A Statistical Analysis of Achievable Resolution in Incoherent Imaging

Morteza Shahram*, Peyman Milanfar†
Department of Electrical Engineering, University of California, Santa Cruz, CA 95064

ABSTRACT

The present paper concerns the statistical analysis of limits to achievable resolution 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 intensities. The objective is to study how far beyond the classical Rayleigh limit of resolution one can reach at a given signal to noise ratio. We consider the definition of resolution limit from a statistical point of view as the ability of the imaging system to distinguish two closely-located sources in presence of additive noise. This problem can be stated in a hypothesis testing framework where the hypotheses consider whether one or two point sources are present. In terms of signal detection/ estimation, this leads to composite detection/estimation problem where a deterministic signal with unknown parameters is being sought. To solve this problem, we use locally optimal statistical tests with respect to a desired range of (small) separations between the point sources. Specifically, we will derive explicit relationships between the minimum detectable distance between two point sources, and the required SNR. For a specific point spread function, the required SNR can be expressed as a function of probabilities of detection and false alarm and the distance between point sources.

Keywords: Point Source Resolution, Resolution Limit, Diffraction Limit, Composite Hypothesis Testing, Estimation/Detection Theory

1. 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 Figure 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.

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.[2] Indeed, at sufficiently high sampling rates, and in the absence of noise, arbitrarily small details can be resolved.

To begin, let us assume that the original signal of interest is the sum of two impulse functions separated by small distances d_x and d_y in horizontal and vertical axes:

$$\sqrt{\alpha}\delta\left(x-\frac{d_x}{2}, y-\frac{d_y}{2}\right)+\sqrt{\beta}\delta\left(x+\frac{d_x}{2}, y+\frac{d_y}{2}\right) \tag{1}$$

*Email: shahram@ee.ucsc.edu
†Email: milanfar@ee.ucsc.edu

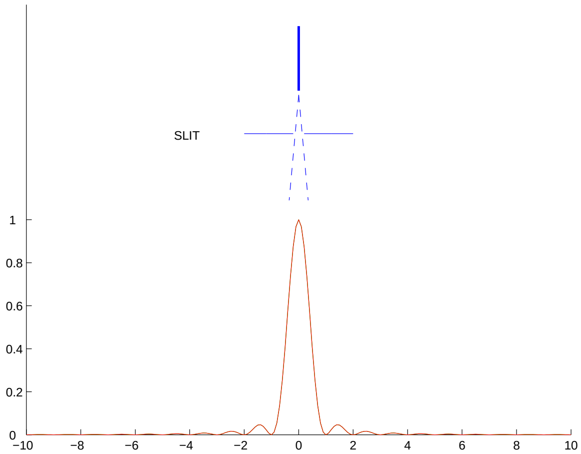

Figure 1. Image of point source captured by diffraction-limited imaging

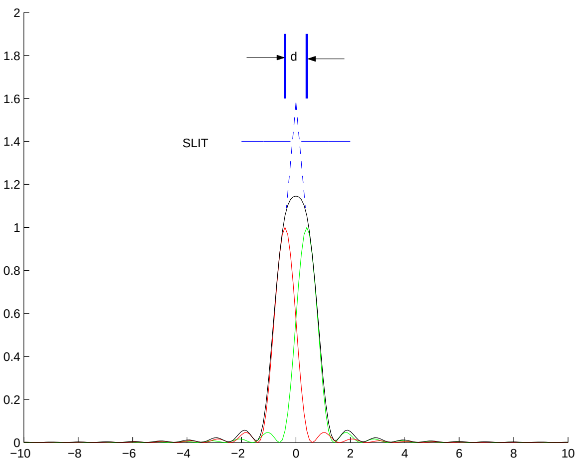

Figure 2. Incoherent imaging of two closely located point sources

where the distance between point sources is $d = \sqrt{d_x^2 + d_y^2}$ and as an energy constraint we require $\alpha + \beta = 2$ throughout the following computations[‡]. As mentioned before, the measured signal will be the incoherent sum of point spread functions, resulting from an imaging aperture with PSF of $p(x, y)$

$$
\begin{aligned}
f(x) &= s(x; \alpha, \beta, d_x, d_y) + w(x, y) && (2) \\
&= \alpha p\left(x - \frac{d_x}{2}, y - \frac{d_y}{2}\right) + \beta p\left(x + \frac{d_x}{2}, y + \frac{d_y}{2}\right) + w(x, y) && (3)
\end{aligned}
$$

where $w(x, y)$ is assumed to be a zero-mean Gaussian white noise process with variance σ^2 . This last assumption is of course not accurate for photon-limited imaging systems, and this case can be considered in the continuation of this work.

