UNIVERSITY OF CALIFORNIA
SANTA CRUZ

STATISTICAL AND INFORMATION-THEORETIC ANALYSIS OF RESOLUTION IN IMAGING AND ARRAY PROCESSING

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ELECTRICAL ENGINEERING

by

Morteza Shahram

June 2005

The Dissertation of Morteza Shahram
is approved:

Professor Peyman Milanfar, Chair

Professor Benjamin Friedlander

Professor Ali Shakouri

Robert C. Miller
Vice Chancellor for Research and
Dean of Graduate Studies

Contents

List of Figures

Abstract

Statistical and Information-Theoretic Analysis of Resolution in Imaging and Array Processing

by

Morteza Shahram

This work investigates some detection-theoretic, estimation-theoretic and information-theoretic methods to analyze the problem of determining resolution limits in imaging systems. The canonical case study is formulated based on a model of the blurred image of two closely-spaced point sources of unknown brightness. To quantify a measure of resolution in statistical terms, we address the following question: What is the minimum detectable separation between two point sources at a given signal-to-noise ratio (SNR), and for pre-specified probabilities of detection and false alarm?". Furthermore, asymptotic performance analysis for the estimation of the unknown parameters is carried out using the Cramer-Rao bound. Also, we analyze the problem of resolution by computing the Kullback-Leibler distance to further confirm the earlier results and to establish a link between the detection-theoretic approach and Fisher information. To study the effects of variation in point spread function (PSF) and model mismatch, we present a perturbation analysis of the detection problem as well. The proposed analysis methodologies presented are carried out for the general two-dimensional model and general sampling scheme. We consider different sampling scenarios and in particular study the case of under-Nyquist (aliased) images.

The approach we have advocated for determining resolution limits in imaging can be similarly used to develop statistical algorithms and performance limits for resolving sinusoids with nearby frequencies, in the presence of noise. Here the problem is that of distinguishing whether the received signal is a single-frequency sinusoid or a double-frequency sinusoid. We derive a locally optimal detection strategy that can be applied in a stand-alone fashion or as a refinement step for existing spectral estimation methods, to yield improved performance.

Dedicated to my parents Hossein Shahram and Fatemeh Moeen,

my uncle Alireza Moeen,

and my brother Mostafa Shahram.

Acknowledgements

This thesis is the result of three and half years of work supported and encouraged by many people and I have been looking forward to this final moment to express my deep gratitude to each of them.

Prof Peyman Milanfar is the person whom I am indebted the most for completing this work. This work could not have been ever done without his incredible support, constant generosity, and excellent criticisms and insights. He has taught me not only many concepts in signal processing (by his wonderful style of teaching and guidance), but also the methodology to carry out the research, to develop analytic thinking, to apply mathematical tools at the right time and also to present and publish my works as clearly as possible. I am extremely grateful for what he has offered me.

I wish to thank the members of my qualifying exam and Ph.D. reading committee Professors Benjamin Friedlander, Ali Shakouri, Joel Yellin for their constructive comments and suggestions on my thesis proposal and the final dissertation. I am very thankful to Prof Louis Scharf of the Colorado State University, Prof Jeffery Fessler of the University of Michigan and Prof Steven Kay of the University of Rhode Island for their very helpful comments on various aspects of my work. I would also like to express my sincere gratitude to Professors Maria Schonbek, Ralph Abraham, Roberto Manduchi and Hongyun Wang of UCSC, Prof Miki Elad from Technion and Prof Homayoun Hashemi from Sharif University of Technology for their great support and

encouragement at several occasions and teaching me many things.

I had incredible enjoyment during the courses I took with Prof Joel Yellin (Information Theory and Combinatorial Mathematics). His classes were always full of pleasure of exploring solvability of many interesting mathematical problems. I also had fascinating experience attending the Statistical Signal Processing course beautifully taught by Prof Milanfar and also the Spectral Estimation and Multirate Signal Processing courses taught by Prof Mohammad Hasan Bastani and Prof Kambiz Nayebi respectively at Sharif University of Technology.

I am highly indebted to Prof Kambiz Nayebi my supervisor during my M.Sc in Sharif University and during my work at the Signal Corporation. He was the first one who taught me signal processing. I can never forget his enormous support and the extreme joy I always had working with him.

I barely find words to express my gratitude to my bright and exceptional friend Daryoosh Vashaee without whom I could not imagine my present state. Since I arrived at Santa Cruz, he was the only one I could bring my endless troubles to. Sadly enough he could not stay long but I could not be more proud of having him as my friend.

Many thanks to my colleagues at Multi-Dimensional Signal Processing Lab Amyn Poonawala for making this long trip much more enjoyable and for many fruitful discussion we had and (now Dr.) Dirk Robinson for his constant encouragement and the insightful discussions on many problems (I also want to thank his wife Emily Robinson, together they offered their valuable friendship). Many thanks to my other

colleagues Lior Zimet, Davy Odom, Hiro Takeda, and Micheal Charest. I also wish to thank other really wonderful friends at UCSC for their support and encouragement: Ahmet Tekin, Rui Li, Javad Shabani, Srikumar Ramalingam, Chandramouli Balasubramanian and many others. I should mention the delightful moments during our daily (sometimes bi-hourly) tea gatherings which was organized by Srikumar who was responsible for making tea in random ways and Amyn who initiated a random discussion at each time!

I would also like to thank my old friends for the unique joy I always had with them: Roozbeh Behzadi, Ahmad Reza Moghaddam, Hossein Kheshtpaz, Abdolkarim Masoumi, Majid Kashefi, Fraydoun Khazaei and Hossein Hajiagha.

When I first arrived US at 2001, Joe and Deedie Calmes were the ones who were kindly waiting for me at the airport and they hosted me in their beautiful house until I could find a place and settle down. I am sincerely grateful for their kindness and continuing warm friendship. I am also thankful to other people in Santa Cruz for their precious friendship and help: Pam and Don Urfer and also Peggy Pollard. I wish to thank Carol Mullane and Jodi Rieger of the graduate office at the UCSC School of Engineering for helping me on many occasions.

My life in Santa Cruz could be much less enjoyable if it was not for the people who accompanied me for sports activities. Many thanks to John Bardos the UCSC racquetball coach who introduced and amazingly taught me racquetball which surprisingly was strangely missing from my life before! I was also excited to be part of the UCSC Racquetball team. Also, I would like to thank other people whom I

played with: Chelsea George (also a wonderful racquetball instructor), Jerome King (he has been unsuccessfully but constantly trying to correct my English!) and also Amyn, Chandu and Rui for the unforgettable spring season of 2004! I should add to all these the amount of great time I have had playing soccer, table tennis, volleyball and wallyball here. In particular, I want to thank my teammates at the UCSC intermural soccer (indoors/outdoors) league: Ahmet Tekin, Gouxing Wang, Pablo Saz Parkinson, Randall Orr, Carolyn Berger, Cecily Chun, Sara Hall, Peter Lippert and Allison Stedry.

Finally, my life has been constantly fulfilled by love and support of my family. I am extremely grateful to my parents for their love, caring and sacrifices, my uncle Alireza Moeen Rezakhanlou for educating and preparing me for my future life since I was a kid, and my wonderful brother Mostafa for his incredible generosity. This work is dedicated to you.

Santa Cruz, CA,

May 24, 2005,

Morteza Shahram

"But in fact, we know nothing from having seen it; for the truth is hidden in the deep."

- Democritus

"Science may be described as the art of systematic over-simplification."

- Karl Popper

Chapter 1

Introduction

1.1 Information-Theoretic Imaging

The problem of resolution historically has been of significant interest in different communities in science and engineering; for example in astronomy, optics, different applications in physics, array processing and imaging. We focus on the problem of achievable resolution in imaging practice for the following reasons. First, the problem of resolving point sources can be considered as a canonical case study to investigate the performance of imaging systems and image restoration techniques. Second, developing different techniques to resolve point sources is indeed a concern in real-world applications, e.g. astronomy.

Resolution in imaging systems can be interpreted as a fundamental composite estimation/detection problem, and we will explain the detail of such a concept in what follows. Resolvability in imaging is therefore considered and categorized as

an information-theoretic problem [48]. Information-theoretic imaging concerns two classes of analyses: to apply fundamental statistical and information-theoretic principles to imaging and to optimally (or suboptimally) extract information from image data. The first type of analysis includes, for example, exploring fundamental performance bounds on detection and estimation or information retrieval [57].

Information theory and statistical methods have been very successfully studied and applied to communication systems. But imaging problems have benefitted much less from the use of information-theoretic methodologies for two reasons: First, the image computing problems are harder; and second, there have not been all-agreed standards of abstract formulation of the problems in imaging [48]. Nevertheless, in the last decade, information theoretic imaging has received remarkable attention to build connections between the advanced theory of information and imaging, and image processing problems. As pointed out by others earlier [57], the broad interpretation of information-theoretic imaging includes every problem of imaging in which information theory plays a role.

1.2 Resolution Limits in Imaging

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 1.1. In two dimensions, this function is the well-known *Airy* diffraction pattern [22]. 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 [22] and [19].

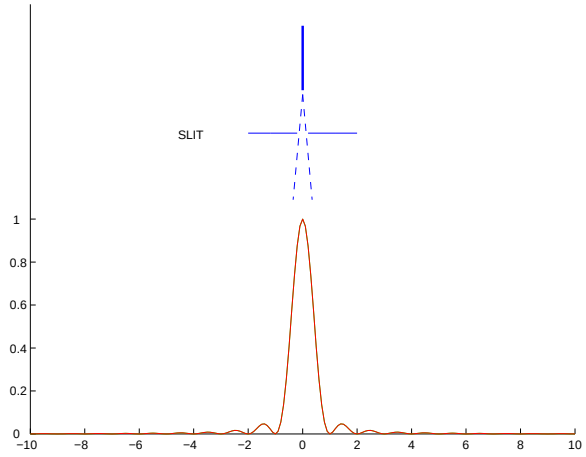

Figure 1.1: Image of point source captured by diffraction-limited imaging

The first part of the research reported in this thesis is concerned with the statistical analysis of the resolution limit in a so-called "diffraction-limited" imaging system. The canonical case study here 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. In other words, we define resolution statistically as the ultimate ability to distinguish whether the captured (noise-corrupted) data originated from one point source or two

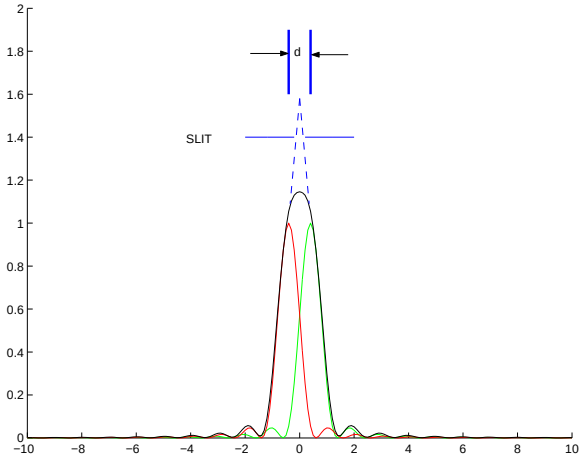

Figure 1.2: Incoherent imaging of two closely located point sources

closely-spaced point sources. The analysis uses tools from statistical detection and estimation theory to specifically find explicit relationships for the minimum detectable distance between two closely-spaced point sources imaged incoherently at a given SNR.

In this chapter, we will first review the basics of a canonical imaging system and identify the signal based on which we will build our analytic framework. After introducing the signal model, the related literature will be discussed.

1.3 Optical Imaging Model

In order to design and study the performance of imaging and image restoration systems, it is essential to characterize the image degradation effects due to the physical parameters of the imaging device. Degradations include noisy measurements, blur, illumination and color imperfection (underexposure, overexposure) and so forth. Especially, accurate image modelling is a key factor in image restoration processing [49, 3]. Fortunately, a deep understanding of optical imaging systems has been devel-

oped using concepts of communication and linear system theories[22]. For example, the effect of imaging components such as apertures, lenses and sensors can be considered as deterministic functions in a linear space-invariant framework.

1.3.1 Blurring by Diffraction: Point Spread Function

A simple schematic of a pinhole imaging system is shown in Figure 1.3. Assuming the rays of light move in a straight line, the output image $I_2(x, y)$ will be a geometric projection of the aperture. But there is in fact no sharp boundary between dark and bright areas in the imaging plane and the observed image is diffusely blurred [49]. Such a system is known as a diffraction-limited (or incoherent) imaging system and the blurring phenomenon is called diffraction. Diffraction is the major cause of blurring in remotely sensed images captured by telescopes, microscopes, infrared or usual cameras [3]. The model of diffraction is mathematically described by a point spread function (PSF) which can be computed by the Fraunhofer far filed approximation of the electromagnetic propagation of the light wave [22]. According to Franhofer approximation, the PSF is modelled by the amplitude square of the Fourier transform of system's aperture $A(\xi, \zeta)$[22]:

$$h(x, y) = \left| \int A(\xi, \zeta) \exp\left(-j\frac{2\pi}{\lambda z}(x\xi + y\zeta)\right) d\xi d\zeta \right|^2 \tag{1.1}$$

where λ is the wavelength of radiation and z is the system focal length (the distance between the aperture and the imaging plane), see Figure 1.3. For instance the PSF of

circular and square apertures have the following forms as shown in Figure 1.4 [22]:

$$\text{Circular:} \quad A(\xi,\zeta) = \Pi\left(\frac{\sqrt{\xi^2+\zeta^2}}{w}\right) \tag{1.2}$$

$$\implies \quad h(x,y) = \left(\frac{2\pi w^2}{\lambda z}\right)^2 \text{jinc}^2\left(\frac{2\pi w\sqrt{x^2+y^2}}{\lambda z}\right) \tag{1.3}$$

$$\text{Square:} \quad A(\xi,\zeta) = \Pi\left(\frac{\xi}{2w_x}\right)\Pi\left(\frac{\zeta}{2w_y}\right) \tag{1.4}$$

$$\implies \quad h(x,y) = \left(\frac{4w_x w_y}{\lambda z}\right)^2 \text{sinc}^2\left(\frac{2w_x x}{\lambda z}\right)\text{sinc}^2\left(\frac{2w_y y}{\lambda z}\right) \tag{1.5}$$

where

$$\Pi(t) = \begin{cases} 1 & t < 1 \\ 0 & t \geq 1 \end{cases} \tag{1.6}$$

$$\text{sinc}(t) = \frac{\sin(\pi t)}{\pi t} \tag{1.7}$$

$$\text{jinc}(t) = \frac{\int_0^{2\pi}\exp(j\theta + jt\cos\theta)d\theta}{2\pi jt} \tag{1.8}$$

Referring to Figure 1.3, the observable image will be obtained by the following integration [42]:

$$I_2(x,y) = \int\int I_1(\xi,\zeta)h\left(x+\frac{z}{z_1}\xi, y+\frac{z}{z_1}\xi\right)d\xi d\zeta \tag{1.9}$$

where $h(x,y)$ is the point spread function and $I_1(x,y)$ and $I_2(x,y)$ denote the input (scene) and output (observation) intensities, respectively. By changing variables in the following form:

$$\xi' = -\frac{z}{z_1}\xi, \tag{1.10}$$

$$\zeta' = -\frac{z}{z_1}\zeta, \tag{1.11}$$

Equation 1.9 can be readily rewritten as a two-dimensional convolution (denoted by $**$) of the input intensity and the point spread function:

$$I_2(x,y) = \left(\frac{z_1}{z}\right)^2 I_1\left(\frac{z_1}{z}x, \frac{z_1}{z}y\right) * *h(x,y) \tag{1.12}$$

Adding a lens to the imaging system as shown in Figure 1.5 will result in:

$$I_2(x,y) = \left(\frac{z_1}{z-z_2}\right)^2 I_1\left(\frac{z_1}{z-z_2}x, \frac{z_1}{z-z_2}y\right) * *h(x,y) \tag{1.13}$$

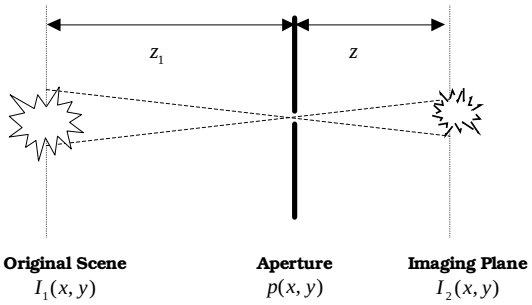

Figure 1.3: Pinhole imaging system

1.3.2 Other Causes of Blurring

Another type of blurring in images is introduced by an out-of-focus lens. The intensity of defocusing is a function of the shape of the aperture and the lens and the distance between the camera and the object. For instance in a circular aperture with radius R, the point spread function due to out-of-focus lens (assuming spatially uniform blur) is given by [3]:

$$p_{def}(x,y) = \begin{cases} \frac{1}{\pi R^2} & \sqrt{x^2+y^2} < R^2 \\ 0 & \sqrt{x^2+y^2} \geq R^2 \end{cases} \tag{1.14}$$

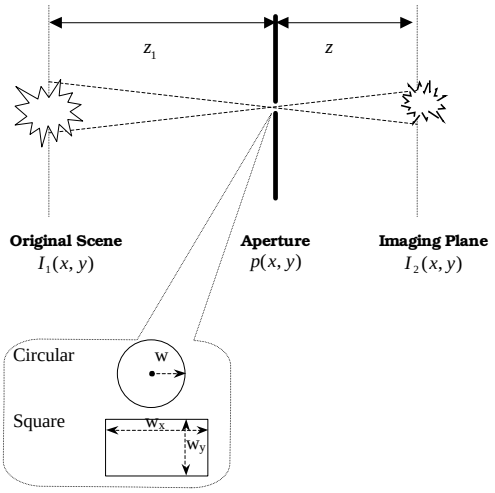

Figure 1.4: Circular and Square Aperture

Also, in underwater or astronomical remote imaging, images may be blurred by a atmospheric or underwater turbulence. Blur in these images in long-term exposures is approximately modelled by a Gaussian function[21, 3]:

$$p_{atm}(x,y) = \frac{1}{\sqrt{2\pi\sigma_{atm}^2}} \exp\left(-\frac{x^2+y^2}{2\sigma_{atm}^2}\right) \tag{1.15}$$

where σ_{atm}^2 parameterizes the amount of blur. In summary, the total blurring effect in an imaging system can be modelled by the convolution of all point spread functions (Figure 1.6):

$$h(x,y) * * p_{atm}(x,y) * * p_{def}(x,y) * * \cdots \tag{1.16}$$

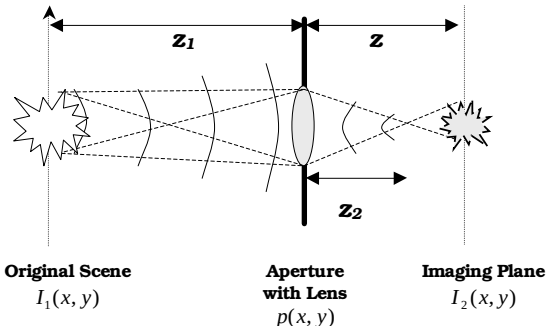

Figure 1.5: Lens in Imaging [19]

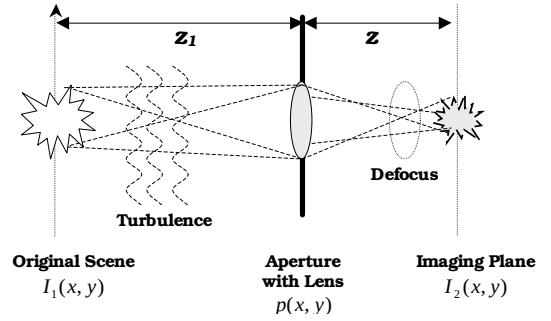

Figure 1.6: Several blurring in imaging[19]

1.3.3 Noise Assumptions

As mentioned before, noise is another type of the degradation in an imaging system. When an optical field reaches a photo-detector, a quantum of energy is absorbed in the form of a photon. The number of photons recorded during a light exposure is a random quantity modelled by a Poisson distribution. However the electric signal generated by charge coupled devices (CCD) or other detectors contains other forms of noise which are mostly caused by thermal noise in the data acquisition process. Thermal noise is modelled by additive zero-mean Gaussian noise. Gaussian

distribution is also used to characterize the behavior of other noises (e.g. photon counting noise) in limiting conditions according to the central-limit theorem [3, 63].

1.3.4 Fundamental Problem: Definition

The direct result of previous sections is that the image of an ideal point source is captured as a spatially extended pattern known as the point-spread function (PSF). This pattern is shown for the one-dimensional case in Figure 1.1. When two closely-located point sources are imaged, the measured signal is the incoherent sum of the respective shifted point spread functions (Figure 1.2). To define the problem in the one-dimensional case, let us assume that the original signal of interest is the sum of two impulse functions with amplitudes $\sqrt{\alpha_1}$ and $\sqrt{\beta_1}$ separated by a small distance d as shown in Figure 1.2:

$$\sqrt{\alpha_1}\delta\left(x-\frac{d}{2}\right)+\sqrt{\beta_1}\delta\left(x+\frac{d}{2}\right) \tag{1.17}$$

From Equation 1.12, given samples at $x_k \quad (k=1,\cdots,N)$ of the measured signal, the measured discrete signal resulting from a slit with length w in one-dimensional case, as seen in Figure 1.2 is [1]:

$$\begin{aligned} g(x) &= s(x_k;\alpha_1,\beta_1,d,w,z,z_1)+w(x_k) && (1.18)\\ &= \left(\frac{z_1}{z}\right)\left(\alpha_1\delta\left(\frac{z_1}{z}x_k-\frac{d}{2}\right)+\beta_1\delta\left(\frac{z_1}{z}x_k+\frac{d}{2}\right)\right)*\left(\frac{w}{\lambda z}\right)^2\operatorname{sinc}^2\left(\frac{wx_k}{\lambda z}\right)\\ &\quad +w(x_k) && (1.19)\\ &= \left(\frac{w^2z_1}{\lambda z^3}\right)\left(\alpha_1\frac{z}{z_1}\delta\left(x_k-\frac{z}{z_1}\frac{d}{2}\right)+\beta_1\frac{z}{z_1}\delta\left(x_k+\frac{z}{z_1}\frac{d}{2}\right)\right)*\operatorname{sinc}^2\left(\frac{wx_k}{\lambda z}\right) \end{aligned}$$

[1] The effect of other blurring function will be considered in next sections.

$$+w(x_k) \tag{1.20}$$

$$= \left(\frac{w}{\lambda z}\right)^2 \left(\alpha_1 \text{sinc}^2\left(\frac{w}{\lambda z}x_k - \frac{w}{\lambda z_1}\frac{d}{2}\right) + \beta_1 \text{sinc}^2\left(\frac{w}{\lambda z}x_k + \frac{w}{\lambda z_1}\frac{d}{2}\right)\right)$$

$$+w(x_k) \tag{1.21}$$

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

$$\mu = \frac{w}{\lambda z}, \tag{1.22}$$

$$\nu = \frac{w}{\lambda z_1}, \tag{1.23}$$

$$\alpha = \left(\frac{w}{\lambda z}\right)^2 \alpha_1 = \mu^2 \alpha_1, \tag{1.24}$$

$$\beta = \left(\frac{w}{\lambda z}\right)^2 \beta_1 = \mu^2 \beta_1, \tag{1.25}$$

Equation 1.21 can be written in the following form

$$g(x_k) = s(x_k; \alpha, \beta, d) + w(x_k) \tag{1.26}$$

$$= \alpha \text{sinc}^2\left(\mu x_k - \nu\frac{d}{2}\right) + \beta \text{sinc}^2\left(\mu x_k + \nu\frac{d}{2}\right) + w(x_k) \tag{1.27}$$

As an example, let us calculate some typical values of the parameters for this imaging scenario. The wavelength of visible light is in the range of 400-900 $\times 10^{-9}$ m. Let the distances ω, z_1 and z be in the range of 1-10 mm, 1-100 m and 5-10 cm, respectively. Then we will have μ in the range 10^4 to 10^6 and ν in the range 10-1000.

Without loss of generality, hereafter we assume that $\mu = 1$ and $\nu = 1$. With the above definition we write the measured image (in one-dimensional case) as

$$g(x_k) = s(x_k; \alpha, \beta, d) + w(x_k) = \alpha h\left(x_k - \frac{d}{2}\right) + \beta h\left(x_k - \frac{d}{2}\right) + w(x_k) \tag{1.28}$$

As we will see later, this model will be generalized in this thesis. For example we will study the 2-D signal model and also asymmetrically located point sources.

For the above model, if the distance between point sources is less than the Rayleigh spacing, 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. 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 [40]. Indeed, at sufficiently high sampling rates, and in the absence of additive noise, arbitrarily small details can be resolved.

1.4 Literature Review

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 [24, 25, 26]. In particular, in [25, 26], he derived lower bounds on the mean-square error of unbiased estimators for the point source positions, the distance between the sources, and the radiance values, using the Cramér-Rao inequality. In [25], 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 work.) 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.

In [40, 39], 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 [39], the analysis of the achievable resolution in deconvolved astronomical images was studied based on a criterion similar to Rayleigh's.

In [8] two-point resolution in imaging systems was studied using a model-fitting theory for coherent and partially coherent imaging, where the probability of resolution (probability of detection) was computed for different models. Also in [2] the Cramér-Rao lower bound formulation was used to study the limits to attainable precision of estimated distances of the two point sources. Assuming a Gaussian PSF, they determined a lower bound for the variance of estimation of the distance. Finally, in [9], the reader can find a very comprehensive review of past and present approaches to the concept of resolution.

Another view of the resolution problem from the information theory perspec-

tive is given in [35]. 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. This paper 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.

Other papers in which some direct applications of information theory to sampled imaging systems have been introduced are [16, 17, 15, 27]. The main feature of these papers is end-to-end (sensor-to-display) analysis of the problem of imaging and image restoration and to unify the design of image gathering and display devices [16]. Huck et al in [27] developed some formulations to study the performance of the imaging process including image gathering, data transmission, and image display. They compute change of information rate for whole system which was claimed to be closely correlated to the fidelity and clarity of final observed images.

The results of our research extend, illuminate, and unify the earlier works in this field using more modern tools in statistical signal processing. As mentioned before, the first part of this work is concerned with the statistical analysis of the ultimate resolving power in incoherent imaging of two closely-located point sources. 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. 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.

We formulate the problem of two-point resolution in terms of statistical estimation/detection. Our approach is to 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.) will be explicitly present in the formulation.

1.5 Organization of the Thesis

To gain maximum intuition and perspective from the foregoing analysis, we will first carry out our detection-theoretic analysis in the one-dimensional case in Chapter 2. We study various cases including unknown and/or unequal intensities, which have not been considered in their full complexity before.

The analysis for the one-dimensional signal model will later be extended to the two-dimensional case in Chapter 3 in which general analytical results are also derived for arbitrary sampling schemes. Our framework in Chapter 3 treats the most general case where all the parameters involved in the signal model are unknown to the detector. Furthermore, we consider two cases, where the value of noise variance (σ^2) is known and where it is unknown. For both cases we develop corresponding detection strategies and obtain the explicit relationship between SNR and the parameters in the

model.

In Chapter 4, the estimation-theoretic and information-theoretic approaches will be discussed. In the interest of completeness, we also compute the Fisher Information matrix (and Cramér-Rao (CR) lower bound) in closed form for two different cases. The Cramér-Rao lower bound formulation is used to study the limits to attainable precision of estimated distances of the two point sources. We carry out this analysis for the case of under-sampled images and for a general PSF. Another appealing and informative analysis is to compute the symmetric Kullback-Leibler Distance (KLD) or Divergence [36, p. 26]. KLD is a measure of discriminating power between two hypotheses, and is directly related to the performance figure of a related optimal detector. To accurately compute the KLD for the underlying problem, we make some essential extensions to the conventional formula of approximating the KLD [36]. We shall see an interesting and important connection between the KLD analysis and the detection-theoretic analysis. A significant question which has not been addressed in the past is to analyze the effect of a known or unknown perturbation of PSF on the detection performance. Variation in PSF can be caused by other blurring elements in the system, for example an out-of-focus lens or atmospheric or underwater turbulence. We first address the problem of finding the change in the required SNR due to a variation of the PSF. This will help us to analyze sensitivity to model inaccuracies. This is the topic of interest in Chapter 5.

Chapter 6 includes an interesting extension of the proposed approach to the filed of spectral analysis and direction finding. The methodology we present for deter-

mining resolution limits in optical imaging is quite similar to that of finding direction of arrivals for incident waves at a receiver array. Finally in Chapter 8, results will be summarized and concluded and future directions will be discussed.

1.6 Important Notations

α and β	point source intensities (brightness)
d	the distance between point sources
$s(x)$ or $s(x,y)$	Respective 1-D or 2-D signal of interest in continuous domain
$s(x_k)$ or $s(x_k,y_l)$	The discretized signal
$\mathbf{s}$	The vector notation of the discretized signal (in 1-D and 2-D models)
$g(x_k)$ or $g(x_k,y_l)$	The discretized measured signal (in 1-D and 2-D models)
$\mathbf{g}$	The vector notation of the discretized measured signal (in 1-D and 2-D models)
θ	The vector of parameters
σ^2	Noise variance
SNR	Signal-to-Noise Ratio
$h(x)$ or $h(x,y)$	The 1-D or 2-D point spread function
$h_i(x)$	The i-th partial derivative of $h(x)$
$\mathbf{h}_i$	The vector notation of $h_i(x)$
$h_{ij}(x,y)$	The i-th partial derivative of $h(x,y)$ with respect to x followed by the j-th partial derivative of $h(x,y)$ with respect to y
$\mathbf{h}_{ij}$	The vector notation of $h_{ij}(x,y)$
E_i	The norm-2 energy of $h_i(x)$
E_{ij}	The norm-2 energy of $h_{ij}(x,y)$
Λ	The Fisher information matrix
$\mathcal{J}$	The Kullback-Leibler distance
$\mathsf{det}(\mathbf{A})$	The determinant of matrix $\mathbf{A}$
$\mathsf{tr}(\mathbf{A})$	The trace of matrix $\mathbf{A}$
$\mathbf{A}^T$	The transpose of matrix $\mathbf{A}$
$\mathbf{A}^{-1}$	The inverse of matrix $\mathbf{A}$

Chapter 2

Detection-Theoretic Approach: One-Dimensional Signal

In this chapter, we focus on the problem of resolution limit for a canonical case of a 1-D imaging scenario. We consider two hypothesis testing problems. First, we mainly concentrate on the problem of detectablity of separation between point sources (spatial super-resolution) and develop detection frameworks for different cases based on the signal parameters being known or unknown to the detector. The second detection problem deals with the problem of resolvability of difference in brightness of closely-spaced point sources (grey-level super-resolution). The latter has application in, for example, determining achievable contrast between adjacent surfaces in thermal imaging. The presented results in this chapter are for the case where the measured image is (unaliased) super-critically and uniformly sampled. These results will be extended to more general sampling schemes in Chapter 3.

2.1 Signal Model

To gain maximum intuition and perspective from the foregoing analysis, all discussion in this chapter will be carried out in the one-dimensional case, which will later (in Chapter 3) 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:

$$\sqrt{\alpha}\delta(x - \frac{d}{2}) + \sqrt{\beta}\delta(x + \frac{d}{2}) \tag{2.1}$$

As mentioned before, the image will be the incoherent sum of two point spread functions[1], resulting from an imaging aperture (or slit in one-dimensional case, as seen in Figure 1.2):

$$s(x; \alpha, \beta, d) = \alpha h\left(x - \frac{d}{2}\right) + \beta h\left(x + \frac{d}{2}\right) \tag{2.2}$$

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

$$g(x_k) \quad = \quad s(x_k; \alpha, \beta, d) + w(x_k) = \alpha h\left(x_k - \frac{d}{2}\right) + \beta h\left(x_k + \frac{d}{2}\right) + w(x_k) \tag{2.3}$$

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

[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 this Chapter.

2.2 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 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 $\mathcal{H}_0$ denote the null hypothesis that $d = 0$ (one peak present) and let $\mathcal{H}_1$ denote the alternate hypothesis that $d > 0$ (two peaks present):

$$\begin{cases} \mathcal{H}_0 & : d = 0 \quad \text{One peak is present} \\ \mathcal{H}_1 & : d > 0 \quad \text{Two peaks are present} \end{cases} \tag{2.4}$$

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

$$\begin{cases} \mathcal{H}_0 : & \mathbf{g} = \mathbf{s}_0 + \mathbf{w} \\ \mathcal{H}_1 : & \mathbf{g} = \mathbf{s} + \mathbf{w} \end{cases} \tag{2.5}$$

where

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

and

$$s(x_k;\alpha,\beta,d) = \alpha h\left(x_k-\frac{d}{2}\right)+\beta h\left(x_k+\frac{d}{2}\right), \tag{2.6}$$

$$s_0(x_k) = s(x_k;\alpha,\beta,d)\,|_{d=0} = (\alpha+\beta)h(x_k). \tag{2.7}$$

This is a problem of detecting a deterministic signal with unknown parameters (α, β, and d, in general). From (2.5), since the probability density function (PDF) under $\mathcal{H}_1$ 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 [33, 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 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 parameter range $0 \le d < 1$.

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 $\mathcal{H}_1$ if

$$L(\mathbf{g}) = \frac{\max\limits_{d} p(\mathbf{g},d,\mathcal{H}_1)}{p(\mathbf{g},\mathcal{H}_0)} = \frac{p(\mathbf{g},\widehat{d},\mathcal{H}_1)}{p(\mathbf{g},\mathcal{H}_0)} > \gamma \tag{2.8}$$

where $\widehat{d}$ denotes the ML estimate of d, and $p(\mathbf{g}, d, \mathcal{H}_1)$ and $p(\mathbf{g}, \mathcal{H}_0)$ are PDFs under $\mathcal{H}_1$ and $\mathcal{H}_0$, respectively. Assuming additive white Gaussian noise (AWGN) with variance σ^2 and $\widehat{\mathbf{s}} = [s(x_1; 1, 1, \widehat{d}), \cdots, s(x_N; 1, 1, \widehat{d})]^T$ we will have:

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

Therefore, $\mathcal{H}_1$ will be chosen if

$$-\|\widehat{\mathbf{s}}\|^2 + 2\mathbf{g}^T(\widehat{\mathbf{s}} - \mathbf{s}_0) > \gamma' \tag{2.9}$$

Equivalently,

$$\begin{aligned} \sum_{k=1}^{N} &- \left[\alpha h\left(x_k - \frac{\widehat{d}}{2}\right) + \beta h\left(x_k + \frac{\widehat{d}}{2}\right)\right]^2 \\ &+ 2\left[\alpha h\left(x_k - \frac{\widehat{d}}{2}\right) + \beta h\left(x_k + \frac{\widehat{d}}{2}\right) - (\alpha + \beta)h(x_k)\right] g(x_k) > \gamma', \end{aligned} \tag{2.10}$$

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

$$\min_d \sum_{k=1}^{N} \left[\alpha h\left(x_k - \frac{d}{2}\right) + \beta h\left(x_k + \frac{d}{2}\right) - g(x_k)\right]^2 \Rightarrow \widehat{d} \tag{2.11}$$

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.

2.3 (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 $d = 0$. 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[2]. More specifically,

$$s(x_k; \alpha, \beta, d) \approx (\alpha + \beta)h(x_k) + \frac{\beta - \alpha}{2} d h_1(x_k) + \frac{\alpha + \beta}{8} d^2 h_2(x_k) \tag{2.12}$$

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

$$h_1(x_k) = \left.\frac{\partial h(x)}{\partial x}\right|_{x=x_k} = \frac{2\sin(\pi x_k)(\sin(\pi x_k) - \pi x_k \cos(\pi x_k))}{\pi^2 x_k^3} \tag{2.13}$$

$$h_2(x_k) = \left.\frac{\partial^2 h(x)}{\partial x^2}\right|_{x=x_k} = \frac{(4\pi^2 x_k^2 - 3)\cos(2\pi x_k) - 4\pi x_k \sin(2\pi x_k) + 3}{2\pi^2 x_k^4} \tag{2.14}$$

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

[2]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 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[33, p. 217] as will be discussed later in Section 2.4.1.

Continuing with vector notation we have:

$$\mathbf{s} \approx (\alpha+\beta)\mathbf{h} + \frac{\beta-\alpha}{2} d\mathbf{h}_1 + \frac{\alpha+\beta}{8} d^2\mathbf{h}_2 \tag{2.15}$$

where

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

Writing in the form of hypotheses described earlier in (2.5):

$$\begin{cases} \mathcal{H}_0: & \tilde{\mathbf{g}} = (\alpha+\beta)\mathbf{h} + \mathbf{w} \\ \mathcal{H}_1: & \tilde{\mathbf{g}} = (\alpha+\beta)\mathbf{h} + \frac{\beta-\alpha}{2} d\mathbf{h}_1 + \frac{\alpha+\beta}{8} d^2\mathbf{h}_2 + \mathbf{w} \end{cases} \tag{2.16}$$

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 as follows:

$$\text{SNR} = \frac{1}{\sigma^2} \left\| (\alpha+\beta)\mathbf{h} + \frac{\beta-\alpha}{2} d\mathbf{h}_1 + \frac{\alpha+\beta}{8} d^2\mathbf{h}_2 \right\|^2 \tag{2.17}$$

For any differentiable symmetric PSF ($h(x)$) and in the case of above-Nyquist sam-

pling, the following relations can be verified[3]

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

Therefore, we can rewrite (2.17) in the following form:

$$\begin{aligned}
\text{SNR} &= \frac{1}{\sigma^2}\left[(\alpha+\beta)^2E_0 + \left(\frac{\beta-\alpha}{2}\right)^2 d^2E_1 + \left(\frac{\alpha+\beta}{8}\right)^2 d^4E_2 + \left(\frac{\alpha+\beta}{2}\right)^2 d^2E_1\right] \\
&= \frac{1}{\sigma^2}\left[(\alpha+\beta)^2E_0 - \alpha\beta d^2E_1 + \left(\frac{\alpha+\beta}{8}\right)^2 d^4E_2\right] \qquad (2.18)
\end{aligned}$$

where we define

$$E_0 = \mathbf{h}^T\mathbf{h} = f_s\int_{-\infty}^{+\infty} h^2(x)dx \qquad (2.19)$$

$$E_1 = \mathbf{h}_1^T\mathbf{h}_1 = f_s\int_{-\infty}^{+\infty}\left[\frac{\partial h(x)}{\partial x}\right]^2 dx \qquad (2.20)$$

$$E_2 = \mathbf{h}_2^T\mathbf{h}_2 = f_s\int_{-\infty}^{+\infty}\left[\frac{\partial^2 h(x)}{\partial x^2}\right]^2 dx \qquad (2.21)$$

as energy terms[4].

---

[3]For instance

$$\mathbf{h}^T\mathbf{h}_1 = \int_{-\infty}^{+\infty} h(x)\frac{\partial h(x)}{\partial x}\mathrm{d}x = \int_{-\infty}^{+\infty}\frac{1}{2}\frac{\partial}{\partial x}[h(x)]^2\mathrm{d}x = \left.\frac{1}{2}h^2(x)\right|_{-\infty}^{+\infty} = 0$$

and

$$\mathbf{h}^T\mathbf{h}_2 + \mathbf{h}_1^T\mathbf{h}_1 = \int_{-\infty}^{+\infty} h(x)\frac{\partial^2 h(x)}{\partial x^2} + \left(\frac{\partial h(x)}{\partial x}\right)^2\mathrm{d}x = \int_{-\infty}^{+\infty}\frac{\partial}{\partial x}h(x)\frac{\partial h(x)}{\partial x}\mathrm{d}x = \left.h(x)\frac{\partial h(x)}{\partial x}\right|_{-\infty}^{+\infty} = 0$$

since the PSF has finite energy and $\lim_{x\to\pm\infty} h(x,y) = 0$

[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 $h(x) = \text{sinc}^2(x)$.

