Optimal strategies for sinusoidal signal detection

Bruce Allen*

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

Maria Alessandra Papa† and Bernard F. Schutz‡

Max-Planck-Institut für Gravitationsphysik, Albert-Einstein-Institut, Am Mühlenberg 1, D-14476 Golm, Germany

We derive and study optimal and nearly optimal strategies for the detection of sinusoidal signals hidden in additive (Gaussian and non-Gaussian) noise. Such strategies are an essential part of algorithms for the detection of the gravitational continuous wave (cw) signals produced by pulsars. Optimal strategies are derived for the case where the signal phase is not known and the product of the signal frequency and the observation time is nonintegral.

I. INTRODUCTION

A key problem in data analysis is to detect sinusoidal signals in noise. Such signals are often called "lines" or "peaks" because in the Fourier domain (frequency space) they appear as spikes (linelike features) or sharp narrow peaks in the energy spectrum of the signal. When the signal is large compared to the noise, such signals are easy to identify. When it is weak, the identification becomes more difficult.

The work in this paper was motivated by the development of algorithms to search for continuous wave (cw) signals in the new generation of interferometric gravitational-wave detectors which are either under construction [1–5] or planned [6]. These signals are produced by rapidly spinning neutron stars (pulsars).

To search for new (previously undetected) pulsars requires a search over possible sky positions, frequencies, and pulsar spin-down parameters. The parameter space is very large and these searches are computationally very intensive. Moreover, the searches will be looking for signals that are (statistically) at the lower limit of detection sensitivity [7].

A brute-force approach (optimally filtering for all possible source parameters) requires unrealistic computational resources (petaflops), so more sophisticated hierarchical approaches have been proposed. When the parameter space is very large, these approaches retain much or all of the sensitivity of the brute-force approach but require fewer computational resources. This is possible because, in the brute-force approach, the number of grid points in parameter space is so large that the detection threshold must be set very high to avoid false alarms and enable confident detection. A hierarchical search visits fewer points in parameter space: it ignores those below the (high) threshold that one must set in order to gain the necessary detection confidence while examining a large parameter space. In other words a hierarchical search method does not "waste" precious computational

cycles examining regions in parameter space where, even if a signal were present, it would not be detected confidently enough.

The hierarchical search techniques [8–11] all involve a second (so-called incoherent) stage. This stage is called "incoherent" because it uses spectral rather than amplitude data. If one neglects polarization, in all of the proposed approaches a putative signal at the second stage would (effectively) appear in a spectrum as a sinusoidal signal at fixed frequency and phase. The third stage of the search works only on the regions in parameter space where significant spectral lines were identified in the second stage.

Our paper addresses the problem of identifying these candidates, that is, "registering" candidate sinusoidal signals. The analysis makes use of the Neyman-Pearson criterion to identify the "best" statistic to use for such identification. In some cases, the best statistic depends upon the expected amplitude of the signal, which is unknown. In these cases, we have used locally optimal methods to identify the best statistic in the weak-signal limit.

The analysis is complicated by several factors.

(1) The signal frequency and phase are not known in advance.

(2) The signal frequency may not lie at an integer multiple of the Rayleigh frequency T^{-1}. A signal of this type does not make an integer number of cycles during the observation time T. We call such frequencies, and the corresponding signal, "unresolved."

(3) The signal frequency must be identified with resolution less than $\pm(2T)^{-1}$, i.e., to within the nearest frequency bin.

(4) The method must handle non-Gaussian noise in an optimal manner.

The analysis presented here addresses all of these concerns.

II. PROBLEM DESCRIPTION AND OPTIMAL STATISTICS

The basic problem that we consider is the following. We are given N samples of a time-domain data stream, sampled at discrete times $t = t_j = j\Delta t$. We denote these data by y_j for $j = 0, 1, \ldots, N-1$. The total observation time is $T = N\Delta t$.

————
*Electronic address: ballen@uwm.edu

†Electronic address: papa@aei.mpg.de

‡Electronic address: schutz@aei.mpg.de

The question that we want to answer is, does the data stream y_j contain a sinusoidal signal

$$y_j = \epsilon \frac{2}{N} \cos(2\pi f t_j - \phi) \qquad (2.1)$$

of constant amplitude[1] and frequency? To address this question, we make use of the theory of optimal signal detection. It is convenient to recast the problem in the Fourier domain. Denote the discrete Fourier transform (DFT) of the data[2] by x_k:

$$x_k = \sum_{j=0}^{N-1} y_j e^{2\pi ijk/N} \quad \text{for} \quad k = -N/2+1, \dots, N/2. \qquad (2.2)$$

Since this transformation is invertible, any question or statement about the y's can also be stated in terms of the x's; hence we will often use the term "data" to refer to the x's rather than to the y's. Here, and elsewhere, the symbols x and y without indices refer to the collective ensemble of all the data. For convenience we will assume that N is a power of 2. The index k will often be referred to as a "frequency bin." The frequencies that these bins correspond to,

$$f_k = \frac{k}{N\Delta t} = \frac{k}{T}, \qquad (2.3)$$

are called "resolved frequencies" for reasons that will become clear later.

In what follows, we will assume that the data y are real. In this case, $x_k = x_{-k}^*$ where the asterisk denotes the complex conjugate, and both x_0 and $x_{N/2}$ are real. The data set y is then exactly equivalent to the set of x_k for $k = 0, \dots, N/2$. To simplify the mathematics, we will assume that the average value of the y's vanishes (i.e., that the dc or average value has been removed from the data) so that $x_0 = 0$. We will also assume that there is no energy at the Nyquist frequency $f_{N/2}$ (which in a real experiment would be enforced by appropriate antialiasing filters) so that $x_{N/2} = 0$. Then, the data set y is exactly equivalent to the set x_k for $k = 1, \dots, N/2-1$.

We use the notation $p(x|\epsilon)$ to denote the probability distribution function (PDF) of the data, in the presence of a signal whose amplitude is ϵ. For example, if the (real and imaginary parts of the) noise in each frequency bin is independent and Gaussian with vanishing mean and unit variance, and the signal is a sinusoid of known phase at resolved frequency f_ℓ given by $y_j = \epsilon(2/N)\cos(2\pi f_\ell t_j - \phi)$, then

$$p(x|\epsilon) = \frac{1}{2\pi} e^{-|x_\ell - \epsilon\, e^{i\phi}|^2/2} \prod_{\substack{k=1 \\ k \neq \ell}}^{N/2-1} \frac{1}{2\pi} e^{-|x_k|^2/2}.$$

_____

[1]The factor $2/N$ in the amplitude of the cosine simplifies the form of the frequency-space PDF, while retaining the standard definitions of the DFT.

[2]This is the traditional "physics" definition. The "engineering" definition has the opposite sign of i.

Note that since ℓ is an integer and ϵ is real the signal only affects the ℓth frequency bin.

If the assumptions are the same as above, but the phase of the signal is unknown and uniformly distributed over the range $\phi \in [0, 2\pi)$, then

$$p(x|\epsilon) = \frac{1}{2\pi} \int_0^{2\pi} d\phi \frac{1}{2\pi} e^{-|x_\ell - \epsilon e^{i\phi}|^2/2}$$

$$\times \prod_{\substack{k=1 \\ k \neq \ell}}^{N/2-1} \frac{1}{2\pi} e^{-|x_k|^2/2}.$$

Somewhat later, we will relax these assumptions, and give more general forms for $p(x|\epsilon)$ where the signal frequency is not a resolved frequency, the noise is not white, and the noise is not Gaussian.

Note that the integration measure for $p(x|\epsilon)$ is

$$\int dx \equiv \prod_{k=1}^{N/2-1} \int_{-\infty}^{\infty} d\Re x_k \int_{-\infty}^{\infty} d\Im x_k,$$

where $\Re$ and $\Im$ denote the real and imaginary parts.

The problem that we wish to solve is well known in the theory of signal detection. The space of possible measurements x_k for $k = 1, \dots, N/2-1$ is $\mathbf{R}^{N-2}$. Our goal is to divide this space of possible measurements into two disjoint regions H_0 and H_1, whose union is all of $\mathbf{R}^{N-2}$. If the observed data lie in H_0 (the "null-hypothesis region") we will conclude that no signal was present in the data. If the data lie in H_1, we will conclude that a signal was present. The problem we need to solve is this: What is the best choice of H_0 and H_1?

The solution we chose is the Neyman-Pearson criterion: the best choice is the one that gives the lowest false dismissal probability for a given false alarm probability. The false alarm probability α is the probability that a signal is detected when none is present,

$$\alpha = \int_{x \in H_1} dx\, p(x|0), \qquad (2.4)$$

and the false dismissal probability $\beta(\epsilon)$ is the probability that a signal of amplitude ϵ is not found,

$$\beta(\epsilon) = \int_{x \in H_0} dx\, p(x|\epsilon). \qquad (2.5)$$

The Neyman-Pearson criterion leads immediately to the following rule to partition the space of possible measurements into H_0 and H_1. Define the likelihood function on the space of possible measurements by

$$\Lambda(x) = \frac{p(x|\epsilon)}{p(x|0)}$$

and consider the surface $\Lambda(x) = \Lambda_0 = $ const. The Neyman-Pearson criterion leads to the following choice. Take H_0 to be the region inside this surface, and H_1 to be the region

outside this surface. The value of Λ_0 that defines the surface determines the false alarm and false dismissal probabilities.

In this paper, we will use the Neyman-Pearson criterion to define an "optimal statistic" which we will denote $\tau(x)$. This is any function of the observed data x whose *level surfaces* are the same as the level surfaces of $\Lambda(x)$. If the statistic is greater than some threshold $\mathcal{T}$ then we conclude that a signal is present, and if the statistic is less than the threshold $\mathcal{T}$ we conclude that no signal was present. The false alarm and false dismissal probabilities are functions of this threshold $\mathcal{T}$: as $\mathcal{T}$ is increased the false alarm probability gets smaller, and the false dismissal probability gets larger. In general this optimal statistic is a function of the signal amplitude ϵ. However, we will see that for the pulsar detection problem, where ϵ is small, the optimal statistic is effectively ϵ independent.

III. A WORKED EXAMPLE

To help make these ideas concrete, we give a complete worked example, demonstrating these ideas for the second PDF described above: a signal of unknown phase at a resolved frequency f_ℓ. The PDF is

$$p(x|\epsilon) = \frac{1}{2\pi} \int_0^{2\pi} d\phi \frac{1}{2\pi} e^{-|x_\ell - \epsilon e^{i\phi}|^2/2}$$

$$\times \prod_{\substack{k=1 \\ k \neq \ell}}^{N/2-1} \frac{1}{2\pi} e^{-|x_k|^2/2}. \tag{3.1}$$

Before continuing, it is convenient to express this in closed form. Writing the complex data sample $x_\ell = |x_\ell|\exp(i\psi_\ell)$ in terms of its modulus $|x_\ell|$ and phase ψ_ℓ, one has

$$\frac{1}{2\pi} \int_0^{2\pi} d\phi \frac{1}{2\pi} e^{-|x_\ell - \epsilon e^{i\phi}|^2/2}$$

$$= \frac{1}{2\pi} \int_0^{2\pi} d\phi \frac{1}{2\pi} e^{-[|x_\ell|^2 + \epsilon^2 - 2\epsilon\Re(x_\ell^* e^{i\phi})]/2}$$

$$= \frac{1}{2\pi} e^{-(|x_\ell|^2 + \epsilon^2)/2} \frac{1}{2\pi} \int_0^{2\pi} d\phi e^{\epsilon|x_\ell|\cos(\phi - \psi_\ell)}$$

$$= \frac{1}{2\pi} e^{-(|x_\ell|^2 + \epsilon^2)/2} I_0(\epsilon|x_\ell|).$$

The final integral has been expressed in terms of a modified Bessel function $I_0(r)$ of the first kind,

$$I_0(r) = \frac{1}{\pi} \int_0^\pi d\theta e^{r\cos\theta}.$$

Thus we obtain a closed form for the PDF (3.1):

$$p(x|\epsilon) = e^{-\epsilon^2/2} I_0(\epsilon|x_\ell|) \prod_{k=1}^{N/2-1} \frac{1}{2\pi} e^{-|x_k|^2/2}. \tag{3.2}$$

The likelihood function is now easily found:

$$\Lambda(x) = \frac{p(x|\epsilon)}{p(x|0)} = e^{-\epsilon^2/2} I_0(\epsilon|x_\ell|). \tag{3.3}$$

While in a general situation the likelihood function depends upon all the different variables, in this particular situation it depends only upon $|x_\ell|$.

