Imaging Below the Diffraction Limit: A Statistical Analysis

Morteza Shahram and Peyman Milanfar, *Senior Member, IEEE*

***Abstract*—The present paper is concerned with the statistical analysis of the resolution limit in a so-called "diffraction-limited" imaging system. The canonical case study is that of incoherent imaging of two closely-spaced sources of possibly unequal brightness. The objective is to study how far beyond the classical Rayleigh limit of resolution one can reach at a given signal to noise ratio. The analysis uses tools from statistical detection and estimation theory. Specifically, we will derive explicit relationships between the minimum detectable distance between two closely-spaced point sources imaged incoherently at a given SNR. For completeness, asymptotic performance analysis for the estimation of the unknown parameters is carried out using the Cramér-Rao bound. To gain maximum intuition, the analysis is carried out in one dimension, but can be well extended to the two-dimensional case and to more practical models.**

***Index Terms*—Cramér-Rao bound, diffraction, estimation, hypothesis test, imaging, Rayleigh limit, resolution, super-resolution.**

I. Introduction

In incoherent optical imaging systems the image of an ideal point source is captured as a spatially extended pattern known as the point-spread function (PSF), as shown for the one-dimensional case in Fig. 1. In two dimensions, this function is the well-known *Airy* diffraction pattern [1]. When two closely-located point sources are measured through this kind of optical imaging system, the measured signal is the incoherent sum of the respective shifted point spread functions. According to the classical Rayleigh criterion, two incoherent point sources are "barely resolved" when the central peak of the diffraction pattern generated by one point source falls exactly on the first zero of the pattern generated by the second one. A more detailed and complete explanation of incoherent imaging and related topics can be found in [1] and [2].

The Rayleigh criterion for resolution in an imaging system is generally considered as an accurate estimate of limits in practice. But under certain conditions related to signal-to-noise ratio (SNR), resolution beyond the Rayleigh limit is indeed possible. This can be called the super-resolution limit [3]. Indeed, at sufficiently high sampling rates, and in the absence of noise, arbitrarily small details can be resolved.

To gain maximum intuition and perspective from the foregoing analysis, all discussion herein will be carried out in the one-dimensional case, which can later be extended to the two-dimensional case. To begin, let us assume that the original signal of interest is the sum of two impulse functions separated by a small distance d:[1]

$$\sqrt{\alpha}\delta\left(x-\frac{d}{2}\right)+\sqrt{\beta}\delta\left(x+\frac{d}{2}\right). \tag{1}$$

As mentioned before, the image will be the incoherent sum of two point spread functions, resulting from an imaging aperture (or slit in one-dimensional case, as seen in Fig. 2)

$$s(x;\alpha,\beta,d)=\alpha h\left(x-\frac{d}{2}\right)+\beta h\left(x-\frac{d}{2}\right) \tag{2}$$

where for our specific case of incoherent imaging $h(x)=\mathrm{sinc}^2(x)=[\sin(\pi x)/\pi x]^2$, but other PSF's can also be considered. Finally, the measured signal includes discretized samples corrupted with additive (readout) noise. Given samples at x_k $(k=1,\ldots,N)$ of the measured signal, we can rewrite the measurement model as

$$\begin{aligned} g(x_k) &= s(x_k;\alpha,\beta,d)+w(x_k) \\ &= \alpha h\left(x_k-\frac{d}{2}\right)+\beta h\left(x_k-\frac{d}{2}\right)+w(x_k) \end{aligned} \tag{3}$$

where $w(x_k)$ is assumed to be a zero-mean Gaussian white noise process with variance σ^2.

With the present definition, the Rayleigh limit corresponds to $d=1$ as can be seen in Figs. 1 and 2. This means that for values $d<1$, the two point sources are (in the classical Rayleigh sense)

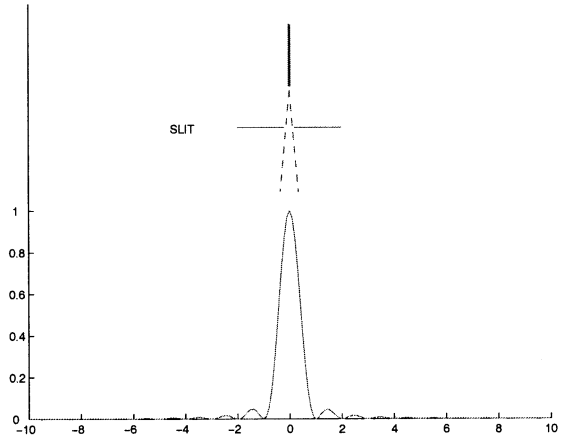

Fig. 1. Image of point source captured by diffraction-limited imaging.

Manuscript received March 3, 2003; revised November 3, 2003. This work was supported in part by NSF CAREER Grant CCR-9984246. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Thierry Blu.

The authors are with the Department of Electrical Engineering, University of California, Santa Cruz, CA 95064 USA (e-mail: shahram@ee.ucsc.edu; milanfar@ee.ucsc.edu).

Digital Object Identifier 10.1109/TIP.2004.826096

[1]From now on we refer to α and β as intensities and also we assume that $\alpha,\beta>0$. Also, note that this model (for now) assumes point sources symmetrically placed about the (known) origin. This model will be generalized later in the paper.

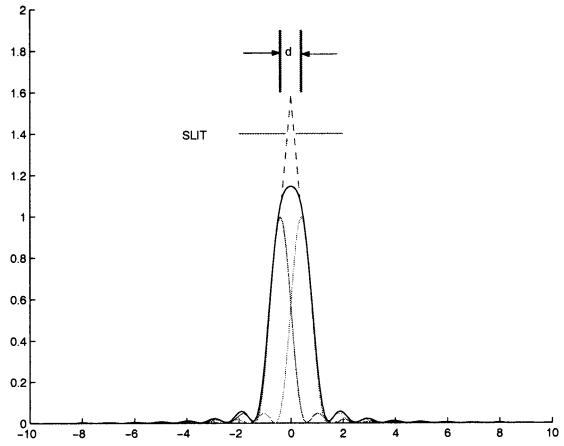

Fig. 2. Incoherent imaging of two closely located point sources.

"unresolvable." It is important to note that the Rayleigh criterion does not consider the presence of noise.

In the last forty years or so, there have been several attempts, and more recently surveys, of the problem of resolution from the statistical viewpoint. Of these, the most significant earliest works were done by Helstrom [4]–[6]. In particular, in [5] and [6], he derived lower bounds on the mean-squared error of unbiased estimators for the source positions, the distance between the sources, and the radiance values, using the Cramér-Rao inequality. In [5], he considered two separate situations. In the first, the problem of whether any signal was present or not was treated, whereas in the second, the question of whether one or two sources were present was treated. (This second scenario is, of course, what interests us in the present paper.) Helstrom described a geometrical optics field model of the problem involving a general radiance distribution and point spread function, for objects with arbitrary shape. To study the case of the circular aperture and point sources, he applied a complex and remarkable set of approximations and simplifications of the initial model. Also, he assumed that the distance between the point sources is known to the detector.

In [3] and [7], an approximate statistical theory was given to compute the required number of detected photons (similar to the notion of signal to noise ratio) for a certain desired resolution, and the value of achievable resolution by image restoration techniques was also investigated by numerical and iterative deconvolution. In these papers the definition of resolution was made as the separation of the two point sources that can be resolved through a deconvolution procedure. In [7], the analysis of the achievable resolution in deconvolved astronomical images was studied based on a criterion similar to Rayleigh's.

In [9] and [12] two-point resolution of imaging systems was studied using a model fitting theory where the probability of resolution was computed based on the structural change of the stationary points of the likelihood function. Also in [11] the Cramér-Rao lower bound formulation was used to study the limits to attainable precision of estimated distance between the two point sources. Assuming a Gaussian PSF, they determined a lower bound for the estimation error variance. Also, in [10], the reader can find a very comprehensive review of past and present approaches to the concept of resolution. In this paper, we also compute the Cramér-Rao (CR) lower bound in exact, closed form for two different cases. This analysis is in fact extendable to any point spread function.

Finally, an interesting, more recent paper [13] views the resolution problem from the information theory perspective. This line of thinking, again with simplifying approximations, is used to compute limits of resolution enhancement using Shannon's theorem of maximum transferable information via a noisy channel. The paper [13] considers the case of equally bright nearby point sources and derives an expression relating resolution (here defined as the inverse of the discernable distance between two equally bright point sources), logarithmically to the SNR.

The results of our paper extend, illuminate, and unify the earlier works in this field using more modern tools in statistical signal processing. Namely, we use locally optimal tests, which lead to more explicit, readily interpreted, and applicable results. In addition, we study various cases including unknown and/or unequal intensities, which have not been considered in their full complexity before.[2] The present results clarify, arguably for the first time, the specific effects of the relevant parameters on the definition of resolution, and its limits, as needed in practice.

In this paper we formulate the problem of two-point resolution in terms of statistical estimation/detection. Our approach is to precisely define a quantitative measure of resolution in statistical terms by addressing the following question: what is the minimum separation between two point sources (maximum attainable resolution limit) that is detectable at a given signal-to-noise ratio (SNR). In contrast to earlier definitions of resolution, there is little ambiguity in our proposed definition, and all parameters (PSF, noise variance, sampling rate, etc.) will be explicitly present in the formulation. Our earlier work on this problem was presented in [14], which essentially covers the material in Section IV-A of this paper.

The organization of the paper is as follows. Section II will explain and formulate our definition, and the corresponding statistical framework and models, in detail. In Section III, in order to use linear detection/estimation structures, we will discuss a signal approximation approach. In Section IV, we will present our statistical analysis for different cases of increasing generality. The asymptotic performance of the maximum likelihood estimate of the unknown parameters in terms of the Cramér-Rao lower bound will be discussed in Section V. Finally, some comments and conclusion will be presented in Section VI.