As an example, let us consider the case of incoherent imaging of two point sources in one-dimensional case, as seen in Figure 2), where the aperture is a slit and hence the resulting point spread function is $\text{sinc}^2(x)$. With the present definition, the Rayleigh limit corresponds to $d = d_x = 1$ as can be seen in Figures 1 and 2.

[‡]From now on we refer to α and β as intensities.

This means that for values $d < 1$, the two point sources are (in the classical Rayleigh sense) "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 Holstrom. In particular, he derived lower bounds on the mean-square error of unbiased estimators for the source positions, the distance between the sources, and the radiance values, using the Cramér-Rao inequality.[5] Also, 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.[4] (This second scenario is, of course, what interests us in the present paper.) However he assumed that the distance d is known to the detector.

Lucy used an approximate statistical theory 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.[2] 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. The analysis of the achievable resolution in deconvolved astronomical images was studied based on a criterion similar to Rayleigh's.[6]

Finally, an interesting, more recent paper[7] 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. That paper 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. These 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 parameters, noise variance, sampling rate, etc.). Our earlier work on this problem was focused on one-dimensional imaging.[8,9]

The organization of the paper is as follows. Section 2 will explain and formulate our definition, and the corresponding statistical framework and models, in detail. In Section 3, in order to use linear detection/estimation structures, we will discuss a signal approximation approach. In Section 4, we will study our statistical analysis for the case of equal intensities.

2. 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 $d = \sqrt{d_x^2 + d_y^2}$ is equal to zero or not. Suppose for underlying imaging case, the diffraction limit is equal to μ. If $d = 0$ then we only have one peak and if $d > \mu$ 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 $0 \leq d < \mu$. Therefore, we can define two hypotheses, which will form the basis of our statistical framework. Namely, let H_0 denote the null hypothesis that $d = 0$ (one peak present) and let H_1 denote the alternate hypothesis that $d > 0$ (two peaks present):

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

Given samples at $(x_k, y_l) \quad (k, l = 1, \cdots, N)$ of the measured signal, we can rewrite the problem as:

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

where $\mathbf{s}$, $\mathbf{s}_0$, $\mathbf{f}$ and $\mathbf{w}$ are vectors arranged in lexicographic order:

$$\begin{aligned} \mathbf{s}[kN + l] &= s(x_k, y_l) \\ \mathbf{s}_0[kN + l] &= s_0(x_k, y_l) \\ \mathbf{f}[kN + l] &= f(x_k, y_l) \\ \mathbf{w}_0[kN + l] &= w(x_k, y_l) \end{aligned}$$

and

$$s(x_k, y_l; \alpha, \beta, d_x, d_y) = \alpha p\left(x_k - \frac{d_x}{2}, y_l - \frac{d_y}{2}\right) + \beta p\left(x_k + \frac{d_x}{2}, y_l + \frac{d_y}{2}\right), \tag{6}$$

$$s_0(x_k, y_l) = s(x_k, y_l; \alpha, \beta, d_x, d_y)|_{d_x, d_y = 0} = 2p(x_k, y_l). \tag{7}$$

This is a problem of detecting a deterministic signal with unknown parameters (α, β, d_x, and d_y, in general). From Equation 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 PDFs.[11] 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 this 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 parameters range $0 \leq d < \mu$.

