Robust statistics for deterministic and stochastic gravitational waves in non-Gaussian noise. II. Bayesian analyses

Bruce Allen and Jolien D. E. Creighton

Department of Physics, University of Wisconsin–Milwaukee, P.O. Box 413, Milwaukee, Wisconsin 53201, USA

Éanna É. Flanagan

Newman Laboratory of Nuclear Studies, Cornell University, Ithaca, New York 14853-5001, USA

Joseph D. Romano

Department of Physical Sciences, University of Texas at Brownsville, Brownsville, Texas 78520, USA

In a previous paper (paper I), we derived a set of near-optimal signal detection techniques for gravitational wave detectors whose noise probability distributions contain non-Gaussian tails. The methods modify standard methods by truncating or clipping sample values which lie in those non-Gaussian tails. The methods were derived, in the frequentist framework, by minimizing false alarm probabilities at fixed false detection probability in the limit of weak signals. For stochastic signals, the resulting statistic consisted of a sum of an autocorrelation term and a cross-correlation term; it was necessary to discard "by hand" the autocorrelation term in order to arrive at the correct, generalized cross-correlation statistic. In the present paper, we present an alternative derivation of the same signal detection techniques from within the Bayesian framework. We compute, for both deterministic and stochastic signals, the probability that a signal is present in the data, in the limit where the signal-to-noise ratio squared per frequency bin is small, where the signal is nevertheless strong enough to be detected (integrated signal-to-noise ratio large compared to 1), and where the total probability in the non-Gaussian tail part of the noise distribution is small. We show that, for each model considered, the resulting probability is to a good approximation a monotonic function of the detection statistic derived in paper I. Moreover, for stochastic signals, the new Bayesian derivation automatically eliminates the problematic autocorrelation term.

I. INTRODUCTION AND SUMMARY

Most of the literature on gravitational-wave data analysis assumes that the detector noise is Gaussian. However, significant non-Gaussian tails have been a characteristic feature of the noise distributions in all gravitational wave detectors constructed to date. Standard detection strategies for both deterministic and stochastic signals, which were designed under the assumption of Gaussian noise, perform more poorly when non-Gaussian noise is present.

In a previous paper in this series [1] (henceforth paper I), we developed a new set of statistical signal-processing techniques to search for deterministic and stochastic gravitational waves in detector data. These techniques are *robust*, meaning that they will work well even if the detector noise is not Gaussian but falls into a broader statistical class that we expect includes realistic detectors. These new methods are similar to the older ones: one constructs matched filters to search for known waveforms or cross-correlates the instrument outputs at the different detector sites to search for a stochastic background. The essential difference is that the statistics are modified by truncation: detector samples that fall outside the central Gaussian-like part of the sample distribution are excluded from (or saturated when constructing) the measurement statistic. For both deterministic and stochastic signals, a robust statistic was found which performs better than the optimal linear filter in the case where the detector noise is non-Gaussian, and almost as well in the

Gaussian-noise case. Alternative methods for dealing with non-Gaussian noise for stochastic signals have been explored by Klimenko and Mitselmakher [2].

In paper I, we derived the statistics using the frequentist criterion of minimizing false alarm probabilities at fixed false detection probabilities in the limit of weak signals. In the present paper, we present an alternative derivation of the same signal detection techniques from within the Bayesian framework.

We start in Sec. II by reviewing the foundations of the two different approaches to determining detection statistics used in paper I and this paper. We review the locally optimal criterion used in paper I in Sec. II A. In Sec. II B we explain how Bayesian considerations lead to a unique choice of detection statistic, as discussed by Finn [3]. In Secs. III and IV of the paper, we compute, for a variety of different models and sets of assumptions, that unique Bayesian statistic. We show that for each case considered, the Bayesian statistic is equivalent to the statistics derived in paper I, in the sense that the false alarm versus false dismissal curves of the two statistics coincide to a good approximation.

The equivalence between the two types of statistic is valid only under certain approximations, discussed below. Under those approximations, the Bayesian statistics which we obtain are equivalent to a particular type of maximum likelihood statistic described in Refs. [4,5]. That type of maximum likelihood statistic differs from the type of maximum likelihood statistic considered previously in gravitational wave

data analysis in its treatment of noise parameters.

Section III deals with known, deterministic signals. In Sec. III A we consider the case of a known signal of unknown amplitude, incident on a single detector with white, Gaussian noise, where the noise variance is assumed to be known. For that case the Bayesian statistic is shown to be equivalent to the standard matched filtering statistic. Section III B generalizes this analysis by allowing the noise variance to be an unknown parameter, to be measured from the data; the same result is obtained. In Secs. III C and III D we consider the cases of white and colored non-Gaussian detector noise. For both of these cases we show that the Bayesian statistic is equivalent to the corresponding locally optimal statistic derived in paper I.

In Sec. IV we give a similar analysis of stochastic signals. In Sec. IV A we compute the Bayesian statistic for the case of a white stochastic signal, and of two co-aligned detectors with white Gaussian noise, where the noise variance is assumed to be known. The maximum likelihood statistic for this case was previously computed, in a more general context, by Finn and Romano [6]. In this case we recover the result of Finn and Romano: the optimal statistic is not the standard cross-correlation statistic, but instead is a sum of the cross-correlation statistic and extra autocorrelation terms. In Sec. IV B we show that, under the more realistic assumption where the noise variances are taken to be unknowns to be determined from the data, then the standard cross-correlation statistic is recovered. Finally, in Sec. IV C we consider two detectors with white, non-Gaussian noise, and we re-derive the generalized cross-correlation statistic of paper I. Section V contains a short conclusion and summary.

II. FOUNDATIONS

We denote the output of a set of gravitational wave detectors by a vector $\mathbf{x}$, with

$$\mathbf{x} = \mathbf{n} + \mathbf{s}. \tag{2.1}$$

Here $\mathbf{n}$ is the detector noise, and $\mathbf{s}$ is a possibly present gravitational wave signal. We can write

$$\mathbf{s} = \epsilon \hat{\mathbf{s}}, \tag{2.2}$$

where ϵ is a parameter governing the signal strength, and the magnitude of $\hat{\mathbf{s}}$ is fixed. As in paper I, we shall be specializing to weak signals and using an expansion in powers of ϵ about $\epsilon = 0$ throughout this paper.

A. Frequentist signal detection

A key quantity is the probability distribution for $\mathbf{x}$ given ϵ, $p(\mathbf{x}|\epsilon)$. This quantity can be used to compute the performance of detection statistics in the frequentist framework. Suppose one is given a statistic

$$\Gamma = \Gamma(\mathbf{x}), \tag{2.3}$$

say, and that one's detection criterion is that a signal is present if $\Gamma(\mathbf{x}) > \Gamma_*$ and not present otherwise, where Γ_* is a threshold. Then the false alarm probability associated with this statistic is

$$\alpha(\Gamma_*) = \int_{\Gamma > \Gamma_*} d\mathbf{x} p(\mathbf{x}|0), \tag{2.4}$$

and the false dismissal probability is

$$\beta(\Gamma_*, \epsilon) = \int_{\Gamma < \Gamma_*} d\mathbf{x} p(\mathbf{x}|\epsilon). \tag{2.5}$$

By eliminating Γ_* between Eqs. (2.4) and (2.5) we obtain the false-dismissal versus false-alarm curve

$$\beta = \beta(\alpha, \epsilon) \tag{2.6}$$

which characterizes the performance of the statistic.

In paper I we showed that there is a unique statistic $\Lambda_{(1)}(\mathbf{x})$ which minimizes $d\beta/d\epsilon$ at $\epsilon = 0$ for fixed α, defined by the expansion

$$p(\mathbf{x}|\epsilon) = p(\mathbf{x}|0)[1 + \epsilon \Lambda_{(1)}(\mathbf{x}) + \epsilon^2 \Lambda_{(2)}(\mathbf{x}) + O(\epsilon^3)]. \tag{2.7}$$

This statistic therefore has the best false-dismissal versus false-alarm curve for weak signals. [If $\Lambda_{(1)}(\mathbf{x})$ vanishes identically, then $\Lambda_{(2)}(\mathbf{x})$ is the unique statistic that minimizes $d^2\beta/d\epsilon^2$ at $\epsilon = 0$ for fixed α.] We applied this class of detection statistics (called locally optimal statistics [8]) to a variety of different gravitational-wave signal detection problems.

B. Bayesian signal detection

In the Bayesian framework, the probability $P^{(1)}$ that a signal is present in the data is given

$$\frac{P^{(1)}}{1 - P^{(1)}} = \Lambda(\mathbf{x}) \frac{P^{(0)}}{1 - P^{(0)}}, \tag{2.8}$$

where $P^{(0)}$ is the *a priori* probability that a signal is present, and $\Lambda(\mathbf{x})$ is the likelihood ratio. In the literature on Bayesian statistics $\Lambda(\mathbf{x})$ is called a *Bayes factor*. The Bayesian framework uniquely determines a detection criterion, which is to threshold on the probability $P^{(1)}$ that a signal is present. From Eq. (2.8) it is clear that $P^{(1)}$ is a monotonic function of $\Lambda(\mathbf{x})$, so one can equivalently threshold on $\Lambda(\mathbf{x})$. Thus, an optimal detection statistic is uniquely determined in the Bayesian framework; however, that statistic does depend on choices of prior probability distributions.