II. Statistical Analysis Framework

The question of whether one or two peaks are present in the measured signal can be formulated in statistical terms. Specifically, for the proposed model the equivalent question is whether the parameter d is equal to zero or not. If $d = 0$ then we only have one peak and if $d > 1$ then there are two resolved peaks according to the Rayleigh criterion. So the problem of interest revolves around values of d in the range of $0 \leq d < 1$. Therefore, we can define two hypotheses, which will form the basis of our statistical framework. Namely, let H_0 denote the null hy-

[2]Reference [9] considered the case of unequal intensities in a different framework.

pothesis that $d = 0$ (one peak present) and let H_1 denote the alternate hypothesis that $d > 0$ (two peaks present)

$$\begin{cases} H_0 : d = 0 & \text{One peak is present} \\ H_1 : d > 0 & \text{Two peaks are present} \end{cases} \tag{4}$$

Given discrete samples of the measured signal, we can rewrite the problem as

$$\begin{cases} H_0 : & \mathbf{g} = \mathbf{s}_0 + \mathbf{w} \\ H_1 : & \mathbf{g} = \mathbf{s} + \mathbf{w} \end{cases} \tag{5}$$

where

$$\begin{aligned} \mathbf{g} &= [g(x_1), \ldots, g(x_N)]^T, \\ \mathbf{w} &= [w(x_1), \ldots, w(x_N)]^T, \\ \mathbf{s} &= [s(x_1; \alpha, \beta, d), \ldots, s(x_N; \alpha, \beta, d)]^T, \\ \mathbf{s}_0 &= [s_0(x_1), \ldots, s_0(x_N)]^T, \end{aligned}$$

and

$$s(x_k; \alpha, \beta, d) = \alpha h\left(x_k - \frac{d}{2}\right) + \beta h\left(x_k + \frac{d}{2}\right) \tag{6}$$

$$s_0(x_k) = s(x_k; \alpha, \beta, d)|_{d=0} = (\alpha + \beta) h(x_k). \tag{7}$$

This is a problem of detecting a deterministic signal with unknown parameters (α, β, and d, in general). From (5), since the probability density function (PDF) under H_1 is not known exactly, it is not possible to design optimal detectors (in the Neyman-Pearson sense) by simply forming the likelihood ratio. The general structure of composite hypothesis testing is involved when unknown parameters appear in the PDF's [16, p. 248]. There are two major approaches for composite hypothesis testing. The first is to use explicit prior knowledge as to the likely values of parameters of interest and apply a Bayesian method to this detection problem. However, there is generally no such a priori information available. Alternately, the second approach, the Generalized Likelihood Ratio Test (GLRT) first computes maximum likelihood (ML) estimates of the unknown parameters, and then will use these estimated value to form the standard Neyman-Pearson (NP) detector. Our focus will be on GLRT-type methods because of less restrictive assumptions and easier computation and implementation; but most importantly, because uniformly most powerful (UMP) and locally most powerful (LMP) tests can be developed for the parameter range $0 \leq d < 1$.

To be a bit more specific, consider the case where it is known that $\alpha = \beta = 1$, with the parameter d unknown. The GLRT approach offers to decide H_1 if

$$L(\mathbf{g}) = \frac{\max_d p(\mathbf{g}, d, H_1)}{p(\mathbf{g}, H_0)} = \frac{p(\mathbf{g}, \hat{d}, H_1)}{p(\mathbf{g}, H_0)} > \gamma \tag{8}$$

where $\hat{d}$ denotes the ML estimate of d, and $p(\mathbf{g}, d, H_1)$ and $p(\mathbf{g}, H_0)$ are PDF's under H_1 and H_0, respectively. Assuming additive white Gaussian noise (AWGN) with variance σ^2 and $\hat{\mathbf{s}} = [s(x_1; 1, 1, \hat{d}), \ldots, s(x_N; 1, 1, \hat{d})]^T$ we will have:

$$\begin{aligned} L(\mathbf{g}) &= \frac{\frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{g} - \hat{\mathbf{s}}\|^2\right)}{\frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{g} - \mathbf{s}_0\|^2\right)} \\ &= \exp\left(-\frac{1}{2\sigma^2}(-\|\hat{\mathbf{s}}\|^2 + \|\mathbf{s}_0\|^2 + 2\mathbf{g}^T(\hat{\mathbf{s}} - \mathbf{s}_0))\right). \end{aligned}$$

Therefore, H_1 will be chosen if

$$-\|\hat{\mathbf{s}}\|^2 + 2\mathbf{g}^T(\hat{\mathbf{s}} - \mathbf{s}_0) > \gamma'. \tag{9}$$

Equivalently,

$$\begin{aligned} \sum_{k=1}^{N} &- \left[\alpha h\left(x_k - \frac{\hat{d}}{2}\right) + \beta h\left(x_k + \frac{\hat{d}}{2}\right)\right]^2 \\ &+ 2\left[\alpha h\left(x_k - \frac{\hat{d}}{2}\right) + \beta h\left(x_k + \frac{\hat{d}}{2}\right)\right. \\ &\left. - (\alpha + \beta) h(x_k)\right] g(x_k) > \gamma' \end{aligned} \tag{10}$$

where the ML estimate of d in the above involves solving the following minimization problem

$$\min_d \sum_{k=1}^{N} \left[\alpha h\left(x_k - \frac{d}{2}\right) + \beta h\left(x_k + \frac{d}{2}\right) - g(x_k)\right]^2 \Rightarrow \hat{d} \tag{11}$$

It should be clear from the above that this detection/estimation problem is highly nonlinear. However, since the range of interest are the values of $0 \leq d < 1$, these representing resolution beyond the Rayleigh limit, it is quite appropriate for the purposes of the our analysis to consider approximating the model of the signal around $d = 0$, and to apply locally optimal detectors. This is the approach we take.

III. (Quadratic) Model Approximation

Much of the complexity we encountered in the earlier formulation of the problem can be remedied by appealing to an approximation of the signal model. This approximate model is derived by expanding the signal about the small parameter values around $d = 0$. As alluded to earlier, this approximation is quite adequate in the sense that all the parameter values of interest for resolution beyond the Rayleigh diffraction limit are contained in the range [0, 1] anyway.

We consider the Taylor series expansion of $s(x_k; \alpha, \beta, d)$ around $d = 0$, with all other variables fixed.[3] More specifically,

$$\begin{aligned} s(x_k; \alpha, \beta, d) \approx (\alpha + \beta) h(x_k) &+ \frac{\beta - \alpha}{2} d h_1(x_k) \\ &+ \frac{\alpha + \beta}{8} d^2 h_2(x_k) \end{aligned} \tag{12}$$

where $h_1(\cdot)$ and $h_2(\cdot)$ denote the first and second order derivatives of $h(\cdot)$ and where for $h(x) = \text{sinc}^2(x)$

$$\begin{aligned} h_1(x_k) &= \left.\frac{\partial h(x)}{\partial x}\right|_{x = x_k} \\ &= \frac{2\sin(\pi x_k)(\sin(\pi x_k) - \pi x_k \cos(\pi x_k))}{\pi^2 x_k^3} \end{aligned} \tag{13}$$

$$\begin{aligned} h_2(x_k) &= \left.\frac{\partial^2 h(x)}{\partial x^2}\right|_{x = x_k} \\ &= \frac{\left(4\pi^2 x_k^2 - 3\right)\cos(2\pi x_k) - 4\pi x_k \sin(2\pi x_k) + 3}{2\pi^2 x_k^4}. \end{aligned} \tag{14}$$

[3]It is important here to note that this is an approximation about the *parameter* of interest d, and not the variable x; as such it therefore is a global approximation of the function.

In the above approximation, we elect to keep terms up to order 2 of the Taylor expansion. This gives a rather more accurate representation of the signal, and more importantly, if we only kept the first order term, then in the case $\alpha = \beta$, the first order term would simply vanish and *no* term in d would appear in the approximation. The reader can find a more detailed discussion on the accuracy of this approximation in Appendix A. The proposed approximation simplifies the hypothesis testing problem to essentially a linear detection problem (as we will see in the next section). The approximation is helpful in that we can carry out our analysis more simply. In addition, it leads to a general form of locally optimum detectors [16, p. 217] as will be discussed later.

Continuing with vector notation we have:

$$\mathbf{s} \approx (\alpha+\beta)\mathbf{h} + \frac{\beta-\alpha}{2} d\mathbf{h}_1 + \frac{\alpha+\beta}{8} d^2\mathbf{h}_2 \tag{15}$$

where

$$\begin{aligned}\mathbf{h} &= [h(x_1), \ldots, h(x_N)]^T \\ \mathbf{h}_1 &= [h_1(x_1), \ldots, h_1(x_N)]^T \\ \mathbf{h}_2 &= [h_2(x_1), \ldots, h_2(x_N)]^T.\end{aligned}$$

Writing in the form of hypotheses described earlier in (5)

$$\begin{cases} H_0: & \tilde{\mathbf{g}} = (\alpha+\beta)\mathbf{h} + \mathbf{w} \\ H_1: & \tilde{\mathbf{g}} = (\alpha+\beta)\mathbf{h} + \frac{\beta-\alpha}{2} d\mathbf{h}_1 + \frac{\alpha+\beta}{8} d^2\mathbf{h}_2 + \mathbf{w} \end{cases} \tag{16}$$

where we distinguish $\tilde{\mathbf{g}}$ from $\mathbf{g}$ due to the approximated model. According to this model, we define the measured signal-to-noise ratio (per sample) as follows:

$$\mathrm{SNR} = \frac{1}{N\sigma^2} \left\| (\alpha+\beta)\mathbf{h} + \frac{\beta-\alpha}{2} d\mathbf{h}_1 + \frac{\alpha+\beta}{8} d^2\mathbf{h}_2 \right\|^2. \tag{17}$$

