University of New Mexico
UNM Digital Repository

Optical Science and Engineering ETDs | Engineering ETDs

9-12-2014

Bayesian Estimators, Error Bounds, and Applications to Imaging

Srikanth Narravula

Follow this and additional works at: https://digitalrepository.unm.edu/ose_etds

Recommended Citation
Narravula, Srikanth. "Bayesian Estimators, Error Bounds, and Applications to Imaging." (2014). https://digitalrepository.unm.edu/ose_etds/29

This Dissertation is brought to you for free and open access by the Engineering ETDs at UNM Digital Repository. It has been accepted for inclusion in Optical Science and Engineering ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.

Candidate

Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

, Chairperson

Bayesian Estimators, Error Bounds, and Applications to Imaging

by

Srikanth Reddy Narravula

B. Tech., Indian Institute of Technology, Madras, 2007
M.S., ECE, University of New Mexico, 2009

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy
Optical Science and Engineering

The University of New Mexico

Albuquerque, New Mexico

July, 2014

©2014, Srikanth Reddy Narravula

Dedication

To That which is not.

Acknowledgments

I would like to thank my advisor, Professor Sudhakar Prasad, for his guidance, patience and support. There are many reasons which make me believe that it is providence which brought me to him. Science was fun when I was a child, but somehow along the road, it was lost. It was only after I worked with Prof. Prasad, just solving problems became exciting again with renewed energy and focus. His insights into various areas of research from quantum physics to statistical estimation theory and anything else that is related to science helped me in broadening my own view on approaching research. Somehow everything is connected and I think I developed an eye to see things a little deeper. I am grateful to him for giving me the freedom to express and pursue my ideas. It has been an enriching experience working with him.

My gratitude goes to Prof. Majeed M. Hayat. In the time I worked with him I learned about being professional and about perseverance which are invaluable assets in an increasingly competitive academic environment. I am grateful to Prof. Keith Lidke and Prof. James Thomas for serving in my committee. I am thankful to my group members Rakesh, Julian, Henry, and Zhixian for their support.

My gratitude goes to my mother, Radha, father, Ramalinga Reddy, and sister, Manogna for their constant support and love through the years. I am thankful to all my friends for making my stay in Albuquerque a memorable one.

Bayesian Estimators, Error Bounds, and Applications to Imaging

by

Srikanth Reddy Narravula

B. Tech., Indian Institute of Technology, Madras, 2007
M.S., ECE, University of New Mexico, 2009

Ph.D., Optical Science and Engineering, University of New Mexico, 2014

Abstract

In a communication system, a signal carrying information about a physical variable, or parameter, is fed into the front end of a channel and noisy or corrupted data are obtained at its back end, data from which one attempts to estimate the physical parameter. The error that accompanies such an estimate is usually characterized by mean squared error (MSE). In a Bayesian setting, the minimum value of the MSE is obtained by the so-called MMSE estimator (MMSEE), which in this sense is the best possible estimator. In general, the MMSE as well as the MMSEE are difficult to compute, which gave early impetus to work on a host of bounds on the MMSE of varying degrees of tightness over the last fifty years. The bounds, if sufficiently tight, help us in evaluating the performance of sub-optimal estimators.

A widely used lower bound on the MMSE is the Ziv-Zakai lower bound, which bounds the MMSE via the minimum probability of error (MPE) of a binary hypothesis testing problem. Extensions of the Ziv-Zakai lower bound and some compu-

tationally efficient approximations are derived. The extensions include a bound in terms of the MPE of an M-ary hypothesis testing problem and another bound for discrete prior probability distributions.

A major focus of the dissertation has been on deriving tight upper bounds on the MMSE, and their applications to imaging and non-imaging problems. An upper bound on the MMSE is derived that has a variational character, is easy to compute, and follows the MMSE tightly. This new upper bound on the MMSE is shown to be the "bias removed mean squared error" of *any* test estimator. By choosing test estimators from suitably parameterized families and then optimizing over the parameters of the families, we obtain tight upper bounds on the MMSE as well as the estimators that achieve the upper bound. A new piecewise quasi-linear estimator which works well particularly when the data depend non-linearly on the parameters being estimated is proposed. The MMSE upper bound for such estimators performs tightly in bounding the MMSE for all values of the signal to noise ratio (SNR).

The upper bounds obtained from the test estimators are applied to illustrative, toy 1-pixel and 2-pixel models of the EM-CCD camera. Further illustrations are provided by applying the upper bound to a highly non-linear and more realistic time-delay estimation (TDE) problem which is basic to many signal processing scenarios, including source detection, array processing, surveillance, synchronization in communications, range finding in RADAR, geo-location and tracking in sensor networks. A comparison with the existing bounds on the MMSE shows that our upper bound performs optimally in terms of its tightness to the MMSE and is the best of all bounds that we considered in all regions of operation. The upper bound was finally applied to characterize the performance of a newly developed rotating point spread function imaging system, which is capable of snapshot 3D imaging, and compared with the performance of a conventional imaging system for localizing point sources beyond the diffraction limit. New asymptotic analyses of the MMSE are also stud-

ied, both in the low SNR and the high SNR regimes, and a useful interpolational approximation of the MMSE is presented which seems to approximate accurately the MMSE for all values of the SNR.

Contents

List of Figures **xiii**

Glossary **xvi**

1 Introduction **1**

2 Review **11**

2.1 The estimation problem 11

2.2 Minimum mean square error (MMSE) 14

2.3 MLE and the maximum *a posteriori* (MAP) estimator 16

2.4 Ziv Zakai bound (ZZB) 18

2.5 Weiss Weinstein bound 22

2.6 Illustrative problems 24

2.6.1 Poisson channel with exponential prior 24

2.6.2 Time-delay estimation (TDE) 28

2.7 Summary 36

3 Extensions of the ZZB **39**

3.1 Approximate ZZB 39

3.2 Beyond the ZZB 41

3.2.1 Relation to the standard ZZB 45

3.3 Discrete ZZB 45

3.4 Summary 50

4 Upper Bound on the MMSE **52**

4.1 Derivation of the upper bound 52

4.1.1 Understanding bias 54

4.2 Multi-parameter generalization of the MMSE and upper bound 57

4.3 Test estimators 59

4.3.1 The polynomial class of estimators 59

4.3.2 The MAP estimator 61

4.4 Results 62

4.4.1 Poisson channel and exponential prior - single pixel case . . . 62

4.4.2 Poisson channel and exponential prior - two pixel case 64

4.4.3 TDE problem 66

4.5 Piecewise quasi-linear estimator and associated upper bound 67

4.5.1 Application of the PQE approach to the TDE problem 74

4.6 Performance analysis of rotating point spread function (RPSF) based imaging system 75

4.7 Summary 81

5 Asymptotic Analyses of the MMSE **83**

5.1 The low SNR regime 84

5.1.1 A second order logarithmic expansion of the data PD 85

5.1.2 The MMSE estimator and the MMSE to second order 86

5.2 The high SNR regime 89

5.2.1 Higher order corrections 90

5.3 TDE problem 95

5.3.1 The Gaussian channel 95

5.3.2 Pseudo-Gaussian channel 98

5.4 An approximate MMSE 100

5.5 Summary 103

6 Conclusions and Future Directions **104**

Appendices **110**

A Poisson channel and exponential prior **111**

A.1 $P(y)$ for the single-pixel problem 111

A.2 Two-pixel problem 112

B The various expectations in the optimal coefficients of the PQE 119

C MMSE approximation in the high-SNR regime 124

D References 127

References 128

List of Figures

1.1 Typical behavior of the MMSE and the different regions of errors. . 3

2.1 Valley filling function, reproduced from [1] 19

2.2 The MMSE, the EZZB and the WWB for the Poisson channel, with bias, b = 10 and exponential prior with mean $\langle X \rangle = 2$. 27

2.3 Log-log plot for the MMSE, the MAP error, the EZZB and the WWB for the Poisson channel, with bias, b = 10 and exponential prior with mean $\langle X \rangle = 2$. 29

2.4 Figure reproduced from [2] representing acoustic source and sensors. The source need not be acoustic, in general. 30

2.5 Figure reproduced from [2] representing acoustic source and sensors. The source need not be acoustic, in general. 31

2.6 Figure reproduced from [3] representing acoustic source and sensors. The source need not be acoustic, in general. 31

2.7 Figure reproduced from [4] representing a planar model of two sensors separated by a distance d. 32

2.8 MMSE, MAP, EZZB and WWB for the passive TDE problem. The MMSE and the MAP are simulated using 1000 noise frames and are plotted with corresponding error bars. 34

3.1 Exact and approximate EZZB for a Poisson channel with exponential prior described in Sec. 2.6.1. The approximate EZZB is generated using $n = 2$ and $n = 5$ in Eq. (3.4). 41

3.2 Construction of the intervals in the derivation of the M-ary EZZB. . 42

3.3 Relation between the continuous variable h and the discrete variable x. 47

4.1 MMSE, MAP-VB, P^3-VB, EZZB and WWB for the Poisson channel, with bias, b = 10 and exponential prior with mean $\langle X \rangle = 2$. 63

4.2 Two-pixel detector being used to resolve the two near-by point sources. 65

4.3 UB with a quadratic estimator. $\alpha + \beta = 1$, $\gamma = 10$, $\langle X_1 \rangle = 2$, $\langle X_2 \rangle = 3$. 66

4.4 UB with quadratic estimator. $\alpha + \beta = 0.4$, $\gamma = 10$, $\langle X_1 \rangle = 2$, $\langle X_2 \rangle = 3$. 66

4.5 MMSE, MAP-VB, linear MMSE, EZZB and WWB for the TDE problem. 68

4.6 Non-linear dependency (dashed curve) of the signal on the prior parameters and a piecewise linear approximation (solid line segments). 69

4.7 Column-wise extraction of elements from a matrix. 71

4.8 MMSE, MAP-VB, PQE, and EZZB for the TDE problem. 74

4.9 Surface plots of the OAR-PSF and the ideal diffraction limited PSF reproduced from [5]. The plots from left to right are for increasing values of defocus, namely -24, -16, -8, 0, 8, 16, and 24 radians at the pupil edge. At a defocus of 16 radians, the conventional PSF, as shown in (m), broadens wide enough to be deemed unfit for imaging. On the other hand, the OAR-PSF simply rotates with defocus and remains compact even at a defocus of 24 radians, as shown in (g). . 77

4.10 MAP-VB for the source super-localization problem with an $M = 16$ fold enhancement. The conventional imager does exceptionally well at best focus, but performs very poorly at large defocus. The OAR-PSF, however remains robust as the defocus changes over a large range. 79

5.1 MMSE, and its approximations for the TDE problem with Gaussian conditional data statistics. 97

5.2 MMSE, and its approximations for the TDE problem with pseudo-Gaussian conditional data statistics. 100

5.3 MMSE and its approximations for the TDE problem with Gaussian conditional data statistics. 101

5.4 MMSE and its approximations for the TDE problem with pseudo-Gaussian conditional data statistics. 102

A.1 Region of support for (ϕ_1, ϕ_2) is the shaded region between the lines $\alpha\phi_1 = \beta\phi_2$ and $\alpha\phi_2 = \beta\phi_1$. 115

Glossary

X	Parameters to be estimated
$\hat{X}$	Estimator
x	Real number, value of the random variable X
$\bar{x}$	Vector of real numbers
$\bar{X}$	Random vector
$\mathbf{A}$	Matrix
Y, $\bar{Y}$, $y(t)$	Statistical data
δX	$X - \mathbb{E}[X]$
$\delta\hat{X}$	$\hat{X} - \mathbb{E}[\hat{X}]$
Δt	Time spacing
$n(t)$	Additive white Gaussian noise
$s(t)$	Signal pulse
Pr	Probability
PD	Probability distribution

Glossary

P_X	Prior probability distribution
P_Y	Probability distribution of data
$P_{X\|Y}$	Posterior probability distribution
$P_{X\|Y}$	Conditional probability distribution
ϵ	Error of estimator
γ, ϵ, η, ξ	Higher order corrections
$\mathbb{E}, \langle . \rangle$	Expectation
σ^2	Variance
G	Valley filling function
RHS	Right hand side
MSE, $\mathcal{E}$	Mean squared error
MMSE, $\mathcal{E}_M$	Minimum mean squared error
MMSEE, $\hat{X}_M$	Minimum mean squared error estimator
MPE, $P_{\min}$	Minimum probability of error
MAP	Maximum *a posteriori*
MLE	Maximum likelihood estimator
CRB	Cramer-Rao bound
UCRB	Uniform Cramer-Rao bound
BB	Barankin bound

ZZB	Ziv-Zakai bound
EZZB	Extended Ziv-Zakai bound
AZZB	Approximate Ziv-Zakai bound
IZZB	Improved Ziv-Zakai bound
DZZB	Discrete Ziv-Zakai bound
MCMC	Markov chain Monte Carlo
WWB	Weiss-Weinstein bound
TDE	Time delay estimation
SNR	Signal-to-noise ratio
FNR	Flux-to-noise ratio
UB	Upper bound
MAP-VB	MAP variational bound
P^N-VB	Variational bound for the N^{th} degree polynomial estimator
PQE	Piecewise quasi-linear estimator
PQE-VB	PQE variational bound
PSF	Point spread function
RPSF	Rotating PSF
LSNR-MMSE	Low-SNR approximation to MMSE
HSNR-MMSE	High-SNR approximation to MMSE
FI, J	Fisher information

Chapter 1

Introduction

Error is a natural outcome in any signal or parameter estimation problem. Generally, the parameter value is known *a priori* to be bounded, which can sometimes be described by a statistical distribution called the prior. Measurements of the parameter, themselves having a statistical character, can provide more information about the prior through inference. One of the ways to statistically infer the value of the parameter from the observations is by using Bayesian estimation which relies on the posterior distribution. The posterior distribution improves upon the information about the parameter contained in the prior, since the observations *a posteriori* restrict the possible values of the parameter.

Two widely used inference measures are the mean and the mode of the posterior distribution. They are, respectively, the minimum mean squared error estimator (MMSEE) and the maximum *a posteriori* (MAP) estimator. The MMSEE is the optimal estimator in the mean squared error (MSE) sense, and its error is the minimum mean squared error (MMSE), which is the lower bound on MSE achieved by *any* estimator. The MMSEE is, in general, notoriously difficult to compute numerically, as is the MMSE, which is an important reason for a host of bounds on the MMSE of

varying degrees of tightness having been introduced and studied over the last fifty years. The bounds, if sufficiently tight, help us in evaluating the performance of sub-optimal estimators. This dissertation is concerned with the derivation of new bounds on the MMSE and with a detailed comparative study of the new and existing bounds in the context of certain applications.

In a typical parameter estimation problem, there are three regions of operation, as shown in Fig. 1.1. In the asymptotic region, where the signal to noise ratio (SNR) is high or many observations are available, any prior information about a parameter is typically dwarfed by the information provided by the high quality of data present, and as a result the MSE can be made small. As the SNR decreases, the MSE rises rapidly at some point. Because of its steepness, this increase in error, also known as the threshold phenomenon, might mimic a phase transition of a thermodynamic system. When the SNR becomes even smaller or when only a few high-quality observations are available, observations provide little information, and the error is controlled largely by the prior knowledge, and the error moderates once again.

A powerful statistical analysis of estimation was developed by Fisher [6, 7] in the 1920's. This was divided into the theory of efficient statistics, namely estimators with no bias provided there are many measurements, and that of sufficient statistics, estimators that provide all the relevant information even when there are only a few measurements. The maximum likelihood estimator (MLE) was shown to be an efficient statistic, and also that it was normally distributed when there are large number of measurements. While this was a significant advance, it was only after the work done by Cramer [8], Rao [9, 10], Bhattacharya [11], Barankin [12] and Chapman-Robbins [13] in the late 1940's that the behavior of error was understood well, and the transition from the high SNR region to the low SNR region was better appreciated. More bounds on the MSE for deterministic parameter estimation, where the unknown parameter is assumed to be deterministic, were derived later [14–18]. They

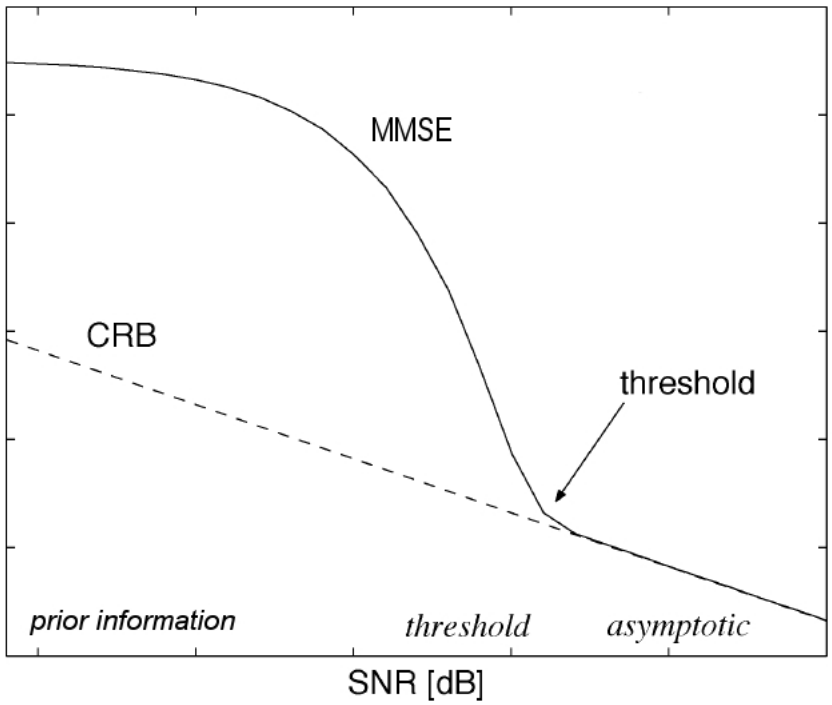

Figure 1.1: Typical behavior of the MMSE and the different regions of errors.

are all lower bounds on the variance of an unbiased estimator, and no prior knowledge about the distribution of the parameter values was taken into account. The simplest of these lower bounds is the Cramer-Rao bound (CRB), which, because it is easily computable, was used widely in the high SNR region. The probability density functions, however, need to obey certain regularity conditions for the calculation of the CRB. Moreover, it is a little too optimistic, underestimating the error, as we move into the "large-error" region [19] as shown in Fig. 1.1. The transition region is precisely this large error region and received a lot of interest [17, 18, 20, 21]. The more complete and computationally intense Barankin bounds (BB) seem to do well in locating the threshold. Rife et al. [22] and Athley [23] used the method of interval errors [21] to locate the threshold. All these lower bounds were shown to be special cases of the Barankin bound by proper selection of test functions [16, 24, 25].

Since most of the estimators tend to be biased in the low SNR limit, these lower bounds are not of much use for values of the SNR in the transition region or lower. Even though the CRB can be formulated for biased estimators [21], which we may call biased CRB, it is of less practical value as it depends on the gradient of the bias of the particular estimator with respect to the parameters under estimation. For performance comparison, a useful lower bound should apply to the entire class of estimators. Usually, when the parameter or the estimator is constrained to lie is a closed space, biased estimators are unavoidable [26]. Nevertheless, biased estimators have been in frequent use as they are observed to reduce the estimation error better than an unbiased estimator [27]. For example, biased estimators were applied widely in image restoration [28], where regularization is implemented to reduce noise amplification (variance) at the expense of reduced spatial resolution (bias). Biased ridge estimators [29] were used in multiple regression to reduce the variance of ordinary least squares estimators.

To compare the performance of biased estimators, Hero et al. [30] developed lower bounds for the MSE, using a multiple parameter Chapman Robbins bound, when the parameter was constrained to lie on a manifold of the parameter space. Later, a new bound called the uniform CRB (UCRB) [31] was developed by the same group, for a scalar function of a deterministic vector parameters. This is a bound on the smallest attainable error that can be achieved using any estimator with a bias gradient whose norm is bounded above by a constant. With the UCRB, the trade-off between the variance and bias of an estimator was analyzed in [32]. The UCRB was extended for vector parameters, and asymptotically optimal estimators that achieve the UCRB were developed in [33]. New lower bounds on the MSE of biased estimators were proposed: in [34] with an integral form of the Barankin bound for biased estimators and using the kernel of the weighted Fourier transform, and in [35] by restricting the estimators to a linear bias sub-class and using a minimax criterion to find an optimal bias vector so that the MSE is minimized uniformly for all the parameter values with

respect to the CRB. A tutorial on the relationship between bias and variance of an estimator can be found in [36].

Another way of approaching the problem of estimation is to look for bounds on the MMSE that are independent of the bias while being applicable under all operating conditions. These are the "Bayesian bounds", which assume the parameter to be a random variable with a known prior probability distribution, and were first developed by Ziv and Zakai [26] in the late 1960's. The Ziv-Zakai bound (ZZB) was initially formulated for a scalar uniformly distributed parameter. Since the ZZB was independent of the bias while accounting, albeit statistically, for any prior knowledge, unlike the previously existing bounds, it represented a significant advance. The ZZB was shown, however, to be asymptotically poor by Seidman [37], who improved the ZZB by introducing a parameter over which an optimization was performed. Chanaz et al. [38] improved the ZZB by avoiding the Chebychev's inequality. Bellini et al. [39] improved it further by using the Kotelnikov inequality [40] for the probability of outage error and by introducing a valley filling function. All of these different versions of the ZZBs relate the MMSE to the probability of error in a binary hypothesis testing problem, and they were further extended to arbitrary prior distributions and vector parameters [1, 41]. A different kind of extension of the ZZB was accomplished by relating the MMSE to an M-ary hypothesis testing problem: in [42] the number of hypothesis depended on the size of the parameter space, and in [41] the number of hypothesis was independent of the size of the parameter space and hence provided more control over the number of hypotheses. The ZZB was also derived for discrete priors in [42], even though the Bayes risk considered there is somewhat different from the traditional MSE.

Another important Bayesian lower bound, the Weiss Weinstein bound (WWB) [43], was developed based on the covariance inequality. The WWB is expressed in terms of a test function, defined on the parameter and observation space, which is or-

thogonal to any function of the observations. By choosing appropriate test functions, new Bayesian bounds can be formulated. In fact, using this approach the Bayesian CRB [21], the Bayesian Chapman Robbins bound, the Bayesian Bhattacharya bound and the Bobrovsky-Zakai bound [44] could all be unified under a single class [45]. In recent years, some new Bayesian bounds have been developed. Renaux et al. [46] and Todros et al. [47] formulated a general class of lower bounds and generated new bounds within this class by using different test functions belonging to the class. Another lower bound on the MMSE was presented in [48] which, while based on a deterministic, biased version of CRB, did not, unlike the Bayesian CRB, require any regularity conditions on the prior distribution . They used an optimal bias function to arrive at the lower bound. A compilation of all the major work done in the field of Bayesian bounds and their applications, till 2007, can be found in [49].

