$ of $\mathcal{D}$ which is orthogonal to the waveform $\mathbf{h}$. Now choose any orthonormal basis in $\mathcal{S}$, say, $\mathbf{e}_\alpha, \alpha = 1, 2, \ldots, p$, so that $(\mathbf{e}_\alpha, \mathbf{e}_\beta) = \delta_{\alpha\beta}$, where $\delta_{\alpha\beta}$ is the Kronecker delta. We easily verify its properties.

(1) For a general data vector $\mathbf{x} \in \mathcal{D}$, we have

$$\chi^2(\mathbf{x}) = \|\mathbf{x}_\mathcal{S}\|^2 = \sum_{\alpha=1}^{p} |(\mathbf{x}, \mathbf{e}_\alpha)|^2. \qquad (3.5)$$

(2) Clearly, $\chi^2(\mathbf{h}) = 0$, because the projection of $\mathbf{h}$ into the subspace $\mathcal{S}$ is zero or $\mathbf{h}_\mathcal{S} = 0$.

(3) Now let us take the noise $\mathbf{n}$ to be Gaussian which satisfies Eq. (2.2), i.e., $\langle \mathbf{n} \rangle = 0$. Therefore, the following is valid:

$$\chi^2(\mathbf{n}) = \|\mathbf{n}_\mathcal{S}\|^2 = \sum_{\alpha=1}^{p} |(\mathbf{n}, \mathbf{e}_\alpha)|^2. \qquad (3.6)$$

Observe that the random variables $(\mathbf{n}, \mathbf{e}_\alpha)$ are independent and Gaussian, with mean zero and variance unity because $\langle (\mathbf{e}_\alpha, \mathbf{n})(\mathbf{n}, \mathbf{e}_\beta) \rangle = (\mathbf{e}_\alpha, \mathbf{e}_\beta) = \delta_{\alpha\beta}$ [we have made use of the identity stated in Eq. (4.11)]. Thus, $\chi^2(\mathbf{n})$ has a χ^2 distribution with p degrees of freedom.

Therefore, the χ^2 statistic satisfies the essential criteria that it is zero on the signal and is distributed χ^2 for Gaussian noise. Moreover, most importantly, the statistic is invariant under the orthogonal group in p dimensions $O(p)$ acting on $\mathcal{S}$. That is, if we choose any other orthonormal basis $\mathbf{e}'_\alpha$ which is related to $\mathbf{e}_\alpha$ by an orthogonal transformation, then the $\chi^2(\mathbf{x})$ for any vector $\mathbf{x} \in \mathcal{D}$ is invariant under this transformation. *This underscores the fact that it is the subspace $\mathcal{S}$ which is relevant, rather than any particular basis.* In order to perform the calculation, we are free to choose any orthonormal basis of $\mathcal{S}$. Choosing an orthonormal basis makes the statistic appear manifestly χ^2 as it can be written as a sum of squares of independent Gaussian random variables with mean zero and variance unity.

However, in the context of the CBC searches, we are in a more complex situation. We do not have just one waveform but a family of waveforms which depend on several parameters, such as masses, spins, and other kinematical parameters. We denote these parameters by ϑ^a, $a = 1, 2, \ldots, m$. As before, we may assume the waveforms to be normalized, i.e., $\|\mathbf{h}(\vartheta^a)\| = 1$. Then the waveforms

trace out a m-dimensional manifold $\mathcal{P}$—the signal manifold–which is a submanifold of $\mathcal{D}$. Each point of $\mathcal{P}$ is a normalized waveform $\mathbf{h}(\vartheta^a)$. We now associate a p-dimensional subspace $\mathcal{S}$ orthogonal to the waveform $\mathbf{h}(\vartheta^a)$ at each point of $\mathcal{P}$; we have a p-dimensional vector space "attached" to each point of $\mathcal{P}$. When done in a smooth manner, this construction produces a fibre bundle with a p-dimensional vector space attached to each point of the m-dimensional manifold $\mathcal{P}$. The fibre bundle so obtained is a vector bundle of dimension $m + p$. Therefore, we have found a very general mathematical structure for the χ^2 discriminator. Any given χ^2 discriminator for a signal waveform $\mathbf{h}(\vartheta^a)$ is the L_2 norm of a given data vector $\mathbf{x}$ projected onto the fibre $\mathcal{S}$ at $\mathbf{h}(\vartheta^a)$. Note that here the space of data trains $\mathcal{D}$ has two distinct roles to play: (i) its subspaces $\mathcal{S}$ orthogonal to the waveforms in $\mathcal{P}$ form the fibres of the vector bundle, and (ii) the signal manifold $\mathcal{P}$ is a submanifold of $\mathcal{D}$.

In the case where the signal or the base manifold has a simple topological structure, the vector bundle reduces to a product manifold—the trivial bundle. Although from the signal models considered here it may seem that the fibre bundle is always trivial, this is not the case in general. For complex signal models, for example, when the orientation of the source enters into the model in a complex manner, the base manifold could involve a spherical topological structure, or, when the sky location is among the search parameters, the celestial sphere could make the topological structure of the base manifold complex and a nontrivial vector bundle would result. (For example, the tangent bundle of a 2-sphere S^2 is nontrivial. See Ref. [22] for an intuitive argument; one cannot comb the hair on the surface of a sphere.)

In effect, we have a large collection of subspaces $\mathcal{S}$ to choose from; in fact, we have literally an infinite number of choices and a plethora of χ^2 discriminators can be constructed. Each χ^2 discriminator gives rise to a different fibre structure, while the base manifold $\mathcal{P}$ of waveforms remains the same. This is a special type of category of vector bundles where the base space remains fixed. We show this situation in Fig. 1, where π and π' are projection maps from the vector bundles $\mathcal{F}$ and $\mathcal{F}'$, respectively, onto the base space $\mathcal{P}$. f is a diffeomorphism of vector bundles and we have $\pi \circ f = \pi'$, that is, the diagram commutes. In fact, f induces a linear map between $\mathcal{S}'$ and $\mathcal{S}$: given a data vector $\mathbf{x} \in \mathcal{D}$ the linear map maps $\mathbf{x}_{\mathcal{S}'}$ to $\mathbf{x}_{\mathcal{S}}$ and therefore further induces a map between the two χ^2's. In vector bundle language, each data vector $\mathbf{x}$ defines a section $\mathcal{P}_\mathbf{x}$ of the fibre bundle, namely, the vector field $\mathbf{x}_\mathcal{S}$ on $\mathcal{P}$, and the χ^2 is a non-negative real-valued function defined on $\mathcal{P}_\mathbf{x}$. Note, however, that it is not absolutely necessary to fix the dimension of $\mathcal{S}$ to be p at each point of $\mathcal{P}$, but it seems convenient to do so.

Although in principle there are an enormous number of choices in selecting the χ^2 discriminator, there are physical
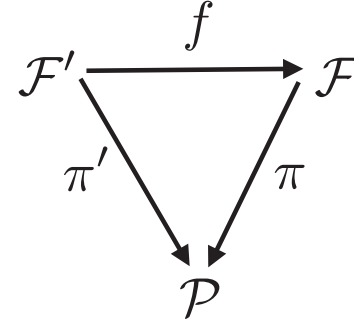

FIG. 1. Two vector bundles $\mathcal{F}$ and $\mathcal{F}'$ having the same base manifold $\mathcal{P}$ and with surjections π and π', respectively. f is a diffeomorphism between the vector bundles such that $\pi \circ f = \pi'$. The diagram commutes.

and practical considerations that limit the choice. Here we mention two such important considerations.

(1) The normal spaces $\mathcal{S}$ chosen must be such that the projection of the glitches on $\mathcal{S}$ is large. Specifically, if $\mathbf{g}$ is a glitch, then if we decompose $\mathbf{g} = \mathbf{g}_\mathcal{S} + \mathbf{g}_{\mathcal{S}^\perp}$ as before, we must have $\chi^2(\mathbf{g}) = \chi^2(\mathbf{g}_\mathcal{S}) \gg p$. This will ensure that the glitch is distinguished from the signal and Gaussian noise. However, this seems to be not so simple a proposition and ingenuity may be required to select the subspaces with this property. The main problem seems to lie in modeling the glitches. On the other hand, our analysis shows that there is a lot of freedom in selecting $\mathcal{S}$. We expect that this facility will help find better performing discriminators.

(2) It is desirable to minimize the computational cost of evaluating the χ^2, that is, to make the discriminator computationally efficient. One way to achieve this is by using precomputed scalar products such as those available from the matched-filtering pipeline so that not much computational overhead is required to evaluate the χ^2. The ambiguity χ^2 and the bank χ^2 make use of this fact.

Finally, we point out that no χ^2 statistic on a finite-dimensional manifold—even with tens (or hundreds) of degrees of freedom—is foolproof. This is because the dimension of $\mathcal{N}_{\chi^2}$ is so large that in a practical situation $\dim(\mathcal{S}^\perp) \gg \dim(\mathcal{S})$ and there can exist glitches (vectors in $\mathcal{D}$) that can have very small projection in $\mathcal{S}$ and large projection in $\mathcal{S}^\perp$. Thus, glitches can get through almost any χ^2 that has a reasonable number of degrees of freedom. Generally speaking, the greater number of degrees of freedom a χ^2 statistic possesses (that is, the more directions it probes), the more effective it will be against glitches. This can again be easily seen from the geometrical picture.

B. The example of the traditional χ^2_t

As a simple demonstration of the general framework described above, we show how the traditional χ^2_t fits into

this framework. The main task is to describe the subspaces $\mathcal{S}$. We obtain not only these but also an orthonormal basis for each subspace. The χ^2_t (due to Allen *et al.* [8,9]) is constructed in the following manner.

One first partitions the detector frequency bandwidth into, say, p nonoverlapping frequency sub-bands $\Delta f_1, \Delta f_2, …, \Delta f_p$, such that the expected signal correlation in each sub-band Δf_α is the fraction $1/p$ of the full correlation c of the trigger. By computing the observed signal correlation c_α in each sub-band and taking differences $\Delta c_\alpha = c_\alpha - c/p$, one defines the χ^2 statistic as

$$\chi^2_t = p \sum_{\alpha=1}^{p} \Delta c_\alpha^2. \tag{3.7}$$

For a real CBC signal in Gaussian noise, this statistic is χ^2 distributed with $p-1$ degrees of freedom [8] when the parameters of the template perfectly match those of the signal. For simplicity, here we consider the case of the known phase ϕ_c, but the argument can be easily extended to the case of an unknown phase.