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

$$L(\mathbf{f}) = \frac{\max_{d_x, d_y} p(\mathbf{f}, d_x, d_y, H_1)}{p(\mathbf{f}, H_0)} = \frac{p(\mathbf{f}, \widehat{d_x}, \widehat{d_y}, H_1)}{p(\mathbf{f}, H_0)} > \gamma \tag{8}$$

where $\widehat{d_x}$ and $\widehat{d_y}$ denote the ML estimate of d_x and d_y, respectively, and $p(\mathbf{f}, d_x, d_y, H_1)$ and $p(\mathbf{f}, H_0)$ are PDFs under H_1 and H_0, respectively. Assuming additive white Gaussian noise (AWGN) with variance σ^2 and $\widehat{\mathbf{s}} = [s(x_1; 1, 1, \widehat{d_x}, \widehat{d_y}), \cdots, s(x_N; 1, 1, \widehat{d_x}, \widehat{d_y})]^T$ we will have:

$$L(\mathbf{f}) = \frac{\frac{1}{(2\pi\sigma^2)^{N/2}} \exp(-\frac{1}{2\sigma^2}\|\mathbf{f} - \widehat{\mathbf{s}}\|^2)}{\frac{1}{(2\pi\sigma^2)^{N/2}} \exp(-\frac{1}{2\sigma^2}\|\mathbf{f} - \mathbf{s}_0\|^2)} = \exp\left(-\frac{1}{2\sigma^2}\left(-\|\widehat{\mathbf{s}}\|^2 + \|\mathbf{s}_0\|^2 + 2\mathbf{f}^T(\widehat{\mathbf{s}} - \mathbf{s}_0)\right)\right)$$

Therefore the test is given by:

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

which is the output of a matched filter. 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 $\{(d_x, d_y) \ni d_x^2 + d_y^2 < \mu^2\}$, 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 $(d_x, d_y) = (0, 0)$, and to apply locally optimal detectors. This is the approach we take.

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

$$\begin{aligned} s(x_k, y_l; \alpha, \beta, d_x, d_y) &\approx s_0(x_k, y_l) + (\beta-\alpha)d_x h_x(x_k, y_l) + (\beta-\alpha)d_y h_y(x_k, y_l) \\ &+ d_x^2 h_{xx}(x_k, y_l) + d_y^2 h_{yy}(x_k, y_l) + 2d_x d_y h_{xy}(x_k, y_l) \end{aligned} \tag{10}$$

where it can be shown that

$$s_0(x_k, y) = 2p(x_k, y_l) \tag{11}$$

$$h_x(x_k, y_l) = \frac{1}{2}\frac{\partial p(x,y)}{\partial x}\bigg|_{x=x_k, y=y_l} \tag{12}$$

$$h_y(x_k, y_l) = \frac{1}{2}\frac{\partial p(x,y)}{\partial y}\bigg|_{x=x_k, y=y_l} \tag{13}$$

$$h_{xx}(x_k, y_l) = \frac{1}{4}\frac{\partial^2 p(x,y)}{\partial x^2}\bigg|_{x=x_k, y=y_l} \tag{14}$$

$$h_{yy}(x_k, y_l) = \frac{1}{4}\frac{\partial^2 p(x,y)}{\partial y^2}\bigg|_{x=x_k, y=y_l} \tag{15}$$

$$h_{xy}(x_k, y_l) = \frac{1}{4}\frac{\partial^2 p(x,y)}{\partial x \partial y}\bigg|_{x=x_k, y=y_l} \tag{16}$$

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 $\alpha = \beta$ *no* term in d_x and d_y would appear in the approximation. The proposed approximation simplifies the hypothesis testing problem to essentially a linear detection problem (as we will see in 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.[11] Refereing to (5) and continuing with vector notation we have:

$$\mathbf{s} \approx \mathbf{s}_0 + (\beta-\alpha)d_x\mathbf{h}_x + (\beta-\alpha)d_y\mathbf{h}_y + d_x^2\mathbf{h}_{xx} + d_y^2\mathbf{h}_{yy} + 2d_xd_y\mathbf{h}_{xy} \tag{17}$$

where $\mathbf{h}_x$, $\mathbf{h}_y$, $\mathbf{h}_{xx}$, $\mathbf{h}_{yy}$ and $\mathbf{h}_{xy}$ are in lexicographic order. For example

$$\mathbf{h}_x[kN + l] = h_x(x_k, y_l) \tag{18}$$

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

$$\begin{cases} H_0: & \tilde{\mathbf{f}} = \mathbf{s_0} + \mathbf{w} \\ H_1: & \tilde{\mathbf{f}} = \mathbf{s}_0 + (\beta-\alpha)d_x\mathbf{h}_x + (\beta-\alpha)d_y\mathbf{h}_y + d_x^2\mathbf{h}_{xx} + d_y^2\mathbf{h}_{yy} + 2d_xd_y\mathbf{h}_{xy} + \mathbf{w} \end{cases} \tag{19}$$