We defined an optimal statistic τ to be any function whose level surfaces are the same as the level surfaces of the likelihood function $\Lambda(x)$. In this simple situation, the likelihood function $\Lambda(x) = \Lambda(x_1, \ldots, x_{N/2-1})$ depends only upon the modulus $|x_\ell|$ of the amplitude in a single (the ℓth) Fourier bin. Since it is a monotonically increasing function of $|x_\ell|$, we can choose as an optimal statistic any monotonic function of $|x_\ell|$, for example $|x_\ell|$ or $|x_\ell|^2$. For historical and later convenience, let us choose as our optimal statistic the function $\tau = |x_\ell|^2$. This is the power in the ℓth bin. The mean value of this statistic, the power in the ℓth bin, is

$$\int dx\, \tau p(x|\epsilon) = \int dx |x_\ell|^2 p(x|\epsilon) = 2 + \epsilon^2. \tag{3.4}$$

In the absence of a signal ($\epsilon = 0$) both the real and imaginary parts of x_ℓ contribute unity.

To complete the analysis of this example, we need to calculate the false alarm and false dismissal probabilities. We will define, for a given value of threshold $\mathcal{T}$, the regions H_0 and H_1 by:

$$H_0 = \{(x_1, \ldots, x_{N/2-1}) \text{ such that } \tau = |x_\ell|^2 \leq \mathcal{T}\},$$

and

$$H_1 = \{(x_1, \ldots, x_{N/2-1}) \text{ such that } \tau = |x_\ell|^2 > \mathcal{T}\}.$$

Thus our choice of statistic gives a decision rule which has a simple physical interpretation. If the power in bin ℓ is greater than $\mathcal{T}$, we conclude that a signal was present. If not, we conclude that no signal was present.

The false alarm probability (2.4) is easy to calculate. It is given by the following function of the threshold $\mathcal{T}$:

$$\alpha(\mathcal{T}) = \int_{x \in H_1} dx\, p(x|0)$$

$$= \int_{|x_\ell|^2 > \mathcal{T}} dx\, p(x|0)$$

$$= \int_{|x_\ell|^2 > \mathcal{T}} dx \prod_{k=1}^{N/2-1} \frac{1}{2\pi} e^{-|x_k|^2/2}$$

$$= \int d\Re x_\ell \int d\Im x_\ell \frac{1}{2\pi} e^{-|x_\ell|^2/2}$$

$$= \int_0^{2\pi} d\psi_\ell \int_{|x_\ell| = \sqrt{\mathcal{T}}}^\infty |x_\ell| d|x_\ell| \frac{1}{2\pi} e^{-|x_\ell|^2/2}$$

$$= \int_{\mathcal{T}/2}^\infty d\left(\frac{1}{2}|x_\ell|^2\right) e^{-|x_\ell|^2/2}$$

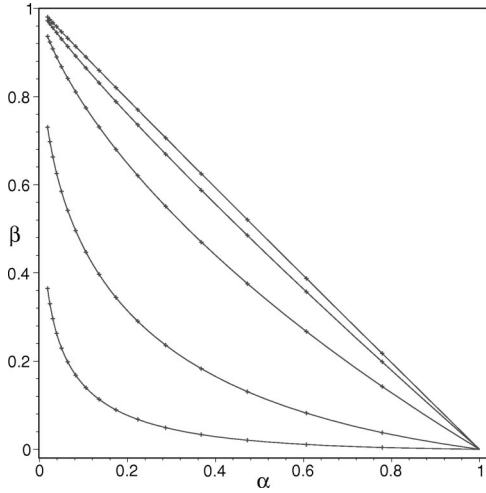$$= e^{-\mathcal{T}/2}. \tag{3.5}$$

FIG. 1. The false dismissal probability $\beta(\mathcal{T})$ as a function of the false alarm probability $\alpha(\mathcal{T})$ for different values of the signal amplitude ϵ. The top curve has $\epsilon = 0.2$. Moving down, the remaining curves have $\epsilon = 0.5, 1.0, 2.0, 3.0$. Along each curve, the threshold $\mathcal{T}$ varies from 0 to 8. In the bottom right of the graph, $\mathcal{T} = 0$. The crosses mark the points where $\mathcal{T} = 1/2, 1, 3/2, \ldots, 8$. For example, with a threshold $\mathcal{T} = 5.5$, if the signal amplitude is $\epsilon = 3$, then the false alarm probability is $\alpha \approx 6.4\%$ and the false dismissal probability is $\beta \approx 20\%$.
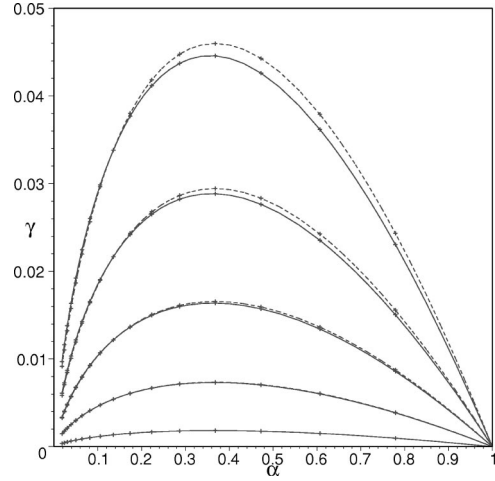

FIG. 2. Solid curves: Detection probability $\gamma = 1 - \alpha - \beta$ as a function of the false alarm probability α for different values of the signal amplitude $\epsilon = 0.1, \ldots, 0.5$ (moving up from the bottom curve). The crosses mark different values of the threshold in the same way as for Fig. 1. Dashed curves: The $O(\epsilon^2)$ approximation is $\gamma = \epsilon^2 \mathcal{T} \exp(-\mathcal{T}/2)/4 = -\epsilon^2 \alpha \ln \alpha/2$. The $O(\epsilon^4)$ approximation to γ is not shown because on this graph it is indistinguishable from the exact result (the solid curves).

In this calculation, the transition from the third to the fourth line is trivial because we integrate over all the coordinates except for x_ℓ. In going from the fourth to the fifth line, we have changed variables from the real and imaginary parts to polar coordinates.

The false dismissal probability (2.5), which depends both upon the signal amplitude ϵ and upon the value $\mathcal{T}$ of the decision statistic threshold, is obtained with a similar calculation:

$$\beta(\mathcal{T}) = \int_{x \in H_0} dx\, p(x|\epsilon)$$

$$= \int_{|x_\ell|^2 \leqslant \mathcal{T}} dx\, p(x|\epsilon)$$

$$= e^{-\epsilon^2/2} \int_0^{\sqrt{\mathcal{T}}} |x_\ell|\, d|x_\ell|\, e^{-|x_\ell|^2/2} I_0(\epsilon |x_\ell|)$$

$$= e^{-\epsilon^2/2} \int_0^{\mathcal{T}/2} du\ e^{-u} I_0(\epsilon \sqrt{2u}). \tag{3.6}$$

This final integral cannot be evaluated in closed form. However, it is easy to check that the limit $\beta(\infty) = 1$: if the threshold is set very large, then the false dismissal probability is unity. In a moment, we will study the behavior of β in the weak-signal limit as $\epsilon \to 0$. However, before this, it is instructive to study the false alarm versus false dismissal curves for this statistic.

The false alarm and false dismissal curves for this optimal detection statistic are illustrated in Fig. 1. Plotting β as a function of α provides a way of describing the optimal sta-

tistic which is completely independent of the actual choice of the statistic.[3] However, the relationship between the threshold $\mathcal{T}$ and the false alarm and false dismissal probabilities does depend upon the choice of optimal statistic. Because this statistic has been chosen by the Neyman-Pearson criterion, any other detection statistic that we choose will have poorer performance. Thus, for a given signal amplitude ϵ, and for a given false alarm probability α, any other detection statistic will have a larger false dismissal probability β: it will lie above the illustrated curves.

Our primary interest is in very weak signals. For the pulsar detection problem, we will have $\epsilon \approx 0.2$ and will be operating on the threshold of detection where $\alpha + \beta$ is only slightly smaller than unity. For such weak signals, it is useful to define the quantity

$$\gamma(\mathcal{T}) = 1 - \alpha(\mathcal{T}) - \beta(\mathcal{T}). \tag{3.7}$$

This may be considered either as a function of the threshold $\mathcal{T}$ or as a function of the false alarm probability $\alpha(\mathcal{T})$. This quantity γ is the difference between the detection probability when a signal is present, $1 - \beta$, and the false alarm probability α. For example, for a very weak signal, the threshold might be set for a false alarm probability of $\alpha = 15\%$. The false dismissal probability for this weak signal might be $\beta = 84\%$. Thus, if no signal is present, the threshold will be

――――――
[3]Remember that any statistic with the same level surfaces as $\Lambda(x)$ is an optimal statistic. There are an infinite number of different but equivalent choices possible.

exceeded $\alpha = 15\%$ of the time. If a signal is present, the threshold will be exceeded $1 - \beta = 16\%$ of the time. Roughly speaking, the difference between these, $\gamma = 1 - \alpha - \beta = 1\%$, is the probability of the threshold being exceeded because the signal was present, rather than because of the detector noise. These weak-signal-limit curves are shown in Fig. 2.

In the small-ϵ (weak-signal) limit, it is easy to obtain an

approximate closed form for β. By substituting the power series representation of the Bessel function,

$$I_0(x) = 1 + \frac{x^2}{4} + \frac{x^4}{64} + \frac{x^6}{2304} + \cdots$$

into Eq. (3.6) and integrating term by term, one obtains

$$\begin{aligned}
\gamma &= 1 - \alpha - \beta \\
&= \frac{\epsilon^2}{4} \mathcal{T} e^{-\mathcal{T}/2} \left[1 + \frac{\epsilon^2}{16}(\mathcal{T} - 4) + \frac{\epsilon^4}{576}(\mathcal{T}^2 - 12\mathcal{T} + 24) + \cdots \right] \\
&= -\frac{1}{2} \epsilon^2 \alpha \ln \alpha \left[1 - \frac{\epsilon^2}{8}(2 + \ln \alpha) + \frac{\epsilon^4}{144}(6 + 6 \ln \alpha + (\ln \alpha)^2) + \cdots \right].
\end{aligned} \tag{3.8}$$

Even at the lowest order in ϵ (the first term in square brackets) this is a very good approximation, as shown by the dashed curves in Fig. 2. At the next order (the first two terms in square brackets) the approximation is indistinguishable from the exact result in Fig. 2—the solid curves. This simplifies matters enormously. Although the statistics of the optimal detection strategy depends upon the signal amplitude ϵ, for small ϵ, this dependence is simple enough to be analytically approximated.

The detection probability plays a key role in the significance of an observation. A hierarchical pulsar search hunts for peaks in the spectra coming from a set of n sequential time series. For example, suppose each time series of length N is one day long. Four months of such data would correspond to $n = 120$. What choice of false alarm probability α (or, equivalently, of detection threshold $\mathcal{T}$) is optimal?

This question is easily answered. One might guess that the best operating point is where the detection probability $\gamma = 1 - \alpha - \beta$ is maximized: in the weak-signal case this is at a threshold of $\mathcal{T} = 2$ corresponding to a false alarm probability $\alpha = 1/e \approx 36.78\%$. However, this is not correct. In the absence of signal, each of the n data sets is independent. The probability of detecting peaks in p of the n data sets is the same as the probability that a coin will come up heads p times in n flips (if the probability of "heads" is the false alarm probability α). This is given by the binomial distribution:

$$\text{probability of } p \text{ peaks} = \binom{n}{p} \alpha^p (1 - \alpha)^{n-p}.$$

Thus, in the absence of a signal, the mean number of peaks is αn, and its variance is $\sigma^2 = \alpha(1 - \alpha)n$. In the presence of a signal, the mean number of peaks registered is $(1 - \beta)n$. A good way to choose a false alarm probability (or threshold) is to maximize the significance s. This is

$$\begin{aligned}
s &= \frac{\langle \# \text{ peaks} \rangle_{\text{signal}} - \langle \# \text{ peaks} \rangle_{\text{no signal}}}{\sigma} \\
&= \frac{(1 - \beta)n - \alpha n}{\sqrt{\alpha(1 - \alpha)n}} \\
&= \frac{1 - \alpha - \beta}{\sqrt{\alpha(1 - \alpha)}} \sqrt{n} \\
&= \frac{\gamma}{\sqrt{\alpha(1 - \alpha)}} \sqrt{n}.
\end{aligned} \tag{3.9}$$

The significance is easily calculated as a function of either α or $\mathcal{T}$. In the weak-signal limit, it is

$$\frac{s}{\sqrt{n}} = \frac{\epsilon^2}{4} \frac{\mathcal{T}}{\sqrt{e^{\mathcal{T}/2} - 1}} = -\frac{\epsilon^2}{2} \sqrt{\frac{\alpha}{1 - \alpha}} \ln \alpha.$$

The significance as a function of either $\mathcal{T}$ or α has a maximum at the threshold value $\mathcal{T} \approx 3.18721$ corresponding to a false alarm probability of $\alpha \approx 20.3188\%$. The significance at this threshold/false alarm probability is $s \approx 0.402371 \epsilon^2 \sqrt{n}$. Note that this exhibits the expected $\sqrt{n}$ scaling in the number n of spectra analyzed. We have numerically verified that this is the optimal statistic.