We now exhibit how the subspace $\mathcal{S}$ of $p-1$ dimensions is constructed for the traditional χ^2, and how it can be imparted with an orthonormal basis in which the statistic is manifestly χ^2 distributed. Let $\mathbf{h}$ be a normalized template waveform, that is, $\|\mathbf{h}\| = 1$. In the frequency domain we write it as $\tilde{h}(f)$, in terms of which we can conveniently define the following p waveform vectors:

$$\tilde{h}_\alpha(f) = \tilde{h}(f), \qquad f \in \Delta f_\alpha,$$
$$= 0 \qquad \text{otherwise}, \tag{3.8}$$

where $\alpha = 1, 2, …, p$. These p vectors are denoted by $\mathbf{h}_\alpha$, and they obey $\|\mathbf{h}_\alpha\|^2 = 1/p$. The waveform vectors corresponding to Δc_α can then be defined as

$$\Delta \mathbf{h}_\alpha = \mathbf{h}_\alpha - \frac{\mathbf{h}}{p}. \tag{3.9}$$

It is easy to check that the $\Delta \mathbf{h}_\alpha$ are orthogonal to $\mathbf{h}$, that is, $(\Delta \mathbf{h}_\alpha, \mathbf{h}) = 0$. We also observe that

$$\sum_{\alpha=1}^{p} \Delta \mathbf{h}_\alpha = 0,$$

$$(\Delta \mathbf{h}_\alpha, \Delta \mathbf{h}_\beta) = -\frac{1}{p^2} \quad \text{for } \alpha \neq \beta. \tag{3.10}$$

Thus the random variables Δc_α (scalar products with the data) corresponding to these waveforms are not only correlated but also algebraically dependent. But, as rigorously shown in Ref. [8], a χ^2 statistic is obtained with $p-1$ degrees of freedom.

Although no orthonormal basis of waveform vectors in $\mathcal{S}$ was explicitly given in Ref. [8] (there is, in fact, no need to do so), nevertheless we list one such basis here:

$$\mathbf{e}_1 = \sqrt{\frac{p}{2}}(\mathbf{h}_1 - \mathbf{h}_2),$$

$$\mathbf{e}_2 = \sqrt{\frac{p}{6}}(\mathbf{h}_1 + \mathbf{h}_2 - 2\mathbf{h}_3),$$

$$\vdots$$

$$\mathbf{e}_{p-1} = \sqrt{\frac{p}{p(p-1)}}(\mathbf{h}_1 + \mathbf{h}_2 + \cdots + \mathbf{h}_{p-1} - (p-1)\mathbf{h}_p). \tag{3.11}$$

The space $\mathcal{S}$ is spanned by the above basis vectors and is, therefore, $p-1$ dimensional. One easily verifies that each $\mathbf{e}_\alpha$ is orthogonal to the waveform $\mathbf{h}$ and $(\mathbf{e}_\alpha, \mathbf{e}_\beta) = \delta_{\alpha\beta}$. As mentioned before, any orthogonal transformation of this basis will also yield another orthonormal basis.

C. Efficiency of χ^2 discriminators for modeled glitches

In the case of glitches that can be modeled it is possible to assign an efficiency to a given χ^2 discriminator. It is in fact the probability that the glitch will be removed by the χ^2. To fix ideas we consider the specific case of sine-Gaussian glitches, even though our analysis can be easily generalized to other glitch models. The sine-Gaussian glitches are described in the Fourier domain as [23,24]

$$g(f; A, f_0, Q) = g_0 \frac{A}{2i} e^{-\frac{Q^2}{4f_0^2}(f-f_0)^2}, \tag{3.12}$$

where g_0 is the normalization factor given by

$$g_0^2 = \int_{f_{\text{lower}}}^{f_{\text{upper}}} df \frac{e^{-\frac{Q^2}{2f_0^2}(f-f_0)^2}}{S_h(f)}, \tag{3.13}$$

where A is the amplitude, f_0 is the central frequency, and Q is the quality factor. The normalization is such that $\|g(f; A, f_0, Q)\| = A$.

Let $\mathcal{G}$ be the set of sine-Gaussian glitches with parameters (A, f_0, Q) in some given range $\mathcal{R}$. Let $p(A, f_0, Q)$ be the probability density function on $\mathcal{R}$ whose integral on $\mathcal{R}$ is unity. It is actually the relative frequency of the occurrence of such glitches in some differential volume $dA df_0 dQ$ that can be estimated from the data. We can also think of $p(A, f_0, Q)$ as a prior on the set of glitches.

Let the templates be denoted by normalized waveforms $\mathbf{h}$, which depend on both kinematic and dynamical parameters. We define the set of triggers $\mathcal{T}$ to be

$$\mathcal{T} = \{(A, f_0, Q) \in \mathcal{R} | \max(\mathbf{g}, \mathbf{h}) > \eta, \mathbf{g} \in \mathcal{G}\}, \tag{3.14}$$

where the maximum is taken over all the parameters and η is a preset threshold. Then $\mathcal{T} \subset \mathcal{R}$, and we define the probability

$$P_{\mathcal{T}} = \int_{\mathcal{T}} p(A, f_0, Q) dA df_0 dQ < 1. \qquad (3.15)$$

Now consider a χ^2 with p degrees of freedom. Then the mean of χ^2 is p and the standard deviation is $\sigma = \sqrt{2p}$. We may then compute $\chi^2(\mathbf{g})$ for a trigger and if we find that this quantity is large—say, q standard deviations more than the mean, i.e., $\chi^2(\mathbf{g}) > p + q\sqrt{2p}$—we may decide to discard it. We may typically take $q \simeq 3$. Accordingly, we define the set

$$\mathcal{B} = \{(A, f_0, Q) \in \mathcal{T} | \chi^2(\mathbf{g}(A, f_0, Q)) > p + q\sqrt{2p}\}, \qquad (3.16)$$

where $\mathcal{B}$ is the set of triggers in $\mathcal{T}$ that are blocked by the χ^2. The probability associated with $\mathcal{B}$ is

$$P_{\mathcal{B}} = \int_{\mathcal{B}} p(A, f_0, Q) dA df_0 dQ. \qquad (3.17)$$

Thus, the conditional probability $P_{\mathcal{B}}/P_{\mathcal{T}}$ is the probability of triggers that are blocked among all the triggers. The closer this quantity is to unity, the better the performance or the efficiency of the χ^2. This would be a quantitative measure that can be used to decide on the choice of χ^2 for sine-Gaussian glitches.

This analysis can be easily generalized in an analogous fashion to any other class of glitches that can be modeled. We plan to study the utility of this quantitative measure for sine-Gaussian and other modeled glitches in a future work.

IV. CONSTRUCTING A GENERIC χ^2 DISCRIMINATOR

Using the framework described in the previous section, we propose a general scheme for constructing χ^2 discriminators. As we saw, constructing a χ^2 involves selecting subspaces $\mathcal{S}$ orthogonal to the signal waveforms. Further, we would like these subspaces to be such that the projection of the glitches on them is as large as possible. One way to achieve this is by choosing vectors along the glitches and using them to construct a field of the subspaces $\mathcal{S}$ on the signal manifold. For example, we may choose the $\mathbf{h}_\alpha$ to be along sine-Gaussians [23,24], or even better, choose the $\mathbf{h}_\alpha$ as a sine-Gaussian basis, if our goal is to rule out sine-Gaussians. We can do the same for other glitches if one can either model them or somehow find their directions in $\mathcal{D}$ and then orient the $\mathbf{h}_\alpha$ along those directions. We present the procedure below.

A. Obtaining the subspaces $\mathcal{S}$

Let us begin with an arbitrary set of vectors $\mathbf{h}_\alpha$, where $\alpha = 1, 2, ..., p$. We choose these vectors to be linearly independent (if they were linearly dependent, we could omit the dependent ones and make our set linearly independent). The χ^2 statistic is constructed by taking the differences between the expected and observed correlations. The expected correlations are obtained when the data is just the signal. Consider a trigger from a template with parameters ϑ and t_c. We denote the trigger waveform vectors by $\mathbf{h}_0(0)$ and $\mathbf{h}_{\pi/2}(0)$ which are time-shifted versions of the templates at ϑ by the amount t_c. The corresponding correlations with the data $\mathbf{x}$ are denoted by $c_0(0) = (\mathbf{x}, \mathbf{h}_0(0))$ and $c_{\pi/2}(0) = (\mathbf{x}, \mathbf{h}_{\pi/2}(0))$. The "zero" in the trigger template is consistent with the notation that $\Delta\vartheta = \Delta t_c = 0$ for this waveform vector. The *observed* value of the correlation of the data $\mathbf{x}$ with the chosen vector $\mathbf{h}_\alpha$ is just given by

$$c_\alpha^o = (\mathbf{x}, \mathbf{h}_\alpha). \qquad (4.1)$$

We now define the expected correlations c_α^e. We consider the situation of the perfect match between the signal and the trigger template. Then the signal can be written as

$$\mathbf{s} = (\mathbf{s}, \mathbf{h}_0(0))\mathbf{h}_0(0) + (\mathbf{s}, \mathbf{h}_{\pi/2}(0))\mathbf{h}_{\pi/2}(0), \qquad (4.2)$$

and its scalar product with the vector $\mathbf{h}_\alpha$ as

$$(\mathbf{s}, \mathbf{h}_\alpha) = (\mathbf{s}, \mathbf{h}_0(0))(\mathbf{h}_0(0), \mathbf{h}_\alpha) \\ + (\mathbf{s}, \mathbf{h}_{\pi/2}(0))(\mathbf{h}_{\pi/2}(0), \mathbf{h}_\alpha). \qquad (4.3)$$