We now describe how the likelihood ratio is computed. It is given by the formula

$$\Lambda(\mathbf{x}) = \int d\epsilon \Lambda(\mathbf{x}, \epsilon) p^{(0)}(\epsilon) \tag{2.9}$$

where $p^{(0)}(\epsilon)$ is the prior probability distribution for the signal strength ϵ, and

$$\Lambda(\mathbf{x}, \epsilon) = \frac{p(\mathbf{x}|\epsilon)}{p(\mathbf{x}|0)}. \tag{2.10}$$

Suppose that the noise $\mathbf{n}$ is described by a probability distribution $p_\mathbf{n}(\mathbf{n})$, and the signal $\mathbf{s}$ by a signal distribution $p_\mathbf{s}(\mathbf{s}|\epsilon)$. Then it follows from Eq. (2.1) that

$$p(\mathbf{x}|\epsilon) = \int d\mathbf{s} \, p_\mathbf{n}(\mathbf{x}-\mathbf{s}) p_\mathbf{s}(\mathbf{s}|\epsilon). \tag{2.11}$$

The formula (2.10) can therefore be written as

$$\Lambda(\mathbf{x}, \epsilon) = \int d\mathbf{s} \frac{p_\mathbf{n}(\mathbf{x}-\mathbf{s})}{p_\mathbf{n}(\mathbf{n})} p_\mathbf{s}(\mathbf{s}|\epsilon). \tag{2.12}$$

The formula for $\Lambda(\mathbf{x}, \epsilon)$ becomes more complex when there are unknown signal and/or noise parameters present. Suppose the signal distribution depends on some parameters $\boldsymbol{\theta}_s$ in addition to the signal amplitude ϵ, which themselves are distributed according to a prior probability distribution $p_{\theta_s}(\boldsymbol{\theta}_s|\epsilon)$. Then the signal distribution $p_\mathbf{s}(\mathbf{s}|\epsilon)$ can be expanded as

$$p_\mathbf{s}(\mathbf{s}|\epsilon) = \int d\boldsymbol{\theta}_s p_\mathbf{s}(\mathbf{s}|\epsilon, \boldsymbol{\theta}_s) p_{\theta_s}(\boldsymbol{\theta}_s|\epsilon), \tag{2.13}$$

where $p_\mathbf{s}(\mathbf{s}|\epsilon, \boldsymbol{\theta}_s)$ is the distribution for $\mathbf{s}$ given both ϵ and $\boldsymbol{\theta}_s$. Similarly suppose that the noise distribution contains unknown parameters $\boldsymbol{\theta}_n$, whose *a priori* distribution is $p_{\theta_n}(\boldsymbol{\theta}_n)$. Then the noise distribution can be expanded as

$$p_\mathbf{n}(\mathbf{n}) = \int d\boldsymbol{\theta}_n p_\mathbf{n}(\mathbf{n}|\boldsymbol{\theta}_n) p_{\theta_n}(\boldsymbol{\theta}_n). \tag{2.14}$$

Inserting the expansions (2.13) and (2.14) into Eq. (2.12) gives the final expression for the likelihood function:

$$\Lambda(\mathbf{x}, \epsilon) = \int d\mathbf{s} \frac{\displaystyle\int d\boldsymbol{\theta}_n p_\mathbf{n}(\mathbf{x}-\mathbf{s}|\boldsymbol{\theta}_n) p_{\theta_n}(\boldsymbol{\theta}_n)}{\displaystyle\int d\boldsymbol{\theta}'_n p_\mathbf{n}(\mathbf{x}|\boldsymbol{\theta}'_n) p_{\theta_n}(\boldsymbol{\theta}'_n)}$$
$$\times \int d\boldsymbol{\theta}_s p_\mathbf{s}(\mathbf{s}|\epsilon, \boldsymbol{\theta}_s) p_{\theta_s}(\boldsymbol{\theta}_s|\epsilon). \tag{2.15}$$

Equations (2.9) and (2.15) are the foundational equations that we will use throughout this paper to compute the likelihood ratio $\Lambda(\mathbf{x})$. A key feature of this formalism is that the noise parameters $\boldsymbol{\theta}_n$ are treated as unknowns, to be measured from the detector data along with the gravitational wave signal, rather than being treated as known *a priori*. That feature underlies the elimination of the autocorrelation terms encountered in paper I in the case of a stochastic gravitational wave background.

In the next few sections we will revisit several of the signal detection problems considered in paper I. In each case, we will show that $\Lambda(\mathbf{x})$ is, to a good approximation, a monotonic function of the locally optimal detection statistic derived in paper I. Since a monotonic function of a detection statistic has the same false alarm versus false dismissal curve as the original statistic, it follows that in each case the uniquely determined Bayesian statistic $\Lambda(\mathbf{x})$ is equivalent to the statistic computed in paper I. Note, however, that the equivalence only applies at the level of choosing the detection statistic, and not at the level of specifying thresholds. The Bayesian and frequentist approaches lead to different detection thresholds for a given specified significance level; see, for example, the discussion in Sec. III C of Ref. [7].

In deriving the formulas for the likelihood ratio $\Lambda(\mathbf{x})$, we shall invoke a number of different approximations. In assessing the validity of those approximations, we shall be concerned only with their effect on the false alarm versus false dismissal curve of the statistic. In other words, the approximations might be very inaccurate for computing the value of $\Lambda(\mathbf{x})$, but might nevertheless be very accurate in the sense that they have only a small effect on the false alarm versus false dismissal curve. [We do need to compute accurate numerical values of $\Lambda(\mathbf{x})$, since we are not concerned here with computing detection thresholds.] We shall use the notation

$$\Lambda_1(\mathbf{x}) \simeq \Lambda_2(\mathbf{x}) \tag{2.16}$$

to mean that the false alarm versus false dismissal curves of the statistics $\Lambda_1(\mathbf{x})$ and $\Lambda_2(\mathbf{x})$ are approximately the same.

III. DETERMINISTIC SIGNALS

A. Single detector, white Gaussian noise, known variance

We first treat the simple case where we are looking for a signal $\mathbf{s}$, in a single detector, whose values in the time domain are [12]

$$s_j = \epsilon \hat{s}_j. \tag{3.1}$$

We assume that the quantities $\hat{s}_j$ are known and fixed, so that the only unknown parameter characterizing the signal is its amplitude ϵ, which can be positive or negative. Without loss of generality we can choose the normalization so that

$$\sum_j |\hat{s}_j|^2 = 1. \tag{3.2}$$

We assume that the detector noise is white and Gaussian with zero mean and unit variance. Then, as shown in paper I, the distribution for the data $\mathbf{x}$ given ϵ is

$$p(\mathbf{x}|\epsilon) = \prod_j \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(x_j - \epsilon\hat{s}_j)^2\right]. \tag{3.3}$$

Inserting this formula into Eq. (2.10) gives

$$\Lambda(\mathbf{x}, \epsilon) = \exp[\epsilon\hat{\epsilon}(\mathbf{x}) - \epsilon^2/2], \tag{3.4}$$

where

$$\hat{\epsilon}(\mathbf{x}) = \sum_j x_j \hat{s}_j \tag{3.5}$$

is the standard matched filtering statistic. Combining this with Eq. (2.9) gives for the likelihood ratio

$$\Lambda(\mathbf{x}) = e^{\hat{\epsilon}(\mathbf{x})^2/2} \int d\epsilon \, p^{(0)}(\epsilon) e^{-[\epsilon - \hat{\epsilon}(\mathbf{x})]^2/2}. \qquad (3.6)$$

Now the quantity $|\hat{\epsilon}(\mathbf{x})|$ is effectively the signal-to-noise ratio. Let us assume that we are in the relevant regime where the signal is detectable with high confidence, so that

$$\exp[\hat{\epsilon}(\mathbf{x})^2/2] \gg 1. \qquad (3.7)$$

Let us also assume that the prior distribution $p^{(0)}(\epsilon)$ is slowly varying and does not strongly constrain the possible values of ϵ. Then, we can approximately evaluate the integral (3.6) using the Laplace approximation to obtain

$$\Lambda(\mathbf{x}) \approx \sqrt{2\pi} p^{(0)}[\hat{\epsilon}(\mathbf{x})] \exp[\hat{\epsilon}(\mathbf{x})^2/2]. \qquad (3.8)$$

Finally, we argue that we can neglect the dependence on $\mathbf{x}$ of the factor $p^{(0)}[\hat{\epsilon}(\mathbf{x})]$ in the expression (3.8). The reason is that the prior distribution $p^{(0)}(\epsilon)$ is a slowly varying function of ϵ, and so this factor has a much weaker dependence on $\mathbf{x}$ than the exponential factor in the regime (3.7). Therefore, dropping the factor $p^{(0)}[\hat{\epsilon}(\mathbf{x})]$ will have a negligible effect on the false alarm versus false dismissal curve of the statistic. In this approximation we see that $\Lambda(\mathbf{x})$ is a monotonic function of the standard detection statistic $|\hat{\epsilon}(\mathbf{x})|$,

$$\Lambda(\mathbf{x}) \simeq \exp[\hat{\epsilon}(\mathbf{x})^2/2], \qquad (3.9)$$

as claimed [13].

We remark that there is a key technical difference between the above computation and the corresponding computation in Sec. II A of paper I. The Bayesian computation presented here requires expanding the quantity $\ln \Lambda(\mathbf{x}, \epsilon)$ to second order in ϵ about $\epsilon = 0$ [Eq. (3.4) above], whereas in paper I it sufficed to compute $\ln \Lambda(\mathbf{x}, \epsilon)$ to linear order in ϵ [Eqs. (2.3) and (2.5) of paper I]. This difference is a common feature of all of our subsequent computations.