2.4 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. It is illuminating to study the various cases of interest in order. Our first assumptions are: 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 (2.16), 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

[5] where point sources are located at $-d_1$ and $+d_2$ instead of $-\frac{d}{2}$ and $\frac{d}{2}$.

independent from d. Therefore we may simplify further:

$$\begin{cases} \mathcal{H}_0: & \mathbf{y} = \mathbf{w} \\ \mathcal{H}_1: & \mathbf{y} = \frac{\beta-\alpha}{2} d\mathbf{h}_1 + \frac{\beta+\alpha}{8} d^2\mathbf{h}_2 + \mathbf{w} \end{cases} \tag{2.22}$$

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.

2.4.1 The Case of Equal Intensities , Symmetrically Located Point Sources

When $\alpha = \beta = 1$, (2.22) is reduced to:

$$\begin{cases} \mathcal{H}_0: & \mathbf{y} = \mathbf{w} \\ \mathcal{H}_1: & \mathbf{y} = \frac{d^2}{4}\mathbf{h}_2 + \mathbf{w} \end{cases} \tag{2.23}$$

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

$$\widehat{d^2} = 4(\mathbf{h}_2^T\mathbf{h}_2)^{-1}\mathbf{h}_2^T\mathbf{y} \tag{2.24}$$

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

$$T(\mathbf{y}) = \frac{1}{\sigma^2}(\mathbf{h}_2^T\mathbf{h}_2)^{-1}\left(\mathbf{h}_2^T\mathbf{y}\right)^2 \tag{2.25}$$

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 [33, p. 194]. Therefore, the above test-statistic can be simply replaced by:

$$T'(\mathbf{y}) = \sqrt{T(\mathbf{y})} = \sqrt{\frac{1}{\sigma^2}(\mathbf{h}_2^T\mathbf{h}_2)^{-1}}\left(\mathbf{h}_2^T\mathbf{y}\right) \tag{2.26}$$

For any given data set $\mathbf{y}$, we decide $\mathcal{H}_1$ if the statistic exceeds a specified threshold:

$$T'(\mathbf{y}) > \gamma \tag{2.27}$$

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 [33, p. 254]:

$$P_d = Q\left(Q^{-1}(P_f) - \sqrt{\eta}\right) \tag{2.28}$$

where

$$\eta = \frac{d^2}{4}\sqrt{\frac{E_2}{\sigma^2}} \tag{2.29}$$

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 [33, p. 20]. A particularly intriguing and useful relationship is the behavior of the smallest peak

[6] In [8, 65] 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.

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 (2.18), (2.28) and (2.29), the relation between d_{min} and required SNR can be made explicit

$$\begin{aligned} \text{SNR} &= (Q^{-1}(P_f) - Q^{-1}(P_d))^2 \frac{64E_0 - 16d^2E_1 + d^4E_2}{d^4E_2} && (2.30) \\ &= (Q^{-1}(P_f) - Q^{-1}(P_d))^2 \left(\frac{64E_0}{E_2}\frac{1}{d^4} - \frac{16E_1}{E_2}\frac{1}{d^2} + 1 \right). && (2.31) \end{aligned}$$

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 $h(x) = \text{sinc}^2(x)$ and for the specified choice of $P_d = 0.99$ and $P_f = 10^{-6}$, if we collect N samples at x_k within the interval $[-10, 10]$, at just above the Nyquist rate, we have

$$\begin{aligned} \text{SNR} &= 50.12 \frac{\frac{140}{\pi^4} - \frac{14}{\pi^2}d^2 + d^4}{d^4} \\ &= \frac{72.04 - 71.1d^2 + 50.12d^4}{Nd^4} \end{aligned}$$

A plot of this function is shown in Figure 2.1. It is worth noting that in (2.31), the term involving d^{-4} dominates for small d. Therefore, a reasonably informative (but approximate) way to write SNR is

$$\text{SNR} \approx (Q^{-1}(P_f) - Q^{-1}(P_d))^2 \frac{E_0}{E_2}\ \frac{1}{d^4} = \frac{c}{Nd^4}, \qquad (2.32)$$

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 (2.32) as

follows:

$$c = 64 \left(Q^{-1}(P_f) - Q^{-1}(P_d) \right)^2 \frac{\int_{-\infty}^{+\infty} h^2(x) dx}{\int_{-\infty}^{+\infty} \left[\frac{\partial^2 h(x)}{\partial x^2} \right]^2 dx} \tag{2.33}$$

A plot of the approximate expression in (2.32) is also shown in Figure 2.1, to be compared against the exact expression (2.31). The above relation (2.32) 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.

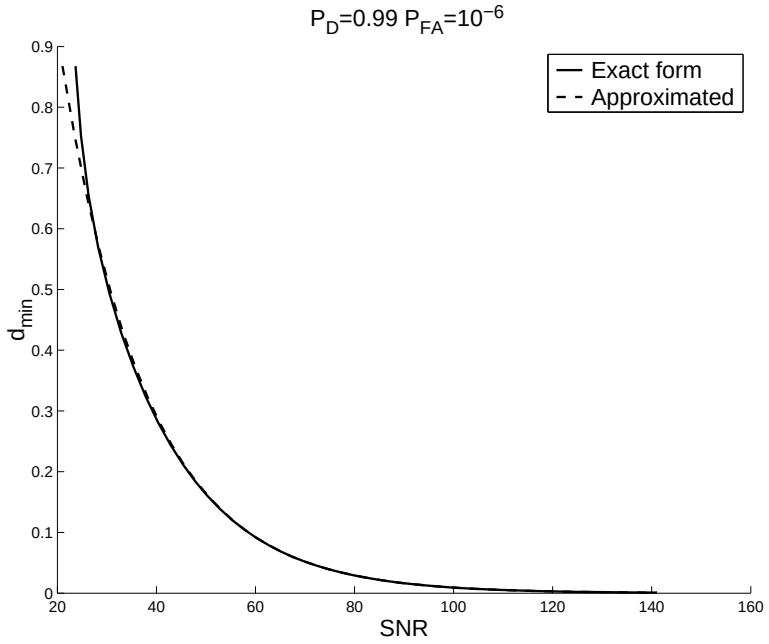

Figure 2.1: Minimum detectable d as a function of SNR (in dB) at the Nyquist rate (exact and approximate)

2.4.2 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[7]. Equation 2.22 leads to a detection problem defined in terms of a linear model over the parameter set θ defined as follows:

$$\mathbf{y} = \mathbf{H}\theta + \mathbf{w} \tag{2.34}$$

$$\mathbf{H} = [\mathbf{h}_1,\ \mathbf{h}_2] \tag{2.35}$$

$$\theta = \begin{bmatrix} d(\alpha - \beta) \\ \dfrac{d^2}{4} \end{bmatrix}, \tag{2.36}$$

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

$$\begin{cases} \mathcal{H}_0 : & \mathbf{A}\theta = \mathbf{b} \\ \mathcal{H}_1 : & \mathbf{A}\theta \neq \mathbf{b} \end{cases} \tag{2.37}$$

where

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{2.38}$$

The GLRT for this problem is given by [33, p. 274]:

$$T(\mathbf{y}) = \frac{1}{\sigma^2}\widehat{\theta}^T \mathbf{A}^T \left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1} \mathbf{A}\widehat{\theta} \tag{2.39}$$

$$= \frac{1}{\sigma^2}\left(\frac{\left(\mathbf{h}_1^T\mathbf{y}\right)^2}{E_1} + \frac{\left(\mathbf{h}_2^T\mathbf{y}\right)^2}{E_2}\right) \tag{2.40}$$

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

where

$$\widehat{\theta} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}\mathbf{y} \tag{2.41}$$

The performance of this detector is characterized by:

$$\begin{aligned} P_f &= Q_{\chi_2^2}(\gamma) && (2.42)\\ P_d &= Q_{\chi_2'^2(\lambda)}(\gamma) && (2.43)\\ \lambda &= \frac{1}{\sigma^2}\theta^T\mathbf{A}^T\left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1}\mathbf{A}\theta && (2.44)\\ &= \frac{1}{\sigma^2}\left(\left(\frac{\alpha-\beta}{2}\right)^2 d^2E_1 + \frac{1}{16}d^4E_2\right) && (2.45) \end{aligned}$$

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 non-central Chi-Squared PDF with 2 degrees of freedom and non-centrality parameter λ. In order to perform the same analysis as Section 2.4.1 (i.e. d_{min} vs. 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 (2.18), we will have:

$$\text{SNR} = \lambda(P_f, P_d)\frac{64E_0 - 16\alpha\beta d^2E_1 + d^4E_2}{4\left(\alpha-\beta\right)^2 d^2E_1 + d^4E_2} \tag{2.46}$$

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

$$\text{SNR} = 56.29\frac{\frac{140}{\pi^4} - \frac{14}{\pi^2}\alpha\beta d^2 + d^4}{\frac{7}{2\pi^2}(\alpha-\beta)^2d^2 + d^4} \tag{2.47}$$

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 Figure 2.2 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.

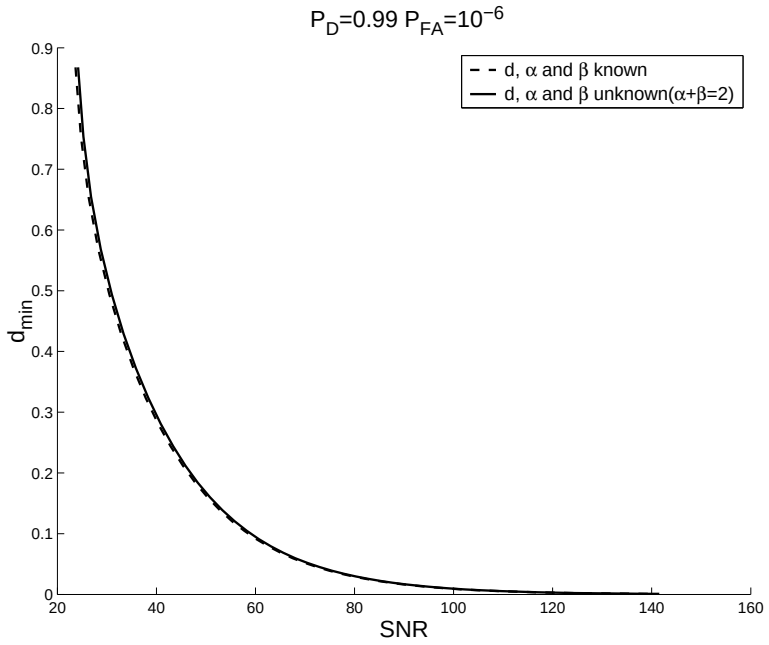

Figure 2.2: $d_{\min}$ versus SNR(dB) for $\alpha = \beta = 1$

An interesting observation arises from a comparison of the minimum detectable d for the cases $\alpha = \beta$ and $\alpha \neq \beta$, shown in Figure 2.3. 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 results 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 signal information and better decision is made possible. This phenomenon is a result of the assumption of symmetry of point sources around the known origin ($x = 0$). If the center of point sources is not known, the results can be different, as we will explained in the next section.

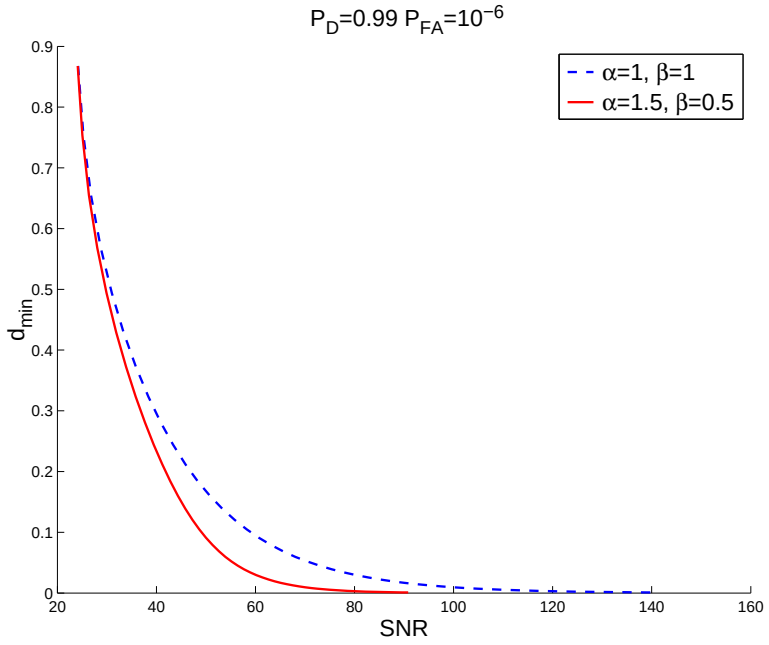

Figure 2.3: GLRT for $\alpha \neq \beta$ and the case $\alpha = \beta$, symmetric sources; $d_{\min}$ versus SNR(dB)

2.4.3 The case of unknown intensities but $\alpha + \beta = 2$, asymmetrically located point sources

With the earlier machinery in place, in this section, we study the case when the point sources are not located symmetrically around the origin ($x = 0$). We consider

the following model for this case:

$$g(x_k) \quad = \quad s(x_k;\alpha,\beta,d_1,d_2)+w(x_k)=\alpha h\left(x_k-d_1\right)+\beta h\left(x_k+d_2\right)+w(x_k) \tag{2.48}$$

where d_1 and d_2 are unknown and $d = d_1 + d_2$ is the distance between point sources. The Taylor expansion for the signal term in (2.48) around $(d_1, d_2) = (0, 0)$ is given by:

$$s(x_k;\alpha,\beta,d_1,d_2)=(\alpha+\beta)h(x_k)+(-\alpha d_1+\beta d_2)h_1(x_k)+\frac{\alpha d_1^2+\beta d_2^2}{2}h_2(x_k) \tag{2.49}$$

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:

$$\begin{cases} \mathcal{H}_0: & [d_1\ d_2]=[0\ 0] \\ \mathcal{H}_1: & [d_1\ d_2]\neq[0\ 0] \end{cases} \tag{2.50}$$

or equivalently

$$\begin{cases} \mathcal{H}_0: & \tilde{g}(x_k)=(\alpha+\beta)h(x_k)+w(x_k) \\ \mathcal{H}_1: & \tilde{g}(x_k)=(\alpha+\beta)h(x_k)+(-\alpha d_1+\beta d_2)h_1(x_k)+\frac{\alpha d_1^2+\beta d_2^2}{2}h_2(x_k)+w(x_k) \end{cases} \tag{2.51}$$

By removing the known common term $(\alpha + \beta)h(x_k)$, the signal can be expressed as the following linear model:

$$\mathbf{y} \quad = \quad \mathbf{H}\theta_a+\mathbf{w}$$

where

$$\begin{aligned} \mathbf{H} &= [\mathbf{h}_1,\ \mathbf{h}_2] \\ \theta_a &= \begin{bmatrix} -\alpha d_1+\beta d_2 \\ \dfrac{\alpha d_1^2+\beta d_2^2}{2} \end{bmatrix} \end{aligned} \tag{2.52}$$

where the subscript *"a"* an θ_a is denoting the asymmetric case, to be distinguished from (2.36). Then, the corresponding hypotheses are given by

$$\begin{cases} \mathcal{H}_0: & \mathbf{A}\theta_a = \mathbf{b} \\ \mathcal{H}_1: & \mathbf{A}\theta_a \neq \mathbf{b} \end{cases} \tag{2.53}$$

where

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

just as in Section 2.4.2. The GLRT for (2.53) will be:

$$T(\mathbf{y}) = \frac{1}{\sigma^2}\left(\frac{\left(\mathbf{h}_1^T\mathbf{y}\right)^2}{E_1} + \frac{\left(\mathbf{h}_2^T\mathbf{y}\right)^2}{E_2}\right) \tag{2.54}$$

From (2.54), the performance of this detector is characterized by:

$$\begin{aligned} P_f &= Q_{\chi_2^2}(\gamma) \\ P_d &= Q_{\chi_2'^2(\lambda)}(\gamma) \\ \lambda &= \frac{1}{\sigma^2}\left(\left(-\alpha d_1 + \beta d_2\right)^2 E_1 + \left(\frac{\alpha d_1^2 + \beta d_2^2}{2}\right)^2 E_2\right) \end{aligned} \tag{2.55}$$

Now, to obtain the relation between SNR and (d_1, d_2), we first need to compute SNR for the model of (2.48), which is given by:

$$\text{SNR} = \frac{1}{\sigma^2}\left[(\alpha+\beta)^2 E_0 - \alpha\beta(d_1+d_2)^2 E_1 + \left(\frac{\alpha d_1^2 + \beta d_2^2}{2}\right)^2 E_2\right] \tag{2.56}$$

The value of σ^2 in (2.55) can be obtained for desired P_d and P_f. By substituting this value in (2.56) we will have:

$$\text{SNR} = \lambda(P_f, P_d)\frac{(\alpha+\beta)^2 E_0 - \alpha\beta(d_1+d_2)^2 E_1 + \left(\frac{\alpha d_1^2+\beta d_2^2}{2}\right)^2 E_2}{\left(-\alpha d_1 + \beta d_2\right)^2 E_1 + \left(\frac{\alpha d_1^2+\beta d_2^2}{2}\right)^2 E_2} \tag{2.57}$$

In order to present the results in this case, let us assume that[8] $\alpha d_1 \approx \beta d_2$ (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 (2.55) 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. Figure 2.4 confirms this result by showing the curves of d_{min} vs. SNR for two cases: equal intensities and unequal intensities (we consider here the case of $h(x) = \text{sinc}^2(x)$). By comparing this result 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 detection performance. Also, it is worth mentioning that with the assumption of $\alpha d_1 \approx \beta d_2$, we can approximate (2.57) for the range of small d_1 and d_2 in the following informative ways:

$$\text{SNR} \approx \lambda(P_f, P_d) \frac{4(\alpha+\beta)^2}{\left(\alpha d_1^2 + \beta d_2^2\right)^2} \frac{E_0}{E_2} = \lambda(P_f, P_d) \frac{4}{d_1^2 d_2^2} \frac{E_0}{E_2} = \lambda(P_f, P_d) \frac{4(\alpha+\beta)^4}{\alpha^2 \beta^2 d^4} \frac{E_0}{E_2} \quad (2.58)$$

More interestingly, one can verify that if the value of $\alpha + \beta$ and the condition $\alpha d_1 = \beta d_2$ (which is quite achievable in practice) is known to the detector, the GLRT framework will result in uniformly most powerful test. The reason for this case is that the first element in (2.52) vanishes and the second element is always positive resulting in optimality.

[8] See Appendix C for a justification

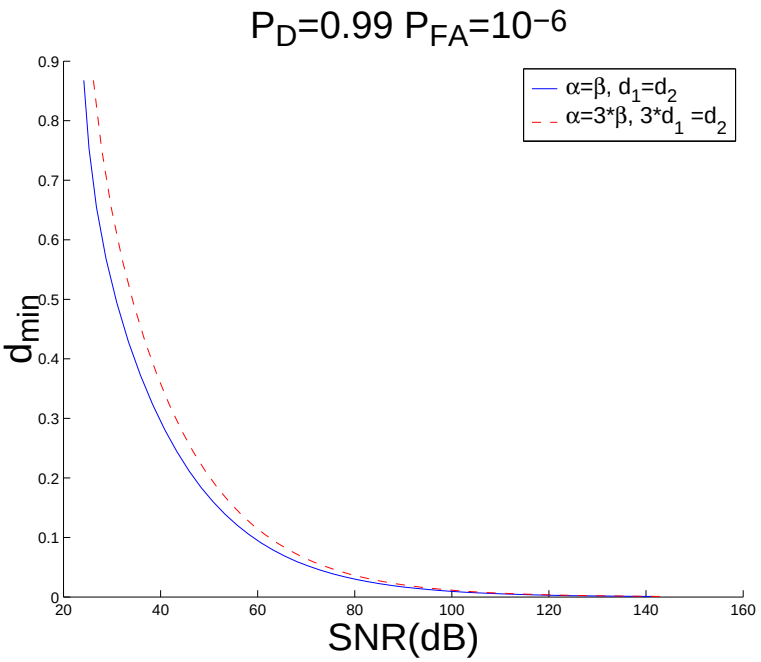

Figure 2.4: $d_{\min}$ versus SNR(dB); $d = d_1 + d_2$ and $\alpha d_1 = \beta d_2$; equal intensities and unequal intensities

2.4.4 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 (2.51), we can set up another linear model as follows:

$$\tilde{\mathbf{g}} \quad = \quad \mathbf{H}_u \theta_u + \mathbf{w}$$

where

$$\mathbf{H}_u \quad = \quad [\mathbf{h},\ \mathbf{h}_1,\ \mathbf{h}_2]$$

$$\theta_u = \begin{bmatrix} \alpha+\beta \\ -\alpha d_1+\beta d_2 \\ \dfrac{\alpha d_1^2+\beta d_2^2}{2} \end{bmatrix}, \tag{2.59}$$

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

$$\begin{cases} \mathcal{H}_0: & \mathbf{A}_u\theta_u = \mathbf{b} \\ \mathcal{H}_1: & \mathbf{A}_u\theta_u \neq \mathbf{b} \end{cases} \tag{2.60}$$

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 (2.60) will be:

$$T(\tilde{\mathbf{g}}) = \frac{1}{\sigma^2}\left(\frac{\left(\mathbf{h}_1^T\tilde{\mathbf{g}}\right)^2}{E_1} + \frac{\left(E_2\mathbf{h}^T\tilde{\mathbf{g}} + E_0\mathbf{h}_2^T\tilde{\mathbf{g}}\right)^2}{E_0\left(E_0E_2 - E_1^2\right)}\right) \tag{2.61}$$

The performance of this detector is given by[9]:

$$\begin{aligned} P_f &= Q_{\chi_2^2}(\gamma) \\ P_d &= Q_{\chi_2'^2(\lambda)}(\gamma) \\ \lambda &= \frac{1}{\sigma^2}\left(\left(-\alpha d_1+\beta d_2\right)^2 E_1 + \left(\frac{\alpha d_1^2+\beta d_2^2}{2}\right)^2\left(E_2 - \frac{E_1^2}{E_0}\right)\right) \end{aligned} \tag{2.62}$$

Consequently the relation between d_1 and d_2 vs. SNR is given by:

$$\mathrm{SNR} = \lambda(P_f, P_d)\frac{(\alpha+\beta)^2E_0 - \alpha\beta(d_1+d_2)^2E_1 + \left(\frac{\alpha d_1^2+\beta d_2^2}{2}\right)^2 E_2}{\left(-\alpha d_1+\beta d_2\right)^2 E_1 + \left(\frac{\alpha d_1^2+\beta d_2^2}{2}\right)^2\left(E_2 - \frac{E_1^2}{E_0}\right)} \tag{2.63}$$

[9]Note that according to the Cauchy-Schwarz inequality $E_0E_2 \geq E_1^2$ always.

By comparing (2.57) and (2.63), it can be readily shown that because of the negative term $-\frac{E_1^2}{E_0}$, the detector without the knowledge of $\alpha+\beta$ performs more poorly than the detector which knows $\alpha+\beta=2$. Figure 2.5 displays the performance of these two different detectors in terms of the minimum detectable d vs SNR for the example of $h(x)$ =sinc$^2(x)$.

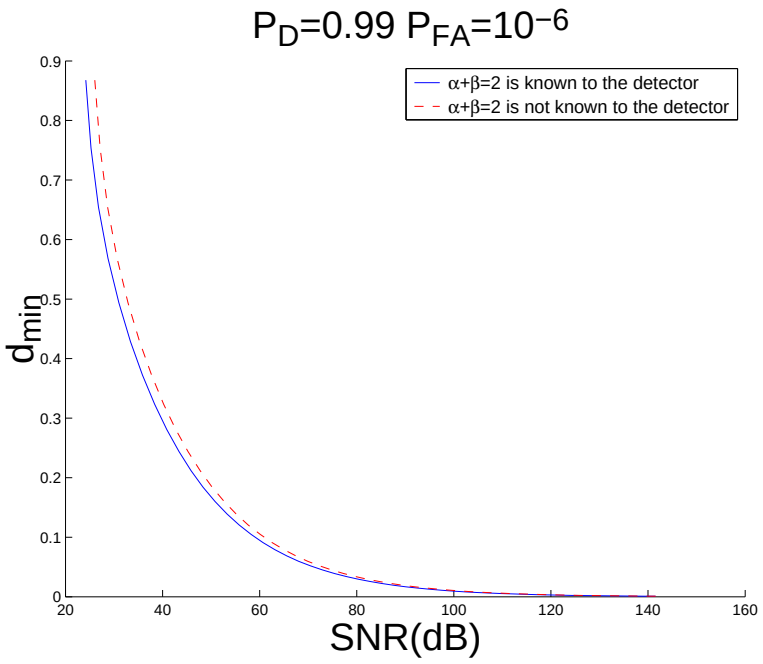

Figure 2.5: $d_{\min}$ versus SNR(dB); $d=d_1+d_2$ and $\alpha d_1=\beta d_2$ detectors with and without the assumption of $\alpha+\beta=2$

2.5 Resolvability in Difference of Brightness

In previous sections we discussed the minimum detectable distance d between two point sources. In this section, we are interested in carrying out the same analysis to determine the minimum detectable difference in brightness $(\alpha-\beta)$; which corresponds

to the following hypothesis test:

$$\begin{cases} \mathcal{K}_0: & \alpha - \beta = 0 \\ \mathcal{K}_1: & \alpha - \beta \neq 0 \end{cases} \tag{2.64}$$

where $\alpha - \beta$ is unknown to the detector. We present our approach here for the simpler case where point sources are located symmetrically and $\alpha + \beta = 2$ is known to the detector. We note that the Taylor expansion in (2.12) is directly applicable for this problem as well, since the unknown variable $\alpha - \beta$ to be decided whether it is zero or not in (2.64) appears as the variable around which the said Taylor expansion is written. Therefore we reconfigure this hypothesis test as

$$\begin{cases} \mathcal{K}_0: & \tilde{\mathbf{g}} = (\alpha + \beta)\mathbf{h} + O(d^2) + \mathbf{w} \\ \mathcal{K}_1: & \tilde{\mathbf{g}} = (\alpha + \beta)\mathbf{h} + \frac{\beta - \alpha}{2} d\mathbf{h}_1 + O(d^2) + \mathbf{w} \end{cases} \tag{2.65}$$

We note that the first-order approximation here will be sufficient to capture the difference between point sources. This first-order approximation is shown to be accurate for this specific problem, since point sources are closely located and also the difference between intensities is small, resulting in a very small $\frac{\beta - \alpha}{2} d$.

Similar to the solution proposed in Section 2.4.1, the following approximate relationship can be derived between the minimum detectable $\alpha - \beta$ and SNR:

$$\text{SNR} \approx \frac{c_1}{N d^2 (\alpha - \beta)^2} \tag{2.66}$$

in which c_1 is a function of the selected P_f and P_d. Therefore as $\alpha \to \beta$ or $d \to 0$, the required SNR grows with a power law.

Finally, we are interested to combine the hypothesis tests in (2.4) and (2.64) into a single multiple hypothesis test, i.e.

$$\begin{cases} \mathcal{H}_0: & \alpha - \beta = 0 \text{ and } d = 0 \\ \mathcal{H}_{10}: & \alpha - \beta = 0 \text{ and } d > 0 \\ \mathcal{H}_{11}: & \alpha - \beta \neq 0 \text{ and } d > 0 \end{cases} . \tag{2.67}$$

What we call here the sequential approach first performs the test for (2.4) which is to decide whether $d = 0$ or not, and then decides between $\mathcal{K}_0$ and $\mathcal{K}_1$. We present a proof here to show that the sequential approach is equivalent to a GLRT directly proposed for the multiple hypothesis testing problem in (2.67). Let $\prec_\gamma$ denote the ordering in GLR sense. For example we write

$$p(\mathbf{g}|\mathcal{H}_0) \prec_\gamma p(\mathbf{g}|\mathcal{H}_1) \tag{2.68}$$

if

$$\frac{p(\mathbf{g}|\mathcal{H}_1)}{p(\mathbf{g}|\mathcal{H}_0)} > \gamma. \tag{2.69}$$

Now we can see that if the above ordering holds meaning $d > 0$, then the sequential test is optimal. The reason is that the acceptance regions for $\mathcal{H}_{10}$ and $\mathcal{H}_{11}$ are the complementary subsets of the acceptance region for $\mathcal{H}_1$. Also we note that

$$\frac{p(\mathbf{g}|\mathcal{H}_1)}{p(\mathbf{g}|\mathcal{H}_0)} = \frac{p(\mathbf{g}|d>0)}{p(\mathbf{g}|d=0)} \geq \frac{p(\mathbf{g}|d>0, \alpha-\beta=0)}{p(\mathbf{g}|d=0)} = \frac{p(\mathbf{g}|\mathcal{H}_{10})}{p(\mathbf{g}|\mathcal{H}_0)} \tag{2.70}$$

and[10]

$$\frac{p(\mathbf{g}|\mathcal{H}_1)}{p(\mathbf{g}|\mathcal{H}_0)} = \frac{p(\mathbf{g}|d>0)}{p(\mathbf{g}|d=0)} \geq \frac{p(\mathbf{g}|d>0, \alpha-\beta\neq 0)}{p(\mathbf{g}|d=0)} \frac{p(\mathbf{g}|\mathcal{H}_{11})}{p(\mathbf{g}|\mathcal{H}_0)} \tag{2.71}$$

[10]because $\alpha - \beta \neq 0$ only constrains the signal space, so in this case the PDF of signal can not be greater.

Hence, if $p(\mathbf{g}|\mathcal{H}_1) \prec_\gamma p(\mathbf{g}|\mathcal{H}_0)$ (it is decided that $d = 0$), we can show that

$$p(\mathbf{g}|\mathcal{H}_{10}) \prec_\gamma p(\mathbf{g}|\mathcal{H}_0) \tag{2.72}$$

$$p(\mathbf{g}|\mathcal{H}_{11}) \prec_\gamma p(\mathbf{g}|\mathcal{H}_0), \tag{2.73}$$

too. This again proves the optimality of the sequential test (in GLRT sense).

2.6 Concluding Remarks

We have set out in this chapter 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? Based on different assumption 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 major conclusion 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 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 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.

The analysis has been carried out in one dimension to facilitate the presentation and to yield maximum intuition. In the next chapter we carry out the analysis for the general 2-D model and also under-Nyquist and non-uniform sampling schemes.

Chapter 3

Detection-Theoretic Approach: Extension to two-Dimensional and Under-Sampled Signals

3.1 Introduction

Having studied the problem with a one-dimensional signal model and uniform over-Nyquist sampling scheme in Chapter 2, we now present an extension to a more general model including two-dimensional scenario and arbitrary non-uniformly-sampled and/or sub-sampled images. To begin, let the original image of interest consist of two impulse functions positioned at points (p_x, p_y) and (q_x, q_y), a small distance of $d = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2}$ apart. That is, the signal model is

$$I(x, y) = \sqrt{\alpha}\delta(x - p_x, y - p_y) + \sqrt{\beta}\delta(x + q_x, y + q_y). \tag{3.1}$$

We consider the following two-dimensional model for the measured (discrete) signal:

$$\begin{aligned} g(x_k, y_l) &= s(x_k, y_l) + w(x_k, y_l) && (3.2) \\ &= \alpha h(x_k - p_x, y_l - p_y) + \beta h(x_k + q_x, y_l + q_y) + w(x_k, y_l) && (3.3) \\ & \quad k, l \in \{1, 2, ..., N\} \end{aligned}$$

where $h(x, y)$ is the blurring kernel[1] (representing the overall point-spread-function (PSF) of the imaging system) and $w(x_k, y_l)$ is assumed to be a zero-mean Gaussian white noise process with variance σ^2.

The statistical analysis is similarly formulated based on the ability to distinguish whether the measured image is generated by one point source or two point sources. This can be posed as a hypothesis testing problem, i.e.

$$\begin{cases} \mathcal{H}_0 : d = 0 & \text{One point source} \\ \mathcal{H}_1 : d > 0 & \text{Two point sources} \end{cases}, \tag{3.4}$$

or equivalently

$$\begin{cases} \mathcal{H}_0 : g(x_k, y_l) = (\alpha + \beta) h(x_k, y_l) + w(x_k, y_l) \\ \mathcal{H}_1 : g(x_k, y_l) = \alpha h(x_k - p_x, y_l - p_y) + \beta h(x_k + q_x, y_l + q_y) + w(x_k, y_l) \end{cases}. \tag{3.5}$$

3.2 Detection-Theoretic Approach

In the test (3.5), when the model parameters are unknown, the probability density function (PDF) under both hypothesis is therefore not known exactly, resulting in a composite hypothesis testing problem. As a common alternative, we use the GLRT

[1]For convenience and ease of presentation only, we assume that the point spread function is a symmetric function throughout this chapter, i.e. $h(x, y) = h(-x, y) = h(x, -y) = h(-x, -y)$.

framework. As will be shown later, the performance of such a detector is very close to that of an ideal detector, to which the values of all the parameters in the model are known. Hence, the performance of the suggested detector can be reasonably considered as an approximate performance bound in practice.

Applying the GLRT approach to the problem of interest directly will produce a highly nonlinear test statistic (see Chapter 2). However, since the range of interest for the value of d is assumed to be small (below the Rayleigh limit[2]), similar to the one-dimensional case, we can benefit from approximating the model of the signal for nearby point sources. The approximate model is obtained by expanding the signal in a Taylor series about the small parameter values around $(p_x, q_x, p_y, q_y) = (0, 0, 0, 0)$. By introducing the partial derivatives of $h(x, y)$ as

$$h_{ij}(x_k, y_l) \quad = \quad \left. \frac{\partial^{i+j} h(x, y)}{\partial x^i \partial y^j} \right|_{x=x_k, y=y_l}, \tag{3.6}$$

we write the signal model in the following (lexicographically scanned) vector form (e.g. $[\mathbf{h}]_{lN+k} = h(x_k, y_l)$):

$$\tilde{\mathbf{s}} = \mathbf{H}\boldsymbol{\theta} \tag{3.7}$$

where

$$\mathbf{H} \quad = \quad [\mathbf{h},\ \mathbf{h}_{10},\ \mathbf{h}_{01},\ \mathbf{h}_{20},\ \mathbf{h}_{02},\ \mathbf{h}_{11}] \tag{3.8}$$

[2]Throughout this chapter, we assume without loss of generality that $d = 1$ corresponds to the Rayleigh limit.

$$\boldsymbol{\theta} = \begin{bmatrix} \alpha + \beta \\ -\alpha p_x + \beta q_x \\ -\alpha p_y + \beta q_y \\ \frac{1}{2}(\alpha p_x^2 + \beta q_x^2) \\ \frac{1}{2}(\alpha p_y^2 + \beta q_y^2) \\ \alpha p_x p_y + \beta q_x q_y \end{bmatrix}. \tag{3.9}$$

We elect to keep terms up to order 2 of the above Taylor expansion which gives a more accurate representation of the signal and avoids trivial approximations in cases where the first order terms would simply vanish. This approximation leads to a linear detection problem and also is equivalent to the framework of locally most powerful tests [33, p. 218].

We first consider the case where the noise variance is known to the detector. By substituting the above approximated model into (3.5), the hypothesis test can be rewritten as

$$\begin{cases} \mathcal{H}_0 : & \mathbf{A}\boldsymbol{\theta} = \mathbf{0} \\ \mathcal{H}_1 : & \mathbf{A}\boldsymbol{\theta} \neq \mathbf{0} \end{cases} \tag{3.10}$$

where

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{3.11}$$

The test (3.10) is a problem of detecting a deterministic signal with unknown parameters. The generalized likelihood ratio test (GLRT) for the approximated model yields [33, p. 274]:

$$T(\mathbf{g}) = \frac{1}{\sigma^2}\widehat{\boldsymbol{\theta}}^T\mathbf{A}^T\left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1}\mathbf{A}\widehat{\boldsymbol{\theta}} \tag{3.12}$$

where

$$\widehat{\boldsymbol{\theta}} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{g} \tag{3.13}$$

is the unconstrained maximum likelihood estimate of $\boldsymbol{\theta}$. $\mathbf{g}$ is the (lexicographically scanned) vector form of the measured signal ($[\mathbf{g}]_{lN+k} = g(x_k, y_l)$). For any given data set $\mathbf{g}$, we decide $\mathcal{H}_1$ if the statistic exceeds a specified threshold,

$$T(\mathbf{g}) > \gamma. \tag{3.14}$$

It is worth mentioning that since the hypothesis test in (3.4) is a one-sided test, the above formulations (the Taylor approximation and the generalized likelihood ratio setup for the problem in (3.10)) can be viewed as a locally most powerful detector [33, p. 218]. From (3.12), the performance of this detector is characterized by

$$P_f = Q_{\chi_5^2}(\gamma) \tag{3.15}$$

$$P_d = Q_{\chi_5'^2(\lambda)}(\gamma) \tag{3.16}$$

$$\lambda = \frac{1}{\sigma^2}\boldsymbol{\theta}^T\mathbf{A}^T\left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1}\mathbf{A}\boldsymbol{\theta}, \tag{3.17}$$

where $Q_{\chi_5^2}$ is the right tail probability for a Central Chi-Squared PDF with 5 degrees of freedom, and $Q_{\chi_5'^2(\lambda)}$ is the right tail probability for a non-central Chi-Squared

PDF with 5 degrees of freedom and non-centrality parameter λ. For a specific desired P_d and P_f, we can compute the implied value for the non-centrality parameter from (3.15) and (3.16). We call this value of the non-centrality parameter $\lambda(P_f, P_d)$. This notation is key in illuminating a very useful relationship between the SNR and the smallest separation which can be detected with high probability, and low false alarm rate. From (3.17) we can write

$$\sigma^2 = \frac{1}{\lambda(P_f, P_d)} \boldsymbol{\theta}^T \mathbf{A}^T \left[\mathbf{A} \left(\mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{A}^T \right]^{-1} \mathbf{A} \boldsymbol{\theta}. \tag{3.18}$$

Also, by defining the SNR as

$$\text{SNR} = \frac{1}{\sigma^2} \boldsymbol{\theta}^T \mathbf{H}^T \mathbf{H} \boldsymbol{\theta}, \tag{3.19}$$

and replacing the value of σ^2 with the right hand side of (3.18), the relation between the parameter set $\boldsymbol{\theta}$ and the required SNR can be made explicit[3]:

$$\text{SNR} = \lambda(P_f, P_d) \frac{\boldsymbol{\theta}^T \mathbf{H}^T \mathbf{H} \boldsymbol{\theta}}{\boldsymbol{\theta}^T \mathbf{A}^T \left[\mathbf{A} \left(\mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{A}^T \right]^{-1} \mathbf{A} \boldsymbol{\theta}} \tag{3.25}$$

[3]To give an insight into the terms in the expression for required SNR, let us denote

$$\mathbf{H}^T \mathbf{H} = \begin{bmatrix} a & \mathbf{b}^T \\ \mathbf{b} & \mathbf{C} \end{bmatrix} \tag{3.20}$$

where

$$a = \mathbf{h}^T \mathbf{h} \tag{3.21}$$

$$\mathbf{b} = \left[\mathbf{h}^T \mathbf{h}_{10},\ \mathbf{h}^T \mathbf{h}_{01},\ \mathbf{h}^T \mathbf{h}_{20},\ \mathbf{h}^T \mathbf{h}_{02},\ \mathbf{h}^T \mathbf{h}_{11} \right]^T \tag{3.22}$$

$$\mathbf{C} = \mathbf{A} \mathbf{H}^T \mathbf{H} \mathbf{A}^T \tag{3.23}$$

Then we can show

$$\left[\mathbf{A} \left(\mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{A}^T \right]^{-1} = \mathbf{C} - \frac{1}{a} \mathbf{b} \mathbf{b}^T \tag{3.24}$$

This form gives a rather better intuition to the established relationship between the required SNR and the resolvability. For instance it can be readily proved that the appearance of the subtracted term in the denominator of (3.25) ($\frac{1}{a}\mathbf{b}\mathbf{b}^T$) is due to the energy of point sources ($\alpha + \beta$) being unknown to the detector. In other words, if only the value of $\alpha + \beta$ is known to the detector, this term would vanish and the related term in the denominator would be given by $\boldsymbol{\theta}^T \mathbf{A}^T \mathbf{C} \mathbf{A} \boldsymbol{\theta}$.