Now the observed values of the correlations of the data with the trigger waveform vectors are

$$c_0(0) = (\mathbf{x}, \mathbf{h}_0(0)), \qquad c_{\pi/2}(0) = (\mathbf{x}, \mathbf{h}_{\pi/2}(0)), \qquad (4.4)$$

which have mean values

$$\langle c_0(0) \rangle = (\mathbf{s}, \mathbf{h}_0(0)), \qquad \langle c_{\pi/2}(0) \rangle = (\mathbf{s}, \mathbf{h}_{\pi/2}(0)), \qquad (4.5)$$

because we have zero mean noise. We then replace the scalar products $(\mathbf{s}, \mathbf{h}_0(0))$ and $(\mathbf{s}, \mathbf{h}_{\pi/2}(0))$ in Eq. (4.3) by the observed $c_0(0) = (\mathbf{x}, \mathbf{h}_0(0))$ and $c_{\pi/2}(0) = (\mathbf{x}, \mathbf{h}_{\pi/2}(0))$, respectively. Accordingly, we define the *expected* correlations as

$$c_\alpha^e = c_0(0)(\mathbf{h}_0(0), \mathbf{h}_\alpha) + c_{\pi/2}(0)(\mathbf{h}_{\pi/2}(0), \mathbf{h}_\alpha), \qquad (4.6)$$

whose mean values coincide with those given in Eq. (4.3). We then take the differences between the expected and observed correlations,

$$\Delta c_\alpha(\mathbf{x}) = c_\alpha^o(\mathbf{x}) - c_\alpha^e \equiv (\mathbf{x}, \Delta\mathbf{h}_\alpha), \qquad (4.7)$$

where we have defined the projected vectors $\Delta\mathbf{h}_\alpha$ as

$$\Delta\mathbf{h}_\alpha = \mathbf{h}_\alpha - (\mathbf{h}_\alpha, \mathbf{h}_0(0))\mathbf{h}_0(0) - (\mathbf{h}_\alpha, \mathbf{h}_{\pi/2}(0))\mathbf{h}_{\pi/2}(0). \qquad (4.8)$$

We readily verify that $(\Delta\mathbf{h}_\alpha, \mathbf{h}_0(0)) = (\Delta\mathbf{h}_\alpha, \mathbf{h}_{\pi/2}(0)) = 0$ for every α, and therefore $\Delta\mathbf{h}_\alpha$ are also orthogonal to the signal $\mathbf{s}$. *Therefore, S is precisely the span of $\Delta\mathbf{h}_\alpha$, and thus we have constructed a generic χ^2 for an arbitrary collection of vectors $\mathbf{h}_\alpha$.* By taking the square of the norm of the projected vector $\mathbf{x}_S$ we obtain the generic χ^2. In the next subsection, we show how this can be done in an explicit manner by constructing an orthonormal basis for S.

First, we note the following properties of the linear functional Δc_α:

$$\Delta c_\alpha(\mathbf{s}) = 0, \qquad \langle\Delta c_\alpha(\mathbf{n})\rangle = 0. \tag{4.9}$$

Then, the Δc_α are Gaussian if the noise is Gaussian because they are linear combinations of Gaussian variables, as can be seen from Eqs. (4.7) and (4.8). However, we cannot immediately take the sum of squares of the Δc_α to build a χ^2 because (i) they are correlated and, moreover, (ii) they do not have unit variance. But these problems can be easily remedied by (i) performing a principal axes transformation and (ii) normalizing the resulting orthogonal vectors. To proceed with the orthogonalization, we need to compute the covariance matrix of the Δc_α, which we do in the next subsection.

B. The covariance matrix, the principal axes transformation, and orthonormal bases

In order to compute the covariance matrix of Δc_α for a general $\mathbf{x}$, it is sufficient to consider the noise alone since $\Delta c_\alpha(\mathbf{s}) = 0$ and therefore $\Delta c_\alpha(\mathbf{x}) = \Delta c_\alpha(\mathbf{n})$. Since $\langle\Delta c_\alpha(\mathbf{n})\rangle = 0$, this computation is simple. From Eq. (4.7) we find that

$$\Delta c_\alpha(\mathbf{n}) = (\mathbf{n}, \Delta\mathbf{h}_\alpha). \tag{4.10}$$

We now make use of the following identity. For any vectors $\mathbf{x}, \mathbf{y} \in \mathcal{D}$ one has

$$\langle(\mathbf{x}, \mathbf{n})(\mathbf{n}, \mathbf{y})\rangle = (\mathbf{x}, \mathbf{y}). \tag{4.11}$$

One can easily prove this identity using Eq. (2.2). Since $\langle\Delta c_\alpha(\mathbf{n})\rangle = 0$, the covariance matrix of Δc_α simplifies to

$$C_{\alpha\beta} \equiv \langle\Delta c_\alpha(\mathbf{n})\Delta c_\beta(\mathbf{n})\rangle. \tag{4.12}$$

One can explicitly compute this matrix using the identity given in Eq. (4.11). We then have

$$C_{\alpha\beta} = (\Delta\mathbf{h}_\alpha, \Delta\mathbf{h}_\beta). \tag{4.13}$$

Note that, in general, the matrix C is not diagonal, which implies that the Δc_α are correlated.

An orthonormal set of vectors may be constructed from $\Delta\mathbf{h}_\alpha$. Let O be the orthogonal transformation which diagonalizes the covariance matrix C, i.e., $O^T C O = \Lambda$,

where Λ is a diagonal matrix whose entries are the eigenvalues, λ_α, $\alpha = 1, 2, ..., p$ of C.

We now assert that C is positive definite if $\Delta\mathbf{h}_\alpha$ are linearly independent (we later comment on the situation where they become linearly dependent after projecting $\mathbf{h}_\alpha$). Choose any orthonormal basis $\mathbf{e}_\alpha$, $\alpha = 1, 2, ..., p$ of S and write $\Delta\mathbf{h}_\alpha = A_{\alpha\beta}\mathbf{e}_\beta$; then, it is easy to see that $C_{\alpha\beta} = A_{\alpha\gamma}A_{\beta\gamma}$, or in matrix notation

$$C = AA^T, \tag{4.14}$$

where A is the matrix whose entries are $A_{\alpha\beta}$. Then for any nonzero vector $\mathbf{v} \in S$, we have $\mathbf{v}^T C \mathbf{v} = \mathbf{v}^T A A^T \mathbf{v} = (A^T\mathbf{v})^T(A^T\mathbf{v}) > 0$. This means that C is positive definite and all its eigenvalues must be positive, that is, $\lambda_\alpha > 0$ for each α.

Now construct

$$\Delta\mathbf{h}'_\alpha = (O^T)_{\alpha\beta}\Delta\mathbf{h}_\beta, \tag{4.15}$$

which then satisfy

$$(\Delta\mathbf{h}'_\alpha, \Delta\mathbf{h}'_\beta) = \lambda_\alpha\delta_{\alpha\beta}. \tag{4.16}$$

The $\Delta\mathbf{h}'_\alpha$ are orthogonal but not orthonormal. We may then define the correlations

$$\Delta c'_\alpha = (\mathbf{x}, \Delta\mathbf{h}'_\alpha). \tag{4.17}$$

If the noise is Gaussian, then $\Delta c'_\alpha$ are independent Gaussian random variables with zero mean and variances λ_α. The next step is to normalize the vectors by defining

$$\mathbf{e}_\alpha = \frac{\Delta\mathbf{h}'_\alpha}{\sqrt{\lambda_\alpha}}. \tag{4.18}$$

Then $\mathbf{e}_\alpha, \alpha = 1, 2, ..., p$ form an orthonormal basis of S. The χ^2 then can be exhibited in terms of this basis as a sum of squares of standard normal random variables. This orthogonal transformation is not unique; any other orthogonal transformation of this basis will suffice to define the χ^2 in this way. The χ^2 discriminator is then defined as

$$\chi^2 = \sum_{\alpha=1}^{p} \frac{\Delta c'^2_\alpha}{\lambda_\alpha} \equiv \Delta c_\alpha[C^{-1}]^{\alpha\beta}\Delta c_\beta. \tag{4.19}$$

For Gaussian noise obeying Eq. (2.2), the above statistic has a χ^2 distribution with p degrees of freedom.

It may happen in certain situations that after projection the vectors $\Delta\mathbf{h}_\alpha$ may become linearly dependent or almost linearly dependent at certain points or regions of the signal manifold, or in other words the matrix C becomes either singular or ill conditioned. In this case, some of the eigenvalues λ_α will either be zero or close to zero. The

remedy is then to remove such eigenvalues or degrees of freedom from the χ^2, and we will then nevertheless obtain a χ^2 although with fewer degrees of freedom. This is acceptable if the number of degrees of freedom is not much smaller than p. It may be possible to avoid such a situation by choosing the original vectors $\mathbf{h}_\alpha$ judiciously.

The χ^2 discriminators constructed so far in the literature (for example, the traditional χ_t^2 and the ambiguity χ^2, which we are going to describe in the next section) are special cases of this generic χ^2. *But the main power of this procedure is that the $\mathbf{h}_\alpha$ may be chosen to be along the glitches that occur in the data*, which could yield a very powerful discriminator. This may be difficult to implement in practice because of the diverse morphology of glitches; nevertheless, we may be able to do so for a particular class of glitches. This may be a direction worth following in the future.

V. THE AMBIGUITY χ^2 DISCRIMINATOR

In this section we construct the ambiguity χ^2 which is a special case of the generic χ^2 described in the previous section. The idea is to (i) choose the vectors $\mathbf{h}_\alpha$ from the template bank and (ii) as far as possible from the vicinity of the trigger template. The reason for choosing the test waveform vectors from the template bank is that we can use the filtered output of the template bank as an input to a χ^2 discriminator and therefore obtain a χ^2 discriminator with very little additional computational cost. The construction essentially then follows the procedure we had adopted in the case of the generic χ^2 in the previous section.

A. The construction

We choose several waveform vectors surrounding this trigger waveform vector. In general we will take the number of waveforms to be p. We choose the vectors at the points $(t_c + \Delta t_{ci}, \vartheta + \Delta\vartheta_i)$ and at any one of the phases $\phi_c = 0$ and $\phi_c = \pi/2$, where $i = 1, 2, \ldots, p$; for example, for $i = 1, 2, \ldots, p_1$, we may choose $\phi_c = 0$ and for the rest from $i = p_1 + 1, \ldots, p$ we may choose $\phi_c = \pi/2$. The $\Delta\vartheta_i$ are chosen such that they correspond to the templates in the bank. On the other hand, we can choose almost any value of Δt_{ci} because the filtered output is almost continuous in the parameter t_c. We denote these vectors by $\mathbf{h}_{0i}$ and $\mathbf{h}_{\pi/2i}$ corresponding to the phases 0 and $\pi/2$, respectively; they are time-shifted versions of the templates at $\vartheta + \Delta\vartheta_i$ by the amount $t_c + \Delta t_{ci}$.