IV. EXAMPLE: LOCAL PEAK DETECTION—A NONOPTIMAL STRATEGY

Section III found and analyzed the optimal (i.e., Neyman-Pearson) peak detection strategy. In this section, we carry out an identical analysis of a different (hence nonoptimal) strategy. The main purpose is to illustrate a side-by-side comparison of different detection statistics.

We will assume that the signal and noise satisfy the same assumptions as in Sec. III, given by Eq. (3.2). There, we showed that the optimal detection strategy was to threshold

on the power $\tau=|x_\ell|^2$ in the ℓth bin. Here, we adopt a different detection strategy. We will say that a peak has been detected if and only if the power $|x_\ell|^2$ in the ℓth bin exceeds the threshold $\mathcal{T}$ *and* is greater than the power in either of the neighboring frequency bins. This strategy looks for "local peaks" that exceed the threshold.

For this peak detection strategy, the detection region H_1 is defined by

$$H_1=\{(x_1,\cdots,x_{N/2-1}) \text{ such that } |x_\ell|^2>\mathcal{T} \text{ and } |x_\ell|^2$$
$$>|x_{\ell-1}|^2 \text{ and } |x_\ell|^2>|x_{\ell+1}|^2\}.$$

In other words, the peak detection strategy is to register a peak if the observed data set lies in H_1. The null-hypothesis or no-signal region H_0 is the set complement $H_0=\mathbf{R}^{N-2}-H_1$: all points not lying in H_1.

To compare this strategy to the optimal one found in Sec. III, we calculate the false alarm and false detection curves as before, and compare them with the optimal strategy. The false alarm probability is

$$\alpha(\mathcal{T})=\int_{x\in H_1}dxp(x|0)$$
$$=\int\frac{dx_{\ell-1}}{2\pi}\int\frac{dx_\ell}{2\pi}\int\frac{dx_{\ell+1}}{2\pi}e^{-(|x_{\ell-1}|^2+|x_\ell|^2+|x_{\ell+1}|^2)/2},$$
$$|x_\ell|^2>\mathcal{T},$$
$$|x_\ell|^2>|x_{\ell-1}|^2,$$
$$|x_\ell|^2>|x_{\ell+1}|^2.$$

In these expressions, $\int dx_k$ denotes $\int_{-\infty}^{\infty}d\Re x_k\int_{-\infty}^{\infty}d\Im x_k$. Putting each of the three integrals into polar coordinates immediately yields

$$\alpha(\mathcal{T})=\frac{1}{2}\int_{\mathcal{T}}^{\infty}d|x_\ell|^2e^{-|x_\ell|^2/2}\left[\frac{1}{2}\int_0^{|x_\ell|^2}d|x_{\ell-1}|^2e^{-|x_{\ell-1}|^2/2}\right]^2$$

$$=\frac{1}{2}\int_{\mathcal{T}}^{\infty}d|x_\ell|^2e^{-|x_\ell|^2/2}[-e^{-u/2}|_{u=0}^{u=|x_\ell|^2}]^2$$

$$=\frac{1}{2}\int_{\mathcal{T}}^{\infty}d|x_\ell|^2e^{-|x_\ell|^2/2}[1-e^{-|x_\ell|^2/2}]^2$$

$$=\int_{\mathcal{T}/2}^{\infty}due^{-u}[1-e^{-u}]^2$$

$$=\frac{1}{3}e^{-3\mathcal{T}/2}-e^{-\mathcal{T}}+e^{-\mathcal{T}/2}. \tag{4.1}$$

The quantity in square brackets that appears in the intermediate steps of this calculation is simply the probability that bins $\ell\pm1$ contain less power than the ℓth bin. This is one minus the false alarm probability (3.5) of the optimal test.

As with the optimal test, the false alarm probability $\alpha(\mathcal{T})$ vanishes at large threshold $\mathcal{T}\to\infty$. However, unlike the optimal test, the false alarm probability at zero threshold is not unity: $\alpha(\mathcal{T}=0)=1/3$. This is because, even if the threshold vanishes, to register as a peak the ℓth bin must contain more power than both adjacent bins. When no signal is present, this happens only 1/3 of the time.

The false dismissal probability for this nonoptimal peak detection strategy can be calculated with the same methods as above. One finds

$$\beta(\mathcal{T})=\int_{x\in H_0}dxp(x|\epsilon)$$

$$=1-\int_{x\in H_1}dxp(x|\epsilon)$$

$$=1-\frac{1}{2}\int_{\mathcal{T}}^{\infty}d|x_\ell|^2e^{-(|x_\ell|^2+\epsilon^2)/2}I_0(\epsilon|x_\ell|)[1-e^{-|x_\ell|^2/2}]^2$$

$$=1-e^{-\epsilon^2/2}\int_{\mathcal{T}/2}^{\infty}du\ I_0(\epsilon\sqrt{2u})e^{-u}[1-e^{-u}]^2$$

$$=e^{-\epsilon^2/4}-\frac{1}{3}e^{-\epsilon^2/3}+e^{-\epsilon^2/2}\int_0^{\mathcal{T}/2}duI_0(\epsilon\sqrt{2u})e^{-u}[1-e^{-u}]^2. \tag{4.2}$$

As for the optimal statistic, this false dismissal probability approaches one at large threshold $\mathcal{T}\to\infty$. However, unlike the optimal test, it does not vanish at zero threshold. Setting $\mathcal{T}=0$ in Eq. (4.2) one finds that

$$\beta(\mathcal{T}=0)=e^{-\epsilon^2/4}-\tfrac{1}{3}e^{-\epsilon^2/3}.$$

If the signal amplitude is small $\epsilon\to0$ then $\beta(\mathcal{T}=0)\to2/3$. There is a 2/3 probability of missing a small signal at zero threshold, because one of the two neighboring frequency bins might contain more power than bin ℓ.

A set of false alarm/false dismissal curves for this nonoptimal statistic is shown in Fig. 3, along with the same curves
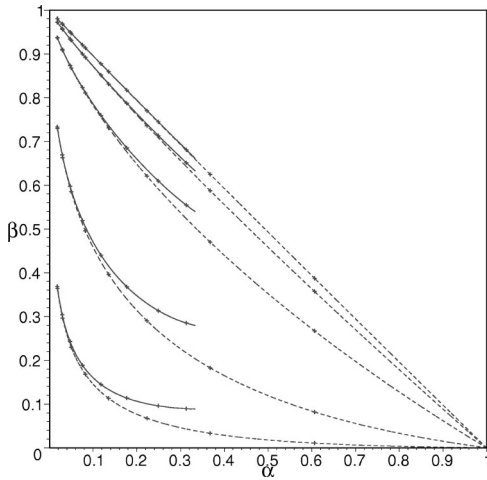
FIG. 3. Solid curves: False dismissal β versus false alarm α for the nonoptimal detection strategy of this section. Moving down from the top, the curves correspond to signal strengths $\epsilon = 0.2, 0.5, 1, 2, 3$. Notice that the false alarm probability α is less than 1/3 for *any* value of the threshold $\mathcal{T}$. For comparison, the dashed curves show the optimal strategy of the previous section. Notice that the optimal strategy always yields a lower false dismissal probability for a given false alarm probability. The crosses mark threshold values $\mathcal{T}=1,2,\ldots,8$ increasing to the left along each curve.

FIG. 4. These graphs are a comparison of two different peak-finding methods, in the weak-signal limit (small ϵ). The dashed curves correspond to the optimal (Neyman-Pearson) test: thresholding on the signal power. The solid curves correspond to the local peak test described in this section. The bottom graph shows the detection probability $\gamma/\epsilon^2 = (1-\alpha-\beta)/\epsilon^2$ as a function of false alarm probability α. The top graph shows the significance $\gamma/\epsilon^2\sqrt{\alpha(1-\alpha)}$. Table I compares the properties of these curves.

for the optimal statistic. Note that for a given signal strength and false alarm probability, the false dismissal probability is always lower for the Neyman-Pearson test. Also notice that at a given threshold, one test statistic does not yield the same false alarm probability as the other statistic. As the false alarm probability decreases, the two statistics have a performance (false dismissal probability) that becomes increasingly similar. This is because, at increasing values of the threshold $\mathcal{T}$, fewer and fewer peaks are rejected because the neighboring peaks are larger.

In the small-signal limit $\epsilon \to 0$, one can use the series expansion of the Bessel function to obtain analytic expressions for the false alarm probability β. The signal detection probability is

$$\gamma = 1 - \alpha - \beta$$
$$= \frac{\epsilon^2 e^{-3\mathcal{T}/2}}{4}\left[\mathcal{T}\left(e^{\mathcal{T}} - e^{\mathcal{T}/2} + \frac{1}{3}\right) + e^{\mathcal{T}/2} - \frac{4}{9}\right] + O(\epsilon^4)$$
$$= \frac{\epsilon^2}{4}\left[\alpha\mathcal{T} + e^{-\mathcal{T}} - \frac{4}{9}e^{-3\mathcal{T}/2}\right] + O(\epsilon^4).$$

This signal detection probability cannot be expressed in analytic form entirely in terms of α given by Eq. (4.1). However we can plot it and compare with the identical curve for the optimal strategy. This is shown in Fig. 4, which also shows the significance as a function of the false alarm probability. The comparison is shown in Table I.

The primary purpose of these last two sections was to demonstrate how a signal detection strategy can be chosen in an optimal fashion, and how it can be compared to a subop-
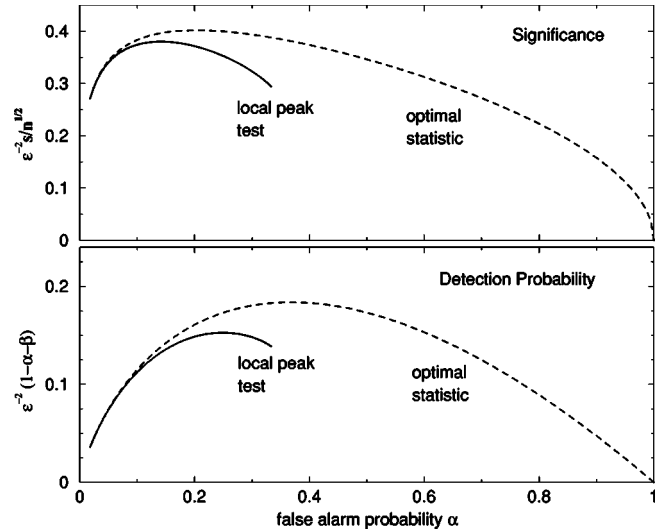
timal strategy. In a "real world" situation, it may be highly desirable to apply a suboptimal strategy, because the mathematical model of the instrumental noise may not be complete, and might not accurately reflect its real behavior. In fact, the suboptimal method discussed in this section has only slightly poorer performance for the simple Gaussian noise model than the optimal test, but may perform much better on "real world" data which have correlations between different frequency bins.

In the following section, we will apply these methods to develop optimal tests for the case where the sinusoidal signal frequency is not one of the exactly resolved frequencies f_k.

TABLE I. A comparison of the optimal Neyman-Pearson detection strategy and the suboptimal local peak detection method, in the weak-signal limit. Most of these values can be read off Fig. 4. The top half of the table gives information about the maximum of $\gamma = 1 - \alpha - \beta$, such as the value of the threshold at the maximum. The bottom half of the table gives the same information for the maximum of $(1-\alpha-\beta)/\sqrt{\alpha(1-\alpha)}$.