For any symmetric PSF $(h(x))$ and in the case of above-Nyquist sampling, the following relations can be verified

$$\begin{aligned}\mathbf{h}^T\mathbf{h}_1 &= 0 \\ \mathbf{h}_2^T\mathbf{h}_1 &= 0 \\ \mathbf{h}^T\mathbf{h}_2 &= -\mathbf{h}_1^T\mathbf{h}_1.\end{aligned}$$

Therefore, we can rewrite (17) in the following form:

$$\begin{aligned}\mathrm{SNR} &= \frac{1}{N\sigma^2}\left[(\alpha+\beta)^2 E_0 + \left(\frac{\beta-\alpha}{2}\right)^2 d^2 E_1 \right. \\ &\quad \left. + \left(\frac{\alpha+\beta}{8}\right)^2 d^4 E_2 - \left(\frac{\alpha+\beta}{2}\right)^2 d^2 E_1\right] \\ &= \frac{1}{N\sigma^2}\left[(\alpha+\beta)^2 E_0 - \alpha\beta d^2 E_1 + \left(\frac{\alpha+\beta}{8}\right)^2 d^4 E_2\right]\end{aligned} \tag{18}$$

where we define

$$E_0 = \mathbf{h}^T\mathbf{h} = f_s \int_{-\infty}^{+\infty} h^2(x)\,dx \tag{19}$$

$$E_1 = \mathbf{h}_1^T\mathbf{h}_1 = f_s \int_{-\infty}^{+\infty} \left[\frac{\partial h(x)}{\partial x}\right]^2 dx \tag{20}$$

$$E_2 = \mathbf{h}_2^T\mathbf{h}_2 = f_s \int_{-\infty}^{+\infty} \left[\frac{\partial^2 h(x)}{\partial x^2}\right]^2 dx \tag{21}$$

as energy terms.[4]

[4]In above-Nyquist sampling, SNR is independent of N (and f_s) since energy terms are all proportional to f_s. See Appendix B for details and explicit computations of these energy terms for the case of $h(x) = \mathrm{sinc}^2(x)$.

IV. Detection Theory for the Approximated Model

In this section, we develop detection strategies for the hypothesis testing problem of interest based upon the approximated model. It is illuminating to study the various cases of interest in order. Our earlier assumptions were equal, known intensities, symmetrically located point sources about a given center, and the energy constraint $\alpha + \beta = 2$. In the interest of clarity and ease of exposition, we start with the case when all these assumptions hold. Then we will extend the discussion in order of increasing levels of generality by relaxing an assumption in each step. Namely, we will treat the problem for the following cases:

- the case of equal, known intensities $\alpha = \beta = 1$, with symmetrically located point sources;
- the case of unknown intensities but $\alpha + \beta = 2$, with symmetrically located point sources;
- the case of unknown intensities but $\alpha + \beta = 2$, asymmetrically[5] located point sources;
- the case of unknown intensities, asymmetrically located point sources.

By considering (16), we notice that when $\alpha+\beta = 2$ is known to the detector (the first three cases), $(\alpha+\beta)\mathbf{h}$ is a common known term in both hypotheses and it is independent from d. Therefore, we may simplify further

$$\begin{cases} H_0: & \mathbf{y} = \mathbf{w} \\ H_1: & \mathbf{y} = \frac{\beta-\alpha}{2} d\mathbf{h}_1 + \frac{\beta+\alpha}{8} d^2\mathbf{h}_2 + \mathbf{w} \end{cases} \tag{22}$$

where $\mathbf{y} = \tilde{\mathbf{g}} - (\alpha+\beta)\mathbf{h}$. As we began to describe earlier, when $\alpha = \beta$, the hypothesis test will be reduced to the case of detecting a known signal with unknown *positive* amplitude $(D = d^2)$. For this case, there exist well-known optimal detection strategies.

A. The Case of Equal Intensities, Symmetrically Located Point Sources

When $\alpha = \beta = 1$, (22) is reduced to

$$\begin{cases} H_0: & \mathbf{y} = \mathbf{w} \\ H_1: & \mathbf{y} = \frac{d^2}{4}\mathbf{h}_2 + \mathbf{w} \end{cases} \tag{23}$$

It is readily shown that given this model, the ML estimate for the parameter d^2 is given by

$$\widehat{d^2} = 4\left(\mathbf{h}_2^T\mathbf{h}_2\right)^{-1}\mathbf{h}_2^T\mathbf{y}. \tag{24}$$

Next, the test statistic resulting from the (generalized) Neyman-Pearson likelihood ratio is given by

$$T(\mathbf{y}) = \frac{1}{\sigma^2}\left(\mathbf{h}_2^T\mathbf{h}_2\right)^{-1}\left(\mathbf{h}_2^T\mathbf{y}\right)^2. \tag{25}$$

We note that the expression for the test-statistic is essentially an energy detector with the condition that the value of d^2 is in fact estimated from the data itself. The detector structure, due to our knowledge of the sign of the unknown distance parameter, is effectively producing a one-sided test, and hence is in fact a Uniformly Most Powerful (UMP) detector in the sense that it produces the highest detection probability for all values of the unknown parameter, and for a given false-alarm rate [16, p. 194]. Therefore, the above test-statistic can be simply replaced by

$$T'(\mathbf{y}) = \sqrt{T(\mathbf{y})} = \sqrt{\frac{1}{\sigma^2}\left(\mathbf{h}_2^T\mathbf{h}_2\right)^{-1}}\left(\mathbf{h}_2^T\mathbf{y}\right). \tag{26}$$

[5]Where point sources are located at $-d_1$ and $+d_2$ instead of $-(d/2)$ and $(d/2)$.

For any given data set $\mathbf{y}$, we decide H_1 if the statistic exceeds a specified threshold

$$T'(\mathbf{y}) > \gamma. \tag{27}$$

The choice of γ is motivated by the level of tolerable false alarm (or false-positive) in a given problem, but is typically kept very low.[6] The detection rate (P_d) and false-alarm rate (P_f) for this detector are related as [16, p. 254]

$$P_d = Q(Q^{-1}(P_f) - \sqrt{\eta}) \tag{28}$$

where

$$\eta = \frac{d^2}{4}\sqrt{\frac{E_2}{\sigma^2}} \tag{29}$$

and Q is the right-tail probability function for a standard Gaussian random variable (zero mean and unit variance); and Q^{-1} is the inverse of this function [16, p. 20]. A particularly intriguing and useful relationship is the behavior of the smallest peak separation d, which can be detected with very high probability (say 0.99), and very low false alarm rate (say 10^{-6}) at a given SNR. According to (18), (28), and (29), the relation between $d_{\min}$ and required SNR can be made explicit

$$\mathrm{SNR} = (Q^{-1}(P_f) - Q^{-1}(P_d))^2 \frac{64E_0 - 16d^2E_1 + d^4E_2}{Nd^4E_2} \tag{30}$$

$$= \frac{1}{N}(Q^{-1}(P_f) - Q^{-1}(P_d))^2 \times \left(\frac{64E_0}{E_2}\frac{1}{d^4} - \frac{16E_1}{E_2}\frac{1}{d^2} + 1\right). \tag{31}$$

The above expression gives an implicit relation between the smallest detectable distance between the two (equal intensity) sources, at the particular SNR. As an example, for $h(x) = \mathrm{sinc}^2(x)$ and for the specified choice of $P_d = 0.99$ and $P_f = 10^{-6}$, if we collect N equally spaced samples at $\{x_k\}$ within the interval $[-10, 10]$, at the Nyquist rate, we have

$$\mathrm{SNR} = 50.12\frac{\frac{140}{\pi^4} - \frac{14}{\pi^2}d^2 + d^4}{Nd^4} = \frac{72.04 - 71.1d^2 + 50.12d^4}{Nd^4}$$

A plot of this function is shown in Fig. 3. It is worth noting that in (31), the term involving d^{-4} dominates for small d. Therefore, a reasonably informative (but approximate) way to write SNR is

$$\mathrm{SNR} \approx \frac{1}{N}(Q^{-1}(P_f) - Q^{-1}(P_d))^2\frac{E_0}{E_2}\frac{1}{d^4} = \frac{c}{Nd^4} \tag{32}$$

where the coefficient c is a function only of the selected P_f and P_d. It is worth noting that for any sampling rate higher than the Nyquist rate, we can rewrite c in (32) as follows:

$$c = 64(Q^{-1}(P_f) - Q^{-1}(P_d))^2\frac{\int_{-\infty}^{+\infty} h^2(x)\,dx}{\int_{-\infty}^{+\infty}\left[\frac{\partial^2 h(x)}{\partial x^2}\right]^2 dx} \tag{33}$$

[6]In [9] and [12] a similar criterion (in a different framework) has been proposed, where they applied a sign test (i.e., a fixed threshold) to decide if there is one or two point sources present. This approach gives a detector with a fixed false alarm rate.

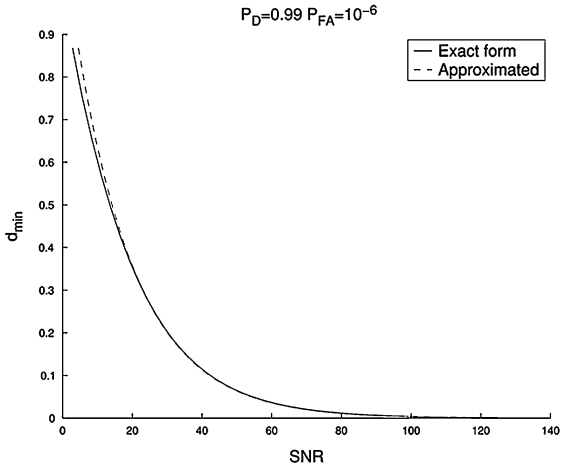

Fig. 3. Minimum detectable d as a function of SNR (in dB) at the Nyquist rate (exact and approximate).