We consider a signal whose ϑ exactly matches with the template. In general, there will be a mismatch between the signal parameters and the template parameters, but because the templates are densely spaced, the mismatch is expected to be small. We will show in Sec. VII that a small mismatch does not change the χ^2 value significantly and, therefore, the test is robust against a small mismatch of parameters. This is also validated by our numerical simulations

discussed in Sec. V C. The waveform vectors satisfy the following orthonormality conditions:

$$(\mathbf{h}_0(0), \mathbf{h}_0(0)) = (\mathbf{h}_{\pi/2}(0), \mathbf{h}_{\pi/2}(0))$$
$$= (\mathbf{h}_{0i}, \mathbf{h}_{0i}) = (\mathbf{h}_{\pi/2i}, \mathbf{h}_{\pi/2i}) = 1, \qquad (5.1)$$

$$(\mathbf{h}_0(0), \mathbf{h}_{\pi/2}(0)) = (\mathbf{h}_{0i}, \mathbf{h}_{\pi/2i}) = 0. \qquad (5.2)$$

It is important to realize that the scalar products other than those listed above *need not vanish*. That is, in general, $(\mathbf{h}_{0i}, \mathbf{h}_{0j}) \neq 0$ or $(\mathbf{h}_{\pi/2i}, \mathbf{h}_{\pi/2j}) \neq 0$, etc., for any i or j. Similarly, scalar products of these waveform vectors with the trigger data vectors need not vanish. This has the important consequence that the scalar products of the data with the waveform vectors can lead to random variables having nonvanishing covariances. This fact is taken into consideration when constructing the ambiguity χ^2 statistic. Referring back to Sec. IV, these are precisely the vectors $\mathbf{h}_\alpha$ pertaining to the generic χ^2, and we may therefore denote the pairs of subscripts of the vectors $(0i)$ or $(\pi/2i)$ by the single greek subscripts α, β.

Since we have chosen the test vectors from the template bank, the scalar products appearing in Eq. (4.3) are just the ambiguity functions. Further, these scalar products corresponding to the two phases are readily available from the output of the search pipeline. Accordingly, we define two ambiguity functions at the waveform vector locations as

$$\mathcal{H}_{0\alpha} = (\mathbf{h}_0(0), \mathbf{h}_\alpha), \qquad \mathcal{H}_{\pi/2\alpha} = (\mathbf{h}_{\pi/2}(0), \mathbf{h}_\alpha). \quad (5.3)$$

We may then write the differences between the expected and observed correlations in terms of the ambiguity functions as follows:

$$\Delta c_\alpha(\mathbf{x}) = c_\alpha^o(\mathbf{x}) - c_\alpha^e = (\mathbf{x}, \mathbf{h}_\alpha)$$
$$- (\mathbf{x}, \mathbf{h}_0(0))\mathcal{H}_{0\alpha} - (\mathbf{x}, \mathbf{h}_{\pi/2}(0))\mathcal{H}_{\pi/2\alpha},$$
$$\equiv (\mathbf{x}, \Delta\mathbf{h}_\alpha), \qquad (5.4)$$

where the projected vector $\Delta\mathbf{h}_\alpha$ now becomes

$$\Delta\mathbf{h}_\alpha = \mathbf{h}_\alpha - \mathbf{h}_0(0)\mathcal{H}_{0\alpha} - \mathbf{h}_{\pi/2}(0)\mathcal{H}_{\pi/2\alpha}. \qquad (5.5)$$

The covariance matrix can also be expressed in terms of the ambiguity functions. We obtain

$$C_{\alpha\beta} = (\Delta\mathbf{h}_\alpha, \Delta\mathbf{h}_\beta) = (\mathbf{h}_\alpha, \mathbf{h}_\beta) - \mathcal{H}_{0\alpha}\mathcal{H}_{0\beta} - \mathcal{H}_{\pi/2\alpha}\mathcal{H}_{\pi/2\beta}. \qquad (5.6)$$

We then just follow the same steps as those for the generic χ^2 and obtain the ambiguity χ^2 as given by Eq. (4.19). For convenience we also present the formula here as well,

$$\chi^2 = \Delta c_\alpha [C^{-1}]^{\alpha\beta} \Delta c_\beta. \qquad (5.7)$$

Let us next estimate the computational costs involved in computing the ambiguity χ^2. We note that all the scalar products with the data are available from the matched-filtering stage of the search pipeline. So no additional FFT cost is involved. In computing Δc_α, we see from Eq. (5.4) that $\mathcal{O}(p)$ operations are involved. The matrix C, the orthogonal matrix O, and the eigenvalues λ_α can be precomputed. This becomes easy in the spinless case where, by choosing chirp times as parameters, the scalar products weakly depend on the parameters. So we could divide the parameter space into a few regions and compute C, O, and λ_α in each region. Then, the computation of $\Delta \mathbf{c}'$ and the multiplication of $\Delta \mathbf{c}$ by O require $\mathcal{O}(p^2)$ operations, and finally computing χ^2 requires $\mathcal{O}(p)$ more operations. Therefore, we have obtained a χ^2 statistic with very little overhead cost. In the next section, we apply this statistic to actual numerical examples and examine how it works.

B. Ambiguity χ^2 for the Newtonian waveform and single phase

To demonstrate how the ambiguity χ^2 works, we consider the simplest possible case of the Newtonian waveform. The Newtonian waveform depends only on a single mass parameter, namely, the chirp mass $\mathcal{M}$ or, equivalently, the chirp time τ_0, as defined in Eq. (2.7). This reduction in one dimension of the parameter space is important because it allows us to exhibit the method conveniently via two-dimensional contour diagrams of ambiguity functions in the parameters τ_0 and t_c. Second, we restrict ourselves to one phase (the dominant phase). There is no fundamental

reason for doing this; we only do this to simplify the procedure and demonstrate it in a lucid manner. Consider a trigger from a template with a certain ϑ and t_c and assume that $c_0 > c_{\pi/2}$. If this is not the case, then we can easily exchange the roles of the templates corresponding to $\phi_c = 0$ and $\phi_c = \pi/2$ in the expressions. Assuming, without loss of generality, that $c_0 > c_{\pi/2}$, we use the vectors $\mathbf{h}_{0i}, i = 1, 2, \ldots, p$ corresponding to zero phase to construct the χ^2 statistic. We drop the greek subscripts for this subsection only.

In Fig. 2, we plot the ambiguity functions $\mathcal{H}_0(\Delta t_c, \Delta\tau_0) = (\mathbf{h}_0(0), \mathbf{h}_0(t_c + \Delta t_c, \tau_0 + \Delta\tau_0))$ and $\mathcal{H}_{\pi/2}(\Delta t_c, \Delta\tau_0) = (\mathbf{h}_{\pi/2}(0), \mathbf{h}_0(t_c + \Delta t_c, \tau_0 + \Delta\tau_0))$ for the aLIGO PSD, where we have used the obvious notation of the subscript identifying the phase ϕ_c. These ambiguity functions can be precomputed.

Let us consider the situation depicted in Fig. 2. Let us consider a trigger signal with amplitude $A = 10$ and that is dominated by c_0. We now need to choose waveform vectors around this trigger. Just to illustrate the workings of this test, we choose only four points, i.e., $p = 4$. The filtered output is available at the points indicated by the dashed lines. Accordingly, we choose the points at $\Delta\tau_0 \simeq \pm 0.1534$ sec. This is twice the distance where the match falls to 0.97. We also choose $\Delta t_c = \pm 0.0159$ sec. This gives four points placed symmetrically about the trigger in the form of a rectangle. We now calculate the expected values of c^e_{0i} from the trigger corsrelations $c_0(0)$, $c_{\pi/2}(0)$ and the values of the ambiguity functions at these points. We then read off observed values c^o_{0i} from the filtered output and take differences to obtain the vector $\Delta \mathbf{c}$. The covariance matrix computed for these points is
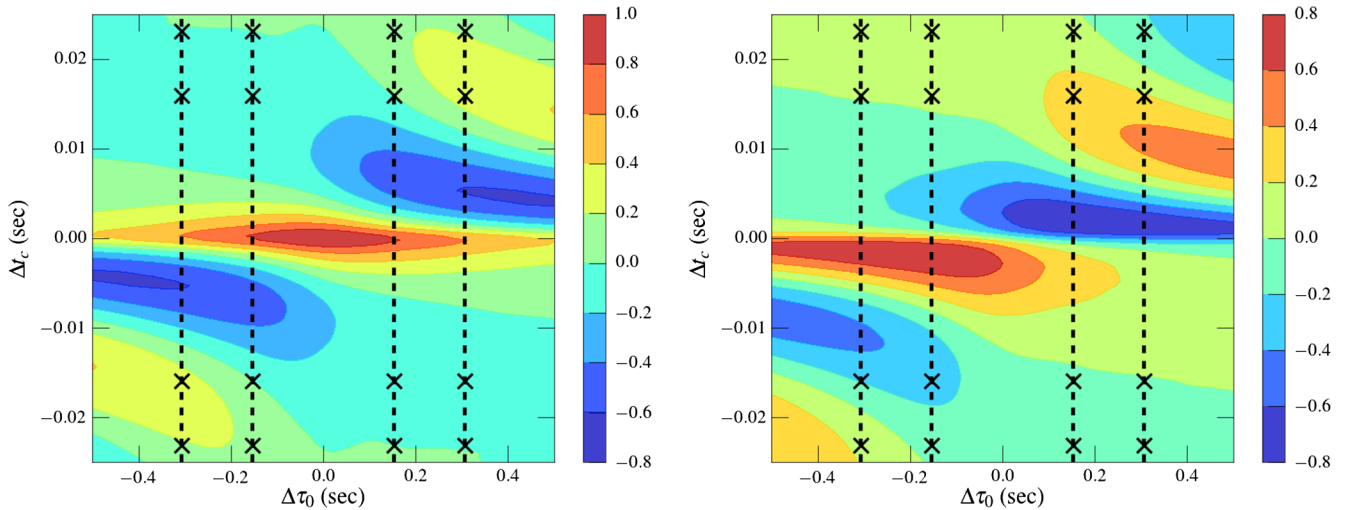

FIG. 2. Plots of the ambiguity functions $\mathcal{H}_0(\Delta t_c, \Delta\tau_0)$ (left) and $\mathcal{H}_{\pi/2}(\Delta t_c, \Delta\tau_0)$ (right) around the trigger template for a $1.4 - 1.4 \, M_\odot$ binary with $\Delta t_c = \Delta\tau_0 = 0$. For a mismatch of $\epsilon = 0.03$ and aLIGO noise, one finds $\Delta\tau_0 \sim 0.076$ sec. The templates in τ_0 are placed at twice this distance. The dashed vertical lines show the points at which the filtered output is available. The crosses mark the templates chosen to construct the χ^2 in the example given here.

$$C = \begin{bmatrix} 0.9243 & 0.0138 & 0.5781 & -0.0161 \\ 0.0138 & 0.9972 & 0.00046 & 0.5781 \\ 0.5781 & 0.00046 & 0.9972 & 0.0138 \\ -0.0161 & 0.5781 & 0.0138 & 0.9243 \end{bmatrix}. \quad (5.8)$$

In order to deduce $\Delta \mathbf{c}'$ we require the orthogonal transformation O that diagonalizes C. This turns out to be

$$O = \begin{bmatrix} -0.5185 & 0.5123 & 0.4808 & -0.4874 \\ 0.4808 & 0.4874 & 0.5185 & 0.5123 \\ 0.4808 & -0.4874 & 0.5185 & -0.5123 \\ -0.5185 & -0.5123 & 0.4808 & 0.4874 \end{bmatrix}. \quad (5.9)$$

As explained before, this transformation is the key to extracting the eigenvalues of C, which in this case are 0.35941223, 0.40363689, 1.53366376, and 1.5465145, in increasing order.