The above analysis can be well extended to the case where σ^2 is unknown a priori, and this will be done next. The corresponding hypotheses for this case are given by

$$\begin{cases} \mathcal{H}_0: & \mathbf{A}\boldsymbol{\theta} = \mathbf{0},\ \sigma^2 > 0 \\ \mathcal{H}_1: & \mathbf{A}\boldsymbol{\theta} \neq \mathbf{0},\ \sigma^2 > 0 \end{cases} \tag{3.26}$$

The GLRT for (3.26) [33, p. 345] gives the following test statistic:

$$T_u(\mathbf{g}) = \frac{\widehat{\boldsymbol{\theta}}^T \mathbf{A}^T \left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1} \mathbf{A}\widehat{\boldsymbol{\theta}}}{\mathbf{g}^T\left[\mathbf{I} - \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\right]\mathbf{g}} > \gamma \tag{3.27}$$

where subscript "u" denotes the case of unknown noise variance, $\mathbf{I}$ is the identity matrix, and

$$\widehat{\boldsymbol{\theta}} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{g} \tag{3.28}$$

is the unconstrained maximum likelihood estimation of $\boldsymbol{\theta}$. For any given data set $\mathbf{y}$, we decide $\mathcal{H}_1$ if the statistic exceeds a specified threshold,

$$T_u(\mathbf{y}) > \gamma'. \tag{3.29}$$

From (3.12), the performance of this detector is characterized by [33, p.186]

$$P_f = Q_{F_{5,N-6}}(\gamma') \tag{3.30}$$

$$P_d = Q_{F'_{5,N-6}(\lambda_u)}(\gamma') \tag{3.31}$$

$$\lambda_u = \frac{1}{\sigma^2}\boldsymbol{\theta}^T\mathbf{A}^T\left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1}\mathbf{A}\boldsymbol{\theta}, \tag{3.32}$$

where $Q_{F_{5,N-6}}$ is the right tail probability for a Central F distribution with 5 numerator degrees of freedom and $N-6$ denominator degrees of freedom; and $Q_{F'_{5,N-6}(\lambda_u)}$ is the

right tail probability for a non-central F distribution with 5 numerator degrees of freedom and $N-6$ denominator degrees of freedom, and non-centrality parameter.

In this GLRT context, the following relation between the parameter set $\boldsymbol{\theta}$ and the required SNR (denoted by a subscript "u") can be obtained:

$$\mathrm{SNR}_u = \lambda_u(P_f, P_d) \frac{\boldsymbol{\theta}^T \mathbf{H}^T \mathbf{H} \boldsymbol{\theta}}{\boldsymbol{\theta}^T \mathbf{A}^T \left[\mathbf{A} \left(\mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{A}^T \right]^{-1} \mathbf{A} \boldsymbol{\theta}}, \tag{3.33}$$

which mirrors (3.25), with the only difference in performance being the change of coefficient from $\lambda(P_f, P_d)$ to $\lambda_u(P_f, P_d)$. It can be easily verified that $\lambda(P_f, P_d) < \lambda_u(P_f, P_d)$ for $P_d > P_f$.

In either case, an important question is to consider how different this obtained performance is from that of the "ideal" clairvoyant detector, to which all the parameters (i.e. $\boldsymbol{\theta}$ and σ^2) are known. We first note that in this case the hypothesis test in (3.10) will be a standard linear Gauss-Gauss detection problem. Also, we can further simplify the problem by seeing that the term $(\alpha + \beta)\mathbf{h}$ in the signal model in (3.7) is a common known term under both hypotheses and can be removed. As a result, the following relationship can be obtained for the completely known case:

$$\mathrm{SNR}_{id} = \eta(P_f, P_d) \frac{\boldsymbol{\theta}^T \mathbf{H}^T \mathbf{H} \boldsymbol{\theta}}{\boldsymbol{\theta}^T \mathbf{A}^T \mathbf{A} \mathbf{H}^T \mathbf{H} \mathbf{A}^T \mathbf{A} \boldsymbol{\theta}}, \tag{3.34}$$

where the subscript "id" denotes the ideal case and $\eta(P_f, P_d)$ is the required deflection coefficient[33, p. 71] computed as

$$\eta = \left(Q^{-1}(P_f) - Q^{-1}(P_d) \right)^2, \tag{3.35}$$

where $Q^{-1}(\cdot)$ is the inverse of the right-tail probability function for a standard Gaussian random variable (zero mean and unit variance). Comparing the expression in (3.25) to that of (3.35), we have

$$\frac{\text{SNR}}{\text{SNR}_{id}} = \frac{\lambda(P_f, P_d)}{\eta(P_f, P_d)} \frac{\boldsymbol{\theta}^T \mathbf{A}^T \mathbf{A} \mathbf{H}^T \mathbf{H} \mathbf{A}^T \mathbf{A} \boldsymbol{\theta}}{\boldsymbol{\theta}^T \mathbf{A}^T \left[\mathbf{A} \left(\mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{A}^T \right]^{-1} \mathbf{A} \boldsymbol{\theta}} \tag{3.36}$$

where we note that $\eta(P_f, P_d) < \lambda(P_f, P_d)$, provided $P_d > P_f$. Also, according to (3.20)-(3.24)

$$\mathbf{A} \mathbf{H}^T \mathbf{H} \mathbf{A}^T - \left[\mathbf{A} \left(\mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{A}^T \right]^{-1} = \mathbf{C} - \mathbf{C} + \frac{1}{a} \mathbf{b} \mathbf{b}^T = \frac{1}{a} \mathbf{b} \mathbf{b}^T \tag{3.37}$$

is a positive definite matrix. As a result, as expected, $\text{SNR}_u > \text{SNR} > \text{SNR}_{id}$, always. However, as we will demonstrate in the following subsection, the difference between SNR_u and SNR_{id} is quite small over most of the parameter range. Also, we note that similar to the analysis presented in section 2.4.1 the proposed detector in the limiting case (where intensities are known and equal[4]) indeed produces a uniformly optimal test (same as ideal detector). Furthermore, for the general case as we will show later, the difference between performances of the proposed detector and ideal detector is very small. Therefore, we argue that the GLRT framework used to suggest a detector can be reasonably accounted to set a performance limit in practice.

The expressions in (3.25) or (3.33) are in general applicable to any sampling scheme including non-uniformly sampled or under-sampled (aliased) images. However, since the energy of a bandlimited signal in the continuous domain is that of its uniformly (and super-critically) discretized version divided by sampling rate, the right

[4]The earlier developed approach for this case can be easily extended to the current 2-D model.

hand side of these expressions should be understood more generally as depending upon the sampling offsets (phases) of the discrete images. In particular, for under-sampled images, the energy terms in $\mathbf{H}^T\mathbf{H}$ will vary significantly as sampling phase changes. We will study this effect in Section 3.2.2.

An interesting related question is how the availability of multiple observations of the image (i.e. several frames with different sampling phases) affects the performance. Let Ω_l denote the l-th set of acquired samples (i.e. l-th frame in a video sequence) out of a total of L frames, and let $\mathbf{H}_l$ and SNR_l represent the corresponding $\mathbf{H}$ and the required SNR of the l-th image. Then, the overall required SNR is given by

$$\mathrm{SNR}_L = \lambda(P_f, P_d)\frac{\boldsymbol{\theta}^T\mathbf{H}_L^T\mathbf{H}_L\boldsymbol{\theta}}{\boldsymbol{\theta}^T\mathbf{A}^T\left[\mathbf{A}\left(\mathbf{H}_L^T\mathbf{H}_L\right)^{-1}\mathbf{A}^T\right]^{-1}\mathbf{A}\boldsymbol{\theta}}, \tag{3.38}$$

where

$$\mathbf{H}_L = \sum_{l=1}^{L}\mathbf{H}_l. \tag{3.39}$$

Furthermore, it can be proved that

$$\mathrm{SNR}_L \leq \frac{1}{L}\sum_{l=1}^{L}\mathrm{SNR}_l \tag{3.40}$$

with equality sign for the case of over-sampled frames. The reason behind this is that the energy of signal in over-Nyquist case is a constant value and does not depend on the phase of sampling. Hence, each (super-critically sampled) frame has the same effect on the detection performance. In other words, in this case $\mathbf{H}_m^T\mathbf{H}_m = \mathbf{H}_n^T\mathbf{H}_n =$

$\frac{1}{L}\mathbf{H}_L^T\mathbf{H}_L$ for $1 < m, n < L$. However in under-Nyquist case, some frames (due to better placement of samples) provide more information for detectability than others.

In the following pages, we further analyze the performance results obtained earlier for over-sampled images. Next we look into a case where the image is under-sampled to see the effect of aliasing on the performance. Having earlier derived the general expression for the performance, to facilitate the presentation, in what follows, we will often consider a particular case with the following set of conditions (we may call this the *symmetric case*):

- $p_x = p_y,\ q_x = q_y = 0$
- $\alpha = \beta = 1$
- $h(x, y) = h(\sqrt{x^2 + y^2}) = h(r)$ (angular symmetric kernel)

Some examples of angular symmetric kernels which will be used later in this thesis are "jinc-squared" and Gaussian windows. The former ($\text{jinc}^2(r)$) is the PSF resulting from a single circular aperture and is characterized by

$$\text{jinc}(r) \quad = \quad \frac{1}{2\pi r\sqrt{-1}} \int_0^{2\pi} \exp[\sqrt{-1}(\theta + r\cos\theta)]\mathrm{d}\theta = \frac{J_1(2\pi r)}{2\pi r} \tag{3.41}$$

where $J_1(\cdot)$ is the first order Bessel function of the first kind[22]. The Gaussian kernel, on the other hand, can be considered as a typical approximation of the overall effect of various elements in the imaging systems (including aperture, CCD, out of focus lens, atmospheric or underwater turbulence etc); it is given by

$$h(r) = \frac{1}{\sqrt{2\pi\sigma_r^2}} \exp\left(-\frac{r^2}{2\sigma_r^2}\right). \tag{3.42}$$

A plot of the above kernels is shown in Figure 3.1, where we note that we have normalized the corresponding Rayleigh spacings for both functions to be 1 (this corresponds to $\sigma_r = 0.35$ in (3.42)).

Due to the existence of sidelobes, under the above normalization, we expect the jinc^2 PSF to provide better resolution (but increased sensitivity to) under fixed SNR conditions. In the coming sections, we will look for confirmation of this intuition.

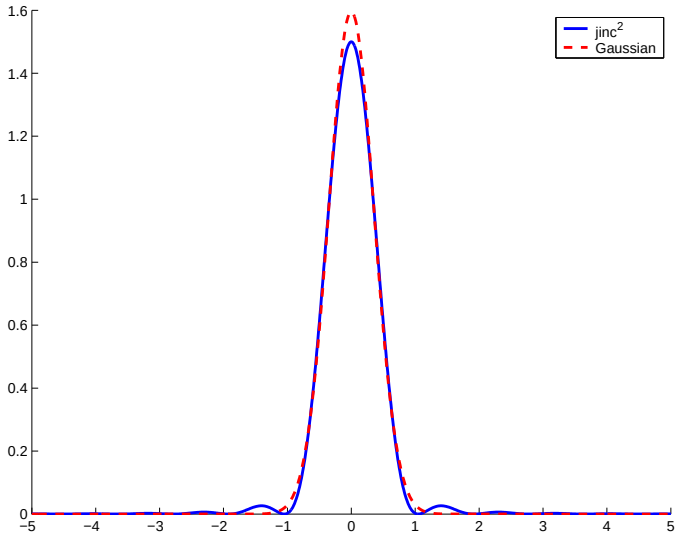

Figure 3.1: Normalized one-dimensional cut of the point spread functions used to present the results

3.2.1 Over-Nyquist sampling

In this section, we further simplify the earlier results for the over-Nyquist case. To begin, we note that the energy terms can be written as

$$\mathbf{P} = \mathbf{H}^T\mathbf{H} = \begin{bmatrix} E_0 & 0 & 0 & -E_{10} & -E_{01} & 0 \\ 0 & E_{10} & 0 & 0 & 0 & 0 \\ 0 & 0 & E_{01} & 0 & 0 & 0 \\ -E_{10} & 0 & 0 & E_{20} & E_{11} & 0 \\ -E_{01} & 0 & 0 & E_{11} & E_{02} & 0 \\ 0 & 0 & 0 & 0 & 0 & E_{11} \end{bmatrix} \tag{3.43}$$

where

$$E_{ij} = \mathbf{h}_{ij}^T\mathbf{h}_{ij} = \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} u^i v^j |H(u,v)|^2 du dv \tag{3.44}$$

$$= f_s^2 \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[\frac{\partial^{i+j}h(x,y)}{\partial x^i \partial y^j}\right]^2 \mathrm{d}x\mathrm{d}y. \tag{3.45}$$

The zero elements in (3.43) are due to the orthogonality of some of the partial derivatives with each other[5]. With the above notation, we will also have

$$\mathbf{Q} = \mathbf{A}^T\left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1}\mathbf{A} \tag{3.46}$$

[5]For instance

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(x,y)\frac{\partial h(x,y)}{\partial x}\mathrm{d}x\mathrm{d}y = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{1}{2}\frac{\partial h^2(x,y)}{\partial x}\mathrm{d}x\mathrm{d}y$$
$$= \int_{-\infty}^{+\infty}\frac{1}{2}h^2(x,y)\Big|_{-\infty}^{+\infty}\mathrm{d}x\mathrm{d}y = 0$$

since the PSF has finite energy and $\lim_{x\to\pm\infty} h(x,y) = 0$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & E_{10} & 0 & 0 & 0 & 0 \\ 0 & 0 & E_{01} & 0 & 0 & 0 \\ 0 & 0 & 0 & E_{20} - \frac{E_{10}^2}{E_0} & E_{11} - \frac{E_{10}E_{01}}{E_0} & 0 \\ 0 & 0 & 0 & E_{11} - \frac{E_{10}E_{01}}{E_0} & E_{02} - \frac{E_{01}^2}{E_0} & 0 \\ 0 & 0 & 0 & 0 & 0 & E_{11} \end{bmatrix} \tag{3.47}$$

Let us now consider the symmetric case. For the over-sampled case, as $d \to 0$, we will have

$$\text{SNR} \approx \frac{\lambda(P_f, P_d)}{N^2} \frac{64}{d^4} \frac{E_{10}^2}{E_0 E_{20} - E_{10}^2} \tag{3.48}$$

where we can show that

$$E_0 = \int_0^{+\infty} r h^2(r) \mathrm{d}r, \tag{3.49}$$

$$E_{10} = \pi \int_0^{+\infty} r \left[\frac{\partial h(r)}{\partial r}\right]^2 \mathrm{d}r, \tag{3.50}$$

$$E_{20} = \frac{\pi}{4} \int_0^{+\infty} \frac{3}{r} \left[\frac{\partial h(r)}{\partial r}\right]^2 + 3r \left[\frac{\partial^2 h(r)}{\partial r^2}\right]^2 \mathrm{d}r \tag{3.51}$$

See Appendix E for detailed calculation of these energy terms.[6] Figure 3.2 shows the minimum detectable d vs SNR for the proposed local detector in (3.27) and the ideal detector. The former detector has been suggested for the case where all the model

[6] We note that due to the Cauchy-Schwartz inequality $E_{10}^2 < E_0 E_{20}$, the right hand side of (3.48) is always positive. Also, as discussed in Section 2.4.1, this term will vanish if the amplitude of the original scene $(\alpha + \beta)$ is known to the detector. This is to say

$$\text{SNR} \approx \frac{\lambda(P_f, P_d)}{N^2} \frac{64}{d^4} \frac{E_0}{E_{20}}.$$

parameters including the noise variance is unknown to the detector. However as can be seen, for any given d, the difference between the required SNR for these detectors is at most 3-4 dB.

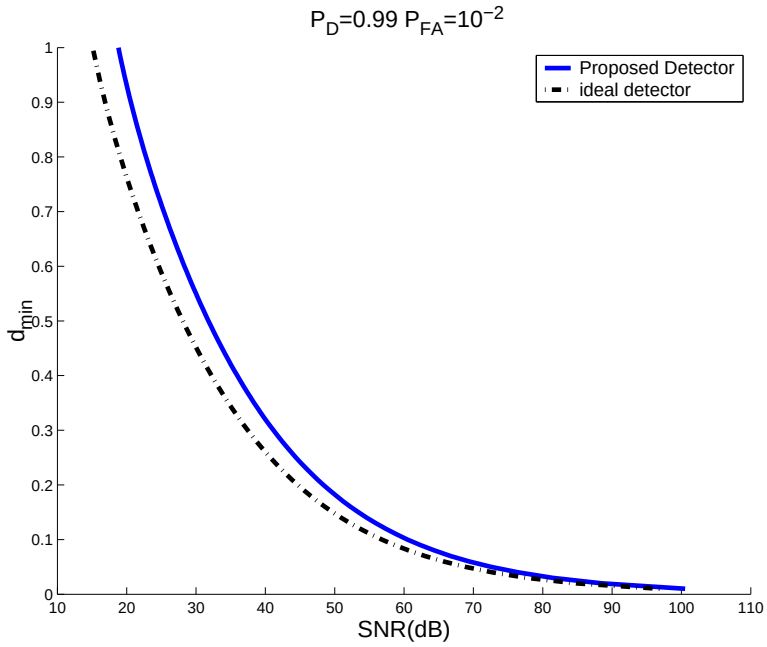

Figure 3.2: Minimum detectable d as a function of SNR (in dB) (just above the Nyquist rate); Gaussian PSF

3.2.2 Under-Nyquist Sampling

Super-critical sampling of a bandlimited function (e.g. $h(x, y)$ and its derivatives) preserves its energy by a factor related to the sampling rate, regardless of sampling offset. However, for the under-Nyquist (aliased) case every element of $\mathbf{H}^T\mathbf{H}$ will be a function of sampling phases in x and y directions (we call these sampling phases ϕ and ψ)[7]. This leads to a rather complicated expression for the required SNR versus d.

[7]See Appendix D

To clarify further, let $\mathbf{P}_{sub}$ be the product $\mathbf{H}^T\mathbf{H}$ for this case. In general, this matrix will have the following form:

$$\mathbf{P}_{sub} = \mathbf{P} + \sum_{m=0}^{L_u}\sum_{n=0}^{L_v}\left[\mathbf{P}_c^{m,n}\sum_{l=0}^{L}\cos(m\phi_l + n\psi_l) + \mathbf{P}_s^{m,n}\sum_{l=0}^{L}\sin(m\phi_l + n\psi_l)\right] \quad (3.52)$$

where $\mathbf{P}$ is defined in (3.43) (for the over-sampled case) and $\mathbf{P}_c^{m,n}$'s and $\mathbf{P}_s^{m,n}$'s are the matrices resulting from aliased components[8]. Specializing to the symmetric case, we have

$$\text{SNR} \approx \lambda(P_f, P_d)\frac{1}{d^4}\frac{4\|\mathbf{h}\|^2 + 4d^2\mathbf{h}^T\mathbf{h}_{20} + d^4\|\mathbf{h}_{02}\|^2}{\left(\|\mathbf{h}_{02}\|^2 - \dfrac{[\mathbf{h}^T\mathbf{h}_{20}]^2}{\|\mathbf{h}\|^2}\right)} \quad (3.53)$$

We note that the energy terms in the above expression involve $\cos(\cdot)$ and $\cos^2(\cdot)$ terms.

Figures 3.3 and 3.4 show the minimum, maximum and average values of the required SNR over the possible range of sampling phases for jinc-squared and Gaussian kernels, respectively. These curves are generated by using (3.25), which is the expression for the required SNR corresponding to the proposed detector in (3.27). For a given value of d, the required SNR is computed for the values of ϕ and ψ drawn uniformly in the range of $[0, 2\pi]$. Then the maximum, minimum, and average of these values are computed.

The curves are shown for the case where only one frame (i.e. one set of uniformly sampled data) is available. Nevertheless, it can be proved that the maximum, minimum and average values of the resolvability (or required SNR) remain the same for arbitrary number of frames (i.e. periodically non-uniform sampled image). This

[8] L_u, L_v, ϕ_l, and ψ_l have been defined in Appendix D.

follows by noting that the number of frames or the sampling rate are embedded inside the "SNR" on the left hand side of (3.25) and (3.33). For instance, for resolving a particular separation, doubling the sampling rate does not change the required SNR, but rather implies that the same detection performance can be achieved with twice the noise variance as compared to the original sampling rate.

It is seen that the required SNR for the case where the PSF is a Gaussian is on average 3 dB (and 16 dB at "worst case") more than for that for the case of simple circular aperture. This can be explained by noting that the jinc-squared kernel contains more energy in its second derivative. The more energy in the second partial derivative means bigger difference between the PDF's under the two hypotheses and therefore better detectability. Also, we note that the jinc-squared window has a larger effective bandwidth which lets more high frequency information through. This phenomenon will also be observed in the following sections. Finally, Figures 3.5 and 3.6 show the average and maximum resolvability at different sampling rates below Nyquist. As observed, change of sampling rate has much less effect on the average performance (i.e. the required *total* SNR) compared to that of the performance at worst case (Figure 3.6). These worst cases occur when at some sampling scenarios (e.g. at 50% Nyquist), the discrete measured signal includes two strong peaks located far from each other (or roughly speaking, where the acquired samples are far from the point sources). This phenomenon clearly degrades the performance of the detector. On the other hand, if there exist some samples which are positioned closely to the point sources, the detector collects more information about the underlying signal.

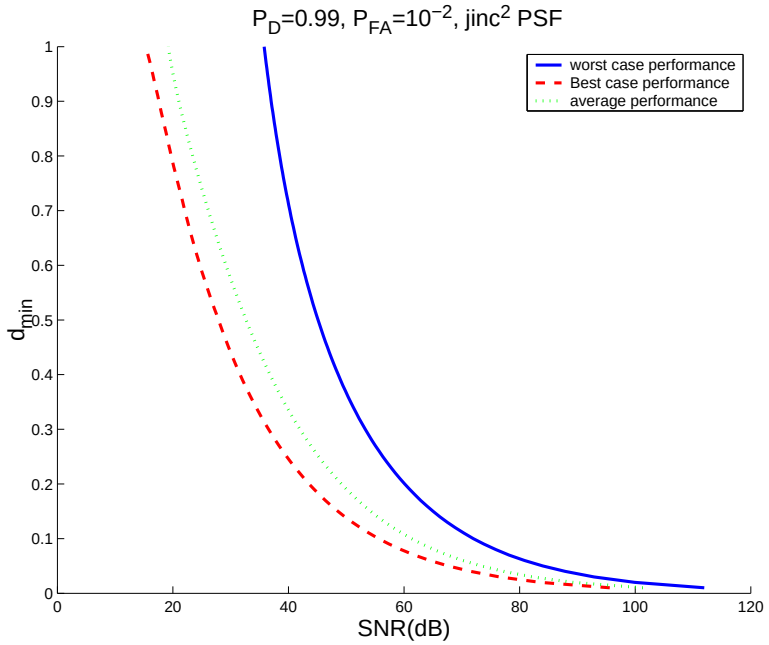

Figure 3.3: Minimum detectable d as a function of SNR (in dB); best, worst, and average performance over the possible range of sampling phases (one set of uniform samples 50% below the Nyquist rate); GLRT detector; $h(r) = \text{jinc}^2(r)$; known σ^2

3.3 Conclusion

In this chapter we investigated the problem of resolving two closely-spaced point sources (beyond the Rayleigh limit) from noise-corrupted, blurred, and possibly under-sampled images. We have studied a detection-theoretic framework to derive performance limits on minimum resolvability. We have discussed the case of under-Nyquist sampling in several parts of this chapter because of its significance in, for example, image super-resolution reconstruction [14, 46, 18]. We have also considered the case where the noise variance is unknown to the detector.

We explicitly found a general, fundamental, and informative relationship to quantify a (statistical) measure of resolution, and also to reveal the effect of point

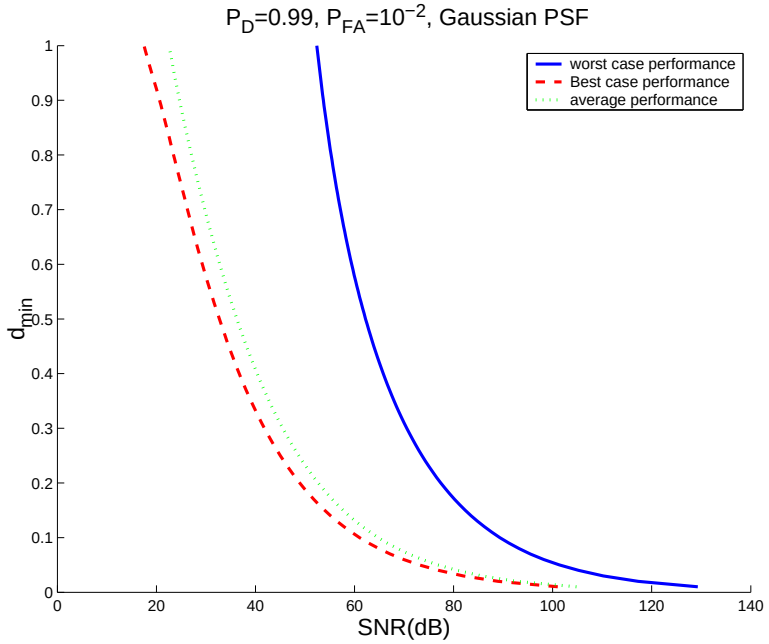

Figure 3.4: Minimum detectable d as a function of SNR (in dB); best, worst and average performance over the possible range of sampling phases (one set of uniform samples 50% below the Nyquist rate); GLRT detector; Gaussian kernel; known σ^2

spread function and sampling parameters on resolvability. The established expressions are in general applicable to any point spread function and any arbitrarily sampling scheme. The analysis for the under-sampled images demonstrates explicitly how the performance is affected by the sampling phases (sampling offset), while in the over-Nyquist case the performance is independent from the sampling offset.

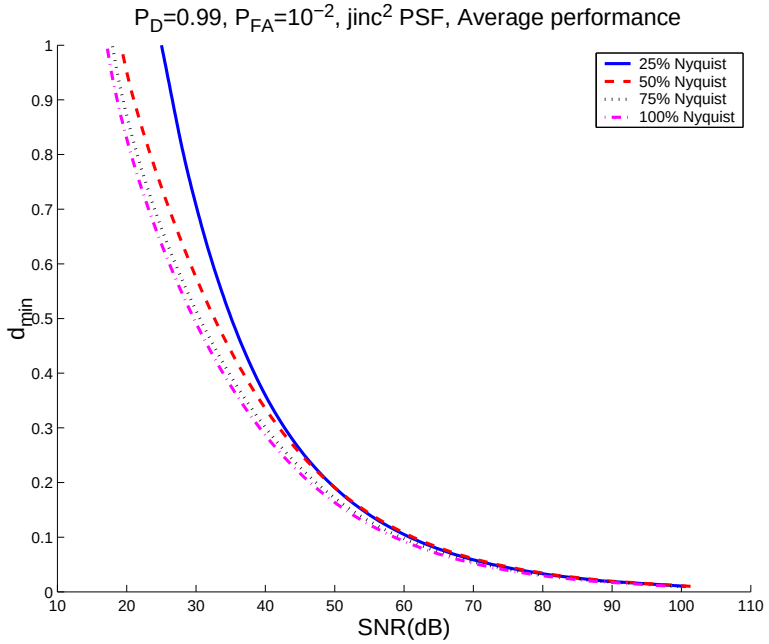

Figure 3.5: Minimum detectable d as a function of SNR (in dB); average performance over the possible range of sampling phases (one set of uniform samples at different sampling rates); GLRT detector; jinc-squared kernel; known σ^2

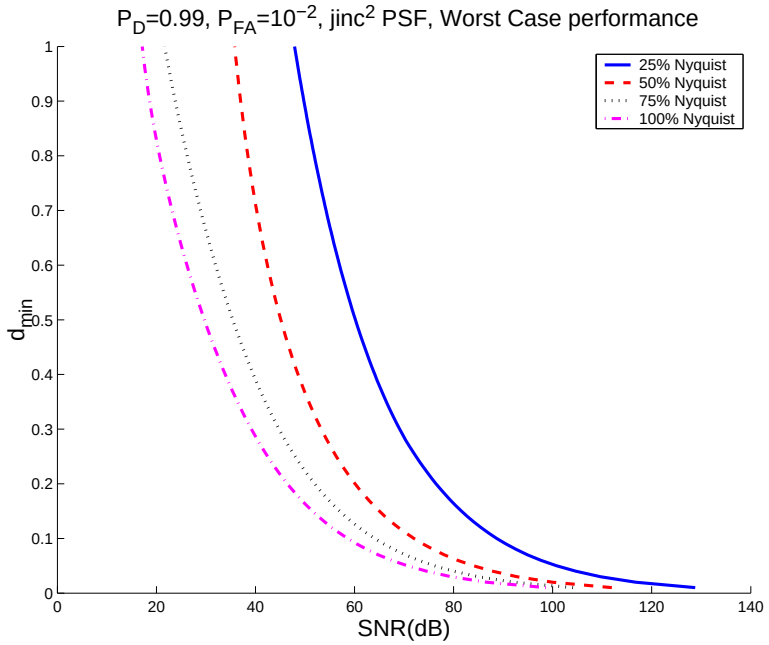

Figure 3.6: Minimum detectable d as a function of SNR (in dB); worst case performance over the possible range of sampling phases (one set of uniform samples at different sampling rates); GLRT detector; jinc-squared kernel; known σ^2

Chapter 4

Estimation-Theoretic and Information-Theoretic Approaches

4.1 Introduction

In this chapter, we re-analyze the resolution problem by using other statistical and information-theoretic tools. Not only do these approaches verify the earlier results of this thesis, but we can also observe some very interesting phenomena connecting all these different frameworks. We first establish a performance bound on estimating the parameters of the signal by computing the Fisher information (and the Cramér-Rao lower bound). We present our calculation for both one-dimensional and two-dimensional signal models. Secondly, we compute a measure of discriminating power between two hypotheses (whether there is one point source or two point

sources) by computing the Kullback-Leibler Distance (KLD). Through this computation, an interesting connection between the Fisher information and the (asymptotic) detectablity is observed.

4.2 Estimation-Theoretic Approach, Fisher Information

In this section we present results on the estimation of the unknown parameters of the model. In particular, we study the asymptotic performance of the maximum likelihood (ML) estimate of the unknown parameters, using the Cramér-Rao lower bound (CRLB). The standard CRLB [32, 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 (unbiased) estimate. Being able to compute a lower bound on the variance of the parameter d, in particular, is rather helpful in verifying and confirming the earlier results of the detection-theoretic framework. For example we shall see how the difference between α and β affects the variance of the estimate in different cases. In what follows we first present the CRLB derivations for a one-dimensional signal model and then will extend it to the two-dimensional scenario.

4.2.1 The CRLB Derivations for One-Dimensional Signal

Here, we compute the CRLB for the following cases:

- the signal model in (2.3), i.e. known intensities but unknown d
- the signal model in (2.48), 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 (2.3), the CRLB for the parameter d (assuming α and β known), is given by [1]:

$$\begin{aligned} \mathrm{var}(\widehat{d}) &\geq \frac{\sigma^2}{\sum_k \left(\frac{\partial s(x_k,d)}{\partial d}\right)^2} = \frac{\sigma^2}{\frac{1}{2\pi}\int_{-\pi}^{\pi}\left|\frac{\partial S(\omega,d)}{\partial d}\right|^2 d\omega} \qquad (4.1) \\ &= \frac{\sigma^2}{f_s}\frac{1}{\frac{\pi^2}{15}(\alpha^2+\beta^2)+\frac{\alpha\beta}{\pi^3 d^5}[(\pi^2d^2-3)\sin(2\pi d)+3\pi d\cos(2\pi d)+3\pi d]} \qquad (4.2) \end{aligned}$$

where $S(\omega, d)$ is the discrete-time Fourier transform (DTFT) of the signal $s(x_k, d)$. To compute the CRLB for the second case, when α, β, d_1 and d_2 are unknown, the Fisher Information matrix is computed[2]. We have

$$\mathrm{cov}(\widehat{d_1},\widehat{d_2},\widehat{\alpha},\widehat{\beta}) \geq \mathbf{\Lambda}^{-1}(\widehat{d_1},\widehat{d_2},\widehat{\alpha},\widehat{\beta}) \qquad (4.3)$$

where $\mathbf{\Lambda}$ is the 4×4 symmetric Fisher Information matrix with its elements defined by

$$\begin{aligned} \mathbf{\Lambda}(1,1) &= \frac{1}{\sigma^2}\sum_k\left(\frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial d_1}\right)^2 = \frac{\alpha^2}{2\pi\sigma^2}\int_{-\pi}^{\pi}|\omega f_s H(\omega,f_s)|^2\,d\omega = \frac{f_s}{\sigma^2}\frac{4\pi^2\alpha^2}{15} \\ \mathbf{\Lambda}(2,2) &= \frac{1}{\sigma^2}\sum_k\left(\frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial d_2}\right)^2 = \frac{\beta^2}{2\pi\sigma^2}\int_{-\pi}^{\pi}|\omega f_s H(\omega,f_s)|^2\,d\omega = \frac{f_s}{\sigma^2}\frac{4\pi^2\beta^2}{15} \\ \mathbf{\Lambda}(3,3) &= \frac{1}{\sigma^2}\sum_k\left(\frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial \alpha}\right)^2 = \frac{1}{2\pi\sigma^2}\int_{-\pi}^{\pi}|H(\omega,f_s)|^2\,d\omega = \frac{f_s}{\sigma^2}\frac{2}{3} \\ \mathbf{\Lambda}(4,4) &= \frac{1}{\sigma^2}\sum_k\left(\frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial \beta}\right)^2 = \frac{1}{2\pi\sigma^2}\int_{-\pi}^{\pi}|H(\omega,f_s)|^2\,d\omega = \frac{f_s}{\sigma^2}\frac{2}{3} \\ \mathbf{\Lambda}(1,2) &= \frac{1}{\sigma^2}\sum_k\frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial d_1}\frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial d_2} \\ &= \frac{-\alpha\beta}{2\pi\sigma^2}\int_{-\pi}^{\pi}|\omega f_s H(\omega,f_s)|^2\cos(\omega f_s(d_1+d_2))d\omega \end{aligned}$$

[1] assuming sampling above Nyquist

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

$$
\begin{aligned}
&= \frac{f_s}{\sigma^2}\frac{2\alpha\beta}{\pi^3}\frac{(\pi^2(d_1+d_2)^2-3)\sin(2\pi(d_1+d_2))+6\pi(d_1+d_2)\cos^2(\pi(d_1+d_2))}{(d_1+d_2)^5}\\
\mathbf{\Lambda}(1,3) &= \frac{1}{\sigma^2}\sum_k \frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial d_1}\frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial \alpha}\\
&= \frac{-\alpha}{2\pi\sigma^2}\int_{-\pi}^{\pi}\omega f_s\left|H(\omega,f_s)\right|^2 d\omega = 0\\
\mathbf{\Lambda}(1,4) &= \frac{1}{\sigma^2}\sum_k \frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial d_1}\frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial \beta}\\
&= \frac{-\alpha}{2\pi\sigma^2}\int_{-\pi}^{\pi}\omega f_s\left|H(\omega,f_s)\right|^2\sin(\omega f_s(d_1+d_2))d\omega\\
&= \frac{f_s}{\sigma^2}\frac{\alpha}{2\pi^3}\frac{3\sin(2\pi(d_1+d_2))-4\pi(d_1+d_2)\cos^2(\pi(d_1+d_2))-2\pi(d_1+d_2)}{(d_1+d_2)^4}\\
\mathbf{\Lambda}(2,3) &= \frac{1}{\sigma^2}\sum_k \frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial d_2}\frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial \alpha}\\
&= \frac{\beta}{2\pi\sigma^2}\int_{-\pi}^{\pi}\omega f_s\left|H(\omega,f_s)\right|^2\sin(\omega f_s(d_1+d_2))d\omega\\
&= \frac{f_s}{\sigma^2}\frac{-\beta}{2\pi^3}\frac{3\sin(2\pi(d_1+d_2))-4\pi(d_1+d_2)\cos^2(\pi(d_1+d_2))-2\pi(d_1+d_2)}{(d_1+d_2)^4}\\
\mathbf{\Lambda}(2,4) &= \frac{1}{\sigma^2}\sum_k \frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial d_2}\frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial \beta} \qquad (4.4)\\
&= \frac{\beta}{2\pi\sigma^2}\int_{-\pi}^{\pi}\omega f_s\left|H(\omega,f_s)\right|^2 d\omega = 0\\
\mathbf{\Lambda}(3,4) &= \frac{1}{\sigma^2}\sum_k \frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial \alpha}\frac{\partial s(x_k;\alpha,\beta,d_1,d_2)}{\partial \beta}\\
&= \frac{1}{2\pi\sigma^2}\int_{-\pi}^{\pi}\left|H(\omega,f_s)\right|^2\cos(\omega f_s(d_1+d_2))d\omega\\
&= \frac{f_s}{\sigma^2}\frac{1}{2\pi^3}\frac{-\sin(2\pi(d_1+d_2))+2\pi(d_1+d_2)}{(d_1+d_2)^4}
\end{aligned}
$$

where $H(\omega, f_s)$ is the DTFT of the PSF $h(x)$ sampled at frequency f_s. 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 $\mathbf{\Lambda}^{-1}$, respectively. Also, the CRLB on $d = d_1 + d_2$ is computed by:

$$
\mathrm{CRLB}(\widehat{d}) = \left[\mathbf{\Lambda}^{-1}\right]_{11} + \left[\mathbf{\Lambda}^{-1}\right]_{22} + 2\left[\mathbf{\Lambda}^{-1}\right]_{12}. \qquad (4.5)
$$

Figure 4.1 shows the square-root[3] of CRLB 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 Figure 4.2 the value of $d = 0.3$ is fixed, and the $\sqrt{\text{CRLB}}$ for $\widehat{d}$ 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 by 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 in Section 2.4.3 where we demonstrated that unequal α and β yields improved detection if the center is known and vice versa.

[3] to maintain the same units as d

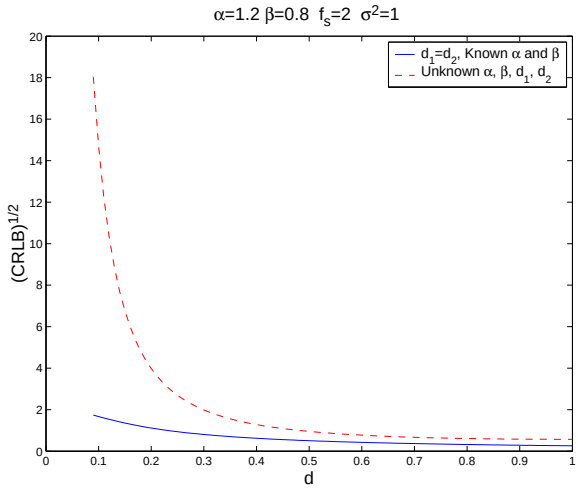

Figure 4.1: CRLB($\widehat{d}$) versus d for two different cases

4.2.2 The CRLB Derivations for the Two-Dimensional Signal

In this section, we carry out the Fisher Information derivations for the general case. Several papers have computed the Cramér-Rao bound to study the mean-square error of unbiased estimators for the distance between the point sources[25, 26, 2, 55, 56]. The CRLB analysis assists us to first confirm the earlier results obtained by the detection-theoretic approach, to better understand the effect of estimation accuracy on the performance of the local detector developed in Chapter 3, and finally to derive the KLD in Section 4.3, when we discuss the information-theoretic framework. We present the analysis for the case where[4] $B_u < f_s < 2B_u$ and $B_v < f_s < 2B_v$ and when two frames are measured, with sampling phases of (ϕ_1, ψ_1) and (ϕ_2, ψ_2), respectively.[5]

[4] B_u and B_v are defined in Appendix D.