For the sake of completeness, we note the intrinsic relation between the Bayesian minimum probability of error (MPE), which is an all encompassing description of error, and the Bayesian MMSE which is a second order error moment. This is suggested by the very formulation of the ZZB. Since the MPE is difficult to compute, many suboptimal distance measures between statistical distributions were proposed and studied. notable among them are the Kullback-Leibler distance [50], the Bhattacharyya distance [51], the Bayesian distance [52] and Ali-Silvey [53] distance. These distance measures enable one to find bounds on the minimum probability of error [52, 54–56]. Recently, Routtenberg et al. [57] presented a general class of bounds on outage errors and, using a new test function, formed a new bound on the outage error, thereby developing new bounds on MMSE.

Studies of expected inverse relationships between Bayesian error and statistical information have also been carried out. Since mutual information [58] represents the amount of coded information that can be passed through a channel reliably for a given prior, and since error measures determine the accuracy of the recovered value

of the parameter (input signal) from the observations (channel output), one expects an inherent relation between mutual information and the various error measures. Inequalities exist between information and the probability of error. They include Fano's bound [59], and a number of other improved inequalities derived by Feder et al. [60], Raviv et al. [61], Seidler [62], Santhi et al. [63] and Golic [64], all based in essence on relations between the equivocation entropy and probability of error. Verdu et al. [65] discovered a relation between the mutual information and MMSE in terms of SNR for Gaussian channels and later extended it to Poisson channels [66]. An inequality between mutual information and the MMSE for any channel and prior was derived by Seidler [62] and more recently by Prasad [67], based on relating the MMSE to the equivocation.

As we see, the literature on statistical error is very rich, having developed largely over the past few decades, but only a few attempts have been made to find useful upper bounds on the MMSE. Seidman [68] derived an upper bound on MMSE by bounding the corresponding probabilities of the MSE. Hawkes et al. [69] developed an upper bound based on the Kullback-Leibler distance measure. A recent work by Flam et al. [70] used an optimal linear estimator to upper bound the MSE which performed well for the Gaussian-mixture model [71].

In this dissertation, we present the derivation and detailed applications of a new upper bound that has a variational character, is easy to compute, and follows the MMSE tightly. Because of these properties, we can use it for a wide range of parameter estimation problems. This new upper bound on the MMSE is shown to be the "bias removed mean squared error" of any test estimator. By means of a properly motivated choice of a class of test estimators and then optimizing over the parameters that define the class, we can obtain the tightest possible upper bound on MMSE as well as the corresponding estimators within this class that achieve the upper bound. Indeed, if good upper and lower bounds can be given for the MMSE

of a particular estimation problem, then it might be possible to answer more readily the question of their tightness in the asymptotic limit of many independently and identically distributed measurements.

The bounds developed in this dissertation are first applied to the time delay estimation (TDE) problem [4] which is fundamental to many scenarios, being intimately linked to detection, array processing, surveillance, synchronization in communications, range finding in RADAR, as well as geo-location and tracking in sensor networks. In the context of sensor networks and geolocation of mobile ad hoc network nodes, node localization, and tracking of position, the TDE is very important [72] where the time delay estimates between the nodes can be translated into position estimates. Localization techniques are also important in understanding biological processes, and with recent advances in fluorescence microscopy [73], localizing single molecules well beyond the diffraction limit is now possible. Seen from a spatial point of view, the TDE problem is formally the same as a one-dimensional version of the source-localization problem in single molecule imaging applications. Typically, the images are sub-divided into regions surrounding suspected fluorescent molecules, and then localization and tracking of molecules is performed region by region [74]. The usually adopted approaches for localization are the fast non-linear least squares fitting [75], which is a centroid based point spread function fitting, and the MLE [76], which is optimal whenever there are many measurements [77].

A quantitative study of the various localization techniques was done using the CRB by Abraham et al. [78], but an error analysis using the CRB is valid only in the asymptotic limit as we have noted before. An asymptotic treatment of sub-diffractive and sub-pixel localization is of importance, for example, to find the minimum source strength required to achieve an M-fold enhancement of source localization accuracy [79,80]. The asymptotic analysis of the minimum probability of error in point-source discrimination was done by treating it as a binary hypothesis testing problem [81,82],

and more generally as a multi-hypothesis testing problem [83–85]. However, for most single molecule experiments, often there are only a few measurements available, especially when the molecules are being tracked, and also the background noise is usually high [86]. Therefore, a Bayesian approach, which fully accounts for any prior information on the spatial bounds on the positions of the molecules and which is valid for all regions of operation, seems surely to be more desirable [86–88].

In its basic form,the TDE problem, consists of estimating the delay of the center of a pulse that has been corrupted in its propagation through a noisy environment [89]. A comparison and survey of the various approaches for estimating the delay can be found in [4, 90–93]. A survey of the performance bounds for the TDE problem can be found in [94]. Even though the ZZB and the WWB provide significantly tighter performance bounds than the family of CRB bounds, they also require more effort to evaluate. The choice of test points is important for the WWB (and the BB), a fact that can make evaluation of the bound significantly more complex as more parameters are added to the problem. The vector ZZB also becomes harder to evaluate due to the more involved calculation of the MPE, involving, in general, multiple integrals which are often difficult to compute. It may be possible to find simplified expressions for both the ZZB and the WWB in order to provide valuable insight. But it has been hard to determine how closely the simplified expressions approximate the MMSE. Moreover, the derivation of these bounds provides no clue to the optimal estimators that achieve them. In our examples, we have calculated the upper bound for different test estimators and compared it to the ZZB and the WWB. Our upper bound, as we shall see, performs optimally and is by far the best in all regions of operation.

The dissertation is organized as follows. In Chapter 2 we will review the MMSE, the MAP estimator and two important lower bounds on the MMSE, the ZZB and the WWB. We will also review the TDE problem and discuss another toy problem

that is used to compare the different bounds on the MMSE. A particular case where the ZZB performs poorly is presented. In Chapter 3, we will extend the ZZB using an M-ary hypothesis testing approach, and also to discrete priors. In Chapter 4, we will derive our new upper bound, and apply it to a polynomial class of estimators, and the MAP estimator. A new piecewise linear estimator that performs well in all the regions of operation will be the centerpiece of the chapter. In chapter 5, we construct asymptotic expressions for the MMSE estimator and the MMSE for high and low values of the SNR. In the high-SNR region, since the MLE estimator is known to achieve the CRB, we expand the MMSE estimator around the MLE in this region, while in the low-SNR region, where data add little information, the mean of the prior seems to be a suitable point about which to expand the MMSE estimator. Using this formalism a different kind of approximation to the MMSE is proposed. Finally, in Chapter 6, we present some ongoing work and future directions.

Chapter 2

Review

The problem of extracting signals and parameters of interest from a set of observations is inherent to many applications, a few of which include amplitude and frequency estimation in RADAR and sonar, direction of arrival of a wireless signal, estimating the positions of individual molecules in single molecule imaging, image restoration, image registration, sub-pixel resolution, and feature extraction in compressive sensing. In this chapter we will set up the problem of parameter estimation and review the MMSE, and the MLE and MAP estimators and some of their properties. We will also review the ZZB and the WWB, and point to some of their limitations.

2.1 The estimation problem

A note on the notation used in this dissertation: capital letters denote scalar random variables and small letters real numbers; capital letters with a bar denote random column vectors and small letters with a bar column vectors; and bold capital letters denote deterministic or random matrices. For example, x denotes a real number, X

a scalar random variable, $\bar{x}$ a column vector, $\bar{X}$ a random column vector, and $\mathbf{A}$ a deterministic matrix. $\mathcal{R}_X$ denotes the set of values that a random variable X takes. Any other notation will be defined at their first occurrence.

Let X be a parameter to be estimated and Y a statistical data variable from which X is being estimated. Let the probability distribution (PD) of X be denoted by P_X, that of Y by P_Y, and that of the conditional, namely of Y, given X, by $P_{Y|X}$. We may also call the latter the PD of the channel, borrowing from the terminology of modern communication theory. However, we shall drop subscripts when the meaning of the associated symbol is clear from its argument. We take these PDs to be densities, thus assuming nominally that X and Y take on continuous real values, but we may easily modify our derivation to apply to either or both of the variables being discrete by treating probability integrals as appropriate discrete sums over probability mass functions (PMFs). The parameter X and the data Y can be a scalar or a vector, but their treatment is a little different in each case. Though the upper bound, the MMSE and all other bounds can be expressed similarly for both cases, we will focus on the simpler case of scalar parameters first. The dimensionality of the data space is arbitrary in this treatment, with a simple reinterpretation of the data integrals as being multi-dimensional.

We note that there are two ways to approach a parameter estimation problem, called the classical and Bayesian approaches. In the classical approach parameter to be estimated is a deterministic quantity, x, and no prior PD is assumed. The estimate $\hat{X}$ of x depends only on the statistics of the measured data, Y. Note that the estimator $\hat{X}$ is a random variable as it is a function of the data, Y, which is a random variable. To avoid the trivial estimator, $\hat{X} = x$, constraints, such as unbiasedness of the estimator, are imposed making the estimation problem a constrained optimization problem. The MLE is a popular choice in the classical approach. Since no information about the prior is considered, at lower signal to

noise ratios, these estimators which rely on fitting to the data, will fit to noise, resulting in poor estimation.

On the other hand, the Bayesian approach involves inferring the parameter from the data through the posterior distribution,

$$P_{X|Y}(x|y) = \frac{P_X(x)P_{Y|X}(y|x)}{P_Y(y)}. \tag{2.1}$$

This is also known as the Bayes rule and hence the name Bayesian estimation for this approach. Note that the posterior distribution assumes a prior distribution on the parameter X. Typically, the posterior distribution, $P_{X|Y}(x|y)$, based on the conditional data statistics, becomes highly peaked at the value x when there are many measurements available. There are two major concerns of using such an approach as opposed to the classical one. The first is on the use of any distribution *a priori* on X, which can introduce subjectivity into the estimation process. This calls into question the appropriateness of a particular prior. The second concern is that the Bayesian approach is inherently computationally intensive, since it is based on the posterior distribution which requires calculating the probabilities through multi-dimensional integrals. The second issue is mostly resolved with the advent of good computational algorithms, for example Markov chain Monte Carlo (MCMC) methods. Moreover, at lower SNR values, a good estimate is arrived at only after a considerable computational expense. In most cases, the first issue, the validity of a prior PD, can be addressed based on the underlying physics of the problem, and it only makes sense to assume a scientifically sound prior rather than none at all.

In this dissertation, we use a Bayesian approach to the parameter estimation problem. This means that the parameter X is assumed to be a random variable with a prior distribution P_X.

2.2 Minimum mean square error (MMSE)

The error in estimating a Bayesian parameter, X, by means of an estimator $\hat{X}(Y)$ is $\epsilon = \hat{X}(Y) - X$, and its MSE is the following expectation over X and Y:

$$\mathcal{E} = \mathbb{E}[\hat{X}(Y) - X]^2 = \int_{\mathcal{R}_X} dx \int_{\mathcal{R}_Y} dy\, [\hat{X}(y) - x]^2 P(x, y). \tag{2.2}$$

We will avoid writing $\mathcal{R}_X$ and $\mathcal{R}_Y$ whenever they are clearly evident from the integrals. We will explicitly state them if needed. Minimizing the MSE with respect to the estimator function $\hat{X}(Y)$ requires setting the corresponding first-order variation of the right-hand side (RHS) of Eq. (2.2) to zero for each data value y,

$$\int dx\, 2[\hat{X}(y) - x]P(x, y) = 0, \tag{2.3}$$

which can be easily simplified to yield the following well-known form of the MMSEE:

$$\hat{X}_M(y) = \frac{\int dx\, x\, P(x, y)}{\int dx\, P(x, y)} = \int dx\, x\, \frac{P(x, y)}{P(y)} = \int dx\, x\, P_{X|Y}(x|y), \tag{2.4}$$

where the marginal, $P_Y(y) = \int dx\, P(x, y)$ and the Bayes rule, Eq. (2.1), are used to arrive at the final expression. The MMSEE is thus simply the posterior mean of the parameter being estimated.

Among its general properties, we note, as a simple averaging of expression (2.4) over y demonstrates, that the mean value of the MMSEE is the mean value of the Bayesian parameter over its prior. The MMSE is the MSE of is optimal estimator. We will use $\langle . \rangle$ to denote $\mathbb{E}[.]$, wherever necessary to make the expressions more readable.

We may relate the MMSE to the variance of the MMSEE,

$$\sigma^2_{MMSEE} = \mathbb{E}[\hat{X}_M(Y) - \langle X \rangle]^2 = \int dy\, [\hat{X}_M(y) - \langle X \rangle]^2 P_Y(y), \tag{2.5}$$

by writing, in expression (2.2), $[\hat{X}_M(y) - x]$ as $[\hat{X}_M(y) - \langle X \rangle] - [x - \langle X \rangle]$ and then squaring it. This procedure is easily seen to yield three terms for the MMSE, $\mathcal{E}_M$:

$$\begin{aligned} \mathcal{E}_M =& \mathbb{E}[\hat{X}_M(Y) - X]^2 && (2.6) \\ =& \sigma_X^2 + \sigma_{MMSEE}^2 - 2 \int dx \int dy \, [\hat{X}_M(y) - \langle X \rangle][x - \langle X \rangle] \, P(x, y). && (2.7) \end{aligned}$$

But, according to Eq. (2.4), $\int dx \, x \, P(x, y) = P_Y(y) \hat{X}_M(y)$, which reduces the third term on the RHS of Eq. (2.7) to $-2\sigma_{MMSEE}^2$, and the following important relation between the MMSE and the variances of the prior and the MMSEE results:

$$\mathcal{E}_M = \sigma_X^2 - \sigma_{MMSEE}^2. \tag{2.8}$$

This result has an important consequence for Bayesian estimation. Since MMSE is non-negative, the MMSEE variance can never exceed the variance of the prior. Indeed, in the limit of infinite SNR, the MMSE is expected to vanish, indicating that the MMSEE, like all other good estimators, provides a perfect estimate of the Bayesian parameter. In the opposite limit of vanishing SNR, the variance of the MMSEE is expected to vanish, since $P(x|y) \to P(x)$ and the MMSEE simply becomes the mean $\langle X \rangle$, which is a constant entirely independent of the data. The data provide no information about the parameter in this limit, and the MMSE reduces to the variance of the prior information about the parameter.

Both the MMSE and the MMSEE are, in general, difficult to compute because the posterior distribution, $P_{X|Y}$, is difficult to compute. The posterior distribution from Eq. (2.1) can be expressed as

$$P_{X|Y}(x|y) = \frac{P_{Y|X}(y|x) P_X(x)}{P_Y(y)} = \frac{P_{Y|X}(y|x) P_X(x)}{\int dx \, P_{Y|X}(y|x) P_X(x)}. \tag{2.9}$$

The statistics of the data, P_Y, are known only through the conditional distribution, $P_{Y|X}$. It is the PD of the data in the denominator that usually causes the problems as it is the expectation of the channel PD over the prior, P_X. The integral cannot, in

general, be evaluated in closed form, but more importantly the posterior is difficult to compute accurately as the ratio (2.9) becomes numerically ill behaved far in the wings of its numerator and denominator. Therefore, many *ad-hoc* estimators have been proposed. In the next section, the widely used MLE and MAP estimator are reviewed.

2.3 MLE and the maximum *a posteriori* (MAP) estimator

The MLE is the maximum of the likelihood function $L(x|y) = P_{Y|X}(y|x)$ with respect to x:

$$\hat{X}_{MLE}(y) = \underset{x \in \mathcal{R}_X}{\operatorname{argmax}} P_{Y|X}(y|x). \tag{2.10}$$

In other words, the MLE picks that value of X from which the data value, Y, is most likely to have arisen. The MLE estimator has been researched extensively in the past few decades. A very powerful iterative algorithm called the expectation-minimization (EM) algorithm was proposed as a numerically expedient approach to obtain the MLE iteratively [95–97]. Many extensions and modifications to EM algorithm have been proposed and applied to a variety of statistical problems [98].

The MLE may also be shown to be optimally efficient [6], i.e., the squared error of the MLE reaches the CRB, which is a lower bound on the variance of any unbiased estimator, when there are many measurements. This can also be understood in terms of the asymptotic behavior of the posterior distribution. The posterior distribution approaches a Gaussian distribution [99] when there are many measurements, a result that we call the Bayesian central limit theorem. As the mean and the mode coincide for a Gaussian distribution, the MLE is nothing but the MMSE estimator in the high SNR limit. The MLE becomes biased, however, when there are only a few

measurements. This departure from being an optimally efficient estimator when there are only a few measurements was studied in [100]. Moreover, as the prior is not taken into account, the MLE is not useful in the transition region. Seen from an SNR point of view, the MLE will fit to the data at high SNR but to the noise, and thus needs to be smoothed, at low SNR. We can incorporate any prior information available to smooth out the noise arising from the observations at low SNR. This will lead us to the MAP estimator which is by the maximum of the posterior distribution,

$$\hat{X}_{\mathrm{MAP}}(y) = \underset{x\in\mathcal{R}_X}{\operatorname{argmax}}\, P(x|y) \tag{2.11}$$

$$= \underset{x\in\mathcal{R}_X}{\operatorname{argmax}}\, P(y|x)P(x) \tag{2.12}$$

Note that the MAP estimator also involves the posterior distribution, (2.11), but the calculation requires finding only the extremum of the posterior distribution and does not utilize the entire posterior distribution, unlike the MMSE estimator Eq. (2.4). The iterative EM algorithm can also be applied to the MAP estimation by modifying the objective function for the maximization step [98]. Even though in most cases the MAP estimator is used in place of the MMSE estimator, it is still an *ad-hoc* estimator in the mean squared sense. Especially as the SNR decreases, the posterior distribution is no longer Gaussian, and generally, its mean and mode do not coincide, which results in a bias.

To understand this biased behavior of the MAP estimate, consider, for example, the estimation of a parameter with a uniformly distributed prior, X, when the SNR is low. Since the observations carry little information when the SNR is low, we have $P(x|\bar{y}) \to P(x)$ and as a result, $\hat{X}_{\mathrm{MAP}}(\bar{y}) = \underset{x}{\operatorname{argmax}}\, P(x|\bar{y}) \to \tilde{X}$, where $\tilde{X}$ is a uniform random variable with the same support as X, but independent of X. Therefore the MSE of the MAP estimator becomes

$$\begin{aligned}\mathcal{E}_{\mathrm{MAP}} &= \mathbb{E}[\hat{X}^2_{\mathrm{MAP}}] + \mathbb{E}[X^2] - 2\mathbb{E}[X\hat{X}_{\mathrm{MAP}}] \\ &= \mathbb{E}[\hat{X}^2_{\mathrm{MAP}}] + \mathbb{E}[X^2] - 2\mathbb{E}[X]\mathbb{E}[\hat{X}_{\mathrm{MAP}}]\end{aligned}$$

$$= 2(\mathbb{E}[X^2] - \mathbb{E}[X]^2)$$
$$= 2\sigma_X^2$$

This simple calculation shows that the MAP estimator is not optimal in the mean squared sense and can be improved. As we will see in Sec. 4.3.2, this apparent bias of the MAP estimator is removed by an application of our new upper bound so that the MAP error follows the MMSE tightly.

We will now review two of the more popular Bayesian lower bounds on MMSE, the ZZB and the WWB.

2.4 Ziv Zakai bound (ZZB)

The ZZB family of lower bounds relate the MSE to the probability of error in a binary hypothesis testing problem. Initially developed by Ziv and Zakai [26], it underwent improvements and modifications in the works of Seidman [37], Chanaz et al. [38], Bellini et al. [39], and Weinstein [101]. All of these were derived for the case when the parameter, X, is uniformly distributed with a finite support. The Bellini-Tartara bound [39] is the tightest in this family. Later on, Bell et al. [1] extended the Bellini-Tartara formulation of the ZZB, which we will call the extended ZZB (EZZB), to include vector parameters and arbitrary priors. Bell [41] also proved some important convergence properties for elliptically distributed priors, and generalized the EZZB to arbitrary distortion measures, including error measures other than the MSE. Some weaker but practical bounds were also formulated. In this dissertation, we will use the EZZB to compare our upper bound performance.