We can now use these results to compute the χ^2 for any data vector $\mathbf{x}$. One can verify that when $\mathbf{x}$ is just the signal, one obtains $\chi^2 = 0$. For a realization of Gaussian stationary noise with zero mean and the aLIGO PSD, we find the average value to be $\langle \chi^2 \rangle \sim 4.27$. The expected value is 4 because we have four degrees of freedom. The value we obtained in this case is close to the expected value. For a sine-Gaussian glitch with the quality factor $Q = 5$ and central frequency $f_0 = 60$ Hz that gives a trigger SNR of 10, the $\chi^2 \sim 187$.

We can increase the number of waveform vectors to a larger number, say, $p = 16$. We then choose the points $\Delta\tau_0 = \pm 0.1534$, ± 0.3068 and $\Delta t_c = \pm 0.01589$, ± 0.02315. These 16 points are also symmetrically placed around the trigger. While we do not explicitly present the covariance matrix, the orthogonal transformation matrix, and the eigenvalues (since it is cumbersome), nonetheless we evaluate the χ^2 for the aforementioned sine-Gaussian glitch with these waveform vectors and find its value to be ~ 365, which is much greater than the expected value of 16. These large values of χ^2 arise because the match of the glitch with the vectors surrounding the trigger is very much different from the ambiguity functions. We show this match for $\phi_c = 0$ surrounding the trigger in Fig. 3.

C. Effect of mismatch

In this subsection we investigate the effect of mismatch between the signal and the template parameters on the ambiguity χ^2. A mismatch can occur because, in principle, the signal can have any parameters continuously distributed over the parameter space, while the templates are necessarily placed at a discrete set of points stipulated by the minimal match [17]. However, owing to the mismatch, in our χ^2 construct the signal will not be subtracted completely, and there will be a residual which we have denoted
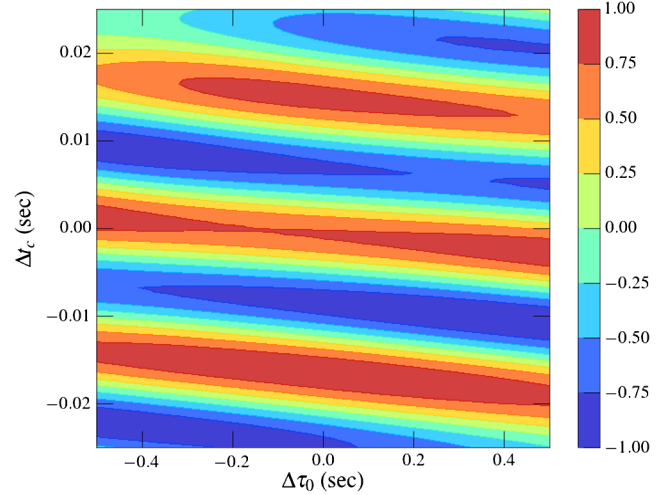

FIG. 3. The match of the sine-Gaussian glitch with $Q = 5$ and $f_0 = 60$ Hz with the normalized phase-zero waveforms. The match is seen to be very different from that of a signal, indicating that the ambiguity χ^2 is effective in separating glitches from signals.

by $\delta\mathbf{h}$ in Sec. VII. This is expected since the signal will not be orthogonal to the chosen subspace.

The results of our investigation into the effects of mismatch are presented in Fig. 4, where we use the spinless TaylorF2 approximant to model both signal and template waveforms. At any given instance, we take the signal to have parameters corresponding to a single point in the left plot of that figure. However, we always assume that the trigger template has parameter values shown by the red star in that plot. We numerically compute the ambiguity χ^2 values (depicted in color) in the vicinity of the trigger template for signals with parameters chosen in the ranges shown in the plot. We assume the noise to be zero. Note that we have chosen the trigger template well inside the allowed space parametrized by the two chirp times τ_0 and τ_3. The templates surrounding the trigger template (shown in white) are search templates, which may or may not be the same templates that are used to calculate the ambiguity χ^2 statistic. The signal can be anywhere in the neighborhood of the trigger template. Here, for the purpose of illustration, we use 14 of these search templates to compute the ambiguity χ^2, in the aLIGO noise PSD and with a lower cutoff frequency of 30 Hz; however, there are only eight degrees of freedom because after diagonalizing the covariance matrix we consider only those eigenvalues that are larger than 0.05.[2] Ellipses for 0.97 (inner) and 0.95 (outer) matches with the trigger template are also shown for reference. The templates that have the best chance of being triggered by a signal with the same parameters as the trigger template lie along or close to the semimajor axis of the

---

[2]We use this threshold on the eigenvalues in all of our numerical simulations in order to select the degrees of freedom.
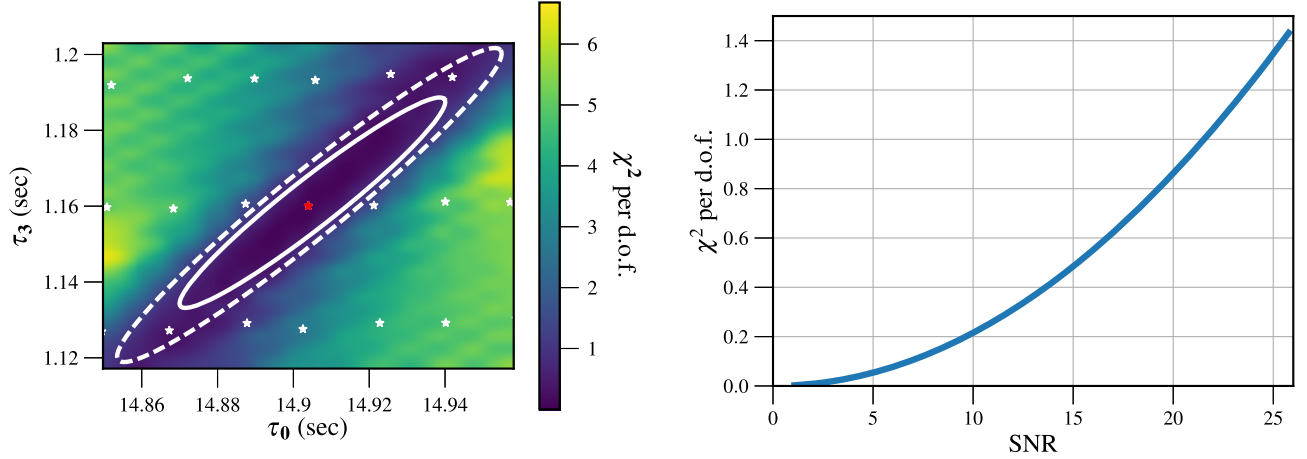
FIG. 4. The left panel shows the effect of mismatch on the ambiguity χ^2 (in color) as a function of the parameters τ_0 and τ_3 in the absence of noise. The trigger template is placed in the center of the plot and shown as a red star. For reference, ellipses at matches of 0.97 (inner) and 0.95 (outer) are shown. The six adjacent templates (shown as white stars) lie in the blue region, indicating that the effect of mismatch is sufficiently small ($\chi^2 \sim 1$ per degree of freedom) for the χ^2 to be effective. The right panel shows the χ^2 per degree of freedom versus the SNR for a fixed mismatch of 0.02. We observe that the χ^2 value increases quadratically with SNR.

match ellipses. One immediately observes that if a signal has the same parameter values as any of the six templates surrounding the trigger template (center) and lying in the blue region, then its χ^2 per degree of freedom will be ~ 1. One concludes from this observation that a small mismatch of this order does not pose any serious problem for implementing the ambiguity χ^2.

We also show how the χ^2 depends on the SNR in the right panel of Fig. 4. We take a fixed match of 0.98 for this purpose and vary the SNR. We find that the χ^2 increases quadratically with the SNR in accordance with Eq. (7.5).

D. χ^2 analysis for simulated data

In this section we test the performance of the ambiguity χ^2 on simulated data. We prepare the simulated data segments with Gaussian colored noise using the aLIGO design PSD. The lower cutoff is set to be 30 Hz, which is the same as that chosen for the first observing run of the aLIGO detectors. In the so-constructed Gaussian noise data we inject simulated GW signals and glitches. We have used nonspinning templates for our simulations. For the binary black hole (BBH) case, we choose component masses uniformly sampled in the range $5\,M_\odot$–$10\,M_\odot$. For the neutron star–black hole (NSBH) case, the black hole masses are uniformly sampled in the range $5\,M_\odot$–$10\,M_\odot$ while neutron star masses are uniformly sampled in the range $1.3\,M_\odot$–$2\,M_\odot$. There are 60 BBH signals and 40 NSBH signals injected in the simulated noise whose SNRs range between 4–40, assuming a uniform volume distribution of sources. Further, we inject 100 sine-Gaussian (SG) glitches [23–25] for both BBH and NSBH cases. One of the reasons for our choice of SG glitches is that the sine-Gaussians form a useful basis on which many of the glitches in real data have strong projections [26,27]. The central frequencies for these SG injections are chosen in the

range [30, 500] Hz, spaced uniformly on a logarithmic scale, while the quality factors are uniform in the range [3,10]. We also inject 100 Gaussian glitches [25] in both cases with standard deviations ranging from 20 to 100. For both the glitch injections, the SNR is maximized over the template banks such that it ranges from 4 to 40, as is done for GW injections. While such glitch injections are useful to an extent, real data must eventually be used for a proper test of the performance of these discriminators.