B. Single detector, white Gaussian noise, unknown variance

We now add one additional complication to the analysis, by taking the noise variance to be an unknown constant σ. We define an inner product $\langle,\rangle$ on the space of signals by

$$\langle \mathbf{x}, \mathbf{y} \rangle \equiv \sum_j x_j y_j. \qquad (3.10)$$

Note that this is *not* the standard inner product used in discussions of matched filtering, which incorporates a weighting factor of σ^{-2}. We define the statistic $\hat{\sigma}(\mathbf{x})$ by

$$\hat{\sigma}(\mathbf{x})^2 \equiv \frac{1}{N} \langle \mathbf{x}, \mathbf{x} \rangle, \qquad (3.11)$$

where N is the number of data points; this is the conventional estimator of σ. We also define the quantity ρ by

$$\rho = \frac{\epsilon}{\sigma}, \qquad (3.12)$$

which from the normalization condition (3.2) is the conventional signal to noise ratio. The corresponding estimator is

$$\hat{\rho}(\mathbf{x}) = \frac{\epsilon}{\hat{\sigma}(\mathbf{x})}. \qquad (3.13)$$

The conventional matched filtering statistic is

$$\frac{1}{\sigma} \langle \mathbf{x}, \hat{s} \rangle, \qquad (3.14)$$

and if we replace the noise variance by its estimator $\hat{\sigma}(\mathbf{x})$ we obtain the statistic

$$\hat{\rho}_1(\mathbf{x}) \equiv \frac{1}{\hat{\sigma}(\mathbf{x})} \langle \mathbf{x}, \hat{s} \rangle. \qquad (3.15)$$

We shall show below that the likelihood ratio $\Lambda(\mathbf{x})$ is to a good approximation equivalent to the conventional statistic (3.15).

The noise distribution given σ is taken to be

$$p_{\mathbf{n}}(\mathbf{n}|\sigma) = \prod_j \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{n_j^2}{2\sigma^2}\right]. \qquad (3.16)$$

The full noise distribution is [cf. Eq. (2.14) above]

$$p_{\mathbf{n}}(\mathbf{n}) = \int_0^\infty d\sigma \, p_\sigma(\sigma) p_{\mathbf{n}}(\mathbf{n}|\sigma), \qquad (3.17)$$

where $p_\sigma(\sigma)$ is the prior probability distribution for σ. Inserting Eq. (3.17) into Eq. (3.16) and using the definition (3.11) we obtain

$$p_{\mathbf{n}}(\mathbf{n}) = \int_0^\infty d\sigma \, p_\sigma(\sigma) \exp[-N\Xi(\sigma)/2], \qquad (3.18)$$

where

$$\Xi(\sigma) = \ln(2\pi\sigma^2) + \frac{\hat{\sigma}(\mathbf{n})^2}{\sigma^2}. \qquad (3.19)$$

We can approximately evaluate the integral (3.18) in the limit where N is large. The function $\Xi(\sigma)$ can be expanded about its local minimum at $\sigma = \hat{\sigma}$ as

$$\Xi(\sigma) = 1 + \ln(2\pi\hat{\sigma}^2) + \frac{2}{\hat{\sigma}^2}(\sigma - \hat{\sigma})^2 + O[(\sigma - \hat{\sigma})^3]. \qquad (3.20)$$

Using this expansion we obtain

$$p_{\mathbf{n}}(\mathbf{n}) = \sqrt{\frac{\pi}{N}}(2\pi e)^{-N/2} p_\sigma[\hat{\sigma}(\mathbf{n})]\hat{\sigma}(\mathbf{n})^{-(N-1)}$$

$$\times \left[1 + O\left(\frac{1}{\sqrt{N}}\right)\right]. \tag{3.21}$$

We assume that $p_\sigma(\sigma)$ is slowly varying, and so have neglected in Eq. (3.21) a fractional error of order the fractional change in $p_{\mathbf{n}}$ over an interval of width $\hat{\sigma}/\sqrt{N}$.

We next insert the formula (3.21) for the noise distribution into the expression (2.12) for the likelihood ratio, using

$$p_{\mathbf{s}}(\mathbf{s}|\epsilon) = \delta^N(\mathbf{s} - \epsilon\hat{\mathbf{s}}). \tag{3.22}$$

The result is

$$\Lambda(\mathbf{x}) = \int_0^\infty d\epsilon\, p^{(0)}(\epsilon)\frac{p_\sigma[\hat{\sigma}(\mathbf{x} - \epsilon\hat{\mathbf{s}})]}{p_\sigma[\hat{\sigma}(\mathbf{x})]}\left[\frac{\hat{\sigma}(\mathbf{x} - \epsilon\hat{\mathbf{s}})}{\hat{\sigma}(\mathbf{x})}\right]^{-(N-1)}$$

$$\tag{3.23}$$

$$= \int_0^\infty d\epsilon\, p^{(0)}(\epsilon)\frac{p_\sigma[\hat{\sigma}(\mathbf{x} - \epsilon\hat{\mathbf{s}})]}{p_\sigma[\hat{\sigma}(\mathbf{x})]}$$

$$\times \exp\left[-\frac{N-1}{2}\ln\left\{1 - 2\epsilon\frac{\langle\mathbf{x},\hat{\mathbf{s}}\rangle}{\langle\mathbf{x},\mathbf{x}\rangle} + \frac{\epsilon^2}{\langle\mathbf{x},\mathbf{x}\rangle^2}\right\}\right]. \tag{3.24}$$

To obtain the second line we used Eqs. (3.2) and (3.11). Expanding the logarithm to second order in ϵ, we can re-express this as

$$\Lambda(\mathbf{x}) = \int_0^\infty d\epsilon\, p^{(0)}(\epsilon)\frac{p_\sigma[\hat{\sigma}(\mathbf{x} - \epsilon\hat{\mathbf{s}})]}{p_\sigma[\hat{\sigma}(\mathbf{x})]}\exp[\hat{a}\epsilon - \hat{b}\epsilon^2 + O(\epsilon^3)],$$

$$\tag{3.25}$$

where

$$\hat{a} = (N-1)\frac{\langle\mathbf{x},\hat{\mathbf{s}}\rangle}{\langle\mathbf{x},\hat{\mathbf{x}}\rangle} \tag{3.26}$$

and

$$\hat{b} = \frac{1}{2}(N-1)\left[\frac{1}{\langle\mathbf{x},\hat{\mathbf{x}}\rangle} - 2\frac{\langle\mathbf{x},\hat{\mathbf{s}}\rangle^2}{\langle\mathbf{x},\hat{\mathbf{x}}\rangle^2}\right]. \tag{3.27}$$

Before proceeding further with the computation of the likelihood ratio, we clarify the domain of validity of the weak signal expansion (expansion in powers of ϵ) used in going from Eq. (3.24) to Eq. (3.25). We will estimate the expected sizes and the scale of statistical fluctuations in the two terms appearing in the argument of the logarithm in Eq. (3.24); the expansion will be good when both of these terms and their fluctuations are small compared to unity. For the purpose of making these estimates we can identify σ and $\hat{\sigma}$, and ρ, $\hat{\rho}$, and $\hat{\rho}_1$.

We can compute the expected value and variance of the statistic $\langle\mathbf{x},\mathbf{x}\rangle$ using Eqs. (2.1), (2.2) and (3.16), which gives

$$\langle\mathbf{x},\mathbf{x}\rangle \sim (N\sigma^2 + \epsilon^2) \pm \sqrt{2N\sigma^4 + 4\epsilon^2\sigma^2}. \tag{3.28}$$

The notation here is that the first term gives the expected value, and the second term gives an estimate of the statistical fluctuations. We can rewrite this formula in terms of the signal-to-noise ratio $\rho = \epsilon/\sigma$ as

$$\langle\mathbf{x},\mathbf{x}\rangle \sim N\sigma^2\left[\left(1 + \frac{\rho^2}{N}\right) \pm \sqrt{\frac{2}{N} + 4\frac{\rho^2}{N^2}}\right]. \tag{3.29}$$

Similarly we have

$$\langle\mathbf{x},\hat{\mathbf{s}}\rangle \sim \epsilon \pm \sigma. \tag{3.30}$$

We assume that $N \gg 1$ and that $\rho \gtrsim 1$. We now consider two different cases:

(i) When $\rho^2/N \ll 1$, the fluctuations in $\langle\mathbf{x},\mathbf{x}\rangle$ are small compared to the expected value, and we have from Eq. (3.29) that $\langle\mathbf{x},\mathbf{x}\rangle \sim N\sigma^2 \pm \sqrt{N}\sigma^2$. Using this together with Eq. (3.30) shows that the first term in the argument of the logarithm in Eq. (3.24) is

$$\epsilon\frac{\langle\mathbf{x},\hat{\mathbf{s}}\rangle}{\langle\mathbf{x},\mathbf{x}\rangle} \sim \frac{\rho^2}{N} \pm \frac{\rho}{N} \ll 1, \tag{3.31}$$

and similarly the second term is

$$\frac{\epsilon^2}{\langle\mathbf{x},\mathbf{x}\rangle} \sim \frac{\rho^2}{N} \pm \frac{\rho^2}{N^{3/2}} \ll 1. \tag{3.32}$$

Thus the approximation is good in this regime.