The derivation of the EZZB involves extending the Kotelnikov's inequality [40] to arbitrary prior distribution. The EZZB is based upon a well known relation between

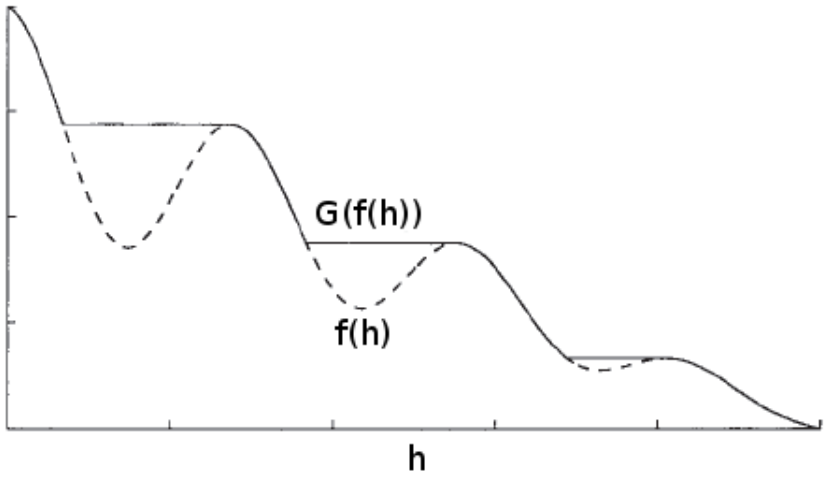

Figure 2.1: Valley filling function, reproduced from [1]

MSE and the probability of outage error [102]:

$$\mathcal{E} = \frac{1}{2}\int_0^\infty dh\, h \Pr\big(|\epsilon| \geq \frac{h}{2}\big). \tag{2.13}$$

By writing $\epsilon = \hat{X} - X$, the probability of outage error is then expressed as

$$\begin{aligned}\Pr\big(|\epsilon| \geq \frac{h}{2}\big) = \int dx \big[P_X(x) + P_X(x+h)\big] \\ \Big[\frac{P_X(x)}{P_X(x)+P_X(x+h)} \Pr\Big(\hat{X} > x + \frac{h}{2}\Big| x\Big) \\ + \frac{P_X(x+h)}{P_X(x)+P_X(x+h)} \Pr\Big(\hat{X} < x + \frac{h}{2}\Big| x+h\Big)\Big],\end{aligned} \tag{2.14}$$

which is related to the binary hypothesis testing problem:

$$\begin{aligned} H_0 : X = x; \quad & \Pr(H_0) = \frac{P_X(x)}{P_X(x)+P_X(x+h)}; \quad Y \sim P(Y|X=x) \\ H_1 : X = x+h; \quad & \Pr(H_1) = 1 - \Pr(H_0); \quad\quad\quad Y \sim P(Y|X=x+h) \end{aligned} \tag{2.15}$$

The term in the square brackets of Eq. (2.14) is the probability of error for the

sub-optimal decision scheme:

$$\begin{aligned} &\text{Decide } \ H_0 : X = x; \qquad \text{if } \ \hat{X}(Y) \leq x + \frac{h}{2} \\ &\text{Decide } \ H_1 : X = x + h; \ \text{ if } \ \hat{X}(Y) \geq x + \frac{h}{2} \end{aligned} \tag{2.16}$$

It is then lower bounded by the minimum probability of error, $P_{\min}(x, x+h)$, for the binary hypothesis testing problem (2.15) yielding the inequality

$$\Pr\big(|\epsilon| \geq \frac{h}{2}\big) \geq G\Big\{ \int dx \big[P_X(x) + P_X(x+h)\big] P_{\min}(x, x+h) \Big\}, \tag{2.17}$$

where $G[.]$ is the valley filling function illustrated in Fig. 2.1. For any function $f(h)$, $G[f(h)]$ is a non-increasing function of h obtained by filling in any valleys in $f(h)$, i.e., for every h,

$$G[f(h)] = \max_{\varepsilon \geq 0} f(h + \varepsilon). \tag{2.18}$$

combining the relations (2.13) and (2.17), we arrive at the EZZB:

$$\mathcal{E} \geq \text{EZZB} = \frac{1}{2} \int dh\, h\, G\Big\{ \int dx \big[P_X(x) + P_X(x+h)\big] P_{\min}(x, x+h) \Big\}, \tag{2.19}$$

where the integration region for x is where $P(x)$ and $P(x+h)$ are non-zero, since $P_{\min}(x, x+h)$ will be zero when one of the hypothesis has zero probability, and the integration region for h is the length of the interval where $P(x)$ is non-zero. The bound on probability of outage error given in Eq. (2.17) reduces to the Kotelnikov's inequality [40] when X is uniformly distributed in $[0, T]$ and the bound on MSE in Eq. (2.19) reduces to the version of Bellini et al. [39]. Further, if the valley-filling function is omitted the version of ZZB due to Chanaz et al. is obtained [38]. Note that the minimum probability of error for the binary hypothesis testing problem in (2.15) can be written as

$$P_{\min}(x, x+h) = \int dy\ \min\{\Pr(H_0) P(y|x), \Pr(H_1) P(y|x+h)\} \tag{2.20}$$

Using (2.20), along with the expressions for the hypothesis probabilities, $\Pr(H_0)$ and $\Pr(H_1)$, in (2.15), in the EZZB (2.19), we get

$$\text{EZZB} = \frac{1}{2}\int dh\, hG\Big(\int dx \int dy \min\{P(x,y), P(x+h,y)\}\Big). \tag{2.21}$$

The EZZB was widely used in all regions of operation and applied to various fields [41, 94]. Even though the ZZB is optimal for posterior distributions that are symmetric and unimodal [41], this condition is not true for finite-support prior PDs, since for such priors, the posterior distribution becomes quite skewed for values of the parameter close to the support boundaries. As a result, the bound on the error probability for the sub-optimal binary decision scheme, in terms of which the EZZB is defined, becomes looser on the whole and thus the EZZB becomes less tight. A widely employed example of such a prior is the uniform prior.

Yet another case of sub-optimality of the EZZB, where it fails to reach the variance of the prior distribution at low SNRs, is obtained when the prior probability distribution has the following properties: a) the prior is defined for positive values only, $\mathcal{R}_X = [0, \infty)$ and is a monotonically decreasing function and b) $\frac{4}{3}\mathbb{E}^2[X] \leq \mathbb{E}[X^2]$ or $\mathbb{E}^2[X] \leq \sigma_X^2$. For such priors, we will show that the EZZB will not reach the prior variance when SNR is low. As SNR is reduced, the dependence of the data on the prior becomes weak and in the limiting case we can assume that $P(y|x) = P(y|x+h)$. Therefore, in the limiting case of zero SNR, EZZB becomes

$$\begin{aligned}
\text{EZZB} &= \frac{1}{2}\int_0^\infty dh\, hG\Big(\int_0^\infty dx \int dy \min\{P(y|x)P(x), P(y|x+h)P(x+h)\}\Big) \\
&= \frac{1}{2}\int_0^\infty dh\, h \int_0^\infty dx\, P_X(x+h) \\
&= \frac{1}{2}\int_0^\infty dh\, h \int_h^\infty d\phi\, P_X(\phi)
\end{aligned} \tag{2.22}$$

$$= \frac{1}{4}\int_0^\infty dh\, h^2 P_X(h)$$
$$= \frac{\sigma_X^2 + \mathbb{E}^2[X]}{4} \leq \sigma_X^2,$$

where the second equality follows from the monotonically decreasing nature of P_X and the fourth equality follows by doing an integration by parts. This simple calculation suggests that the EZZB is not optimal in general. Moreover, we will see from the results in Sec. 4.4, even for cases for which the EZZB is expected to perform well, the EZZB is not tight as our upper bound in the transition region.

2.5 Weiss Weinstein bound

Weiss and Weinstein developed lower bound on the MMSE using the Schwarz inequality [43]. Consider a function $\psi(x, y)$ satisfying the following equality:

$$\int_{\mathcal{R}_X} dx\, \psi(x, y) P(x, y) = 0, \tag{2.23}$$

for any y. Such functions are orthogonal to any transformation, $g(Y)$, of the data:

$$\mathbb{E}\big[(g(y)\psi(x, y)\big] = \mathbb{E}_Y\big[g(y)\mathbb{E}_{X|Y}[\psi(x, y)]\big] = 0, \tag{2.24}$$

This equation holds for any estimator $\hat{X}(Y)$, as it depend only the data Y. From Eq. (2.24), we can write

$$\mathbb{E}\big[(x - g(y))\psi(x, y)\big] = \mathbb{E}\big[x\psi(x, y)\big] \tag{2.25}$$

By applying Schwarz inequality to the left hand side of Eq. (2.25), we arrive at the WWB:

$$\mathcal{E} \geq \frac{\mathbb{E}^2[x\psi(x, y)]}{\mathbb{E}[\psi^2(x, y)]}. \tag{2.26}$$

By choosing a function ψ satisfying Eq. (2.23), various bounds on MMSE can be generated. For example, a test function involving the MMSE estimator,

$$\psi(x,y) = x - \mathbb{E}_{X|Y}(x|y), \tag{2.27}$$

generates the inequality $\mathcal{E} \geq \mathcal{E}_M$, which proves that the MMSE is indeed the minimum MSE.

By selecting

$$\psi(x,y) = \frac{\partial \ln P(x,y)}{\partial x} \tag{2.28}$$

the Bayesian version of the CRB [21] is generated, given by

$$\mathcal{E} \geq \left\{ \mathbb{E}\Big[\Big(\frac{\partial \ln P(x,y)}{\partial x} \Big)^2 \Big] \right\}^{-1}. \tag{2.29}$$

It should be noted that the Bayesian CRB is very restrictive as it requires the joint PD to be differentiable. This implies that a very common prior, the uniform distribution, does not affect the Bayesian CRB.

Using various test functions ψ, the Bayesian version of the Bhattacharya bound and the Bobrowsky-Zakai bound can be derived using this formulation [45]. The Bayesian Bhattacharya bound needs to satisfy some regularity conditions, but the Bobrowsky-Zakai bound avoids the regularity conditions by using finite differences instead of local derivatives.

To obtain the original Weiss-Weinstein bound, choose

$$\psi(x,y) = \mathcal{L}^s(y; x+h, x) - \mathcal{L}^{1-s}(y; x-h, x), \tag{2.30}$$

where $0 \leq s \leq 1$ and h are free parameters and

$$\mathcal{L}(y; x_1, x_2) = \frac{P(x_1, y)}{P(x_2, y)}. \tag{2.31}$$

The resulting lower bound on the MSE does not suffer from any regularity conditions, and is given by:

$$\begin{aligned}\mathcal{E} \geq & \text{WWB} \\ = & \max_{0\leq s\leq 1, h} \frac{h^2 \exp\big(2\mu(s,h)\big)}{\exp\big(\mu(2s,h)\big) + \exp\big(2\mu(2-2s,-h)\big) - 2\exp\big(\mu(s,2h)\big)},\end{aligned} \tag{2.32}$$

where

$$\begin{aligned}\mu(s,h) = & \ln \mathbb{E}[\mathcal{L}^s(y;x,x+h)] \\ = & \ln \int dy \int dx\, P^s(x+h,y) P^{1-s}(x,y),\end{aligned} \tag{2.33}$$

which can be related to the Chernoff distance [103] between two statistical distributions, $f_0(x)$ and $f_1(x)$, given by

$$\int dx\, f_1^s(x) f_0^{1-s}(x). \tag{2.34}$$

The computation of the WWB involves choosing a large number of test points, s and h. But, it is generally observed that $s = \frac{1}{2}$ is a good choice. For $s = \frac{1}{2}$, the WWB can be written as

$$\text{WWB} = \max_h \frac{h^2 \exp\big(2\mu(\frac{1}{2},h)\big)}{2\big(1-\exp\big(\mu(\frac{1}{2},2h)\big)\big)}. \tag{2.35}$$

The MMSEE, the MAP estimation error, the ZZB, and the WWB are computed for two specific problems which we will review in the following sections.

2.6 Illustrative problems

2.6.1 Poisson channel with exponential prior

A linear Poisson channel or the discrete-time Poisson channel appears naturally in many optical systems, e.g., an electron multiplied charge coupled device (EM-CCD)

imaging system can be modeled using a Poisson channel and an exponential prior. A linear Poisson channel with linear gain, a, and bias, b, is one where the conditional mean of the data, Y, given the input, X, is $\mathbb{E}[Y|X = x] = ax + b$. The conditional probability distribution is given by

$$P(y|x) = \frac{(ax+b)^y}{y!} \exp[-(ax+b)], \quad y = 0, 1, 2, \ldots \tag{2.36}$$

The Bayesian parameter X being estimated is taken to have a negative-exponential prior,

$$P_X(x) = \begin{cases} \dfrac{1}{\langle X \rangle} \exp\left(-\dfrac{x}{\langle X \rangle}\right) & \text{for } x > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2.37}$$

The posterior PD for X, namely $P_{X|Y}(x|y)$, may be calculated by applying the Bayes theorem (2.9) to the probability distributions (2.36) and (2.37),

$$P_{X|Y}(x|y) = \frac{1}{P_Y(y)} \frac{1}{\langle X \rangle} \frac{(ax+b)^y}{y!} \exp\left[-\left(a + \frac{1}{\langle X \rangle}\right)x - b\right], \quad x \geq 0; \quad y = 0, 1, \ldots, \tag{2.38}$$

where $P_Y(y)$ is given by,

$$P_Y(y) = \int_0^\infty \frac{dx}{\langle X \rangle} \frac{(ax+b)^y}{y!} \exp\left[-\left(a + \frac{1}{\langle X \rangle}\right)x - b\right], \quad y = 0, 1, \ldots, \tag{2.39}$$

$$= \frac{1}{y!} \frac{(a\langle X \rangle)^y}{(1 + a\langle X \rangle)^{y+1}} \exp\left(\frac{b}{a\langle X \rangle}\right) \Gamma\left[y+1, \frac{b(1+a\langle X \rangle)}{a\langle X \rangle}\right], \tag{2.40}$$

where

$$\Gamma(y+1, u) = \int_u^\infty dx\, x^y \exp(-x), \tag{2.41}$$

is the incomplete Gamma function. The detailed derivation of $P_Y(y)$ can be found in Appendix A.1. From equations (2.38), (2.39) and (2.4), we have the following

expression for the MMSEE,

$$\hat{X}_M(y) = \frac{1}{P_Y(y)} \int_0^\infty dx\, x \frac{(ax+b)^y}{\langle X \rangle\, y!} \exp\Big[-\Big(a + \frac{1}{\langle X \rangle}\Big)x - b \Big] \tag{2.42}$$

$$= \frac{1}{P_Y(y)} \int_0^\infty dx\, \frac{(ax+b)-b}{a} \frac{(ax+b)^y}{\langle X \rangle\, y!} \exp\Big[-\Big(a + \frac{1}{\langle X \rangle}\Big)x - b \Big] \tag{2.43}$$

$$= \frac{1}{P_Y(y)} \int_0^\infty dx\, \frac{(ax+b)^{y+1}}{a\,\langle X \rangle\, y!} \exp\Big[-\Big(a + \frac{1}{\langle X \rangle}\Big)x - b \Big] - \frac{b}{a} \tag{2.44}$$

$$= \frac{(y+1)P_Y(y+1)}{aP_Y(y)} - \frac{b}{a}, \tag{2.45}$$

where in going from (2.42) to (2.43), we used $x = \dfrac{(ax+b)-b}{a}$. Taking the squared mean of the MMSEE yields

$$\mathbb{E}[\hat{X}_M^2] = \frac{1}{a} \sum_{y=0}^\infty \frac{(y+1)^2 P^2(y+1)}{P(y)} - 2 \sum_{y=0}^\infty \frac{b}{a^2}(y+1)P(y+1) + \frac{b^2}{a^2} \sum_{y=0}^\infty P(y)$$

$$= \frac{1}{a} \sum_{y=0}^\infty \frac{(y+1)^2 P^2(y+1)}{P(y)} - \frac{2b}{a} \langle X \rangle - \frac{b^2}{a^2}. \tag{2.46}$$

According to Eq. (2.8), subtracting this expression from $\langle X^2 \rangle$, which is $2\langle X \rangle^2$ for an exponential PD, yields the desired MMSE,

$$\mathcal{E}_M = 2\langle X \rangle^2 + \frac{2b}{a} \langle X \rangle + \frac{b^2}{a^2} - \frac{1}{a^2} \sum_{y=0}^\infty \frac{(y+1)^2 P^2(y+1)}{P(y)}. \tag{2.47}$$

The MAP estimator, from Eq. (2.11), is given by

$$\hat{X}_{\text{MAP}}(y) = \frac{\langle X \rangle}{a\langle X \rangle + 1} y - \frac{b}{a}, \tag{2.48}$$

which is linear in y. The MAP estimation error, $\mathcal{E}_{\text{MAP}}$, can be calculated by substituting expression (2.48) for the MAP estimator into expression (2.2) for the MSE. To calculate the EZZB, Eq. (2.21), we first need to calculate $\min\{P(x,y), P(x+h,y)\}$. To facilitate this calculation we first observe that the equivalency of the following two inequalities:

$$P(x,y) \geq P(x+h,y) \Leftrightarrow y \leq \Big\lceil \frac{a + \frac{1}{\langle X \rangle} h}{\ln(1 + \frac{ah}{a\langle X \rangle + b})} \Big\rceil, \tag{2.49}$$

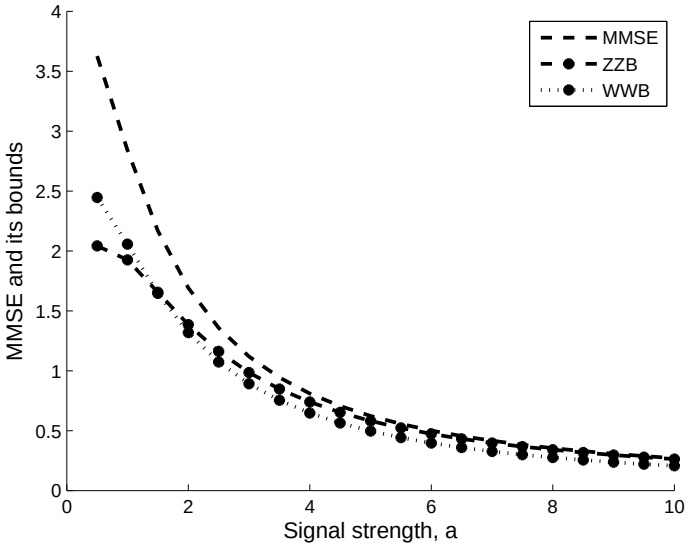

Figure 2.2: The MMSE, the EZZB and the WWB for the Poisson channel, with bias, b = 10 and exponential prior with mean $\langle X \rangle = 2$.

where $\Leftrightarrow$ means "if and only if", and [.] represents the integer part of a real number. We can now write the EZZB as

$$\text{EZZB} = \frac{1}{2}\int_0^\infty dh\, h\, G\Big\{\int_0^\infty dx\big[\sum_{y=0}^{Y_l} P(x+h,y) + \sum_{y=Y_l+1}^{\infty} P(x,y)\big]\Big\}, \tag{2.50}$$

which will be calculated numerically. To calculate the WWB we need to calculate the mu-function, $\mu(s,h)$, from (2.33). It can be written as

$$\exp(\mu(s,h)) = \sum_{y=0}^{\infty}\int_0^\infty dx\, P^s(x+h,y)P^{1-s}(x,y) \tag{2.51}$$

$$= \int_0^\infty dx\, P^s(x+h)P^{1-s}(x)\sum_{y=0}^{\infty} P^s(y|x+h)P^{1-s}(y|x). \tag{2.52}$$

After using the expressions for $P(x)$ and $P(y|x)$ from (2.37) and (2.36), respectively, and some simple arithmetic, we can carry out the summation over y using a Taylor

series expansion for the exponential function to give

$$\exp(\mu(s,h)) = \frac{\exp\left(\frac{b-sah(1+a\langle X\rangle)}{a\langle X\rangle}\right)}{a\langle X\rangle}\int_b^\infty d\theta \exp\Big(\theta\big(1+\frac{ah}{\theta}\big)^s - \frac{1+a\langle X\rangle}{a\langle X\rangle}\theta\Big) \quad (2.53)$$

where $\theta = ax + b$. By using the following approximation

$$\big(1+\frac{ah}{\theta}\big)^s = 1 + \frac{sah}{\theta}, \quad \text{for } \theta > Nah, \quad (2.54)$$

where N is a large number, the integral in Eq. (2.53) can be simplified by writing it as a sum over the two intervals $[b, Nah]$ and $[Nah, \infty)$, the latter of which can be easily calculated. The exponential of the mu-function is then approximated as

$$\begin{aligned}\exp(\mu(s,h)) \approx & \exp\Big(\frac{b-ah(s+N)}{a\langle X\rangle}\Big) \\ & + \frac{\exp\left(\frac{b-sah(1+a\langle X\rangle)}{a\langle X\rangle}\right)}{a\langle X\rangle}\int_b^{Nah} d\theta \exp\Big(\theta\big(1+\frac{ah}{\theta}\big)^s - \frac{1+a\langle X\rangle}{a\langle X\rangle}\theta\Big)\end{aligned} \quad (2.55)$$

Using this expression in Eq. (2.32), and carrying out the optimization over s and h numerically, we can calculate the WWB. The results are plotted in Fig. 2.2 where we can see that there is considerable gap between the MMSE and its lower bounds, the ZZB and the WWB. Since the MAP estimator is an unoptimized linear estimator, its error can not approximate the MMSE tightly, resulting in a large error at low SNRs. We used a log-log plot in Fig. 2.3 to include the MAP error alongside the other results. This clearly shows that the MAP estimator is not able to handle bias very well, but as we will see in Sec. 4.3.2, a simple application of our upper bound brings, by a simple linear modification, the MAP estimation error close to the MMSE.

2.6.2 Time-delay estimation (TDE)

Time delay estimation (TDE) problem arises in a variety of signal processing scenarios, such as radar, communications, and geo-location, and TDE performance analysis

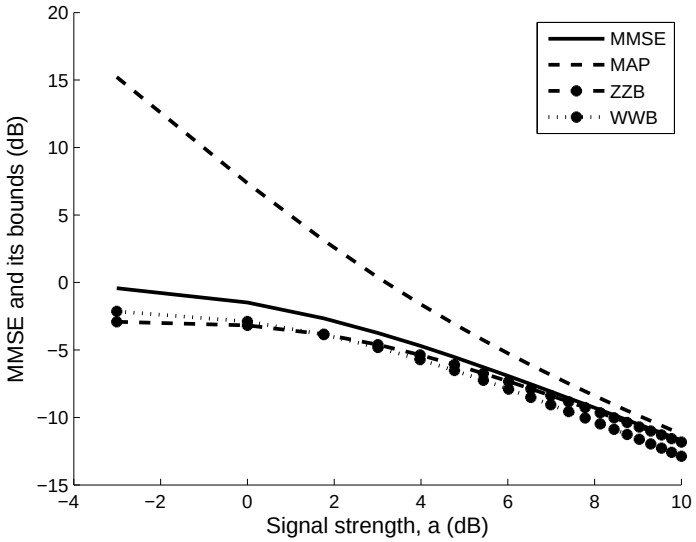

Figure 2.3: Log-log plot for the MMSE, the MAP error, the EZZB and the WWB for the Poisson channel, with bias, b = 10 and exponential prior with mean $\langle X \rangle = 2$.

is needed to find limits in these applications. Typically, the location of a radiating source can be determined by passive observation of its signal at two or more spatially separated receivers. All the relevant information about source location (i.e., bearing and range) is contained in the relative (differential) delay of the signal wavefront to the various receiver pairs. TDE and localization have therefore attracted a great deal of interest in the literature [3, 4, 72, 89, 91–93, 104–107].

To motivate the problem of TDE , consider two spatially separated sensors that receive a signal from a radiating point source, as shown in Fig. 2.4. Assuming a constant velocity, v, of propagation for the signal, the difference of the distances of the source and the two sensors is directly proportional to the time difference of signal arrival times at the sensors. So the time delay is nothing but the the difference in path lengths (from the source to each sensor) divided by the propagation velocity. Relative to the sensors, the source lies on a well defined locus of points for which the

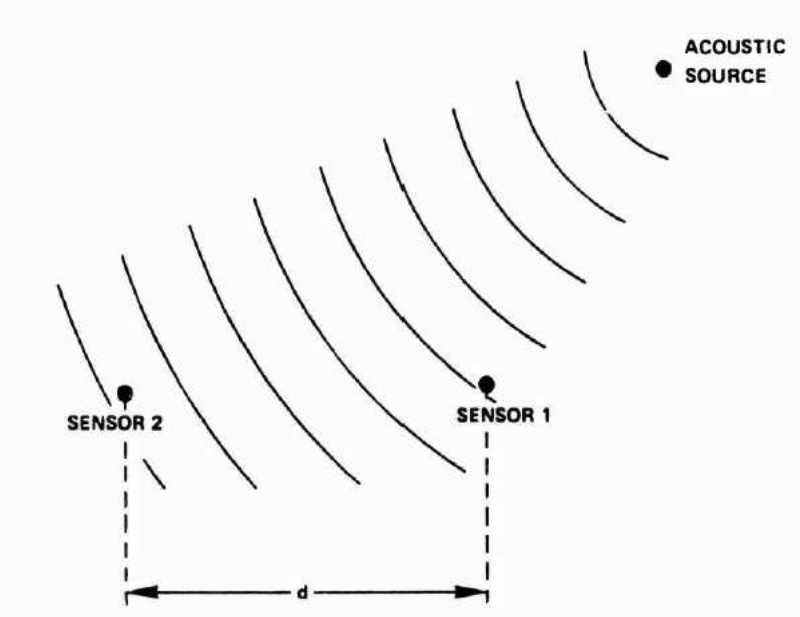

Figure 2.4: Figure reproduced from [2] representing acoustic source and sensors. The source need not be acoustic, in general.

time delay is constant. Thus, knowing the time delay, one can locate the source on the locus of points, which is a hyperboloid as shown in Fig. 2.5. When the source and the sensors are in a single plane, the source location is estimated from the intersection of two or more hyperbolas as shown in Fig. 2.6. The remaining ambiguity can be resolved by the use of *a priori* information or by using a fourth sensor or with the bearing estimates.