We also study the performance of the ambiguity χ^2 when the data contain only colored Gaussian noise. For this purpose we take 100 independent realizations. We compute the ambiguity χ^2 for these injections as follows. In the ambiguity χ^2 (or more generally in a bank χ^2), the trigger template is a random variable. However, when there is a trigger with SNR above a threshold of 7 or 8, the distribution for the trigger template is highly peaked and the trigger template is essentially pinned down. This happens for both GW signals and glitches. Therefore, the ambiguity χ^2 has an approximately χ^2 distribution for signals, and a noncentral χ^2 distribution for glitches with the noncentrality parameter $\|\mathbf{g}_S\|^2$. However, in the case of Gaussian noise where the SNRs are low, the trigger template jumps from one noise realization to another, such that the subspace $\mathcal{S}$ changes violently from one noise realization to another. Thus, in the χ^2 the terms of the form $\mathbf{n} \cdot \mathbf{e}_\alpha$ are not Gaussian distributed at all as $\mathbf{e}_\alpha$ is a random variable. Consequently, the χ^2 possibly has some other distribution (which so far we have not found) giving rise to high values of the statistic. Note that in the traditional χ^2 the subspace $\mathcal{S}$ is fixed before the search such that Gaussian noise gives a true χ^2 distribution. We therefore arbitrarily fix the trigger template. This gives a true χ^2 distribution for Gaussian noise injections.
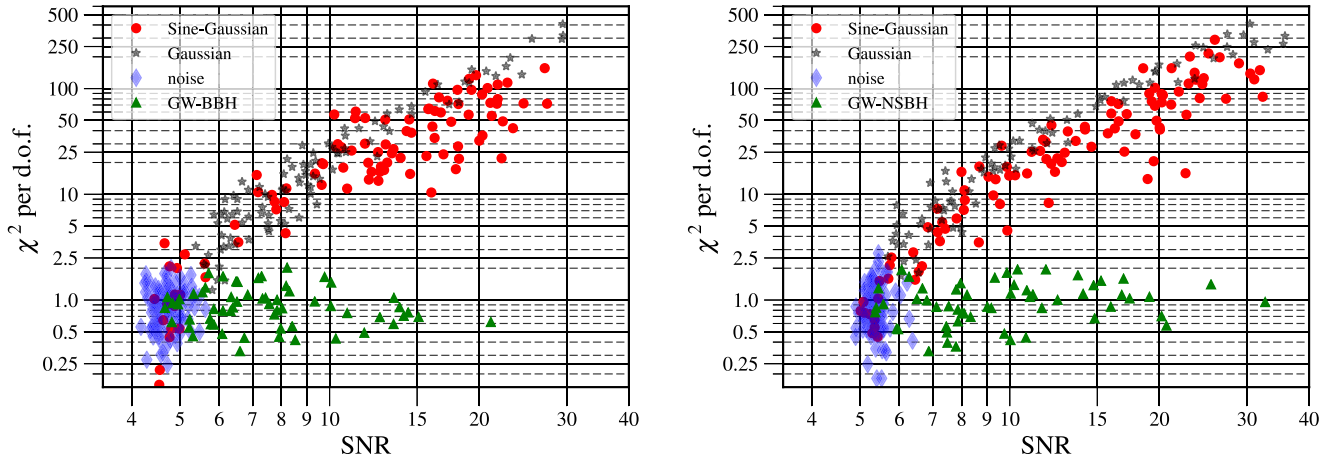
FIG. 5.   The χ^2 per degree of freedom versus SNR for triggers of CBC signals (green triangles), sine-Gaussian glitches (red circles), Gaussian glitches (grey stars), and Gaussian noise (blue diamonds). The left panel pertains to BBH injections and the right one to NSBH injections. We observe that in general the ambiguity χ^2 separates out CBC signal triggers from noise triggers.

In Fig. 5, we plot the ambiguity χ^2 per d.o.f. as a function of SNR for the BBH case (left) and NSBH case (right). The green triangles are triggers from the CBC signals and we see that they generally have low χ^2 values. The red circles are the triggers from the sine-Gaussian glitches and are seen to have high χ^2 values. The triggers from Gaussian noise are shown as blue diamonds which lie in the neighborhood of the origin; they have both low SNR and χ^2 values. We also plot triggers from Gaussian glitches which also have high χ^2 per d.o.f., as expected. Therefore, we observe that the ambiguity χ^2 has the important property of separating the noise triggers from the signal triggers and thus is a good candidate as a χ^2 discriminator. For better performance it may be possible to tune the ambiguity χ^2.

Note that fixing the trigger template when the data are purely Gaussian noise was done only to show that we get a χ^2 distribution in such a case. If we do not fix the trigger template as remarked before, we only get a higher value for the χ^2 statistic, but since it is in fact noise, it will not affect the implementation of the ambiguity χ^2.

VI. GENERALIZATION TO THE COHERENT MULTIDETECTOR CASE

It is easy to generalize the framework to the coherent multidetector case [18,19]. We use the results from the above works to generalize to the multidetector case. We do not however go into the details. Following Refs. [11,28], we can write the signal in detector I as

$$h^I(t) = \mathcal{A}^\mu h^I_\mu(t), \qquad (6.1)$$

where $\mathcal{A}^\mu$ are amplitudes depending on the distance to the source, initial phase, polarization angle, and inclination of the source, and $h^I_\mu(t)$ are defined by

$$h^I_1(t) = F^I_+ h_0(t^I), \qquad h^I_2(t) = F^I_\times h_0(t^I),$$
$$h^I_3(t) = F^I_+ h_{\pi/2}(t^I), \qquad h^I_4(t) = F^I_\times h_{\pi/2}(t^I), \qquad (6.2)$$

where $F^I_{+,\times}$ are the antenna pattern functions of detector I and t^I is the retarded time in detector I.

Consider that there are M detectors; then, the multi-detector signal vector can be written as

$$\mathbf{h}(t) = (h^1(t), h^2(t), \dots, h^M(t)), \qquad (6.3)$$

where now the signal vector belongs to the direct sum of Hilbert spaces $\mathcal{D}_I$, where $\mathcal{D}_I$ is the Hilbert space of the data trains pertaining to the detector I. We denote this space by

$$\mathcal{D}_{\text{network}} = \mathcal{D}_1 \oplus \mathcal{D}_2 \oplus \dots \oplus \mathcal{D}_M. \qquad (6.4)$$

The vector space $\mathcal{D}_{\text{network}}$ becomes a Hilbert space if we define a scalar product on it. For uncorrelated noise between different detectors, a natural scalar product on $\mathcal{D}_{\text{network}}$ is just the sum of the scalar products corresponding to individual detectors [11,18]. This is the scalar product we use on $\mathcal{D}_{\text{network}}$ in the context of the network.

From Eq. (6.2), the four network signal waveform vectors $\mathbf{h}_\mu \in \mathcal{D}_{\text{network}}$ are given by

$$\mathbf{h}_\mu = (h^1_\mu(t), h^2_\mu(t), \dots, h^M_\mu(t)), \qquad \mu = 1, 2, 3, 4. \qquad (6.5)$$

As shown in Ref. [11], in the dominant polarization frame (which was also used in Ref. [18]), these four vectors are mutually orthogonal (though not necessarily orthonormal). This construct then leads to the coherent statistic which is also essentially the F statistic (except for a factor of 2) first defined for continuous wave sources in Ref. [29].

To construct the χ^2 for a network, it is only necessary to select a subspace $\mathcal{S}_{\text{network}} \subset \mathcal{D}_{\text{network}}$ which is orthogonal to

each of the $\mathbf{h}_\mu$. This can be easily done. Then the network χ^2 is just the square of the L_2 norm of a data vector $\mathbf{x} \in \mathcal{D}_{\text{network}}$ projected into the subspace $\mathcal{S}_{\text{network}}$.

In the single-detector case we had just two basis waveforms corresponding to the phases $\phi_c = 0, \pi/2$. In the coherent multidetector case we have four basis waveforms corresponding to the two phases and the two polarizations $+$ and $\times$. If so desired, an ambiguity χ^2 for the multidetector case can be designed based on selecting only one of the signal waveforms $\mathbf{h}_\mu$ as was done in the single-detector case in Sec. V B. We do not pursue this point here any further and leave it for future investigation.

VII. EFFECT OF SMALL MISMATCH IN PARAMETERS BETWEEN THE SIGNAL AND THE TEMPLATE

Although the kinematical parameters t_c and ϕ_c can be searched over continuously by the FFT and analytical maximization, the mass parameters are searched with a discrete bank of templates. In practice, the templates are laid such that the match maximized over t_c and ϕ_c is at least 97% or the mismatch is at most 3%. It is very unlikely that a signal will have exactly the parameters as those of the templates in the bank; in general, there will be a mismatch. In this section, we estimate the effect of this mismatch on the value of the χ^2. Although the χ^2 vanishes for a signal matching the template (the subspace $\mathcal{S}$ is orthogonal to the template), it will not in general be orthogonal to the signal because of the mismatch. Nevertheless, if the mismatch is small, we will show that this effect on χ^2 is small and we will derive a bound on this value.