(ii) When $\rho^2/N \gg 1$, a similar computation gives that $\langle\mathbf{x},\mathbf{x}\rangle \sim \rho^2\sigma^2 \pm \rho\sigma^2$. The first term in the argument of the logarithm in Eq. (3.24) now scales as

$$\epsilon\frac{\langle\mathbf{x},\hat{\mathbf{s}}\rangle}{\langle\mathbf{x},\mathbf{x}\rangle} \sim 1 \pm \frac{1}{\rho} \tag{3.33}$$

and similarly the second term scales as

$$\frac{\epsilon^2}{\langle\mathbf{x},\mathbf{x}\rangle} \sim 1 \pm \frac{1}{\rho}. \tag{3.34}$$

Thus, the approximation breaks down in this regime.

We now return to computing the likelihood ratio $\Lambda(\mathbf{x})$. We can approximately evaluate the integral (3.25) using the Laplace approximation to obtain

$$\Lambda(\mathbf{x}) \approx p^{(0)}(\hat{\epsilon})\frac{p_\sigma[\hat{\sigma}(\mathbf{x} - \hat{\epsilon}\hat{\mathbf{s}})]}{p_\sigma[\hat{\sigma}(\mathbf{x})]}\sqrt{\frac{\hat{a}\pi}{\hat{b}}}\exp\left[\frac{\hat{a}^2}{4\hat{b}}\right], \tag{3.35}$$

where

$$\hat{\epsilon}(\mathbf{x}) = \frac{\hat{a}(\mathbf{x})}{2\hat{b}(\mathbf{x})}. \tag{3.36}$$

This approximation will be good whenever the exponential factor in Eq. (3.35) is large, which it will be in the regime where the signal is detectable (see below), and when the

prior probability distribution $p^{(0)}(\epsilon)$ is slowly varying. Using Eqs. (3.11), (3.15), (3.26) and (3.27), we can write the exponential factor as

$$\exp\left[\frac{N-1}{2}g(\hat{\rho}_1^2/N)\right], \qquad (3.37)$$

where the function g is given by $g(x)=x/(1-2x)$. Since we are in the regime $\rho^2/N\ll 1$, the argument of the function g is small, and we can replace $g(\hat{\rho}_1^2/N)$ by $\hat{\rho}_1^2/N$. This gives, using $N\gg 1$,

$$\Lambda(\mathbf{x})\approx p^{(0)}(\hat{\epsilon})\frac{p_\sigma[\hat{\sigma}(\mathbf{x}-\hat{\epsilon}\hat{\mathbf{s}})]}{p_\sigma[\hat{\sigma}(\mathbf{x})]}\sqrt{\frac{\hat{a}\pi}{\hat{b}}}\exp\left[\frac{1}{2}\hat{\rho}_1(\mathbf{x})^2\right]. \qquad (3.38)$$

Finally, we argue as before that in the regime $\exp[\rho^2/2]\gg 1$ where the signal is detectable, the dependence on $\mathbf{x}$ of all the other factors in Eq. (3.38) can be neglected in comparison to the exponential factor, assuming that the prior distributions are slowly varying. This gives

$$\Lambda(\mathbf{x})\simeq\exp[\hat{\rho}_1(\mathbf{x})^2/2], \qquad (3.39)$$

as claimed.

Our final answer (3.39) is essentially the same as the answer (3.9) obtained when the noise variance σ is assumed to be known. Therefore, treating σ as an unknown parameter rather than as a fixed, known parameter does not make much difference in this case. However, we will see below for the case of stochastic signals that treating the properties of the noise distribution as unknowns does have a significant effect on the analysis, and that the correct answer is obtained only when those properties are treated as unknowns.

We end this subsection by recapitulating the various approximations and assumptions we have invoked:

(i) The large N approximation $N\gg 1$.

(ii) The assumption that we are in the regime where the signal is detectable, $\exp(\rho^2/2)\gg 1$. This is necessary for evaluating the integral over ϵ to obtain Eq. (3.35), and also for the validity in neglecting the prefactors in deriving Eq. (3.39). From a practical point of view the assumption $\exp(\rho^2/2)\gg 1$ is not a serious restriction, as it does not matter how our statistics perform in the regime $\exp(\rho^2/2)\sim 1$ where signals are not detectable.

(iii) The assumption that the prior probability distributions $p^{(0)}(\epsilon)$ and $p_\sigma(\sigma)$ are slowly varying.

(iv) We have clarified the "weak signal" assumption of paper I; it is the assumption that the signal-to-noise ratio squared per data point is small, $\rho^2/N\ll 1$. This requirement ensures that the presence of the signal does not significantly bias the estimate (3.11) of the noise variance. In practice we can always choose segments of data large enough to satisfy this assumption.

C. Single detector, white non-Gaussian noise

We now turn to the case where the noise has a known, non-Gaussian distribution. In this subsection we follow

Sec. II A of paper I, and assume that the noise samples in the time domain are statistically independent but identically distributed, with a known distribution. We can write the noise probability distribution as

$$p_\mathbf{n}(\mathbf{n})=\prod_j e^{-f(n_j)}. \qquad (3.40)$$

We assume that the probability distribution $e^{-f(x)}$ has a central Gaussian region $|x|<x_b$ in which

$$f(x)=\frac{x^2}{2\sigma^2}, \qquad (3.41)$$

which contains most of the probability, and a tail region $|x|\geq x_b$ containing a total probability p_{tail} with $p_{\text{tail}}\ll 1$.

As before, the signal is assumed to be known up to an overall amplitude parameter. From Eqs. (2.11) and (3.22) we obtain the following modified version of Eq. (3.3):

$$p(\mathbf{x}|\epsilon)=\prod_j \exp[-f(x_j-\epsilon\hat{s}_j)], \qquad (3.42)$$

and inserting this into Eqs. (2.9) and (2.10) gives

$$\Lambda(\mathbf{x})=\int d\epsilon p^{(0)}(\epsilon)\prod_j \exp[-f(x_j-\epsilon\hat{s}_j)+f(x_j)]. \qquad (3.43)$$

Expanding to second order in ϵ gives

$$\Lambda(\mathbf{x})=\int d\epsilon p^{(0)}(\epsilon)\exp[\hat{a}(\mathbf{x})\epsilon-\hat{b}(\mathbf{x})\epsilon^2+O(\epsilon^3)], \qquad (3.44)$$

where

$$\hat{a}(\mathbf{x})=\sum_j f'(x_j)\hat{s}_j \qquad (3.45)$$

and

$$\hat{b}(\mathbf{x})=\frac{1}{2}\sum_j f''(x_j)\hat{s}_j^2. \qquad (3.46)$$

Evaluating the integral over ϵ using the same types of arguments as in Sec. III B gives

$$\Lambda(\mathbf{x})\simeq\exp\left[\frac{\hat{a}(\mathbf{x})^2}{2\hat{b}(\mathbf{x})}\right]. \qquad (3.47)$$

Now the statistic $\hat{a}(\mathbf{x})$ is the locally optimal statistic computed in paper I [Eq. (2.9) of paper I]. Therefore it remains to show that we can neglect the $\mathbf{x}$ dependence of the factor $\hat{b}(\mathbf{x})$ in the argument of the exponential in Eq. (3.47).

We can split the sum (3.46) into contributions from the Gaussian region and from the tail. Using the fact that $f''(x)=1/\sigma^2$ in the Gaussian region, we obtain

$$\hat{b}(\mathbf{x}) = \frac{1}{2} \sum_{x_j < x_b} \frac{\hat{s}_j^2}{\sigma^2} + \frac{1}{2} \sum_{x_j \geqslant x_b} f''(x_j) \hat{s}_j^2. \qquad (3.48)$$

Most of the values of x_j will fall in the central Gaussian region, since $x_j = n_j + \epsilon \hat{s}_j$, and $\epsilon \hat{s}_j \ll \sigma$ for each individual j [14]. Since $\Sigma_j \hat{s}_j^2 = 1$, the first term gives $1/(2\sigma^2)$, up to fractional corrections of order p_{tail}. Similarly the second term will be bounded above by $\sim p_{\text{tail}}/(2\sigma^2)$, since $f''(x)$ will be smaller in the tails than in the central Gaussian region. We conclude that

$$\hat{b}(\mathbf{x}) = \frac{1}{2\sigma^2}[1 + O(p_{\text{tail}})]. \qquad (3.49)$$

It follows in particular that the **x**-dependent fluctuations in $\hat{b}(\mathbf{x})$ are smaller than its expected value by a factor of $p_{\text{tail}} \ll 1$, and therefore we can neglect the **x** dependence of $\hat{b}(\mathbf{x})$ in Eq. (3.47), as required.

D. Single detector, colored non-Gaussian noise

We next consider the model of colored, non-Gaussian noise of Sec. II B of paper I, where each frequency bin is assumed to be statistically independent. This is given by

$$p_{\mathbf{n}}(\mathbf{n}) = \prod_{k=1}^{[(N-1)/2]} \frac{2}{\pi P_k} \exp\left[-2g_k\left(\frac{|\tilde{n}_k|^2}{P_k}\right)\right], \qquad (3.50)$$

where the volume element is understood to be

$$\prod_{k=1}^{[(N-1)/2]} d(\operatorname{Re}\tilde{n}_k) d(\operatorname{Im}\tilde{n}_k). \qquad (3.51)$$

Here

$$\tilde{n}_k = \frac{1}{\sqrt{N}} \sum_j e^{2\pi i j k/N} n_j \qquad (3.52)$$

are the components of the discrete Fourier transform of the time domain samples n_j. The quantities P_k describe the noise spectrum. For each frequency bin k, the function $g_k(x)$ is arbitrary except for the normalization conditions

$$\int_0^\infty dx e^{-g_k(x)} = \int_0^\infty dx x e^{-g_k(x)} = 1, \qquad (3.53)$$

and the requirement that $g_k(x) = x$ in a central Gaussian region containing most of the probability.