In its simplest form the received signals at two sensors in a TDE problem can be stated as follows:

$$\begin{aligned} y_1(t) =& s(t) + n_1(t), \\ y_2(t) =& s(t-\tau) + n_2(t), \end{aligned} \tag{2.56}$$

where τ is the delay parameter, $n_1(t)$ and $n_2(t)$ are uncorrelated additive white Gaussian processes and the signal $s(t)$ is assumed to be uncorrelated to the noises. This assumption is valid when the receivers are widely separated and the source is

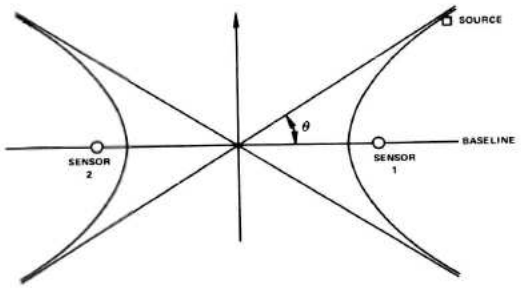

Figure 2.5: Figure reproduced from [2] representing acoustic source and sensors. The source need not be acoustic, in general.

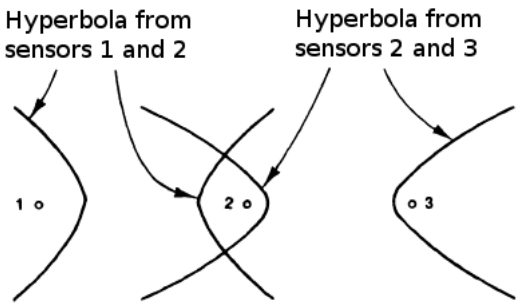

Figure 2.6: Figure reproduced from [3] representing acoustic source and sensors. The source need not be acoustic, in general.

sufficiently far off from the sensors. The sensor-to-sensor delay τ is assumed to be uniformly distributed and confined to:

$$\tau_l \leq \tau \leq \tau_u. \tag{2.57}$$

Such a domain of the prior could be based on the knowledge from the known receiver separation and the known velocity of pulse propagation in the medium. A further assumption is made on the observation time, $t \in [0, T]$, that the delayed pulse $s(t-\tau)$ be contained in $[0, T]$.

The time delay estimate $\hat{\tau}$ can be used to estimate the bearing angle, as shown

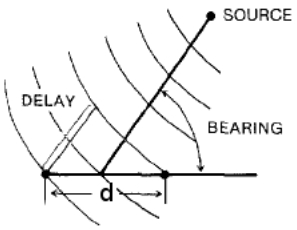

Figure 2.7: Figure reproduced from [4] representing a planar model of two sensors separated by a distance d.

in Fig. 2.7, approximately as

$$\hat{\theta} = \arccos\left(\frac{v\hat{\tau}}{d}\right). \tag{2.58}$$

This approximation becomes more accurate, the more distant the point source. Thus, from the discussion in the previous paragraphs, the time delay estimates can be used to calculate the bearing of the source and also the range of the source, thereby localizing the source. Hence, in a typical TDE problem, we are interested in estimating the time delay, τ in Eq. (2.56), corresponding to a signal that is transmitted either intentionally or unintentionally. Though there exists an intrinsic relationship between the two scenarios, there are some important differences. When the transmission is unintentional, no prior knowledge of the signal $s(t)$ is available and the time delay estimate is obtained from measurements at two or more spatially separated sensors. This type of problem is referred to as time difference of arrival (TDOA) estimation and is of concern to passive sonar and microphone array systems. When the transmission is intentional, the TDE problem is also referred to as time of arrival (TOA) estimation where we are interested in estimating the time-delay between the transmission of a pulse and reception after an echo from a possible target. Knowledge of the signal pulse $s(t)$ is assumed and this type of problem is encountered in active RADAR and SONAR systems. Such a time delay estimate can be obtained based on measurements from a single sensor and the TDE problem from Eq. (2.56) becomes

$$y(t) = s(t - \tau) + n(t), \tag{2.59}$$

where $n(t)$ is a sample function of zero-mean stationary Gaussian process with double-sided noise spectral density, N_0. In this dissertation, we will be content with using Eq. (2.59), by assuming a knowledge about the signal pulse shape transmitted, which will serve our purpose of comparing the different bounds on the MMSE and also in evaluating different estimator performances.

We sample the data, uniformly, at N_t time instants $\{t_i | i = 1, 2, \ldots, N_t\}$ such that the time spacing between successive samples is $\Delta t = t_{i+1} - t_i$. The data will then be the set, $\{x(t_1), \ldots, x(t_{N_t})\}$ denoted by the vector $\bar{x}$. The conditional probability distribution of the data, given the delay parameter, can be written as:

$$P(\bar{y}|\tau) = \left(\frac{1}{\sqrt{2\pi\sigma_n^2}}\right)^{N_t} \exp\Big\{-\frac{1}{2\sigma_n^2}\sum_{i=1}^{N_t}[y(t_i) - s(t_i - \tau)]^2\Big\}, \tag{2.60}$$

where σ_n is related to the noise spectral density, N_0, as $\sigma_n^2 = N_0/\Delta t$.

For this problem, analytical expressions for the MMSEE and the MAP estimator are not available. Both $\mathcal{E}_M$ and $\mathcal{E}_{\text{MAP}}$ were therefore calculated through simulations, and are plotted in Fig. 2.8. For the simulations, 1000 noise frames were used and the error bars are also plotted in the figure. Since the parameter τ is uniformly distributed in $\mathcal{R}_\tau : [\tau_l, \tau_u]$, the MAP estimator is given by:

$$\hat{\tau}_{\text{MAP}}(\bar{y}) = \underset{\tau\in\mathcal{R}_\tau}{\text{argmax}}\, P(\tau|\bar{y}), \tag{2.61}$$

$$= \underset{\tau\in\mathcal{R}_\tau}{\text{argmax}}\, P(\bar{y}|\tau), \tag{2.62}$$

$$= \underset{\tau\in\mathcal{R}_\tau}{\text{argmax}}\, \ln P(\bar{y}|\tau), \tag{2.63}$$

$$= \underset{\tau\in\mathcal{R}_\tau}{\text{argmax}} \sum_{i=1}^{N_t} y_i s(t_i - \tau), \tag{2.64}$$

where we assumed that $s(t - \tau)$, for all $\tau \in [\tau_l, \tau_u]$, is contained in [0,T], so that $\sum_i s^2(t_i - \tau)$ is independent of τ. This is the discrete matched filter evaluation of the first term of the Karhunen-Loeve expansion [108] ehich dominates all the other terms when SNR is high.

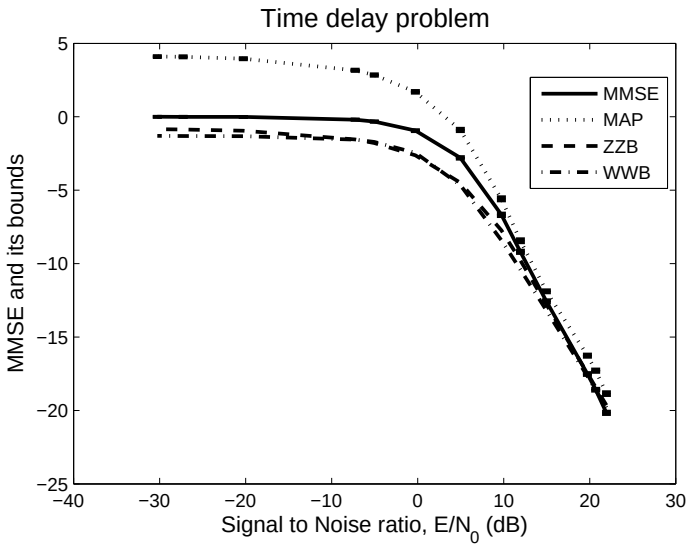

Figure 2.8: MMSE, MAP, EZZB and WWB for the passive TDE problem. The MMSE and the MAP are simulated using 1000 noise frames and are plotted with corresponding error bars.

The expression for the MMSEE, $\hat{\tau}_M$, is a bit more involved as it needs the evaluation of $P(\tau|\bar{y})$ explicitly:

$$\hat{\tau}_M(\bar{y}) = \int_{\mathcal{R}_\tau} d\tau\, \tau P(\tau|\bar{y}), \tag{2.65}$$

$$= \frac{\int_{\mathcal{R}_\tau} d\tau\, \tau P(\bar{y}|\tau)}{\int_{\mathcal{R}_\tau} d\tau\, P(\bar{y}|\tau)}, \tag{2.66}$$

$$= \frac{\int_{\mathcal{R}_\tau} d\tau\, \tau \exp(\frac{1}{\sigma_n^2}\sum_{i=1}^{N_t} y_i s(t_i-\tau))}{\int_{\mathcal{R}_\tau} d\tau\, \exp(\frac{1}{\sigma_n^2}\sum_{i=1}^{N_t} y_i s(t_i-\tau))}. \tag{2.67}$$

The EZZB can be calculated using the form given in (2.19). We will reproduce it here with appropriate limits for the TDE problem:

$$\text{EZZB} = \frac{1}{2}\int_0^{\tau_u-\tau_l} dh\, h\, G\Big\{\int_{\tau_l}^{\tau_u-h} dx\big[P(\tau)+P(\tau+h)\big]P_{\min}(\tau,\tau+h)\Big\}. \tag{2.68}$$

For this particular problem, $P_{\min}(\tau,\tau+h)$ can be calculated analytically [21], and is

given by

$$P_{\min}(\tau, \tau + h) = Q\left(\sqrt{\frac{E}{2N_0}(1 - \rho(h))}\right), \tag{2.69}$$

where E is the signal energy,

$$E = \int dt\, s^2(t), \tag{2.70}$$

$\rho(h)$ is the correlation of the deterministic signal, $s(t)$,

$$\rho(h) = \frac{1}{E} \int dt\, s(t)s(t - h), \tag{2.71}$$

and $Q(z)$ is the complementary error function,

$$Q(z) = \int_z^\infty \frac{dv}{\sqrt{2\pi}} \exp(-\frac{v^2}{2}). \tag{2.72}$$

Note that $P_{\min}(\tau, \tau + h)$ is independent of τ, so that we can denote it by $P_{\min}(h)$ and remove it from the integral inside the valley-filling function, $G[.]$, in Eq. (2.68). By calling the remaining integral inside the valley filling function $A(h)$, we can write it as

$$\begin{aligned} A(h) =& \int_{\tau_l}^{\tau_u - h} dx\big(P(\tau) + P(\tau + h)\big) \\ =& 2\Big(1 - \frac{h}{\tau_u - \tau_l}\Big) \end{aligned} \tag{2.73}$$

Since $A(h)$ and $P_{\min}(h)$ are monotonically decreasing functions of h, we can omit the valley-filling function from Eq. (2.68) and the EZZB becomes

$$\text{EZZB} = \frac{1}{2} \int_0^{\tau_u - \tau_l} dh\, hA(h)P_{\min}(h). \tag{2.74}$$

This final form for the EZZB can be calculated numerically.

Finally, to calculate the WWB we will first evaluate the exponential of the mu-function $\mu(s, h)$ defined in (2.33):

$$\exp \mu(s, h) = \int_0^{\tau_u - h} \frac{d\tau}{\tau_u - \tau_l} \int \cdots \int d\bar{y}\, P^s(\bar{y}|\tau + h)P^{1-s}(\bar{y}|\tau) \tag{2.75}$$

The inner multiple integral being a Gaussian, since $P(\bar{y}|\tau)$ is of a Gaussian from Eq. (2.60), we can carry out simple algebraic manipulations to arrive at

$$\exp \mu(s,h) = \int_0^{\tau_u - h} \frac{d\tau}{\tau_u - \tau_l} \exp[-s(1-s)\frac{E}{N_0}(1-\rho(h))], \tag{2.76}$$

where E and ρ are given by equations (2.70) and (2.71). Since the exponential is not dependent on τ, we can carry out the integration over τ to obtain

$$\exp \mu(s,h) = \frac{1}{2}A(h) \exp[-s(1-s)\frac{E}{N_0}(1-\rho(h))], \tag{2.77}$$

where $A(h)$ is defined in (2.73). Using this expression in Eq. (2.32), and carrying out the optimization over s and h numerically, we can calculate the WWB.

The EZZB and the WWB as well as the MMSE and the MAP estimation error are plotted in Fig. 2.8. We can see that both lower bounds and the MAP estimation error reach the MMSE at high SNR, but as the SNR decreases, there is considerable gap between the MMSE and the rest.

2.7 Summary

In this chapter, we reviewed the MMSE, the MAP estimator, and two of the most popular lower bounds on the MMSE, namely the EZZB and the WWB. Since the MMSEE, as well as the MMSE, are difficult to compute, *ad-hoc* estimators like the MLE and MAP estimators are used. To evaluate their performance, tight bounds on the MMSE are sought, which avoid any explicit calculation of the posterior distribution, $P(x|y)$, given by Eq. (2.9). Both the EZZB and the WWB provide such a formulation avoiding the posterior distribution while still making use of the prior information. Both these bounds can be applied to arbitrary prior distributions and to vector parameters. An attempt was made to relate the two bounds in [41], but there are no other results relating the two bounds analytically. These bounds were

compared numerically for various problems of interest, and it is generally observed that the WWB is tighter in the low SNR region while the EZZB does well in the asymptotic region providing a better estimate of the threshold region [109]. Nevertheless, these two families of lower bounds are difficult to compute [70, 94] and their tightness to the MMSE has not been established analytically. Moreover, optimal estimators achieving these bounds, when they are achievable, or even their achievability are not determined by their general theory.

With the help of illustrative problems where we can compute the MMSE and its bounds with relative ease, we saw that the MAP estimator is not optimal in the MSE sense. It is for this reason and also due to the inability of the MAP estimator to handle bias at low values of the SNR, the MAP estimation error departs from the MMSE as the SNR decreases. Even though for some problems, mainly in the high SNR region, the MAP estimator and the MMSE estimator can be shown to be the same [21], as we move into the low SNR region this is no longer a valid assumption. This is evident from both problems considered in this section, as shown graphically in Fig. 2.2 and Fig. 2.8. The two popular lower bounds, the EZZB and the WWB are also not particularly tight at low to moderate values of the SNR.

In this dissertation, extensions of the EZZB are derived in Chapter 3. Even though theoretical in nature, these bounds could provide valuable insights in deriving better bounds. Our upper bound on MMSE is derived in Chapter 4 and applied to different classes of test estimators. A new piecewise linear estimator is proposed that does well in all the regions of operation. In general, our upper bound is shown to be tighter than the EZZB and the WWB. By using the upper bound, the performance of a new rotating point spread function imager is analyzed in terms of the 2D subpixel localization of a single molecule. Asymptotic expressions for the MMSE and the MMSEE are constructed in Chapter 5 and a different possible approximation to the MMSE is proposed. In Chapter 6 possible further developments involving the

piecewise linear estimator and iterative algorithms for the upper bound are discussed.

Chapter 3

Extensions of the ZZB

3.1 Approximate ZZB

Implementing the EZZB depends on the calculation of the minimum probability of error (MPE) in a binary, or more generally an M-ary, hypothesis testing problem. The EZZB is useful only when the MPE is known or can be bounded tightly. Expressions for the MPE were derived, and numerous approximations and bounds which vary in complexity and tightness for many problems were also evaluated by other researchers [21, 41, 52, 54–57]. All of them are based on certain distance measures between statistical distributions. Here we present an alternative approach that does not depend on any specific distance measure. We will start with the form of the EZZB presented in Eq. (2.21) except for the valley-filling function,

$$\mathcal{E} \geq \frac{1}{2}\int dh\, h \int dx \big(P_X(x) + P_X(x+h)\big) P_{\min}(x, x+h) \tag{3.1}$$

$$= \frac{1}{2}\int dh\, h \int dx \int dy \min\{P(x,y), P(x+h,y)\}. \tag{3.2}$$

It is the MPE $P_{\min}(x, x+h)$, that is usually difficult to compute, and one usually employs bounds on it to facilitate the evaluation of the ZZB [41, 94]. We present a

simple form for the EZZB using the following lower bound on the minimum of two positive numbers, a, b:

$$\min(a,b) \geq \frac{ab}{(a^n + b^n)^{1/n}}, \quad n > 0. \tag{3.3}$$

This inequality provides an accurate approximation, for sufficiently large n and dissimilar values of a, b, for the minimum error probability in Eq. (3.2). Indeed, relation Eq. (3.3) becomes an equality in the limit $n \to \infty$. Use of this inequality yields a lower bound on the EZZB itself,

$$\text{EZZB} \geq \frac{1}{2} \int dh\, h \int \int dx\, dy\, \frac{P(x,y)\, P(x+h,y)}{[P^n(x,y) + P^n(x+h,y)]^{1/n}}, \quad n > 0. \tag{3.4}$$

This is our first result proposed in this dissertation. We will call this the approximate ZZB (AZZB). The advantage of this result will become evident when we apply this result to the problem considered in Sec. 2.6.1. The AZZB for this problem can be expressed as:

$$\text{AZZB} = \frac{1}{2} \int_0^\infty dh\, h\, G\Big\{ \int_0^\infty dx \sum_{y=0}^{\infty} \big[P(x,y)^{-n} + P(x+h,y)^{-n} \big]^{-\frac{1}{n}} \Big\}, \tag{3.5}$$

A quick comparison of Eq. (3.5) with the exact EZZB calculation in Eq. (2.50) reveals that there is no increase in numerical complexity. However, note that, for more complicated problems, the exact calculation of the MPE might not be feasible. In such scenarios, since we have access to $P(x)$ and $P(y|x)$, the joint distribution can be calculated and a MCMC approach can be used to evaluate the integral in (3.4). As we can see from Fig. 3.1 the results are quite accurate. The high accuracy of these results reflects the fact that, in general, the regions of the joint sample space of X, Y where the two joint PDs are comparable and the approximation (3.3) is not accurate, only make small contributions when compared to those made by the rest of the sample space where that approximation can be quite accurate for sufficiently large n.

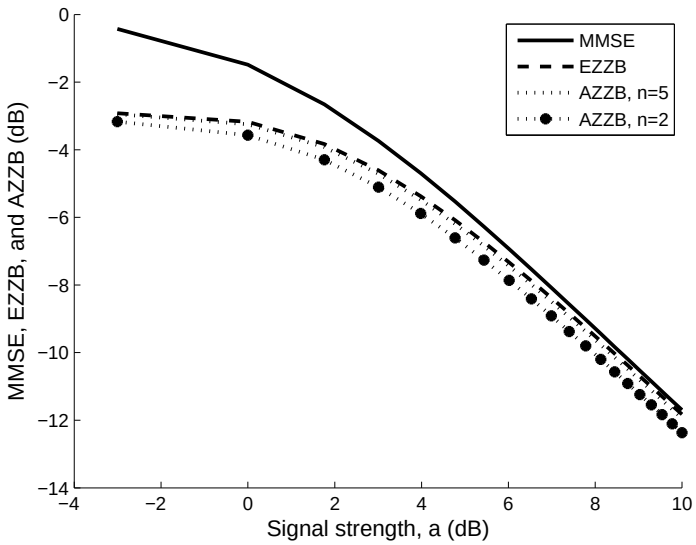

Figure 3.1: Exact and approximate EZZB for a Poisson channel with exponential prior described in Sec. 2.6.1. The approximate EZZB is generated using $n = 2$ and $n = 5$ in Eq. (3.4).

3.2 Beyond the ZZB

It should be possible, in general, to improve upon the EZZB whenever that bound is not tight relative to the MMSE. A possible strategy for doing this is suggested by the progression of ever improving local lower bounds of estimation, starting from the CRB, which is inversely related to a single-point derivative measure of sensitivity, namely Fisher information, to the Chapman-Robbins bound (ChRB), which is inversely related to a single-point first-order difference measure, and ultimately to the Barankin bound (BB), which is a multi-point generalization of the ChRB. Since the ZZB depends on the minimum probability of error of a certain binary-hypothesis testing problem, a possible way of improving upon the EZZB is to express the exact MSE as an M-ary ($M > 2$) hypothesis testing problem. This should enable us to derive a tighter lower bound than the "binary" EZZB. This is what we do next. Some results were derived using this same strategy in [41, 42]. In [42] the Bayes risk consid-

ered is somewhat different from the traditional MSE, and the number of hypotheses depends on the size of the parameter space. In [41] the number of hypotheses was independent of the size of the parameter space, which provided more control over the choice of the number of hypotheses.