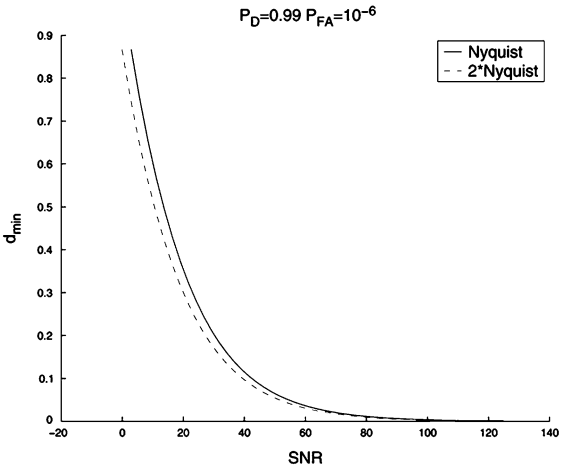

Fig. 4. Minimum detectable d versus SNR (in dB) at Nyquist rate, and at twice Nyquist rate.

A plot of the approximate expression in (32) is also shown in Fig. 3 to be compared against the exact expression (31). The above relation (32) is a neat and rather intuitive power law that one can use to, for instance, understand the required SNR to achieve a particular resolution level of interest below the diffraction limit. Fig. 4 shows the curves defined by (30) for different sampling rates; namely Nyquist rate and twice Nyquist. As one would expect, the minimum detectable d becomes smaller as the number of samples increases, but it does not do so at a very fast rate because of the proportionality between SNR and the sampling rate.[7]

B. The Case of Unknown α and β, Symmetrically Located Point Sources

In this section we discuss a more general case where neither the intensities α and β, nor the distance d, are known.[8] Equation

[7]Similar analysis for the two-dimensional extension of this problem is presented in [22].

[8]But we assume that $\alpha + \beta = 2$ is known to the detector.

(22) leads to a detection problem defined in terms of a linear model over the parameter set $\boldsymbol{\theta}$ defined as follows:

$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \mathbf{w} \tag{34}$$

$$\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2] \tag{35}$$

$$\theta = \begin{bmatrix} d(\alpha - \beta) \\ \frac{d^2}{4} \end{bmatrix} \tag{36}$$

where we note that the matrix H has orthogonal columns. Specifically, the detection problem is now posed as

$$\begin{cases} H_0: & \mathbf{A}\boldsymbol{\theta} = \mathbf{b} \\ H_1: & \mathbf{A}\boldsymbol{\theta} \neq \mathbf{b} \end{cases} \tag{37}$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{38}$$

The GLRT for this problem is given by ([16], p. 274):

$$T(\mathbf{y}) = \frac{1}{\sigma^2}\hat{\boldsymbol{\theta}}^T \mathbf{A}^T [\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}\mathbf{A}\hat{\boldsymbol{\theta}} \tag{39}$$

$$= \frac{1}{\sigma^2}\left(\frac{\left(\mathbf{h}_1^T\mathbf{y}\right)^2}{E_1} + \frac{\left(\mathbf{h}_2^T\mathbf{y}\right)^2}{E_2}\right) \tag{40}$$

where

$$\hat{\boldsymbol{\theta}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}\mathbf{y}. \tag{41}$$

The performance of this detector is characterized by

$$P_f = Q_{\chi_2^2}(\gamma) \tag{42}$$

$$P_d = Q_{\chi_2'^2(\lambda)}(\gamma) \tag{43}$$

$$\lambda = \frac{1}{\sigma^2}\boldsymbol{\theta}^T \mathbf{A}^T [\mathbf{A}(\mathbf{H}^T\mathbf{H})^{-1}\mathbf{A}^T]^{-1}\mathbf{A}\boldsymbol{\theta} \tag{44}$$

$$= \frac{1}{\sigma^2}\left(\left(\frac{\alpha - \beta}{2}\right)^2 d^2 E_1 + \frac{1}{16}d^4 E_2\right) \tag{45}$$

where $Q_{\chi_2^2}$ is the right tail probability for a Central Chi-Squared PDF with 2 degrees of freedom, and $Q_{\chi_2'^2(\lambda)}$ is the right tail probability for a noncentral Chi-Squared PDF with 2 degrees of freedom and noncentrality parameter λ. In order to perform the same analysis as Section 4.1 (i.e., $d_{\min}$ versus SNR curve), we start by computing the required λ from the above expressions, based on the fixed values of P_d and P_f. Then, using the relation (18), we will have

$$\mathrm{SNR} = \frac{\lambda(P_f, P_d)}{N}\frac{64E_0 - 16\alpha\beta d^2 E_1 + d^4 E_2}{4(\alpha - \beta)^2 d^2 E_1 + d^4 E_2} \tag{46}$$

where $\lambda(P_f, P_d)$ represents the required value of noncentrality parameter as a function of the desired P_f and P_d. For instance, for the case of $h(x) = \mathrm{sinc}^2(x)$, with $P_d = 0.99$ and $P_f = 10^{-6}$ we have

$$\mathrm{SNR} = \frac{56.29}{N}\frac{\frac{140}{\pi^4} - \frac{14}{\pi^2}\alpha\beta d^2 + d^4}{\frac{7}{2\pi^2}(\alpha - \beta)^2 d^2 + d^4}. \tag{47}$$

It is useful to compare the performance of this detector (in terms of minimum detectable d) against the "best" case where the parameters d, α and β are actually known. In fact, a comparison in Fig. 5 demonstrates that, happily (and perhaps rather unexpectedly), the curves are very close, implying that the performance of GLRT is very close to the optimal detector for which all parameters are known.

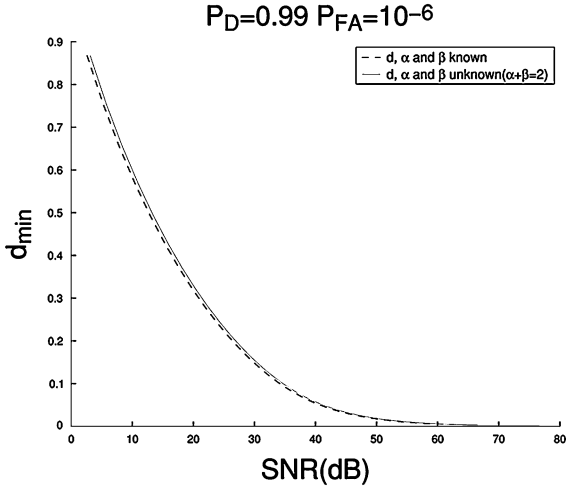

Fig. 5. $d_{\min}$ versus SNR (dB) for $\alpha = 1.2$ and $\beta = 0.8$.

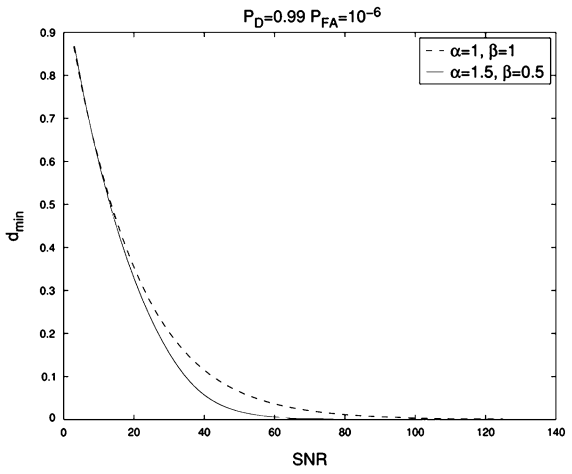

Fig. 6. GLRT for $\alpha \neq \beta$ and the case $\alpha = \beta$, symmetric sources; $d_{\min}$ versus SNR(dB).

An interesting observation arises from a comparison of the minimum detectable d for the cases $\alpha = \beta$ and $\alpha \neq \beta$, shown in Fig. 6. It is seen that unequal α and β yield better detection. That is, for a fixed d, the required SNR for resolving two closely-spaced unequally bright point sources is *smaller* than the SNR required to resolve two *equally* spaced sources. This result seems counter-intuitive. Yet, the reason behind it is somewhat clear in hindsight. Equal α and β produce a perfectly symmetric signal (without noise) and therefore result in redundancy in the measured signal content. With unequal α and β, an anti-symmetric part is added to the signal information and better decision is made possible. This phenomenon is a result of by the assumption of symmetry of point sources around the origin ($x = 0$). If the center of the point sources is not known, the results can be different, as we will explain in the next section.

C. The Case of Unknown Intensities But $\alpha + \beta = 2$ with Asymmetrically Located Point Sources

With the earlier machinery in place, in this section, we study the case where the point sources are not located symmetrically

around the origin ($x = 0$). We consider the following model for this case:

$$\begin{aligned} g(x_k) &= s(x_k; \alpha, \beta, d_1, d_2) + w(x_k) \\ &= \alpha h(x_k - d_1) + \beta h(x_k + d_2) + w(x_k) \end{aligned} \quad (48)$$

where d_1 and d_2 are unknown and $d = d_1 + d_2$ is the distance between the point sources. The Taylor expansion for the signal term in (48) around $(d_1, d_2) = (0, 0)$ is given by

$$\begin{aligned} s(x_k; \alpha, \beta, d_1, d_2) = (\alpha + \beta)h(x_k) + (-\alpha d_1 + \beta d_2)h_1(x_k) \\ + \frac{\alpha d_1^2 + \beta d_2^2}{2} h_2(x_k). \end{aligned} \quad (49)$$

Here we consider the general case of unknown α and β but $\alpha + \beta = 2$ is known to the detector. However, we assume that the test for determining whether one peak is present or two peaks are present is performed at *some* point located between the two point sources. Hence, the hypothesis test can be expressed as

$$\begin{cases} H_0: & [d_1 \;\; d_2] = [0 \;\; 0] \\ H_1: & [d_1 \;\; d_2] \neq [0 \;\; 0] \end{cases} \quad (50)$$

or equivalently (see (51) at the bottom of the page). By removing the known common term $(\alpha + \beta)h(x_k)$, we have the following linear model:

$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta}_a + \mathbf{w}$$

where

$$\begin{aligned} \mathbf{H} &= [\mathbf{h}_1, \mathbf{h}_2] \\ \boldsymbol{\theta}_a &= \begin{bmatrix} -\alpha d_1 + \beta d_2 \\ \frac{\alpha d_1^2 + \beta d_2^2}{2} \end{bmatrix} \end{aligned} \quad (52)$$

and where the subscript "a" an θ_a is denoting the asymmetric case, to be distinguished from (36). Then, the corresponding hypotheses are given by

$$\begin{cases} H_0: & \mathbf{A}\boldsymbol{\theta}_a = \mathbf{b} \\ H_1: & \mathbf{A}\boldsymbol{\theta}_a \neq \mathbf{b} \end{cases} \quad (53)$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

just as in Section IV-B. The GLRT for (53) will be

$$T(\mathbf{y}) = \frac{1}{\sigma^2} \left(\frac{\left(\mathbf{h}_1^T \mathbf{y}\right)^2}{E_1} + \frac{\left(\mathbf{h}_2^T \mathbf{y}\right)^2}{E_2} \right). \quad (54)$$

From (54), the performance of this detector is characterized by

$$\begin{aligned} P_f &= Q_{\chi_2^2}(\gamma) \\ P_d &= Q_{\chi_2'^2(\lambda)}(\gamma) \\ \lambda &= \frac{1}{\sigma^2} \left((-\alpha d_1 + \beta d_2)^2 E_1 \right. \\ &\qquad \left. + \left(\frac{\alpha d_1^2 + \beta d_2^2}{2} \right)^2 E_2 \right). \end{aligned} \quad (55)$$

Now, to obtain the relation between SNR and (d_1, d_2), we first need to compute the SNR for the model of (48), which is given by

$$\begin{aligned} \mathrm{SNR} = \frac{1}{N\sigma^2} \left[(\alpha + \beta)^2 E_0 - \alpha\beta(d_1 + d_2)^2 E_1 \right. \\ \left. + \left(\frac{\alpha d_1^2 + \beta d_2^2}{2} \right)^2 E_2 \right]. \end{aligned} \quad (56)$$

The value of σ^2 in (55) can be obtained for the desired P_d and P_f. By substituting this value in (56) we will have (57), shown at the bottom of the page. In order to present the results in this case, let us assume that[9] $\alpha d_1 \approx \beta d_2$ (i.e., we perform the test at a point which is closer to the stronger peak.). It can be easily shown that the value of λ in (55) is maximized for the case of $\alpha = \beta$. This shows that when $\alpha d_1 \approx \beta d_2$, the performance for the case of equal intensities is better than the performance of the case with unequal intensities. Fig. 7 confirms this result by showing the curves for $d_{\min}$ versus SNR for two cases: equal intensities and unequal intensities (we assume $h(x) = \mathrm{sinc}^2(x)$). By comparing this results and that of the previous section, we conclude that the assumption of symmetrically located point sources around the test point plays a very important role in the performance of the detector. Also, it is worth mentioning that with the assumption of $\alpha d_1 \approx \beta d_2$, we can approximate (57) for the range of small d_1 and d_2 in the following informative ways:

$$\begin{aligned} \mathrm{SNR} &= \frac{\lambda(P_f, P_d)}{N} \frac{4(\alpha + \beta)^2}{\left(\alpha d_1^2 + \beta d_2^2\right)^2} \frac{E_0}{E_2} = \frac{\lambda(P_f, P_d)}{N} \frac{4}{d_1^2 d_2^2} \frac{E_0}{E_2} \\ &= \frac{\lambda(P_f, P_d)}{N} \frac{4(\alpha + \beta)^4}{\alpha^2 \beta^2 d^4} \frac{E_0}{E_2} \end{aligned} \quad (58)$$

[9]See Appendix C for a justification.

$$\begin{cases} H_0: & \tilde{g}(x_k) = (\alpha + \beta)h(x_k) + w(x_k) \\ H_1: & \tilde{g}(x_k) = (\alpha + \beta)h(x_k) + (-\alpha d_1 + \beta d_2)h_1(x_k) + \frac{\alpha d_1^2 + \beta d_2^2}{2} h_2(x_k) + w(x_k) \end{cases} \quad (51)$$

$$\mathrm{SNR} = \frac{\lambda(P_f, P_d)}{N} \frac{(\alpha + \beta)^2 E_0 - \alpha\beta(d_1 + d_2)^2 E_1 + \left(\frac{\alpha d_1^2 + \beta d_2^2}{2} \right)^2 E_2}{(-\alpha d_1 + \beta d_2)^2 E_1 + \left(\frac{\alpha d_1^2 + \beta d_2^2}{2} \right)^2 E_2} \quad (57)$$

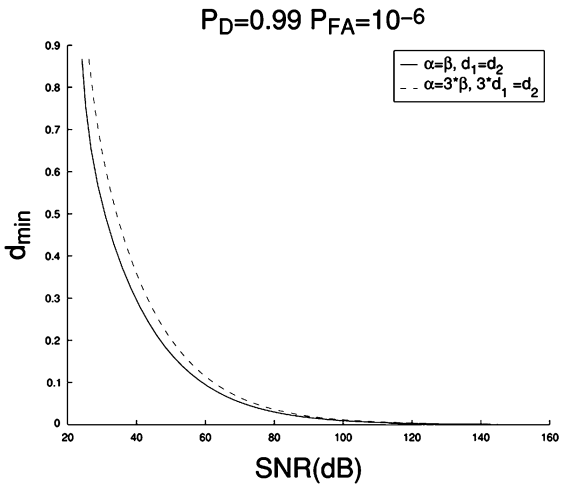

Fig. 7. $d_{\min}$ versus SNR(dB); $d = d_1 + d_2$ and $\alpha d_1 = \beta d_2$; equal intensities and unequal intensities.

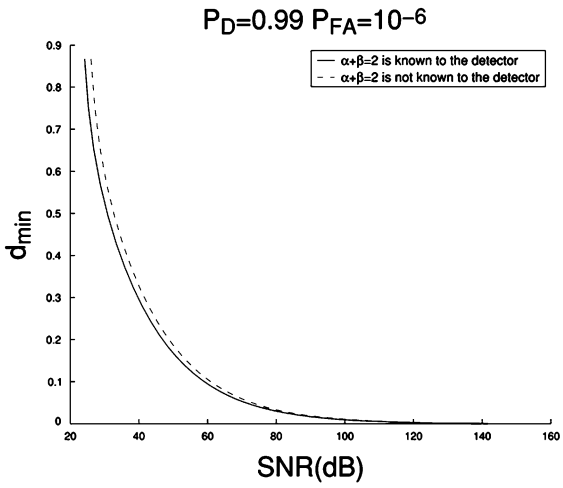

Fig. 8. $d_{\min}$ versus SNR(dB); $d = d_1 + d_2$ and $\alpha d_1 = \beta d_2$ detectors with and without the assumption of $\alpha + \beta = 2$.

D. The Case of Unknown Intensities, Asymmetrically Located Point Sources

Here, we analyze the most general case in which we assume that the energy of point sources $(\alpha + \beta)$ is unknown to the detector, as well as the individual α, β, d_1, and d_2. Recalling (51), we can set up another linear model as follows:

$$\tilde{\mathbf{g}} = \mathbf{H}_u \boldsymbol{\theta}_u + \mathbf{w}$$

where

$$\begin{aligned} \mathbf{H}_u &= [\mathbf{h}, \mathbf{h}_1, \mathbf{h}_2] \\ \boldsymbol{\theta}_u &= \begin{bmatrix} \alpha + \beta \\ -\alpha d_1 + \beta d_2 \\ \frac{\alpha d_1^2 + \beta d_2^2}{2} \end{bmatrix} \end{aligned} \tag{59}$$

and the subscript "u" denotes the completely unknown parameters. The above setup leads to the following hypothesis test:

$$\begin{cases} H_0: & \mathbf{A}_u \boldsymbol{\theta}_u = \mathbf{b} \\ H_1: & \mathbf{A}_u \boldsymbol{\theta}_u \neq \mathbf{b} \end{cases} \tag{60}$$

where

$$\mathbf{A}_u = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The GLRT for (60) will be

$$T(\tilde{\mathbf{g}}) = \frac{1}{\sigma^2} \left(\frac{\left(\mathbf{h}_1^T \tilde{\mathbf{g}}\right)^2}{E_1} + \frac{\left(E_2 \mathbf{h}^T \tilde{\mathbf{g}} + E_0 \mathbf{h}_2^T \tilde{\mathbf{g}}\right)^2}{E_0 \left(E_0 E_2 - E_1^2\right)} \right). \tag{61}$$

The performance of this detector is given by[10]

$$\begin{aligned} P_f &= Q_{\chi_2^2}(\gamma) \\ P_d &= Q_{\chi_2'^2(\lambda)}(\gamma) \\ \lambda &= \frac{1}{\sigma^2} \left((-\alpha d_1 + \beta d_2)^2 E_1 \right. \\ &\quad \left. + \left(\frac{\alpha d_1^2 + \beta d_2^2}{2} \right)^2 \left(E_2 - \frac{E_1^2}{E_0} \right) \right). \end{aligned} \tag{62}$$

Consequently, the relation between (d_1, d_2) and SNR is given by (63) as shown at the bottom of the page. By comparing (57) and (63), it can be readily shown that because of the negative term $-(E_1^2/E_0)$, the detector without the knowledge of $\alpha + \beta$ performs more poorly than the detector which knows $\alpha + \beta = 2$. Fig. 8 displays the performance of these two different detectors in terms of the minimum detectable d versus SNR for the case of $h(x) = \operatorname{sinc}^2(x)$.