We perform the analysis in general. We consider a signal $\mathbf{s}$ with parameters $\vartheta^a + \Delta\vartheta^a$ and amplitude A; then, we have $\mathbf{s}(\vartheta^a + \Delta\vartheta^a) = A\mathbf{h}(\vartheta^a + \Delta\vartheta^a)$, where $\mathbf{h}(\vartheta^a)$ denotes a normalized waveform. The trigger template has parameters ϑ^a and we write $\mathbf{h}(\vartheta^a + \Delta\vartheta^a) = \mathbf{h}(\vartheta^a) + \delta\mathbf{h}$. Since both $\mathbf{h}(\vartheta^a + \Delta\vartheta^a)$ and $\mathbf{h}(\vartheta^a)$ are normalized, to the first order, $\delta\mathbf{h}$ is orthogonal to $\mathbf{h}(\vartheta^a)$. *However, it is important to realize that $\delta\mathbf{h}$ may not lie in the subspace $\mathcal{S}$ corresponding to the trigger template. In fact, to the first order, $\delta\mathbf{h} \in \mathcal{N}_{\chi^2}$.* We will have more to say about this later in this section. For now, we relate the norm of $\delta\mathbf{h}$ to the ambiguity function. We have the following relation:

$$\begin{aligned}\|\delta\mathbf{h}\|^2 &= \|\mathbf{h}(\vartheta^a + \Delta\vartheta^a) - \mathbf{h}(\vartheta^a)\|^2 \\ &= \|\mathbf{h}(\vartheta^a + \Delta\vartheta^a)\|^2 + \|\mathbf{h}(\vartheta^a)\|^2 \\ &\quad - 2(\mathbf{h}(\vartheta^a), \mathbf{h}(\vartheta^a + \Delta\vartheta^a)) \\ &\equiv 2(1 - \mathcal{H}(\vartheta^a, \Delta\vartheta^a)),\end{aligned} \quad (7.1)$$

where $\mathcal{H}(\vartheta^a, \Delta\vartheta^a) = (\mathbf{h}(\vartheta^a), \mathbf{h}(\vartheta^a + \Delta\vartheta^a))$ is the ambiguity function. The mismatch ϵ is related to the ambiguity function by

$$\mathcal{H}(\vartheta^a, \Delta\vartheta^a) \geq 1 - \epsilon. \quad (7.2)$$

Therefore, we have $1 - \mathcal{H}(\vartheta^a, \Delta\vartheta^a) \leq \epsilon$ which leads to the following inequality:

$$\|\delta\mathbf{h}\|^2 \leq 2\epsilon. \quad (7.3)$$

From the above relation, we can now obtain a bound for χ^2. The χ^2 of the signal $\mathbf{s}(\vartheta + \Delta\vartheta^a)$ mismatching with the template $\mathbf{h}(\vartheta^a)$ is

$$\chi^2(\mathbf{s}) = \chi^2(A(\mathbf{h} + \delta\mathbf{h})). \quad (7.4)$$

Projecting orthogonal to $\mathbf{h}$, we must have

$$\chi^2(\mathbf{s}) = A^2\|\delta\mathbf{h}_\mathcal{S}\|^2 \leq A^2\|\delta\mathbf{h}\|^2 \leq 2A^2\epsilon. \quad (7.5)$$

This bound was obtained in Ref. [8] [Eq. (6.24) therein] for χ_t^2. But here we have obtained this bound generally without referring to any specific χ^2. Even with this bound, one can see that for $\epsilon = 0.03$ and $A = 10$, $\chi^2(\delta\mathbf{h}) \leq 6$. For the ambiguity χ^2 proposed here with 16 templates, a numerical evaluation leads to $\chi^2 \sim 4.3$.

The above result does not depend on the degrees of freedom because we have obtained this bound from $\delta\mathbf{h}$ which is in $\mathcal{N}_{\chi^2}(\mathbf{h})$. However, in Ref. [8] a tighter bound was obtained. In general, we may be able to obtain a better bound by projecting onto the subspace $\mathcal{S}$ since $\|\mathbf{h}_\mathcal{S}\|^2 < \|\mathbf{h}\|^2$. But now the bound could depend on p, the dimension of $\mathcal{S}$. This is what happens, as is seen from the results obtained in Ref. [8]. Here we sketch the arguments given in Ref. [8] in the context of our framework.

For brevity, let us write the normalized mismatched signal vector as $\mathbf{h}' = \mathbf{h} + \delta\mathbf{h}$. We now project $\mathbf{h}'$ onto the $p - 1$-dimensional subspace $\mathcal{S}$ and obtain $\mathbf{h}'_\mathcal{S}$; then, $\chi^2 = A^2\|\mathbf{h}'_\mathcal{S}\|^2$. In order to make connection with Ref. [8] we need to go to p dimensions. We can write the basis vectors $\mathbf{e}_\alpha, \alpha = 1, 2, ..., p$, as linear combinations of $\Delta\mathbf{h}_\beta$ in terms of an orthogonal $p \times p$ matrix O given by

$$O = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & ... & 0 \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} & 0 & ... & 0 \\ ... & ... & ... & ... & ... & ... \\ \frac{1}{\sqrt{p}} & \frac{1}{\sqrt{p}} & ... & ... & ... & \frac{1}{\sqrt{p}} \end{bmatrix}. \quad (7.6)$$

That is,

$$[\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_{p-1}, \mathbf{e}_p]^T = \sqrt{p}O[\Delta\mathbf{h}_1, \Delta\mathbf{h}_2, ..., \Delta\mathbf{h}_p]^T, \quad (7.7)$$

where the symbol T denotes the transpose; it transposes the row vectors to column vectors and also effects $O^T O = 1$. Note that the last vector $\mathbf{e}_p = 0$ because of Eq. (3.10). Following the notation in Ref. [8] one finds that

$$\mathbf{h}' \cdot \Delta \mathbf{h}_\beta = \mathbf{h}' \cdot \left(\mathbf{h}_\beta - \frac{\mathbf{h}}{p}\right) = \left(\vartheta_\beta - \frac{1}{p}\right)\cos\theta \equiv \omega_\beta \cos\theta, \tag{7.8}$$

where $\cos\theta = 1 - \epsilon$ is the match, and θ is the "angle" between $\mathbf{h}$ and $\mathbf{h}'$, i.e., $\mathbf{h} \cdot \mathbf{h}' = \cos\theta$. Writing $\boldsymbol{\Omega} = [\omega_1, \omega_2, ..., \omega_p]^T$ and $\mathbf{h}'_\mathcal{S}$ as column vectors, we easily get

$$\mathbf{h}'_\mathcal{S} = [\mathbf{h}' \cdot \mathbf{e}_1, \mathbf{h}' \cdot \mathbf{e}_2, ..., \mathbf{h}' \cdot \mathbf{e}_p]^T = \sqrt{p}O\boldsymbol{\Omega}\cos\theta. \tag{7.9}$$

The contribution of the mismatch to χ_t^2 is

$$\chi_t^2 = A^2\|\mathbf{h}'_\mathcal{S}\|^2 = A^2\mathbf{h}'^T_\mathcal{S}\mathbf{h}'_\mathcal{S} = pA^2\boldsymbol{\Omega}^T O^T O\boldsymbol{\Omega}\cos^2\theta$$

$$= pA^2\cos^2\theta\boldsymbol{\Omega}^T\boldsymbol{\Omega} = pA^2\cos^2\theta\sum_{\beta=1}^{p}\omega_\beta^2. \tag{7.10}$$

In Ref. [8], the constraints on ω_β were derived by assuming that both the signal and template contribute the same power to each frequency bin. To the lowest order in ϵ, they are the following:

$$-\frac{\epsilon}{p} \le \omega_\beta \le \frac{2}{p}, \qquad \sum_{\beta=1}^{p}\omega_\beta = 0. \tag{7.11}$$

Subject to these constraints the χ_t^2 can be maximized. This happens when $\omega_1 = \epsilon(p-1)/p$, $\omega_2 = \omega_3 = ... = \omega_p = -\epsilon/p$ and the result to the lowest order in ϵ is $\chi_t^2 \sim (p-1)\epsilon^2 A^2$. (The $\cos^2\theta$ factor drops out because we are interested in the lowest order in ϵ.) This bound is smaller than the general bound $2\epsilon A^2$ only when $p\epsilon < 2$, that is, the degrees of freedom p are not too large. This is the second bound obtained for χ_t^2. We do not know whether some analogous bound exists in general.

We could go even further. Let us first pick a waveform $\mathbf{h}_0 \in \mathcal{P}$. If we were to choose the subspace $\mathcal{S}_0$ in such a way that it is orthogonal to *both* $\mathbf{h}_0$ and $\mathbf{h}_0 + \delta\mathbf{h}$ (this could be easily achieved since there is a lot of freedom in the choice of $\mathcal{S}$), then the effect of the mismatch could be completely nullified. However, there is not one $\delta\mathbf{h}$, but several waveforms $\mathbf{h}$ centered at say $\mathbf{h}_0$, with $\delta\mathbf{h} = \mathbf{h} - \mathbf{h}_0$, which lie within the region of mismatch determined by ϵ. This region is given by

$$E_0 = \{\mathbf{h} \in \mathcal{P}/\|\mathbf{h} - \mathbf{h}_0\|^2 < 2\epsilon\}. \tag{7.12}$$

We denote it by E_0 since it is essentially shaped like an ellipsoid (or hyperellipsoid in higher dimensions) for small mismatches, because expanding the ambiguity function to the quadratic order is sufficient. Note that $\|\delta\mathbf{h}\|^2$ is a quadratic form in $\Delta\vartheta^a$ and is in fact $2 \times g_{ab}\Delta\vartheta^a\Delta\vartheta^b$, where g_{ab} is the metric. (The factor of 2 appears because the metric is defined with this factor in Ref. [17]. However, if one uses the definition of the metric from Ref. [13], this factor of 2 does not appear.) Topologically, E_0 is an open

sphere with radius $\sqrt{2\epsilon}$. The value of χ^2 on E_0 will not be identically zero but could be made close to zero, by judiciously choosing the subspace $\mathcal{S}_0$ at $\mathbf{h}_0$ to mitigate the effect of the mismatch. In the maximization procedure of the statistic over the kinematical parameters, one may in fact need to focus only on a subset of E_0 so that the problem could be less involved.

We remark that the mismatch parameter ϵ used here is quadratic in the amplitude just as in Ref. [8] and, thus, it appears linearly in the bound in Eq. (7.4). However, in Ref. [11], the mismatch parameter ϵ was used in the amplitude and, therefore, it appeared as ϵ^2 in the bounds on the χ^2. It is also noteworthy that there may be reasons for mismatch other than the discreteness of the template bank; for instance, the model of the waveform may differ from the actual astrophysical signal, in which case a systematic error would be introduced, thereby leading to higher values of χ^2. We have not addressed this case here.