Since $\mathbf{s}_0$ is a common term in both hypotheses and it is independent from d, we may simplify further:

$$\begin{cases} H_0: & \mathbf{y} = \mathbf{w} \\ H_1: & \mathbf{y} = (\beta-\alpha)d_x\mathbf{h}_x + (\beta-\alpha)d_y\mathbf{h}_y + d_x^2\mathbf{h}_{xx} + d_y^2\mathbf{h}_{yy} + 2d_xd_y\mathbf{h}_{xy} + \mathbf{w} \end{cases} \tag{20}$$

where $\mathbf{y} = \tilde{\mathbf{f}} - \mathbf{s_0}$ and the parameters d_x and d_y are unknown. At this point we define the measured SNR per sample as follows:

$$\text{SNR} = \frac{1}{N^2\sigma^2}\|\mathbf{s}_0 + (\beta-\alpha)d_x\mathbf{h}_x + (\beta-\alpha)d_y\mathbf{h}_y + d_x^2\mathbf{h}_{xx} + d_y^2\mathbf{h}_{yy} + 2d_xd_y\mathbf{h}_{xy}\|^2 \tag{21}$$

[§]It is important here to note that this is an approximation about the *parameters* of interest d_x and d_y, and not the variables x or y; as such it therefore is a global approximation of the function.

4. 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. We will treat the problem when the intensity coefficients α and β are equal. Analysis for the case of unequal intensities or the case of unknown unequal intensities in general two-dimensional imaging has been carried and we have elected to present the results in next publication. However, the results for one-dimensional imaging scenario were presented before.[9]

When $\alpha = \beta$, (20) is reduced to:

$$\begin{cases} H_0: & \mathbf{y} = \mathbf{w} \\ H_1: & \mathbf{y} = d_x^2 \mathbf{h}_{xx} + d_y^2 \mathbf{h}_{yy} + 2 d_x d_y \mathbf{h}_{xy} + \mathbf{w} \end{cases} \tag{22}$$

Equation 22 can be reexpressed as:

$$\begin{cases} H_0: & d_x = 0 \quad \text{and} \quad d_y = 0 \\ H_1: & d_x \neq 0 \quad \text{or} \quad d_y \neq 0 \end{cases} \tag{23}$$

Suppose that $D = [d_x \, d_y]^T$, the eigenvalues of the matrix DD^T are $d_x^2 + d_y^2$ and 0. Therefore (23) can be reduced to the following one-sided parameter test:

$$\begin{cases} H_0: & d_x^2 + d_y^2 = 0 \\ H_1: & d_x^2 + d_y^2 > 0 \end{cases} \tag{24}$$

It is readily shown that the ML estimate for the parameter $D = d_x^2 + d_y^2$ is given by[10]:

$$\widehat{D} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y} \tag{25}$$

where:

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_{xx}^T \mathbf{h}_{xx} & \mathbf{h}_{xy}^T \mathbf{h}_{xy} \\ \mathbf{h}_{xy}^T \mathbf{h}_{xy} & \mathbf{h}_{yy}^T \mathbf{h}_{yy} \end{bmatrix} \tag{26}$$

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

$$T(\mathbf{y}) = \widehat{D}^2 \tag{27}$$

We note that the expression for the test-statistic is essentially an energy detector with the condition that the value of $D = d_x^2 + d_y^2$ is in fact estimated from the data itself. The detector structure, due to our knowledge of the sign of the unknown distance parameter, is in fact a Uniformly Most Powerful (UMP) detector.[11–13] For any given data set $\mathbf{y}$, we decide H_1 in (24) if the statistic exceeds a specified threshold:

$$T(\mathbf{y}) > \gamma \tag{28}$$

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. The detection rate (P_d) and false-alarm rate (P_f) for this detector are defined and related as[11]:

$$P_d = P(H_1|H_1) \quad = \quad Q\left(\frac{\gamma}{\sqrt{\sigma^2 \eta}}\right) \tag{29}$$

$$P_f = P(H_1|H_0) \quad = \quad Q\left(\frac{\gamma - (d_x^2 + d_y^2)\eta}{\sqrt{\sigma^2 \eta}}\right) \tag{30}$$

$$\Longrightarrow \frac{1}{\sigma^2}(d_x^2 + d_y^2)^2 \eta \quad = \quad \left(Q^{-1}(P_f) - Q^{-1}(P_d)\right)^2 \tag{31}$$

where 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[11] and¶

$$\eta \quad = \quad \mathbf{h}_{xx}^T \mathbf{h}_{xx} + \mathbf{h}_{yy}^T \mathbf{h}_{yy} + 2\mathbf{h}_{xy}^T \mathbf{h}_{xy} \tag{32}$$