We will start the derivation from the relation between the MSE and the probability of outage error as given in Eq. (2.13):

$$\mathcal{E} = \frac{1}{2}\int_0^\infty dh\, h \Pr\big(|\epsilon| \geq \frac{h}{2}\big) \tag{3.6}$$

$$= \frac{1}{2}\int_0^\infty dh\, h \Big(1 - \Pr\big(|\epsilon| < \frac{h}{2}\big)\Big) \tag{3.7}$$

Assuming that $x \in [X_{\min}, X_{max}]$, where the limits can, in general, be $(-\infty, \infty)$, we divide the interval $[X_{\min}, X_{max}]$ into regular sub-intervals, each of size h, for a given h in the outer integrals as shown in Fig. 3.2. This is possible for all $h \leq (X_{max} - X_{\min})$. Note that since $(X_{max} - X_{\min})$ is not, in general, an exact multiple of an arbitrary h, the last sub-interval of $[X_{\min}, X_{max}]$ in the proposed division into sub-intervals of length h must in general be smaller than h, but this should present no problems in the improved bound we derive below. With such a construction, we can express the

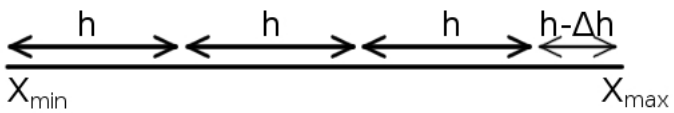

Figure 3.2: Construction of the intervals in the derivation of the M-ary EZZB.

probability in Eq. (3.7) as

$$\Pr\big(|\epsilon| < \frac{h}{2}\big) = \int_{X_{\min}}^{X_{max}} dx\, P_X(x) \Pr(|\hat{X}(Y) - x| \leq \frac{h}{2}) \tag{3.8}$$

$$= \sum_{n=0}^{M-1} \int_{X_{\min}}^{X_{\min}+h} dx\, P_X(x+nh) \Pr(|\hat{X}(Y) - x - nh| \leq \frac{h}{2}) + \int_{X_{\min}}^{X_{\min}+h-\Delta h} dx\, P_X(x+Mh) \Pr(|\hat{X}(Y) - x - Mh| \leq \frac{h}{2}), \tag{3.9}$$

$$= \int_{X_{\min}}^{X_{\min}+h} dx \sum_{n=0}^{M} P_X(x+nh) \Pr(|\hat{X}(Y) - x - nh| \leq \frac{h}{2}), \tag{3.10}$$

where, in arriving at Eq. (3.9), we used the substitution $x \to x + nh$ in the n^{th} sub-interval to reduce the range of the integral in each sub-interval to $(X_{\min}, X_{\min}+h)$. In the third equation, the last sub-interval is artificially extended to have the same size by noting that $P_X(x+Mh) = 0$ for $x > X_{max} - Mh$. The sum inside the outer integral above may be regarded as being proportional to a suboptimal probability of making the right decision in an $(M+1)$-ary hypothesis testing problem involving the hypotheses $\{x, x+h, \ldots x+Mh\}$, with a prior PD equal to $\{\frac{P_X(x)}{\mathcal{N}}, \frac{P_X(x+h)}{\mathcal{N}}, \ldots, \frac{P_X(x+Mh)}{\mathcal{N}}\}$ where

$$\mathcal{N} = \frac{1}{\sum_{n=0}^{M} P_X(x+nh)} \tag{3.11}$$

is a normalizing function. The quantity

$$\mathcal{N} \sum_{n=0}^{M} P_X(x+nh) \Pr(|\hat{X}(Y) - x - nh| \leq \frac{h}{2}) \tag{3.12}$$

may be regarded as the probability of a correct decision when the decision rule is based on the estimator, $\hat{X}(Y)$, falling closest to the point in whose favor the decision is made. This suboptimal decision rule based probability can be upper-bounded by the maximum probability of correct decision, $P_{max}(\{x+nh | n = 0, 1, \ldots, M\})$, obtained by using the MAP rule. Thus,

$$\Pr\big(|\epsilon| < \frac{h}{2}\big) \leq \int_{X_{\min}}^{X_{\min}+h} \frac{dx}{\mathcal{N}}\, P_{max}(\{x+nh | n = 0, 1, \ldots, M\}). \tag{3.13}$$

Therefore, from Eq. (3.7), we can write a new lower bound on the MSE as

$$\mathcal{E} \geq \frac{1}{2}\int_0^\infty dh\, h\left[1 - \int_{X_{\min}}^{X_{\min}+h} \frac{dx}{\mathcal{N}}\, P_{max}(\{x+nh|n=0,1,\ldots,M\})\right], \tag{3.14}$$

$$= \frac{1}{2}\int_0^\infty dh\, h\left[1 - \int_{X_{\min}}^{X_{\min}+h} dx \int dy \max\{P(x+nh,y)|n=0,1,\ldots,M\}\right], \tag{3.15}$$

where $M = 1 + [\frac{X_{max}-X_{\min}}{h}]$, with [.] representing the integer part of a real number, is the smallest integer larger than $\frac{X_{max}-X_{\min}}{h}$. The RHSs of inequalities (3.14) and (3.15) represent two equivalent forms of our proposed new lower bound for the MSE. We call it an improved Ziv-Zakai bound (IZZB). It is expected to be tighter, in general, than the standard EZZB since the maximum probability of a correct decision from amongst M hypotheses is expected to decrease – and thus the RHS of inequality (3.15) is expected to increase – with increasing M.

We can avoid calculating the probability of correct decision using a similar technique as given in Sec. 3.1. An approximate form of the improved ZZB (3.15) may be given by noting that the maximum of any M non-negative quantities, $a_1, \ldots, a_M$, is bounded above by the p-norm of the vector formed from these quantities,

$$\max(a_1,\ldots,a_M) \leq \|\mathbf{a}\|_p, \quad p \geq 0; \quad \mathbf{a} \equiv (a_1,\ldots,a_M), \tag{3.16}$$

with the above inequality becoming an equality in the limit $p \to \infty$. In view of this upper bound, we have from Eq. (3.15) the following weaker lower bound on MSE:

$$\mathcal{E} \geq \frac{1}{2}\int_0^\infty dh\, h\left[1 - \int_h dx \int dy\, \|\mathbf{\Pi}(x, x+h, \ldots, x+Mh; y)\|_p\right], \tag{3.17}$$

where $\mathbf{\Pi}$ stands for the vector $\Big(P(x,y),\ P(x+h,y),\ \ldots,\ P(x+Mh,y)\Big)$. This approximate form of the bound may provide a useful numerical approach to evaluate the improved ZZB by taking p to be sufficiently large compared to 1.

3.2.1 Relation to the standard ZZB

The standard EZZB (3.2) may also be expressed in a form involving the max function. Use of the identity,

$$\min(a,b) + \max(a,b) = a + b, \tag{3.18}$$

valid for any two real quantities a and b, in expression (3.2) implies the following alternate form for the ZZB:

$$\begin{aligned}\text{EZZB} =& \frac{1}{2}\int_0^\infty dh\, h \int\int dx\, dy \Big\{P(x,y) + P(x+h,y) \\ & \qquad - \max[P(x,y), P(x+h,y)]\Big\} \\ =& \frac{1}{2}\int_0^\infty dh\, h\Big\{2 - \int\int dx\, dy\ \max[P(x,y), P(x+h,y)]\Big\},\end{aligned} \tag{3.19}$$

where we used the normalization of the joint PD to arrive at the second form. Note that the form given in Eq. (3.19) is different from the IZZB when there are only two hypothesis. From Eq. (3.15), for a binary hypothesis testing problem,

$$\text{IZZB} = \frac{1}{2}\int_0^\infty dh\, h\Big(1 - \int_{X_{\min}}^{X_{\min}+h} dx \int dy \max\{P(x,y), P(x+h,y)\}\Big) \tag{3.20}$$

3.3 Discrete ZZB

The discrete parameter case is already considered in [42], but the Bayes risk considered there is somewhat different from the traditional MSE, and scalar, uniformly distributed parameters were considered. Here we derive the discrete EZZB for a scalar parameter with an arbitrary prior distribution. First of all, note that even though the parameter to be estimated is discrete, the estimate itself can be continuous. Let the discrete random variable, X, be defined on a discrete set with a uniform

spacing δx and let N_X be the number of values that X can take. The MSE for an estimator $\hat{X}(y)$ can be written as

$$\mathcal{E} = \mathbb{E}[(\hat{X}(\bar{y}) - X)^2] \tag{3.21}$$

$$= \sum_x P(x) \int_{\hat{X}} d\hat{X}\, P(\hat{X}|x)(\hat{X} - x)^2 \tag{3.22}$$

$$= \sum_x P(x) \int_{-\infty}^{\infty} dh\, P(x+h|x)h^2, \tag{3.23}$$

where $\hat{X} = X + h$. Let the inner integral be denoted by $\mathcal{E}_{|x}$, so that

$$\mathcal{E}_{|x} = \int_{-\infty}^{\infty} dh\, P(x+h|x)h^2 \tag{3.24}$$

$$= \int_0^{\infty} dh\, P(x+h|x)h^2 + \int_0^{\infty} dh\, P(x-h|x)h^2 \tag{3.25}$$

$$= 2\int_0^{\infty} dh\, h(\Pr(\hat{X} > x+h|x) + \Pr(\hat{X} < x-h|x)), \tag{3.26}$$

where the last equation is obtained by doing an integration by parts and Pr denotes the cumulative probability distribution, so that

$$\frac{d}{dh}\Pr(\hat{X} > x+h|x) = -P(x+h|x), \tag{3.27}$$

and

$$\frac{d}{dh}\Pr(\hat{X} < x-h|x) = -P(x-h|x). \tag{3.28}$$

The total error can then be written as:

$$\mathcal{E} = \frac{1}{2}\int_0^{\infty} dh\, h \sum_x P(x)(\Pr(\hat{X} > x + \frac{h}{2}|x) + \Pr(\hat{X} > x - \frac{h}{2})|x), \tag{3.29}$$

where the redefinition $h \to \frac{h}{2}$ is used. The derivation of the discrete EZZB has some elements of the continuous EZZB derivation [1]. We can see that Eq. (3.29) is the starting point of for the derivation of the continuous EZZB Eq. (2.13), where we can now define the outage probability for the discrete prior as

$$\Pr(|\epsilon| \geq \frac{h}{2}) = \sum_x P(x)(\Pr(\hat{X} > x + \frac{h}{2}|x) + \Pr(\hat{X} > x - \frac{h}{2})|x), \tag{3.30}$$

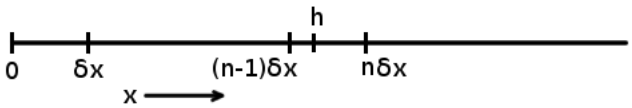

Figure 3.3: Relation between the continuous variable h and the discrete variable x.

where $\epsilon = \hat{X} - X$ is the error in estimating the discrete parameter X. This is also the point where we depart from the continuous EZZB derivation, since P_X is not defined at points $x + h$ as h is continuous. We will expand the integral of Eq. (3.29) in intervals of δx and aim to arrive at a lower bound on each of the individual integrals. For the n^{th} such integral:

$$I_n = \frac{1}{2}\int_{(n-1)\delta x}^{n\delta x} dh\, h \sum_x P(x)(\Pr(\hat{X} > x + \frac{h}{2}|x) + \Pr(\hat{X} < x - \frac{h}{2})|x). \tag{3.31}$$

Knowing that $h \in [(n-1)\delta x, n\,\delta x]$, as shown in Fig. 3.3, we have the following relations

$$x + \frac{h}{2} \leq x + \frac{n\delta x}{2} \tag{3.32}$$

$$x - \frac{n\delta x}{2} \leq x - \frac{h}{2}. \tag{3.33}$$

The individual probabilities in Eq. (3.31) can then be lower bounded as

$$\Pr(\hat{X} > x + \frac{h}{2}|x) = \Pr(\hat{X} > x + \frac{n\,\delta x}{2}|x) + \Pr(x + \frac{h}{2} < \hat{X} < x + \frac{n\,\delta x}{2}|x) \tag{3.34}$$

$$\geq \Pr(\hat{X} > x + \frac{n\,\delta x}{2}|x), \tag{3.35}$$

where relation (3.32) is used in the first equation, and

$$\Pr(\hat{X} < x - \frac{h}{2}|x) = \Pr(\hat{X} < x - \frac{n\,\delta x}{2}|x) + \Pr(x - \frac{n\delta x}{2} < \hat{X} < x - \frac{h}{2}|x) \tag{3.36}$$

$$\geq \Pr(\hat{X} < x - \frac{n\,\delta x}{2}|x), \tag{3.37}$$

where relation (3.33) is used in the first equation. Hence, I_n from Eq. (3.31) can be bounded below as:

$$I_n \geq \frac{1}{2}\int_{(n-1)\delta x}^{n\delta x} dh\, h\Big[\sum_x P(x)\Pr(\hat{X} > x + \frac{n\,\delta x}{2}|x) + \sum_x P(x)\Pr(\hat{X} < x - \frac{n\,\delta x}{2}|x)\Big] \tag{3.38}$$

Note that the inner summation is now independent of h and it can be expressed as $\Pr(|\epsilon| \geq \frac{n\,\delta x}{2})$. Carrying out the integration over h, we can express the relation (3.38) as

$$I_n \geq \frac{2n-1}{4}(\delta x)^2 \Pr(|\epsilon| \geq \frac{n\,\delta x}{2}), \tag{3.39}$$

where

$$\Pr(|\epsilon| \geq \frac{n\,\delta x}{2}) = \sum_x P(x)\Pr(\hat{X} > x + \frac{n\,\delta x}{2}|x) + \sum_x P(x)\Pr(\hat{X} < x - \frac{n\,\delta x}{2}|x). \tag{3.40}$$

A change of variable from $x \to x + n\,\delta x$ in the second sum of Eq. (3.40) yields

$$\Pr(|\epsilon| \geq \frac{n\,\delta x}{2}) = \sum_x \Big[P(x)\Pr(\hat{X} > x + \frac{n\,\delta x}{2}|x) + P(x + n\,\delta x)\Pr(\hat{X} < x + \frac{n\,\delta x}{2}|x + n\,\delta x)\Big] \tag{3.41}$$

Multiplying and dividing by $P(x) + P(x + n\,\delta x)$,

$$\Pr(|\epsilon| \geq \frac{n\,\delta x}{2}) = \sum_x [P(x) + P(x + n\,\delta x)] \Big[\frac{P(x)}{P(x) + P(x + n\,\delta x)}\Pr(\hat{X} > x + \frac{n\,\delta x}{2}|x) + \frac{P(x + n\,\delta x)}{P(x) + P(x + n\,\delta x)}\Pr(\hat{X} < x + \frac{n\,\delta x}{2}|x + n\,\delta x)\Big] \tag{3.42}$$

Now, consider the binary hypothesis testing problem:

$$\begin{aligned} &H_0 : X = x; && \Pr(H_0) = \frac{P(x)}{P(x) + P(x + n\,\delta x)}; && Y \sim P(Y|X = x) \\ &H_1 : X = x + n\,\delta x; && \Pr(H_1) = 1 - \Pr(H_0); && Y \sim P(Y|X = x + n\,\delta x) \end{aligned} \tag{3.43}$$

and the suboptimal decision scheme of estimating X and then a nearest neighbor decision being made:

$$\begin{aligned} &\text{Decide } \ H_0 : X = x; && \text{if } \ \hat{X}(Y) \leq x + \frac{n\,\delta x}{2} \\ &\text{Decide } \ H_1 : X = x + n\,\delta x; && \text{if } \ \hat{X}(Y) \geq x + \frac{n\,\delta x}{2} \end{aligned} \tag{3.44}$$

The probability of error for this suboptimal decision scheme is the term inside the summation of Eq. (3.42). This can be lower bounded by the MPE, $P_{\min}(x, x + n\,\delta x)$. Thus

$$\Pr(|\epsilon| \geq \frac{n\,\delta x}{2}) \geq \sum_x (P(x) + P(x + n\,\delta x)) P_{\min}(x, x + n\,\delta x), \tag{3.45}$$

and I_n from (3.39)can be bounded below as,

$$I_n \geq \frac{2n-1}{4} (\delta x)^2 \sum_x (P(x) + P(x + n\,\delta x)) P_{\min}(x, x + n\,\delta x), \tag{3.46}$$

from which we can write the discrete ZZB as:

$$\mathcal{E} \geq \text{DZZB} = \sum_{n=1}^{N_X - 1} \frac{2n-1}{4} (\delta x)^2 \sum_x (P(x) + P(x + n\,\delta x)) P_{\min}(x, x + n\,\delta x), \tag{3.47}$$

where the upper limit for the summation on n is introduced. Also note that the region for summation for x is over those values where $P(x)$ and $P(x + n\,\delta x)$ are non-zero. A slightly weaker bound can be achieved which is computationally simpler if we consider equally likely hypotheses along similar lines by Bell et. al [1]. Towards this end, we will start by lower bounding the RHS of Eq. (3.41),

$$\begin{aligned} \Pr(|\epsilon| \geq \frac{n\,\delta x}{2}) &\geq \sum_x \min\{P(x), P(x + n\,\delta x)\} \Big[\Pr(\hat{X} > x + \frac{n\,\delta x}{2} | x) \\ &\qquad + \Pr(\hat{X} < x + \frac{n\,\delta x}{2} | x + n\,\delta x) \Big] \qquad (3.48) \\ &= 2 \sum_x \min\{P(x), P(x + n\,\delta x)\} \Big[\frac{1}{2} \Pr(\hat{X} > x + \frac{n\,\delta x}{2} | x) \\ &\qquad + \frac{1}{2} \Pr(\hat{X} < x + \frac{n\,\delta x}{2} | x + n\,\delta x) \Big] \qquad (3.49) \end{aligned}$$

The expression in the square brackets in Eq. (3.49) can be recognized as a probability of error for the binary hypothesis testing problem given in (3.43) using the suboptimal decision rule specified in (3.44) with equal probability for the hypotheses H_0 and H_1. This can be lower bounded by the MPE for an equi-probable hypothesis testing problem, $P^{el}_{\min}(x, x + n\,\delta x)$. Thus,

$$\Pr(|\epsilon| \geq \frac{n\,\delta x}{2}) \geq 2 \sum_x \min(P(x), P(x + n\,\delta x)) P^{el}_{\min}(x, x + n\,\delta x), \tag{3.50}$$

from which we can lower bound the MSE as

$$\mathcal{E} \geq \text{DZZB}_{el} = \sum_{n=1}^{N_X - 1} \frac{2n-1}{2} (\delta x)^2 \sum_x \min(P(x), P(x+n\,\delta x)) P^{el}_{\min}(x, x+n\,\delta x). \tag{3.51}$$

We can also use the relation (3.3) presented in Sec. 3.1 to derive an approximation to the DZZB in Eq. (3.47),

$$\text{DZZB} = \sum_{n=1}^{N_X - 1} \frac{2n-1}{4} (\delta x)^2 \sum_x \sum_y \min\{P(x,y), P(x + n\,\delta x, y)\} \tag{3.52}$$

$$\geq \sum_{n=1}^{N_X - 1} \frac{2n-1}{4} (\delta x)^2 \sum_x \sum_y \frac{P(x,y) P(x + n\,\delta x, y)}{[P^n(x,y) + P^n(x + n\,\delta x, y)]^{\frac{1}{n}}}. \tag{3.53}$$

3.4 Summary

In this chapter, three new results for the Ziv-Zakai bounds are presented. The first one is a practical result for the EZZB through an approximation to the MPE which involves a free parameter. To evaluate the integrals in the AZZB an MCMC approach can be utilized and a free parameter can be made arbitrary large to get tight approximations. The advantage of this result is that there is no need to calculate the different regions in the binary hypothesis testing problem. The second one is an extension of the EZZB using an M-ary hypothesis formulation. Since the M-ary formulation, in general, is more comprehensive than the binary one, we expect this

extension to improve the EZZB. An approximation to the IZZB is also provided avoiding the explicit calculation of the MPE. The third result considers the case when the parameter takes only discrete values. These results form the basis for our paper [110] which is under preparation.

Chapter 4

Upper Bound on the MMSE

Most of the existing bounds on the MMSE are lower bounds, but only a few attempts have been made to find useful upper bounds on the MMSE. Seidman [68] derived an upper bound on the MMSE by bounding the corresponding probabilities of the squared error in an expansion of the MSE. Hawkes et al. [69] developed an upper bound based on the information theoretic Kullback-Leibler distance measure. A recent work by Flam et al. [70] used an optimal linear estimator to derive an upper bound on the MSE, which performs well for the Gaussian-mixture model. In this chapter a new upper bound (UB) on the MMSE is derived. The derivation uses Jensen's inequality, and is valid for both scalar and vector parameters. We will first derive the result for scalar parameters and extend it to vector parameters in Sec. 4.2.

4.1 Derivation of the upper bound

Following the same notation given in Chapter 2, let $\hat{X}(Y)$ be an estimator of a Bayesian parameter X, $\mathcal{E}$ its MSE, and $\hat{X}_M(Y)$ and $\mathcal{E}_M$, the MMSEE and the MMSE, respectively. In the following derivation, we will denote $\hat{X}(Y)$ by just $\hat{X}$ to enhance

readability.

By writing the estimation error as $\hat{X} - X = (\hat{X} - \hat{X}_M) + (\hat{X}_M - X)$ and then taking its mean squared value (2.2), we may express its MSE in the expanded form

$$\begin{aligned}\mathcal{E} =& \mathcal{E}_M + \mathbb{E}[(\hat{X} - \hat{X}_M)^2] + 2\mathbb{E}[(\hat{X} - \hat{X}_M)(\hat{X}_M - X)] \\ =& \mathcal{E}_M + \mathbb{E}[(\hat{X} - \hat{X}_M)^2], \end{aligned} \tag{4.1}$$

where the first equality follows from Eq. (2.6) and the second equality from the fact that the last term in the first equality vanishes identically, since

$$\mathbb{E}[(\hat{X} - \hat{X}_M)(\hat{X}_M - X)] = \mathbb{E}_Y\big[(\hat{X} - \hat{X}_M)\mathbb{E}_{X|Y}[\hat{X}_M - X]\big], \tag{4.2}$$

$$= \mathbb{E}_Y\big[(\hat{X} - \hat{X}_M)(\hat{X}_M - \mathbb{E}_{X|Y}[X])\big], \tag{4.3}$$

$$= 0, \tag{4.4}$$

where we used the relation that the MMSEE is the mean of the posterior, (2.4), in going from Eq. (4.3) to Eq. (4.4).

The second term in the second equality in (4.1) is the amount by which the MSE for an estimator must exceed the MMSE. By subtracting this term from both sides of the equality, we note that this term cannot be smaller than the square of the mean $\mathbb{E}[\hat{X} - \hat{X}_M]$, as follows immediately from applying the Jensen's inequality to the function $f(\eta) = \eta^2$. We have thus the following UB on the MMSE:

$$\begin{aligned}\mathcal{E}_M \leq& \mathcal{E} - \mathbb{E}^2[\hat{X} - \hat{X}_M] \\ =& \mathbb{E}[(\hat{X} - X)^2] - \mathbb{E}^2[\hat{X} - X] \qquad (4.5) \\ =& \mathbb{E}\left[\left(\delta\hat{X} - \delta X\right)^2\right] \stackrel{\text{def}}{=} \mathcal{E}_{UB}, \qquad (4.6)\end{aligned}$$

where the second line follows from the fact that the mean value of the MMSEE is simply the mean value of the prior, $\mathbb{E}_X[X]$. The third line represents a convenient way of combining the two terms in the second line and then regrouping them inside

the square. The symbol δ preceding a variable denotes its deviation from its mean value, e.g.,

$$\delta X \equiv X - \mathbb{E}[X]. \tag{4.7}$$

The rather simple upper bound (4.6) has the immediate benefit that it can be computed readily since, unlike the MMSE, it does not require any knowledge of the posterior PD, $P_{X|Y}$, the chief bane of any MMSE calculation. Furthermore, since according to Eq. (4.1) the MSE of any estimator differs from the MMSE in the quadratic order in the deviation of an estimator from the MMSE estimator, a variational approach to estimate the MMSE by minimizing the UB (4.6) with respect to classes of estimators should yield excellent accuracy even when the estimator itself does not approximate the MMSEE as well. Note further that being smaller than the MSE, which is the first term in (4.5), the UB obtained from any trial estimator is already closer to the MMSE than its own MSE, and thus potentially furnishes a better starting point and a faster convergence to the MMSE. This difference $\mathbb{E}^2[\hat{X} - X]$ in (4.5) is the squared Bayesian bias for an estimator and by its subtraction from its MSE yields the upper bound.