[5] Such results can be straightforwardly extended to the case where a higher number of frames are available.

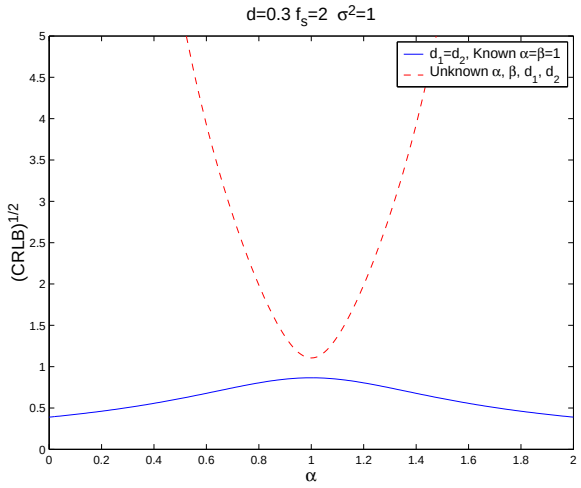

Figure 4.2: CRLB($\widehat{d}$) versus α for two different cases

As such, the vector of unknown parameters of the signal model in (3.3) is

$$\mathbf{t} = \left[\underbrace{p_x,\ q_x,\ p_y,\ q_y}_{\mathbf{t}_r},\ \underbrace{\alpha,\ \beta,\ \phi_2,\ \psi_2}_{\mathbf{t}_s} \right]^T , \tag{4.6}$$

in which we identify two sets of parameters: parameters of interest ($\mathbf{t}_r$) and nuisance parameters ($\mathbf{t}_s$). We have not assumed ϕ_1 and ψ_1 to be unknown, since the first frame can be considered as the reference frame and the uncertainty about the sampling phases can be absorbed by the separation between point sources being unknown.

The CRLB for the separation $d = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2}$ (i.e a nonlinear function of $\mathbf{t}$) is given by [32]:

$$\text{var}(\widehat{d}) \geq \begin{bmatrix} \frac{\partial d}{\partial p_x} \\ \frac{\partial d}{\partial q_x} \\ \frac{\partial d}{\partial p_y} \\ \frac{\partial d}{\partial q_y} \end{bmatrix}^T \begin{bmatrix} [\mathbf{\Lambda}^{-1}]_{11} & [\mathbf{\Lambda}^{-1}]_{12} & [\mathbf{\Lambda}^{-1}]_{13} & [\mathbf{\Lambda}^{-1}]_{14} \\ [\mathbf{\Lambda}^{-1}]_{21} & [\mathbf{\Lambda}^{-1}]_{22} & [\mathbf{\Lambda}^{-1}]_{23} & [\mathbf{\Lambda}^{-1}]_{24} \\ [\mathbf{\Lambda}^{-1}]_{31} & [\mathbf{\Lambda}^{-1}]_{32} & [\mathbf{\Lambda}^{-1}]_{33} & [\mathbf{\Lambda}^{-1}]_{34} \\ [\mathbf{\Lambda}^{-1}]_{41} & [\mathbf{\Lambda}^{-1}]_{42} & [\mathbf{\Lambda}^{-1}]_{43} & [\mathbf{\Lambda}^{-1}]_{44} \end{bmatrix} \begin{bmatrix} \frac{\partial d}{\partial p_x} \\ \frac{\partial d}{\partial q_x} \\ \frac{\partial d}{\partial p_y} \\ \frac{\partial d}{\partial q_y} \end{bmatrix} \tag{4.7}$$

or

$$\mathrm{var}(\widehat{d}) \geq \frac{1}{d^2} \begin{bmatrix} p_x(p_x+q_x) \\ q_x(p_x+q_x) \\ p_y(p_y+q_y) \\ q_y(p_y+q_y) \end{bmatrix}^T \begin{bmatrix} [\mathbf{\Lambda}^{-1}]_{11} & [\mathbf{\Lambda}^{-1}]_{12} & [\mathbf{\Lambda}^{-1}]_{13} & [\mathbf{\Lambda}^{-1}]_{14} \\ [\mathbf{\Lambda}^{-1}]_{21} & [\mathbf{\Lambda}^{-1}]_{22} & [\mathbf{\Lambda}^{-1}]_{23} & [\mathbf{\Lambda}^{-1}]_{24} \\ [\mathbf{\Lambda}^{-1}]_{31} & [\mathbf{\Lambda}^{-1}]_{32} & [\mathbf{\Lambda}^{-1}]_{33} & [\mathbf{\Lambda}^{-1}]_{34} \\ [\mathbf{\Lambda}^{-1}]_{41} & [\mathbf{\Lambda}^{-1}]_{42} & [\mathbf{\Lambda}^{-1}]_{43} & [\mathbf{\Lambda}^{-1}]_{44} \end{bmatrix} \begin{bmatrix} p_x(p_x+q_x) \\ q_x(p_x+q_x) \\ p_y(p_y+q_y) \\ q_y(p_y+q_y) \end{bmatrix} \quad (4.8)$$

where $\mathbf{\Lambda}$ is the Fisher information matrix defined by

$$[\mathbf{\Lambda}]_{ij} = -E\left[\frac{\partial^2 \ln p(\mathbf{g},\mathbf{t})}{\partial \mathbf{t}_i \partial \mathbf{t}_j}\right] \quad (4.9)$$

$$= \frac{1}{\sigma^2}\sum_k \sum_l \left[\frac{\partial}{\partial \mathbf{t}_i}\alpha h(x-p_x, y-p_y) + \beta h(x+q_x, y+q_y)\Big|_{x=x_k, y=y_l}\right. \quad (4.10)$$

$$\left. \times \frac{\partial}{\partial \mathbf{t}_j}\alpha h(x-p_x, y-p_y) + \beta h(x+q_x, y+q_y)\Big|_{x=x_k, y=y_l}\right]$$

The matrix $\mathbf{\Lambda}$ can be partitioned with respect to the parameter sets $\mathbf{t}_r$ and $\mathbf{t}_s$ as

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}_{rr} & \mathbf{\Lambda}_{rs} \\ \mathbf{\Lambda}_{rs} & \mathbf{\Lambda}_{ss} \end{bmatrix}. \quad (4.11)$$

The derivation of the Fisher information matrix for the general sampling scheme is presented in Appendix F. For the over-Nyquist case, we note that the summations in Fisher information matrix can be simply substituted by continuous integrations. Furthermore these integrations can be rather easily computed in the frequency domain for a given point spread function (see Section 4.2 for some examples in 1-D case).

As for the under-Nyquist case, similar to the earlier calculation of energy terms in Chapter 3, we can see that the Fisher Information Matrix has the following

components

$$\mathbf{\Lambda}_{sub} = \mathbf{\Lambda} + \sum_{n=0}^{L_u} \sum_{m=0}^{L_v} \left[\mathbf{\Lambda}_c^{n,m} \sum_{l=0}^{L} \cos(n\phi_l + m\psi_l) + \mathbf{\Lambda}_s^{n,m} \sum_{l=0}^{L} \sin(n\phi_l + m\psi_l) \right] \quad (4.12)$$

where $\mathbf{\Lambda}$ is the Fisher information matrix provided there is no aliasing (i.e. same as what was computed for the over-Nyquist case) and $\mathbf{\Lambda}_c^{n,m}$'s and $\mathbf{\Lambda}_s^{n,m}$'s are the related matrices due to aliasing. For a given value of d, the square root of the CRLB is computed for the values ϕ and ψ in the range of $[0, 2\pi]$. Then the maximum, minimum and average values are calculated to be shown in the following figures. For example these values are displayed in Figure 4.3 for a Gaussian kernel when $\alpha = \beta = 1$. As seen, the estimation task becomes much harder as d decreases.

Figures 4.4 and 4.5 show the average of the square-root of CRLB vs d for two cases, where $\alpha = \beta$ and where $\alpha \neq \beta$. For the case where the PSF is jinc-squared, the estimation accuracy is better than the case of Gaussian PSF, due to the larger energy in the higher order derivatives of a jinc-squared function. The effect of the difference between intensities $(\alpha - \beta)$ on the estimation variance is shown in Figures 4.6 and 4.7. These curves indicate that the estimation task is harder for the case of unequal intensities, as expected.

So far we have explored some detection-theoretic and estimation-theoretic approaches to the problem of achievable resolution. We have investigated the corresponding performance figures (required SNR for a specific resolvability and the lower bound on error of estimating the separation). In the following section we use (and extend) a well-known information-theoretic measure in distinguishing two hypothe-

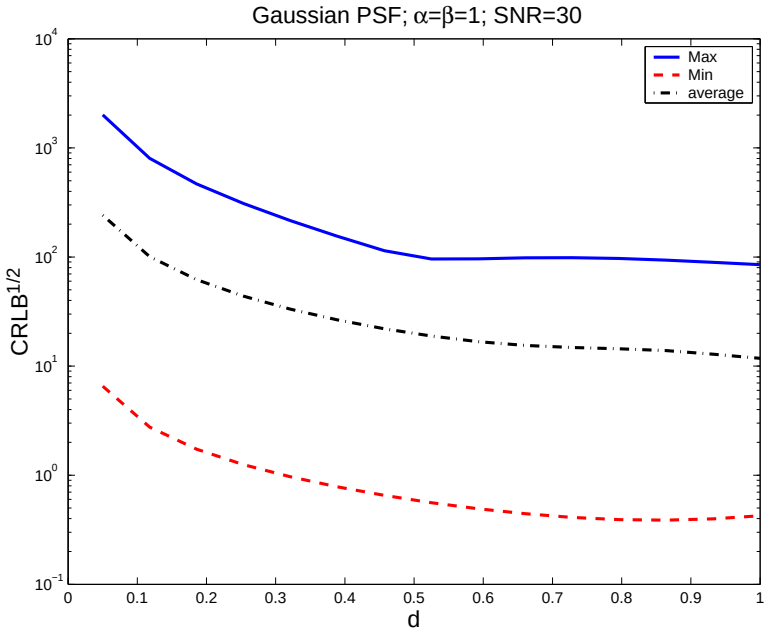

Figure 4.3: $\sqrt{\text{CRLB}(d)}$ vs d; Maximum, minimum and average values over the possible range of sampling phases resulted from two sets of uniform samples 50% below the Nyquist rate (Periodically non-uniform sampling); $\alpha = \beta = 1$; Gaussian PSF; known σ^2

ses. This measure nicely links the asymptotic detection performance to the Fisher information derived in this section.

4.3 Information-Theoretic Analysis, The Kullback-Leibler Distance

In the interest of completeness and also verifying the earlier result from yet another perspective, we investigate the problem of the achievable resolution by an information-theoretic approach. Namely, we compute the symmetric Kullback-Leibler Distance (KLD) or Divergence [36, p. 26] for the underlying hypothesis testing problem. KLD is a measure of discrimination between two hypotheses, and can be directly

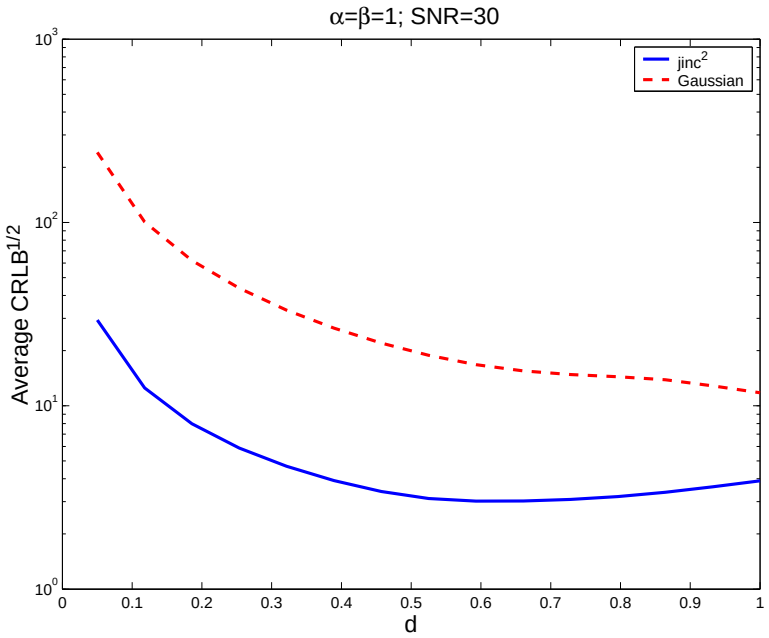

Figure 4.4: $\sqrt{\mathrm{CRLB}(d)}$ vs d; average values over the possible range of sampling phases resulted from two sets of uniform samples 50% below the Nyquist rate (Periodically non-uniform sampling) for Gaussian and jinc-squared PSFs; $\alpha = \beta = 1$; known σ^2

related to the performance of the optimal detector. However, since KLD analysis does not take the effect of nuisance parameters[6] into account, it will indicate a somewhat loose bound on the detection performance for our problem.

To gain better insight, we first carry out an analysis for the case where the point sources are symmetric (that is to say $p_x = q_x$, $p_y = q_y$ and $\alpha = \beta$). Having computed this simpler case, we will extend the result to the general case. To begin, let $p(\mathbf{g}, d)$ and $p(\mathbf{g}, 0)$ be the PDFs of the measured signal under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ in equation (3.5). Then, we will have (See Appendix G for a proof)

$$\mathcal{J}(d) \quad = \quad \int_{\mathcal{D}} [p(\mathbf{g}, d) - p(\mathbf{g}, 0)] \log \left(\frac{p(\mathbf{g}, d)}{p(\mathbf{g}, 0)} \right) \mathrm{d}\mathbf{g} \tag{4.13}$$

[6] those which are unknown to the detector but are common under both hypotheses

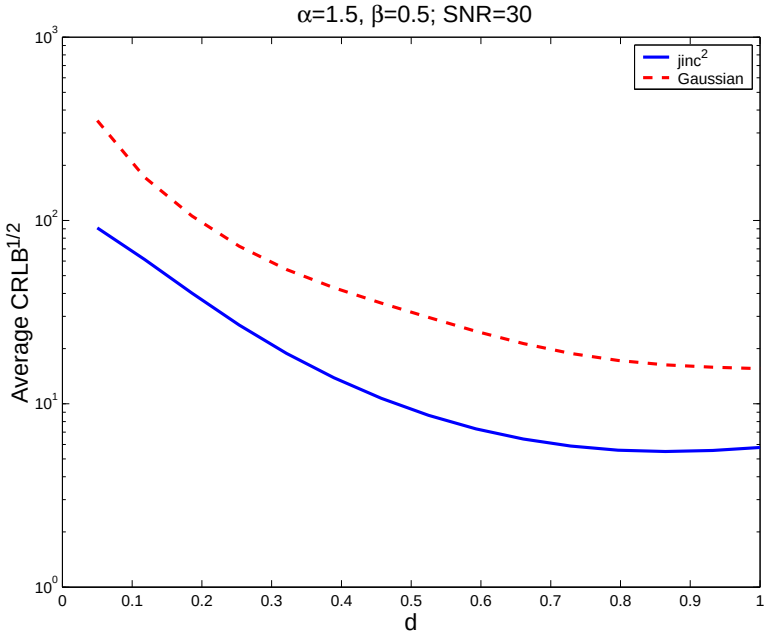

Figure 4.5: $\sqrt{\mathrm{CRLB}(d)}$ vs d; average values over the possible range of sampling phases resulted from two sets of uniform samples 50% below the Nyquist rate (Periodically non-uniform sampling) for Gaussian and jinc-squared PSFs; $\alpha = 1.5, \beta = 0.5$; known σ^2

$$\approx \frac{d^4}{4\sigma^2} \sum_k \sum_l \left(\frac{\partial^2 h_{02}(x_k, y_l)}{\partial d^2} \right)^2 = \frac{d^4}{4\sigma^2} \mathbf{h}_{02}^T \mathbf{h}_{02} \tag{4.14}$$

as $d \to 0$, where $\mathcal{D}$ is the observation (signal) space and we recall that h_{02} is the partial second derivative defined in Chapter 3. We note that the KLD measure behaves as the minimum detectable d raised to the power of 4 (confirming the power law we have derived for the inverse of the required SNR in earlier sections.). We also note that the KLD is proportional to the energy of the second derivative of the PSF, indicating its major role in any measure of detection performance.

Now, let us consider the more general model of unequal and asymmetric point sources. First, we observe from the above analysis that for the underlying problem, KLD computation requires an extension to higher order terms (See Appendix G again).

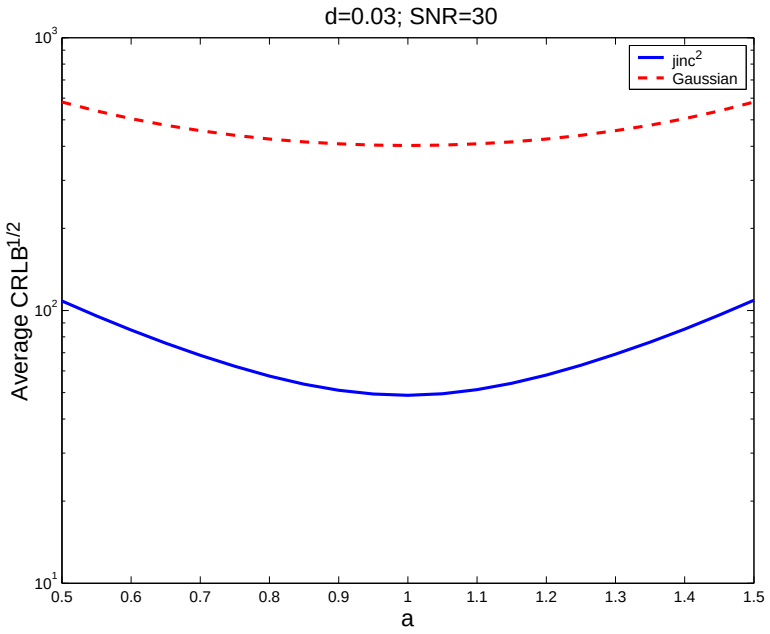

Figure 4.6: $\sqrt{\mathrm{CRLB}(d)}$ vs $\alpha(=2-\beta)$; average values over the possible range of sampling phases resulted from two sets of uniform samples 50% below the Nyquist rate (Periodically non-uniform sampling) for Gaussian and jinc-squared PSFs; $d=0.03$; known σ^2

We utilize this fact to construct our KLD computation for the general case.

To this end, we extend the (low order) formula typically used,

$$\mathcal{J}(\mathbf{t}_r) \approx \mathbf{t}_r^T \, \mathbf{\Lambda}_{rr}|_{\mathbf{t}_r=\mathbf{0}} \, \mathbf{t}_r \tag{4.15}$$

as in for example [36, p. 26], by a second order approximation

$$\mathcal{J}(\mathbf{t}_r) \approx \mathbf{t}_r^T \, \mathbf{\Lambda}_{rr}|_{\mathbf{t}_r=\mathbf{0}} \, \mathbf{t}_r + \left[\mathsf{col}(\mathbf{t}_r\mathbf{t}_r^T)\right]^T \left.\frac{\partial^2 \mathbf{\Lambda}_{rr}}{\partial \mathbf{t}_r^2}\right|_{\mathbf{t}_r=\mathbf{0}} \left[\mathsf{col}(\mathbf{t}_r\mathbf{t}_r^T)\right] \tag{4.16}$$

where $\mathsf{col}(\cdot)$ denotes the lexicographical (columnwise) scanning operator. After some algebra (4.16) will lead to[7]

$$\mathcal{J}(\boldsymbol{\theta}) \approx \frac{1}{\sigma^2}\boldsymbol{\theta}^T\mathbf{A}^T\mathbf{A}\mathbf{H}^T\mathbf{H}\mathbf{A}^T\mathbf{A}\boldsymbol{\theta} \tag{4.17}$$

[7]Note the difference between $\boldsymbol{\theta}$ and $\mathbf{t}_r$ which are defined in (3.9) and (4.6), respectively.

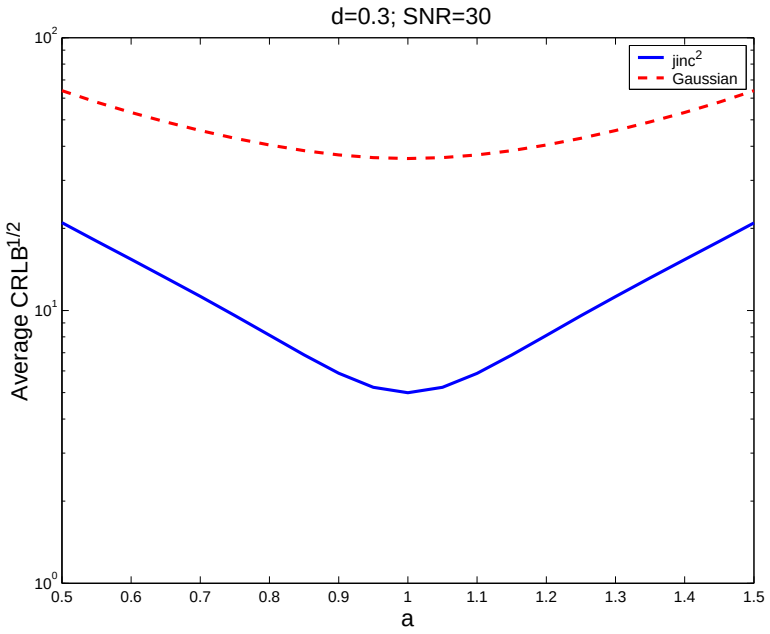

Figure 4.7: $\sqrt{\text{CRLB}(d)}$ vs $\alpha(= 2 - \beta)$; average values over the possible range of sampling phases resulted from two sets of uniform samples 50% below the Nyquist rate (Periodically non-uniform sampling) for Gaussian and jinc-squared PSFs; $d = 0.03$; known σ^2

This is again in general applicable to any arbitrary sampling scenario and point spread function. It is worth mentioning that the matrix $\mathbf{A}^T\mathbf{A}\mathbf{H}^T\mathbf{H}\mathbf{A}^T\mathbf{A}$ is in fact the Fisher information for the parameter set $\boldsymbol{\theta}$ in the quadratic approximated model in (3.7). Interestingly, the above framework shows that computing KLD in the context of [36] (that is, an approximation based on small variations of parameter(s) of interest) is in spirit similar to computing the original KLD for the approximated model. The latter approach, of course, does not require any concern about higher order terms.

For an alias-free signal (4.17) can be further simplified to

$$\mathcal{J}(\boldsymbol{\theta}) \approx \frac{1}{\sigma^2} \begin{bmatrix} -\alpha p_x + \beta q_x \\ -\alpha p_y + \beta q_y \\ \frac{1}{2}(\alpha p_x^2 + \beta q_x^2) \\ \frac{1}{2}(\alpha p_y^2 + \beta q_y^2) \\ \alpha p_x p_y + \beta q_x q_y \end{bmatrix}^T \begin{bmatrix} E_{10} & 0 & 0 & 0 & 0 \\ 0 & E_{01} & 0 & 0 & 0 \\ 0 & 0 & E_{20} & E_{11} & 0 \\ 0 & 0 & E_{11} & E_{02} & 0 \\ 0 & 0 & 0 & 0 & E_{11} \end{bmatrix} \begin{bmatrix} -\alpha p_x + \beta q_x \\ -\alpha p_y + \beta q_y \\ \frac{1}{2}(\alpha p_x^2 + \beta q_x^2) \\ \frac{1}{2}(\alpha p_y^2 + \beta q_y^2) \\ \alpha p_x p_y + \beta q_x q_y \end{bmatrix} \tag{4.18}$$

whereas for the aliased case, the right hand side of (4.17) will depend on the sampling phases.

For under-sampled images, similar to previous analyses, KLD varies with sampling phases,

$$\mathcal{J}_{sub}(\boldsymbol{\theta}) = \mathcal{J}(\boldsymbol{\theta}) + \sum_{m=0}^{L_u} \sum_{n=0}^{L_v} \left[\mathcal{J}_c^{m,n}(\boldsymbol{\theta}) \sum_{l=0}^{L} \cos(m\phi_l + n\psi_l) + \mathcal{J}_s^{m,n}(\boldsymbol{\theta}) \sum_{l=0}^{L} \sin(m\phi_l + n\psi_l) \right] \tag{4.19}$$

where $\mathcal{J}(\boldsymbol{\theta})$ is the Fisher information matrix for actual over-sampled image and $\mathcal{J}_c^{m,n}(\boldsymbol{\theta})$'s and $\mathcal{J}_s^{m,n}(\boldsymbol{\theta})$'s are the related terms caused by aliasing. As sampling rate increases, the terms resulting from aliasing ($\mathcal{J}^{m,n}(\boldsymbol{\theta})$ $m+n>0$) will vanish and the expression in (4.18) is obtained. Figure 4.8 shows variation of KLD over the range of sampling phases for the symmetric case. We have used the expression in (4.13) to find the maximum, minimum and average values of KLD for any given separation d. KLD is beneficial from another perspective, the value of KLD can be asymptotically related to the probability of detection and false alarm rate and provides an upper bound on the detection performance[36]. Namely,

$$P_f(1 - P_d) = \exp(-N\mathcal{J}) \tag{4.20}$$

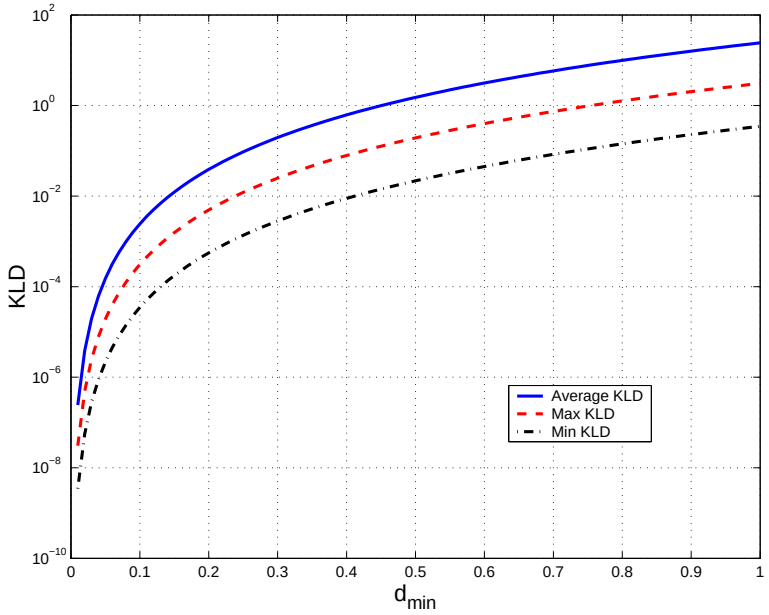

Figure 4.8: Maximum, minimum and average values of KLD vs d for the possible range of sampling phases (50% below the Nyquist rate); $h(r) = \text{jinc}^2(r)$; symmetric case

as N (number of independent samples) goes to infinity.

4.4 Conclusions

We studied the asymptotic performance of ML estimate of the unknown parameters, using the Fisher information matrix. Deriving a lower bound on the variance of the parameter d, in particular, is rather helpful in confirming the results of detection-theoretic analysis and also justifying the effect of estimation accuracy on the performance of the proposed detectors. We also derived the symmetric Kullback-Leibler distance for the underlying problem by extending its standard form to higher order terms. This analysis provides an upper bound on the detection performance we

have derived in Chapter 3 and also connects the Fisher information matrix with this performance bound.

Chapter 5

Perturbation and Sensitivity Analysis

5.1 Introduction

The purpose of this chapter is to analyze how other parameters in real-world imaging can affect the performance of achievable resolution in imaging. The analysis here helps us to compute the effect of small variations in PSF or changes of PSF by other blurring functions (e.g. the effect of lens or charge coupled devices (CCD)). In this section we are interested in studying three cases. In the first case we present an analysis for an imaging system in which samples are acquired through a CCD. We study how the detection performance is affected by using such a system as compared to the idealized point sampling. For the second case, we study the effect of variations in PSF on the resolvability in a general framework. We assume that the variation in PSF is known to the detector and we derive the sensitivity of the required SNR vs

the small change in PSF. The third, and perhaps most important case, is the scenario where the model (PSF) based on which we design our detector is slightly different from the actual PSF. The result of these analyses will (for example) help quantify the importance of precisely knowing the blurring kernel on the resolving power. To gain more intuition, we concentrate on alias-free images throughout this section.

5.2 Imaging with spatial Integration: CCD Sampling

In real-world imaging there are other possible blurring sources which change the total PSF of the imaging system. These blurring effects are usually modelled as a space-invariant functions and can be therefore represented by a linear convolution with the PSF of imaging system. For instance, imaging with a CCD can be properly modelled in such a way. To see this effect, let us first recall that in uniform standard (point) sampling scheme, we have

$$s(x_k, y_l) = s(x, y)|_{x=k/f_s, y=l/f_s} \tag{5.1}$$

where f_s is the sampling frequency. On the other hand, using a CCD in image gathering will result in spatial integration of the light-field coming from a (continuous) physical scene. As an example, CCDs with square area which are uniformly sensitive to light will generate the following discretized output:

$$s_{\text{ccd}}(x_k, y_l) = \frac{1}{W^2} \int_{(l+1/2)W}^{(l-1/2)W} \int_{(k+1/2)W}^{(k-1/2)W} s(x, y) \mathrm{d}x \mathrm{d}y \tag{5.2}$$

where W is the dimension of each CCD cell[1], as depicted in Figure 5.1. By defining

$$\text{rect}(x,y) = \begin{cases} \frac{1}{W^2} & |x|,|y| \leq W/2 \\ 0 & \text{otherwise} \end{cases} \tag{5.3}$$

as the blurring kernel of the above CCD, it can be seen that the expression in (5.2) directly leads to the following

$$s_{\text{ccd}}(x_k, y_l) = s(x,y) * *\text{rect}(x,y)|_{x=k/f_s,\, y=l/f_s} \tag{5.4}$$

$$= I(x,y) * * \underbrace{h(x,y) * *\text{rect}(x,y)}_{h_{\text{ccd}}(x,y)} |_{x=k/f_s,\, y=l/f_s} \tag{5.5}$$

where $I(x,y)$ is the original scene to be imaged, ** denotes the two-dimensional convolution operator and $h_{\text{ccd}}(\cdot,\cdot)$ is the overall PSF. Now let SNR_{ccd} denote the required SNR for the image collected by the above CCD sampling scheme. Figure 5.2 shows the relative difference between this quantity and the required SNR for the point sampling as a function of W. In this example we have considered the symmetric and over-Nyquist case and have presented the results for both jinc-squared and Gaussian PSFs.

From a system design point of view, since typically a CCD with larger effective area has better noise characteristics (i.e. smaller σ^2), one can optimize the size of the CCD by using such a curve in effecting a trade-off with other parameters (like noise variance vs sampling rate).

In what follows, we investigate the effect of any (small) variations in the PSF on the required SNR in a general framework using perturbation analysis.

[1]We ignore the effect of fill factor not being 100%.

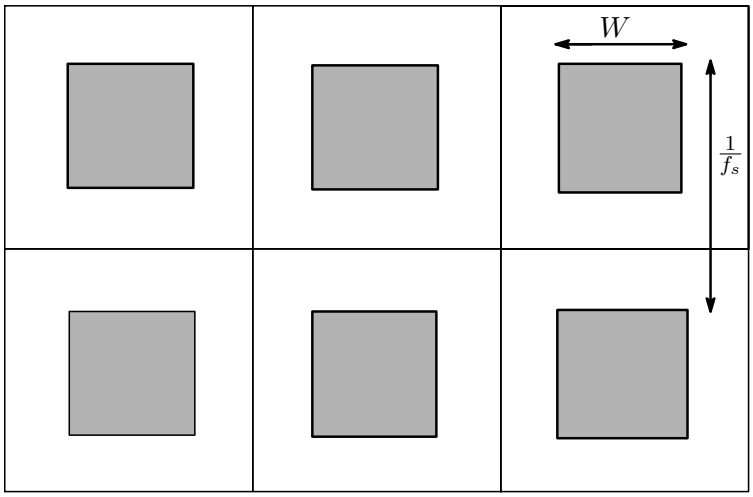

Figure 5.1: A simple structure illustrating the spatial integration caused by CCD

5.3 Variational Derivative Approach

In this section we concentrate on calculating the sensitivity of the required SNR to (known) variations in PSF. We recall from Chapter 3 that the expression for required SNR is a function of the PSF and in particular its partial derivatives (up to the second order). We use well-known techniques in calculus of variations to compute the overall variation in the required SNR.

To begin, consider the expression for SNR in (3.25) and suppose that the point spread function is changed by $h(x,y) \to h(x,y)+\Delta h(x,y)$. By using the concept of variational derivative [20, 60, 70], we can compute the variation in SNR ($\Delta\text{SNR}(h)$) due to the variation $\Delta h(x,y)$. Namely,

$$\Delta\text{SNR}(h) = \frac{\lambda(P_f,P_d)}{N^2}\frac{\boldsymbol{\theta}^T\Delta\mathbf{P}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{Q}\boldsymbol{\theta}-\boldsymbol{\theta}^T\Delta\mathbf{Q}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{P}\boldsymbol{\theta}}{\boldsymbol{\theta}^T\mathbf{Q}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{Q}\boldsymbol{\theta}} \tag{5.6}$$

$$= \frac{\lambda(P_f,P_d)}{N^2}\frac{\boldsymbol{\theta}^T\left(\Delta\mathbf{P}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{Q}-\Delta\mathbf{Q}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{P}\right)\boldsymbol{\theta}}{\boldsymbol{\theta}^T\mathbf{Q}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{Q}\boldsymbol{\theta}} \tag{5.7}$$

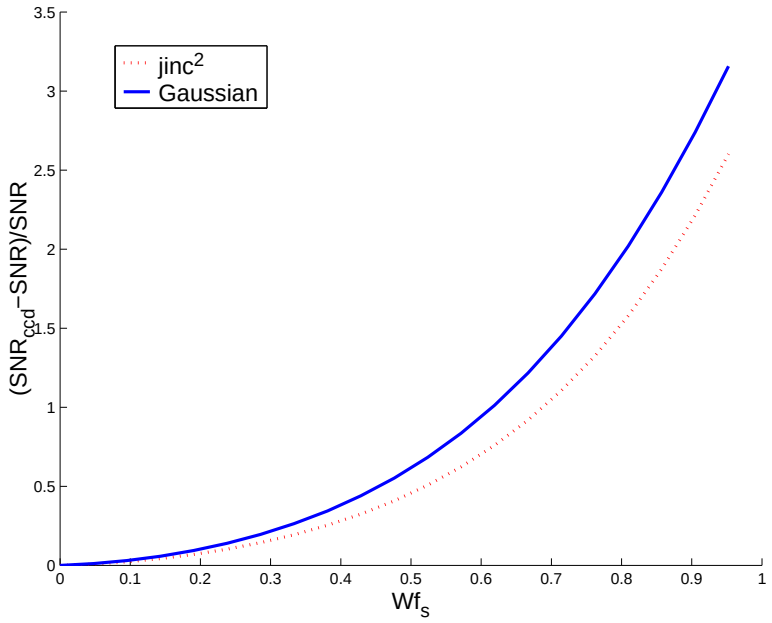

Figure 5.2: Relative difference between the required SNR for CCD sampling and that of point sampling

where $\Delta\text{SNR}(h) = \text{SNR}(h + \Delta h) - \text{SNR}(h)$ and $\Delta\mathbf{P}$ and $\Delta\mathbf{Q}$ are perturbations of matrices $\mathbf{P}$ and $\mathbf{Q}$ defined in (3.43) and (3.47). We have detailed the derivation of $\Delta\mathbf{Q}$ and $\Delta\mathbf{P}$ in Appendix H. Following these derivations, we present the result for the case where point sources are located symmetrically , $p_x = p_y,\ q_x = q_y = 0$, $\alpha = \beta = 1$, and $h(x, y) = h(\sqrt{x^2 + y^2}) = h(r)$ (angular symmetric kernel). In this case the required SNR is given by (3.48). Then the variation is computed as

$$\Delta\text{SNR}(h) \approx \frac{\lambda(P_f, P_d)}{N^2} \times \frac{64}{d^4} \frac{2E_0\Delta E_0 \left(E_0E_{20} - E_{10}^2\right) - E_0^2 \left(E_0\Delta E_{20} + \Delta E_0E_{20} - 2E_{10}\Delta E_{10}\right)}{\left(E_0E_{20} - E_{10}^2\right)^2}$$

As a result, the relative change in SNR is given by[2]

$$\frac{\Delta\mathrm{SNR}(h)}{\mathrm{SNR}} \approx \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} A_h(x,y)\Delta h(x,y)\mathrm{d}x\mathrm{d}y, \tag{5.8}$$

where

$$A_h(x,y) = \frac{\left(E_0E_{20} - 2E_{10}^2\right)h(x,y) + 2E_0E_{10}\dfrac{\partial^2 h(x,y)}{\partial x^2} - E_0^2\dfrac{\partial^4 h(x,y)}{\partial x^4}}{E_0E_{20} - E_{10}^2}. \tag{5.9}$$

Invoking the Cauchy-Schwartz inequality, we have

$$\frac{\Delta\mathrm{SNR}(h)}{\mathrm{SNR}} \leq \sqrt{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[\Delta h(x,y)]^2\mathrm{d}x\mathrm{d}y}\sqrt{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[A_h(x,y)\right]^2\mathrm{d}x\mathrm{d}y}$$

or

$$\frac{\Delta\mathrm{SNR}(h)}{\mathrm{SNR}} \leq \frac{\sqrt{E_0}\sqrt{E_0^3E_{40} - E_0^2E_{20}^2 - 4E_0^2E_{10}E_{30} + 8E_0E_{10}^2E_{20} - 4E_{10}^4}}{E_0E_{20} - E_{10}^2} \tag{5.10}$$

$$\times\sqrt{\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[\Delta h(x,y)]^2\mathrm{d}x\mathrm{d}y} \tag{5.11}$$

which indicates that the relative change in SNR is bounded by the energy in the variation times a term related to energies of the PSF and its derivatives. As an example, consider the following variation which corresponds to a "stretching" ($\epsilon < 0$) or "compressing" ($\epsilon > 0$) of the PSF:

$$\Delta h(x,y) = \rho(\epsilon)h([1+\epsilon]x,[1+\epsilon]y) - h(x,y) \tag{5.12}$$

where

$$\rho(\epsilon) = \frac{\displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h^2(x,y)\mathrm{d}x\mathrm{d}y}{\displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h^2([1+\epsilon]x,[1+\epsilon]y)\mathrm{d}x\mathrm{d}y} \tag{5.13}$$

$$= \frac{E_0}{\displaystyle\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h^2([1+\epsilon]x,[1+\epsilon]y)\mathrm{d}x\mathrm{d}y}. \tag{5.14}$$

[2]See Appendix H for details

is merely an energy normalization factor[3]. A plot of 1-D cut of stretched and compressed versions of the Guassian PSF is depicted in Figure (5.3).