VIII. CONCLUSIONS

We have presented for the first time a general framework for unifying χ^2 discriminators that also unravels how new ones can be generated straightforwardly. We showed that these discriminators have the underlying mathematical structure of a vector bundle. The χ^2 statistic is just the L_2 norm of the data vector projected onto a subspace of the Hilbert space of data vectors orthogonal to the trigger template. This underscores the fact that the basis of this subspace is unimportant. Since the trigger could occur anywhere in the parameter space, we are dealing with a collection of subspaces orthogonal to each vector tracing out the signal manifold; a smooth choice leads to a vector bundle. The χ^2 can be visualized as a non-negative real-valued function on sections of the fibre bundle, which are in fact the vector fields obtained by the projection of the data vectors onto the subspaces. Apart from its elegance, the important practical insight that emerges from this formulation is the enormous flexibility available in defining χ^2 discriminators. We expect that this freedom could be used in *tuning* the χ^2 so that it can discriminate more decisively against frequently occurring glitches or glitches which pose difficulties to the search algorithms. In this context, we would like to point out the work which relates template bank triggers to the sine-Gaussian glitch parameters [23,24]. This may turn out to be useful in tuning the χ^2 to sine-Gaussian glitches, because it relates regions of signal parameter space which respond to the parameters of the glitches. The problem is of course more general: since we have a family of χ^2 to choose from, the tuning of the χ^2 to real data is intimately linked. One needs to optimize the χ^2 to those which are most effective on real data. This also raises the question of why the traditional χ_t^2 performs so well. We believe our general framework would be useful in understanding this issue. Our near future goal is to make headway in these general directions.

Using this formulation, we have constructed a generic χ^2 discriminator by choosing an arbitrary set of vectors in $\mathcal{D}$ and then projecting out their components parallel to the trigger templates. For instance, we could choose these vectors in the direction of the glitches—say, sine-Gaussians —and by subtracting out the projections on the trigger template construct $\mathcal{S}$ orthogonal to the trigger template. We could in fact choose the vectors to be a sine-Gaussian basis. This would be a useful χ^2 to discriminate against sine-Gaussians. Such a χ^2 would be viable if the cost in computing the χ^2 is small compared with the cost of the CBC search. Further, we have also defined a performance measure for a generic χ^2—a conditional probability—if the glitches can be modeled. We have obtained this probability explicitly for the sine-Gaussian glitches.

By choosing the vectors in $\mathcal{D}$ from the template bank, and carrying out the construction as in the generic case, we have proposed a new χ^2—the ambiguity χ^2—based on the behavior of the ambiguity function. This is because the matched filter outputs of a bank of templates centered around the parameters of a strong enough signal sample the ambiguity function of that signal. If a trigger corresponds to a glitch and not a signal then those outputs will not, in general, follow the ambiguity function of a signal with the parameters of the triggered templates. The advantage of the ambiguity χ^2 is that its computation entails negligible overhead cost since most of the inputs required to compute it are available from other steps of the search [9,20].

We have also investigated the effect of the mismatch between the signal and the templates on the ambiguity χ^2. The reason for the mismatch is generic because in the intrinsic parameters (like masses or spins) the parameter space is covered with a discrete bank of templates, and therefore the signal parameters are unlikely to match with any particular template. This results in a higher value of χ^2. We have shown that the effect of the mismatch does not pose any serious problem for the implementation if the templates lie sufficiently close to one another—say, within the usual mismatch of 0.03. The χ^2 value per degree of freedom is at most of the order of unity in the noise-free case if the signal parameters mismatch up to the adjacent template. In fact, although we have not optimized for this, another advantage of the freedom in the choice of the ambiguity χ^2 is that we may be able to mitigate the effects of mismatch of the signal with the templates

To demonstrate the power of the ambiguity χ^2, we carried out an extensive set of simulations with CBC signals and mock glitches separately added to data that was otherwise statistically Gaussian and stationary. The CBC signals were simulated with a wide range of astrophysical parameters corresponding to BBH and NSBH systems, involving stellar-mass black holes. The glitches were simulated to be of the Gaussian and sine-Gaussian types, with parameters (such as quality factor or central frequency) observed in real glitches. These data were then matched-filtered with

NSBH and BBH template banks, as in real searches. The results presented here show that the glitches are clearly separated from the signals. This is true even when the signals do not match the template parameters. Also, the triggers from Gaussian noise lie in the neighborhood of the origin. Thus, in this simulated numerical experiment, the ambiguity χ^2 is effective and yields encouraging results. Although our numerical simulations have been performed for the spinless TaylorF2 waveform, our discriminator is applicable to a wider class of waveforms: it has been successfully applied to other spinless waveforms, for example, IMRPhenomD [16]. We expect that the ambiguity χ^2 should be applicable more generally to spinning waveforms.

We also point out how the ambiguity χ^2 has an advantage over the bank χ^2 [10] and the autocorrelation χ^2 [30]. In the bank χ^2, in order to have the test templates as mutually orthogonal as possible, the templates must be well spread out in the parameter space. But then several of the templates tend to be away from the trigger (almost zero match with the trigger template), but these may also give zero match with the noise transient as well, reducing the discriminatory power of the χ^2; these degrees of freedom of the χ^2 do not contribute to the χ^2 value. On the other hand, for the ambiguity χ^2 we have the freedom to choose templates near the trigger template, as we do not have to worry about orthogonality. Therefore, we believe that it would be more effective in discriminating signals against the noise transient; at least the signal will have nonzero projection on the selected template vectors. It is clear that the directions probed by the χ^2 should be those of either the signal or the glitch in order for the χ^2 to be effective. Moreover, if the test templates are not chosen to be sufficiently orthogonal, the effective number of degrees of freedom could be less than the number of templates; the matrix C could have some small eigenvalues, leading to an overestimate of the number of degrees of freedom of the χ^2 discriminator. In the ambiguity χ^2, we remove these degrees of freedom by putting a lower cutoff on the eigenvalues, so we have a true estimate of the effectiveness of the χ^2. In the autocorrelation χ^2 [10,30], the SNR time series is considered such that only the vectors along the time dimension are considered for consistency, whereas in the ambiguity χ^2 the consistency is also evaluated in the other dimensions of the parameter space besides time, such as the masses.

We also indicated how the ambiguity χ^2 discriminator could be generalized to the coherent multidetector case. For the case of uncorrelated noise between different detectors, the generalization is straightforward. We have not pursued this point here any further but may do so in the future.

Finally, we would like to mention that the vector bundle structure of the χ^2 is more generally applicable. In this paper we have discussed a specific signal model of the

CBC, but if the signal model is different (either due to a modification of the CBC model used here or because of a completely different astrophysical source) the signal manifold $\mathcal{P}$ will differ, but the overall mathematical structure of the χ^2 (namely, that of the vector bundle) will remain the same. Therefore, the formalism for the χ^2 presented here is valid in more general settings.

ACKNOWLEDGMENTS

[1] B. P. Abbott *et al.* (LIGO Scientific and VIRGO Collaborations), Phys. Rev. Lett. **116**, 061102 (2016).

[2] B. P. Abbott *et al.* (LIGO Scientific and VIRGO Collaborations), Phys. Rev. Lett. **116**, 241103 (2016).

[3] D. Martynov *et al.* (LIGO Scientific and VIRGO Collaborations), Phys. Rev. D **93**, 112004 (2016).

[4] J. Aasi *et al.* (LIGO Scientific and VIRGO Collaborations), Classical Quantum Gravity **32**, 115012 (2015).

[5] S. Bose, B. Hall, N. Mazumder, S. Dhurandhar, A. Gupta, and A. Lundgren, J. Phys. Conf. Ser. **716**, 012007 (2016).

[6] C. W. Helstrom, *Statistical Theory of Signal Detection*, 2nd ed. (Pergamon, London, 1968).

[7] B. S. Sathyaprakash and S. V. Dhurandhar, Phys. Rev. D **44**, 3819 (1991).

[8] B. Allen, Phys. Rev. D **71**, 062001 (2005).

[9] B. Allen, W. G. Anderson, P. R. Brady, D. A. Brown, and J. D. E. Creighton, Phys. Rev. D **85**, 122006 (2012).

[10] C. Hanna, Ph.D. Thesis, Louisiana State University, 2008.

[11] I. W. Harry and S. Fairhurst, Phys. Rev. D **83**, 084002 (2011).

[12] S. Babak, H. Grote, M. Hewitson, H. Lck, and K. A. Strain, Phys. Rev. D **72**, 022002 (2005).

[13] R. Balasubramanian, B. S. Sathyaprakash, and S. V. Dhurandhar, Phys. Rev. D **53**, 3033 (1996).

[14] A. Nitz, Ph.D. Thesis, Syracuse University, 2015.

[15] A. Buonanno, B. Iyer, E. Ochsner, Y. Pan, and B. S. Sathyaprakash, Phys. Rev. D **80**, 084043 (2009).

[16] S. Khan, S. Husa, M. Hannam, F. Ohme, M. Pürrer, X. Forteza, and A. Bohe, Phys. Rev. D **93**, 044007 (2016).

[17] B. Owen, Phys. Rev. D **53**, 6749 (1996).

[18] A. Pai S. Dhurandhar, and S. Bose, Phys. Rev. D **64**, 042004 (2001).

[19] S. Bose, A. Pai, and S. V. Dhurandhar, Int. J. Mod. Phys. D **09**, 325 (2000).

[20] S. A. Usman *et al.*, Classical Quantum Gravity **33**, 215004 (2016)..

[21] S. V. Dhurandhar and B. S. Sathyaprakash, Phys. Rev. D **49**, 1707 (1994).

[22] B. F. Schutz, *Geometrical methods of mathematical physics* (Cambridge University Press, Cambridge, England, 1980).

[23] S. Bose, S. Dhurandhar, A. Gupta, and A. Lundgren, Phys. Rev. D **94**, 122004 (2016).

[24] T. D. Canton, S. Bhagwat, S. V. Dhurandhar, and A. Lundgren, Classical Quantum Gravity **31**, 015016 (2014).

[25] J. Powell, D. Trifirò, E. Cuoco, I. S. Heng, and M. Cavaglià, Classical Quantum Gravity **32**, 215012 (2015).

[26] S. Chatterji, L. Blackburn, G. Martin, and E. Katsavounidis, Classical Quantum Gravity **21**, S1809 (2004).

[27] S. Chatterji, Ph.D. Thesis, Massachusetts Institute of Technology, 2005.

[28] S. Bose, T. Dayanga, S. Ghosh, and D. Talukder, Classical Quantum Gravity **28**, 134009 (2011).

[29] P. Jaranowski, A. Krolak, and B. F. Schutz, Phys. Rev. D **58**, 063001 (1998).

[30] C. Messick *et al.*, Phys. Rev. D **95**, 042001 (2017).