By paralleling the analysis of Sec. III C, we again arrive at the formulas (3.44) and (3.47), where now the statistics $\hat{a}(\mathbf{x})$ and $\hat{b}(\mathbf{x})$ are given by

$$\hat{a}(\mathbf{x}) = 4 \sum_k g_k'\left(\frac{|\tilde{x}_k|^2}{P_k}\right) \frac{\operatorname{Re}(\tilde{x}_k^* \tilde{s}_k)}{P_k} \qquad (3.54)$$

and

$$\hat{b}(\mathbf{x}) = 2 \sum_k g_k'\left(\frac{|\tilde{x}_k|^2}{P_k}\right) \frac{|\tilde{s}_k|^2}{P_k} + 4 \sum_k g_k''\left(\frac{|\tilde{x}_k|^2}{P_k}\right) \frac{[\operatorname{Re}(\tilde{x}_k^* \tilde{s}_k)]^2}{P_k^2}. \qquad (3.55)$$

As before, the statistic $\hat{a}(\mathbf{x})$ coincides with the locally optimal statistic derived in paper I [Eq. (2.21) of paper I], and it suffices to show that the **x** dependence of the factor $\hat{b}(\mathbf{x})$ can be neglected in Eq. (3.47). We evaluate the sums in Eq. (3.55) by splitting them into Gaussian and tail contributions as before. Since $g_k'(x) = 1$ in the Gaussian region, the first term in Eq. (3.55) yields $2\Sigma_k |\tilde{s}_k|^2/P_k[1 + O(p_{\text{tail}})]$. Also the second term is proportional to p_{tail} since $g_k''(x)$ vanishes in the Gaussian region. Thus we obtain

$$\hat{b}(\mathbf{x}) = \left[2 \sum_k \frac{|\tilde{s}_k|^2}{P_k}\right][1 + O(p_{\text{tail}})], \qquad (3.56)$$

and the rest of the argument follows as before.

E. Signals with unknown parameters

We now generalize the analysis of the preceding subsections by allowing the signals to depend on additional parameters other than the overall amplitude parameter ϵ. We write

$$\mathbf{s} = \epsilon \hat{\mathbf{s}}(\boldsymbol{\theta}_s), \qquad (3.57)$$

where the signal parameters $\boldsymbol{\theta}_s$ are distributed according to the distribution $p_{\theta_s}(\boldsymbol{\theta}_s|\epsilon)$. Then from Eq. (2.15) we can write

$$\Lambda(\mathbf{x}) = \int d\epsilon \int d\boldsymbol{\theta}_s p^{(0)}(\epsilon) p_{\theta_s}(\boldsymbol{\theta}_s|\epsilon) \Lambda(\mathbf{x}, \epsilon, \boldsymbol{\theta}_s), \qquad (3.58)$$

where $\Lambda(\mathbf{x}, \epsilon, \boldsymbol{\theta}_s)$ is given by Eq. (2.11) with $p_{\mathbf{s}}(\mathbf{s}|\epsilon)$ replaced by $p_{\mathbf{s}}(\mathbf{s}|\epsilon, \boldsymbol{\theta}_s)$. In the regime where the signal is detectable, we can repeat the arguments of the preceding subsections to approximately evaluate the integrals as

$$\Lambda(\mathbf{x}) \simeq \max_{\boldsymbol{\theta}_s} \max_\epsilon \Lambda(\mathbf{x}, \epsilon, \boldsymbol{\theta}_s). \qquad (3.59)$$

Thus the result is to take the statistics previously derived and to maximize over the signal parameters. Such a maximization is the standard thing to do for linear matched filtering; the above argument indicates that it is also the appropriate procedure for the more general class of locally optimal statistics.

IV. STOCHASTIC SIGNALS

The standard method of detecting a stochastic background is to compute a cross-correlation between two different instruments [9]; see Ref. [10] for a detailed description. In Sec. IV A below we compute the likelihood ratio $\Lambda(\mathbf{x})$ for the simplest case of a white stochastic signal, and of two coaligned detectors with white Gaussian noise, where the noise variance is assumed to be known. For this case we do not recover the standard cross-correlation statistic, but instead

we obtain a statistic with extra autocorrelation terms. That statistic was first derived and has been investigated in detail in a more general context by Finn and Romano [6]. We then argue that it is unrealistic to take the noise variances to be known parameters. In Sec. IV B we show that, when the noise variances are taken to be unknowns to be determined from the data, then the likelihood ratio $\Lambda(\mathbf{x})$ is to a good approximation equivalent to the standard cross-correlation statistic. This computation is again in the simple context of coincident aligned detectors with white noise. The computation of Sec. IV B is a simplified version of the computation in Appendix A of Ref. [11].

We then turn to non-Gaussian noise models. In paper I, we derived a generalized cross-correlation (GCC) statistic appropriate for non-Gaussian noise, which is a modification of the standard cross-correlation statistic. In Sec. IV C below we re-derive that statistic using the Bayesian approach.

A. Two coincident co-aligned detectors, white Gaussian noise, known variances

The output of the pair of detectors is $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$, where

$$\mathbf{x}_1 = \mathbf{n}_1 + \mathbf{s} \tag{4.1}$$

is the output of the first detector, and

$$\mathbf{x}_2 = \mathbf{n}_2 + \mathbf{s} \tag{4.2}$$

is the output of the second. We assume, for simplicity, that the noise in each detector is white and Gaussian with unit variance:

$$p_{\mathbf{n}_1}(\mathbf{n}_1) = \prod_j \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{n_{1j}^2}{2}\right], \tag{4.3}$$

with a similar equation for the second detector. We assume that the stochastic background signal is also white and Gaussian with variance ϵ:

$$p_{\mathbf{s}}(\mathbf{s}|\epsilon) = \prod_j \frac{1}{\sqrt{2\pi\epsilon}} \exp\left[-\frac{s_j^2}{2\epsilon}\right], \tag{4.4}$$

where $\epsilon \geq 0$. As in earlier sections, ϵ parametrizes the signal strength, and we will be using a weak signal expansion of expanding in powers of ϵ about $\epsilon = 0$.

By inserting the distributions (4.3) and (4.4) into the formulas (2.11) and (2.14) we obtain

$$p(\mathbf{x}|\epsilon) = \exp\left[-\frac{N}{2}\Xi(\epsilon, \mathbf{x})\right], \tag{4.5}$$

where

$$\Xi(\epsilon, \mathbf{x}) = 2\ln(2\pi) + \ln[1 + 2\epsilon]$$
$$+ \frac{(1+\epsilon)\hat{\sigma}_2^2 + (1+\epsilon)\hat{\sigma}_1^2 - 2\epsilon\hat{\epsilon}}{1+2\epsilon}, \tag{4.6}$$

and where

$$\hat{\sigma}_1^2 = \frac{1}{N}\sum_j x_{1j}^2, \tag{4.7}$$

$$\hat{\sigma}_2^2 = \frac{1}{N}\sum_j x_{2j}^2, \tag{4.8}$$

and

$$\hat{\epsilon} = \frac{1}{N}\sum_j x_{1j}x_{2j}. \tag{4.9}$$

The statistic $\hat{\epsilon}$ is the standard cross-correlation statistic.

Before proceeding further we discuss the validity of the weak signal approximation in the context of a stochastic background signal. Suppose that a stochastic background is present and just barely detectable by cross-correlating between the two detectors. Then we have $\hat{\epsilon} \sim 1/\sqrt{N}$. In the context of ground based detectors such as the Laser Interferometric Gravitational Wave Observatory (LIGO), when this analysis is generalized to colored noise, N is replaced by the product $T\Delta f$, where T is the observation time and Δf is the effective bandwidth in the usual formula for the signal-to-noise ratio [Eq. (1.2) of Ref. [11]]. Using the estimates $T \sim 1/3$ year and $\Delta f \sim 50$ Hz we find

$$\hat{\epsilon} \sim \frac{1}{\sqrt{T\Delta f}} \sim 10^{-4}. \tag{4.10}$$

Therefore in all our analyses it will be sufficient to work to first order in $\hat{\epsilon}$. The approximation would only break down if $\hat{\epsilon} \sim 1$, that is, if the stochastic background could be seen in a single detector, which is thought to be very unlikely.