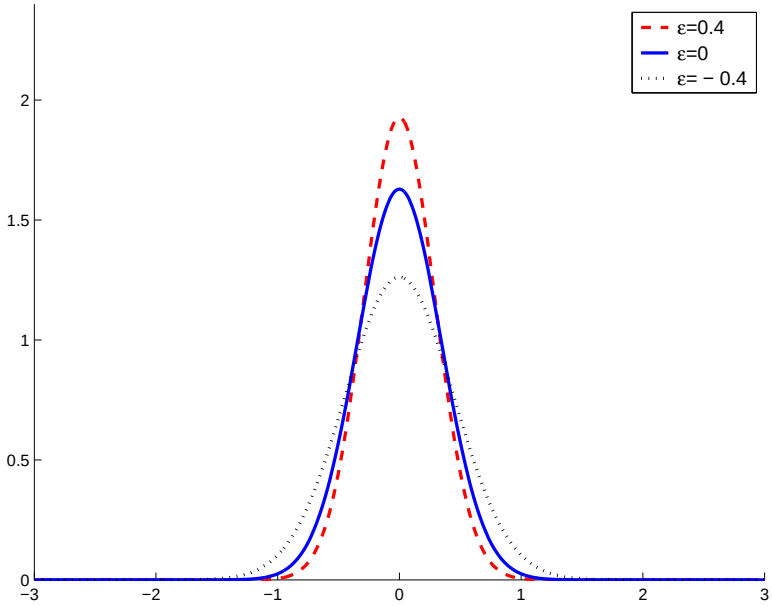

Figure 5.3: 1-D cut of stretched ($\epsilon = 0.4$) and compressed ($\epsilon = -0.4$) versions of the Guassian PSF

Figure 5.4 shows the normalized variation in SNR vs ϵ for the jinc-squared kernel. As expected, with a narrower point spread function ($\epsilon > 0$) less SNR is required to resolve the point sources and vice versa. In fact, we can obtain a closed form relationship for ΔSNR in this case assuming that the image is sampled supercritically. Let us consider the expression in (3.48) which includes the energy of PSF and its partial derivatives. Now let $E_{ij}(\epsilon)$ denote the energy terms for the new kernel

[3]The expression in (5.12) can be approximated by

$$\Delta h(x,y) \approx \epsilon \left(-\frac{2\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(x,y)\left[x\frac{\partial h(x,y)}{\partial x} + y\frac{\partial h(x,y)}{\partial y}\right]\mathrm{d}x\mathrm{d}y}{E_0^2} h(x,y) + \left[x\frac{\partial h(x,y)}{\partial x} + y\frac{\partial h(x,y)}{\partial y}\right] \right)$$

$\rho(\epsilon)h([1+\epsilon]x, [1+\epsilon]y)$. It can be shown that

$$E_{ij}(\epsilon) = (1+\epsilon)^{i+j} E_{ij}(0) \tag{5.15}$$

After some algebra we will have

$$\frac{\Delta\text{SNR}}{\text{SNR}} = \left(\frac{1}{(1+\epsilon)^4} - 1\right) \tag{5.16}$$

which holds true for any PSF.

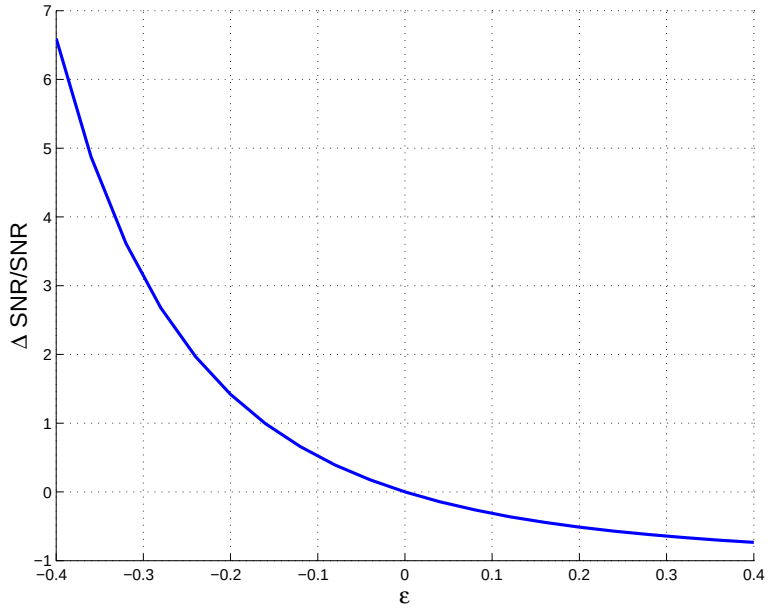

Figure 5.4: Variation in the required SNR vs parameter variation in PSF

5.4 Effects of model mismatch on performance

Another type of analysis is to study the case when the assumed model of the measured signal does not follow the true model. This is an interesting case study

which provides answers to the question of how much performance degrades due to modelling inaccuracies or mismatch.

Let us assume that the actual point spread function is $h(x,y) + \Delta h(x,y)$, whereas the optimal detector is designed for the point spread function $h(x,y)$; so that $\Delta h(x,y)$ is the mismatch (unknown to the detector) between the actual PSF and the assumed PSF. We observe that the (approximated) signal model in this case is now given by

$$\tilde{\mathbf{s}} + \Delta\tilde{\mathbf{s}} = (\mathbf{H} + \Delta\mathbf{H})\,\boldsymbol{\theta} = \tilde{\mathbf{s}} + \Delta\mathbf{H}\boldsymbol{\theta}, \tag{5.17}$$

where

$$\Delta\mathbf{H} \;=\; [\Delta\mathbf{h},\ \Delta\mathbf{h}_{10},\ \Delta\mathbf{h}_{01},\ \Delta\mathbf{h}_{20},\ \Delta\mathbf{h}_{02},\ \Delta\mathbf{h}_{11}].$$

The measured signal will then be $\mathbf{g} + \Delta\mathbf{H}\boldsymbol{\theta}$. Consequently, the estimate of the parameter vector will be changed to

$$\widehat{\boldsymbol{\theta}} + \Delta\widehat{\boldsymbol{\theta}} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{g} + \underbrace{\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\Delta\mathbf{H}\boldsymbol{\theta}}_{\mathbf{b}} \tag{5.18}$$

in which the second term on the right hand side is identified as the "estimator bias". We are now able to show that $\widehat{\boldsymbol{\theta}} + \Delta\widehat{\boldsymbol{\theta}}$ follows a Gaussian PDF with mean $\boldsymbol{\theta} + \mathbf{b}$ and variance $\sigma^2\left(\mathbf{H}^T\mathbf{H}\right)^{-1}$. Using this, we can compute the PDF of $\mathbf{A}\widehat{\boldsymbol{\theta}} + \Delta\widehat{\boldsymbol{\theta}}$ which is again a Gaussian PDF with mean $\mathbf{A}\boldsymbol{\theta} + \mathbf{b}$ and variance $\sigma^2\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T$. The PDF of the suggested test statistics which is compared to a pre-specified threshold similar to the expression in (3.14),

$$\frac{1}{\sigma^2}\left(\widehat{\boldsymbol{\theta}} + \Delta\widehat{\boldsymbol{\theta}}\right)^T\mathbf{A}^T\left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1}\mathbf{A}\left(\widehat{\boldsymbol{\theta}} + \Delta\widehat{\boldsymbol{\theta}}\right) > \gamma \tag{5.19}$$

is characterized by a non-central Chi-squared PDF under both hypotheses [33, p. 32]. To this end, we conclude that the detection performance in the presence of mismatch is characterized by

$$P_f(h+\Delta h) = Q_{\chi_5'^2(\lambda_1)}(\gamma) \tag{5.20}$$

$$P_d(h+\Delta h) = Q_{\chi_5'^2(\lambda_2)}(\gamma) \tag{5.21}$$

where

$$\lambda_1 = \frac{1}{\sigma^2}\mathbf{b}^T\mathbf{A}^T\left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1}\mathbf{A}\mathbf{b} \tag{5.22}$$

$$\lambda_2 = \frac{1}{\sigma^2}(\boldsymbol{\theta}+\mathbf{b})^T\mathbf{A}^T\left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1}\mathbf{A}(\boldsymbol{\theta}+\mathbf{b}) \tag{5.23}$$

are the resulting non-centrality parameters. It is also worth emphasizing here that in order to obtain λ_1 and λ_2, the value of σ^2 in the expression above needs to be computed according to the desired P_d and P_f by using the formula in (3.17).

Now as an example we again use (5.12) to compute the variation in probability of detection and false alarm rate. Hereafter, we present the results for the case where the desired detection and false alarm rates are 0.99 and 0.01 respectively. Figure 5.5 and 5.6 show the variation in probability of error vs ϵ (which controls the stretching or compressing the PSF as in (5.12)) for two different kernels, each of which at different values of d. The change in the total probability of error is computed by

$$P_e(h) = \frac{1}{2}[1 - P_d(h) + P_f(h)] \tag{5.24}$$

$$P_e(h+\Delta h) = \frac{1}{2}[1 - P_d(h+\Delta h) + P_f(h+\Delta h)] \tag{5.25}$$

$$\Delta P_e = P_e(h+\Delta h) - P_e(h)$$

$$= \frac{1}{2}[\underbrace{P_f(h+\Delta h) - P_f(h)}_{\Delta P_f} + \underbrace{P_d(h) - P_d(h+\Delta h)}_{-\Delta P_d}] \qquad (5.26)$$

Firstly we observe that the detector performance is severely affected for a smaller d (e.g. $d = 0.1$). Roughly speaking, for the range of $d < 0.3$, the proposed detector completely fails if $|\epsilon|$ exceeds $d^2/2$.

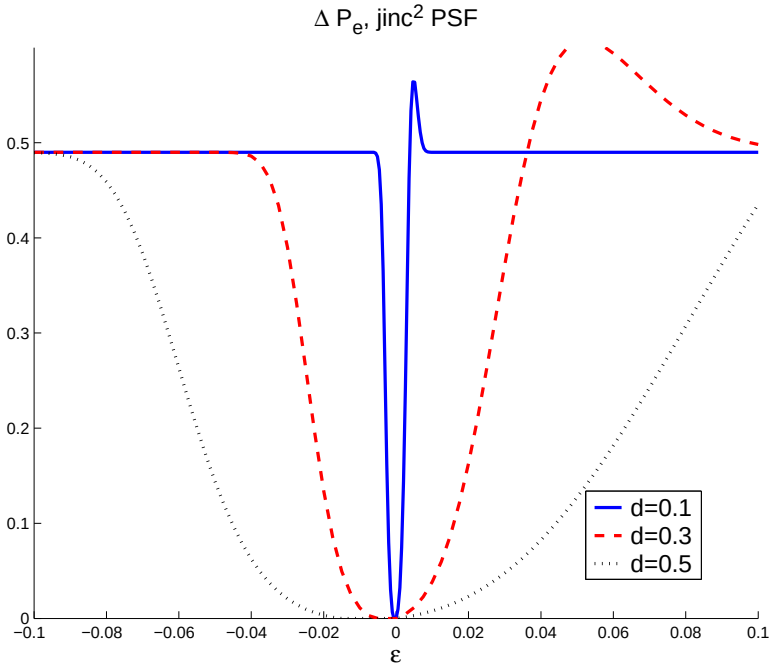

Figure 5.5: Variation in the error rate vs parameter variation in PSF; $h(r) = \text{jinc}^2(r)$

In Figure 5.7 we observe how the probability of error changes as a function of d for a given ϵ (i.e. variation in PSF)[4]. Also Figure 5.8 depicts the variations in PSF which can be tolerated such that the probability of error is lower than a certain level. Clearly, this amount highly depends on the separation between point sources. For a small separation, even a minimal variation in PSF can cause dramatically unpleasant

[4]Hereafter, we only consider negative ϵ.

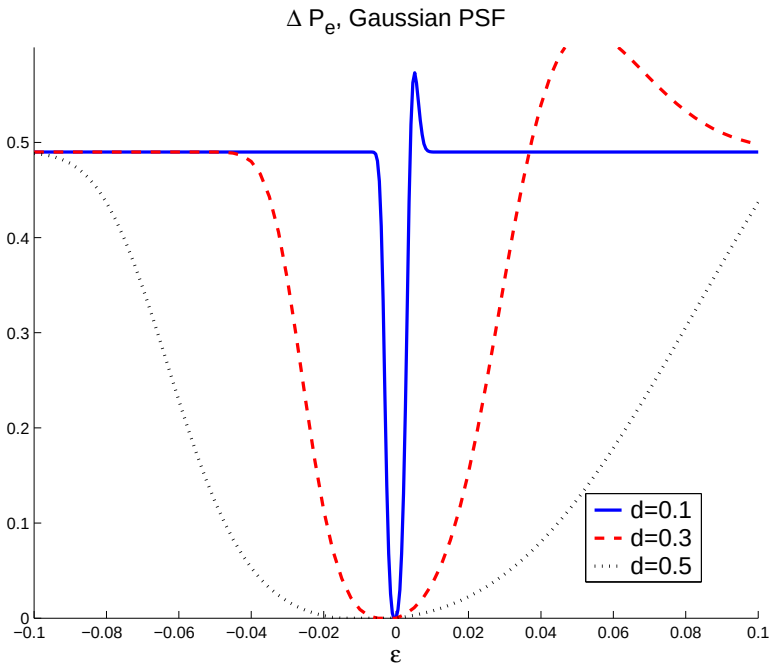

Figure 5.6: Variation in the error rate vs parameter variation in PSF; Gaussian window

results. it is worth noting that the SNR used to generate these figures is the same as that required for $P_e = 0.01$ under no model mismatch.

Another interesting question in this regard would be how much extra SNR is required to compensate the error caused by a model mismatch. To answer this question let us consider the case where $P_d = 0.99$ and $P_f = 0.01$ are the desired detection and false alarm rate respectively. To satisfy these conditions we see that the threshold γ must be equal to 15.1 in (3.14). Howevr if there exists any mismatch caused by the variation in PSF, according to (5.22) and (5.23) we should have $\lambda_1 > 33.5$ and $\lambda_2 < 0.008$ in order to achieve the error rates above. In other words, for a given $\Delta h(x, y)$, increasing the SNR can provide the desired detection accuracy only if the

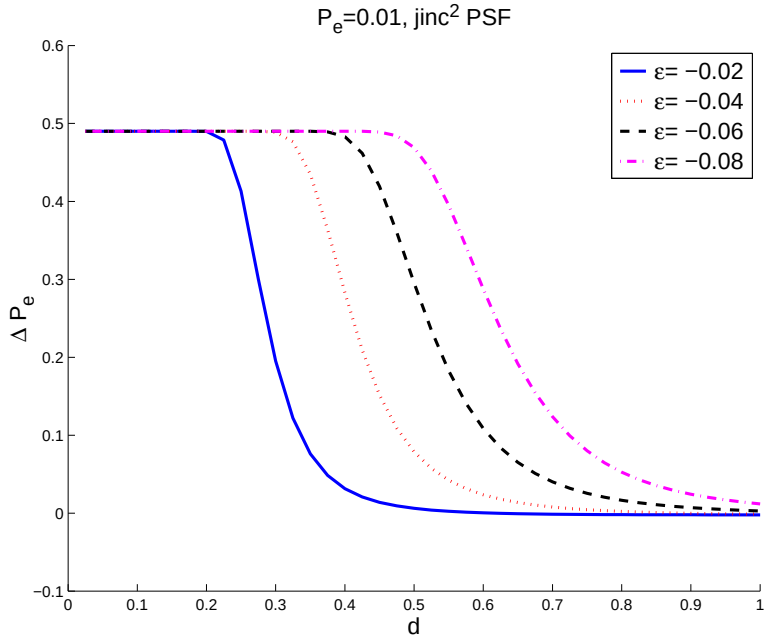

Figure 5.7: Variation in the error rate vs d for fixed ϵ; jinc-squared PSF

following inequalities hold simultaneously:

$$\frac{1}{\sigma^2}\mathbf{b}^T\mathbf{A}^T\left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1}\mathbf{A}\mathbf{b} \;<\; 0.008 \tag{5.27}$$

$$\frac{1}{\sigma^2}(\boldsymbol{\theta}+\mathbf{b})^T\mathbf{A}^T\left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1}\mathbf{A}(\boldsymbol{\theta}+\mathbf{b}) \;>\; 33.5 \tag{5.28}$$

We note that the first inequality enforces the PDF of the test statistic under $\mathcal{H}_0$ to have smaller non-centrality parameter so that the required P_d is accessible, whereas the second inequality plays reverse role for the PDF under $\mathcal{H}_1$. We can unify the inequalities in (5.27) and (5.28) as

$$\frac{(\boldsymbol{\theta}+\mathbf{b})^T\mathbf{A}^T\left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1}\mathbf{A}(\boldsymbol{\theta}+\mathbf{b})}{\mathbf{b}^T\mathbf{A}^T\left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1}\mathbf{A}\mathbf{b}} > 4187.5. \tag{5.29}$$

Under the above condition a sufficiently high value of SNR can compensate the effect of model mismatch. To be more realistic let us also carry out the analysis for the

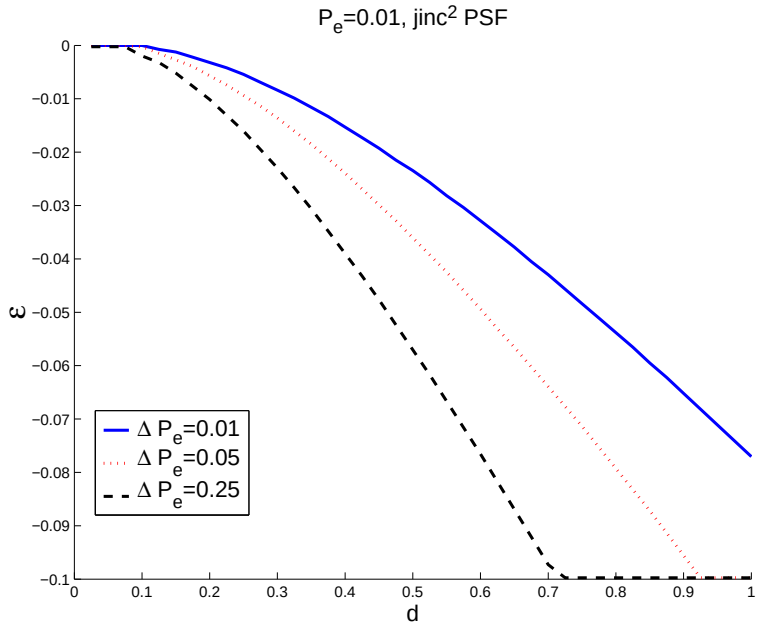

Figure 5.8: the maximum tolerable ϵ vs d for the fixed probability of error; jinc-squared PSF

case where we allow the probability of error to be equal to 0.02 (i.e. $P_d = 0.98$ and $P_f = 0.02$). The threshold γ remains the same. However, to satisfy the new conditions we only need $\lambda > 30.5$ and $\lambda < 0.66$. In other words it is possible to have an error rate less than 0.02 if

$$\frac{(\boldsymbol{\theta}+\mathbf{b})^T\mathbf{A}^T\left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1}\mathbf{A}(\boldsymbol{\theta}+\mathbf{b})}{\mathbf{b}^T\mathbf{A}^T\left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1}\mathbf{A}\mathbf{b}} > 452.23. \tag{5.30}$$

As an illustration of the above analysis, let us consider the case where PSF undergoes the same effect as in (5.12). Figure 5.9 shows the amount of mismatch which can be compensated by presumably high SNR (theoretically as SNR$\rightarrow \infty$) for a given distance d. We demonstrate two cases: the case where no extra error can be tolerated ($\Delta P_e = 0$) and the case where we allow $\Delta P_e = 0.01$. The results clearly indicate that

specially for small distance between point sources the detector is extremely sensitive to (unknown) variation of PSF.

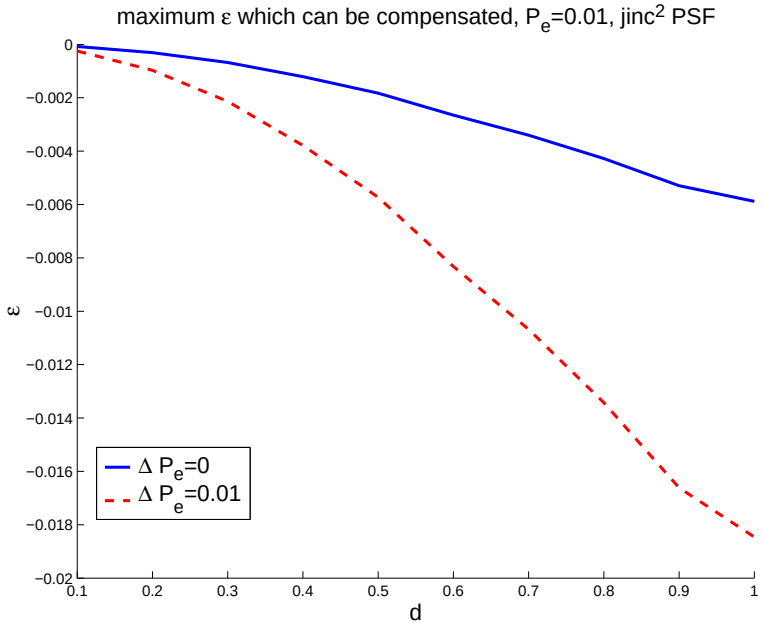

Figure 5.9: Maximum tolerable ϵ which can be compensated by sufficiently increasing SNR vs given d; jinc-squared PSF

Chapter 6

Application of the Proposed Approach to Array Processing: Resolution in the Spectral Domain

Spectral estimation has a long history and significant applications in signal processing. In many areas of application, including the vast body of knowledge in array processing [29], resolving sinusoidal signals with nearby frequencies has been of special interest. In particular, the problem in array signal processing arises in several contexts, including direction-of-arrival estimation, when two incoherent plane waves are incident upon a linear equi-spaced array of sensors [41]. In the past, the vast majority of the techniques in this area have been based on matrix decomposition methods. Notable instances of the relevant literature are found in [41, 51, 28, 4, 31, 30].

These approaches are based principally on second order statistical analysis which relies on the covariance structure of the measured signal. Extensive work has been done to determine the performance of such methods [41, 67, 44, 52, 23, 59, 58, 68, 38].

A common question in this area has been to investigate the relationship between resolution and SNR. Nearly all papers that have addressed this question, either directly or in a related framework, have been focused on the celebrated MUSIC algorithm [53] or its variants (e.g. root MUSIC [1]). The earliest related work was done by Kaveh and Barabell [41] to determine the (minimum) threshold SNR required to resolve two equi-powered sinusoids in the asymptotic regime. In the context of array processing, recent work has employed Cramér-Rao bound analysis to investigate the relationship between resolvability and SNR [55, 56]. While the methods employed are somewhat different, the results obtained are consistent both with our earlier work on establishing detection and estimation-theoretic bounds for resolution in imaging systems [45]-[54], and with the results reported in this chapter.

Without being limited to subspace methods or to asymptotic regimes, a relatively similar question interests us here. We employ a local model-based hypothesis-testing approach to determine the limits to the resolution of frequencies of nearby tones in signals measured in the presence of noise, and over short observation intervals. 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 frequencies of nearby tones (maximum attainable resolution) that is detectable at a

given signal-to-noise ratio (SNR), and for pre-specified probabilities of detection and false alarm (P_d and P_f)?"

As we will demonstrate, in the process of addressing the above question, the machinery of the analysis will also suggest a corresponding detection strategy that can be applied in practice. In other words, the final computed performance limit is simply the result of employing these locally, uniformly, most powerful detectors. In order to illustrate the relevance of the results, we present comparisons against the general class of subspace methods, and in particular the MUSIC algorithm, which is perhaps the most commonly used subspace-based technique in practice. We demonstrate that the proposed detectors yield significantly improved performance in distinguishing frequencies of nearby tones.

We begin by defining the signal of interest as

$$s(x;\delta_1,\delta_2) = a_1 \sin\left(2\pi(f_c-\delta_1)x+\phi_1\right) + a_2 \sin\left(2\pi(f_c+\delta_2)x+\phi_2\right) \tag{6.1}$$

in the range $x \in [-\frac{B}{2}, \frac{B}{2}]$, where for convenience we consider the two frequencies $f_c-\delta_1$ and $f_c+\delta_2$ to be around a "center" frequency[1] f_c. The measured signal is a sampled, and noise-corrupted version of (6.1) as follows:

$$\begin{aligned} f(k;\delta_1,\delta_2) &= s(k;\delta_1,\delta_2) + w(k) \qquad (6.2)\\ &= a_1 \sin\left(2\pi\left(f_c-\delta_1\right)\frac{k}{f_s}+\phi_1\right) + a_2 \sin\left(2\pi\left(f_c+\delta_2\right)\frac{k}{f_s}+\phi_2\right) + w(k), \end{aligned}$$

where the sampling frequency is f_s (Hz), assumed to be sufficiently high to avoid aliasing, and the integer index k is in the range $k \in \{-\frac{N-1}{2}, \cdots, \frac{N-1}{2}\}$, where $N =$

[1]We note that this center frequency can be assumed to be known or estimated, or the detection procedure can be repeated at various candidate center frequencies.

Bf_s. The term $w(k)$ is assumed to be a zero-mean Gaussian white noise process with variance σ^2.

According to the so-called Rayleigh criterion [52], the two peaks in the frequency domain corresponding to $f_c - \delta_1$ and $f_c + \delta_2$ are barely resolvable if

$$\delta_1 + \delta_2 = \frac{1}{B}. \tag{6.3}$$

We are interested in studying the scenario in which the two frequency components are, in this "classical" sense, *unresolvable*. In practice, this corresponds to the situation in which the main-lobe of the Fourier transform of the (sum of) two sinusoids is located in the same FFT bin. So in this context, what we mean by "signals with short observation interval" is simply those signals in which the values of B, δ_1 and δ_2 satisfy the inequality $\delta_1 + \delta_2 < \frac{1}{B}$. A scheme of frequency representation of such signal is depicted in Figure 6.1.

With the above framework in place, we treat the problem of resolution by formulating a hypothesis test. In particular, the corresponding hypotheses for this problem are

$$\begin{cases} \mathcal{H}_0: \quad \delta_1 = 0 \quad \text{and} \quad \delta_2 = 0 \\ \mathcal{H}_1: \quad \delta_1 > 0 \quad \text{or} \quad \delta_2 > 0 \end{cases} \tag{6.4}$$

where $\mathcal{H}_0$ embodies the case where only one spectral component is present, whereas $\mathcal{H}_1$ captures the case where two distinct frequencies are present.[2] We note here that as

[2]Note that the hypothesis test in (6.4) is a one-sided test.

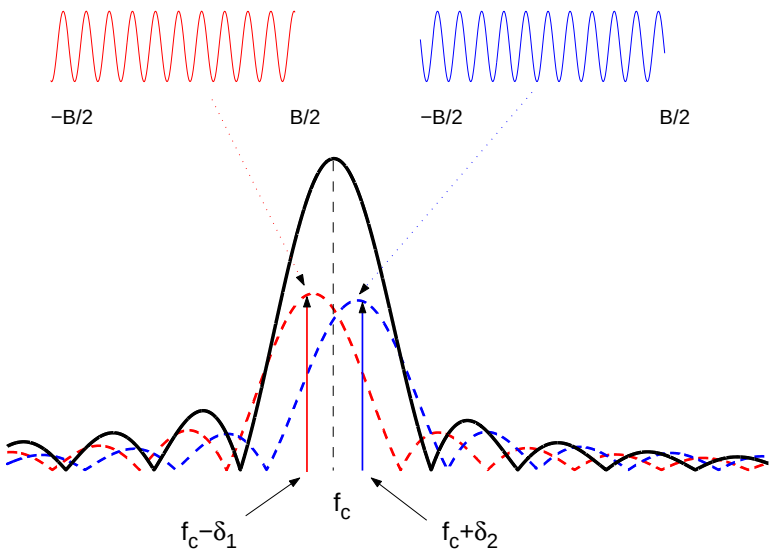

Figure 6.1: Two sinusoids with short observation interval having nearby frequencies: FFT domain

earlier in this framework we consider δ_1 and δ_2 to be unknown to the detector, so that this is a composite hypothesis testing problem. Our approach in this work will again be to take advantage of the small separation between the frequency components to effect an approximation that will yield a detector which is locally uniformly most powerful. As before, this analysis will enable us to explicitly compute the relationship between minimum detectable frequency separation and SNR. Naturally the methodology we present here is quite similar to the approach we advocated in earlier chapters for determining resolution limits in optical imaging [45]-[54].

In Sections 6.1 and 6.2 we study the problem in the case where the noise variance is *known* to the detector. In Section 6.3 we also treat the case of *unknown* noise variance which is perhaps a more practical scenario. In other words, all the parameters of the received signal model are considered unknown to the detector. The

main result of this additional analysis, as we shall see, is that there is little loss in performance when the noise variance is unknown.

6.1 The Case of Equal and Known Amplitude and Phase

To gain maximum intuition and perspective from the foregoing analysis, we first consider a simple case with the following assumptions

- $a_1 = a_2 = 1$.
- $\phi_1 = \phi_2 = 0$
- $\delta_1 = \delta_2 = \delta$

In this case, the measured signal model is given by

$$\begin{aligned} f(k;\delta) &= s(k;\delta) + w(k) && (6.5) \\ &= \sin\left(2\pi\left(f_c - \delta\right)\frac{k}{f_s}\right) + \sin\left(2\pi\left(f_c + \delta\right)\frac{k}{f_s}\right) + w(k) && (6.6) \end{aligned}$$

Since the range of interest for the values of δ is small ($\delta < \frac{1}{2B}$), (these representing one wide peak in the frequency domain,) it is quite appropriate for the purposes of the our analysis (even in the more general case treated in the next Section) to consider approximating the model of the signal around $\delta = 0$. The second order Taylor expansion of (6.5) about $\delta = 0$, with all other variables fixed, is

$$s(k;\delta) = h_0(k) + \delta^2 h_2(k) + O(\delta^4) \tag{6.7}$$

where

$$h_0(k) = 2\sin\left(\frac{2\pi f_c k}{f_s}\right) \tag{6.8}$$

$$h_2(k) = -\frac{4\pi^2 k^2}{f_s^2}\sin\left(\frac{2\pi f_c k}{f_s}\right) \tag{6.9}$$

By ignoring the $O(\delta^4)$ terms in (6.7), the approximate measured signal model can then be written as

$$\widetilde{f}(k;\delta) = h_0(k) + \delta^2 h_2(k) + w(k). \tag{6.10}$$

It is worth noting that in the above approximation, we elect to make explicit use of terms up to order 2 of the Taylor series, since *no* linear term in δ appears in the approximation. By neglecting higher order $O(\delta^4)$ terms, the hypotheses in vector form will be

$$\begin{cases} \mathcal{H}_0: & \widetilde{\mathbf{f}} = \mathbf{h_0} + \mathbf{w} \\ \mathcal{H}_1: & \widetilde{\mathbf{f}} = \mathbf{h_0} + \delta^{\mathbf{2}}\mathbf{h_2} + \mathbf{w} \end{cases} \tag{6.11}$$

where

$$\widetilde{\mathbf{f}} = \left[\widetilde{f}\left(-\frac{N-1}{2}\right), \cdots, \widetilde{f}\left(\frac{N-1}{2}\right)\right]^T, \tag{6.12}$$

and $\mathbf{h}_0$, $\mathbf{h}_2$, and $\mathbf{w}$ are similarly defined. Since $\mathbf{h}_0$ is a common (known) term in both hypotheses, we may simplify further:

$$\begin{cases} \mathcal{H}_0: & \mathbf{y} = \mathbf{w} \\ \mathcal{H}_1: & \mathbf{y} = \delta^2\mathbf{h}_2 + \mathbf{w} \end{cases} \tag{6.13}$$

where we have defined $\mathbf{y} = \tilde{\mathbf{f}} - \mathbf{h_0}$, and the parameter δ is unknown. This is a

problem of detecting a deterministic signal with an unknown parameter (δ^2). We again apply the GLRT which uses the maximum likelihood (ML) estimates of the unknown parameters to form the standard Neyman-Pearson (NP) likelihood ratio detector. The (unconstrained) ML estimate for the parameter δ^2 is given by [32]

$$\widehat{\delta^2} = (\mathbf{h}_2^T \mathbf{h}_2)^{-1} \mathbf{h}_2^T \mathbf{y}, \tag{6.14}$$

which leads to the following GLRT detector:

$$T(\mathbf{y}) = \frac{\widehat{\delta^2}^2}{\sigma^2} \mathbf{h}_2^T \mathbf{h}_2 = \frac{1}{\sigma^2} (\mathbf{h}_2^T \mathbf{h}_2)^{-1} \left(\mathbf{h}_2^T \mathbf{y} \right)^2 \tag{6.15}$$

where we decide $\mathcal{H}_1$ if the statistic exceeds a specified threshold $T(\mathbf{y}) > \gamma_1$. It is worth noting that $T(\mathbf{y})$ is in fact a quadratic form in a rank-one projection. While it may seem troublesome to use the unconstrained ML estimate to form the GLRT, in fact, due to the (known) positivity of δ^2, the detector structure is effectively producing a one-sided test, and hence is in fact a Uniformly Most Powerful (UMP) detector [33, p. 194], [50, p. 124]. The detector can therefore be described simply as a normalized matched filter ($\mathbf{h}_2^T \mathbf{y}$), giving the best detection rate for a given false alarm rate, and for all small values of δ^2. Hence we can write

$$T'(\mathbf{y}) = \sqrt{T(\mathbf{y})} = \sqrt{\frac{1}{\sigma^2} (\mathbf{h}_2^T \mathbf{h}_2)^{-1}} \left(\mathbf{h}_2^T \mathbf{y} \right) \tag{6.16}$$

For any given data set $\mathbf{y}$, we decide $\mathcal{H}_1$ if the statistic exceeds a specified threshold[3]

$$T'(\mathbf{y}) > \gamma. \tag{6.17}$$

[3]Due to the known positivity of δ^2, the threshold (γ) should be positive. In fact another way of writing the GLRT is: $\max\{\widehat{\delta^2}, 0\}(\mathbf{h}_2^T \mathbf{h}_2/\sigma^2) > \gamma_2$. This will result in deciding $\mathcal{H}_0$ for any negative estimate of δ^2 (i.e. $\mathbf{h}_2^T \mathbf{y} < 0$).

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. For this matched filter structure, the detection rate (P_d) and false-alarm rate (P_f) are related as

$$Q(P_d) = Q(\delta^2\eta + \gamma) = Q\left(\delta^2\eta + Q^{-1}(P_f)\right) \tag{6.18}$$

where

$$\eta = \sqrt{\frac{\mathbf{h}_2^T\mathbf{h}_2}{\sigma^2}} \tag{6.19}$$

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. From (6.18) we can write

$$\delta^2\eta = Q^{-1}(P_f) - Q^{-1}(P_d). \tag{6.20}$$

The above expression is key in illuminating a very useful relationship between the SNR and the smallest δ which can be detected with very high probability, and very low false alarm rate. To see this, it is convenient to define the output (total) SNR as follows:

$$\text{SNR} \quad = \quad \frac{\|\mathbf{h}_0 + \delta^2\mathbf{h}_2\|^2}{\sigma^2}. \tag{6.21}$$

Using (6.19) and (6.21), the relation between minimum resolvable δ (i.e. δ_{min}) and the required SNR can be made explicit. Namely, SNR can be computed as

$$\text{SNR} = \frac{\|\mathbf{h}_0 + \delta^2\mathbf{h}_2\|^2}{\|\mathbf{h}_2\|^2} \times \eta^2 \tag{6.22}$$

By substituting the required value of η from (6.20), we have

$$\text{SNR} \quad = \quad (Q^{-1}(P_f) - Q^{-1}(P_d))^2 \left[\frac{\mathbf{h}_0^T \mathbf{h}_0}{\mathbf{h}_2^T \mathbf{h}_2} \frac{1}{\delta^4} + 2 \frac{\mathbf{h}_2^T \mathbf{h}_0}{\mathbf{h}_2^T \mathbf{h}_2} \frac{1}{\delta^2} + 1 \right]. \tag{6.23}$$

This is a fundamental relationship relating minimum resolvable δ to SNR. To make the expressions more explicit, the energy terms in (6.23) can be approximated by[4]

$$\begin{aligned} \mathbf{h}_0^T \mathbf{h}_0 &\approx 2N \\ \mathbf{h}_2^T \mathbf{h}_2 &\approx \frac{\pi^4 N^5}{10 f_s^4} = \frac{\pi^4 N B^4}{10} \\ \mathbf{h}_0^T \mathbf{h}_2 &\approx -\frac{\pi^2 N^3}{6 f_s^2} = -\frac{\pi^2 N B^2}{6}. \end{aligned} \tag{6.24}$$

With these approximations, it is readily seen that for the range of $2\delta B < 1$ the relation (6.23) can be properly summarized by

$$\text{SNR} \approx \frac{320}{\pi^4} \frac{(Q^{-1}(P_f) - Q^{-1}(P_d))^2}{(2\delta B)^4}. \tag{6.25}$$

A plot of (6.23) and its approximation (6.25) are shown in Figure 6.2. The result clearly shows that the minimum resolvable separation is essentially proportional to the inverse of the SNR to the fractional power of 1/4 for the range of $2\delta B < 1$. Note that the frequencies here are separated by 2δ.

Looking at (6.23) or (6.25), one may study the effect of sampling rate on these relationships. It should be mentioned that the sampling rate is embedded inside the "SNR" on the left hand side of (6.23) and (6.25). For instance, for resolving a particular frequency separation (2δ), doubling the sampling rate does not change the required SNR, but rather implies that the same detection performance can be achieved with twice the noise variance as compared to the original sampling rate.

[4]See Appendix I for a justification of these approximations.

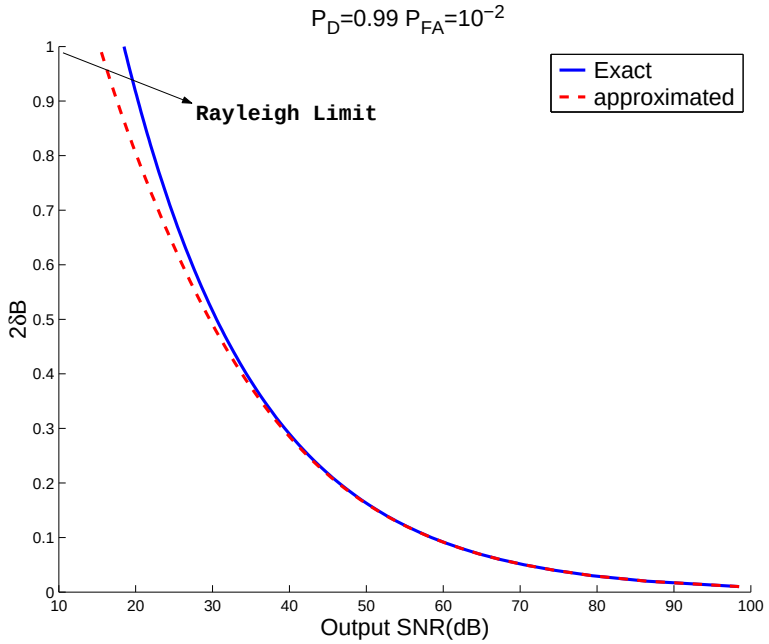

Figure 6.2: $2\delta B$ vs. required SNR.