We observe that the detection rate can be written as a function of the pre-specified false alarm rate, η, D and σ^2. It is worth noting here that the UMP detector produces the highest detection probability for all values of the unknown parameter, and for a given false-alarm rate.[11] A particularly intriguing and useful relationship is the behavior of the smallest peak separation $d = \sqrt{d_x^2 + d_y^2}$, 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 (21), (31) and (32) the relation between d_{min} and required SNR can be made explicit:

$$\text{SNR} = \frac{1}{N^2}\left(Q^{-1}(P_f) - Q^{-1}(P_d)\right)^2 \frac{64E_0 - 16E_x d_x^2 - 16E_y d_y^2 + E_{xx} d_x^4 + E_{yy} d_y^4 + 4E_{xy} d_x^2 d_y^2}{(E_{xx} + E_{yy} + 2E_{xy})(d_x^2 + d_y^2)^2} \tag{33}$$

where

$$E_0 = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p^2(x,y)dxdy \qquad E_{xx} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left(\frac{\partial^2 p(x,y)}{\partial x^2}\right)^2 dxdy \tag{34}$$

$$E_x = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left(\frac{\partial p(x,y)}{\partial x}\right)^2 dxdy \qquad E_{yy} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left(\frac{\partial^2 p(x,y)}{\partial y^2}\right)^2 dxdy \tag{35}$$

$$E_y = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left(\frac{\partial p(x,y)}{\partial y}\right)^2 dxdy \qquad E_{xy} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \left(\frac{\partial^2 p(x,y)}{\partial x \partial y}\right)^2 dxdy \tag{36}$$

It is worth noting that in (33), for small values of d_x and d_y a reasonably informative (but approximate) way to write SNR is

$$\text{SNR} \quad \approx \quad \frac{64}{N^2}(Q^{-1}(P_f) - Q^{-1}(P_d))^2 \frac{E_0}{E_{xx} + E_{yy} + 2E_{xy}} \frac{1}{(d_x^2 + d_y^2)^2} \tag{37}$$

As a specific example, let us consider the case of circular aperture with following point spread function[1]:

$$p(x,y) = 4\frac{J_1\left(2\pi\sqrt{x^2+y^2}\right)}{\pi(x^2+y^2)} \tag{38}$$

where $J_1(.)$ is the first order bessel function of the first kind. For this case (33) is given by:

$$\text{SNR} = \frac{1}{N^2}\left(Q^{-1}(P_f) - Q^{-1}(P_d)\right)^2 \frac{0.58 - 1.57d^2 + 1.6d^4}{4.8d^4} \tag{39}$$

where $d = \sqrt{d_x^2 + d_y^2}$. The expression in (33) gives an implicit relation between the smallest detectable distance between the two (equal intensity) sources, at the particular SNR. For the specified choice of $P_d = 0.99$ and $P_f = 10^{-6}$, if we collect N^2 samples at at points (x_k, y_l) within the interval region $[-10, 10] \times [-10, 10]$, at just above the Nyquist rate, we have

$$\text{SNR} = \frac{1}{N^2}\left[\frac{6.12}{d^4} - \frac{16.38}{d^2} + 16.7\right] \tag{40}$$

A plot of this function is shown in Figure 3. For a small d, the term involving d^{-4} dominates in (40). Therefore

$$\text{SNR} \approx \frac{6.12}{N^2 d^4} \tag{41}$$

¶For the case when $p(x,y) = p(x,-y) = p(-x,y) = p(-x,-y)$

A plot of this approximate expression is also shown in Figure 3 to be compared against the exact expression 40. The above relation (41) 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. 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.