	Optimal test	Local peak test
Maximum of $1-\alpha-\beta$	$0.1839\epsilon^2$	$0.1529\epsilon^2$
is at threshold value $\mathcal{T}$	2.0	2.0
and false alarm probability α	36.79%	24.91%
Maximum of $\dfrac{1-\alpha-\beta}{\sqrt{\alpha(1-\alpha)}}$	$0.4024\epsilon^2$	$0.3806\epsilon^2$
is at threshold value $\mathcal{T}$	3.187	3.567
and false alarm probability α	20.32%	14.14%

V. COMMENTS ON THE WEAK-SIGNAL APPROXIMATION

In the previous sections, we studied the validity of the weak-signal limit $\epsilon \to 0$, and made use of it when appropriate. We will continue to take this limit throughout the paper. This brings up several interesting issues.

These types of weak-signal approximations have been studied extensively under the rubric of "locally optimal statistics" [12]. Later in this paper, they will make treatment of non-Gaussian noise models tractable.

In practice, the weak-signal approximation is well justified for the pulsar detection problem. This is dramatically illustrated in Fig. 2. This is a typical case: for $\epsilon < 1/2$ only the lowest-order terms in ϵ need to be retained in order to have a good approximation. Keeping the next order terms as well gives an extremely good approximation even for $\epsilon \approx 1$. Typical detectable signal strengths will be $\epsilon \approx 0.2$.

In the weak-signal limit, the PDF can be well approximated by the first nonvanishing term in its Taylor series in ϵ. The first derivative of $p(x|\epsilon)$ with respect to ϵ vanishes at $\epsilon = 0$, because p is an even function of ϵ. This is because the phase ϕ of the signal is uniformly distributed in the range $[0, 2\pi)$. The PDF is well approximated by

$$p(x|\epsilon) = p(x|0) + \frac{1}{2}\epsilon^2 p''(x|0) + O(\epsilon^4), \qquad (5.1)$$

where the prime denotes $\partial/\partial\epsilon$. The likelihood function is then approximated by

$$\Lambda(x) = \frac{p(x|\epsilon)}{p(x|0)} = 1 + \frac{1}{2}\epsilon^2 \frac{p''(x|0)}{p(x|0)}. \qquad (5.2)$$

Thus in the weak-signal case (neglecting fourth order terms in the signal amplitude ϵ) the optimal detection statistic is independent of signal strength, and can be found from the second derivative of the PDF at zero signal strength. This tremendously simplifies the analysis.

The likelihood function itself, or the likelihood function minus a constant, can be used as the optimal statistic τ (for example, threshold on $\Lambda - 1$). In the absence of signal, the mean value of this statistic must vanish. This follows immediately from the definition of Λ, since

$$\int dx p(x|0)(\Lambda - 1) = \int dx[p(x|\epsilon) - p(x|0)] = 0. \qquad (5.3)$$

In the weak-signal case, keeping only terms up to a given order (say ϵ^2) in $\Lambda - 1$, it is easy to show that the same relation holds. Hence, in the absence of a signal, the mean value of $\Lambda(x) - 1$ vanishes. This will be useful later.

VI. OPTIMAL DETECTION OF UNRESOLVED FREQUENCY SIGNALS

We now begin to address one of our key concerns. The previous sections showed how to systematically derive and characterize a detection strategy for the case where the weak sinusoidal signal had unknown phase, but where, if present, the signal's frequency precisely corresponded to one of the Fourier bins. We now suppose that the frequency is also a random variable, whose value is uniformly distributed between $(f_\ell + f_{\ell-1})/2$ and $(f_\ell + f_{\ell+1})/2$. In other words, the signal of interest lies somewhere between a half bin to the left and a half bin to the right of the ℓth frequency bin.

Before delving into the details of the analysis, it will be helpful to briefly examine the appearance (in frequency space) of an unresolved sinusoidal signal in the absence of noise. Take the signal frequency to be

$$f_l = \frac{l}{N\Delta t} \qquad (6.1)$$

where we do *not* assume that l is an integer (corresponding to one of the resolved frequencies). Let ℓ denote the nearest bin to l, so that

$$l = \ell - \delta \quad \text{for} \quad \delta \in (-1/2, 1/2]. \qquad (6.2)$$

Without loss of generality, we assume that the frequency f_l is between the dc and Nyquist frequencies, corresponding to the range $l \in (0, N/2)$. In the absence of noise, the signal in the time domain is given by

$$y_j = \epsilon \frac{2}{N}\cos(2\pi f_l j\Delta t - \phi) = \epsilon \frac{2}{N}\cos(2\pi jl/N - \phi).$$

Substituting this into the DFT (2.2) and using the sum of the geometric series

$$\sum_{j=0}^{N-1} z^j = \frac{1-z^N}{1-z} \qquad (6.3)$$

gives Fourier amplitudes

$$x_k = \epsilon[e^{i\phi}D_N(k-l) + e^{-i\phi}D_N(k+l)], \qquad (6.4)$$

where the function D_N is the *Dirichlet kernel*:

$$D_N(z) = e^{i\pi z(1-1/N)}\frac{\sin(\pi z)}{N\sin(\pi z/N)}. \qquad (6.5)$$

As described following Eq. (2.3), the range of the frequency index k is $1, \ldots, N/2 - 1$. Since $D_N(z)$ vanishes for all integer arguments except for zero, where its value is $D_N(0) = 1$, in the resolved-frequency case where l is an integer, one has $x_k = 0$ for $k \neq l$, and $x_l = \epsilon e^{i\phi}$. In the unresolved case, the signal energy is not confined to the ℓth bin, and forms a characteristic pattern of "sidelobes" in the nearby frequency bins.

If the signal frequency is unresolved (l noninteger) the optimal statistical test will not involve only data from the ℓth bin. The adjacent frequency bins also contain part of the signal energy, and we will shortly find that the statistically optimal search also takes into account their content (in the sense of energy and information).

One can simplify the form of the Dirichlet kernel with several approximations.[4] Our primary interest is to extract as much useful information as possible from the Fourier amplitudes in the bins near bin ℓ. Because $D_N(z)$ is strongly peaked at $z=0$ and falls off $\sim z^{-1}$ away from it, one may neglect the second term in Eq. (6.4) and concentrate on the first term. In addition, in practical applications, N will be large enough (greater than 10^5) that the $1/N$ term in the exponential of D_N can be neglected. Finally, since we will be interested in the Fourier amplitudes in nearby bins, $|z| \ll N$, which means that the denominator $N \sin(\pi z/N)$ is well approximated by πz. This leaves us with

$$x_k \approx \epsilon e^{i\phi} \omega(k-l),$$

where the coefficients

$$\omega(z) = e^{i\pi z} \frac{\sin \pi z}{\pi z}$$
$$= e^{i\pi z} j_0(\pi z)$$
$$= e^{i\pi z} \mathrm{sinc}(z). \qquad (6.6)$$

Here j_0 is a spherical Bessel function, and we have used Woodward and Bracewell's definition of the sampling function sinc.

We now suppose that the signal of interest is distributed, with equal probability, anywhere between $\pm 1/2$ a frequency bin from the ℓth bin, and write an expression for the PDF of the data. If, as before, the signal phase ϕ is a uniformly distributed random variable, and if the instrument noise is Gaussian and satisfies the same assumptions as before, one has

$$p(x|\epsilon) = \frac{1}{2\pi} \int_{-1/2}^{1/2} d\delta \int_0^{2\pi} d\phi \prod_{k=1-\ell}^{N/2-1-\ell} \frac{1}{2\pi}$$
$$\times e^{-|x_{k+\ell} - \epsilon\omega(k+\delta)e^{i\phi}|^2/2}. \qquad (6.7)$$

In this expression, which involves a product over all frequency bins, the index k has been shifted so that $k=0$ labels the ℓth bin.

When searching for a signal peak in the vicinity of the ℓth bin, there are practical reasons (computational efficiency and algorithm structure) why it is desirable to use only information from (some small number of) nearby bins.[5] Fortunately

---

[4]Further justification for these approximations may be found in Sec. X and Fig. 6.

[5]Section X and Fig. 6 show that virtually all the information is within a few bins from the ℓth bin.

for us, the Neyman-Pearson criterion can be easily derived for this more limited information: we merely write down the PDF for the parts of the data (the nearby bins) which are available to us. From this point on, we will assume that our search for a signal in the vicinity of the ℓth frequency bin is restricted to $2P+1$ bins. These are the ℓth bin itself, and P frequency bins to its left and to its right. For this restricted data set, the PDF is

$$p(x|\epsilon) = \frac{1}{2\pi} \int_{-1/2}^{1/2} d\delta \int_0^{2\pi} d\phi \prod_{k=-P}^{P} \frac{1}{2\pi}$$
$$\times e^{-|x_{k+\ell} - \epsilon\omega(k+\delta)e^{i\phi}|^2/2}. \qquad (6.8)$$

One may now easily write down the likelihood function, and an optimal statistic, in the weak-signal limit, making use of Eqs. (5.1) and (5.2). It is easily verified that there are no terms of order ϵ. Writing the PDF in the form

$$p(x|\epsilon) = \frac{1}{2\pi} \int_{-1/2}^{1/2} d\delta \int_0^{2\pi} d\phi \, e^{W(\epsilon)} \qquad (6.9)$$

where

$$W(\epsilon) \equiv \sum_{k=-P}^{P} \left\{ -\frac{1}{2} \left| x_{k+\ell} - \epsilon\omega(k+\delta)e^{i\phi} \right|^2 - \ln 2\pi \right\},$$

and taking two derivatives with respect to ϵ, one has

$$p''(x|0) = \frac{1}{2\pi} \int_{-1/2}^{1/2} d\delta \int_0^{2\pi} d\phi \, e^{W(0)} \{ [W'(0)]^2 + W''(0) \}$$
$$= p(x|0) \int_{-1/2}^{1/2} d\delta \int_0^{2\pi} \frac{d\phi}{2\pi} \{ [W'(0)]^2 + W''(0) \}. \qquad (6.10)$$

We will do similar calculations later, in much less detail. The derivatives are easily evaluated:

$$W'(0) = \left. \frac{dW}{d\epsilon} \right|_{\epsilon=0} = \sum_{k=-P}^{P} \Re[x_{k+\ell}^* \omega(k+\delta)e^{i\phi}], \qquad (6.11)$$

$$W''(0) = \left. \frac{d^2W}{d\epsilon^2} \right|_{\epsilon=0} = -\sum_{k=-P}^{P} |\omega(k+\delta)|^2. \qquad (6.12)$$

The integral $d\phi$ of $W'(0)^2$ is evaluated by noting that for any complex numbers A and B

$$\int_0^{2\pi} \frac{d\phi}{2\pi} \Re(Ae^{i\phi})\Re(Be^{i\phi})$$

$$= |A||B| \int_0^{2\pi} \frac{d\phi}{2\pi} \cos(\phi - \psi_A)\cos(\phi - \psi_B)$$

$$= \frac{|A||B|}{2} \int_0^{2\pi} \frac{d\phi}{2\pi} [\cos(\psi_B - \psi_A) + \cos(2\phi - \psi_A - \psi_B)]$$

$$= \tfrac{1}{2}|A||B|\cos(\psi_B - \psi_A)$$

$$= \tfrac{1}{2}\Re(AB^*). \tag{6.13}$$

Making use of this, the inner integral in Eq. (6.10) gives

$$\int_0^{2\pi} \frac{d\phi}{2\pi}\{[W'(0)]^2 + W''(0)\}$$

$$= \frac{1}{2}\Re \sum_{r=-P}^{P} \sum_{r'=-P}^{P} x_{r+\ell}^* x_{r'+\ell}\omega(r+\delta)\omega^*(r'+\delta)$$

$$- \sum_{r=-P}^{P} |\omega(r+\delta)|^2.$$

Substituting this back into expression (6.10) for the second derivative of the PDF yields

$$\frac{p''(x|0)}{p(x|0)} = \frac{1}{2} \sum_{r,r'=-P}^{P} x_{r+\ell}^* M_{rr'} x_{r'+\ell} - \sum_{r=-P}^{P} M_{rr}. \tag{6.14}$$

Here, $M_{rr'}$ is a $(2P+1)$-dimensional square, symmetric, real, positive-definite matrix. Making use of the definition of ω in Eq. (6.6) gives

$$\mathbf{M} = M_{rr'} = \int_{-1/2}^{1/2} d\delta\, \omega(r+\delta)\omega^*(r'+\delta) \tag{6.15}$$

$$= (-1)^{r-r'} \int_{-1/2}^{1/2} d\delta\, j_0[\pi(r+\delta)]j_0[\pi(r'+\delta)].$$

Adopting the Einstein summation convention (repeated indices r and r' are summed from $-P$ to P) and substituting Eq. (6.14) into the weak-signal approximation (5.2) of the likelihood function, one obtains

$$\Lambda(x) - 1 = \frac{\epsilon^2}{2}\left(\frac{1}{2}x_{r+\ell}^* M_{rr'} x_{r'+\ell} - M_{rr}\right)$$

$$= \frac{\epsilon^2}{2}\left(\frac{1}{2}x_{r+\ell}^* x_{r'+\ell} - \delta_{rr'}\right)M_{rr'}. \tag{6.16}$$

In the absence of a signal, Eq. (5.3) shows that the mean value of $\Lambda - 1$ must vanish. This is clearly the case, since under our assumptions, in the absence of a signal, the mean value of $x_{r+\ell}^* x_{r'+\ell}$ is $2\delta_{rr'}$, where $\delta_{rr'}$ is the Kronecker delta.