6.2 The General Case: Known Noise Variance

With the results of the previous section in place, we now follow a similar analysis and extend the results in this section to the general signal model of (6.2), with unknown amplitudes, phases, and also unknown frequency parameters δ_1 and δ_2.[5] The second-order Taylor expansion of the signal model around $(\delta_1, \delta_2) = (0, 0)$ is

$$s(k;\delta) \approx \alpha_0 p_0(k) + \beta_0 q_0(k) + \alpha_1 p_1(k) + \beta_1 q_1(k) + \alpha_2 p_2(k) + \beta_2 q_2(k) \quad (6.26)$$

where

$$p_i(k) = \left(\frac{k}{f_s}\right)^i \sin\left(2\pi f_c \frac{k}{f_s}\right) \quad (6.27)$$

$$q_i(k) = \left(\frac{k}{f_s}\right)^i \cos\left(2\pi f_c \frac{k}{f_s}\right) \quad (6.28)$$

[5]Another more general, and perhaps more practical problem would be to consider the case where σ^2 is also unknown. This is presented in Section 6.3.

and

$$\begin{aligned}
\alpha_0 &= a_1\cos(\phi_1) + a_2\cos(\phi_2) \qquad (6.29)\\
\beta_0 &= a_1\sin(\phi_1) + a_2\sin(\phi_2)\\
\alpha_1 &= 2\pi(a_1\delta_1\sin(\phi_1) - a_2\delta_2\sin(\phi_2))\\
\beta_1 &= 2\pi(-a_1\delta_1\cos(\phi_1) + a_2\delta_2\cos(\phi_2))\\
\alpha_2 &= -2\pi^2(a_1\delta_1^2\cos(\phi_1) + a_2\delta_2^2\cos(\phi_2))\\
\beta_2 &= -2\pi^2(a_1\delta_1^2\sin(\phi_1) + a_2\delta_2^2\sin(\phi_2))
\end{aligned}$$

Rewriting (6.26) in vector form will result in

$$\mathbf{s} \approx \alpha_0\mathbf{p}_0 + \beta_0\mathbf{q}_0 + \alpha_1\mathbf{p}_1 + \beta_1\mathbf{q}_1 + \alpha_2\mathbf{p}_2 + \beta_2\mathbf{q}_2 \qquad (6.30)$$

Now, the hypotheses in (6.4) appear in the following form:

$$\begin{cases} \mathcal{H}_0: \quad \mathbf{z} = \alpha_0\mathbf{p}_0 + \beta_0\mathbf{q}_0 + \mathbf{w} \\ \mathcal{H}_1: \quad \mathbf{z} = \alpha_0\mathbf{p}_0 + \beta_0\mathbf{q}_0 + \alpha_1\mathbf{p}_1 + \beta_1\mathbf{q}_1 + \alpha_2\mathbf{p}_2 + \beta_2\mathbf{q}_2 + \mathbf{w} \end{cases} \qquad (6.31)$$

where $\mathbf{z}$ denotes the approximate measured signal model. Equation (6.31) leads to a linear model for testing the parameter set θ defined as follows:

$$\begin{aligned}
\mathbf{z} &= \mathbf{H}\theta + \mathbf{w} \qquad (6.32)\\
\mathbf{H} &= [\mathbf{p}_0|\ \mathbf{q}_0|\ \mathbf{p}_1|\ \mathbf{q}_1|\ \mathbf{p}_2|\ \mathbf{q}_2] \qquad (6.33)\\
\theta &= [\alpha_0\ \beta_0\ \alpha_1\ \beta_1\ \alpha_2\ \beta_2]^T \qquad (6.34)
\end{aligned}$$

where $\mathbf{H}$ and θ are an $N \times 6$ matrix, and a 6×1 vector, respectively. The corresponding hypotheses are[6]

$$\begin{cases} \mathcal{H}_0: & \mathbf{A}\theta = \mathbf{0} \\ \mathcal{H}_1: & \mathbf{A}\theta \neq \mathbf{0} \end{cases} \tag{6.35}$$

where

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}. \tag{6.36}$$

The GLRT for (6.35) will be

$$T = \frac{1}{\sigma^2} \widehat{\theta}^T \mathbf{A}^T \left[\mathbf{A} \left(\mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{A}^T \right]^{-1} \mathbf{A} \widehat{\theta} \tag{6.37}$$

where

$$\widehat{\theta} = \left(\mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{z} \tag{6.38}$$

From (6.37), the performance of this detector is characterized by

$$P_f = Q_{\chi_4^2}(\gamma) \tag{6.39}$$

$$P_d = Q_{\chi_4'^2(\lambda)}(\gamma) \tag{6.40}$$

$$\lambda = \frac{1}{\sigma^2} \theta^T \mathbf{A}^T \left[\mathbf{A} \left(\mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{A}^T \right]^{-1} \mathbf{A}\theta, \tag{6.41}$$

[6]Two inequalities constrain the values of the parameters in (6.34): $\alpha_0 \alpha_2 \leq 0$ and $\beta_0 \beta_2 \leq 0$. For the detector development in Section 6.2 we have ignored these constraints. We note that ignoring these constraints will still yield a detector, while invoking the constraints will yield (slightly) better detection performance. At an operating point where very high P_d and low P_f are considered, the performance of the detector will not be affected much at all by applying these constraints. Indeed, in such cases, the implied high value of SNR will effectively enforce the constraints with very high likelihood. In other words, for high SNR cases, the probability of violating these inequality constraints is negligible.

where $Q_{\chi_4^2}$ is the right tail probability for a Central Chi-Squared PDF with 4 degrees of freedom, and $Q_{\chi_4'^2(\lambda)}$ is the right tail probability for a non-central Chi-Squared PDF with 4 degrees of freedom and non-centrality parameter λ. For a specific desired P_d and P_f, we can compute the implied value for the non-centrality parameter from (6.39) and (6.40). We call this value of the non-centrality parameter $\lambda(P_f, P_d)$ and explicitly denote it as a function of desired probability of detection and false alarm rate. Meanwhile, similar to the simpler case in the previous section, the SNR in this case is given by

$$\text{SNR} = \frac{\theta^T \mathbf{H}^T \mathbf{H} \theta}{\sigma^2} \tag{6.42}$$

Together, the above yield the relation between the parameter set θ and the required SNR as follows:

$$\text{SNR} = \lambda(P_f, P_d) \left(\theta^T \mathbf{A}^T \left[\mathbf{A} \left(\mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{A}^T \right]^{-1} \mathbf{A} \theta \right)^{-1} \theta^T \mathbf{H}^T \mathbf{H} \theta \tag{6.43}$$

It is instructive to simplify (6.43) by approximating the elements of the matrix $\mathbf{H}^T\mathbf{H}$. These approximations (again, justified in Appendix I), yield

$$\mathbf{H}^T\mathbf{H} \approx \begin{bmatrix} \frac{N}{2} & 0 & 0 & \frac{-N}{4}\mu & \frac{N^3}{24 f_s^2} & 0 \\ 0 & \frac{N}{2} & \frac{-N}{4}\mu & 0 & 0 & \frac{N^3}{24 f_s^2} \\ 0 & \frac{-N}{4}\mu & \frac{N^3}{24 f_s^2} & 0 & 0 & \frac{-N^3}{16}\mu \\ \frac{-N}{4}\mu & 0 & 0 & \frac{N^3}{24 f_s^2} & \frac{-N^3}{16}\mu & 0 \\ \frac{N^3}{24 f_s^2} & 0 & 0 & \frac{-N^3}{16}\mu & \frac{N^5}{160 f_s^4} & 0 \\ 0 & \frac{N^3}{24 f_s^2} & \frac{-N^3}{16}\mu & 0 & 0 & \frac{N^5}{160 f_s^4} \end{bmatrix}$$

where $\mu = \frac{\cos\left(\frac{2\pi f_c}{f_s}N\right)}{\sin\left(\frac{2\pi f_c}{f_s}\right)}$. With this approximation, after some algebra and replacing $\frac{N}{f_s}$ by B, (6.43) will result in

$$\text{SNR} \approx \lambda(P_f, P_d)\frac{E_1 + E_2\mu + E_3B^2 + E_4B^2\mu + E_5B^4}{F_1B^2 + F_2B^2\mu + F_3B^4} \tag{6.44}$$

where

$$\begin{aligned}
E_1 &= 16(\alpha_0^2 + \beta_0^2) \qquad (6.45)\\
E_2 &= -16(\alpha_0\beta_1 + \beta_0\alpha_1)\\
E_3 &= \frac{4}{3}\left[\alpha_1^2 + \beta_1^2 + 2\alpha_0\alpha_2 + 2\beta_0\beta_2\right]\\
E_4 &= -4(\alpha_1\beta_2 + \beta_1\alpha_2)\\
E_5 &= \frac{1}{5}\left(\alpha_2^2 + \beta_2^2\right)\\
F_1 &= \frac{4}{3}\left(\alpha_1^2 + \beta_1^2\right)\\
F_2 &= \frac{-8}{3}\left(\alpha_1\beta_2 + \beta_1\alpha_2\right)\\
F_3 &= \frac{4}{45}\left(\alpha_2^2 + \beta_2^2\right)
\end{aligned}$$

It is useful to note that for the case where $\phi_1 \approx \phi_2$ the first two terms in the numerator of (6.44) dominate its size for small δ_1 and δ_2 (i.e. δ_1, $\delta_2 \ll \frac{1}{B}$), as the other terms are $O(\delta_1^2)$ and $O(\delta_2^2)$. Hence, (6.44) can be further approximated to

$$\text{SNR} \approx \lambda(P_f, P_d)\frac{E_1 + E_2\mu}{F_1B^2 + F_2B^2\mu + F_3B^4} \tag{6.46}$$

To gain further insight, we can consider yet another special case. By assuming $a_1\delta_1 \approx a_2\delta_2$, which results from a proper choice of the center frequency f_c (See Appendix

C.2), the values $E_2,\ F_1,\ F_2$ are also negligibly small and we have

$$\begin{aligned} \text{SNR} &\approx \lambda(P_f, P_d)\frac{E_1}{F_3 B^4} && (6.47)\\ &= \lambda(P_f, P_d)\frac{16(\alpha_0^2+\beta_0^2)}{\frac{4}{45}\left(\alpha_2^2+\beta_2^2\right)B^4} && (6.48)\\ &= \frac{45}{\pi^4}\frac{\lambda(P_f, P_d)}{B^4}\frac{a_1^2+a_2^2+2a_1a_2\cos(\phi_1-\phi_2)}{a_1^2\delta_1^4+a_2^2\delta_2^4+2a_1a_2\delta_1^2\delta_2^2\cos(\phi_1-\phi_2)} && (6.49) \end{aligned}$$

A plot of (6.49) is shown in Figure 6.3 for the case of equal amplitude and for the case of $a_1 = 4a_2$ (In either case, the amplitudes and phases are not known to the detector.). As expected, the case of equal amplitudes produces better detection performance.

In order to compare (6.49) with (6.25), let us set $\delta_1 = \delta_2 = \delta$, $a_1 = a_2 = 1$, $\phi_1 = \phi_2 = 0$ to get[7]

$$\text{SNR} \approx \frac{720}{\pi^4}\frac{\lambda(P_f, P_d)}{(2B\delta)^4} \tag{6.50}$$

As an example, let $P_d = 0.99$ and $P_f = 10^{-2}$. Comparing (6.25) and (6.50) shows that the required SNR for the second case (general case) is increased by a multiplicative factor of 2.72.

Similar to Sections 2 and 3, we can show that that if the value of $a_1 + a_2$ and $\phi_1 = \phi_2$ and the condition $a_1 d_1 = a_2 d_2$ are known to the detector, the GLRT will result in uniformly most powerful test. The reason for this case is that the third and forth elements in (2.52) vanish and the fifth and sixth elements can be combined to produce a single negative coefficient for the vector $cos(\phi_1)\mathbf{p}_2 + sin(\phi_1)\mathbf{q}_2$. Therefore an optimal test statistics is achievable in this case.

[7]It should be noted that these parameter values are unknown to the detector in the general case, therefore we expect poorer performance, as observed.

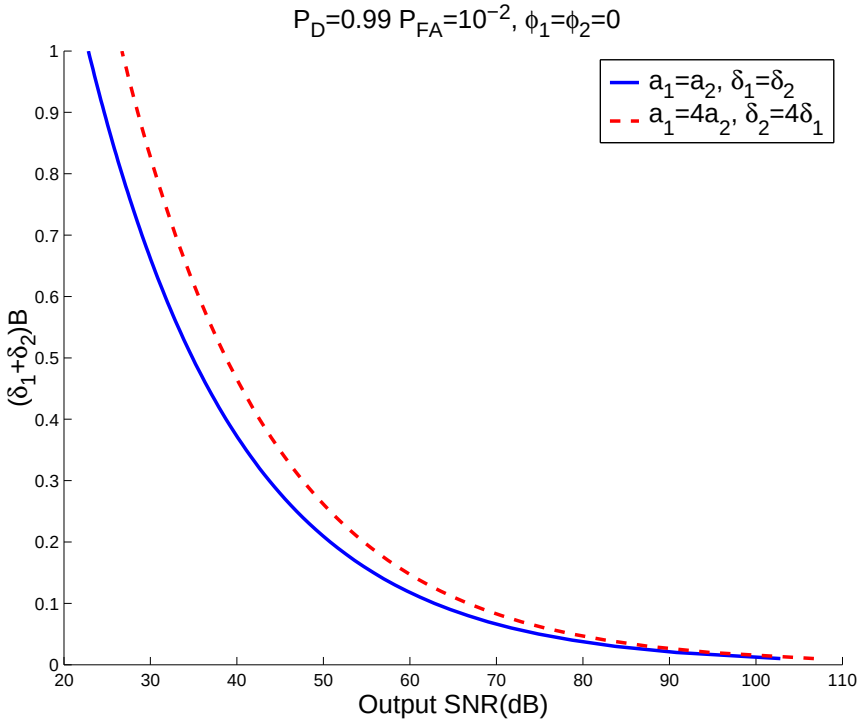

Figure 6.3: $(\delta_1 + \delta_2)B$ vs. required SNR for equal and unequal amplitudes.

6.3 The General Case: Unknown Noise Variance

The hypothesis test in this case is represented by

$$\begin{cases} \mathcal{H}_0: & \mathbf{A}\theta = \mathbf{0},\ \sigma^2 > 0 \\ \mathcal{H}_1: & \mathbf{A}\theta \neq \mathbf{0},\ \sigma^2 > 0 \end{cases} \tag{6.51}$$

The GLRT for (6.51) [33, p. 345] gives the following test statistic:

$$T_u(\mathbf{z}) = \frac{\widehat{\theta}^T \mathbf{A}^T \left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1} \mathbf{A}\widehat{\theta}}{\mathbf{z}^T\left[\mathbf{I} - \mathbf{H}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\right]\mathbf{z}} > \gamma \tag{6.52}$$

where subscript u denote the case of "unknown noise variance" and $\mathbf{I}$ is the identity matrix. For any given data set $\mathbf{z}$, we decide $\mathcal{H}_1$ if the statistic exceeds a specified

threshold,

$$T_u(\mathbf{z}) > \gamma_u. \tag{6.53}$$

From (6.52), the performance of this detector is characterized by [33, p.186]

$$\begin{aligned} P_f &= Q_{F_{4,N-6}}(\gamma) && (6.54)\\ P_d &= Q_{F'_{4,N-6}(\lambda)}(\gamma) && (6.55)\\ \lambda_u &= \frac{1}{\sigma^2}\theta^T\mathbf{A}^T\left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1}\mathbf{A}\theta, && (6.56) \end{aligned}$$

where $Q_{F_{4,N-6}}$ is the right tail probability for a Central F distribution with 4 numerator degrees of freedom and $N-6$ denominator degrees of freedom, and $Q_{F'_{4,N-6}(\lambda)}$ is the right tail probability for a non-central F distribution with 4 numerator degrees of freedom and $N-6$ denominator degrees of freedom, and non-centrality parameter λ. Similar to the previous section, the required SNR is given by

$$\mathrm{SNR}_u = \lambda_u(P_f, P_d)\frac{\theta^T\mathbf{H}^T\mathbf{H}\theta}{\theta^T\mathbf{A}^T\left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1}\mathbf{A}\theta}. \tag{6.57}$$

For the sake of comparison, let us also consider the ideal "clairvoyant" detector, to which all the parameters (amplitudes, phases, frequencies and noise variance) are known. The hypothesis test for such a detector (6.31) will be a standard Gauss-Gauss detection problem. In that case, we also note that term $\alpha_0\mathbf{p}_0 + \beta_0\mathbf{q}_0$ is a common known term under both hypotheses and can be removed. As a result, the required SNR for the ideal detector is given by

$$\mathrm{SNR}_{id} = \eta(P_f, P_d)\frac{\theta^T\mathbf{H}^T\mathbf{H}\theta}{\theta^T\mathbf{A}^T\mathbf{A}\mathbf{H}^T\mathbf{H}\mathbf{A}^T\mathbf{A}\theta}, \tag{6.58}$$

where the subscript "id" denotes the ideal case and $\eta(P_f, P_d)$ is the required deflection coefficient[33, p. 71] computed as

$$\eta = \left(Q^{-1}(P_f) - Q^{-1}(P_d)\right)^2, \tag{6.59}$$

where $Q^{-1}(\cdot)$ is the inverse of the right-tail probability function for a standard Gaussian random variable (zero mean and unit variance). Comparing the expressions in (6.43), (6.57) and (6.58), we conclude that SNR_u >SNR>SNR_{id} always since (1) $\eta(P_f, P_d) < \lambda(P_f, P_d) < \lambda_u(P_f, P_d)$ and (2) $\mathbf{A}\mathbf{H}^T\mathbf{H}\mathbf{A}^T - \left[\mathbf{A}\left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{A}^T\right]^{-1}$ is a positive definite matrix. In Figure 6.4, we compare the performance of the proposed detector for the unknown σ^2 case with those of the detector for the known σ^2 case and the ideal detector (6.58) for the case where $a_1 = a_2$ and $\delta_1 = \delta_2 = \delta$ (curves of $2\delta_2 B$ versus required SNR) . We observe that knowledge of the noise variance makes little difference to the performance (around 1 dB in required SNR). It is worth mentioning that the estimate of the noise variance used in (6.52) is known to be unbiased [33, p.346]. Comparing the ideal (unrealizable) detector, the GLR detector in (6.52) requires 3-5 dB more SNR to achieve the same resolvability.

6.4 Comparison with Existing Subspace-Based Methods

A very significant question is how the above results compare with existing methods for spectral estimation. Since we claim that the proposed detector structures are optimal, we expect that, at least for the particular signal model studied here, we should observe improved performance over existing subspace-based methods. As we

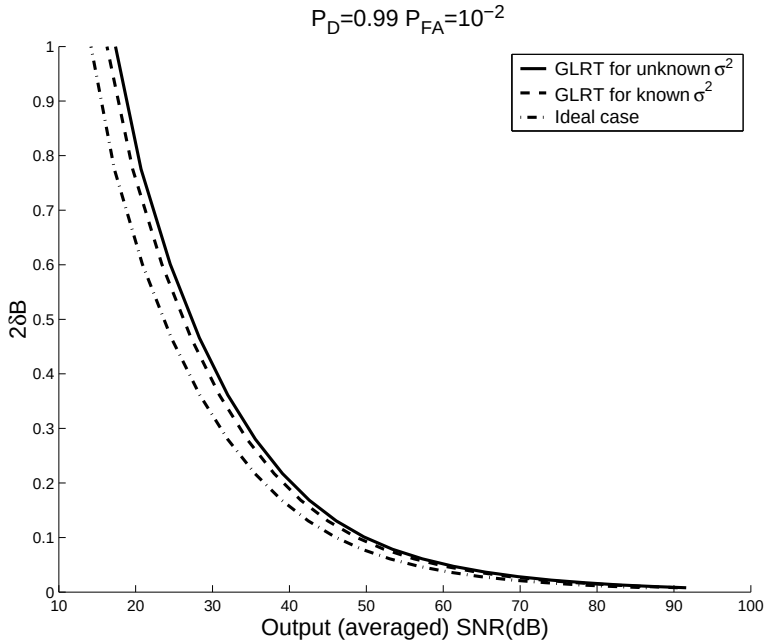

Figure 6.4: $2\delta B$ vs. required average SNR for known and unknown noise variance, $a_1 = a_2$ and $\delta_1 = \delta_2 = \delta$. Averaging (integration) is done over possible range of $[0, 2\pi]$ for both ϕ_1 and ϕ_2

demonstrate below, this is indeed the case. The subspace methods (e.g. MUSIC) for spectral estimation are based on eigen-decomposition of the autocorrelation matrix into orthogonal signal and noise subspaces [30]. In practice, however, since typically only the time series are available, one uses an estimate of the autocorrelation matrix derived from the signal samples.

In any event, much work has been done to study the performance and sensitivity of subspace methods (specifically MUSIC) [43, 5, 37, 41, 67, 44, 52, 23, 59, 58, 68, 38, 34]. Here, we make some comparisons to the existing methods. First, we consider the general class of subspace methods, in which we, very optimistically, assume that the exact autocorrelation matrix is known to the subspace detector under both

hypotheses. As we will see, the proposed approach outperforms the subspace methods even in this (unrealistic) situation. Next, we present a comparison to the performance of the MUSIC algorithm in resolving sinusoids with nearby frequencies.

Throughout this section, we assume that $a_1 = a_2 = 1$ and $\delta_1 = \delta_2 = \delta$. However, we will use our detector structure described in Section 6.2, where we assume that amplitudes, frequencies and phases in the signal model are unknown to the detector[8]. Note that for subspace detectors, the phase is typically assumed to be a uniformly distributed random variable in $[0, 2\pi]$. Meanwhile the "required SNR" computed in Section 6.2 is in general a function of the phases of the sinusoids. Thus, in order to set up a fair comparison to subspace methods, we perform the following averaging for the required SNR over the possible range of ϕ_1 and ϕ_2:

$$\mathrm{SNR}_{\mathrm{avg}} = \frac{1}{4\pi^2} \int_0^{2\pi} \int_0^{2\pi} \mathrm{SNR} \;\; \mathrm{d}\phi_1 \mathrm{d}\phi_2 \tag{6.60}$$

where subscript "avg" denotes the averaged value and the integrand (SNR) is the right hand side of (6.43).

6.4.1 General class of subspace methods; completely known autocorrelation matrix

We first consider the most idealistic subspace detector structure, to which the amplitudes $a_1 = a_2 = 1$ and frequency variables $\delta_1 = \delta_2 = \delta$ of the signal model $f(k, \delta_1, \delta_2)$ in (6.2) are known, and where ϕ_1 and ϕ_2 are assumed to be uniformly distributed random variables in the range of $[0, 2\pi]$. To decide whether the received

[8]We assume that the noise variance is known.

signal contains a single frequency component or two frequency components, we set up the following hypothesis test:

$$\begin{cases} \mathcal{H}_0: & \mathbf{f} \sim \mathcal{N}\left(0, \mathbf{R}_0 + \sigma^2\mathbf{I}\right) \\ \mathcal{H}_1: & \mathbf{f} \sim \mathcal{N}\left(0, \mathbf{R}_1 + \sigma^2\mathbf{I}\right) \end{cases} \tag{6.61}$$

where $\mathbf{R}_0$ and $\mathbf{R}_1$ are the autocorrelation matrices of the signal part in (6.2),[9]

$$\mathbf{R}_0 = \mathsf{Re}\left[\mathbf{r}(f_c)\mathbf{r}^H(f_c)\right] \tag{6.62}$$

$$\mathbf{R}_1 = \frac{1}{2}\mathsf{Re}\left[\mathbf{r}(f_c+\delta)\mathbf{r}^H(f_c+\delta)\right] + \frac{1}{2}\mathsf{Re}\left[\mathbf{r}(f_c-\delta)\mathbf{r}^H(f_c-\delta)\right] \tag{6.63}$$

where $\mathsf{Re}[\cdot]$ denotes the real part and $\mathbf{r}(\cdot)$ is the vector form of

$$r(k; f_c) = \exp\left(j2\pi f_c \frac{k}{f_s}\right)$$

An NP detector for (6.61) decides $\mathcal{H}_1$ if

$$T_c(\mathbf{f}) = \mathbf{f}^T\left[\left(\mathbf{R}_1 + \sigma^2\mathbf{I}\right)^{-1} - \left(\mathbf{R}_0 + \sigma^2\mathbf{I}\right)^{-1}\right]\mathbf{f} > \gamma \tag{6.64}$$

where subscript "c" denotes the "completely known" case. The performance of this detector can be calculated analytically [33] or through Monte-Carlo simulations, the result of which is shown in Figure 6.5. For the purpose of simulation, the performance of (6.64) was computed by averaging over the possible range of ϕ_1 and ϕ_2 similar to (6.60). It is observed that the *required* SNR of this idealistic subspace method is generally between 5-10 dB higher than the required SNR for the proposed GLRT detector in (6.37). An interesting analysis related to the subspace framework is to compute the symmetric Kullback-Leibler Distance (KLD) or Divergence ($J(\cdot)$) [36, p.

[9]Superscript "H" in (6.62) and (6.63) denotes conjugate transpose.

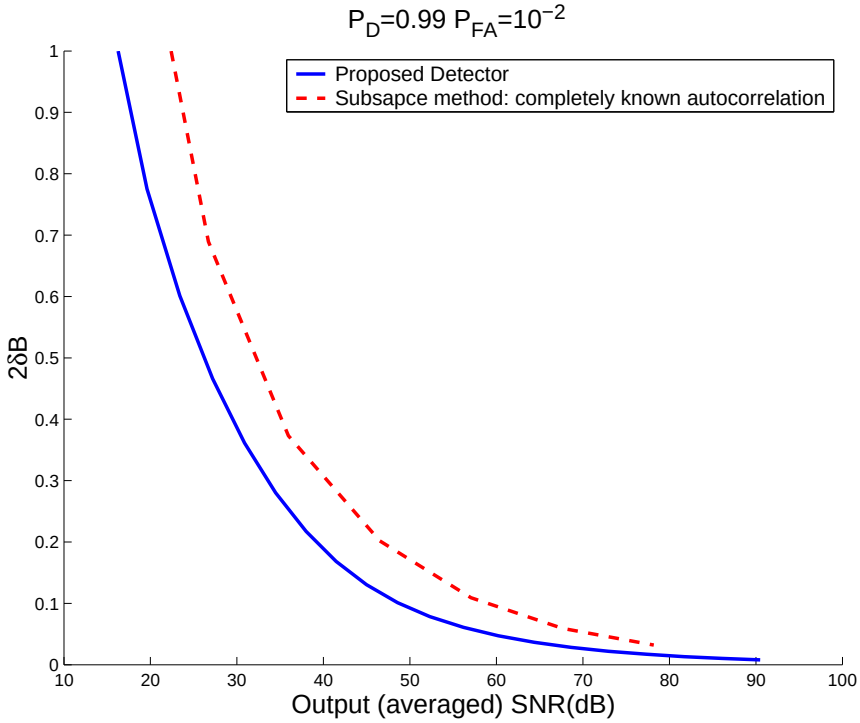

Figure 6.5: $2\delta B$ vs. required output SNR for the subspace detector with completely known autocorrelation.

26]. KLD is a measure of difficulty in discriminating between two hypotheses, and is directly related to the performance figure of the subspace detector. More specifically, let $p(\mathbf{f}, 0)$ and $p(\mathbf{f}, \delta)$ be the PDFs of the measured signal under hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ in equation (6.61). Then, we will have[10]

$$J(\delta) = \int_{\mathcal{D}} [p(\mathbf{f}, \delta) - p(\mathbf{f}, 0)] \log \left(\frac{p(\mathbf{f}, \delta)}{p(\mathbf{f}, 0)} \right) \mathrm{d}\mathbf{f} \approx \frac{\delta^4}{8} \mathtt{tr} \left[\left(\left[\mathbf{R}_1 + \sigma^2 \mathbf{I} \right]^{-1} \frac{\partial^2 \mathbf{R}_1}{\partial \delta^2} \bigg|_{\delta=0} \right)^2 \right] \quad (6.65)$$

as $\delta \to 0$, where $\mathtt{tr}[\cdot]$ is the trace operator and $\mathcal{D}$ is the observation (signal) space. We note that the KLD measure behaves as the minimum detectable δ raised to the power of 4, confirming the power law we have derived for the inverse of the required SNR in

[10] See Appendix G.2.

earlier sections. A comprehensive analysis of the relationship between divergence and resolution in a related framework can be found in [52].

6.4.2 Comparison with MUSIC

For further comparison, we simulated the behavior of the MUSIC algorithm for resolving sinusoids with nearby frequencies. In simulation of MUSIC, the signal is declared to be resolvable if the output of MUSIC produces two distinct peaks within an interval around the true frequencies ($f_c \pm \delta$). The simulations for MUSIC are carried out for cases in which either a single snapshot, or multiple snapshots, are available. Naturally, we consider the output SNR in the latter case as the sum of SNR's of each snapshot.

Here, we develop two different comparison procedures. First, we compare the performance of MUSIC with the performance of the detector in (6.37), where we assume that the center frequency f_c, at which we perform the hypothesis test, is known a priori. Since this might be seen as an unfair comparison, we have put forward an alternative (perhaps more practical) scenario, too. In this scenario, we first seek assistance from MUSIC to estimate the center frequency and then apply the proposed detector in (6.37) centered at the peak estimated by MUSIC.

The results of these experiments are shown in Figure 6.6. First, we observe that the proposed detector significantly outperforms MUSIC in both cases (using known or estimated center frequency). More interestingly, we see that the result of the proposed detector with estimated center frequency (provided by MUSIC) is very close to the

performance of the same detector with known center frequency, the latter representing the ultimate performance bound. This implies that the MUSIC algorithm does a very promising job in locating the center frequency (i.e. the candidate location where we can perform a refinement step using our proposed approach). Intuitively, the reason for this behavior is that for the case where a high probability of resolution (say 0.99) is considered, a fairly high value of SNR should be provided. This value of SNR will effectively guarantee a condition under which the MUSIC algorithm will produce the peak in its spectrum within the range of $[f_c - \delta, f_c + \delta]$. This observation is essentially in agreement with what has been noted in the past about the stability of MUSIC for single-sinusoid signals[11]. See for example [41, 59].

Figure 6.7 shows an example of implementing the proposed approach and also the results obtained by using MUSIC and FFT methods. We observe that the computed FFT representation produces a wide pattern around the frequencies present in the signal (although its maximum is located exactly on the center of frequencies of the sinusoids denoted by stars). Also, MUSIC results in a single sharp peak at the same location. To show the effectiveness of the proposed detector, we plot the probability of presence of a tone at each point in frequency domain. We have used the estimated values of f_c (obtained by either FFT method or MUSIC) and δ_1 and δ_2 (form the ML estimate of θ) and we have assumed that probability of presence of a tone follows the Gaussian distribution (asymptotic pdf for maximum likelihood

[11] Although the signal in our case is a double-sinusoid, since the frequencies are very close and we expect MUSIC to produce one peak, this is indeed a similar situation.

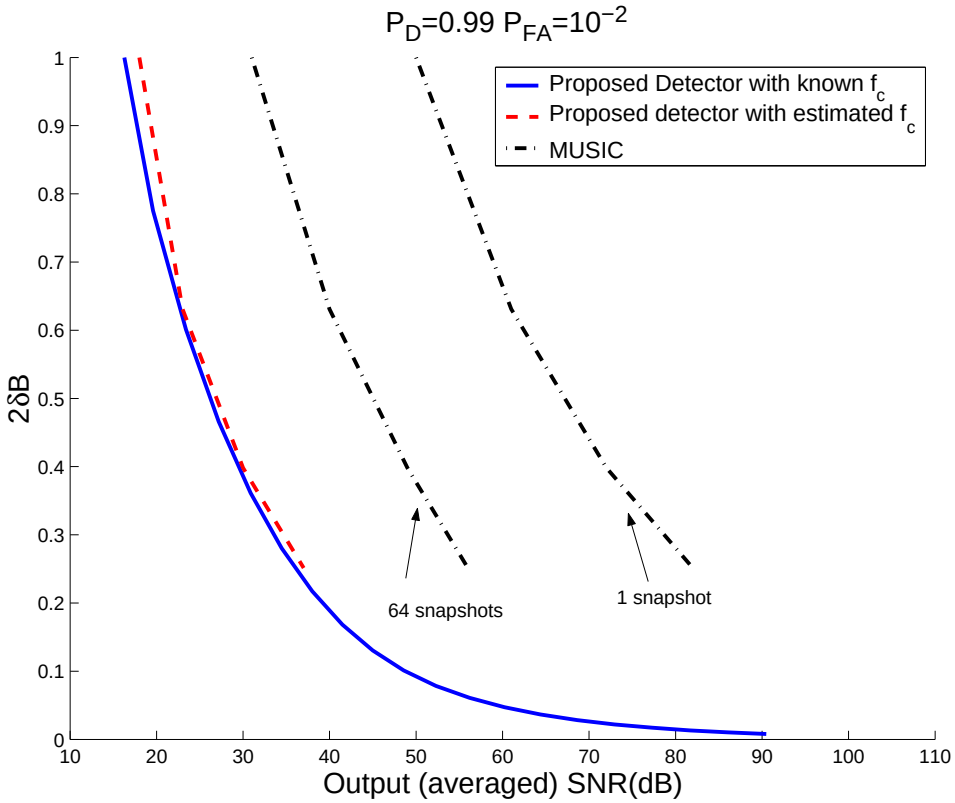

Figure 6.6: $2\delta B$ vs. required output SNR for the MUSIC algorithm.

estimate) which can be computed by CRLB analysis [59, 58]. As seen from Figure 6.7, the Gaussian profiles are very well separated, showing the promising performance of the proposed approach.

6.5 Conclusion

In this chapter, we have derived a performance bound for the minimum resolvable frequency separation between two tones in the presence of noise. We carried out the analysis in the context of locally optimal detectors, and developed corresponding detection strategies that can in practice produce significant improvements over

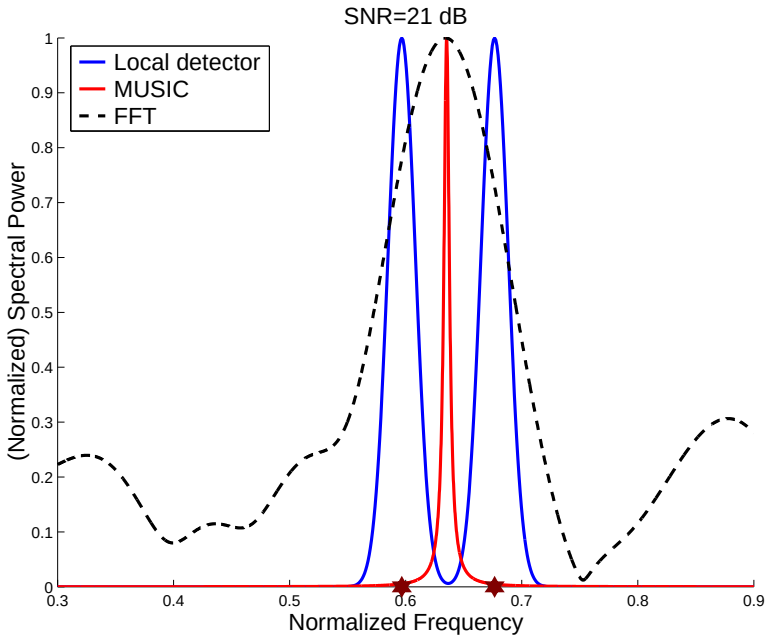

Figure 6.7: An example illustrating the output of MUSIC and standard FFT and also the output obtained based on the proposed approach

existing spectral estimation methods. For the task of bounding performance, we have answered a very practical question: "What is the minimum detectable frequency difference between two sinusoids at a given signal-to-noise ratio?" Or equivalently: "What is the minimum SNR required to discriminate these two sinusoids?"

Compared to existing spectral estimation methods, the proposed locally optimal detectors yield significantly improved detection of very nearby frequencies. It is worth noting that as a matter of implementation, one can always apply an existing method for spectral estimation (such as MUSIC etc.) to the given signal, and then apply the proposed detector as a post-processing operation intended to further improve resolution. As discussed in Section 6.4.2, the application of such a detector, which uses (for example) MUSIC to estimate the center frequency as the test point, is nearly

as effective as applying the proposed detector with a known center frequency.

Furthermore, we clearly observed that noise variance being unknown has little effect on the detection performance. This is a useful observation, since in practice the variance is often unknown to the receiver.

In closing, we mention that the strategy for the analysis of resolution we have put forward here is very generally applicable to other types of signal models such as damped sinusoidal signals. Once the signal model is decided upon, the same line of reasoning including approximations and the development of locally optimal detectors can be carried out.

Chapter 7

Conclusions and Future Directions

7.1 Summary and Concluding Remarks

This thesis has proposed statistical and information-theoretic frameworks to study the resolution limits in typical imaging systems and also in spectral estimation (or equivalently direction finding). The approaches have included detection-theoretic, estimation-theoretic, information-theoretic frameworks and also variational and perturbation analyses.

First, we considered the problem of resolvability from a detection-theoretic point of view. The underlying hypothesis testing problem here was whether a discrete, noisy, and blurred signal is originated from a single point source or two closely-spaced point sources. We developed detection strategies to solve such hypothesis testing problem and carried out the performance analysis of these strategies. The suggested perfor-

mance criterion is the relation (or trade-off) between the required SNR for resolving a given separation between point sources at a pre-specified probability of detection and false alarm rate, or alternatively the minimum detectable separation at a given SNR. One of the fundamental results we obtained is that the required SNR is proportional to the inverse of the separation between point sources raised to the fourth power. Even if the point sources do not have the same level of brightness, with an intuitively reasonable choice of the location at which the detector is (locally) implemented, this statement still holds.

Second, as for the estimation-theoretic framework, we employed the broadly used Fisher information approach to establish lower bounds on estimation of the separation between point sources and other possibly unknown parameters (such as intensity of the two point sources). We have observed that the results generated in such a framework not only verify the earlier results of the detection-theoretic methodology, but also build an interesting link between the hypothesis testing methodology and the information-theoretic measures derived later in Chapter 4.

Moreover, as a third method, we computed the Kullback-Leibler distance for the underlying hypothesis testing problem, which provides an asymptotic measure of resolvability power in connection to the Fisher information results.

Finally, we have put forward the variational analysis of the detection performance where we are interested in studying the effect of any variation of the parameters in a given imaging system on the performance.For instance we were able to compute the sensitivity of the proposed performance figure as a function of PSF variations.

Most of the developed results are presented for both one-dimensional and two-dimensional cases and also for any general scheme for sampling (sub-Nyquist and non-uniform sampling). For sub-Nyquist sampling, we have observed that the (detection or estimation) performance depends on the sampling phases (or offsets).

Overall, we wish to mention that this thesis, we hope, represents one step forward in an overall methodology for studying problems in 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.

Moreover, the approach we have advocated for determining resolution limits in imaging can be similarly used to develop statistical algorithms and performance limits for resolving sinusoids with nearby frequencies, in the presence of noise. We formulate the problem as a hypothesis test, the aim of which is to distinguish whether the received signal contains a single-tone or double-tone. We have considered the most general case where the amplitudes, frequencies and phases of sinusoids and also the value of noise variance is unknown to the detector.

By utilizing a quadratic approximation, we in fact carried out the analysis in the context of locally optimal detectors, and developed corresponding detection strategies. The performance figure of resolution has been quantified by the following practical question: "What is the minimum detectable frequency difference between two sinusoids at a given signal-to-noise ratio?"

We clearly observed that noise variance being unknown has little effect on

the detection performance. This is a useful observation, since in practice the variance is often unknown to the receiver.

Also, the proposed locally optimal detectors yield significantly improved detection of very nearby frequencies, as compared to the existing subspace methods. In terms of implementing the suggested detection algorithm, we merely need to estimate the center frequency. Fortunately, as we confirmed by some experiments, this task can be very effectively performed by using one of the myriad of existing methods for spectral estimation and then running the proposed detector at the estimated peak. The performance of such a detector is nearly identical to that of the detector with a known center frequency.

7.2 Future Directions

In this section, we briefly discuss some future aspects of the analyses developed in the thesis.