We define the statistics $\hat{d}_1$ and $\hat{d}_2$ by

$$\hat{d}_1(\mathbf{x}) = \hat{\sigma}_1(\mathbf{x})^2 - 1 - \hat{\epsilon}(\mathbf{x}) \tag{4.11}$$

and

$$\hat{d}_2(\mathbf{x}) = \hat{\sigma}_2(\mathbf{x})^2 - 1 - \hat{\epsilon}(\mathbf{x}). \tag{4.12}$$

Now $\langle \hat{\sigma}_1^2 \rangle = 1 + \epsilon$, so the quantities $\hat{d}_1$ and $\hat{d}_2$ will be small;

$$|\hat{d}_{1,2}(\mathbf{x})| \lesssim O(1/\sqrt{N}) + O(\hat{\epsilon}). \tag{4.13}$$

We now insert Eqs. (4.11) and (4.12) into the formula (4.6) for the function Ξ. We expand to second order in ϵ, $\hat{\epsilon}$, $\hat{d}_1$, and $\hat{d}_2$, treating these quantities as formally all of the same order. We then insert the result into Eqs. (2.9), (2.10), and (4.5), which yields

$$\Lambda(\mathbf{x}) = \int d\epsilon\, p^{(0)}(\epsilon) \exp[-N(\epsilon - \hat{\epsilon}_b)^2 + N\hat{\epsilon}_b^2], \tag{4.14}$$

where

$$\hat{\epsilon}_b(\mathbf{x}) = \hat{\epsilon}(\mathbf{x}) + \frac{1}{4}[\hat{d}_1(\mathbf{x}) + \hat{d}_2(\mathbf{x})]$$

$$= \frac{1}{2}\hat{\epsilon}(\mathbf{x}) + \frac{1}{4}[\hat{\sigma}_1(\mathbf{x})^2 - 1]$$

$$+ \frac{1}{4}[\hat{\sigma}_2(\mathbf{x})^2 - 1]. \tag{4.15}$$

Evaluating the integral over ϵ using the same types of arguments as in previous sections gives, using $p^{(0)}(\epsilon) = 0$ for $\epsilon < 0$,

$$\Lambda(\mathbf{x}) \approx \Theta(\hat{\epsilon}_b)\sqrt{\frac{\pi}{N}}p^{(0)}(\hat{\epsilon}_b)\exp[N\hat{\epsilon}_b^2]. \tag{4.16}$$

Here Θ is the step function. If follows by the same type of arguments as before that $\Lambda(\mathbf{x}) \approx \Theta(\hat{\epsilon}_b)\exp[N\hat{\epsilon}_b^2]$.

Our final result identifies the statistic $\hat{\epsilon}_b$ as the optimal detection statistic; see Ref. [6] for a more general version of this statistic. From Eq. (4.15) this statistic is not the standard cross-correlation statistic $\hat{\epsilon}$, but instead contains the autocorrelation terms $\hat{\sigma}_1^2 - 1$ and $\hat{\sigma}_2^2 - 1$. The interpretation of these terms is that it is possible, under the assumptions of this subsection, to measure the stochastic background signal with just one detector. If the detector's noise variance is known, then one can just measure the variance of the detector's output and subtract the known noise variance to reveal the stochastic background contribution.

Of course, in reality, the noise in detectors is not known *a priori*, and is measured from the data. In particular, there is no way that the detectors noise can be known beforehand to a fractional accuracy of 10^{-4}. Therefore we have to generalize the preceding analysis by allowing the noise variances to be unknown parameters.

B. Two coincident co-aligned detectors, white Gaussian noise, unknown variances

We assume that the noise in each detector is white and Gaussian with variances σ_1 and σ_2. We replace Eq. (4.3) with

$$p_{\mathbf{n}_1}(\mathbf{n}_1|\sigma_1) = \prod_j \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left[-\frac{n_{1j}^2}{2\sigma_1^2}\right], \tag{4.17}$$

with a similar equation for the second detector. The prior distribution for the parameters σ_1, σ_2 will be written as $p_\sigma(\sigma_1, \sigma_2)$. By inserting the distributions (4.17) and (4.4) into the formulas (2.11) and (2.14) we obtain

$$p(\mathbf{x}|\epsilon) = \int_0^\infty d\sigma_1 \int_0^\infty d\sigma_2 p_\sigma(\sigma_1, \sigma_2)\exp\left[-\frac{N}{2}\Xi(\epsilon, \sigma_1, \sigma_2)\right], \tag{4.18}$$

where the function Ξ is now given by

$$\Xi(\epsilon, \sigma_1, \sigma_2) = 2\ln(2\pi) + \ln[\sigma_1^2\sigma_2^2 + \epsilon(\sigma_1^2 + \sigma_2^2)]$$

$$+ \frac{(\sigma_1^2 + \epsilon)\hat{\sigma}_2^2 + (\sigma_2^2 + \epsilon)\hat{\sigma}_1^2 - 2\epsilon\hat{\epsilon}}{\sigma_1^2\sigma_2^2 + \epsilon(\sigma_1^2 + \sigma_2^2)}. \tag{4.19}$$

To evaluate the integral over σ_1 and σ_2 in Eq. (4.18), we make a change of variables to variables f_1, f_2 defined by

$$\sigma_1^2 + \epsilon = f_1^2\hat{\sigma}_1^2 \tag{4.20}$$

and

$$\sigma_2^2 + \epsilon = f_2^2\hat{\sigma}_2^2. \tag{4.21}$$

We also define the rescaled variables

$$\alpha = \frac{\epsilon}{\hat{\sigma}_1\hat{\sigma}_2}, \quad \hat{\alpha} = \frac{\hat{\epsilon}}{\hat{\sigma}_1\hat{\sigma}_2}. \tag{4.22}$$

We expand Ξ to second order around its local minimum at $f_1 = f_2 = 1$, $\alpha = \hat{\alpha}$:

$$\Xi = 2\ln(2\pi\hat{\sigma}_1\hat{\sigma}_2) + 2 + \ln(1 - \hat{\alpha}^2) + \frac{1 + \hat{\alpha}^2}{(1 - \hat{\alpha}^2)^2}\Delta\alpha^2$$

$$- \frac{4\hat{\alpha}}{(1 - \hat{\alpha}^2)^2}\Delta\alpha(\Delta f_1 + \Delta f_2) + \frac{2}{(1 - \hat{\alpha}^2)^2}(\Delta f_1^2 + \Delta f_2^2$$

$$+ 2\hat{\alpha}^2\Delta f_1^2\Delta f_2^2). \tag{4.23}$$

Here $\Delta f_1 = f_1 - 1$, $\Delta f_2 = f_2 - 1$, and $\Delta\alpha = \alpha - \hat{\alpha}$. At fixed α, Ξ is minimized at

$$f_1 = f_2 = 1 + \frac{\hat{\alpha}}{(1 + \hat{\alpha}^2)}(\alpha - \hat{\alpha}). \tag{4.24}$$

We now perform the Gaussian integral over σ_1, σ_2 or f_1, f_2, which gives

$$p(\mathbf{x}|\epsilon) = \frac{p_\sigma[\hat{\sigma}_{1b}(\epsilon), \hat{\sigma}_{2b}(\epsilon)]}{(2\pi\hat{\sigma}_1\hat{\sigma}_2)^{N-1}}\frac{(1 - \hat{\alpha}^2)^{3/2}}{4(1 + \hat{\alpha}^2)^{1/2}}\mathcal{J}(\epsilon)$$

$$\times \exp\left[-\frac{N}{2(1 + \hat{\alpha}^2)}(\alpha - \hat{\alpha})^2\right]. \tag{4.25}$$

Here $\hat{\sigma}_{1b}(\epsilon)$ is the value of σ_1 at the peak (4.24) of the integrand, given from Eqs. (4.20) and (4.24) by

$$\hat{\sigma}_{1b}^2(\epsilon) = -\epsilon + \left[1 + \frac{\hat{\alpha}(\alpha - \hat{\alpha})}{1 + \hat{\alpha}^2}\right]^2\hat{\sigma}_1^2, \tag{4.26}$$

and similarly for $\hat{\sigma}_{2b}(\epsilon)$. The factor $\mathcal{J}(\epsilon)$ is a Jacobian factor given by

$$\mathcal{J}(\epsilon) = \left(1 - \frac{\epsilon}{f_1^2 \hat{\sigma}_1^2}\right)^{-1/2} \left(1 - \frac{\epsilon}{f_2^2 \hat{\sigma}_2^2}\right)^{-1/2}, \quad (4.27)$$

where f_1 and f_2 are given in terms of ϵ by Eqs. (4.22) and (4.24).

We next insert the result (4.25) for $p(\mathbf{x}|\epsilon)$ into Eqs. (2.9) and (2.10). The result is

$$\Lambda(\mathbf{x}) = \int d\epsilon p^{(0)}(\epsilon) \frac{p_\sigma[\hat{\sigma}_{1b}(\epsilon), \hat{\sigma}_{2b}(\epsilon)]}{p_\sigma[\hat{\sigma}_{1b}(0), \hat{\sigma}_{2b}(0)]} \frac{\mathcal{J}(\epsilon)}{\mathcal{J}(0)}$$

$$\times \exp\left[-\frac{N}{2(1+\hat{\alpha}^2)}(\alpha^2 - 2\alpha\hat{\alpha})\right]. \quad (4.28)$$

Finally, integrating over ϵ gives

$$\Lambda(\mathbf{x}) = \sqrt{\frac{2\pi(1+\hat{\alpha}^2)}{N}} p^{(0)}(\hat{\epsilon}) \frac{p_\sigma[\hat{\sigma}_{1b}(\hat{\epsilon}), \hat{\sigma}_{2b}(\hat{\epsilon})]}{p_\sigma[\hat{\sigma}_{1b}(0), \hat{\sigma}_{2b}(0)]} \frac{\mathcal{J}(\hat{\epsilon})}{\mathcal{J}(0)}$$

$$\times \exp\left[\frac{N\hat{\alpha}^2}{2(1+\hat{\alpha}^2)}\right] \Theta(\hat{\alpha}), \quad (4.29)$$

and invoking the arguments of Secs. III A and III B above for neglecting the prefactors gives

$$\Lambda(\mathbf{x}) \simeq \exp\left[\frac{N\hat{\alpha}^2}{2(1+\hat{\alpha}^2)}\right] \Theta(\hat{\alpha}). \quad (4.30)$$

Thus, in the limit $\hat{\alpha} \ll 1$, $\Lambda(\mathbf{x})$ is equivalent to the usual cross-correlation statistic $\Theta(\hat{\alpha})\hat{\alpha}$ defined by Eqs. (4.7), (4.8), (4.9), and (4.22).