We note that the formalism of this section can be trivially adapted to the case where the frequency of the signal lies in any desired range $\pm\Delta$ around the ℓth bin. The only change is that in Eq. (6.15) one makes the transformation

$$\int_{-1/2}^{1/2} d\delta \rightarrow \frac{1}{2\Delta}\int_{-\Delta}^{\Delta} d\delta. \tag{6.17}$$

In the limit $\Delta \rightarrow 0$, it is easy to see that $M_{00} = 1$ and all other components of $M_{rr'} = 0$. The results are then identical to the resolved-frequency case of Sec. III.

The results of this section can be summarized in a few lines. In Sec. III we studied the case where the signal frequency was exactly resolved. In that case, we found that the optimal statistic was the power in the corresponding bin. Thresholding on this statistic gave the lowest false dismissal probability for a given false alarm probability. In this section, after assuming that the signal frequency is uniformly distributed around bin ℓ, we have found that the optimal statistic (in the weak-signal case) is to threshold on the bilinear quantity (6.16). We can choose (from the value of P) how many of the data around the given bin to use. If $P = 0$ we recover the power statistic of Sec. III. If P is larger, then additional information from neighboring bins also gets added, and the test performs better. In the following sections, we will analyze the performance of this test, using the methods of Sec. IV to compare the optimal statistics for different values of P.

VII. PROPERTIES OF THE MATRIX M

Let us begin by exhibiting the $(2P+1)$-dimensional matrix $\mathbf{M}$, given by Eq. (6.15). It is easy to integrate Eq. (6.15) to get an exact expression for the matrix in terms of sine- and cosine-integral functions Si and Ci. On the diagonal (no summation convention on n)

$$M_{nn} = \frac{4}{\pi^2(4n^2-1)} + \frac{1}{\pi}\text{Si}[\pi(2n+1)] - \frac{1}{\pi}\text{Si}[\pi(2n-1)],$$

and off the diagonal

$$M_{mn} = \frac{\text{C}(2m-1) - \text{C}(2m+1) - \text{C}(2n-1) + \text{C}(2n+1)}{2\pi^2(n-m)},$$

where $\text{C}(x) \equiv \text{Ci}(\pi x) - \ln x$. In these equations, the range of the subscripts n, m is $-P, \dots, P$.

The ''central'' element of $\mathbf{M}$ has row and column numbers zero. The matrix extends away from this central element by an amount determined by the value of P. For example, if $P = 2$ one has the five-dimensional matrix

$$\mathbf{M} = 0.7737 \begin{bmatrix} 0.0181 & 0.0422 & -\mathbf{0.0169} & -0.0366 & -0.0169 \\ 0.0422 & 0.1017 & -\mathbf{0.0761} & -0.0761 & -0.0366 \\ -\mathbf{0.0169} & -\mathbf{0.0761} & \mathbf{1.0000} & -\mathbf{0.0761} & -\mathbf{0.0169} \\ -0.0366 & -0.0761 & -\mathbf{0.0761} & 0.1017 & 0.0422 \\ -0.0169 & -0.0366 & -\mathbf{0.0169} & 0.0422 & 0.0181 \end{bmatrix},$$

where the 0th row and column are highlighted, and we have taken out an overall factor of M_{00}. Note that this matrix is invariant under reflection about both diagonals, so it can be presented by listing just the $(P+1)$-dimensional block of elements with non-negative row and column numbers.

Because the matrix $\mathbf{M}$ is real and symmetric, it can be diagonalized by a similarity transformation

$$\mathbf{D} = \mathbf{O}^{-1}\mathbf{MO}, \tag{7.1}$$

where $\mathbf{O}$ is an orthogonal square matrix $\mathbf{O}^t = \mathbf{O}^{-1}$, and $\mathbf{D}$ is diagonal. Because $\mathbf{M}$ is positive, its eigenvalues are all real and positive. To six decimal places of accuracy, for the first few values of P, the eigenvalues of $\mathbf{M}$ are given by

$$\lambda_0 = 7.73695 \times 10^{-1} \quad \text{for} \quad P=0, \tag{7.2}$$

$$\lambda_0 = 7.82774 \times 10^{-1},$$

$$\lambda_1 = 1.37549 \times 10^{-1},$$

$$\lambda_2 = 1.07687 \times 10^{-2} \quad \text{for} \quad P=1, \tag{7.3}$$

$$\lambda_0 = 7.83230 \times 10^{-1},$$

$$\lambda_1 = 1.64608 \times 10^{-1},$$

$$\lambda_2 = 1.12358 \times 10^{-2},$$

$$\lambda_3 = 8.16859 \times 10^{-5},$$

$$\lambda_4 = 1.53779 \times 10^{-6} \quad \text{for} \quad P=2, \tag{7.4}$$

$$\lambda_0 = 7.83317 \times 10^{-1},$$

$$\lambda_1 = 1.76172 \times 10^{-1},$$

$$\lambda_2 = 1.13227 \times 10^{-2},$$

$$\lambda_3 = 1.20531 \times 10^{-4},$$

$$\lambda_4 = 1.91042 \times 10^{-6},$$

$$\lambda_5 = 3.03979 \times 10^{-9},$$

$$\lambda_6 = 2.72000 \times 10^{-11} \quad \text{for} \quad P=3. \tag{7.5}$$

We will see shortly that these eigenvalues determine the false alarm and false dismissal probabilities for the corresponding threshold statistics/tests.

The case analyzed in Sec. III, where the signal frequency is resolved, and a one-point test is used, corresponds to setting $P=0$ and having $\lambda_0=1$. This is the limit when the frequency band (6.17) over which the signal is distributed is very small, and centered around a bin frequency. In the opposite limit where the frequency band $\pm\Delta$ is large, the matrix $\mathbf{M}$ approaches something proportional to the identity matrix, with a large number of nearly equal eigenvalues.

VIII. PERFORMANCE OF THE OPTIMAL TEST FOR UNRESOLVED SIGNALS

The situation we are considering is defined by the PDF given in Eq. (6.7). We will suppose that we have implemented a search for sinusoidal signals (in the weak-signal limit) using the thresholding statistic defined by Eq. (6.16), for a particular value of P. We will call such a test the "$(2P+1)$-point test." For example, the "five point test" makes use of the data samples in the five bins nearest to some central bin, to determine if a sinusoidal signal is present within $\pm$ half a bin of that central bin.

Our goal is to determine the false alarm and false dismissal curves for different values of P. In this way, one can quantify the loss of performance that arises from throwing away the additional information coming from bins located away from the bin of interest.

Let us first calculate the false alarm probability for the $(2P+1)$-point test. This is easy because it only involves the probability distribution $p(x|0)$ (and its second derivative) for vanishing signal strength, which is an independent Gaussian in each frequency bin. We choose, as our optimal statistic, the quantity

$$\tau \equiv \mathbf{x}^\dagger \mathbf{Mx} \tag{8.1}$$

where $\mathbf{x}$ is a vector of (frequency space) data around the bin of interest. This differs from $\Lambda - 1$ by a data-independent constant term $\epsilon^2/2$, so it has the same level surfaces. Thus, for the three-point test, the optimal statistic to threshold on would be

$$\tau = [x^*_{\ell-1} \quad x^*_\ell \quad x^*_{\ell+1}]$$
$$\times \begin{bmatrix} 0.0787 & -0.0589 & -0.0589 \\ -0.0589 & 0.7737 & -0.0589 \\ -0.0589 & -0.0589 & 0.0787 \end{bmatrix} \begin{bmatrix} x_{\ell-1} \\ x_\ell \\ x_{\ell+1} \end{bmatrix}.$$

In the absence of signal, each of the x_j is an independent random Gaussian variable with zero mean and unit variance.

Thus, if U is a unitary matrix, the column vector of variables $\mathbf{Ux}$ are also independent random Gaussian variables with zero mean and unit variance. Since the orthogonal matrix $\mathbf{U}=\mathbf{O}^{-1}$ that diagonalizes $\mathbf{M}$ is unitary, the statistical properties of the optimal statistic τ (8.1) are the same as those of a random variable,

$$\tau = \sum_{r=0}^{2P} \lambda_r |z_r|^2,$$

where each z_r is an independent variable whose real and imaginary parts have independent Gaussian PDFS with zero mean and unit variance. Note that the PDF of $u=|z_r|^2$ is exponential with mean$=2$ and variance$=4$.

The PDF of the statistic τ is easily computed using generating functions. Suppose that τ is any random variable, and $p(\tau)d\tau$ is its probability density. We define the generating function $\bar{p}(\xi)$ to be the expected value of $e^{i\xi\tau}$:

$$\bar{p}(\xi) = \int_{-\infty}^{\infty} d\tau p(\tau) e^{i\xi\tau}.$$

This is basically the Fourier transform of the PDF. It makes it simple to compute the PDF of a random variable that is a sum of other random variables. Since

$$\tau = \sum_{r=0}^{2P} \lambda_r u_r$$

where each u_r is a real random variable with PDF

$$p(u)du = \begin{cases} 0 & \text{for} \quad u<0, \\ \frac{1}{2} e^{-u/2} du & \text{for} \quad u \geq 0, \end{cases}$$

the generating function for the PDF of τ (in the absence of a signal) is

$$\bar{p}(\xi) = \prod_{r=0}^{2P} \left[\int_0^{\infty} du_r \frac{1}{2} e^{-u_r/2} \right] e^{i\xi\tau}$$

$$= \prod_{r=0}^{2P} \left[\int_0^{\infty} du_r \frac{1}{2} e^{-u_r/2} \right] e^{i\xi(\lambda_0 u_0 + \cdots + \lambda_{2P} u_{2P})}$$

$$= \prod_{r=0}^{2P} \frac{1}{2} \int_0^{\infty} du_r e^{(i\xi\lambda_r - 1/2)u_r}$$

$$= \prod_{r=0}^{2P} (1 - 2i\xi\lambda_r)^{-1}. \tag{8.2}$$

This closed form for the generating function $\bar{p}$ makes it possible to find the probability distribution of the optimal statistic τ in the absence of a signal.

To determine p from $\bar{p}$, we invert the Fourier transform

$$p(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\xi \bar{p}(\xi) e^{-i\xi\tau}.$$

This gives

$$p(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\xi e^{-i\xi\tau} \prod_{r=0}^{2P} \frac{i}{2\lambda_r} \left(\xi + \frac{i}{2\lambda_r} \right)^{-1}. \tag{8.3}$$

The integral clearly vanishes for $\tau<0$, because the integrand has all of its poles in the complex ξ plane below the real-ξ axis. If $\tau<0$, the sign of the exponential term permits the contour of integration to be closed in the upper half-ξ-plane. Since there are then no poles contained inside the integration path, Cauchy's theorem implies that $p(\tau)=0$ for $\tau<0$.

To find a closed form for $p(\tau)$ when $\tau>0$, one must close the integration contour in the lower half-ξ-plane. The residue theorem then implies that $p(\tau)$ is a sum over the resides of the poles, which are located at $\xi=-i/2\lambda_r$. One obtains

$$p(\tau) = \sum_{r=0}^{2P} \left[\frac{e^{-\tau/2\lambda_r}}{2\lambda_r} \prod_{\substack{r'=0 \\ r' \neq r}}^{2P} \left(1 - \frac{\lambda_{r'}}{\lambda_r} \right)^{-1} \right]$$

$$= \sum_{r=0}^{2P} \frac{c_r}{2\lambda_r} e^{-\tau/2\lambda_r}. \tag{8.4}$$

Here, we have introduced the set of $2P+1$ weights $c_0, \ldots, c_{2P}$ defined by

$$c_r \equiv \prod_{\substack{r'=0 \\ r' \neq r}}^{2P} \left(1 - \frac{\lambda_{r'}}{\lambda_r} \right)^{-1}.$$

(Note: if $P=0$ then $c_0=1$.) These weights have several interesting properties. In particular,

$$\sum_{r=0}^{2P} c_r = 1, \text{ and} \tag{8.5}$$

$$\sum_{r=0}^{2P} c_r \lambda_r = \sum_{r=0}^{2P} \lambda_r = M_{rr}. \tag{8.6}$$

These weights simplify the notation in what follows.