V. The Cramér-Rao Lower Bound on Estimation of the Unknown Parameters

In the interest of completeness, in this section we present results on the estimation of the unknown parameters of the model. In particular, we study the asymptotic performance of ML estimate of the unknown parameters, using the Cramér-Rao lower bound (CRLB). CRLB [15, p. 27] is a covariance inequality bound which treats the parameters as unknown deterministic quantities and provides a local bound on the mean square error (MSE) of their estimate. Being able to compute a lower bound

[10]Note that according to the Cauchy-Schwarz inequality $E_0 E_2 \geq E_1^2$.

$$\mathrm{SNR} = \frac{\lambda(P_f, P_d)}{N} \frac{(\alpha + \beta)^2 E_0 - \alpha\beta(d_1 + d_2)^2 E_1 + \left(\frac{\alpha d_1^2 + \beta d_2^2}{2} \right)^2 E_2}{(-\alpha d_1 + \beta d_2)^2 E_1 + \left(\frac{\alpha d_1^2 + \beta d_2^2}{2} \right)^2 \left(E_2 - \frac{E_1^2}{E_0} \right)} \tag{63}$$

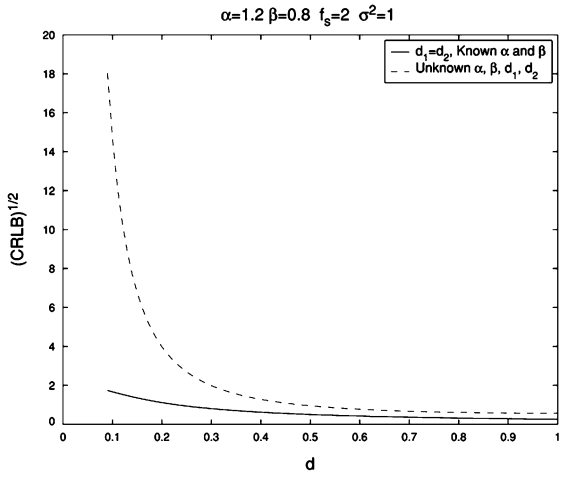

Fig. 9. $\sqrt{\mathrm{CRLB}(\hat{d})}$ versus $\hat{d}$ for two different cases.

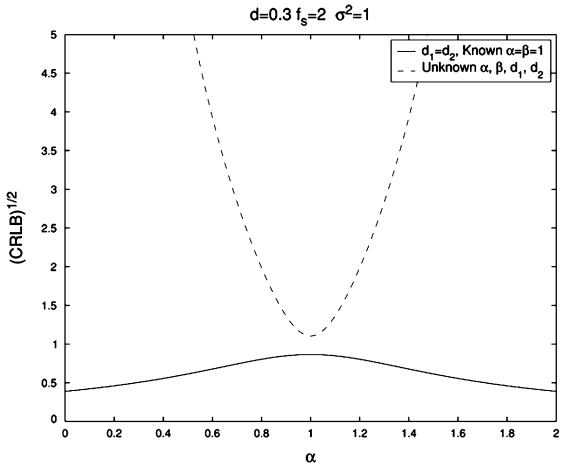

Fig. 10. $\sqrt{\mathrm{CRLB}(\hat{d})}$ versus α for two different cases.

on the variance of the parameter d, in particular, is rather helpful in verifying and confirming the earlier results of this paper. For example we shall see how the difference between α and β affects the variance of the estimate in different cases. Here, we compute the CRLB for following cases:

- the signal model in (3), i.e., known intensities but unknown d;
- the signal model in (48), i.e., unknown α, β, d_1, and d_2.

To verify the details of the calculations (carried out mostly in the frequency domain), we refer the reader to Appendix B. Recalling (3), the CRLB for the parameter d (assuming α and β known), is given by (64) and (65) at the bottom of the page. To compute the CRLB for the second case, when α, β, d_1, and d_2 are unknown, the Fisher Information matrix is computed.[11] We have

$$\mathrm{cov}(\widehat{d_1}, \widehat{d_2}, \hat{\alpha}, \hat{\beta}) \geq \Psi^{-1}(d_1, d_2, \alpha, \beta) \tag{66}$$

where Ψ is the 4×4 symmetric Fisher Information matrix with its elements defined by the equations at the bottom of the next page. The bound on the variance of $\widehat{d_1}$ and $\widehat{d_2}$ can be obtained by taking the elements (1, 1) and (2, 2) of the inverse Fisher information matrix Ψ^{-1}, respectively. Also, the CRLB on $d = d_1 + d_2$ is computed from

$$\mathrm{CRLB}(\hat{d}) = [\Psi^{-1}]_{11} + [\Psi^{-1}]_{22} + 2[\Psi^{-1}]_{12}. \tag{67}$$

[11]We thank Prof. Jeff Fessler for sharing with us his calculations for the continuous data case.

Fig. 9 shows the square-root of the CRLB (to maintain the same units as d) for d, for fixed values of the intensities α and β, versus the parameter value d, for two different cases; namely, the known intensity case with symmetrically located point sources, and the unknown α, β, d_1 and d_2 case. In this figure, we observe that the curves in each case are rather close for $d > 0.5$, and they are distinct when α is unknown and d is smaller than 0.5. In Fig. 10, the value of $d = 0.3$ is fixed, and the square-root of CRLB for $\hat{d}$ is shown over a range of values of α. The graph demonstrates the effect of the difference of α and β on the CRLB. As seen in this figure, the CRLB for the second case (unknown α, β, d_1 and d_2) increases rapidly when moving away from $(\alpha, \beta) = (1, 1)$; but for known α and β, there is a (rather slow) decay away from the position $\alpha = \beta = 1$. The observed phenomenon is counter-intuitive, but can be readily explained by looking at the derivatives we computed in the calculation of the CRLB. When point sources are located symmetrically, with unequal intensities, the shape of the overall signal is dramatically different than the case when $\alpha = \beta = 1$. This difference is accentuated further as the value of $\alpha - \beta$ becomes larger. Whereas for second case, because of uncertainty about the center and intensities of point sources, if $\alpha - \beta \neq 0$, the overall shape looks more like a single peak is present. The observed behavior is consistent with what we saw before where we demonstrated that unequal α and β yields improved detection if the center is known and vice versa.

VI. Conclusion

We have set out in this paper to address the question of resolution from a sound statistical viewpoint. In particular, we

$$\mathrm{var}(\hat{d}) \geq \frac{\sigma^2}{\sum_k \left(\frac{\partial s(x_k,d)}{\partial d}\right)^2} = \frac{\sigma^2}{\frac{1}{2\pi}\int_{-\pi}^{\pi} \left|\frac{\partial S(\omega,d)}{\partial d}\right|^2 d\omega} \tag{64}$$

$$= \frac{\sigma^2}{f_s} \frac{1}{\frac{\pi^2}{15}(\alpha^2+\beta^2) + \frac{\alpha\beta}{\pi^3 d^5}[(\pi^2 d^2 - 3)\sin(2\pi d) + 3\pi d\cos(2\pi d) + 3\pi d]} \tag{65}$$

have explicitly answered a very practical question: What is the minimum detectable distance between two point sources imaged incoherently at a given signal-to-noise ratio? Or equivalently, what is the minimum SNR required to discriminate two point sources separated by a distance smaller than the Rayleigh limit? Based on different assumptions and models, we explicitly studied four different cases in our detection-theoretic approach, from the simplest to the most general case. We employed a hypothesis testing framework using like locally most powerful tests, where the original highly nonlinear problem was approximated using a quadratic model in the parameter d. We also discussed asymptotic performance for estimation of the unknown parameters. The analysis has been carried out in one dimension to facilitate the presentation and to yield maximum intuition. We have begun the analysis in 2-D, including studies as a function of different aperture shapes and lenses, and the complete 2-D (spatial integration) sampling model. This 2-D analysis is not so different in spirit from the

$$\Psi(1,1) = \frac{1}{\sigma^2}\sum_k \left(\frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial d_1}\right)^2 = \frac{\alpha^2}{2\pi\sigma^2}\int_{-\pi}^{\pi} |\omega f_s H(\omega,f_s)|^2\, d\omega = \frac{f_s}{\sigma^2}\frac{4\pi^2\alpha^2}{15}$$

$$\Psi(2,2) = \frac{1}{\sigma^2}\sum_k \left(\frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial d_2}\right)^2 = \frac{\beta^2}{2\pi\sigma^2}\int_{-\pi}^{\pi} |\omega f_s H(\omega,f_s)|^2\, d\omega = \frac{f_s}{\sigma^2}\frac{4\pi^2\beta^2}{15}$$

$$\Psi(3,3) = \frac{1}{\sigma^2}\sum_k \left(\frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial \alpha}\right)^2 = \frac{1}{2\pi\sigma^2}\int_{-\pi}^{\pi} |H(\omega,f_s)|^2\, d\omega = \frac{f_s}{\sigma^2}\frac{2}{3}$$

$$\Psi(4,4) = \frac{1}{\sigma^2}\sum_k \left(\frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial \beta}\right)^2 = \frac{1}{2\pi\sigma^2}\int_{-\pi}^{\pi} |H(\omega,f_s)|^2\, d\omega = \frac{f_s}{\sigma^2}\frac{2}{3}$$

$$\begin{aligned}\Psi(1,2) &= \frac{1}{\sigma^2}\sum_k \frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial d_1}\frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial d_2}\\ &= \frac{-\alpha\beta}{2\pi\sigma^2}\int_{-\pi}^{\pi} |\omega f_s H(\omega,f_s)|^2 \cos(\omega f_s(d_1+d_2))\, d\omega\\ &= \frac{f_s}{\sigma^2}\frac{2\alpha\beta}{\pi^3}\frac{(\pi^2(d_1+d_2)^2-3)\sin(2\pi(d_1+d_2)) + 6\pi(d_1+d_2)\cos^2(\pi(d_1+d_2))}{(d_1+d_2)^5}\end{aligned}$$