We end this subsection by recapitulating the various approximations necessary to obtain the result:

(i) The large N approximation $N \gg 1$, necessary for the validity of the Laplace approximation in integrating over σ_1, σ_2.

(ii) The assumption that we are in the regime where the signal is detectable, $\exp[N\hat{\alpha}^2/2] \gg 1$. This is necessary for the evaluation of the integral over ϵ in Eq. (4.28), and for neglecting the prefactors in deriving Eq. (4.30).

(iii) The assumption, as before, that the prior probability distributions $p^{(0)}(\epsilon)$ and $p_\sigma(\sigma_1, \sigma_2)$ are slowly varying.

(iv) The weak signal approximation $\hat{\alpha} \ll 1$, which will be satisfied unless the stochastic background contribution to the output of one of the detectors becomes comparable to the noise in that detector. As discussed above, for signal strengths at the margin of detectability, and for several month searches for a stochastic background with ground based interferometers, we have $\hat{\alpha} \sim 10^{-4}$.

C. Two coincident, co-aligned detectors, white non-Gaussian noise

We next turn to the non-Gaussian noise model of Sec. III A of paper I. The noise in each detector is assumed to be white, with each sample statistically independent and identically distributed, so that

$$p_\mathbf{n}(\mathbf{n}_1, \mathbf{n}_2) = \prod_j \exp[-f_1(n_{1j}) - f_2(n_{2j})]. \quad (4.31)$$

However, it is clear that we cannot assume that the noise distributions e^{-f_1} and e^{-f_2} in each detector are known in advance. Otherwise, as explained in Sec. IV A, the analysis would predict that one can measure the stochastic background in a single detector by measuring the noise distribution and subtracting from it the "known" noise distribution.

Therefore, in this subsection, we will allow the functions $f_1(x)$ and $f_2(x)$ to be arbitrary except for the normalization conditions

$$\int e^{-f_1(x)} dx = \int e^{-f_2(x)} dx = 1. \quad (4.32)$$

Formally, there are an infinite number of parameters to specify to determine the functions f_1 and f_2. However, in practice these distributions will be measured as histograms, determined by a finite set of numbers or parameters. We identify this finite set of parameters with the noise parameters $\boldsymbol{\theta}_n$ of Eq. (2.14). We rewrite Eq. (2.14) as

$$p_\mathbf{n}(\mathbf{n}_1, \mathbf{n}_2) = \int \mathcal{D}f_1 \int \mathcal{D}f_2 p_\mathbf{n}(\mathbf{n}_1, \mathbf{n}_2 | f_1, f_2) p_f[f_1, f_2]. \quad (4.33)$$

Here for simplicity we have used a functional or path integral notation for the integral over the noise parameters (even though the integral is only over a finite number of parameters). In Eq. (4.33), $p_f[f_1, f_2]$ is the prior probability density functional for the functions f_1 and f_2, and $p_\mathbf{n}(\mathbf{n}_1, \mathbf{n}_2 | f_1, f_2)$ is given by the expression on the right hand side of Eq. (4.31). The probability distribution $p_f[f_1, f_2]$ encodes our assumption that the noise distributions will have central Gaussian regions, with unknown variances σ_1 and σ_2, and arbitrary tail regions containing a small fraction p_tail of the total probability.

We now insert the distributions (4.4) and (4.33) into Eqs. (2.11) and (2.14) and expand to second order in ϵ. The result is

$$p(\mathbf{x}|\epsilon) = \int \mathcal{D}f_1 \mathcal{D}f_2 p_f[f_1, f_2] \prod_j e^{-f_1(x_{1j}) - f_2(x_{2j})}$$

$$\times \prod_j [1 + \epsilon \mathcal{A}_j + \epsilon \mathcal{C}_j + \epsilon^2 \mathcal{E}_j + O(\epsilon^3)], \quad (4.34)$$

where

$$\mathcal{A}_j = \frac{1}{2}[f_1'(x_{1j})^2 + f_2'(x_{2j})^2 - f_1''(x_{1j}) - f_2''(x_{2j})],$$

$$(4.35)$$

and

$$\mathcal{C}_j = f_1'(x_{1j})f_2'(x_{2j}).\qquad(4.36)$$

The quantity $\mathcal{E}_j$ is a sum of terms of the form $f_1^{(n)}(x_{1j})^m$, $f_2^{(n)}(x_{2j})^m$, and $f_1^{(n)}(x_{1j})^m f_2^{(l)}(x_{2j})^r$ for integers n, m, l, r, whose exact form will not be needed here.

Working to second order in ϵ, we can re-express Eq. (4.34) as

$$p(\mathbf{x}|\epsilon) = \int \mathcal{D}f_1 \int \mathcal{D}f_2 p_f[f_1, f_2]\exp[-\Xi(\epsilon, f_1, f_2)],$$

$$(4.37)$$

where

$$\Xi(\epsilon, f_1, f_2) = \Xi_0[f_1, f_2] + \epsilon\Xi_1[f_1, f_2] + \epsilon^2\Xi_2[f_1, f_2]$$

$$+ O(\epsilon^3),\qquad(4.38)$$

where

$$\Xi_0[f_1, f_2] = -\sum_j f_1(x_{1j}) + f_2(x_{2j}),\qquad(4.39)$$

$$\Xi_1[f_1, f_2] = \sum_j (\mathcal{A}_j + \mathcal{C}_j),\qquad(4.40)$$

and

$$\Xi_2[f_1, f_2] = \sum_j \left[\mathcal{E}_j - \frac{1}{2}(\mathcal{A}_j + \mathcal{C}_j)^2\right].\qquad(4.41)$$

We now note that we can eliminate the autocorrelation terms $\mathcal{A}_j$ from Eq. (4.40) by making a change of variables. We define the operator $\mathcal{P}_\epsilon$ that acts on functions via

$$(\mathcal{P}_\epsilon f)(x) = f(x) - \frac{1}{2}\epsilon f'(x)^2 + \frac{1}{2}\epsilon f''(x),\qquad(4.42)$$

and we define the functions F_1 and F_2 by

$$F_1 = \mathcal{P}_\epsilon f_1, \quad F_2 = \mathcal{P}_\epsilon f_2.\qquad(4.43)$$

Using Eqs. (4.35), (4.38)–(4.40), (4.42) and (4.43) the functional Ξ can be rewritten as

$$\Xi(\epsilon, f_1, f_2) = \Xi_0[F_1, F_2] + \epsilon\tilde{\Xi}_1[F_1, F_2] + \epsilon^2\tilde{\Xi}_2[F_1, F_2]$$

$$+ O(\epsilon^3).\qquad(4.44)$$

Here the first order piece $\tilde{\Xi}_1$ consists only of the cross-correlation term,

$$\tilde{\Xi}_1[F_1, F_2] = \sum_j F_1'(x_{1j})F_2'(x_{2j});\qquad(4.45)$$

the corrections to this cross-correlation expression due to changing from f_1, f_2 to F_1, F_2 do not appear at this (linear) order in ϵ and instead contribute to the second order term $\tilde{\Xi}_2[F_1, F_2]$. The exact form of the functional $\tilde{\Xi}_2[F_1, F_2]$ will not be needed for our arguments below.

We define the functions $\hat{f}_1(x)$ and $\hat{f}_2(x)$ to be the functions corresponding to the measured noise distributions at the two detectors. That is, they are step functions defined by the requirement

$$\int_{-\infty}^x e^{-\hat{f}_1(u)}du = \frac{1}{N}\sum_{j \text{ with } x_{1j} \leq x} 1,\qquad(4.46)$$

with a similar equation for $\hat{f}_2$.

Consider now the evaluation of the integral (4.37) with Ξ given by the expression (4.44). Consider first the $\epsilon \to 0$ limit. In this limit one can show from the normalization conditions (4.32) that Ξ will be minimized at the measured noise distributions:

$$F_1 = \hat{f}_1, \quad F_2 = \hat{f}_2.\qquad(4.47)$$

For nonzero ϵ, the leading order correction to the $\epsilon = 0$ result will be given by evaluating the function (4.44) at the local minimum (4.47). We thus arrive at

$$p(\mathbf{x}|\epsilon) = \mathcal{J}(\mathbf{x})p_f[\mathcal{P}_\epsilon^{-1}\hat{f}_1, \mathcal{P}_\epsilon^{-1}\hat{f}_2]\exp\{-\Xi_0[\hat{f}_1, \hat{f}_2]\}$$

$$\times \exp\{-\epsilon\tilde{\Xi}_1[\hat{f}_1, \hat{f}_2] - \epsilon^2\tilde{\Xi}_2[\hat{f}_1, \hat{f}_2]\}.\qquad(4.48)$$

Here $\mathcal{J}(\mathbf{x})$ is a width factor whose origin is approximating the integrals over f_1 and f_2 as the value of the integrand at the peak times the "width" of the peak [15]. This factor is analogous to the various factors that appear in front of the exponential in Eq. (4.25). It depends weakly on $\mathbf{x}$ in comparison to the exponential factors. The second order functional $\tilde{\tilde{\Xi}}_2$ in Eq. (4.48) will differ from the corresponding functional $\tilde{\Xi}_2$ in Eq. (4.44) since the location of the peak of the integrand will receive a correction of order ϵ away from the value (4.47) which will give a correction of order $O(\epsilon^2)$ to the value of the integral.