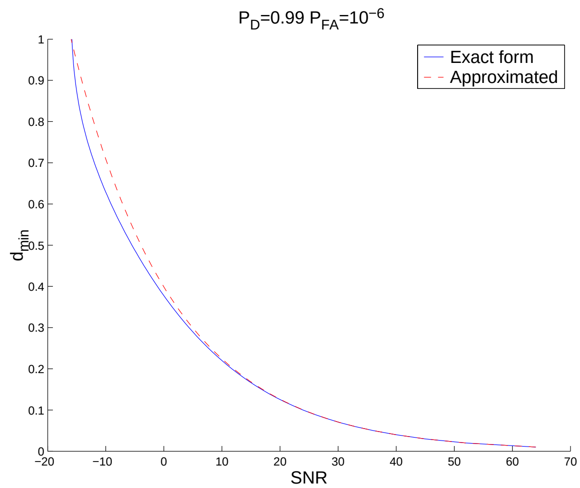

Figure 3. Minimum detectable d as a function of SNR (in dB) at Nyquist rate (exact and approximate)

5. CONCLUSIONS

We have set out in this paper to address the question of resolution from a sound statistical viewpoint. In particular, we 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? In order to use standard hypothesis testing frameworks like locally most powerful tests, the original highly nonlinear problem was approximated using a quadratic model. To gain maximum intuition, we studied the case of equal intensities in this paper, but we have presently carried out the analysis for the case where the two sources are of unequal and unknown brightness and also for different scenarios of imaging such as: imaging by charge coupled devices (CCD), the case of additional sources of blurring in the imaging process and so forth.

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, 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. 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.

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 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.[14–16] 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.

Acknowlegement: This work was supported in part by U.S. National Science Foundation CAREER Grant CCR-9984246. Also, we wish to acknowledge Prof. Ali Shakouri of U.C. Santa Cruz for providing the early practical inspiration.

REFERENCES

1. J. W. Goodman, *Introduction to Fourier Optics.* New York: McGraw-Hill, 1996.
2. L. B. Lucy, "Statistical Limits to Super-resolution" Astron. Astrophys. vol. 261 pp. 706-710, 1992
3. C. W. Helstrom, "The Detection and Resolution of Optical Signals." IEEE Trans. Inf. Theory IT-10, pp. 275-287, 1964.
4. C. W. Helstrom, "Detection and Resolution of Incoherent Objects by a Background-Limited Optical System." J. Opt. Soc. Am. 59, pp. 164-175, 1969.
5. C. W. Helstrom, "Resolvability of Objects from the Standpoint of Statistical Parameter Estimation." J. Opt. Soc. Am. 60, pp. 659-666, 1970.
6. L. B. Lucy, "Resolution Limits for Deconvolved Images" Astron. J. vol. 104 pp. 1260–1265, 1992.
7. E. L. Kosarev, "Shannon's Superresolution Limit for Signal Recovery" Inverse Problem, vol. 6, pp. 55-76, 1990
8. P. Milanfar, A. Shakouri, "A Statistical Analysis of Diffraction-Limited Imaging" Proceeding of the International Conference on Image Processing, pp. 864-867, September 2002.
9. M. Shahram, P. Milanfar "Imaging below the Diffraction Limit, A Statistical Analysis" submitted to IEEE Transactions on Image Processing
10. S. M. Kay, *Fundamentals of Statistical Signal Processing, Estimation Theory.* Prentice-Hall, Inc, 1998
11. S. M. Kay, *Fundamentals of Statistical Signal Processing, Detection Theory.* Prentice-Hall, Inc, 1998
12. H. L. Van trees, *Detection, Estimation and Modulation Theory.* Wiley Inter-science, Inc, 2001
13. L. L. Scharf, *Statistical Signal Processing* Addison-WesleyWiley Inter-science, Inc, 1990
14. M. Elad, A. Feuer, "Restoration of single super-resolution image from several blurred, noisy and down-sampled measured images." IEEE Trans. on Image Processing, vol. 6, no. 12, pp.1646-1658, Dec. 1997.
15. N. Nguyen, P. Milanfar, G.H. Golub, "A computationally efficient image superresolution algorithm." IEEE Trans. on Image Processing, vol.10, no.4, pp. 573-583, Apr. 2001.
16. S. Farsiu, D. Robinson, M. Elad, P. Milanfar, "Robust Superresolution" To appear in Proceedings of the 2003 IEEE International Conference on Image Processing, Barcelona