7.2.1 An application to physical fault detection in IC manufacturing

Integrated circuit manufacturing processes require transferring a circuit pattern (including logic gates, memory cells, etc) onto the silicon wafers. This process is referred to as lithography. However, because of several factors (like irregularity of surface or mask imperfection) the above process may introduce some errors in replicating the pattern. These errors translate into short or open circuits which make the manufactured IC useless. Hence, there is a need for metrology and inspection to de-

tect these physical defects or faults. The commonly employed method is to image the printed wafer using an SEM (Scanning Electron Microscope) and to inspect this image by matching with the desired, expected pattern. Over time, the dimensions of the printed circuits are becoming increasingly small (currently 65nm), and this, coupled with the resolution limits of the SEM, lead to a heavily blurred and noisy version of the desired image of the wafer.

In this section, we employ and extend the detection frameworks developed earlier in Chapter 2 to such inspection tasks. The models we put forward here suggests a detector which evaluates the correctness of a circuit or mask in a simple way. For comprehensive and practical purposes the method we propose needs to be generalized. Such generalization requires better understanding of the physical models of the underlying systems.

Let us first start with the mask inspection problem. Consider a typical mask pattern with two isolated holes as shown in Figure 7.1. The hypothesis test here is to decide whether these bars are connected or not[1]. In 1-D model, we can write the following model for the collected image:

$$g(x_k) = \alpha_m h_m\left(x_k - \frac{d}{2}\right) + \alpha_m h_m\left(x_k + \frac{d}{2}\right) + w(x_k) \tag{7.1}$$

where α_m is the peak intensity of a single bar and

$$h_m(x) = h(x) * \mathrm{rect}(x, w), \quad \mathrm{rect}(x, w) = \begin{cases} 1 & |x| < \frac{w}{2} \\ 0 & \text{otherwise} \end{cases} \tag{7.2}$$

[1] or whether the distance between the boundaries of these holes is smaller than a required threshold.

is the overall blurring kernel resulting from the imaging PSF ($h(x)$) convolved with the aperture function (or hole shape). w is the width of each hole. We can now use the developed detectors in Section 2.4 by replacing $h(x)$ with $h_m(x)$.

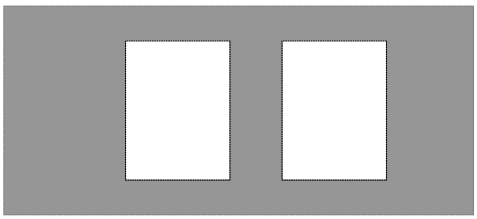

Figure 7.1: A simple scheme of a typical mask pattern

The problem in wafer inspection is however more complicated, since the printing process does not usually produce a sharp edge (perfect rectangle) in the resist profile. See for example the pattern in Figure 7.2. Therefore to have a faithful model, one should first estimate the structure of the produced pattern generated from for example a single point or bar. Perhaps the difficulty is that this pattern is changed in time and at different locations in the circuit. Also, to decide whether a wafer should be categorized as a defective one or not there are at least two relevant conditions: the distance between the patterns and the depth of the valley between them.

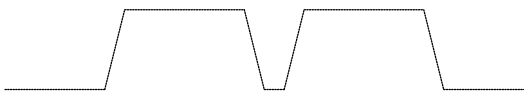

Figure 7.2: 1-D version of a resist profile of the two closely spaced bars

To attempt to overcome these difficulties, we take a rather different viewpoint from what we have proposed so far in this thesis. Let $\text{rect}(\cdot, w)$(a rectangle with width w), $h_p(\cdot)$ and $h(\cdot)$ represent the shape of the original bar to be printed, the blurring effect caused by the printing process, and the imaging PSF, respectively. First, we note that if $h_p(\cdot)$ is itself a rectangular pulse, then the resulting burned pattern $\text{rect}(\cdot) * h_p(\cdot)$ will have a trapezoidal shape[2], similar to what is shown in Figure 7.2. A block diagram of the model is shown in Figure 7.3. Now, let us assume that we know the general structure of $h_p(\cdot)$ beforehand and that there however exists a different uncertainty about stretching (or compressing) of this kernel for each printed component. The reason is that every printed component maybe affected by adjacent element and space- and time-varying manufacturing processes. To this end, the uncertainty is not about the distance between the components to be printed (we in fact know a-priori about their desired locations and other specifications), but about the stretching or compressing of the blurring function.

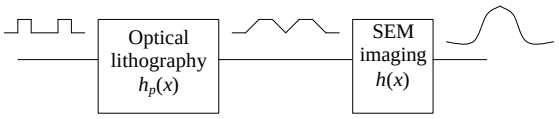

Figure 7.3: Block diagram of the underlying model

Without loss of generality, we suggest the following methodology to remarkably simplify the hypothesis test (to be specified later). To facilitate presentation, let $h_{pn}(x)$ be a canonical form of the kernel $h_p(x)$ (where for example its energy is normalized

[2]We should mention that we have ignored the effect of soft thresholding in this model.

to 1). By means of some experiments or perhaps analytical solutions, we will be able to find a scaled version of $h_{pn}(x)$ (i.e. to find the variable c_p in $h_{pn}(c_p x)$) such that it just satisfies (user-defined minimum acceptability) all the validity conditions (the distance and depth criteria and possibly others) for the printed circuits. Then we set $h_p(x) = h_{pn}(c_p x)$ as our acceptable kernel. In other words, any further stretching of $h_p(x)$ will produce incorrect resist profile. Note that as $c \to +\infty$, the kernel will shrink enough so that it will eventually pass the validity test.

Now consider two closely-spaced bars in the underlying layout, so that

$$\text{rect}_w\,(x - d_1) + \text{rect}_w\,(x + d_2) \tag{7.3}$$

is the 1-D original pattern[3] and that each of these bars experiences different blurring as discussed earlier which we denote by $h_p([1+\epsilon_1]x)$ and $h_p([1+\epsilon_2]x)$. If $\epsilon_1, \epsilon_2 > 0$, then we have correctly printed components. The measured discrete image is given by[4]

$$\begin{aligned} g(x_k) &= s(x_k, \epsilon_1, \epsilon_2) + w(x_k) && (7.4)\\ &= \text{rect}_w\,(x - d_1) * h_p([1+\epsilon_1]x) * h(x)\big|_{x=x_k} && (7.5)\\ &+ \text{rect}_w\,(x + d_2) * h_p([1+\epsilon_2]x) * h(x)\big|_{x=x_k} + w(x_k). && (7.6) \end{aligned}$$

The signal $s(x_k)$ can be globally approximated by the Taylor series around $(\epsilon_1, \epsilon_2) = (0, 0)$ as

$$s(x_k, \epsilon_1, \epsilon_2) \approx s_0(x_k, 0, 0) + \epsilon_1 s_1(x_k, 0, 0) + \epsilon_2 s_2(x_k, 0, 0) \tag{7.7}$$

[3]It is worth emphasizing that d_1 and d_2 are known here.

[4]We have assumed an additive Gaussian noise here, however a more precise model should include the effect of both Gaussian and Poisson noises[71].

where

$$s_0(x_k,0,0) = [\text{rect}_w(x-d_1)+\text{rect}_w(x+d_2)] * h_p(x) * h(x)\big|_{x=x_k} \quad (7.8)$$

$$s_1(x_k,0,0) = x\text{rect}_w(x-d_1) * \frac{\partial h_p(x)}{\partial x} * h(x)\Big|_{x=x_k} \quad (7.9)$$

$$s_2(x_k,0,0) = x\text{rect}_w(x+d_2) * \frac{\partial h_p(x)}{\partial x} * h(x)\Big|_{x=x_k} \quad (7.10)$$

or in vector form

$$\mathbf{s} = \mathbf{H}_p\theta_p \quad (7.11)$$

where

$$\mathbf{H}_p = [\mathbf{s}_0|\mathbf{s}_1|\mathbf{s}_2] \quad (7.12)$$

$$\theta_p = \begin{bmatrix} 1 \\ \epsilon_1 \\ \epsilon_2 \end{bmatrix}. \quad (7.13)$$

Finally, the hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ are respectively defined based on whether ϵ_1, ϵ_2 (or both) are negative (the printed circuit is unacceptable) or otherwise (the printed circuit is acceptable,

$$\begin{cases} \mathcal{H}_0: & \epsilon_1 < 0 \text{ or } \epsilon_2 < 0 \\ \mathcal{H}_1: & \epsilon_1 > 0 \text{ and } \epsilon_2 > 0 \end{cases}. \quad (7.14)$$

See the critical region in Figure 7.4. A careful reader will notice that this hypothesis suggests somewhat a pessimistic detector. One may think of proposing a more precise and critical region (like a smooth curve as the decision boundary).

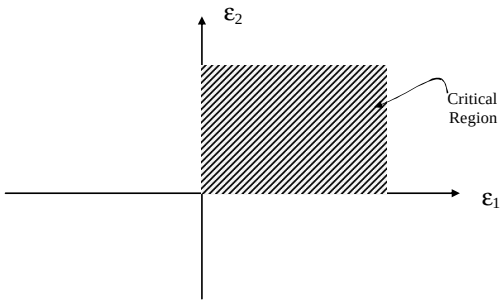

Figure 7.4: The critical region of the hypothesis test in (7.14)

7.2.2 Sparsity and Resolution

The theory of overcomplete representations has recently become a very popular area of research [6, 12, 11, 13]. Many papers have been published to establish the mathematical and statistical theories behind sparse representations and also to properly use it in signal processing applications (such as compression, inverse problems in imaging and many more). Basically, sparsity constraint for a given problem (e.g. image denoising) provides prior information which enables us to recover the underlying signal even from a set of weak observations about the signal, such as under-sampled measurements or lower-dimensional projections of the signal.

A natural problem related to sparse representations is to study the question of resolution limits in imaging or array processing applications [10]. One example is in resolving frequencies of multiple sinusoids which are placed closer than the Rayleigh interval. However, as indicated in [69], this leads to extremely ill-posed problems. What we are suggesting here to overcome the ill-posedness is to utilize the proposed local detectors. We can readily show that for the case of two sinusoids, use of the local

detector is quite straight-forward. To solve the problem for the case of more than two tones in the signal, an extension of the framework is required. As an example, one may think of modifying and using the multi-stage greedy algorithms [61, 62] in a local fashion to recover the underlying sparse multi-tone signal.

Given a unitary dictionary Φ and signal $\mathbf{g}$, we want to solve the following problem[5] (i.e. searching the "sparsest" representation):

$$\min_{\mathbf{c}} \|c\|_0 \text{ subject to } \mathbf{g} = \Phi\mathbf{c} \tag{7.15}$$

The (orthogonal) greedy algorithm is a iterative method which at step l gives an approximate l-atom representation (a representation in which exactly l atoms contribute). The process initializes by first finding the most dominant present component (atom) in the signal and then computing the residual. In general for step l we have

$$\mathbf{r}_{(l)} = \mathbf{g} - \widehat{\mathbf{g}}_{(l)}, \tag{7.16}$$

where $\mathbf{r}_{(l)}$ and $\widehat{\mathbf{g}}_{(l)}$ denote the residual and the approximated signal at step l, respectively. Also,

$$\widehat{\mathbf{g}}_{(l)} = \Phi\mathbf{c}_{(l)} \tag{7.17}$$

is the representation at step l, where $\mathbf{c}_{(l)}$ is the estimated coefficients of decomposition and has only l non-zero elements. Then at step $l+1$, a new atom is chosen which has the maximum correlation with the residual resulted from l-th step. Iteration stops when the level of the residual error is less then a pre-specified threshold. To

[5] norm 0 of a vector simply means number of non-zero elements in that vector.

employ this approach locally, one requirement is to decide whether at any stage, the most recently added atom (tone) can be further resolved to two closely-spaced atoms (or tones). Note that this hypothesis test is performed at the frequency of the most recently added component. Moreover some theoretical results as in [11] can be similarly derived for such specific problem.

7.2.3 Performance Analysis of Resolution in Indirect Imaging

It is worthwhile to note that the strategy for the analysis of resolution we have put forward in this thesis is very generally applicable to other types of imaging systems. Once the signal model of a given imaging system is determined 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.

In computed tomography, it is appealing to study the effect of SNR as well as the number of projections on the resolvability. The underlying signal can be chosen in some other forms rather than point sources. It could be for instance chosen as two concentric circles with slightly different radii.

7.2.4 Statistical Bounds on Distinguishing Details

Let $s(x;\theta)$ represent a set of signals parameterized by a "feature" vector θ. Now given a discrete, blurry, noisy version of such signal we setup the following hypothesis testing to distinguish whether the measured signal is generated by a set of parameters θ_0 or $\theta_0 + \Delta\theta$:

$$\begin{cases} \mathcal{H}_0: & \theta = \theta_0 \\ \mathcal{H}_1: & \theta = \theta + \Delta\theta \end{cases}. \tag{7.18}$$

or

$$\begin{cases} \mathcal{H}_0: & \Delta\theta = \mathbf{0} \\ \mathcal{H}_1: & \Delta\theta \neq \mathbf{0} \end{cases}. \tag{7.19}$$

We can now pose the following question "what is the smallest resolvable detail(s) at a given SNR and detection accuracy?". One possible solution is to develop locally powerful detectors by means of the Taylor approximation around a nominal value of the parameter of interest (for example $\Delta\theta = \mathbf{0}$ in the above hypothesis testing). The hypotheses here are represented by testing the value of these parameters. Alternative ways can be proposed based on the estimation-theoretic approach by computing the Fisher information matrix for the unknown parameters. Onother way is to compute the KLD and establish an asymptotic performance figure. We carried out a similar analysis in Section 4.3. Another approach is to derive and utilize the CRLB as we explain in the following example. In [55, 56], the resolution is characterized based on the minimum attainable variance (in terms of the CR bound) of estimation of the underlying parameter(s). For example in point source separation, the minimum

requirement for resolvability is defined based on whether the standard deviation of the source separation is less than the actual value of the source separation itself. In other words, the resolution limit is obtained as the separation at which such equality is achieved. Interestingly, the results in [55, 56] are similar to that of the proposed local detector we have developed mainly in Chapters 2 and 3.

We note that the presence of nuisance parameters in the model or clutter embedded in the original signal needs to be treated carefully. It is worth mentioning that the approach based on the CR bound is able to formulate and take into account the presence of the nuisance parameters.

7.2.5 Imaging System Design

A preliminary example of a system design problem is given in Section 5.2 where we discussed the effect of CCD fill factor on detection performance in a trade-off with noise level. Another interesting scenario is to study the effect of pupil filters on resolvability. Overall, the statistical analysis machinery we have developed in this thesis can be thought as a well-defined roadmap to be used in so-called PSF engineering [7].

7.2.6 Generalization to Different Noise Characteristic or Other Statistical Uncertainties

Resolution limit in photon-limited imaging systems is another important case study that needs to be investigated. The noise in these systems is no longer additive

readout noise and follows the Poisson process. Also, in many cases the underlying image itself may contain different forms of noise and in particular clutter or unwanted interference which are also described by a probabilistic function.

Another challenging direction is to investigate the performance of the imaging systems with time-varying characteristics (for example time-varying PSF particularly in astronomical applications). Such behavior may be modelled by some probabilistic descriptions as well. The analysis here requires a somewhat more sophisticated machinery.

We now briefly study the resolution problem for the case where the measured signal is contaminated by Poisson noise (for example in photon-limited imaging). This is an important case study in astronomical applications and SEM imaging.

Let $g(x_k)$ be the measured signal of the underlying (blurred) signal

$$s(x_k, d_1, d_2) = \alpha h(x_k - d_1) + \beta h(x_k + d_2)$$

and let $\alpha d_1 = \beta d_2$. The log-likelihood ratio for the hypothesis testing problem (i.e. whether $d = 0$ or $d > 0$) is given by

$$\begin{aligned} \text{LLR} &= \ln \prod_k \frac{s(x_k, d_1, d_2)^{g(x_k)} \exp(-s(x_k, d_1, d_2))/g(x_k)}{s(x_k, 0, 0)^{g(x_k)} \exp(-s(x_k, 0, 0))/g(x_k)} && (7.20) \\ &= \sum_k g(x_k) \ln \left(\frac{s(x_k, d_1, d_2)}{s(x_k, 0, 0)} \right) - s(x_k, d_1, d_2) + s(x_k, 0, 0). && (7.21) \end{aligned}$$

Approximating the log-likelihood ratio above by the Taylor expansion around $(d_1, d_2) = (0, 0)$ will result in

$$\text{LLR} \approx \frac{1}{2}(\alpha d_1^2 + \beta d_2^2) \sum_k \left(\frac{g(x_k)}{s(x_k, d_1, d_2)} - 1 \right) \left. \frac{\partial^2 s(x, d_1, d_2)}{\partial x^2} \right| x = x_k. \qquad (7.22)$$

We note that

$$\alpha d_1^2 + \beta d_2^2 = (\alpha + \beta) d_1 d_2 = \frac{\alpha\beta}{\alpha + \beta} d^2, \tag{7.23}$$

confirming the proportionality of the log-likelihood function with d^2. The remaining steps include computing the probability of detection and false alarm rate for the approximated test statistics. It is worth mentioning that this approximation is directly related to an extension of the locally most powerful test [33, p. 218] and one can straight-forwardly use the result of applying such a framework to study the performance for the current case.

Appendix A

Accuracy of Signal Approximation Using the Taylor Series

A.1 On the Accuracy of the Quadratic Approximation

Here, we present an analysis to demonstrate the accuracy of the Taylor expansion proposed in Section 2.3. We consider the general model of (2.48) and its Taylor expansion in (2.49). Let us define residual percentage error of the approximation as follows:

$$\epsilon = \frac{\left\| \mathbf{s} - (\alpha+\beta)\mathbf{h} - (-\alpha d_1 + \beta d_2)\mathbf{h}_1 - \frac{\alpha d_1^2 + \beta d_2^2}{2}\mathbf{h}_2 \right\|^2}{\|\mathbf{s}\|^2} \tag{A.1}$$

Consider the case when $\alpha d_1 = \beta d_2$ (See Appendix C). Figure A.1 shows the upper bound ($d = d_1 + d_2 = 1$) for ϵ as a function of α for $h(x) = \text{sinc}^2(x)$ (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%. Figure A.2 shows the curve for ϵ vs d for This figure 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 Figure A.3 for two different values of d over the range of the variable x in $[-10, 10]$.

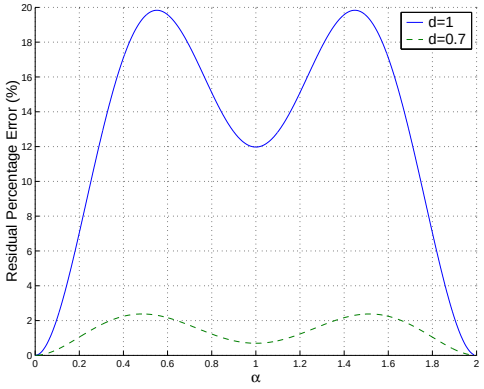

Figure A.1: Residual percentage error of the quadratic model; $\alpha d_1 = \beta d_2$

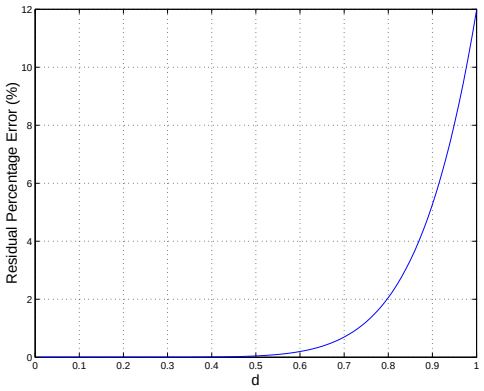

Figure A.2: Residual percentage error of the quadratic model; $\alpha = \beta = 1$

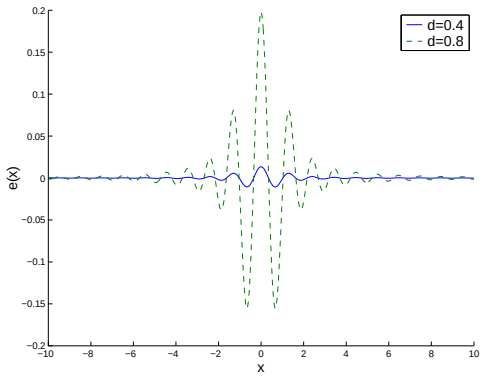

Figure A.3: Difference between the actual signal and the quadratic model; $\alpha = \beta = 1$

Appendix B

Frequency Domain Derivations for Computing the Energy Terms

B.1 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:

$$s(n;\alpha,\beta,d_1,d_2) = s(x;\alpha,\beta,d_1,d_2)|_{x=\frac{n}{f_s}} = \alpha h\left(\frac{n}{f_s} - d_1\right) + \beta h\left(\frac{n}{f_s} + d_2\right) \quad \text{(B.1)}$$

For the case of above-Nyquist sampling[1], in the frequency domain we will have the following 2π-periodic representation:

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

where $H(\omega, f_s) = \frac{f_s^2}{2\pi}(\frac{2\pi}{f_s} - |\omega|)$ is the DTFT of $h(x_k)$ when $h(x) = \text{sinc}^2(x)$ 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

$$H_1(\omega, f_s) = \begin{cases} j\frac{\omega f_s^3}{2\pi}(\frac{2\pi}{f_s} - |\omega|) & |\omega| < \frac{2\pi}{fs} \\ 0 & \frac{2\pi}{fs} \leq |\omega| \leq 2\pi \end{cases} \tag{B.3}$$

$$H_2(\omega, f_s) = \begin{cases} -\frac{\omega^2 f_s^4}{2\pi}(\frac{2\pi}{f_s} - |\omega|) & |\omega| < \frac{2\pi}{fs} \\ 0 & \frac{2\pi}{fs} \leq |\omega| \leq 2\pi \end{cases} \tag{B.4}$$

Using Parseval's identities [47]:

$$\sum_{n=-\infty}^{\infty} |x(n)|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |X(\omega)|^2 d\omega \tag{B.5}$$

$$\sum_{n=-\infty}^{\infty} x(n)y^*(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} X(\omega)Y^*(\omega) d\omega \tag{B.6}$$

we can easily compute the following terms

$$E_0 = \mathbf{h}^T\mathbf{h} = f_s\frac{2}{3} \tag{B.7}$$

$$E_1 = \mathbf{h}_1^T\mathbf{h}_1 = f_s\frac{4\pi^2}{15} \tag{B.8}$$

$$E_2 = \mathbf{h}_2^T\mathbf{h}_2 = f_s\frac{32\pi^4}{105} \tag{B.9}$$

[1]To recover exactly $s(x; d)$ would mathematically require an infinite number of measurements (or samples) $s(n; d)$ [66]. 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.

and

$$\mathbf{h}_1^T \mathbf{s}_0 = \mathbf{h}_1^T \mathbf{h}_2 = 0 \tag{B.10}$$

Note that in every case the energy terms are proportional to the sampling rate. It can be shown [64] that the energy of any uniformly (super-critically) sampled version of a band-limited signal is proportional to sampling rate.

Appendix C

Is $\alpha d_1 \approx \beta d_2$ a reasonable assumption?

This appendix includes a proof for justifying the assumption $\alpha d_1 \approx \beta d_2$ which states the estimated test point is (linearly proportionally) closer to the stronger peak. We present the proof for both cases of localizing the center for resolution in imaging and direction finding.

C.1 Proof for determining the test point in resolving two point sources

Suppose that we first wish to determine a location at which in signal 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).

$$\begin{aligned} R_{sh}(|\tau|, \alpha, \beta, d_1, d_2) &= \int_{-\infty}^{+\infty} \left(s(x; \alpha, \beta, d_1, d_2) + w(x)\right) h(x+\tau) dx \qquad \text{(C.1)} \\ &= \int_{-\infty}^{+\infty} \left(\alpha h\left(x - d_1\right) + \beta h\left(x + d_2\right) + w(x)\right) h(x+\tau) dx \\ &= \alpha R_{hh}(|\tau| - d_1) + \beta R_{hh}(|\tau| + d_2) + u(|\tau|) \qquad \text{(C.2)} \end{aligned}$$

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

$$u(|\tau|) = \int_{-\infty}^{+\infty} w(x) h(x+\tau) dx \qquad \text{(C.3)}$$

is a noise term (with zero mean). It must 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:

$$R_{hh}(|\tau| - d_1) = \xi_0 + (|\tau| - d_1)\xi_1 + (|\tau| - d_1)^2 \xi_2 \qquad \text{(C.4)}$$

$$R_{hh}(|\tau| + d_2) = \xi_0 + (|\tau| + d_2)\xi_1 + (|\tau| + d_2)^2 \xi_2 \qquad \text{(C.5)}$$

where ξ_0, ξ_1 and ξ_3 are some constant coefficients of the above Taylor expansion. Also, it can be shown that $\xi_1 = 0$. Therefore, we can write (C.2) as follows:

$$R_{sh}(|\tau|, \alpha, \beta, d_1, d_2) = (\alpha + \beta)\xi_0 + \left(\alpha(|\tau| - d_1)^2 + \beta(|\tau| - d_2)^2\right) \xi_2 + u(|\tau|) \qquad \text{(C.6)}$$

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

$$(\alpha + \beta)|\tau| = \alpha d_1 - \beta d_2 \qquad \text{(C.7)}$$

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

C.2 Proof for determining the center frequency in resolving two harmonic signals

Suppose $\phi_1 \approx \phi_2$. Then the magnitude of the discrete-Time Fourier Transform of the signal is given by [47]

$$\begin{aligned} F(f, f_c, a_1, a_2, \delta_1, \delta_2) &= S(f, f_c, a_1, a_2, \delta_1, \delta_2) + W(f) \\ &= a_1 \frac{\sin\left[\pi N\left(f - \frac{f_c - \delta_1}{f_s}\right)\right]}{\sin\left[\pi\left(f - \frac{f_c - \delta_1}{f_s}\right)\right]} + a_2 \frac{\sin\left[\pi N\left(f - \frac{f_c + \delta_2}{f_s}\right)\right]}{\sin\left[\pi\left(f - \frac{f_c + \delta_2}{f_s}\right)\right]} + W(f). \end{aligned} \tag{C.8}$$

where $W(f)$ is the Fourier domain representation of $w(x_k)$. A reasonable way to find a good candidate center frequency (f_c), where we can perform our test, is to compute the correlation of the signal with the following window in the frequency domain,

$$G(f, f_x) = \frac{\sin\left[\pi N\left(f - \frac{f_x}{f_s}\right)\right]}{\sin\left[\pi\left(f - \frac{f_x}{f_s}\right)\right]} \tag{C.9}$$

and find the point where the correlation is maximum (this would yield a point near the stronger of the two peaks). Consider

$$\begin{aligned} R_{SG}(|f_c - f_x|, a_1, a_2, \delta_1, \delta_2) &= \int_{-\infty}^{+\infty} \left[S(f, f_c, a_1, a_2, \delta_1, \delta_2) + W(f)\right] G(f, f_x) \mathrm{d}f \\ &= a_1 R_{GG}(|f_c - \delta_1 - f_x|) + a_2 R_{GG}(|f_c + \delta_2 - f_x|) \\ &+ R_{WG}(|f_x|) \end{aligned}$$

where R_{SG}, R_{WG} and R_{GG} are the cross-correlation and autocorrelation functions defined as:

$$R_{GG}(|f_c - \delta_2 - f_x|) = \int_{-\infty}^{+\infty} G(f, f_x) G(f, f_c - \delta_2) \mathrm{d}f. \tag{C.10}$$

$$R_{GG}(|f_c+\delta_2-f_x|) = \int_{-\infty}^{+\infty} G(f,f_x)G(f,f_c+\delta_2)\mathrm{d}f. \quad \text{(C.11)}$$

$$R_{WG}(|f_c-\delta_1-f_x|) = \int_{-\infty}^{+\infty} G(f,f_x)W(f)\mathrm{d}f \quad \text{(C.12)}$$

Since $\delta_1, \delta_2 \ll \frac{1}{B}$ and f_x is expected to be close to f_c, we can again use the Taylor expansion for (C.10) around $(f_c-\delta_1+f_x, f_c+\delta_2-f_x) = (0,0)$,

$$R_{GG}(|f_c-\delta_1-f_x|) \approx \xi_0 + (|f_c-\delta_1-f_x|)\xi_1 + (f_c-\delta_1-f_x)^2\xi_2 \quad \text{(C.13)}$$

$$R_{GG}(|f_c+\delta_2-f_x|) \approx \xi_0 + (|f_c+\delta_2-f_x|)\xi_1 + (f_c+\delta_2-f_x)^2\xi_2, \quad \text{(C.14)}$$

where ξ_0, ξ_1 and ξ_3 are some constant coefficients of the above Taylor expansion. Also, it can be shown that $\xi_1 = 0$. Therefore, we can write (C.10) as follows:

$$\begin{aligned} R_{SG}(|f_c-f_x|, a_1, a_2, \delta_1, \delta_2) &\approx (a_1+a_2)\xi_0 \quad \text{(C.15)} \\ &+ \left(a_1(f_c-f_x-\delta_1)^2 + a_2(f_c-f_x+\delta_2)^2\right)\xi_2 \end{aligned}$$

Taking derivative of the right hand side of (C.15) with respect to f_x and setting it to zero will result in

$$(a_1+a_2)(f_c-f_x) \approx a_1\delta_1 - a_2\delta_2. \quad \text{(C.16)}$$

which suggests to select f_c such that lead to $a_1\delta_1 \approx a_2\delta_2$. It is worth noting that application of any subspace-based method to the data will also provide a reasonable candidate for the center frequency f_c, as discussed in Section 6.4.2.

Appendix D

Sampling Theory, Derivations for Under-Nyquist Case

D.1 Sampling theory for Under-Nyquist Measurements

To begin, let us first mention that for the case of over-Nyquist sampling, following relationships are hold:

$$E_{ij} = \mathbf{h}_{ij}^T \mathbf{h}_{ij} = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} u^i v^j |H(u,v)|^2 \mathrm{d}u\mathrm{d}v \tag{D.1}$$

$$= f_s^2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left[\frac{\partial^{i+j} p(x,y)}{\partial x^i \partial y^j} \right]^2 \mathrm{d}x\mathrm{d}y \tag{D.2}$$

Whereas in under-Nyquist case, assuming that $H(u,v)$ is band limited to $-B_u < u < B_u$ and $-B_v < v < B_v$, we will have:

$$E_0 = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{1}{L_u L_v} \left| \sum_{i=0}^{L_u-1} \sum_{j=0}^{L_v-1} H(u-2\pi i, v-2\pi j) \right|^2 \mathrm{d}u\mathrm{d}v \tag{D.3}$$

$$= \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{1}{L_u L_v} \sum_{i=0}^{L_u-1} \sum_{j=0}^{L_v-1} |H(u-2\pi j, v-2\pi j)|^2$$

$$
\begin{aligned}
&+ \quad \frac{2}{L_u L_v} \sum_{i=0}^{L_u-1} \sum_{i=0}^{L_v-1} \sum_{i'=i+1}^{L_u-1} \sum_{j'=j+1}^{L_v-1} \Re\left[H(u-2\pi i, v-2\pi j) H^*(u-2\pi i', v-2\pi j')\right] \mathrm{d}u\mathrm{d}v \\
&= \quad f_s^2 \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h^2(x,y)\mathrm{d}x\mathrm{d}y \\
&+ \quad \left\{ \begin{array}{l} \dfrac{2}{L_u L_v} \displaystyle\sum_{i=0}^{L_u-1} \sum_{i=0}^{L_v-1} \sum_{i'=i+1}^{L_u-1} \sum_{j'=j+1}^{L_v-1} \cos\left((i-i')\phi + (j-j')\psi\right) \\ \displaystyle\int_{-\pi}^{\pi} \int_{-\pi}^{\pi} H(u-2\pi i, v-2\pi j) H(u-2\pi i', v-2\pi j')\mathrm{d}u\mathrm{d}v \end{array} \right\}
\end{aligned}
\tag{D.4}
$$

where $L_u = \lceil \frac{2B_u}{f_s} \rceil$, $L_v = \lceil \frac{2B_v}{f_s} \rceil$, and ϕ and ψ are the sampling phases in vertical and horizontal directions $0 \leq \phi,\ \psi \leq 2\pi$. For instance if $B_u < f_s < 2B_u$ and $B_v < f_s < 2B_v$ we will similarly have:

$$
E_{ij} \quad = \quad E_{ij}^0 + E_{ij}^v \cos(\phi) + E_{ij}^u \cos(\psi) + E_{ij}^{uv} \cos(\phi + \psi) \tag{D.5}
$$

Appendix E

Angular-Symmetric PSF

E.1 Angular Symmetric PSF

Let us consider the case of angular symmetric PSF ($h(x,y) = q(\sqrt{x^2+y^2}) = q(r)$) such as that corresponding to circular aperture. Following relationships for the partial derivatives can be written in the polar form (r, θ):

$$\begin{aligned}
\frac{\partial h(x,y)}{\partial x} &= \frac{\partial q(r)}{\partial r}\cos(\theta) \\
\frac{\partial h(x,y)}{\partial y} &= \frac{\partial q(r)}{\partial r}\sin(\theta) \\
\frac{\partial^2 h(x,y)}{\partial x^2} &= \frac{\partial^2 q(r)}{\partial r^2}\cos^2(\theta) + \frac{\partial q(r)}{\partial r}\frac{\sin^2(\theta)}{r} \\
\frac{\partial^2 h(x,y)}{\partial y^2} &= \frac{\partial^2 q(r)}{\partial r^2}\sin^2(\theta) + \frac{\partial q(r)}{\partial r}\frac{\cos^2(\theta)}{r} \\
\frac{\partial^2 h(x,y)}{\partial y \partial x} &= \frac{\partial^2 q(r)}{\partial r^2}\sin(\theta)\cos(\theta) - \frac{\partial q(r)}{\partial r}\frac{\sin(\theta)\cos(\theta)}{r}
\end{aligned}$$

After doing some math and by applying

$$\int_0^{2\pi} \sin(\theta)\cos(\theta)\mathrm{d}\theta = \int_0^{2\pi} \sin(\theta)\cos^3(\theta)\mathrm{d}\theta = 0$$

$$\int_0^{2\pi} \sin^2(\theta)\mathrm{d}\theta = \int_0^{2\pi} \cos^2(\theta)\mathrm{d}\theta = \pi$$
$$\int_0^{2\pi} \sin^4(\theta)\mathrm{d}\theta = \int_0^{2\pi} \cos^4(\theta)\mathrm{d}\theta = \frac{3\pi}{4}$$
$$\int_0^{2\pi} \sin^2(\theta)\cos^2(\theta)\mathrm{d}\theta = \frac{\pi}{4}$$

We can show that

$$E_0 = \int_0^{+\infty} rq^2(r)\mathrm{d}r \tag{E.1}$$
$$E_{10} = E_{01} = \pi \int_0^{+\infty} r\left[\frac{\partial q(r)}{\partial r}\right]^2 \mathrm{d}r \tag{E.2}$$
$$E_{20} = E_{02} = \frac{\pi}{4}\int_0^{+\infty} \frac{3}{r}\left[\frac{\partial q(r)}{\partial r}\right]^2 + 3r\left[\frac{\partial^2 q(r)}{\partial r^2}\right]^2 \mathrm{d}r \tag{E.3}$$
$$E_{11} = \frac{\pi}{4}\int_0^{+\infty} \frac{1}{r}\left[\frac{\partial q(r)}{\partial r}\right]^2 + r\left[\frac{\partial^2 q(r)}{\partial r^2}\right]^2 \mathrm{d}r \tag{E.4}$$

by using

$$\int_0^{+\infty} \frac{\partial q(r)}{\partial r}\frac{\partial^2 q(r)}{\partial r^2}\mathrm{d}r = 0$$

Therefore

$$\mathrm{SNR} = \frac{\lambda(P_f, P_d)}{N^2}\frac{64E_0 - 16d^2E_{10} + d^4E_{20}}{d^4\left(E_{20} - \frac{E_{10}^2}{E_0}\right)} \tag{E.5}$$

Appendix F

Derivations of CRLB

F.1 Derivation of the Fisher Information Matrix

Let Ω_1 and Ω_2 denote the sets of samples of the first and the second frames, respectively. From (4.10), the Fisher Information matrix can be written as

$$\mathbf{\Lambda} = \mathbf{Z}^T\mathbf{Z} \tag{F.1}$$

where $\mathbf{Z} = [\mathbf{z}_1,\ \mathbf{z}_2,\ \mathbf{z}_3,\ ...,\ \mathbf{z}_8]$ is defined in the following form

$$[\mathbf{z}_1]_{lN+k} = \left.\frac{\partial s(x,y)}{\partial p_x}\right|_{x=x_k,y=y_l} = -\alpha h_{10}(x_k - p_x, y_l - p_y) \tag{F.2}$$

$$[\mathbf{z}_2]_{lN+k} = \left.\frac{\partial s(x,y)}{\partial q_x}\right|_{x=x_k,y=y_l} = \beta h_{10}(x_k + q_x, y_l + q_y) \tag{F.3}$$

$$[\mathbf{z}_3]_{lN+k} = \left.\frac{\partial s(x,y)}{\partial p_y}\right|_{x=x_k,y=y_l} = -\alpha h_{01}(x_k - p_x, y_l - p_y) \tag{F.4}$$

$$[\mathbf{z}_4]_{lN+k} = \left.\frac{\partial s(x,y)}{\partial q_y}\right|_{x=x_k,y=y_l} = \beta h_{01}(x_k + q_x, y_l + q_y) \tag{F.5}$$

$$[\mathbf{z}_5]_{lN+k} = \left.\frac{\partial s(x,y)}{\partial \alpha}\right|_{x=x_k,y=y_l} = h(x_k - p_x, y_l - p_y) \tag{F.6}$$

$$[\mathbf{z}_6]_{lN+k} = \left.\frac{\partial s(x,y)}{\partial \beta}\right|_{x=x_k,y=y_l} = h(x_k + q_x, y_l + q_y) \tag{F.7}$$

$$[\mathbf{z}_7]_{lN+k} = \left.\frac{\partial s(x,y)}{\partial \phi_2}\right|_{x=x_k, y=y_l} \tag{F.8}$$

$$= \begin{cases} 0 & (x_k, y_l) \in \Omega_1 \\ \alpha h_{10}(x_k - p_x, y_l - p_y) + \beta h_{10}(x_k + q_x, y_l + q_y) & (x_k, y_l) \in \Omega_2 \end{cases}$$

$$[\mathbf{z}_8]_{lN+k} = \left.\frac{\partial s(x,y)}{\partial \psi_2}\right|_{x=x_k, y=y_l} \tag{F.9}$$

$$= \begin{cases} 0 & (x_k, y_l) \in \Omega_1 \\ \alpha h_{01}(x_k - p_x, y_l - p_y) + \beta h_{01}(x_k + q_x, y_l + q_y) & (x_k, y_l) \in \Omega_2 \end{cases}$$

Appendix G

Details of KLD computation

In this appendix we present the derivations of KLD for two somewhat similar problems discussed in this thesis. The former is for the case of resolution limit in a 2-D imaging scenario and the latter is for the case of resolvability is spectral domain.