We now insert the formula (4.48) for $p(\mathbf{x}, \epsilon)$ into Eqs. (2.9) and (2.10). This gives

$$\Lambda(\mathbf{x}) = \int_0^\infty d\epsilon p^{(0)}(\epsilon)\frac{p_f[\mathcal{P}_\epsilon^{-1}\hat{f}_1, \mathcal{P}_\epsilon^{-1}\hat{f}_1]}{p_f[\hat{f}_1, \hat{f}_1]}\exp\{-\epsilon\tilde{\Xi}_1[\hat{f}_1, \hat{f}_2]$$

$$- \epsilon^2\Xi_2[\hat{f}_1, \hat{f}_2]\}.\qquad(4.49)$$

Evaluating the integral over ϵ gives

$$\Lambda(\mathbf{x}) = \sqrt{\frac{\pi}{\tilde{\tilde{\Xi}}_2}}p^{(0)}(\hat{\epsilon})\frac{p_f[\mathcal{P}_{\hat{\epsilon}}^{-1}\hat{f}_1, \mathcal{P}_{\hat{\epsilon}}^{-1}\hat{f}_1]}{p_f[\hat{f}_1, \hat{f}_1]}$$

$$\times \exp\left\{\frac{\tilde{\Xi}_1[\hat{f}_1, \hat{f}_2]^2}{4\tilde{\Xi}_2[\hat{f}_1, \hat{f}_2]}\right\}\Theta(\tilde{\Xi}_1),\qquad(4.50)$$

and as before we can neglect the prefactors to give

$$\Lambda(\mathbf{x}) \simeq \exp\left\{\frac{\widetilde{\Xi}_1[\hat{f}_1, \hat{f}_2]^2}{4\widetilde{\widetilde{\Xi}}_2[\hat{f}_1, \hat{f}_2]}\right\} \Theta(\widetilde{\Xi}_1). \qquad (4.51)$$

Now the statistic $\widetilde{\Xi}_1[\hat{f}_1, \hat{f}_2]$ defined by Eqs. (4.45) and (4.46) coincides with the locally optimal statistic obtained in paper I [Eq. (3.8) of paper I, specialized to a white stochastic background], except for the following modification. One first measures the noise probability distributions in each detector separately [cf. Eq. (4.46)]. Then, one computes the generalized cross-correlation statistic (4.45) using those distributions and the measured data.

From Eq. (4.51), we see that $\Lambda(\mathbf{x})$ will be approximately equivalent to the locally optimal statistic $\widetilde{\Xi}_1[\hat{f}_1, \hat{f}_2]$ if the statistic $\widetilde{\widetilde{\Xi}}_2[\hat{f}_1, \hat{f}_2]$ has a weak dependence on the data $\mathbf{x}$. To establish this, consider the limit $p_{\text{tail}} \rightarrow 0$, where $p_{\text{tail}} \ll 1$ is the total probability in the noise distribution tails. In that limit our assumptions imply that the noise distributions in the two detectors are Gaussians with unknown variances σ_1, σ_2, and therefore the analysis of this subsection reduces to the analysis of Sec. IV B above. Therefore we can read off the $p_{\text{tail}} \rightarrow 0$ limit of the statistic $\widetilde{\widetilde{\Xi}}_2[\hat{f}_1, \hat{f}_2]$ by comparing Eqs. (4.28) and (4.49) and identifying the coefficients of ϵ^2 in the arguments of the exponentials. We thus obtain

$$\widetilde{\widetilde{\Xi}}_2[\hat{f}_1, \hat{f}_2] = \frac{N}{2(1 + \hat{\alpha}^2)} \frac{1}{\hat{\sigma}_1^2 \hat{\sigma}_2^2} [1 + O(p_{\text{tail}})]. \quad (4.52)$$

In the limit $\hat{\alpha} \ll 1$ we can neglect the $\hat{\alpha}$ dependence in Eq. (4.52). We then see that the $\mathbf{x}$-dependent fluctuations in the statistic are suppressed by either the small parameter p_{tail}, as in Secs. III C and III D above, or by the parameter $1/\sqrt{N}$ governing the size of the fractional fluctuations of the statistics $\hat{\sigma}_1$, $\hat{\sigma}_2$. Thus, we can neglect the $\mathbf{x}$ dependence of $\widetilde{\widetilde{\Xi}}_2[\hat{f}_1, \hat{f}_2]$ in Eq. (4.51), and the approximate equivalence of $\Lambda(\mathbf{x})$ and the locally optimal statistic $\widetilde{\Xi}_1[\hat{f}_1, \hat{f}_2]$ follows.

Finally, we remark that we have not analyzed, in this paper, the most general situation for stochastic signals of separated, nonaligned detectors with colored noise, which was analyzed in Sec. IV B of paper I [16]. However, the results we have obtained make it very plausible that, for that more general situation, the Bayesian statistic $\Lambda(\mathbf{x})$ should again be equivalent to the generalized cross-correlation statistic derived in paper I.

We also note that our assumption that the stochastic signal be Gaussian is necessary for our analysis. Modifying the signal by making it non-Gaussian instead of Gaussian would alter Eqs. (4.6), (4.19) and (4.38) at $O(\epsilon^2)$. Therefore, the derivation here does not generalize straightforwardly to non-Gaussian stochastic signals, unlike the corresponding derivations in paper I. In Ref. [5] it is shown that one can find detection techniques tailored to non-Gaussian stochastic signals that perform better, for such signals, than the methods considered here.

V. CONCLUSION

The derivation in this paper, from a different framework, of the detection strategies obtained in paper I gives us increased confidence in the utility of those strategies. In addition, the analysis of this paper has clarified the regime in which we expect the strategies to work well. For deterministic signals, data segments to be analyzed should be long enough that the signal-to-noise squared per data point be small. This requirement is easy to satisfy in practice, as signal-to-noise thresholds are usually in the range $5 - 10$. In addition, the strategies will only be close to optimal in the regime where signals are strong enough to be detectable; this restriction is unimportant in practice, as the performance of detection statistics in the regime where signals are far too weak to be detected is not important. Finally, for stochastic signals, the signal must be small compared to the noise in each individual detector, and the total probability in the tail part of the noise distributions must be small.

---

[1] B. Allen, J.D.E. Creighton, É.É. Flanagan, and J.D. Romano, Phys. Rev. D **65**, 122002 (2002).

[2] S. Klimenko and G. Mitselmakher, LIGO technical report LIGO-T010125-00-D.

[3] L.S. Finn, Phys. Rev. D **46**, 5236 (1992).

[4] P.J. Bickel and K.A. Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics* (Holden-Day, California, 1977), Sec. 6.4.

[5] S. Drasco and É.É. Flanagan, Phys. Rev. D **67**, 082003 (2003).

[6] L.S. Finn and J.D. Romano (in preparation).

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

[8] Saleem A. Kassam, *Signal Detection in Non-Gaussian Noise* (Springer-Verlag, New York, 1988).

[9] R.F. Michelson, Mon. Not. R. Astron. Soc. **227**, 933 (1987); N.L. Christensen, Phys. Rev. D **46**, 5250 (1992).

[10] B. Allen, in *Proceedings of the Les Houches School on Astrophysical Sources of Gravitational Waves*, edited by Jean-Alain Marck and Jean-Pierre Lasota (Cambridge University Press, Cambridge, England, 1997).

[11] É.É. Flanagan, Phys. Rev. D **48**, 2389 (1993). Note that the second term on the RHS of Eq. (B6) should read $-10j_1(\alpha)$ rather than $-2j_1(\alpha)$, that the sliding delay function shown in Fig. 2 is incorrect, and that the right hand side of Eq. (2.8) should be divided by π.

[12] The quantity denoted by $\hat{s}_j$ here was denoted by s_j in paper I.

[13] In the case where the prior distribution $p^{(0)}(\epsilon)$ restricts the sign of ϵ to be positive, one can similarly show that, in the relevant regime given by $\hat{\epsilon}(\mathbf{x}) > 0$ together with the condition (3.7), $\Lambda(\mathbf{x})$ is to a good approximation a monotonic function of the standard detection statistic $\Theta[\hat{\epsilon}(\mathbf{x})]\hat{\epsilon}(\mathbf{x})$, where Θ is the step function. Similar comments apply to the statistics (3.39) and (3.47).

[14] This follows from the relation $\rho^2/N = \hat{s}_{\mathrm{rms}}^2 \epsilon^2/\sigma^2$, where $\hat{s}_{\mathrm{rms}}$ is the rms average of the quantities $\hat{s}_j$, and from the assumption $\rho^2/N \ll 1$.

[15] In order that the integrand in Eq. (4.37) be sharply peaked, it is necessary to make the number of parameters specifying the functions f_1, f_2, $\hat{f}_1$, and $\hat{f}_2$ considerably less than the number N of data points, by modifying Eq. (4.46) to incorporate a suitable coarse-graining of the binning. This modification does not affect our argument.

[16] For Gaussian noise, the equivalence of $\Lambda(\mathbf{x})$ and the standard cross-correlation statistic has been demonstrated for separated, nonaligned detectors with colored noise in Appendix A of Ref. [11], in the limit of weak signals.