4.1.1 Understanding bias

In the high SNR region, it is well known that the CRB is tight and the MLE estimator achieves the CRB asymptotically. In this region, estimators can be constructed that are unbiased and still reach the MMSE. But, as the SNR decreases introducing bias has been observed to be more advantageous in decreasing the MSE [29,30,33,36,111]. In this section, we will briefly present this theory which can provide valuable insights into developing iterative methods to find the best upper bound on the MMSE.

Note that the definition of bias in a deterministic setting is different from that in the Bayesian setting. In a deterministic setting, the parameter x is assumed to be

non-random and unknown. Bias, $b(\hat{X}, x)$, of an estimator $\hat{X}$ is defined as:

$$b(\hat{X}, x) = \mathbb{E}[\hat{X}|x] - x, \tag{4.8}$$

where no prior distribution, $P_X(x)$, on the parameter is assumed and the expectation is taken with $P(y|x)$. This difference from the Bayesian setting leads to a very useful relation involving the MSE, the bias and the variance of the estimator. We can express the MSE as,

$$\begin{aligned}
\mathcal{E} = \mathbb{E}[(\hat{X} - x)^2|x] &= \mathbb{E}[\big((\hat{X} - \mathbb{E}[\hat{X}|x]) + (\mathbb{E}[\hat{X}|x] - x)\big)^2|x] && (4.9)\\
&= \mathbb{E}[(\hat{X} - \mathbb{E}[\hat{X}|x])^2|x] + \mathbb{E}[b^2(\hat{X}, x)|x] \\
&\quad + 2\mathbb{E}[(\hat{X} - \mathbb{E}[\hat{X}|x])b(\hat{X}, x)|x] && (4.10)\\
&= \mathbb{E}[(\hat{X} - \mathbb{E}[\hat{X}|x])^2|x] + b^2(\hat{X}, x) && (4.11)
\end{aligned}$$

The cross term in Eq. (4.10) is zero because the bias, $b(\hat{X}, x)$, is independent of y and removed from the expectation resulting in the mean linear deviation, $\mathbb{E}[(\hat{X} - \mathbb{E}[\hat{X}|x])|x]$, which is zero. Thus, the MSE of an estimator in a deterministic setting is equal to the sum of its squared bias and variance.

Since in the deterministic approach the parameter x is assumed to be non-random, trivial estimators such as $\hat{X}(Y) = x$ can be formulated to yield zero MSE. However, the parameter x is assumed to be unknown which makes the trivial estimator useless on an average. Constraints such as unbiasedness, $b(\hat{X}(Y), x) = 0$, are imposed in the classical approach. This amounts to reducing the variance of the estimator from Eq. (4.11). The resulting optimal estimators are called minimum variance unbiased estimators (MVUE) and were studied extensively [112]. But, reducing the MSE is, in general, of greater importance in an estimation problem than reducing just the variance of an estimator. We will consider a particular example of "shrinkage estimator" which introduces bias linearly to an unbiased estimator so that the overall MSE is reduced. More details regarding the following presentation can be found in [36].

Construct a biased estimator, $\hat{X}_b(Y)$, as

$$\hat{X}_b(Y) = (1+m)\hat{X}_u(Y), \tag{4.12}$$

where $\hat{X}_u(Y)$ is the MVUE with a known variance, σ_u^2 and m is a free parameter which will be optimized to reduce the MSE of the biased estimator. From Eq. (4.11), we can express the MSE of the biased estimator as the sum of its variance and squared bias,

$$\mathcal{E}_b = \sigma_b^2 + (\mathbb{E}[\hat{X}_b(Y)|x] - x)^2. \tag{4.13}$$

Since $\mathbb{E}[\hat{X}_u(Y)|x] = x$, using Eq. (4.12), we can express $\mathcal{E}_b$ as

$$\mathcal{E}_b = (1+m)^2\sigma_u^2 + m^2x^2. \tag{4.14}$$

Since the MSE of the unbiased estimator is its variance, $\mathcal{E}_u = \sigma_u^2$, by optimizing over m, it is possible to reduce the overall MSE, $\mathcal{E}_b$ further. This amounts to reducing the variance of the estimator more than the increase in the squared bias. The optimal value of m can be found using a minimax approach as suggested in [111]:

$$m_{opt} = \arg\min_m \max_x (\mathcal{E}_b - \mathcal{E}_u), \tag{4.15}$$

where m_{opt} is chosen so that $\mathcal{E}_b \leq \mathcal{E}_u$ for all $|x| \leq x_0$, and that the MSE is reduced maximally. This approach presents a way of introducing bias into the estimator such that the overall MSE can be reduced uniformly.

In this dissertation, we are considering a Bayesian bias, where the expectations are with respect to both the prior and the conditional distribution. As a consequence, a simple relation between the MSE, the variance and the bias of a Bayesian estimator is not possible, unlike Eq. (4.11) for the deterministic estimator. However, the minimax structure presents a way of calculating the upper bound iteratively. We reserve this as a future direction where a non-linear bias, one that is possibly dependent on the

parameter, can be introduced in each iteration and the upper bound minimized in each step resulting in a tighter bound on the MMSE.

Even without the iterative scheme, we observe that when the SNR decreases, unbiasedness is no longer guaranteed, and to reduce the error, a nonzero squared Bayesian bias must be subtracted out in general, as in (4.5). As we shall see in sections 4.3.1 and 4.3.2, it is possible, however to construct unbiased estimators, even in the low SNR region that can reach the MMSE. Knowledge about the channel and the prior in a particular problem will help us in determining an appropriate test estimator so that the corresponding upper bound is as small as possible. In general, the MMSE is a non-linear, highly complicated function of the data, and a test estimator that carries similar complexity will yield a tighter upper bound. In this regard, a test estimator which is linear in the MAP estimator, for example, will have the needed complexity, at least at moderate to large values of the SNR, and the calculated upper bound is expected to be much closer to the MMSE than the upper bound resulting from even the most general linear estimator that can not track a non-linear parameter even at very high SNR.

4.2 Multi-parameter generalization of the MMSE and upper bound

As is well known, the form (2.4) of the MMSEE remains valid for the most general definition of the MSE for multivariate parameter and data vectors,

$$\mathcal{E} \stackrel{\text{def}}{=} \mathbb{E}[(\bar{\hat{X}} - \bar{X})^T \mathbf{R} \, (\bar{\hat{X}} - \bar{X})], \tag{4.16}$$

where $\mathbf{R}$ is a positive-definite, symmetric, real matrix. As for a scalar parameter, the MMSEE is that which minimizes (4.16). By a similar calculation as that performed

to arrive at (2.4), we may easily show that the MMSEE for each parameter,

$$\hat{X}_{nM}(\bar{y}) = \mathbb{E}_{X|Y}[X_n] = \int x_n \, P(x_n|\bar{y}) \, dx_n, \quad n = 1, \ldots, N_x, \tag{4.17}$$

where N_x is the total number of parameters being estimated. The MMSE can be written as

$$\mathcal{E}_M = \mathbb{E}[(\bar{\hat{X}}_M - \bar{X})^T \mathbf{R} \, (\bar{\hat{X}}_M - \bar{X})]. \tag{4.18}$$

Since $\mathbf{R}$ is a positive-definite symmetric matrix, there exist a unique positive-definite symmetric matrix $\mathbf{Q}$ such that $\mathbf{R} = \mathbf{Q}^2$. We can now derive the upper bound for multivariate parameter by first transforming the MSE (4.16) into a sum of the MSEs of scalar parameters,

$$\begin{aligned}
\mathcal{E} &= \mathbb{E}[(\bar{\hat{X}} - \bar{X})^T \mathbf{R} \, (\bar{\hat{X}} - \bar{X})] && (4.19)\\
&= \mathbb{E}[(\bar{\hat{X}} - \bar{X})^T \mathbf{Q}\mathbf{Q} \, (\bar{\hat{X}} - \bar{X})] && (4.20)\\
&= \mathbb{E}[(\bar{\hat{Z}} - \bar{Z})^T (\bar{\hat{Z}} - \bar{Z})] && (4.21)\\
&= \sum_{n=1}^{N_x} \mathbb{E}[(\hat{Z}_n - Z_n)^2], && (4.22)
\end{aligned}$$

where

$$\bar{Z} = \mathbf{Q}\bar{X}, \text{ and } \bar{\hat{Z}} = \mathbf{Q}\bar{\hat{X}}. \tag{4.23}$$

We used $\mathbf{Q} = \mathbf{Q}^T$, since it is a symmetric matrix, in going from (4.20) to (4.21). From the upper bound relation for scalar parameters (4.5), we have

$$\mathbb{E}[(\hat{Z}_{nM} - Z_n)^2] \leq \mathbb{E}[(\delta\hat{Z}_n - \delta Z_n)^2], \tag{4.24}$$

where

$$\bar{\hat{Z}}_M = \mathbf{Q}\bar{\hat{X}}_M. \tag{4.25}$$

By summing over n on both sides of the inequality (4.24), and doing an inverse transformation using relations (4.23) and (4.25), we arrive at the upper bound

$$\mathcal{E}_M \leq \mathcal{E}_{UB},$$

where

$$\mathcal{E}_{UB} = \mathbb{E}[(\delta\bar{\hat{X}} - \delta\bar{X})^T \mathbf{R} (\delta\bar{\hat{X}} - \delta\bar{X})]. \tag{4.26}$$

We will now apply the upper bound formulation to the polynomial class of estimators as well as the MAP estimator.

4.3 Test estimators

4.3.1 The polynomial class of estimators

Consider first the simple problem of a single parameter being estimated from a single observation and trial estimators belonging to the class of Nth-order polynomials, i.e.,

$$\hat{X}_N(y) = a_0 + \sum_{n=1}^{N} a_n y^n. \tag{4.27}$$

We shall regard $\{a_n\}$ as variational parameters that need to be chosen to minimize the UB (4.6). Substituting the form (4.27) into the formula (4.6), we may express the UB as

$$\begin{aligned} \mathcal{E}_{UB} =& \mathbb{E}\Big[\Big(\sum_{n=1}^{N} a_n \delta y^n - \delta X\Big)^2\Big] \\ =& \bar{a}^T \mathbf{M} \bar{a} - 2\bar{a}^T \bar{v} + \sigma_X^2, \end{aligned} \tag{4.28}$$

where $\bar{a}$ denotes the vector $(a_1, \ldots, a_N)^T$, $\mathbf{M}$ an $N \times N$ symmetric matrix with elements, $M_{mn} = \mathbb{E}[\delta y^m \delta y^n]$, $m, n = 1, \ldots, N$, and $\bar{v}$ an $N \times 1$ vector of elements $\mathbb{E}[\delta X\, \delta y^n]$, $n = 1, \ldots, N$.

Since the matrix $\mathbf{M}$ is positive definite, the quadratic problem (4.28) has a single extremum that is its absolute minimum. Its location $\bar{a}_*$ in the space of the parameter vector $\bar{a}$ is determined by setting the gradient of the expression (4.28) with respect

to the parameter vector $\bar{a}$ equal to 0, namely $\mathbf{M}\bar{a}_* = \bar{v}$, or equivalently $\bar{a}_* = \mathbf{M}^{-1}\bar{v}$. The corresponding minimum value of the UB (4.28), which we call P^N-VB, can then be simplified as

$$\mathrm{P}^N\text{-VB} = \sigma_X^2 + (\mathbf{M}^{-1}\bar{v})^T\mathbf{M}(\mathbf{M}^{-1}\bar{v}) - 2(\mathbf{M}^{-1}\bar{v})^T\bar{v}, \tag{4.29}$$

$$= \sigma_X^2 - \bar{v}^T\mathbf{M}^{-1}\bar{v}, \tag{4.30}$$

where in arriving at the final equation we used the fact that $\mathbf{M}^{-1}$ is symmetric and $\mathbf{M}\mathbf{M}^{-1}$ is the identity matrix.

The minimum value of the UB (4.30) is both smaller than the prior variance, σ_X^2, and lowered in general as the order of the polynomial, N, increases. To prove the latter assertion, we simply note that the class of $(N+1)$th-order polynomials includes, as a proper subset, the class of Nth order polynomials. Thus the minimum of the UB over the former class cannot be larger than the that over the latter class. A more rigorous mathematical proof may also be given based on Schur's inversion formula for the block-matrix form of a square matrix.

Curiously, the overall additive constant a_0 in the trial form (4.27) of the estimator is left undetermined. This is not a surprise since the UB form (4.28) is clearly insensitive to any overall additive constants. Yet, such a constant is in general included in the actual form of the MMSEE, $\hat{X}_M$, e.g., for a Gaussian channel and a Gaussian prior, $\hat{X}_M$ is a weighted sum of the prior mean, which is a constant, and the data. We can thus estimate the form of the MMSEE only up to an arbitrary additive constant even when the UB approximates the MMSE very well. To fix this constant, we need an additional constraint of unbiasedness obeyed by the MMSEE, namely that its mean be the same as the prior mean.

4.3.2 The MAP estimator

As we saw in Sec. 2.3, the MAP estimator must be modified as it does not reach the prior variance at low SNR values. We will now apply the upper bound variationally to a MAP estimator. A linear extension of the MAP estimator is formulated as the following test estimator for the upper bound:

$$\hat{X}(y) = \alpha \hat{X}_{\mathrm{MAP}}(y) + \beta, \tag{4.31}$$

where α and β are the variational parameters that need to be chosen to minimize the upper bound. Substituting the form (4.31) into the formula (4.5) we may express the UB as

$$\mathcal{E}_{UB} = \mathbb{E}\left[\left(\alpha \delta \hat{X}_{\mathrm{MAP}}(y) - \delta X\right)^2\right] \tag{4.32}$$

and a simple differentiation of the Eq. (4.32) with respect to α and setting it to zero yields the optimal as

$$\alpha = \frac{\mathbb{E}[\delta \hat{X}_{\mathrm{MAP}} \delta X]}{\mathbb{E}[\delta \hat{X}^2_{\mathrm{MAP}}]}. \tag{4.33}$$

The second derivative at this optimum can easily be checked to be positive so that the minimum value of the upper bound, the variational bound involving the MAP estimator (MAP-VB), becomes

$$\text{MAP-VB} = \sigma_X^2 - \frac{\mathbb{E}^2\left[\delta \hat{X}_{\mathrm{MAP}} \delta X\right]}{\mathbb{E}\left[\delta \hat{X}^2_{\mathrm{MAP}}\right]}. \tag{4.34}$$

From Eq. (4.34) then, when SNR is low independent of the prior PD, $\hat{X}_{\mathrm{MAP}}$ becomes independent of X, and the expectation in the numerator of the subtraction term, $\mathbb{E}^2\left[\delta \hat{X}_{\mathrm{MAP}} \delta X\right] = \mathbb{E}^2\left[\delta \hat{X}_{\mathrm{MAP}}\right] \mathbb{E}^2\left[\delta X\right] = 0$, resulting in MAP-VB $= \sigma_X^2$. When SNR is high, $\hat{X}_{\mathrm{MAP}} \to X \Rightarrow \mathbb{E}^2\left[\delta \hat{X}_{\mathrm{MAP}} \delta X\right] \to \mathbb{E}^2\left[\delta X^2\right] (= \sigma_X^2)$, which results in MAP-VB $\to 0$. By using a variational upper bound of the form (4.34), both ends of

the error curve are made to follow the MMSE tightly. This further constrains the error, in all the SNR regions, to be as close as possible to the MMSE, as shown in the next section with simple illustrative examples. Note that the constant additive parameter β, as before, does not have any effect in the value of the minimum upper bound. By noting that the upper bound reduces the mean square error of the estimator by removing the bias, if we chose β so that the estimator becomes unbiased, we have a way to construct an estimator so that its MSE is the closest to the MMSE.

We will now calculate the upper bound for linear and cubic estimators, and also the variational MAP upper bound for the problems already considered in Sec. 2.6.

4.4 Results

4.4.1 Poisson channel and exponential prior - single pixel case

As we saw in Sec. 2.6.1, neither the MAP estimation error, not the ZZB and the WWB are adequate in approximating the MMSE. The MAP estimator, especially, is not able to adapt to the highly biased problem and performs very poorly for all values of SNR. In this section, we will apply the upper bound variationally on the MAP estimator, Eq. (4.34) and also calculate the upper bound for polynomial class of estimators, Eq. (4.30). We will see a dramatic improvement in the performance of the MAP estimator as it is capable of addressing bias and we will further notice that a simple cubic estimator is able to follow the MMSE nearly perfectly for all values of SNR.

For the case of a Poisson channel and exponential prior, with a and b being the gain and bias of the single-pixel detector, the MAP estimator, as we noted in

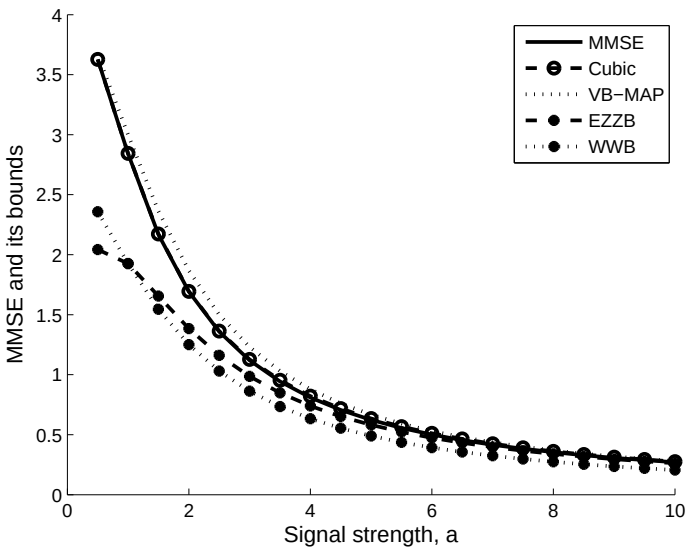

Figure 4.1: MMSE, MAP-VB, P^3-VB, EZZB and WWB for the Poisson channel, with bias, b = 10 and exponential prior with mean $\langle X \rangle = 2$.

Sec. 2.6.1, is the following linear estimator,

$$\hat{X}_{\text{MAP}} = \frac{\langle X \rangle}{a \langle X \rangle + 1} y - \frac{b}{a}, \tag{4.35}$$

where the coefficients are not optimized. By variationally calculating the UB for this estimator we are optimizing the linear coefficients, there by attaining the best linear MMSE estimator. An equivalent equation to Eq. (4.34) is

$$\text{MAP-VB} = \sigma_X^2 - \frac{\mathbb{E}[\hat{X}_{\text{MAP}} X] - \mathbb{E}[\hat{X}_{\text{MAP}}]\mathbb{E}[X]}{E[\hat{X}_{\text{MAP}}^2] - E^2[\hat{X}_{\text{MAP}}]}. \tag{4.36}$$

Calculating the MAP-VB, would involve in calculating the various expectations in Eq. (4.36):

$$\langle \hat{X}_{\text{MAP}} X \rangle - \langle \hat{X}_{\text{MAP}} \rangle \langle X \rangle = \frac{\langle XY \rangle - \langle X \rangle \langle Y \rangle}{a + \frac{1}{\langle X \rangle}} \tag{4.37}$$

$$\langle \hat{X}_{\text{MAP}}^2 \rangle - \langle \hat{X}_{\text{MAP}} \rangle^2 = \frac{\langle Y^2 \rangle - \langle Y \rangle^2}{(a + \frac{1}{\langle X \rangle})^2} \tag{4.38}$$

Thus the final expression for MAP-VB can be written as

$$\text{MAP-VB} = \sigma_X^2 - \frac{\big(\langle XY \rangle - \langle X \rangle \langle Y \rangle\big)^2}{\langle Y^2 \rangle - \langle Y \rangle^2}. \tag{4.39}$$

The result is plotted in Fig. 4.1 and we can see that the MAP-VB estimator brings down the pure-MAP-estimator MSE dramatically at low SNR, and the estimation error bound follows the MMSE rather tightly.

We also calculated the UB for a cubic estimator :

$$\hat{X}_c = a_0 + a_1 y + a_2 y^2 + a_3 y^3, \tag{4.40}$$

where the UB is minimized over the parameters $\{a_1, a_2, a_3\}$. The resulting variational upper bound is P^3-VB, obtained by setting $n = 3$ in Eq. (4.30). As we can see from Fig. 4.1, P^3-VB follows the MMSE essentially indistinguishably in all regions of operation.

This simple toy problem shows the power of the variational calculation. A cubic estimator itself may not be sufficient to approximate the MMSEE, Eq. (2.45), yet its MSE being of quadratic order approximates the MMSE tightly.

Encouraged by the performance of the UB for the single pixel case, we considered a two-pixel problem which we discuss in the following section.

4.4.2 Poisson channel and exponential prior - two pixel case

As a simple extension to the single pixel case, we consider a two pixel detector to resolve two near-by point sources, S_1 and S_2, as shown in Fig. 4.2. We assume that the photon counts at the two sensor pixels are independently captured, given the two source intensities. We further assume a multiplicative gain and a multiplicative cross-talk with an additive bias, representing the background intensity, expressed by the following equations:

$$P(y_1, y_2|x_1, x_2) = P(y_1|x_1, x_2)P(y_2|x_1, x_2), \tag{4.41}$$

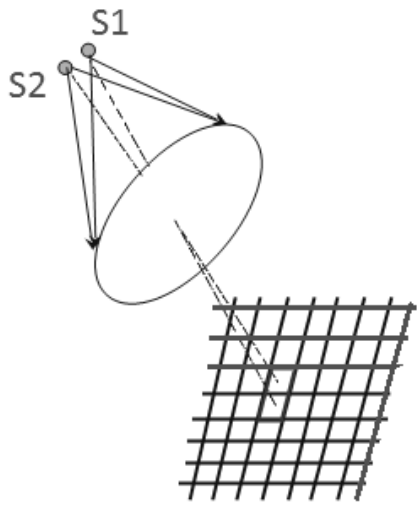

Figure 4.2: Two-pixel detector being used to resolve the two near-by point sources.

$$P(y_1|x_1,x_2) = \frac{(\alpha x_1 + \beta x_2 + \gamma)^{y_1}}{y_1!}\exp[-(\alpha x_1 + \beta x_2 + \gamma)], \quad y_1 = 0,1,\ldots, \tag{4.42}$$

$$P(y_2|x_1,x_2) = \frac{(\beta x_1 + \alpha x_2 + \gamma)^{y_2}}{y_2!}\exp[-(\beta x_1 + \alpha x_2 + \gamma)], \quad y_2 = 0,1,\ldots., \tag{4.43}$$

where α is the on-axis signal strength, β is the cross-talk signal strength, and γ is a uniform bias. We assume X_1 and X_2 are independent and exponentially distributed random variables.

Since the polynomial class of estimators performed well for the single-pixel case, we calculated the UB for a linear and quadratic test estimators. The MMSE and the bounds are evaluated as a function of the signal strength α keeping the overall intensity received at a pixel $\alpha + \beta$ fixed. As we can see from figures 4.3 and 4.4 the quadratic estimator bounds the MMSE nearly perfectly, being indistinguishable from the MMSE over the entire range of gain values. Exact evaluations of the MMSE and the various terms in the UB are presented in Appendix A.2.