The false alarm probability $\alpha(\mathcal{T})$ can now be obtained by straightforward integration:

$$\alpha(\mathcal{T}) = \int_{\mathcal{T}}^{\infty} d\tau p(\tau)$$

$$= \sum_{r=0}^{2P} c_r e^{-\mathcal{T}/2\lambda_r}. \tag{8.7}$$

It follows from Eq. (8.5) that $\alpha(0)=1$.

Our calculations assume that the eigenvalues λ_r are distinct (as is the case here). If m of them were equal then a polynomial of order $m-1$ in τ would appear on the right-hand side (RHS) of Eq. (8.4) and a polynomial of order $m-1$ in $\mathcal{T}$ would appear on the RHS of Eq. (8.7).

For concreteness, we give the numerical form of the false alarm functions for the first few values of P. The subscript on α denotes $2P+1$: the number of points used in the test.

$$\alpha_1(\mathcal{T}) = e^{-0.646249\mathcal{T}},$$

$$\alpha_3(\mathcal{T}) = e^{-0.63875\mathcal{T}+0.207097} - e^{-3.6351\mathcal{T}-1.46410}$$

$$+ e^{-46.430\mathcal{T}-6.73815},$$

$$\alpha_5(\mathcal{T}) = e^{-0.63840\mathcal{T}+0.250487} - e^{-3.0375\mathcal{T}-1.25272}$$

$$+ e^{-44.500\mathcal{T}-6.83620} - e^{-6121.0\mathcal{T}-21.6738}$$

$$+ e^{-325140.0\mathcal{T}-37.5716}.$$

The false dismissal probability β is a bit more challenging to calculate. However, for the weak-signal case of interest, it is still possible.

To find the false dismissal probability β we begin by writing the PDF for the weak signal case as

$$p(x|\epsilon) = p(x|0) + \tfrac{1}{2}\epsilon^2 p''(x|0)$$

$$= p(x|0)\left(1 + \frac{1}{2}\epsilon^2 \frac{p''(x|0)}{p(x|0)}\right)$$

$$= p(x|0)[1 + \tfrac{1}{2}\epsilon^2(\tfrac{1}{2}x^*_{r+\ell}M_{rr'}x_{r'+\ell} - M_{rr})]$$

$$= p(x|0)\left[1 + \frac{1}{2}\epsilon^2\left(\frac{\tau}{2} - M_{rr}\right)\right]$$

where τ is the optimal statistic (8.1). From this, we can immediately write an expression for the generating function of $p(\tau|\epsilon)$ to lowest order in ϵ,

$$\bar{p}(\xi|\epsilon) = \prod_{r=0}^{2P}\left[\int_0^\infty du_r \frac{1}{2}e^{-u_r/2}\right]e^{i\xi\tau}\left[1 + \frac{1}{2}\epsilon^2\left(\frac{\tau}{2} - M_{rr}\right)\right],$$

where as before $\tau = \lambda_0 u_0 + \cdots + \lambda_{2P}u_{2P}$. Since differentiating with respect to ξ brings down a factor of $i\tau$, one has

$$\bar{p}(\xi|\epsilon) = \left[1 + \frac{1}{2}\epsilon^2\left(\frac{1}{2i}\frac{d}{d\xi} - M_{rr}\right)\right]p(\xi|0). \qquad (8.8)$$

This relation is easily inverted to find a lowest-order formula for $p(\tau|\epsilon)$. We simply integrate the new term by parts:

$$\frac{1}{2\pi}\int_{-\infty}^\infty d\xi e^{-i\xi\tau}\frac{d\bar{p}}{d\xi}$$

$$= \frac{1}{2\pi}\int_{-\infty}^\infty \frac{d}{d\xi}[e^{-i\xi\tau}\bar{p}(\xi)] + i\tau e^{-i\xi\tau}\bar{p}(\xi)d\xi$$

$$= \tau p(\tau) = \tau p(\tau|0).$$

Thus we find a formula for the PDF of the optimal statistic τ in the small-ϵ limit:

$$p(\tau|\epsilon) = p(\tau|0)\left[1 + \frac{1}{2}\epsilon^2\left(\frac{\tau}{2} - M_{rr}\right)\right].$$

Since the PDFs on both sides are normalized, an important consequence of this is that the mean value of the test statistic in the absence of a signal is

$$M_{rr} = \sum_{r=0}^{2P}\lambda_r = \frac{1}{2}\int_0^\infty d\tau\, \tau\, p(\tau).$$

This is because the mean value of the likelihood function in the absence of a signal is unity. It is also easy to show that $\int_0^\infty \alpha(\mathcal{T})d\mathcal{T} = 2M_{rr}$.

From this it is straightforward to calculate the false dismissal probability

$$\beta(\mathcal{T}) = \int_0^\mathcal{T} d\tau\, p(\tau|\epsilon)$$

$$= \left(1 - \frac{\epsilon^2}{2}M_{rr}\right)\int_0^\mathcal{T} d\tau\, p(\tau|0) + \frac{1}{4}\epsilon^2\int_0^\mathcal{T} d\tau\, \tau\, p(\tau|0)$$

$$= \left(1 - \frac{\epsilon^2}{2}M_{rr}\right)[1 - \alpha(\mathcal{T})]$$

$$+ \frac{\epsilon^2}{4}\sum_{r=0}^{2P} c_r[2\lambda_r - (\mathcal{T} + 2\lambda_r)e^{-\mathcal{T}/2\lambda_r}]$$

$$= 1 - \alpha(\mathcal{T}) - \frac{\epsilon^2}{4}\left[(\mathcal{T} - 2M_{rr})\alpha(\mathcal{T}) + \int_\mathcal{T}^\infty d\tau\, \alpha(\tau)\right].$$

A bit of rearrangement gives us the weak-signal detection probability $\gamma(\mathcal{T}) = 1 - \alpha(\mathcal{T}) - \beta(\mathcal{T})$ as a function of the threshold:

$$\epsilon^{-2}\gamma(\mathcal{T}) = \frac{1}{2}\left(\frac{\mathcal{T}}{2} - M_{rr}\right)\alpha(\mathcal{T}) + \frac{1}{4}\int_\mathcal{T}^\infty d\tau\, \alpha(\tau)$$

$$= \frac{1}{2}\left(\frac{\mathcal{T}}{2} - M_{rr}\right)\alpha(\mathcal{T}) + \frac{1}{2}\sum_{r=0}^{2P} c_r\lambda_r e^{-\mathcal{T}/2\lambda_r}$$

$$= \frac{1}{2}\sum_{r'=0}^{2P}\left[\frac{\mathcal{T}}{2} - M_{rr} + \lambda_{r'}\right]c_{r'}e^{-\mathcal{T}/2\lambda_{r'}}. \qquad (8.9)$$

These formulas make it clear that $\gamma = 1 - \alpha - \beta$ vanishes as $\mathcal{T} \to 0$ and as $\mathcal{T} \to \infty$.

It is instructive to return briefly to the $P = 0$ (one-point) test. Equations (8.7) and (8.9) give false alarm and signal detection probabilities:

$$\alpha_1(\mathcal{T}) = e^{-\mathcal{T}/2\lambda_0}$$

and

$$\gamma_1(\mathcal{T}) = 1 - \alpha_1 - \beta_1 = \frac{\epsilon^2}{4}\mathcal{T}e^{-\mathcal{T}/2\lambda_0} = \epsilon^2\left[-\frac{\lambda_0}{2}\alpha_1\ln\alpha_1\right].$$

These should be compared with the resolved-frequency case, given in Eqs. (3.5) and (3.8). As expected, the formulas are identical if $\lambda_0 = 1$. However, for the unresolved-frequency case of this section, Eq. (7.2) gives $\lambda_0 \approx 0.773695$. Hence the signal detection probability at a given false alarm probability α is *lower* than in the resolved-frequency case:

For a resolved signal $\gamma = -\frac{1}{2}\epsilon^2 \alpha \ln \alpha$.

For an unresolved signal $\gamma = -0.3868\epsilon^2 \alpha \ln \alpha$.

Thus, for weak signals, the detection probability of a one-point test for unresolved signals is 77% the probability of detection of a 1-point test for resolved signals. This can also be seen by comparing the maxima of the one-point detection probabilities shown in Figs. 4 and 5.

For the first few values of P, the detection probability is given by

$$\epsilon^{-2}\gamma_1(\mathcal{T}) = \mathcal{T}e^{-1.38629 - 0.646250\mathcal{T}},$$

$$\epsilon^{-2}\gamma_3(\mathcal{T}) = (\mathcal{T}-0.29663)e^{-1.17920 - 0.638755\mathcal{T}}$$
$$- (\mathcal{T}-1.58708)e^{-2.85039 - 3.63507\mathcal{T}}$$
$$+ (\mathcal{T}-1.84064)e^{-8.12445 - 46.4309\mathcal{T}},$$

$$\epsilon^{-2}\gamma_5(\mathcal{T}) = (\mathcal{T}-0.35186)e^{-1.13581 - 0.638380\mathcal{T}}$$
$$- (\mathcal{T}-1.58910)e^{-2.63902 - 3.03752\mathcal{T}}$$
$$+ (\mathcal{T}-1.89585)e^{-8.22249 - 44.5006\mathcal{T}}$$
$$- (\mathcal{T}-1.91816)e^{-23.0601 - 6121.0\mathcal{T}}$$
$$+ (\mathcal{T}-1.91832)e^{-38.9580 - 325142.0\mathcal{T}},$$

where the subscript on γ is $2P+1$: the number of points used in the test. Figure 5 shows the detection probability and significance as a function of false alarm probability α for the one-, three-, five- and seven-point tests, for this case, where the signal frequency is uniformly distributed in the range $\delta \in \pm 1/2$ a bin. It is clear from this figure and from Table II that, while adding the additional information from the nearby frequency bins does improve the detection probability and significance slightly, the gain is relatively small. In practice, there is little to be gained from going beyond the three- or five-point tests, as can be seen by noting that the eigenvalues of $\mathbf{M}$ drop to small values very quickly with increasing P. This means that for sensible values of the threshold, the terms that they add to α and β have very small effects: the dominant terms are from the largest eigenvalues.

IX. INTERPRETATION OF RESULTS AS FREQUENCY-SPACE "INTERPOLATION"

In this section, the optimal statistic τ of the previous section is shown to have a simple intuitive interpretation: It is the total power contained in a continuous spectrum in the frequency range $f_{\ell-1/2} < f < f_{\ell+1/2}$. The continuous spectrum is obtained from the discrete spectrum x_j via frequency-space interpolation .

This frequency-space interpolation may be understood in terms of "zero padding," as follows.

Start with the low-resolution frequency-domain Fourier amplitudes x_k defined by Eq. (2.2). Here, "low resolution" indicates that the frequency spacing between successive bins is $1/T$.
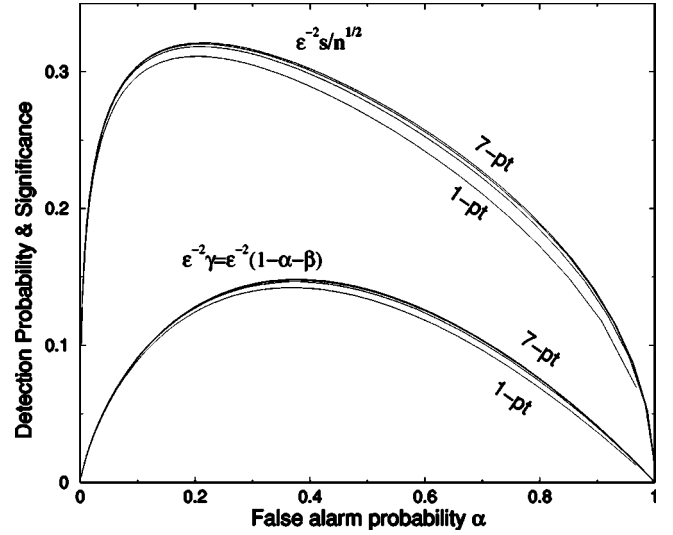

FIG. 5. Bottom four curves: The detection probability $\epsilon^{-2}\gamma = \epsilon^{-2}(1-\alpha-\beta)$ is plotted as a function of the false alarm probability α, for the one-, three-, five-, and seven-point optimal tests defined by Eq. (8.1), in the weak-signal limit. While using the additional information in the neighboring bins does improve the detection probability, the improvement is slight. Top four curves: The significance $\epsilon^{-2}s/\sqrt{n}$ is plotted for the same one-, three-, five-, and seven-point tests, in the weak-signal limit. The maxima of the eight curves are given in Table II.