G.1 Computing the Kullback-Leibler Distance in (4.13)

Directly using the results in [36, p. 26], we can obtain the following expression for KLD:

$$\mathcal{J}(d) \approx d^2 I(0), \tag{G.1}$$

where $I(d)$ is the Fisher Information measure [32, p 40],

$$I(d) = \frac{1}{\sigma^2} \sum_k \sum_l \left(\frac{\partial s(x_k, y_l)}{\partial d} \right)^2 \tag{G.2}$$

However, for the hypothesis test of interest in (3.5), $I(0)$ is zero and (G.1) is not directly applicable. Here we extend the approach in [36, p. 26] by considering higher

order terms. Consider the following Taylor expansion:

$$\begin{aligned}\mathcal{J}(d) &= \int_{\mathcal{D}} [p(\mathbf{g},d) - p(\mathbf{g},0)] \log\left(\frac{p(\mathbf{g},d)}{p(\mathbf{g},0)}\right) \mathrm{d}\mathbf{g} \qquad \text{(G.3)}\\ &= \mathcal{J}(0) + d\,\frac{\partial \mathcal{J}}{\partial d}\bigg|_{d=0} + \frac{d^2}{2}\,\frac{\partial^2 \mathcal{J}}{\partial d^2}\bigg|_{d=0} + \frac{d^3}{6}\,\frac{\partial^3 \mathcal{J}}{\partial d^3}\bigg|_{d=0} + \frac{d^4}{24}\,\frac{\partial^4 \mathcal{J}}{\partial d^4}\bigg|_{d=0} + O(d^6).\end{aligned}$$

Noting that[1]

$$\frac{\partial^i p(\mathbf{g},d)}{\partial d^i}\bigg|_{d=0} = 0 \quad i = 1,\ 3, \qquad \text{(G.4)}$$

We will have

$$\begin{aligned}\mathcal{J}(0) &= 0,\\ \frac{\partial \mathcal{J}}{\partial d}\bigg|_{d=0} &= \int_{\mathcal{D}} \frac{\partial p(\mathbf{g},d)}{\partial d} \log\left(\frac{p(\mathbf{g},d)}{p(\mathbf{g},0)}\right) + [p(\mathbf{g},d) - p(\mathbf{g},0)]\,\frac{\dfrac{\partial p(\mathbf{g},d)}{\partial d}}{p(\mathbf{g},d)}\bigg|_{d=0} \mathrm{d}\mathbf{g} = 0,\\ \frac{\partial^2 \mathcal{J}}{\partial d^2}\bigg|_{d=0} &= \int_{\mathcal{D}} \frac{2\left[\dfrac{\partial p(\mathbf{g},d)}{\partial d}\right]^2}{p(\mathbf{g},d)}\bigg|_{d=0} \mathrm{d}\mathbf{g} = 0,\\ \frac{\partial^3 \mathcal{J}}{\partial d^3}\bigg|_{d=0} &= \int_{\mathcal{D}} \frac{6\dfrac{\partial p(\mathbf{g},d)}{\partial d}\dfrac{\partial^2 p(\mathbf{g},d)}{\partial d^2}}{p(\mathbf{g},d)} - \frac{3\left[\dfrac{\partial p(\mathbf{g},d)}{\partial d}\right]^3}{[p(\mathbf{g},d)]^2}\bigg|_{d=0} \mathrm{d}\mathbf{g} = 0,\\ \frac{\partial^4 \mathcal{J}}{\partial d^4}\bigg|_{d=0} &= \int_{\mathcal{D}} \frac{8\dfrac{\partial p(\mathbf{g},d)}{\partial d}\dfrac{\partial^3 p(\mathbf{g},d)}{\partial d^3} + 6\left[\dfrac{\partial^2 p(\mathbf{g},d)}{\partial d^2}\right]^3}{p(\mathbf{g},d)}\\ &\qquad - \frac{18\left[\dfrac{\partial p(\mathbf{g},d)}{\partial d}\right]^2 \dfrac{\partial^2 p(\mathbf{g},d)}{\partial d^2}}{[p(\mathbf{g},d)]^2} + \frac{8\left[\dfrac{\partial p(\mathbf{g},d)}{\partial d}\right]^4}{[p(\mathbf{g},d)]^3}\bigg|_{d=0} \mathrm{d}\mathbf{g},\\ &= \int_{\mathcal{D}} \frac{6\left[\dfrac{\partial^2 p(\mathbf{g},d)}{\partial d^2}\right]^2}{p(\mathbf{g},d)}\bigg|_{d=0} \mathrm{d}\mathbf{g}.\end{aligned}$$

[1]Since $p(\mathbf{g},d)$ is an even (and differentiable) function around $d = 0$.

As a result, we can write (G.3) as

$$\mathcal{J}(d) \approx \frac{d^4}{4} \int_{\mathcal{D}} \left. \frac{\left[\dfrac{\partial^2 p(\mathbf{g},d)}{\partial d^2}\right]^2}{p(\mathbf{g},d)} \right|_{d=0} \mathrm{d}\mathbf{g}. \tag{G.5}$$

On the other hand, we observe that

$$\begin{aligned}
\left.\frac{\partial^2 I(d)}{\partial d^2}\right|_{d=0} &= -\frac{\partial^2}{\partial d^2} \int_{\mathcal{D}} \left. \frac{\partial^2 \ln p(\mathbf{g},d)}{\partial d^2} p(\mathbf{g},d) \right|_{d=0} \mathrm{d}\mathbf{g} \\
&= \frac{\partial^2}{\partial d^2} \int_{\mathcal{D}} \left. \frac{\left[\dfrac{\partial \ln p(\mathbf{g},d)}{\partial d}\right]^2}{p(\mathbf{g},d)} \right|_{d=0} \mathrm{d}\mathbf{g} \\
&= \int_{\mathcal{D}} \left. \frac{2\dfrac{\partial p(\mathbf{g},d)}{\partial d}\dfrac{\partial^3 p(\mathbf{g},d)}{\partial d^3} + 2\left[\dfrac{\partial^2 p(\mathbf{g},d)}{\partial d^2}\right]^3}{p(\mathbf{g},d)} \right. \\
&\qquad \left. -\frac{5\left[\dfrac{\partial p(\mathbf{g},d)}{\partial d}\right]^2 \dfrac{\partial^2 p(\mathbf{g},d)}{\partial d^2}}{[p(\mathbf{g},d)]^2} + \frac{2\left[\dfrac{\partial p(\mathbf{g},d)}{\partial d}\right]^4}{[p(\mathbf{g},d)]^3} \right|_{d=0} \mathrm{d}\mathbf{g} \\
&= \int_{\mathcal{D}} \left. \frac{2\left[\dfrac{\partial^2 p(\mathbf{g},d)}{\partial d^2}\right]^2}{p(\mathbf{g},d)} \right|_{d=0} \mathrm{d}\mathbf{g}.
\end{aligned} \tag{G.6}$$

Therefore,

$$\begin{aligned}
\mathcal{J}(d) &\approx \frac{d^4}{8} \left.\frac{\partial^2 I(d)}{\partial d^2}\right|_{d=0} \qquad (G.7) \\
&= \frac{d^4}{4\sigma^2} \sum_k \sum_l \left(\left.\frac{\partial^2 s(x_k,y_l)}{\partial d^2}\right)^2 \right|_{d=0} + \left. \frac{\partial s(x_k,y_l)}{\partial d} \frac{\partial^3 s(x_k,y_l)}{\partial d^3} \right|_{d=0} \\
&= \frac{d^4}{4\sigma^2} \sum_k \sum_l \left(\left.\frac{\partial^2 s(x_k,y_l)}{\partial d^2}\right)^2 \right|_{d=0} + 0 \\
&= \frac{d^4}{4\sigma^2} \sum_k \sum_l \left(\left.\frac{\partial^2 \left[h(x_k+d/2,y_l)+h(x_k-d/2,y_l)\right]}{\partial d^2}\right)^2 \right|_{d=0}
\end{aligned}$$

$$= \frac{d^4}{4\sigma^2} \sum_k \sum_l h_{20}^2(x_k, y_l)\Big|_{d=0} \tag{G.8}$$

As we see from (G.7), the divergence for the underlying hypothesis testing problem is directly related to the second derivative of the Fisher information matrix evaluated at $d = 0$.

G.2 Computing the Kullback-Leibler Distance in (6.65)

Similar to the previous section, using the results in [36, p. 26], we have

$$J(\delta) \approx \delta^2 I(0), \tag{G.9}$$

where $I(\delta)$ is the Fisher Information measure defined as,

$$I(\delta) = -\mathsf{E}\left[\frac{\partial^2 \ln p(\mathbf{f}, \delta)}{\partial \delta^2}\right] = \frac{1}{2}\mathsf{t}r\left[\left(\left[\mathbf{R}_1 + \sigma^2\mathbf{I}\right]^{-1} \frac{\partial \mathbf{R}_1}{\partial \delta}\right)^2\right] \tag{G.10}$$

We again note that for the hypothesis test of interest in (6.61), $I(0)$ is zero and that an extension is required. Consider the following Taylor expansion:

$$\begin{aligned} J(\delta) &= \int_{\mathcal{D}} [p(\mathbf{f}, \delta) - p(\mathbf{f}, 0)] \log\left(\frac{p(\mathbf{f}, \delta)}{p(\mathbf{f}, 0)}\right) \mathrm{d}\mathbf{f} \qquad \text{(G.11)} \\ &= J(0) + \delta \left.\frac{\partial J}{\partial \delta}\right|_{\delta=0} + \frac{\delta^2}{2} \left.\frac{\partial^2 J}{\partial \delta^2}\right|_{\delta=0} + \frac{\delta^3}{6} \left.\frac{\partial^3 J}{\partial \delta^3}\right|_{\delta=0} + \frac{\delta^4}{24} \left.\frac{\partial^4 J}{\partial \delta^4}\right|_{\delta=0} + O(\delta^6). \end{aligned}$$

Also

$$\left.\frac{\partial^i p(\mathbf{f}, \delta)}{\partial \delta^i}\right|_{\delta=0} = 0 \quad i = 1,\ 3,$$

Therefore

$$J(0) = 0,$$

$$
\begin{aligned}
\left.\frac{\partial J}{\partial \delta}\right|_{\delta=0} &= \int_{\mathcal{D}} \frac{\partial p(\mathbf{f},\delta)}{\partial \delta} \log\left(\frac{p(\mathbf{f},\delta)}{p(\mathbf{f},0)}\right) + [p(\mathbf{f},\delta) - p(\mathbf{f},0)] \left.\frac{\frac{\partial p(\mathbf{f},\delta)}{\partial \delta}}{p(\mathbf{f},\delta)}\right|_{\delta=0} \mathrm{d}\mathbf{f} = 0, \\
\left.\frac{\partial^2 J}{\partial \delta^2}\right|_{\delta=0} &= \int_{\mathcal{D}} \left.\frac{2\left[\frac{\partial p(\mathbf{f},\delta)}{\partial \delta}\right]^2}{p(\mathbf{f},\delta)}\right|_{\delta=0} \mathrm{d}\mathbf{f} = 0, \\
\left.\frac{\partial^3 J}{\partial \delta^3}\right|_{\delta=0} &= \int_{\mathcal{D}} \left.\frac{6\frac{\partial p(\mathbf{f},\delta)}{\partial \delta}\frac{\partial^2 p(\mathbf{f},\delta)}{\partial \delta^2}}{p(\mathbf{f},\delta)} - \frac{3\left[\frac{\partial p(\mathbf{f},\delta)}{\partial \delta}\right]^3}{[p(\mathbf{f},\delta)]^2}\right|_{\delta=0} \mathrm{d}\mathbf{f} = 0, \\
\left.\frac{\partial^4 J}{\partial \delta^4}\right|_{\delta=0} &= \int_{\mathcal{D}} \left.\frac{8\frac{\partial p(\mathbf{f},\delta)}{\partial \delta}\frac{\partial^3 p(\mathbf{f},\delta)}{\partial \delta^3} + 6\left[\frac{\partial^2 p(\mathbf{f},\delta)}{\partial \delta^2}\right]^3}{p(\mathbf{f},\delta)}\right. \\
&\quad \left. - \frac{18\left[\frac{\partial p(\mathbf{f},\delta)}{\partial \delta}\right]^2\frac{\partial^2 p(\mathbf{f},\delta)}{\partial \delta^2}}{[p(\mathbf{f},\delta)]^2} + \frac{8\left[\frac{\partial p(\mathbf{f},\delta)}{\partial \delta}\right]^4}{[p(\mathbf{f},\delta)]^3}\right|_{\delta=0} \mathrm{d}\mathbf{f}, \\
&= \int_{\mathcal{D}} \left.\frac{6\left[\frac{\partial^2 p(\mathbf{f},\delta)}{\partial \delta^2}\right]^2}{p(\mathbf{f},\delta)}\right|_{\delta=0} \mathrm{d}\mathbf{f}.
\end{aligned}
$$

As a result, we can write (G.11) as

$$
J(\delta) \approx \frac{\delta^4}{4} \int_{\mathcal{D}} \left.\frac{\left[\frac{\partial^2 p(\mathbf{f},\delta)}{\partial \delta^2}\right]^2}{p(\mathbf{f},\delta)}\right|_{\delta=0} \mathrm{d}\mathbf{f}. \tag{G.12}
$$

On the other hand, we observe that

$$
\begin{aligned}
\left.\frac{\partial^2 I(\delta)}{\partial \delta^2}\right|_{\delta=0} &= -\frac{\partial^2}{\partial \delta^2}\int_{\mathcal{D}} \left.\frac{\partial^2 \ln p(\mathbf{f},\delta)}{\partial \delta^2} p(\mathbf{f},\delta)\right|_{\delta=0} \mathrm{d}\mathbf{f} \\
&= \frac{\partial^2}{\partial \delta^2}\int_{\mathcal{D}} \left.\frac{\left[\frac{\partial \ln p(\mathbf{f},\delta)}{\partial \delta}\right]^2}{p(\mathbf{f},\delta)}\right|_{\delta=0} \mathrm{d}\mathbf{f}
\end{aligned}
$$

$$
\begin{aligned}
&= \int_{\mathcal{D}} \frac{2\dfrac{\partial p(\mathbf{f},\delta)}{\partial \delta}\dfrac{\partial^3 p(\mathbf{f},\delta)}{\partial \delta^3} + 2\left[\dfrac{\partial^2 p(\mathbf{f},\delta)}{\partial \delta^2}\right]^3}{p(\mathbf{f},\delta)} \\
&\quad - \frac{5\left[\dfrac{\partial p(\mathbf{f},\delta)}{\partial \delta}\right]^2 \dfrac{\partial^2 p(\mathbf{f},\delta)}{\partial \delta^2}}{[p(\mathbf{f},\delta)]^2} + \frac{2\left[\dfrac{\partial p(\mathbf{f},\delta)}{\partial \delta}\right]^4}{[p(\mathbf{f},\delta)]^3}\Bigg|_{\delta=0} \mathrm{d}\mathbf{f} \\
&= \int_{\mathcal{D}} \frac{2\left[\dfrac{\partial^2 p(\mathbf{f},\delta)}{\partial \delta^2}\right]^2}{p(\mathbf{f},\delta)}\Bigg|_{\delta=0} \mathrm{d}\mathbf{f}.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
J(\delta) &\approx \frac{\delta^4}{8}\left.\frac{\partial^2 I(\delta)}{\partial \delta^2}\right|_{\delta=0} \\
&= \frac{\delta^4}{16}\mathsf{t}r\left[\frac{\partial^2}{\partial \delta^2}\left(\left[\mathbf{R}_1+\sigma^2\mathbf{I}\right]^{-1}\frac{\partial \mathbf{R}_1}{\partial \delta}\right)^2\Bigg|_{\delta=0}\right] \\
&= \frac{\delta^4}{16}\mathsf{t}r\left[2\left(\frac{\partial \mathbf{R}}{\partial \delta}\right)^2 + \frac{\partial^2 \mathbf{R}}{\partial \delta^2}\mathbf{R} + \frac{\partial^2 \mathbf{R}}{\partial \delta^2}\mathbf{R}\Bigg|_{\delta=0}\right],
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{R} &= \left[\mathbf{R}_1+\sigma^2\mathbf{I}\right]^{-1}\frac{\partial \mathbf{R}_1}{\partial \delta} \\
\frac{\partial \mathbf{R}}{\partial \delta} &= -\left(\left[\mathbf{R}_1+\sigma^2\mathbf{I}\right]^{-1}\frac{\partial \mathbf{R}_1}{\partial \delta}\right)^2 + \left[\mathbf{R}_1+\sigma^2\mathbf{I}\right]^{-1}\frac{\partial^2 \mathbf{R}_1}{\partial \delta^2} \\
\frac{\partial^2 \mathbf{R}}{\partial \delta^2} &= 2\left(\left[\mathbf{R}_1+\sigma^2\mathbf{I}\right]^{-1}\frac{\partial \mathbf{R}_1}{\partial \delta}\right)^3 - 2\left[\mathbf{R}_1+\sigma^2\mathbf{I}\right]^{-1}\frac{\partial \mathbf{R}_1}{\partial \delta}\left[\mathbf{R}_1+\sigma^2\mathbf{I}\right]^{-1}\frac{\partial^2 \mathbf{R}_1}{\partial \delta^2} \\
&\quad - \left[\mathbf{R}_1+\sigma^2\mathbf{I}\right]^{-1}\frac{\partial^2 \mathbf{R}_1}{\partial \delta^2}\left[\mathbf{R}_1+\sigma^2\mathbf{I}\right]^{-1}\frac{\partial \mathbf{R}_1}{\partial \delta} + \left[\mathbf{R}_1+\sigma^2\mathbf{I}\right]^{-1}\frac{\partial^3 \mathbf{R}_1}{\partial \delta^3}
\end{aligned}
$$

Finally, since

$$
\left.\frac{\partial^i \mathbf{R}_1}{\partial \delta^i}\right|_{\delta=0} = 0 \quad i = 1,\ 3,
$$

we will have

$$J(\delta) = \frac{\delta^4}{8} \mathtt{t}r \left[\left(\left[\mathbf{R}_1 + \sigma^2 \mathbf{I} \right]^{-1} \left. \frac{\partial^2 \mathbf{R}_1}{\partial \delta^2} \right|_{\delta=0} \right)^2 \right]. \tag{G.13}$$

As we see from (G.13) and (G.13), the divergence for this case is also related to the second derivative of the Fisher information matrix evaluated at $\delta = 0$.

Appendix H

Derivation of ΔQ and ΔP

H.1 Derivation of ΔQ and ΔP)

Computing $\Delta\mathbf{Q}$ and $\Delta\mathbf{P}$ requires the following

$$
\begin{aligned}
\Delta E_0(h) &= f_s \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} h(x,y)\Delta h(x,y)\mathrm{d}x\mathrm{d}y \\
\Delta E_{10}(h) &= f_s \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{\partial h(x,y)}{\partial x}\frac{\partial \Delta h(x,y)}{\partial x}\mathrm{d}x\mathrm{d}y \\
&= f_s \int_{-\infty}^{+\infty}\left[\lim \frac{\partial h(x,y)}{\partial x}\Delta h(x,y)\Big|_{x\to-\infty}^{x\to+\infty} - \int_{-\infty}^{+\infty}\frac{\partial^2 h(x,y)}{\partial x^2}\Delta h(x,y)\mathrm{d}x\right]\mathrm{d}y \\
&= -f_s \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{\partial^2 h(x,y)}{\partial x^2}\Delta h(x,y)\mathrm{d}x\mathrm{d}y \\
\Delta E_{01}(h) &= f_s \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{\partial h(x,y)}{\partial y}\frac{\partial \Delta h(x,y)}{\partial y}\mathrm{d}x\mathrm{d}y \\
&= -f_s \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{\partial^2 h(x,y)}{\partial y^2}\Delta h(x,y)\mathrm{d}x\mathrm{d}y \\
\Delta E_{20}(h) &= f_s \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \frac{\partial^2 h(x,y)}{\partial x^2}\frac{\partial^2 \Delta h(x,y)}{\partial x^2}\mathrm{d}x\mathrm{d}y \\
&= f_s \int_{-\infty}^{+\infty}\left[\lim \frac{\partial^2 h(x,y)}{\partial x^2}\frac{\partial \Delta h(x,y)}{\partial x}\Big|_{x\to-\infty}^{x\to+\infty} - \int_{-\infty}^{+\infty}\frac{\partial^3 h(x,y)}{\partial x^3}\frac{\partial \Delta h(x,y)}{\partial x}\mathrm{d}x\right]\mathrm{d}y \\
&= f_s \int_{-\infty}^{+\infty}\left[-\lim \frac{\partial^3 h(x,y)}{\partial x^3}\Delta h(x,y)\Big|_{x\to-\infty}^{x\to+\infty} + \int_{-\infty}^{+\infty}\frac{\partial^4 h(x,y)}{\partial x^4}\Delta h(x,y)\mathrm{d}x\right]\mathrm{d}y
\end{aligned}
$$

$$
\begin{aligned}
&= f_s \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\partial^4 h(x,y)}{\partial x^4} \Delta h(x,y) \mathrm{d}x\mathrm{d}y \\
\Delta E_{02}(h) &= f_s \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\partial^2 h(x,y)}{\partial y^2} \frac{\partial^2 \Delta h(x,y)}{\partial y^2} \mathrm{d}x\mathrm{d}y \\
&= f_s \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\partial^4 h(x,y)}{\partial y^4} \Delta h(x,y) \mathrm{d}x\mathrm{d}y \\
\Delta E_{11}(h) &= f_s \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\partial^2 h(x,y)}{\partial x \partial y} \frac{\partial^2 \Delta h(x,y)}{\partial x \partial y} \mathrm{d}x\mathrm{d}y \\
&= f_s \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \frac{\partial^4 h(x,y)}{\partial x^2 \partial y^2} \Delta h(x,y) \mathrm{d}x\mathrm{d}y
\end{aligned}
$$

where we assume that

$$
\lim h(x,y) = \lim \frac{\partial^{i+j} h(x,y)}{\partial x^i \partial y^j} = 0
$$

as x or $y \to \pm\infty$. Therefore

$$
\Delta \mathbf{P} = \begin{bmatrix}
\Delta E_0 & 0 & 0 & -\Delta E_{10} & -\Delta E_{01} & 0 \\
0 & \Delta E_{10} & 0 & 0 & 0 & 0 \\
0 & 0 & \Delta E_{01} & 0 & 0 & 0 \\
-\Delta E_{10} & 0 & 0 & \Delta E_{20} & \Delta E_{11} & 0 \\
-\Delta E_{01} & 0 & 0 & \Delta E_{11} & \Delta E_{02} & 0 \\
0 & 0 & 0 & 0 & 0 & \Delta E_{11}
\end{bmatrix} \tag{H.1}
$$

and

$$\Delta\mathbf{Q} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \Delta E_{10} & 0 & 0 & 0 & 0 \\ 0 & 0 & \Delta E_{01} & 0 & 0 & 0 \\ 0 & 0 & 0 & \Delta\left(E_{20} - \dfrac{E_{10}^2}{E_0}\right) & \Delta\left(E_{11} - \dfrac{E_{10}E_{01}}{E_0}\right) & 0 \\ 0 & 0 & 0 & \Delta\left(E_{11} - \dfrac{E_{10}E_{01}}{E_0}\right) & \Delta\left(E_{02} - \dfrac{E_{01}^2}{E_0}\right) & 0 \\ 0 & 0 & 0 & 0 & 0 & \Delta E_{11} \end{bmatrix} \quad \text{(H.2)}$$

where

$$\begin{aligned}
\Delta\left(E_{20} - \frac{E_{10}^2}{E_0}\right) &= \Delta E_{20} - 2\frac{E_{10}}{E_0}\Delta E_{10} + \frac{E_{10}^2}{E_0^2}\Delta E_0 \\
&= f_s \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[h_{40}(x,y) - 2\frac{E_{10}}{E_0}h_{20}(x,y) + \frac{E_{10}^2}{E_0^2}h(x,y)\right]\Delta h(x,y)\mathrm{d}x\mathrm{d}y \\
\Delta\left(E_{11} - \frac{E_{10}E_{01}}{E_0}\right) &= \Delta E_{11} - \frac{E_{01}}{E_0}\Delta E_{10} - \frac{E_{10}}{E_0}\Delta E_{01} + \frac{E_{10}E_{01}}{E_0^2}\Delta E_0 \\
&= f_s \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[h_{22}(x,y) - \frac{E_{01}}{E_0}h_{20}(x,y) - \frac{E_{10}}{E_0}h_{02}(x,y) + \frac{E_{10}E_{01}}{E_0^2}h(x,y)\right] \\
&\qquad \times\Delta h(x,y)\mathrm{d}x\mathrm{d}y \\
\Delta\left(E_{02} - \frac{E_{01}^2}{E_0}\right) &= \Delta E_{02} - 2\frac{E_{01}}{E_0}\Delta E_{01} + \frac{E_{01}^2}{E_0^2}\Delta E_0 \\
&= f_s \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left[h_{04}(x,y) - 2\frac{E_{01}}{E_0}h_{02}(x,y) + \frac{E_{01}^2}{E_0^2}h(x,y)\right]\Delta h(x,y)\mathrm{d}x\mathrm{d}y
\end{aligned}$$

Appendix I

Computing the Energy Terms

I.1 Computing the Energy Terms

In this appendix, we explain the general process for the approximate computation of the energy terms. We will utilize the following identities for the calculation:

$$\sum_{k=0}^{L} x^k = \frac{1-x^{L+1}}{1-x} \tag{I.1}$$

$$\sum_{k=0}^{L} k^p x^k = \sum_{m=1}^{p} x^m \frac{\partial^m}{\partial x^m}\left(\frac{1-x^{L+1}}{1-x}\right) \tag{I.2}$$

$$\sum_{k} k^{p+1} \sin(xk)\cos(xk) = \frac{1}{2}\frac{\partial}{\partial x}\sum_{k} k^p \sin^2(xk) \tag{I.3}$$

Instead of showing all the calculations, for the sake of brevity we discuss, as an example, the calculation of the term $\mathbf{h}_0^T\mathbf{h}_0$:

$$\begin{aligned}\mathbf{h}_0^T\mathbf{h}_0 &= 4\sum_{k=-(N-1)/2}^{(N-1)/2} \sin^2\left(\frac{2\pi f_c}{f_s}k\right) \\ &= \sum_{k=-(N-1)/2}^{(N-1)/2} -\left[\exp\left(j\frac{2\pi f_c}{f_s}k\right) - \exp\left(-j\frac{2\pi f_c}{f_s}k\right)\right]^2\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k=-(N-1)/2}^{(N-1)/2} 2 - \exp\left(j\frac{4\pi f_c}{f_s}k\right) - \exp\left(-j\frac{4\pi f_c}{f_s}k\right) \\
&= 2N - 2\frac{1-\exp\left(j\frac{2\pi f_c}{f_s}(N+1)\right)}{1-\exp\left(j\frac{4\pi f_c}{f_s}\right)} - 2\frac{1-\exp\left(-j\frac{2\pi f_c}{f_s}(N+1)\right)}{1-\exp\left(-j\frac{4\pi f_c}{f_s}\right)} + 2 \\
&= 2N + 2 - 2\frac{1-\cos\left(\frac{4\pi f_c}{f_s}\right)+\cos\left(\frac{2\pi f_c}{f_s}(N-1)\right)-\cos\left(\frac{2\pi f_c}{f_s}(N+1)\right)}{1-\cos\left(\frac{4\pi f_c}{f_s}\right)} \\
&= 2N - \underbrace{2\frac{\sin\left(\frac{2\pi f_c}{f_s}N\right)}{\sin\left(\frac{2\pi f_c}{f_s}\right)}}_{C}
\end{aligned}
\tag{I.4}
$$

Since $\sin(x) \geq 1 - \left|\frac{2}{\pi}x - 1\right|$ for $0 \leq x \leq \pi$, and $\frac{2\pi f_c}{f_s} < \pi$, by upper and lower bounding the numerator and the denominator of $|C|$, respectively, we have

$$
|C| = 2\left|\frac{\sin\left(\frac{2\pi f_c}{f_s}N\right)}{\sin\left(\frac{2\pi f_c}{f_s}\right)}\right| \leq \frac{2}{\sin\left(\frac{2\pi f_c}{f_s}\right)} \leq \frac{2}{1-\left|\frac{4f_c}{f_s}-1\right|} \tag{I.5}
$$

Thus for the range of $\epsilon \leq \frac{2f_c}{f_s} \leq 1-\epsilon$ (representing the range of f_s from just above the Nyquist rate ($2f_c$) to $1/\epsilon$ times the Nyquist rate), we will have $|C| < 1/\epsilon$ and therefore for $\epsilon < \frac{5}{N}$

$$
\mathbf{h}_0^T \mathbf{h}_0 \approx 2N \tag{I.6}
$$

A similar approach can be followed to compute other energy terms.

Bibliography

[1] A. J. Barabell, *Improving the resolution performance of eigenstructurebased direction-finding algorithms*, in Proc. ICASSP, Boston, MA, 1983, pp. 336–339.

[2] 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 **77** (1999), 37–48.

[3] A. Bovik, *Handbook of Image and Video Processing*, Academic Press, 2000.

[4] J. A. Cadzow, *Direction finding: a signal subspace approach*, IEEE Trans Systems, Man and Cybernetics **21** (Sept.-Oct. 1991), no. 5, 1115–1124.

[5] P. Charge, Yide Wang, and J. Saillard, *An extended cyclic MUSIC algorithm*, IEEE Transactions on Signal Processing **51** (July 2003), no. 7, 1605 – 1613.

[6] S.S. Chen, D.L. Donoho, and M.A. Saunders, *Atomic decomposition by basis pursuit*, SIAM Review **43** (2001), no. 1, 129159.

[7] B. J. Davis, W. C. Karl, A. K. Swan, M. S. Unlu, , and B. B. Goldberg, *Capabilities and limitations of pupil-plane filters for superresolution and image enhancement*, Opt. Express **12** (2004), 4150–4156.

[8] A. J. den Dekker, *Model-based optical resolution.*, IEEE Trans. Instr. Measurement **46** (1997), 798–802.

[9] A. J. den Dekker and A. van den Bos, *Resolution, a survey.*, J. Opt. Soc. Amer. **14** (1997), 547–557.

[10] D.L. Donoho, *Superresolution via sparsity constraints*, SIAM J. Math. Anal., **23** (2001), 1309–1331.

[11] D.L. Donoho, M. Elad, , and V.N. Temlyakov, *Stable recovery of sparse overcomplete representations in the presence of noise*, submitted to the IEEE Trans. On Information Theory on February 2004. (2004).

[12] D.L. Donoho and M. Elad, *Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ_1 minimization*, Proc. Nat. Aca. Sci. **100** (2003), 2197–2202.

[13] D.L. Donoho and X. Huo, *Uncertainty principles and ideal atomic decomposition*, IEEE Trans. on Inf. Theory **47** (2001), no. 7, 28452862.

[14] M. Elad and A. Feuer, *Restoration of single super-resolution image from several blurred, noisy and down-sampled measured images.*, IEEE Trans. on Image Processing **6** (Dec. 1997), no. 12, 1646–1658.

[15] C. L. Fales and F. O. Huck, *An information theory of image gathering*, Inf. Sci. **57-58** (1991), 245–285.

[16] C. L. Fales, F. O. Huck, R. Alter-Gartenberg, and Z. Rahman, *Image gathering and digital restoration*, Philos. Trans. Roy. Soc. London a **354** (1996), 2249–2287.

[17] C. L. Fales, F. O. Huck, J. A. McCormick, and S. K. Park, *Wiener restoration of sampled image data: End-to-end analysis*, J. Opt. Soc. Am. A5 (1988), 300–315.

[18] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, *Fast and robust multi-frame super-resolution*, IEEE Transactions on Image Processing **13** (October 2004), no. 10, 1327–1344.

[19] J. D. Gaskill, *Linear Systems, Fourier Transforms, and Optics*, John Wiley and Sons, Inc, 1978.

[20] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*, Dover Publications, Inc., 2000.

[21] R. Gonzales and R. Woods, *Digital Image Processing*, Prentice-Hall International, 2003.

[22] J. W. Goodman, *Introduction to Fourier Optics.*, New York: McGraw-Hill, 1996.

[23] H. V. Hamme, *A stochastical limit to the resolution of least squares estimation of the frequencies of a double complex sinusoid*, IEEE Transactions on Signal Processing **39** (Dec. 1991), no. 12, 2652 –2658.

[24] C. W. Helstrom, *The detection and resolution of optical signals.*, IEEE Trans. Inf. Theory **IT-10** (1964), 275–287.

[25] ———, *Detection and resolution of incoherent objects by a background-limited optical system.*, J. Opt. Soc. Am. **59** (1969), 164–175.

[26] _____, *Resolvability of objects from the standpoint of statistical parameter estimation.*, J. Opt. Soc. Am. **60** (1970), 659–666.

[27] F. O. Huck, C. L. Fales, R. Alter-Gartenberg, S. K. Park, and Z. Rahman, *An information-theoretic assessment of sampled imaging systems*, Opt. Eng. **25** (1999), 742–762.

[28] D. C. Shiue J. A. Cadzow, Y. S. Kim, *General direction-of-arrival estimation: a signal subspace approach*, IEEE Trans. Aerospace and Electronic Systems **25** (Jan. 1989), no. 1, pp. 31–47.

[29] Don H. Johnson and Dan E. Dugeon, *Array signal processing: Concepts and techniques*, Prentice Hall, 1993.

[30] S. M. Kay, *Spectrum analysis: A modern perspective.*, Proc. IEEE **69** (1981), no. 11, 1380–1418.

[31] _____, *Modern spectral estimation: Theory and application.*, Prentice-Hall, 1988.

[32] _____, *Fundamentals of statistical signal processing, estimation theory*, Prentice-Hall, 1993.

[33] _____, *Fundamentals of statistical signal processing, detection theory*, Prentice-Hall, 1998.

[34] A. Salim Kayhan and Moeness G. Amin, *Spatial evolutionary spectrum for* DOA *estimation and blind signal separation*, IEEE Transactions on Signal Processing **48** (March 2000), no. 3, 791–798.

[35] E. L. Kosarev, *Shannon's superresolution limit for signal recovery.*, Inverse Problem **6** (1990), 55–76.

[36] S. Kullback, *Information theory and statistics*, Dover Publications, 1968.

[37] E. G. Larsson and P. Stoica, *High-resolution direction finding: the missing data case*, IEEE Transactions on Signal Processing **49** (May 2001), no. 5, 950 – 958.

[38] H. B. Lee and M. S. Wegrovitz, *Resolution threshold of beamspace MUSIC for two closely spaced emitters*, IEEE Trans. on Acoustic, Speech, Signal Processing **38** (Sep. 1990), 1545–1559.

[39] L. B. Lucy, *Resolution limits for deconvolved images.*, Astron. J. **104** (1992), 1260–1265.

[40] _____, *Statistical limits to super-resolution*, Astron. Astrophys. **261** (1992), 706–710.

[41] A. J. Barabell M. Kaveh, *The statistical performance of the MUSIC and the minimum-norm algorithms in resolving plane waves in noise*, IEEE Trans. Acoust., Speech and Signal Processing **ASSP-34** (April 1986), 331–341.

[42] A. Macovski, *Mediacl Imaging Systems*, Prentice-Hall Inc, 1983.

[43] M. L. McCloud and L. L. Scharf, *A new subspace identification algorithm for high-resolution* DOA *estimation*, IEEE Transactions on Antennas and Propagation **50** (1382 - 1390), no. 10, Oct. 2002.

[44] L.T. McWhorter and L.L. Scharf, *Cramer-Rao bounds for deterministic modal analysis*, IEEE Trans. Signal Proc **41** (May 1993), 1847–1862.

[45] P. Milanfar and A. Shakouri, *A statistical analysis of diffraction-limited imaging*, Proceeding of the International Conference on Image Processing. (September 2002, Rochester, NY), 864–867.

[46] N. Nguyen, P. Milanfar, and G.H. Golub, *A computationally efficient image superresolution algorithm.*, IEEE Trans. on Image Processing **10** (Apr. 2001), no. 4, 573–583.

[47] A. V. Oppenheim and R. W. Schafer, *Discrete-time Signal Processing*, Prentice-Hall, Inc, 1989.

[48] J.A. O'Sullivan, R.E. Blahut, and D.L. Snyder, *Information-theoretic image formation*, IEEE Transactions on Information Theory **44** (1998), no. 6, 2094 –2123.

[49] W. K. Pratt, *Digital Image Processing*, John-Wiley and Sons, Inc, 2001.

[50] L. L. Scharf, *Statistical signal processing, detection, estimation, and time series analysis*, Addison-Wesley, 1991.

[51] L. L. Scharf and B. Friedlander, *Matched subspace detectors*, IEEE Trans. on Signal Processing **42** (August 1994), 2146–2157.

[52] L. L. Scharf and P. H. Moose, *Information measures and performance bounds for array processors*, IEEE Trans. on Information Theory **IT-22** (Jan 1976), no. 1, 11–21.

[53] R. O. Schmidt, *Multiple emitter location and signal parameter estimation*, IEEE Trans. Antennas Prop. **AP-34** (1986), 276–280.

[54] M. Shahram and P. Milanfar, *A statistical analysis of achievable resolution in incoherent imaging.*, Proceedings of the SPIE Annual Meeting, Signal and Data Processing of Small Tergets 2003, San Diego, CA **5204** (August 2003), 1–9.

[55] S.T. Smith, *Accuracy and resolution bounds for adaptive sensor array processing.*, 9th IEEE Signal Proc. Workshop Statist. Signal Array Proc. (Portland, OR) (1998), 3740.

[56] ———, *Statistical resolution limits and the complexified Cramer-Rao bound.*, to appear in IEEE Transactions on Signal Processing (2005).

[57] D.L. Snyder, A.O. Hero, P. Moulin, J.M.F Moura, and J.A. O'Sullivan, *Introduction to the special issue on information-theoretic imaging*, IEEE Transactions on Information Theory **46** (Aug 2000), no. 5, 1709 –1713.

[58] P. Stoica and A. Nehorai, *MUSIC, maximum likelihood, and cramr-Rao bound: Further results and comparisons*, IEEE Trans. Acoust., Speech, Signal Processing **38** (Dec. 1990), 2140–2150.

[59] ———, *MUSIC, maximum likelihood, and cramr-Rao bound*, IEEE Trans. Acoust., Speech Signal Processing **37** (May 1989), 720–741.

[60] G. Strang, *Introduction to Applied Mathematics*, Wellesley-Cambridge Press, 1986.

[61] V.N. Temlyakov, *Greedy algorithms and m-term approximation*, J. Approx. Theory **98** (1999), 117145.

[62] J.A. Tropp, *Greed is good: algorithmic results for sparse approximation*, IEEE Trans IT **50** (2004), 22312242.

[63] R. Tyson, *Adaptive Optics Engineering Handbook*, Marcel-Dekker Inc, 2000.

[64] P. P. Vaidyanathan, *Generalizations of the sampling theorem: Seven decades after nyquist.*, IEEE Trans. Circuits and Systems - 1 **48(9)** (Sep. 2001), 1094–1109.

[65] A. van den Bos, *Resolution in model-based measurements*, IEEE Trans. Instr. Measurement **51** (2002), 1055–1060.

[66] M. Vetterli, P. Marziliano, and T. Blu, *Sampling signals with finite rate of innovation*, IEEE Transactions on Signal Processing **50** (June 2002), 1417 –1428.

[67] L.S. Biradar V.U. Reddy, *Svd-based information theoretic criteria for detection of the number of damped/undamped sinusoids and their performance analysis*, IEEE Transactions on Signal Processing **41** (Sept. 1993), no. 9, 2872 –2881.

[68] A. Weiss and B. Friedlander, *Effects of modeling error on the resolution threshold of the MUSIC algorithm*, IEEE Transactions on Signal Processing **42** (1994), no. 1, 147–155.

[69] B. Wohlberg, *Noise sensitivity of sparse signal representations: Reconstruction error bounds for the inverse problem*, IEEE Trans. Signal Processing **51** (2003), no. 12, 3053–3060.

[70] C. Ray Wylie, *Advanced Engineering Mathematics*, Mc-Graw Hill, 1975.

[71] J. M. Zuo, *Electron detection characteristics of a slow-scan CCD camera, imaging plates and film, and electron image restoration*, MICROSCOPY RESEARCH AND TECHNIQUE **49** (2000), 245–268.