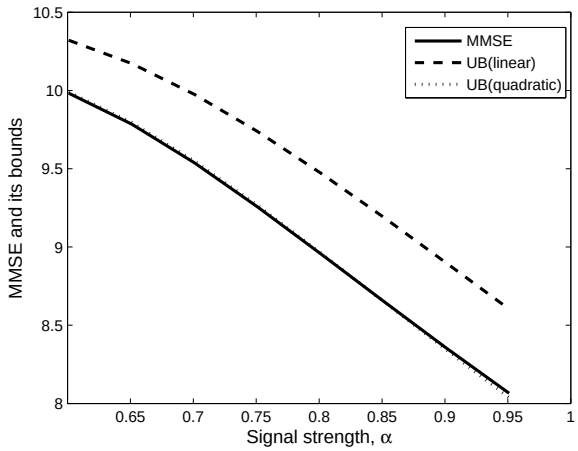

Figure 4.3: UB with a quadratic estimator. $\alpha + \beta = 1$, $\gamma = 10$, $\langle X_1 \rangle = 2$, $\langle X_2 \rangle = 3$.

4.4.3 TDE problem

In Sec. 2.6.2 we noted that the MAP estimation error, the ZZB and the WWB in the transition region depart from the MMSE. In this section, we will calculate the

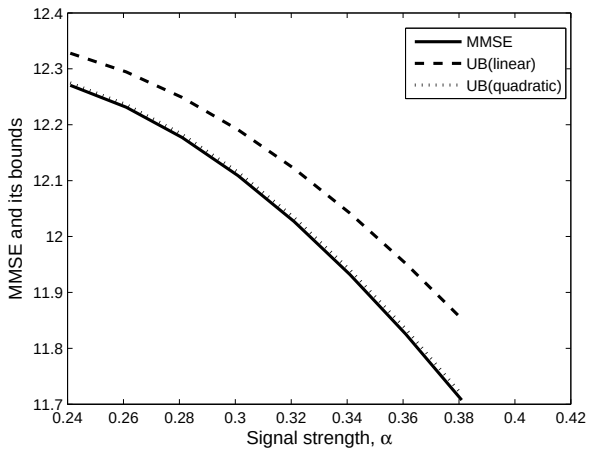

Figure 4.4: UB with quadratic estimator. $\alpha + \beta = 0.4$, $\gamma = 10$, $\langle X_1 \rangle = 2$, $\langle X_2 \rangle = 3$.

MAP-VB as well as the upper bound with linear estimator, and compare them with the existing bounds.

The MAP estimator can be estimated from Eq. (2.64):

$$\hat{\tau}_{\mathrm{MAP}}(\bar{y}) = \underset{\tau \in \mathcal{R}_\tau}{\operatorname{argmax}} \sum_{i=1}^{N_t} y_i s(t_i - \tau) \tag{4.44}$$

Once we evaluate the MAP estimator, we can proceed to find the upper bound, by assuming a test estimator that is linear in the MAP estimator, as in Eq. (4.31). The upper bound is given by Eq. (4.34). We numerically simulated the MAP estimator to calculate this upper bound. The result is plotted in Fig. 4.5, from which we can see that the MAP-VB hugs the MMSE tightly even in the transition region.

We next consider a linear estimator,

$$\hat{\tau}_l(\bar{y}) = \bar{a}'\bar{y} + b, \tag{4.45}$$

where the UB is optimized over the variational parameters $\bar{a}, b$. The resulting UB, P^1-VB from Eq. (4.30), is plotted in Fig. 4.5. Since the parameter τ depends non-linearly on the data for this problem, the error from a linear, or in general a polynomial, estimator is insufficient to approximate the MMSE, particularly at high SNR where it fails to track the nonlinear parameter accurately. A further improvement on the linear estimator is needed in such cases. A particular approach that avoids expensive calculations of the MAP estimator, and the MAP-VB based on it, is discussed in detail in the next section.

4.5 Piecewise quasi-linear estimator and associated upper bound

Even though the polynomial class of test estimators bound the MMSE tightly in certain cases, they do not have the required complexity to approximate the MMSE,

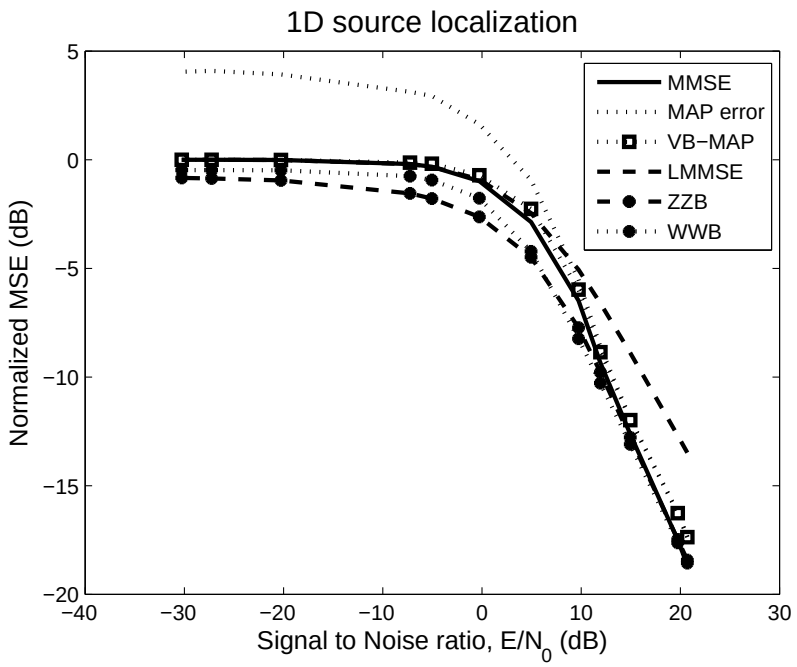

Figure 4.5: MMSE, MAP-VB, linear MMSE, EZZB and WWB for the TDE problem.

whenever the parameter depends non-linearly on the signal. This non-linearity in the parameters renders linear and polynomial estimation inadequate to approximate the MMSE. In this section, we will formulate a piecewise quasi-linear estimator designed for a Gaussian channel, which will overcome this drawback.

The prior interval, $[\tau_l, \tau_u]$, is divided into I intervals, as shown in Fig. 4.6 and in each of the sub-intervals, the estimator is approximated to be linear about the mid-point of each interval, labeled as $\tau_\mu, \mu = 1, 2, 3 \ldots, I$. The collection of these linear estimators are then combined using a Gaussian weighting function, $e_\mu(\bar{y})$, which has different width parameters for different sub-intervals, which we label as $w_\mu, \mu = 1, 2, 3 \ldots, I$, that need to be optimized. In formulating the estimator in this way, the chief advantage we gain is that it is simple enough that the optimization with respect to the various linear parameters becomes easy and can be carried out analytically. At the same time sufficient complexity is brought in by the weighting

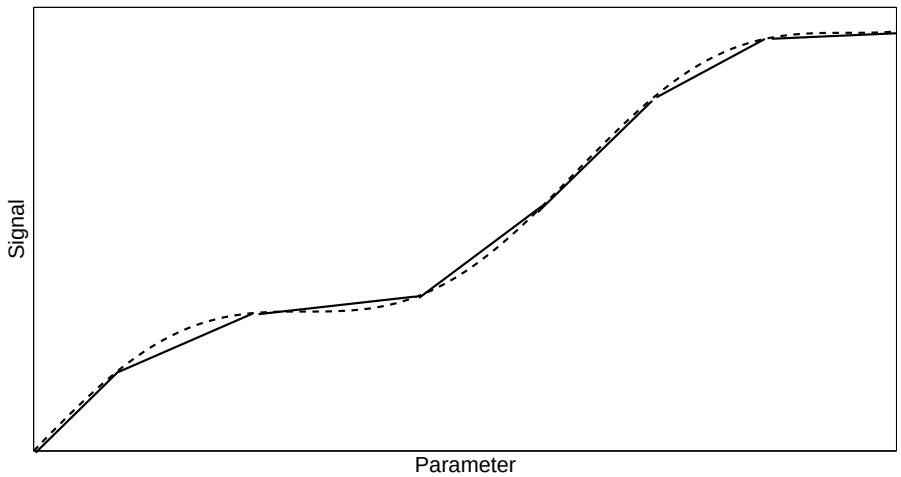

Figure 4.6: Non-linear dependency (dashed curve) of the signal on the prior parameters and a piecewise linear approximation (solid line segments).

functions which allow for the adaptation of the linear parameters of the estimator from one interval to the next. Literally, the linear estimator is allowed to evolve and adjust from one sub-interval to the next in the space of the parameter being estimated. It is for this reason that such a piecewise linear estimator performs well in the high SNR region as well.

The piecewise qausi-linear estimator (PQE) is given by:

$$\begin{aligned}\hat{\tau}(y) =& \sum_{\mu=1}^{I}(b_\mu + \bar{y}'\bar{A}_\mu)e_\mu \\ =& \bar{b}'\bar{e} + \bar{y}'\mathbf{A}\,\bar{e},\end{aligned} \tag{4.46}$$

where $\mathbf{A}$ is an $N_t \times I$ matrix ($\bar{A}_\mu$ is the corresponding $N_t \times 1$ column vector), $\bar{b}$ is an I element vector of coefficients, and $\bar{e}$ is a vector of Gaussian weighting functions, with widths $\bar{w}$:

$$e_\mu(\bar{y}) = \exp\left\{-\frac{1}{2w_\mu^2}\sum_{i=1}^{N_t}\Big[y(t_i) - s(t_i - \tau_\mu)\Big]^2\right\}. \tag{4.47}$$

To understand heuristically the basic idea behind such a formulation, consider a very high value for the SNR for which a good estimator is expected to track the value of the prior parameter precisely. In the PQE formulation, the weighting function (4.47) is significant only in the sub-interval containing the true value of the parameter, and becomes exponentially small for all the other intervals. In this way, the weighting function "picks" that interval among the I sub-intervals that contains the true value of the prior parameter. Since the linear variational parameters $\bar{A}_i$ and b_i are optimized over the respective sub-intervals, we are assured that the overall PQE tracks the value of the prior parameter optimally in all sub-intervals.

To calculate the upper bound, first substitute the form (4.46) for the test estimator into the upper bound Eq. (4.5), and then set its gradient with respect to the variational parameters $\mathbf{A}$ and $\bar{b}$ to zero. This can be shown to yield the following set of equations:

$$\begin{aligned} \mathbb{E}[(\hat{\tau}-\tau)\frac{\partial\hat{\tau}}{\partial\bar{b}}] &= (\mathbb{E}[\hat{\tau}]-\mathbb{E}[\tau])\frac{\partial\mathbb{E}[\hat{\tau}]}{\partial\bar{b}} \\ \mathbb{E}[(\hat{\tau}-\tau)\frac{\partial\hat{\tau}}{\partial\mathbf{A}}] &= (\mathbb{E}[\hat{\tau}]-\mathbb{E}[\tau])\frac{\partial\mathbb{E}[\hat{\tau}]}{\partial\mathbf{A}} \end{aligned} \tag{4.48}$$

From Eq. (4.46), we can calculate the various partial derivatives as

$$\frac{\partial\hat{\tau}}{\partial\bar{b}} = \bar{e} \tag{4.49}$$

$$\frac{\partial\mathbb{E}[\hat{\tau}]}{\partial\bar{b}} = \mathbb{E}[\bar{e}] \tag{4.50}$$

$$\frac{\partial\hat{\tau}}{\partial\mathbf{A}} = \bar{y}\bar{e}' \tag{4.51}$$

$$\frac{\partial\mathbb{E}[\hat{\tau}]}{\partial\mathbf{A}} = \mathbb{E}[\bar{y}\bar{e}'], \tag{4.52}$$

so that we can rewrite the set of equations (4.48) more explicitly as

$$\begin{aligned} \{\mathbb{E}[\bar{e}\bar{e}'] - \mathbb{E}[\bar{e}]\mathbb{E}[\bar{e}']\}\bar{b} + \mathbb{E}[\bar{e}\bar{y}'\mathbf{A}\bar{e}] - \mathbb{E}[\bar{e}]\mathbb{E}[\bar{y}'\mathbf{A}\bar{e}] &= \mathbb{E}[\tau\bar{e}] - \mathbb{E}[\tau]\mathbb{E}[\bar{e}] \\ \mathbb{E}[(\bar{y}\bar{e}')\bar{e}'\bar{b}] - \mathbb{E}[\bar{y}\bar{e}']\mathbb{E}[\bar{e}'\bar{b}] + \mathbb{E}[\bar{y}'\mathbf{A}\bar{e}\bar{y}'\bar{e}] - \mathbb{E}[\bar{y}'\mathbf{A}\bar{e}]\mathbb{E}[\bar{y}'\bar{e}] &= \mathbb{E}[\tau\bar{y}\bar{e}'] - \mathbb{E}[\tau]\mathbb{E}[\bar{y}\bar{e}']. \end{aligned} \tag{4.53}$$

These are the set of matrix equations that need to be solved to arrive at the UB for our piecewise quasi-linear estimator.

We note that, by lexicographically transforming the indices, we can greatly simplify the set of equations (4.53). We will give a detailed explanation of this procedure for one of the terms, $\mathbb{E}[\bar{y}'\mathbf{A}\bar{e}\bar{y}'\bar{e}]$. A particular element of this expectation can be written as $\mathbb{E}[y_i A_{i\mu} e_\mu y_j e_\nu] = A_{i\mu}\mathbb{E}[y_i e_\mu y_j e_\nu]$. By denoting the new indices lexicographically as $\tilde{i} = i + (\mu - 1)N_t$ and $\tilde{j} = j + (\nu - 1)N_t$, we can express the overall expectation as a true vector-matrix product,

$$\mathbb{E}[\bar{y}'\mathbf{A}\bar{e}\bar{y}'\bar{e}] = \bar{a}'\mathbf{T}, \tag{4.54}$$

where

$$a_{\tilde{i}} = A_{i\mu}, \tag{4.55}$$

$$\mathbf{T}_{\tilde{i}\tilde{j}} = \mathbb{E}[y_i e_\mu y_j e_\nu]. \tag{4.56}$$

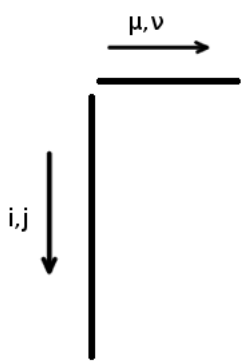

Figure 4.7: Column-wise extraction of elements from a matrix.

Hence, transforming the indices,

$$\begin{aligned} &i, j \in \{1, 2, \ldots, N_t\}, \text{ and} \\ &\mu, \nu \in \{1, 2, \ldots, I\} \end{aligned} \tag{4.57}$$

in this way to $\tilde{i}, \tilde{j}$ as

$$\tilde{i} = i + (\mu - 1)N_t \;\text{ and }\; \tilde{j} = j + (\nu - 1)N_t, \quad \tilde{i}, \tilde{j} \in \{1, 2, \ldots, N_t\, I\}, \tag{4.58}$$

we can express all the expectations in Eq. (4.53) in terms of the new indices. The resulting set of equations can be written in a concise form as the following matrix-vector equations:

$$\begin{aligned} \mathbf{M}\bar{b} + \mathbf{N}\bar{a} &= \bar{v} \\ \mathbf{N}'\bar{b} + \mathbf{U}\bar{a} &= \mathbf{Q}, \end{aligned} \tag{4.59}$$

where

$$\begin{aligned} \mathbf{M} &= \mathbb{E}[\bar{e}\bar{e}'] - \mathbb{E}[\bar{e}]\mathbb{E}[\bar{e}'], \\ \bar{v} &= \mathbb{E}[\tau\bar{e}] - \mathbb{E}[\tau]\mathbb{E}[\bar{e}], \\ U_{\tilde{i}\tilde{j}} &= \mathbb{E}[y_i e_\mu y_j e_\nu] - \mathbb{E}[y_i e_\mu]\mathbb{E}[y_j e_\nu] \\ N_{\nu\tilde{i}} &= \mathbb{E}[e_\nu y_i e_\lambda] - \mathbb{E}[e_\nu]\mathbb{E}[y_i e_\lambda] \\ Q_{\tilde{i}} &= \mathbb{E}[\tau y_i e_\mu] - \mathbb{E}[\tau]\mathbb{E}[y_i e_\mu]. \end{aligned} \tag{4.60}$$

Note that we can define a new variable $\bar{d} = \bar{y}'\bar{e}$ under this lexicographic transformation. The estimator (4.46) can then be written as,

$$\hat{\tau}(\bar{y}) = \bar{b}'\bar{e} + \bar{a}'\bar{d}. \tag{4.61}$$

The unoptimized upper bound (4.5) itself can be expressed, after the transformation, as

$$\mathcal{E}_{UB} = \mathbb{E}\big[(\delta\hat{\tau} - \delta\tau)^2\big], \tag{4.62}$$

$$= \mathbb{E}\big[(\delta\tau)^2\big] + \mathbb{E}\big[(\delta\hat{\tau})^2\big] - 2\mathbb{E}\big[\delta\hat{\tau}\,\delta\tau\big], \tag{4.63}$$

$$= \sigma_\tau^2 + \mathbb{E}\big[(\bar{b}'\bar{e} + \bar{a}'\bar{d})^2\big] - 2\mathbb{E}\big[\bar{b}'\delta\bar{e}\delta\tau + \bar{a}'\delta\bar{d}\delta\tau\big], \tag{4.64}$$

$$\begin{aligned} = \sigma_\tau^2 &+ \bar{b}'\mathbb{E}\big[\delta\bar{e}\,\delta\bar{e}'\big]\bar{b} + \bar{a}'\mathbb{E}\big[\delta\bar{d}\,\delta\bar{d}'\big]\bar{a} + 2\bar{b}'\mathbb{E}\big[\delta\bar{e}\,\delta\bar{d}'\big]\bar{a} \\ &- 2\bar{b}'\mathbb{E}\big[\delta\bar{e}\,\delta\tau\big] - 2\mathbb{E}\big[\delta\bar{d}'\,\delta\tau\big]\bar{a}, \end{aligned} \tag{4.65}$$

$$= \sigma_\tau^2 + \bar{b}'\mathbf{M}\bar{b} - 2\bar{b}'(\bar{v} - \mathbf{N}\bar{a}) + \bar{a}'\mathbf{U}\bar{a} - 2\mathbf{Q}'\bar{a}, \tag{4.66}$$

where we used the relations in (4.60) to go from Eq. (4.65) to Eq. (4.66).

We can now solve the set of equations Eq. (4.59) for $\bar{b}$ and $\bar{a}$. First, express $\bar{b}$ in terms of $\bar{a}$ from the first equation in (4.59),

$$\bar{b} = \mathbf{M}^{-1}(\bar{v} - \mathbf{N}\bar{a}) \tag{4.67}$$

and then substitute it in the second equation of (4.59) to solve for the optimal value for $\bar{a}_*$ as

$$\bar{a}_* = \chi^{-1}(\mathbf{Q} - \mathbf{N}'\mathbf{M}^{-1}v). \tag{4.68}$$

where

$$\chi = \mathbf{U} - \mathbf{N}'\mathbf{M}^{-1}\mathbf{N}. \tag{4.69}$$

The optimal value for $\bar{b}_*$ can then be calculated from Eq. (4.67) in terms of the optimal value for $\bar{a}_*$.

We will now substitute $\bar{b}_*$ in terms of $\bar{a}_*$ into the expression for the upper bound 4.66 to evaluate it as,

$$\begin{aligned}
\text{VB-PQE} &= \sigma_\tau^2 - (\mathbf{M}^{-1}\bar{v} - \mathbf{M}^{-1}\mathbf{N}\bar{a})'(\bar{v} - \mathbf{N}\bar{a}) + \bar{a}'\mathbf{U}\bar{a} - 2\mathbf{Q}'\bar{a}, && (4.70)\\
&= \sigma_\tau^2 - \bar{v}'\mathbf{M}^{-1}\bar{v} + \bar{a}'\chi^{-1}\bar{a} + \bar{v}'\mathbf{M}^{-1}\mathbf{N}\bar{a} + (\bar{v}'\mathbf{M}^{-1}\mathbf{N}\bar{a})' - 2\mathbf{Q}'\bar{a}, && (4.71)\\
&= \sigma_\tau^2 - \bar{v}'\mathbf{M}^{-1}\bar{v} + \bar{a}'\chi^{-1}\bar{a} - 2(\mathbf{Q}' - 2\bar{v}'\mathbf{M}^{-1}\mathbf{N})\bar{a}, && (4.72)\\
&= \sigma_\tau^2 - \bar{v}'\mathbf{M}^{-1}\bar{v} - (\mathbf{Q} - \mathbf{N}'\mathbf{M}^{-1}\bar{v})'\chi^{-1}(\mathbf{Q} - \mathbf{N}'\mathbf{M}^{-1}\bar{v}), && (4.73)
\end{aligned}$$

where in going from (4.70) to (4.71) we used the fact that $\mathbf{M}$ is symmetric and then re-grouped the terms. In (4.72) we used the fact that $\bar{v}'\mathbf{M}^{-1}\mathbf{N}\bar{a}$ is a real number so that it is equal to its transpose. Finally we substituted in (4.73) the optimal value of $\bar{a}_*$ (4.68) to arrive at the optimum upper bound for the PQE, which we abbreviate as PQE-VB.

Note that the upper bound, (4.73), still contains the the width parameters $\{w_i\}$, of the weighting functions $\{e_i\}$, which need to be optimized. We carried this remaining optimization numerically, as we now illustrate for the TDE problem.

4.5.1 Application of the PQE approach to the TDE problem

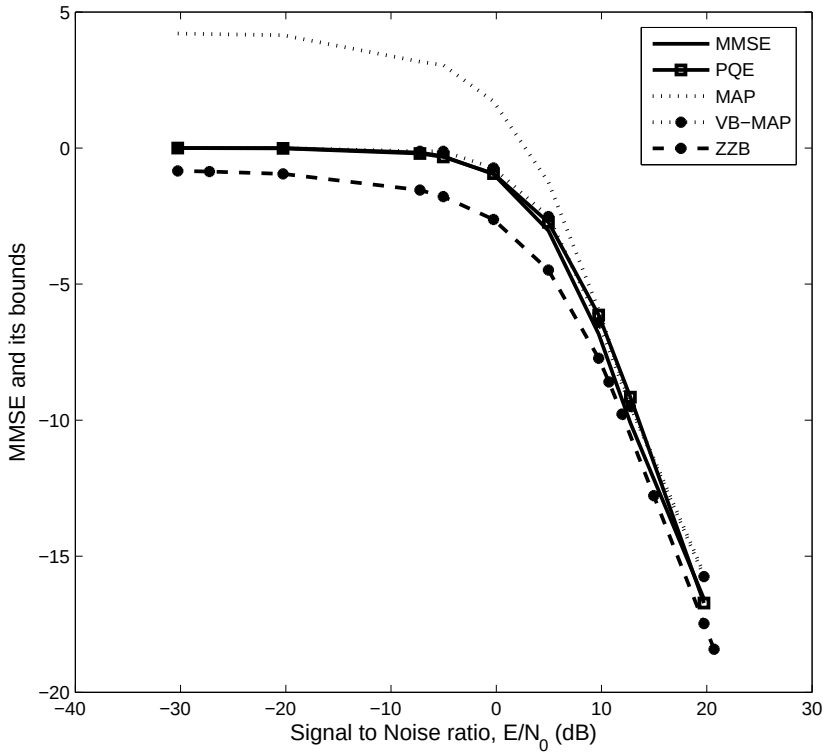

Figure 4.8: MMSE, MAP-VB, PQE, and EZZB for the TDE problem.