Transform these into time-domain y_j for $j=0,\dots,N-1$.

Zero-pad the time-domain data to L times its original length N, by appending $(L-1)N$ zeros, for $j=N,\dots,NL-1$.

Now transform back into the frequency domain to get a higher-frequency-resolution set of Fourier amplitudes $\bar{x}_k$. Here "high resolution" indicates that the frequency spacing between successive bins is $1/LT$.

In the limit $L\to\infty$ this gives rise to a continuous spectrum $\bar{x}(f)$. The optimal statistic τ of the previous section is ex-

TABLE II. The maximum detection probability γ and significance s of the optimal $(2P+1)$-point peak detection tests, for $P=0,1,2,$ and 3. These correspond to the curves of Fig. 5. The top half of the table lists the maximum value of the detection probability $\gamma=1-\alpha-\beta$, and the values of the threshold $\mathcal{T}$ and false alarm probability α for which that maximum is obtained. The bottom half of the table lists the maximum value of the significance s, and the values of the threshold $\mathcal{T}$ and false alarm probability α for which that maximum is obtained.

	One point	Three point	Five point	Seven point
Max(γ/ϵ^2)	0.1424	0.1465	0.1477	0.1483
$\mathcal{T}$	1.548	1.863	1.918	1.942
α	0.3679	0.3739	0.3767	0.3775
Max($s/\epsilon^2\sqrt{n}$)	0.3113	0.3188	0.3204	0.3211
$\mathcal{T}$	2.467	2.773	2.821	2.840
α	0.2031	0.2093	0.2121	0.2135

actly the signal power contained in this continuous spectrum in the range from $f_{\ell-1/2} < f < f_{\ell+1/2}$. This quantity depends only on the Fourier amplitudes x_k because the zero padding has not added any information to the original data set.

To prove this assertion, we first derive a formula for the high-resolution DFT in terms of the lower-resolution one, following the procedure above. The Fourier amplitudes of the time-domain samples y_j are given by Eq. (2.2) as

$$x_k = \sum_{j=0}^{N-1} y_j e^{2\pi ijk/N} \quad \text{for } k = -N/2+1, \ldots, N/2.$$
(9.1)

The inverse relationship gives the time-domain samples in terms of the Fourier amplitudes as

$$y_k = \frac{1}{N} \sum_{j=-N/2+1}^{N/2} x_j e^{-2\pi ijk/N}, \quad \text{for } k = 0, \ldots, N-1.$$
(9.2)

Zero-pad these time-domain samples by appending $(L-1)N$ zeros, so that the total number of time-domain samples is now NL. Taking this back into the frequency domain gives the high-resolution Fourier amplitudes (for $k = -NL/2+1, \ldots, NL/2$)

$$\bar{x}_k = \sum_{j=0}^{NL-1} y_j e^{2\pi ijk/NL}$$

$$= \sum_{j=0}^{N-1} y_j e^{2\pi ijk/NL}$$

$$= \frac{1}{N} \sum_{j=0}^{N-1} \sum_{r=-N/2+1}^{N/2} x_r e^{-2\pi ijr/N} e^{2\pi ijk/NL}$$

$$= \sum_{r=-N/2+1}^{N/2} D_N\left(\frac{k}{L} - r\right) x_r.$$
(9.3)

In the third line, we have carried out the sum over j by using the geometric series in Eq. (6.3). The last line is the desired result giving the high-resolution Fourier amplitudes $\bar{x}$ in terms of the low-resolution x's. The Dirichlet kernel D_N in Eq. (6.5) is responsible for doing the interpolation.

The high-resolution spectrum has exactly as many degrees of freedom as the low-resolution spectrum, although it has L times as many frequency bins. This is because the amplitudes in the high-resolution spectrum are correlated with each other. The high-resolution spectrum also contains an exact duplicate of the low-resolution spectrum. Since D_N vanishes for nonzero integer arguments, and $D_N(0) = 1$, every Lth high-resolution bin contains the same value as one of the low-resolution bins: $\bar{x}_{Lr} = x_r$ for all integer r.

To finish proving the assertion, we calculate the average power in the high-resolution frequency bins $k = L(\ell-1/2), \ldots, L(\ell+1/2)-1$. These L high-resolution bins cover the frequency range from $f_{\ell-1/2}$ to $f_{\ell+1/2}$, which is $\pm 1/2$ a bin around the ℓth bin. Anticipating the final result, this quantity is denoted "τ." It is

$$\tau = \frac{1}{L} \sum_{k=0}^{L-1} |\bar{x}_{L\ell - L/2+k}|^2$$

$$= \frac{1}{L} \sum_{k=0}^{L-1} \left| \sum_{r=-N/2+1}^{N/2} D_N\left(\ell + \frac{k}{L} - r - \frac{1}{2}\right) x_r \right|^2.$$

Since $D_N(x)$ is peaked around $x = 0$, in the spirit of the previous section, this may be approximated as the sum over the $2P+1$ bins around the ℓth bin. Further justification can be found in Sec. X and in Fig. 6 below. This gives

$$\tau = \frac{1}{L} \sum_{k=0}^{L-1} \left| \sum_{r=-P}^{P} D_N\left(\frac{k}{L} - \frac{1}{2} - r\right) x_{\ell+r} \right|^2.$$
(9.4)

In the continuous limit, when the number of high-resolution frequency bins $L \to \infty$, the outer sum can be converted into an integral over $\delta = k/L - 1/2$, giving

$$\tau = \int_{-1/2}^{1/2} d\delta \left| \sum_{r=-P}^{P} D_N(\delta - r) x_{\ell+r} \right|^2$$

$$= \sum_{r,r'=-P}^{P} x_{\ell+r} S_{rr'} x_{\ell+r'}^*.$$

Here, the matrix $S_{rr'}$ is a $(2P+1)$-dimensional Hermitian matrix defined by

$$S_{rr'} = \int_{-1/2}^{1/2} d\delta \ D_N(\delta - r) D_N^*(\delta - r').$$
(9.5)

This equation should be compared to the definition of $L_{rr'}$ given in Eq. (6.15). Making the same large-N approximation as earlier gives

$$S_{rr'} = e^{i\pi(r-r')(1-1/N)} \int_{-1/2}^{1/2} d\delta j_0[\pi(\delta-r)] j_0[\pi(\delta-r')]$$

$$\approx e^{i\pi(r-r')} \int_{-1/2}^{1/2} d\delta j_0[\pi(\delta+r)] j_0[\pi(\delta+r')] = M_{rr'}.$$
(9.6)

Thus, the optimal statistic τ of the previous section is just the average power in a continuous interpolated spectrum within a frequency band of width $\pm 1/2$ a bin around f_ℓ.

X. WHY "WINDOWING" DOES NOT GIVE A BETTER TEST

Windowing is a well-known method for reducing the bias in a power spectrum, particularly for frequencies that are not resolved. It is natural to ask if this technique might provide a better test than the Neyman-Pearson test.

For large P (the number of bins used on either side of bin ℓ) the answer is clearly "no." In this case, the Neyman-Pearson test *is* (by its very definition) the optimal test. However, if P is very small, one might wonder if windowing could provide a better test, or if for large P windowing might provide a more efficient implementation of the optimal

Neyman-Pearson test. The reason is that in frequency space the amplitudes $|x_k|$ fall off proportionally to k^{-1} away from the peak. One might then wonder if windowing can "concentrate" more of the power close to the peak, to provide a better test when P has small values. As we shall show, the answer to the question is still "no" even when P is small.

"Windowing" is the process of multiplying the time-domain data y_j by a time-domain window function w_j, and then transforming the data into frequency space. Thus $y_j \rightarrow w_j y_j$ in Eq. (2.2). This is also referred to as "apodizing" or "tapering." Note: in addition, one may zero-pad the data set before taking it into the frequency domain. But, as described in Sec. IX the optimal test already effectively does this, in the limit of infinite zero padding.

Common choices of windowing functions are given such names as "Hamming," "Parzen," "Welch," and so on. These window functions are chosen for their properties: quickest sidelobe falloff, narrowest -3 db range, minimum spectral bias, and so on. As an example here, to explain why windowing the data first does not provide a better test, we take as a window function the cosine window

$$ w_j = \sqrt{\frac{2}{3}} \left[1 - \cos\frac{2\pi j}{N} \right]. \tag{10.1} $$

The situation for other windowing functions is similar.

The window function is normalized so that the total power in the spectrum is the same with or without the window. This is ensured by the condition (true for large N)

$$ \sum_{j=0}^{N-1} w_j^2 = N. \tag{10.2} $$

This condition ensures that, for stationary noise, the statistical properties of the noise in the frequency bins are the same with or without the windowing. Thus, for example, the expected power spectra of independent Gaussian-distributed time-domain samples (white Gaussian noise) are exactly the same for this window and for the rectangular window $w_j = 1$.

Shown in Fig. 6 are the spectra of sinusoidal signals (2.1) for the frequency bins near the peak. In the unwindowed case, a resolved signal ($\delta = 0$) has all its power in the ℓth bin: $|x_\ell| = \epsilon^2$. As the frequency shifts upward to $\delta = -0.5$, the magnitude of $|x_\ell|^2$ drops to $0.40\epsilon^2$. The adjacent $(\ell+1)$th bin also contains 40% of the energy. The remaining bins contain the other 20% of the energy, mostly in bins $\ell-1$ and $\ell+2$. The large magnitude of this ratio $1/0.40 = 2.5$ is one reason why rectangular windows are often undesirable: a peak at a resolved frequency can be as much as a factor of 2.5 times higher than a peak from a signal of the same amplitude at an unresolved frequency. In contrast, in the windowed case, the magnitude of $|x_\ell|^2 = 0.67\epsilon^2$ when $\delta = 0$ and only drops to $|x_\ell|^2 = 0.48\epsilon^2$ when $\delta = -0.5$. The ratio $0.67/0.48 = 1.38$ is much smaller; hence the cosine window produces a less biased power spectrum than the rectangular window.

But Fig. 6 also makes it clear why windowing does not result in a better test for sinusoidal signals buried in noise
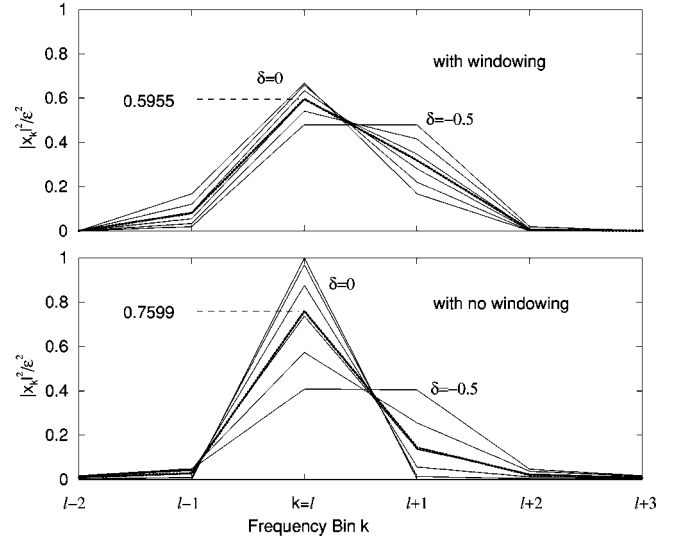

FIG. 6. The frequency-domain effects of windowing sinusoidal signals of amplitude ϵ are shown in the absence of noise. The bottom graph uses a rectangular window $w_i = 1$ (no windowing). The top graph uses the cosine window defined by Eq. (10.1). The solid curves show how the power $|x_k|^2$ is distributed bin by bin around the peak at $k = \ell$, for five different frequencies defined by $\delta = 0, -0.1, \ldots, -0.5$ in Eqs. (6.1), (6.2). The dotted line shows the average. Windowing greatly reduces the difference in $|x_\ell|^2$ between resolved frequencies ($\delta = 0$) and unresolved frequencies, so it reduces the *bias* in a spectrum. However, it also reduces the power in the peak substantially: the mean value is $0.60\epsilon^2$ with windowing compared to $0.76\epsilon^2$ without windowing. This means that windowing does not give a better test: at a given threshold $\mathcal{T}$ it yields a larger false dismissal probability.

than the Neyman-Pearson test, even for small P. The reason is that windowing "lowers the peak" for signals that are near a resolved frequency even more than it "raises the peak" for signals that are far from a resolved frequency. The dotted lines in Fig. 6 show the average power (averaged over the six values $\delta = 0, 0.1, \ldots, 0.5$). In the windowed case the average power in the peak is only $0.60\epsilon^2$ compared to $0.76\epsilon^2$ for the unwindowed case. This reduction in peak power results in a tremendous loss of significance for small ϵ, when the signals are buried in noise. For a given value of the threshold $\mathcal{T}$ (corresponding to a fixed false alarm probability), the windowed signal is far less likely to cross the threshold when a signal is present than the nonwindowed signal. Thus, it has a higher false dismissal probability than the Neyman-Pearson test.