$$\Psi(1,3) = \frac{1}{\sigma^2}\sum_k \frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial d_1}\frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial \alpha} = \frac{-\alpha}{2\pi\sigma^2}\int_{-\pi}^{\pi} \omega f_s |H(\omega,f_s)|^2\, d\omega = 0$$

$$\begin{aligned}\Psi(1,4) &= \frac{1}{\sigma^2}\sum_k \frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial d_1}\frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial \beta}\\ &= \frac{-\alpha}{2\pi\sigma^2}\int_{-\pi}^{\pi} \omega f_s |H(\omega,f_s)|^2 \sin(\omega f_s(d_1+d_2))\, d\omega\\ &= \frac{f_s}{\sigma^2}\frac{\alpha}{2\pi^3}\frac{3\sin(2\pi(d_1+d_2)) - 4\pi(d_1+d_2)\cos^2(\pi(d_1+d_2)) - 2\pi(d_1+d_2)}{(d_1+d_2)^4}\end{aligned}$$

$$\begin{aligned}\Psi(2,3) &= \frac{1}{\sigma^2}\sum_k \frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial d_2}\frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial \alpha}\\ &= \frac{\beta}{2\pi\sigma^2}\int_{-\pi}^{\pi} \omega f_s |H(\omega,f_s)|^2 \sin(\omega f_s(d_1+d_2))\, d\omega\\ &= \frac{f_s}{\sigma^2}\frac{-\beta}{2\pi^3}\frac{3\sin(2\pi(d_1+d_2)) - 4\pi(d_1+d_2)\cos^2(\pi(d_1+d_2)) - 2\pi(d_1+d_2)}{(d_1+d_2)^4}\end{aligned}$$

$$\Psi(2,4) = \frac{1}{\sigma^2}\sum_k \frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial d_2}\frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial \beta} = \frac{\beta}{2\pi\sigma^2}\int_{-\pi}^{\pi} \omega f_s |H(\omega,f_s)|^2\, d\omega = 0$$

$$\begin{aligned}\Psi(3,4) &= \frac{1}{\sigma^2}\sum_k \frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial \alpha}\frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial \beta}\\ &= \frac{1}{2\pi\sigma^2}\int_{-\pi}^{\pi} |H(\omega,f_s)|^2 \cos(\omega f_s(d_1+d_2))\, d\omega\\ &= \frac{f_s}{\sigma^2}\frac{1}{2\pi^3}\frac{-\sin(2\pi(d_1+d_2)) + 2\pi(d_1+d_2)}{(d_1+d_2)^4}\end{aligned}$$

1-D case, but is significantly more messy; so we have elected to defer its presentation to the near future.

The major conclusion of this paper is that for a given imaging scenario (in this case, incoherent imaging through a slit), with required probabilities of detection and false alarm, the minimum resolvable separation between two sources from uniformly sampled data can be derived *explicitly* as a function of the SNR per sample of the imaging array, and the sampling rate. The most useful rule of thumb we glean from these results is that for the case of equal intensities (or for the case of unequal intensities with a proper choice of test point), the minimum resolvable distance is essentially proportional to the inverse of the SNR to the fractional power of $1/4$. The proportionality constant was shown to be a function of the probabilities of detection and false alarm, and the point spread function. In deriving these results, we have unified and generalized much of the literature on this topic that, while sparse, has spanned the course of roughly four decades.

Many interesting questions remain to be studied. Of these, the analysis of the problem as a function of the sampling rate and sampling strategy come to mind. For instance, it is useful to study the performance in the presence of aliasing (i.e., sub-Nyquist sampling). It would also be interesting to study the effect of nonuniform sampling on performance.

It is important to note that the strategy for the analysis of resolution we have put forward here is very generally applicable to other types of imaging systems. Once the point-spread function of the imaging system is known, the signal model $s(x;d)$ is determined, and the same line of reasoning can be carried out. The optical imaging scenario we have described here should really be thought of as a canonical example of the application of the general strategy we propose for studying resolution. Extensions of these ideas can also be considered to study limits to resolution for indirect imaging such as in computed tomography.

As for other extensions and applications in optical imaging, an appealing direction is to study the limits to super-resolution from video [23]–[25]. The analysis presented here can help answer questions regarding the ability of image super-resolution methods to integrate multiple low resolution frames to produce a high resolution image from aliased data.

Finally, we wish to mention that this paper, we hope, represents one step forward in an overall methodology for studying imaging and image processing that appeals directly to concepts in information theory. This approach and point of view has been sorely lacking in the imaging community, and we hope that it will become more pervasive in the years to come.

Appendix A
On the Accuracy of the Quadratic Approximation

Here, we present an analysis to demonstrate the accuracy of the Taylor expansion proposed in Section 3. We consider the general model of (48) and its Taylor expansion in (49). Let us define residual percentage error of the approximation as follows:

$$\epsilon = \frac{\left\| \mathbf{s} - (\alpha+\beta)\mathbf{h} - (-\alpha d_1 + \beta d_2)\mathbf{h}_1 - \frac{\alpha d_1^2 + \beta d_2^2}{2}\mathbf{h}_2 \right\|^2}{\|\mathbf{s}\|^2} \tag{68}$$

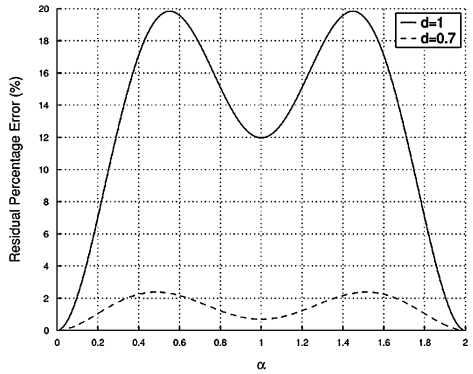

Fig. 11. Residual percentage error of the quadratic model; $\alpha d_1 = \beta d_2$.

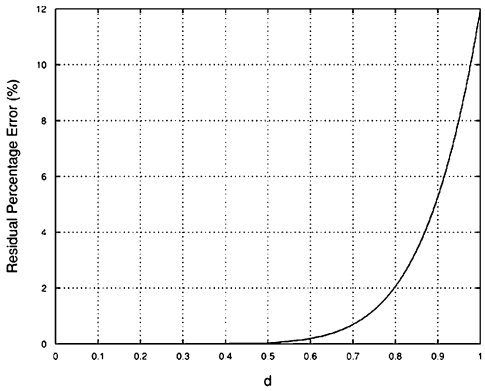

Fig. 12. Residual percentage error of the quadratic model; $\alpha = \beta = 1$.

Consider the case when $\alpha d_1 = \beta d_2$ (See Appendix C). Fig. 11 shows the upper bound when $(d = d_1 + d_2 = 1)$ for ϵ as a function of α for $h(x) = \mathrm{sinc}^2(x)$ (Note that again for above-Nyquist sampling, ϵ is independent from the sampling rate.). The maximum of ϵ is less than 20% in any case. Also, as seen in this figure, the approximation error for $d = 0.7$ is always less than 2.5%. Fig. 12 shows the curve for ϵ versus d which indicates that the approximation error is quite acceptable for the range of interest near $d = 0$. To have a picture of the local error in the approximation, the error term

$$e(x;\alpha,\beta,d_1,d_2) = s(x;\alpha,\beta,d_1,d_2) - (\alpha+\beta)h(x) - (-\alpha d_1 + \beta d_2)h_1(x) - \frac{\alpha d_1^2 + \beta d_2^2}{2}h_2(x)$$

is shown in Fig. 13 for two different values of d over the range of the variable x in $[-10, 10]$.

Appendix B
Frequency Domain Representation; Parseval's Theorem for the Signal $s(x;d)$

Considering the sampled signal of the general model, where the point sources are located at $-d_1$ and d_2 we have

$$\begin{aligned} s(n;\alpha,\beta,d_1,d_2) &= s(x;\alpha,\beta,d_1,d_2)|_{x=\frac{n}{f_s}} \\ &= \alpha h\left(\frac{n}{f_s} - d_1\right) + \beta h\left(\frac{n}{f_s} + d_2\right). \end{aligned} \tag{69}$$

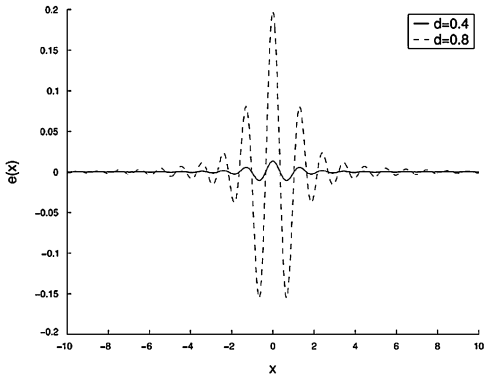

Fig. 13. Difference between the actual signal and the quadratic model; $\alpha = \beta = 1$.