Figure 6 also demonstrates that in the unwindowed case almost all of the power is within a few bins of the peak. Consequently, even small values of P will give a nearly optimal test. For example, even for the worst-case signal ($\delta = -0.5$) over 92% of the power is contained in just the range of bins from $\ell - 2$ to $\ell + 2$. Averaging over δ, these bins contain more than 96% of the signal power. When P is increased this rises rapidly: in the worse case ($\delta = -0.5$) for $P = 10$, the 21 bins around the peak contain more than 98% of the total power. There is effectively nothing to be gained by increasing P to larger values.

XI. OPTIMAL TESTS IN THE PRESENCE OF NON-GAUSSIAN NOISE

Section V showed how the weak-signal assumption of small ϵ permitted several useful simplifying approximations. One important simplification was that the optimal statistical test does not depend upon the amplitude ϵ.

This same weak-signal assumption also makes it possible to find the optimal statistical test for signals hidden in certain types of non-Gaussian noise as described, for example, in [13,14]. Consider the following generalization for the PDF (6.8):

$$p(x|\epsilon) = \int_{-1/2}^{1/2} d\delta \int_0^{2\pi} \frac{d\phi}{2\pi} \prod_{k=-P}^{P} \frac{1}{2\pi S_k}$$

$$\times e^{-g_k[|x_{k+\ell} - \epsilon\omega(k+\delta)e^{i\phi}|^2/2S_k]}. \qquad (11.1)$$

The Gaussian case treated in Sec. VI is a special case of this, for which $g_k(x) = x$ and $S_k = 1$. These types of non-Gaussian noise models, and the methods that are being used here (locally optimal tests), are discussed in more detail in [13,14], where they are used to construct optimal search techniques for stochastic background detection and for matched filtering.

This form of the PDF assumes that the noise in the different frequency bins is independent, but it allows each bin to have its own, different, arbitrary statistical distribution. For example, this can describe a very common situation, where the PDF has a central Gaussian region, plus a non-Gaussian tail. Typically there is a "knee" at some characteristic signal amplitude, where the slope of the distribution changes, or the non-Gaussian tail begins. Some preliminary work [15] has shown that it is straightforward to approximate these functions given a real data stream.

The functions g_k are not completely arbitrary. In order that Eq. (11.1) be properly normalized, one must have

$$\int_0^\infty du\, e^{-g_k(u)} = 1.$$

For any functional form of g, this can be satisfied by adding the correct constant term to g. We also require that g satisfy the additional normalization condition

$$\int_0^\infty du\; u\, e^{-g_k(u)} = 1,$$

which can always be satisfied by rescaling the argument of g. One then has

$$\int dx\, p(x|0) x_k^* x_r = 2\delta_{kr} S_k,$$

so the positive weights S_k can be interpreted as the mean-squared noise power in the kth frequency bin. This formula should be compared with Eq. (3.4). For example, one might have
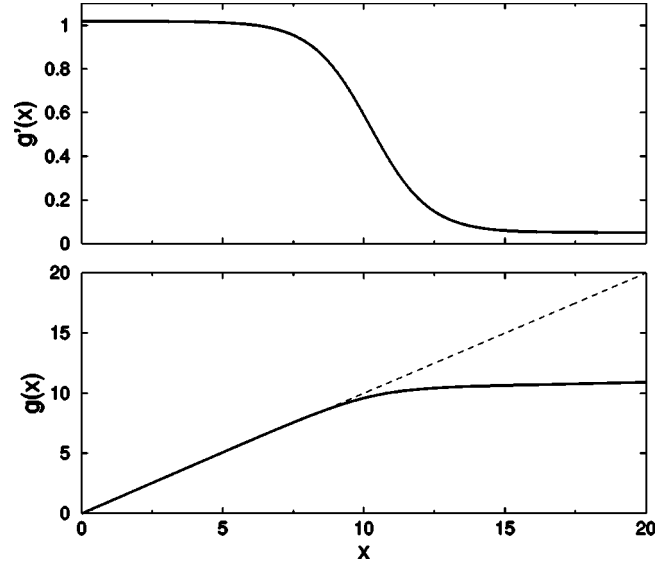

FIG. 7. An example of a function $g(x)$ corresponding to non-Gaussian statistical behavior, given by Eq. (11.2) with $\sigma^2 = 20$ and $p = 0.999$. Notice that in the central Gaussian region, $g'(x) \approx 1$, whereas $g'(x) \to \sigma^{-2}$ when the argument x is larger than $\approx \sigma^2/2$. The dotted line in the bottom graph shows (for comparison purposes) $g(x) = x$.

$$e^{-g(x)} = \kappa\left[p e^{-\kappa x} + \frac{1-p}{\sigma^2} e^{-\kappa x/\sigma^2} \right], \qquad (11.2)$$

where $\kappa = p + (1-p)\sigma^2$. Here we assume that p is positive and less than unity. The cases of most interest are when $1-p$ is very small, and σ^2 is large, so that $\kappa \approx 1$. Shown in Fig. 7 is a graph of $g(x)$ and $g'(x)$ for the case where $p = 0.999$ and $\sigma^2 = 20$. This corresponds to a case where 99.9% of the data is described by a Gaussian distribution with unit variance. The other 0.1% of the data samples are outlier points, described by a different Gaussian distribution with a variance of 20.

It is straightforward to derive the optimal peak-detection statistic in the weak signal limit, by proceeding exactly as in the Gaussian case of Sec. VI. We write

$$p(x|\epsilon) = \int_{-1/2}^{1/2} d\delta \int_0^{2\pi} \frac{d\phi}{2\pi} e^{W(\epsilon)}, \qquad (11.3)$$

where

$$W(\epsilon) = \sum_{k=-P}^{P} \left\{ -g_k\left(\frac{|x_{k+\ell} - \epsilon\omega(k+\delta)e^{i\phi}|^2}{2S_k} \right) - \ln 2\pi S_k \right\}. \qquad (11.4)$$

As before, it is easy to see that $p'(x|\epsilon)$ vanishes at $\epsilon = 0$. So the first nonvanishing derivative is

$$\frac{p''(x|0)}{p(x|0)} = \int_{-1/2}^{1/2} d\delta \int_0^{2\pi} \frac{d\phi}{2\pi} \{ [W'(0)]^2 + W''(0) \}.$$

$$\text{(11.5)}$$

The derivatives of W that appear are

$$W'(0) = \sum_{k=-P}^{P} \frac{1}{S_k} g'_k \left(\frac{|x_{k+\ell}|^2}{2S_k} \right) \Re[x^*_{k+\ell} \omega(k+\delta) e^{i\phi}]$$

and

$$W''(0) = -\sum_{k=-P}^{P} \left\{ \frac{|\omega(k+\delta)|^2}{S_k} g'_k \left(\frac{|x_{k+\ell}|^2}{2S_k} \right) + \left[\frac{\Re[x^*_{k+\ell} \omega(k+\delta) e^{i\phi}]}{S_k} \right]^2 g''_k \left(\frac{|x_{k+\ell}|^2}{2S_k} \right) \right\},$$

where g'_k and g''_k are the first and second derivatives of the function g_k with respect to its arguments. Using Eq. (6.13) to evaluate the integral over ϕ and Eq. (6.15) to evaluate the integral over δ gives

$$\frac{p''(x|0)}{p(x|0)} = \frac{1}{2} \sum_{k,r=-P}^{P} \frac{g'_k(|x_{k+\ell}|^2/2S_k) g'_r(|x_{r+\ell}|^2/2S_r)}{S_k S_r} x^*_{k+\ell} M_{kr} x_{r+\ell} - \frac{1}{2} \sum_{k=-P}^{P} \frac{g''_k(|x_{k+\ell}|^2/2S_k)}{S_k^2} M_{kk} |x_{k+\ell}|^2$$

$$- \sum_{k=-P}^{P} \frac{g'_k(|x_{k+\ell}|^2/2S_k)}{S_k} M_{kk}.$$

$$\text{(11.6)}$$

A good algebraic check is to verify that in the absence of a signal the mean value of this quantity vanishes.

Thus we arrive at the final result: the optimal weak-signal detection statistic in the non-Gaussian case. Leaving out the data-independent constant term, it is

$$\tau = \sum_{k,r=-P}^{P} \frac{g'_k(|x_{k+\ell}|^2/2S_k) g'_r(|x_{r+\ell}|^2/2S_r)}{S_k S_r} x^*_{k+\ell} M_{kr} x_{r+\ell} - \sum_{k=-P}^{P} \frac{g''_k(|x_{k+\ell}|^2/2S_k)}{S_k^2} M_{kk} |x_{k+\ell}|^2.$$

$$\text{(11.7)}$$

This reduces to the original expression (8.1) in the Gaussian case, where $g' = 1$ and $g'' = 0$. In the non-Gaussian case (refer to Fig. 7) the effect of the g' and g'' terms is to ''clip'' or ''truncate'' the effects of outlier points.

[1] A. Abramovici, W.E. Althouse, R.W.P. Drever, Y. Gürsel, S. Kawamura, F.J. Raab, D. Shoemaker, L. Sievers, R.E. Spero, K.S. Thorne, R.E. Vogt, R. Weiss, S.E. Whitcomb, and M.E. Zucker, Science **256**, 325 (1992).

[2] Barry C. Barish and Rainer Weiss, Phys. Today **52**, 44 (1999).

[3] C. Bradaschia *et al.*, Nucl. Instrum. Methods Phys. Res. A **289**, 518 (1990); also in *Gravitation 1990*, Proceedings of the Banff Summer Institute, Banff, Alberta, 1990, edited by R. Mann and P. Wesson (World Scientific, Singapore, 1991).

[4] K. Danzmann *et al.*, in *Gravitational Wave Experiments*, Proceedings of the Edoardo Amaldi Conference (World Scientific, Singapore, 1995), p. 100.

[5] K. Tsubono, in *Gravitational Wave Experiments* [4], p. 112.

[6] Instruments currently in the planning and proposal stage include the second-generation LIGO detector (LIGO-II), an advanced European detector, and a 3-km scale-up of the TAMA-300 detector.

[7] K. S. Thorne, in *300 Years of Gravitation*, edited by S. W. Hawking and W. Israel (Cambridge University Press, Cambridge, England, 1987), pp. 330–458.

[8] Patrick R. Brady, Teviet Creighton, Curt Cutler, and Bernard F. Schutz, Phys. Rev. D **57**, 2101 (1998).

[9] Piotr Jaranowski, Andrzej Królak, and Bernard F. Schutz, Phys. Rev. D **58**, 063001 (1998).

[10] Bernard F. Schutz and M. Alessandra Papa, ''End-to-end Algorithm for Hierarchical Area Searches for long-duration GW Sources for GEO 600,'' in Proceedings of January 1999 Moriond meeting, ''Gravitational Waves and Experimental Gravity,'' gr-qc/9905018.

[11] Patrick R. Brady and Teviet Creighton, Phys. Rev. D **61**, 082001 (2000).

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

[13] Bruce Allen, Jolien Creighton, Éanna É. Flanagan, and Joseph Romano, Phys. Rev. D **65**, 122002 (2002).

[14] Bruce Allen, Jolien Creighton, Éanna É. Flanagan, and Joseph Romano, ''Robust Statistics for Deterministic and Stochastic Gravitational Waves in Non-Gaussian Noise: Bayesian Analyses,'' gr-qc/0205015.

[15] Susan Scott and Bernard Whiting, work reported at the LIGO Scientific Collaboration Meetings in August 1999, and March 2000, and at the TAMA Meeting in October 1999.