For the case of above-Nyquist sampling,[12] in the frequency domain we will have the following 2π-periodic representation (see (70) at the bottom of the page) where $H(\omega, f_s) = (f_s^2/2\pi)((2\pi/f_s) - |\omega|)$ is the DTFT of $h(x_k)$ when $h(x) = \mathrm{sinc}^2(x)$ and sampling rate is f_s. Correspondingly, for this case, the functions $h_1(x)$ and $h_2(x)$ can be written in the frequency domain as

$$H_1(\omega, f_s) = \begin{cases} j\frac{\omega f_s^3}{2\pi}\left(\frac{2\pi}{f_s} - |\omega|\right) & |\omega| < \frac{2\pi}{fs} \\ 0 & \frac{2\pi}{fs} \le |\omega| \le 2\pi \end{cases} \tag{71}$$

$$H_2(\omega, f_s) = \begin{cases} -\frac{\omega^2 f_s^4}{2\pi}\left(\frac{2\pi}{f_s} - |\omega|\right) & |\omega| < \frac{2\pi}{fs} \\ 0 & \frac{2\pi}{fs} \le |\omega| \le 2\pi \end{cases} \tag{72}$$

Using Parseval's identities [19]:

$$\sum_{n=-\infty}^{\infty} |x(n)|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |X(\omega)|^2\, d\omega \tag{73}$$

$$\sum_{n=-\infty}^{\infty} x(n)y^*(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(\omega)Y^*(\omega)\, d\omega \tag{74}$$

we can easily compute the following terms:

$$E_0 = \mathbf{h}^T\mathbf{h} = f_s\frac{2}{3} \tag{75}$$

$$E_1 = \mathbf{h}_1^T\mathbf{h}_1 = f_s\frac{4\pi^2}{15} \tag{76}$$

$$E_2 = \mathbf{h}_2^T\mathbf{h}_2 = f_s\frac{32\pi^4}{105} \tag{77}$$

and

$$\mathbf{h}_1^T\mathbf{s}_0 = \mathbf{h}_1^T\mathbf{h}_2 = 0 \tag{78}$$

[12]To recover exactly $s(x; d)$ would mathematically require an infinite number of measurements (or samples) $s(n; d)$ [21]. But since we have considered a fairly large range (-10 to 10) for sampling, and since the energy in the tails of the function in the range is very small, the effect of aliasing is essentially negligible.

Note that in every case the energy terms are proportional to the sampling rate. It can be shown [20] that the energy of any uniformly (super-critically) sampled version of a band-limited signal is proportional to the sampling rate.

Appendix C
Is $\alpha d_1 \approx \beta d_2$ a Reasonable Assumption?

Suppose that we first wish to determine a location at which we carry out our hypothesis test. A reasonable way to find a good candidate is to compute the correlation of the signal with a shifted version of $h(x)$ and find the point where the correlation is maximum (this would yield a point near the brighter of the two peaks). Consider

$$R_{sh}(|\tau|, \alpha, \beta, d_1, d_2) = \int_{-\infty}^{+\infty} (s(x; \alpha, \beta, d_1, d_2) + w(x))h(x+\tau)\, dx \tag{79}$$

$$= \int_{-\infty}^{+\infty} (\alpha h(x - d_1) + \beta h(x + d_2) + w(x))h(x+\tau)\, dx \tag{80}$$

$$= \alpha R_{hh}(|\tau| - d_1) + \beta R_{hh}(|\tau| + d_2) + u(|\tau|) \tag{81}$$

where R_{sh} and R_{hh} are the cross-correlation and autocorrelation functions, respectively, and

$$u(|\tau|) = \int_{-\infty}^{+\infty} w(x)h(x+\tau)\, dx \tag{82}$$

is a noise term (with zero mean). It should be clear form the model that R_{sh} would be maximized at $\tau = 0$. Also, since d_1 and d_2 are assumed to be small, by using the Taylor expansion around $|\tau| - d_1 = 0$ and $|\tau| + d_2 = 0$, we will have

$$R_{hh}(|\tau| - d_1) = \xi_0 + (|\tau| - d_1)\xi_1 + (|\tau| - d_1)^2\xi_2 \tag{83}$$

$$R_{hh}(|\tau| + d_2) = \xi_0 + (|\tau| + d_2)\xi_1 + (|\tau| + d_2)^2\xi_2 \tag{84}$$

where ξ_0, ξ_1, and ξ_2 are some constant coefficients of the above Taylor expansion. Also, it can be shown that $\xi_1 = 0$. Therefore, we can write (81) as follows:

$$R_{\mathrm{sh}}(|\tau|, \alpha, \beta, d_1, d_2) = (\alpha+\beta)\xi_0 + (\alpha(|\tau| - d_1)^2 + \beta(|\tau| - d_2)^2)\xi_2 + u(|\tau|) \tag{85}$$

Taking derivative of $R_{\mathrm{sh}}(|\tau|, \alpha, \beta, d_1, d_2)$ with respect to τ and setting it to zero will result in:

$$(\alpha+\beta)|\tau| = \alpha d_1 - \beta d_2 \tag{86}$$

Hence, a proper selection of τ (i.e., the test point) will lead to $\alpha d_1 \approx \beta d_2$.

$$S(\omega, d) = \begin{cases} H(\omega, f_s)(\alpha\exp(-j\omega f_s d_1) + \beta\exp(j\omega f_s d_2)) & |\omega| < \frac{2\pi}{fs} \\ 0 & \frac{2\pi}{fs} \le |\omega| \le 2\pi \end{cases} \tag{70}$$

ACKNOWLEDGMENT

The authors wish to acknowledge Prof. A. Shakouri of U.C., Santa Cruz, for providing the early practical inspiration from the laboratory bench that led them to consider the questions addressed in this paper. They thank Prof. J. Fessler of the University of Michigan for his helpful suggestions for the CRLB analysis. They also thank the reviewers for their constructive comments and suggestions.

REFERENCES

[1] J. W. Goodman, *Introduction to Fourier Optics*. New York: McGraw-Hill, 1996.
[2] J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*. New York: Wiley, 1978.
[3] L. B. Lucy, "Statistical limits to super-resolution," *Astron. Astrophys*, vol. 261, pp. 706–710, 1992.
[4] C. W. Helstrom, "The detection and resolution of optical signals," *IEEE Trans. Inf. Theory*, vol. IT-10, pp. 275–287, 1964.
[5] ——, "Detection and resolution of incoherent objects by a background-limited optical system," *J. Opt. Soc. Amer.*, vol. 59, pp. 164–175, 1969.
[6] ——, "Resolvability of objects from the standpoint of statistical parameter estimation," *J. Opt. Soc. Amer.*, vol. 60, pp. 659–666, 1970.
[7] L. B. Lucy, "Resolution limits for deconvolved images," *Astron. J.*, vol. 104, pp. 1260–1265, 1992.
[8] A. van den Bos, "Ultimate resolution: A mathematical framework," *Ultramicroscopy*, vol. 47, pp. 298–306, 1992.
[9] A. J. den Dekker, "Model-based optical resolution," *IEEE Trans. Instrum. Meas.*, vol. 46, pp. 798–802, 1997.
[10] A. J. den Dekker and A. van den Bos, "Resolution, a survey," *J. Opt. Soc. Amer.*, vol. 14, pp. 547–557, 1997.
[11] E. Bettens, D. Van Dyck, A. J. den Dekker, J. Sijbers, and A. van den Bos, "Model-based two-object resolution from observations having counting statistics," *Ultramicroscopy*, vol. 77, pp. 37–48, 1999.
[12] A. van den Bos, "Resolution in model-based measurements," *IEEE Trans. Instrum. Meas.*, vol. 51, pp. 1055–1060, 2002.
[13] E. L. Kosarev, "Shannon's superresolution limit for signal recovery," *Inverse Problem*, vol. 6, pp. 55–76, 1990.
[14] P. Milanfar and A. Shakouri, "A Statistical analysis of diffraction-limited imaging," in *Proc. Int. Conf. Image Processing*, Sept. 2002, pp. 864–867.
[15] S. M. Kay, *Fundamentals of Statistical Signal Processing, Estimation Theory*: Prentice-Hall, Inc., 1998.
[16] ——, *Fundamentals of Statistical Signal Processing, Detection Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1998.
[17] ——, *Modern Spectral Estimation, Theory and Application*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
[18] ——, "Spectrum analysis, a modern perspective," *Proc. IEEE*, vol. 69, no. 11, pp. 1380–1418, 1981.
[19] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[20] P. P. Vaidyanathan, "Generalizations of the sampling theorem: Seven decades after Nyquist," *IEEE Trans. Circuits Syst.*, vol. 48, pp. 1094–1109, Sept. 2001.
[21] M. Vetterli, P. Marziliano, and T. Blu, "Sampling signals with finite rate of innovation," *IEEE Trans. Signal Processing*, vol. 50, pp. 1417–1428, June 2002.
[22] M. Shahram and P. Milanfar, "A statistical analysis of achievable resolution in incoherent imaging," in *Proc. SPIE Annual Meeting*, San Diego, CA, Aug. 2003, URL: http://www.soe.ucsc.edu/~milanfar/publications.htm.
[23] M. Elad and A. Feuer, "Restoration of single super-resolution image from several blurred, noisy and down-sampled measured images," *IEEE Trans. Image Processing*, vol. 6, pp. 1646–1658, Dec. 1997.
[24] N. Nguyen, P. Milanfar, and G. H. Golub, "A computationally efficient image superresolution algorithm," *IEEE Trans. Image Processing*, vol. 10, pp. 573–583, Apr. 2001.
[25] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe superresolution," *IEEE Trans. Image Processing*, to be published.

Morteza Shahram received the B.S. degree from the Amir-Kabir University of Technology, Tehran, Iran, in 1996 and the M.S. degree from the Sharif University of Technology, Tehran, in 1998 both in electrical engineering. He is currently pursuing the Ph.D. degree in electrical engineering at the University of California, Santa Cruz.

He was with the Signal Company, Tehran, as a Research Engineer from 1996 to 2001. His research interests are statistical signal and image processing and information-theoretic imaging.

Peyman Milanfar (S'90–M'93–SM'98) received the B.S. degree in electrical engineering/mathematics from the University of California, Berkeley, in 1988, and the S.M., E.E., and Ph.D. degrees in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1990, 1992, and 1993, respectively.

Until 1999, he was a Senior Research Engineer at SRI International, Menlo Park, CA. He is currently Associate Professor of Electrical Engineering at the University of California, Santa Cruz. He was a Consulting Assistant Professor of computer science at Stanford University from 1998–2000, and a visiting Associate Professor there from June to December 2002. His technical interests are in statistical signal and image processing, and inverse problems.

Dr. Milanfar won a National Science Foundation CAREER award in 2000. He was an associate editor for the IEEE SIGNAL PROCESSING LETTERS from 1998 to 2001.