We now calculate the PQE-VB for the TDE problem. We assume $I = 4$ and a uniform-width weighting function, i.e., $w_\mu = w_0 \quad \forall \mu = 1, \ldots, I$, so that the final numerical optimization is only over a single variable w_0. The various expectations in (4.60), presented in Appendix B, can be calculated numerically.

The PQE-VB is plotted in Fig. 4.8 which shows that the PQE performs really well in tightly bounding the MMSE. The detailed manner in which we have structured the PQE serves its purpose in tracking the value of the parameter at high SNRs. This includes, among other things, a proper selection of the weighting functions

that enables the estimator to evolve into an appropriate optimum linear estimator estimator, sub-interval by sub-interval, and thus adjust optimally according to the corresponding parameter value.

We can see that the calculation of VB-PQE involves inverting the matrix χ, given by Eq. (4.66). This is an $IN_t \times IN_t$ matrix, involving a quadratic scaling in the product of number of data samples, N_t, times the number of prior intervals, I. For high dimensional estimation this will pose a computational challenge. Therefore, more sophisticated methods which rely on a sparse sampling from a limited number of subregions into which the support region of the prior is divided might be needed for high dimensional problems. However, note that for compressive imaging systems, the number of data samples are far fewer and therefore the scaling problem may not be as severe. The PQE might provide an effective solution in such instances.

In the following section, the statistical error analysis is extended to 2D imaging by means of the upper bound to approximate the MMSE to quantify the performance of different imaging systems.

4.6 Performance analysis of rotating point spread function (RPSF) based imaging system

Point spread function (PSF) engineering for encoding depth in the rotation of the PSF was considered widely in the last decade [113–116]. The rotating-PSF (RPSF) approach presented in [116] is based on orbital angular momentum states of light, which we will call OAR-PSF and those presented in [113, 114] are based on Gauss-Laguerre modes, which we will call GLR-PSF. These computational imaging approaches simultaneously encode the axial and transverse positions of a point source, the former by means of the rotation of the PSF. While the two classes of the RPSFs

are similar, the OAR-PSF, nevertheless, is superior to the GLR-PSF in terms of generating a more compact PSF, having a larger depth of field, and an ability to encode information about spherical aberrations of the imaging optics. The other advantage of this OAR-PSF is that it has a 100% transmission efficiency as opposed to the older GLR-PSFs, making it more sensitive to depth information even under low-light levels. Further, since this OAR-PSF has an ability to generate single-lobe character, it is easier to resolve two nearby point sources which is an attractive property to have when extracting defocus from a densely populated 3D field. All these properties make it a very important 3D imaging modality for at least two important imaging applications, namely, single molecule biological microscopy and 3D space debris tracking for space situational awareness [5].

More generally, both kinds of the rotating-PSFs have been applied to 3D super-localization problems [117–119] and their performance has been evaluated [115, 120] using CRB. Results showed the superior performance of the RPSFs [117] in terms of their localization accuracy.

In this dissertation, however, we will consider only the OAR-PSF. In future, it will be useful to compare the different kinds of 3D imaging modalities. The OAR-PSF maintains its shape and compact size over a large axial depth, as shown in Fig. 4.9, thereby presenting an approach with high sensitivity over a large range of values of the axial depth. The axial depth, or defocus, is encoded in the amount of rotation of the off-center PSF. We will compare the performance of the OAR-PSF for localizing a point source to sub-diffractive scales, which we call the problem of super-localization, with that of a conventional imager. The conventional imager is expected to perform really well at the best focus, but its performance, in terms of sensitivity and transverse resolution, deteriorates rapidly with defocus as shown, e.g., in Fig. 4.9(m). In a 3D imaging scenario, the use of conventional imaging would require scanning, slice by slice, in focus to acquire highly resolvable and sensitive

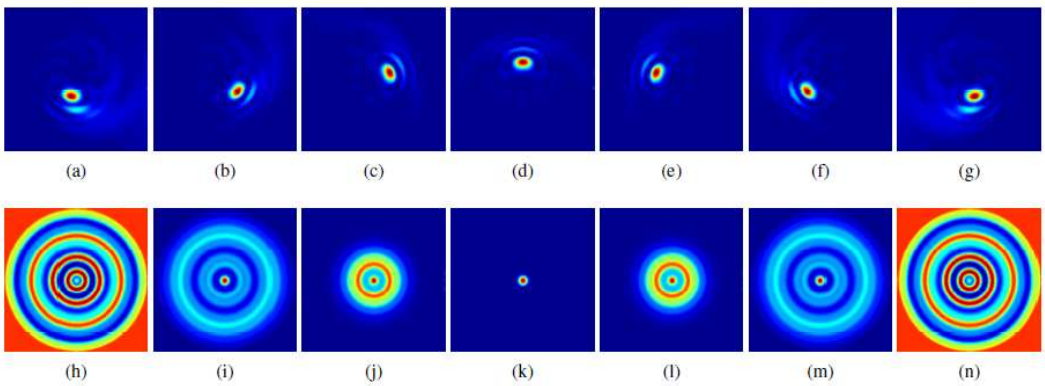

Figure 4.9: Surface plots of the OAR-PSF and the ideal diffraction limited PSF reproduced from [5]. The plots from left to right are for increasing values of defocus, namely -24, -16, -8, 0, 8, 16, and 24 radians at the pupil edge. At a defocus of 16 radians, the conventional PSF, as shown in (m), broadens wide enough to be deemed unfit for imaging. On the other hand, the OAR-PSF simply rotates with defocus and remains compact even at a defocus of 24 radians, as shown in (g).

data, which results in a poor temporal resolution. Unlike the conventional imager, a RPSF based imager encodes the depth in terms of the rotation of its PSF which enables a 3D snap-shot imaging modality in which image information from different depth planes are captured in a single shot.

Unlike a CRB based analysis [117], which is useful mainly in the asymptotic limit, we will use the upper bound on the MMSE in a Bayesian analysis. For a general 3D localization performance characterization, the CRB is not adequate. For example, the conventional PSF is not sensitive to defocus when imaging in focus and this causes to the first order sensitivity to go to zero, resulting in CRB divergence. Therefore a Bayesian analysis provides a more robust alternative in evaluating the performance of various imaging modalities. In this section, we will characterize and demonstrate the robustness of the OAR-PSF in terms of the MMSE for localizing a point source to sub-diffractive errors in the 2D transverse plane only, leaving the axial localization problem for future work.

Specifically, consider an $N_p \times N_p$ nominal diffraction limited area which represents the base resolution cell of an imaging system. The problem of super-localizing a point source may then be phrased in terms of the MSE for localizing the source to within one of M^2 possible sub-cells into which the base resolution cell is subdivided uniformly. Note that this amounts to finding the 2D location of the center of the "blur-spot" that corresponds to a point source. The prior PD was chosen to be uniform over all possible sub-cells. We considered a pseudo-Gaussian conditional data PD which accurately describes, under rather general conditions, the statistics of image data acquired under combined photon-number fluctuations and sensor read-out noise. An appropriate background flux was also taken into consideration. The conditional data statistics on a set of pixels $\{x_i, y_j | i = 1 \ldots, N, j = 1, \ldots, N\}$, given the location (τ_x, τ_y) of the point source can be written as

$$P\big(D(x_i, y_j)|(\tau_x, \tau_y)\big) = \frac{1}{\prod_i \prod_j \sqrt{2\pi v_{i,j}^2}} \exp\left\{ - \sum_i \sum_j \frac{\big[D(x_i, y_j) - m_b - KS(x_i - \tau_x, y_j - \tau_y)\big]^2}{2v_{i,j}^2} \right\}, \tag{4.74}$$

where $\mathbf{D}$ is the observed data, K is the source flux density, $\mathbf{S}$ is the PSF of an imaging system, m_b is the mean background (BG) flux present in the system, and

$$v_{i,j}^2 = KS(x_i - \tau_x, y_j - \tau_y) + m_b + \sigma_n^2. \tag{4.75}$$

The variance of the read out noise, σ_n^2, is defined in terms of its noise spectral density, N_0, and the pixel area, $\Delta x \Delta y$, as

$$\sigma_n^2 = \frac{N_0}{\Delta x \Delta y}. \tag{4.76}$$

Typically, a region of size $N_s \times N_s$ is selected from a larger $N \times N$ image data region containing the possible point source. This can be done by cropping a sub-image around the pixel with maximum intensity. The selected sub-image is then

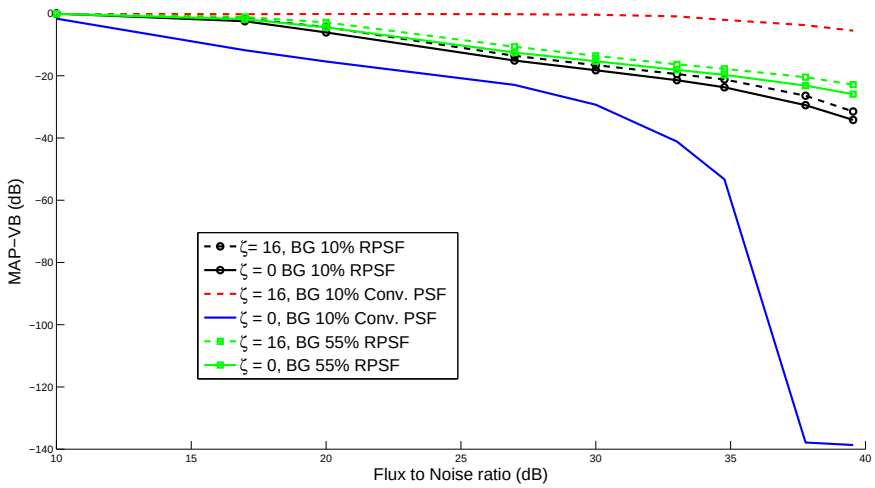

Figure 4.10: MAP-VB for the source super-localization problem with an $M = 16$ fold enhancement. The conventional imager does exceptionally well at best focus, but performs very poorly at large defocus. The OAR-PSF, however remains robust as the defocus changes over a large range.

processed to find the location of the point source. The MSE for this 2D problem is defined as the sum of the MSEs for the two source position coordinates,

$$\mathcal{E} = \mathbb{E}[(\hat{\tau}_x - \tau_x)^2] + \mathbb{E}[(\hat{\tau}_y - \tau_y)^2]. \tag{4.77}$$

We used the MAP-VB as an approximation to the MMSE. Test estimators involving the MAP estimators, $(\hat{\tau}_x, \hat{\tau}_y)$, and variational parameters $\{\alpha_x, \beta_x, \alpha_y, \beta_y\}$ can be written as

$$\tilde{\tau}_x = \alpha_x \hat{\tau}_x + \beta_x, \tag{4.78}$$

$$\tilde{\tau}_y = \alpha_y \hat{\tau}_y + \beta_y. \tag{4.79}$$

The upper bound

$$\mathcal{E}_{UB} = \mathbb{E}\left[(\alpha_x \delta\tilde{\tau}_x - \delta\tau_x)^2\right] + \mathbb{E}\left[(\alpha_y \delta\tilde{\tau}_y - \delta\tau_y)^2\right] \tag{4.80}$$

is then optimized over the variational parameters $\{\alpha_x, \alpha_y\}$ to yield the MAP-VB as

$$\text{MAP-VB} = \sigma_x^2 + \sigma_y^2 - \frac{\mathbb{E}^2\left[\delta\hat{\tau}_x \delta\tau_x\right]}{\mathbb{E}\left[\delta\hat{\tau}_x^2\right]} - \frac{\mathbb{E}^2\left[\delta\hat{\tau}_y \delta\tau_y\right]}{\mathbb{E}\left[\delta\hat{\tau}_y^2\right]}, \tag{4.81}$$

where σ_x^2 and σ_y^2 are the prior variances $\mathbb{E}[\delta\tau_x]$ and $\mathbb{E}[\delta\tau_y]$ respectively. Note that, the upper bound (4.80) is independent of $\{\beta_x, \beta_y\}$ as already noted in Sec. 4.3.2, the latter simply provide a means to impose the unbiasedness of the estimators without affecting the optimal upper bound.

For $M = 16$ fold linear enhancement in the localization accuracy, the prior on (τ_x, τ_y) will be uniform on the $M^2 = 256$ possible positions for the point source. Since the parameters are uniformly distributed, the MAP estimator is obtained by maximizing just the conditional data statistics, (4.74), as

$$(\hat{\tau}_x, \hat{\tau}_y) = \underset{(\tau_x,\tau_y)}{\text{argmax}}\, P\big((\tau_x, \tau_y)|D(x_i, y_i)\big), \tag{4.82}$$

$$= \underset{(\tau_x,\tau_y)}{\text{argmax}}\, P\big(D(x_i, y_i)|(\tau_x, \tau_y)\big), \tag{4.83}$$

$$= \underset{(\tau_x,\tau_y)}{\text{argmax}}\, \ln P\big(D(x_i, y_i)|(\tau_x, \tau_y)\big), \tag{4.84}$$

$$= \underset{(\tau_x,\tau_y)}{\text{argmin}} \sum_i^{N_s} \sum_j^{N_s} \left\{ \ln v_{i,j}^2 + \frac{\left[D(x_i, y_j) - m_b - KS(x_i - \tau_x, y_j - \tau_y)\right]^2}{v_{i,j}^2} \right\}. \tag{4.85}$$

The SNR is redefined as the flux to noise ratio (FNR), $\frac{K}{\sigma_n^2}$, involving the total number of photons incident on the imager. The mean BG, m_b is chosen to be a fraction of the peak value of the total photon number K. As Fig. 4.10 shows, the conventional imager yields a rapid decrease in 2D-localization MMSE with increasing FNR for best focus, but its behavior with respect to the OAR-PSF-based imager is reversed, as we shall see presently, at a large defocus phase,[1] $\zeta = 16$ radians, with a rather small reduction in that MMSE even at FNR = 40 dB. However, the OAR-PSF

[1]The defocus phase, ζ, is the quadratic phase at the edge of the pupil corresponding to a defocus distance δz in the object space.

imager performs robustly over a defocus phase between 0 and 16 (and in fact for much greater values of the defocus phase). We can define the minimum value of FNR needed to achieve an M-fold transverse super-localization as that FNR value at which the MMSE is reduced by a factor M^2.

This analysis is part of a comprehensive 3D super-localization performance analysis, including an MPE analysis, of the OAR-PSF imager to be presented in [121]. The transverse and longitudinal resolution trade-off limits of the OAR-PSF were studied in [5], in which a robust recovery of transverse and longitudinal resolution over an increased depth of field was demonstrated. The upper bound thus presents a convenient way of approximating the MMSE and thus a valuable tool in a statistical performance analysis of different imaging systems.

4.7 Summary

In this chapter, an upper bound on the MMSE was derived. The upper bound was also generalized for vector parameters. The so derived upper bound is numerically efficient to evaluate, and allows for a removal of the Bayesian bias from any estimator. Since the MSE of any estimator differs from the MMSE by a quadratic deviation of the estimator from the MMSE estimator, a variational approach provides an accurate approximation for the MMSE via such an upper bound. The test estimator was formulated in terms of certain variational parameters over which the upper bound was minimized.

The upper bound was considered for the polynomial class of estimators and to the MAP estimator for a variety of applications. The variational upper bound involving the MAP estimator ties both the high SNR and low SNR ends of the MSE to the value of the MMSE in those regions, and thus tightly bounds the MMSE for all values of the SNR. The polynomial class of estimators performs well for certain cases, but,

in general, does not have the necessary complexity whenever the data depend rather nonlinearly on a parameter. A piecewise quasi-linear estimator, which is more useful in such situations, was presented. It accounts for the non-linear complexity through certain data-dependent weighting functions. Our results show that the piecewise quasi-linear estimator based UB and the MAP-VB tightly bound the MMSE for all values of SNR, and outperform the existing popular lower bounds, namely the EZZB and the WWB, on the MMSE in terms of their tightness. The MAP-VB was used to analyze and compare the performance of a newly developed rotating-PSF imager with that of a conventional imager in localizing a point source beyond the diffraction limit. Based on our results, it was shown that the OAR-PSF imager is quite robust to large changes of defocus, encoding defocus in PSF rotation, and thus provides a practical method for 3D snap-shot imaging.

The formulation of the upper bound, the polynomial class of test estimators, and the MAP-VB form the basis for our paper [122] which is under preparation. The piecewise quasi-linear upper bound will be presented in another paper [123]. Statistical error analyses of the OAR-PSF imager appear in [121, 124].

Chapter 5

Asymptotic Analyses of the MMSE

Asymptotic analyses of the posterior PD have been widely studied in the literature [99, 125–129]. All of them use the Laplace method for an aymptotic evaluation of integrals [130]. An asymptotic analysis of posterior PD is important as it provides insights into the behavior of a particular estimation rule and of the various bounds on the MMSE when there are many measurements available. As we already saw in the previous chapter, the MAP-VB performs well, mainly because it matches the MMSE at both low and high SNR values. The error is small at high SNR values where the CRB, which is based on a first order sensitivity metric of information, the Fisher information, proves sufficient and tightly bounds the MMSE. As the SNR decreases, thereby increasing the error, a bound that includes higher order derivatives as well will provide a much tighter approximation to the MMSE.

In this chapter, we will develop a framework to approximate the MMSE for low values of the SNR first. A subsequent analysis will be presented for the asymptotic regime of high SNR. Since the data become less informative at lower SNR values,

a good estimator is expected to attain the mean value of the prior, which is the MMSEE, in the limiting case of zero SNR. Therefore, a Taylor series expansion is used to expand the MMSEE about the mean value of the prior, from which, we construct an approximate expression for both the MMSEE and the MMSE, which is expected to be valid for low SNR values. At high SNR values, unlike the Laplace method widely used in the literature [127], a different implementation of the Laplace method is used for asymptotically evaluating integrals. This, as we will see, yields an excellent approximate expression for the MMSE. A new parameter, related to the joint distribution, is introduced and expanded about the MAP estimator. The resulting MMSE approximation is expected to be valid for a wider range of moderate to high SNR values. Finally, a new approximation to the MMSE is proposed that interpolates between the low SNR and high SNR values of the MMSE. Our results show that it is able to capture the behavior of error for all SNR values.

5.1 The low SNR regime

Equation (2.8) shows that the MMSE is always lower than the variance of the prior. It improves with the information provided by Y about X, and can be shown to go to 0 in the asymptotic limit of infinitely high SNR. In the limit of zero SNR, the MMSE tends to the variance of the prior, σ_X^2 which is attained by the mean of the prior, $\hat{X}_M = E[X]$, as the data, Y, carry no information about X in this limit.

We present below approximate expressions for the MMSE and the MMSE estimator that are valid in the low SNR regime. These expressions are of a perturbative nature, with the small perturbative parameter given by the ratio of a first-order data sensitivity metric, specifically the mean-squared score function, and the prior variance σ_X^2. While it is possible to push our perturbative expansions to higher order sensitivity parameters, we shall be content, in the following analysis, with the lowest-

order correction to the prior mean and variance that represent the MMSE estimator and the MMSE in the zero-SNR limit.

5.1.1 A second order logarithmic expansion of the data PD

At low SNR values, the data PD, $P(y|x)$, is expected to be relatively insensitive to the value of the random variable X. This suggests the following approximate Taylor expansion for the log-likelihood function (LLF) around the prior mean:

$$\ln P(y|x) \approx \ln P(y|\langle X\rangle) + V(y)(\delta x) - \frac{1}{2}S(y)(\delta x)^2 \tag{5.1}$$

$$\equiv \ln \bar{P}(y|x), \tag{5.2}$$

where $\delta x = x - \langle X\rangle$ represents a particular realization of the zero mean random variable $\delta X = X - \langle X\rangle$. Here $V(y)$ is the score function, which is a first-order sensitivity function, and $S(y)$ is a second-order sensitivity function, both evaluated at the prior mean and defined by the relations

$$V(y) = \frac{\partial}{\partial \langle X\rangle} \ln P(y|\langle X\rangle), \quad S(y) = -\frac{\partial^2}{\partial \langle X\rangle^2} \ln P(y|\langle X\rangle). \tag{5.3}$$

We can express Eq. (5.1) equivalently as

$$\bar{P}(y|x) = P(y|\langle X\rangle) \exp[V(y)(\delta x) - \frac{1}{2}S(y)(\delta x)^2]. \tag{5.4}$$

The expected value of the score function over $P(y|\langle X\rangle)$, namely $\mathbb{E}[V(Y)|\langle X\rangle]$, vanishes,

$$\begin{aligned}\mathbb{E}[V(Y)|\langle X\rangle] =& \int_{R_Y} P(y|\langle X\rangle) \frac{\partial}{\partial \langle X\rangle} \ln P(y|\langle X\rangle) dy \\ =& \frac{d}{d\langle X\rangle} \int_{S_Y} P(y|\langle X\rangle) dy = 0,\end{aligned} \tag{5.5}$$

in which the unit normalization of the conditional data PD $P(y|X)$, for any realization of X, was used to arrive at the final equality. With the negative sign chosen in

the definition of S, the expected values of S and V^2 over $P(y|\langle X\rangle)$ may be shown [21] to be equal to each other and to $J(Y|\langle X\rangle)$, the Fisher information (FI) of the data PD with respect to X at its mean value $x = \langle X\rangle$:

$$\mathbb{E}[S(Y)|\langle X\rangle] = \mathbb{E}[V^2(Y)|\langle X\rangle] = J(Y|\langle X\rangle), \tag{5.6}$$

where

$$J(Y|x) = \mathbb{E}\Big[\Big(\frac{\partial \ln P(Y|x)}{\partial x}\Big)^2|x\Big] \tag{5.7}$$

As is well known, the inverse of the FI is the CRB that lower-bounds the variance of any unbiased estimator of X, so it represents a measure of the maximum first-order sensitivity with which the data yield information about the signal.

The advantage of expanding the LLF, instead of the likelihood function $P(y|x)$ itself, about the prior mean is that the approximate PD, $\bar{P}(y|x)$ is in fact already normalized correctly to the second order in data sensitivity, i.e., to the same order as S and V^2. To see this, we expand expression (5.4) to the second order to obtain

$$\bar{P}(y|x) = P(y|\langle X\rangle)\Big[1 + V(y)(\delta x) + \frac{1}{2}[V^2(y) - S(y)](\delta x)^2\Big]. \tag{5.8}$$

Since, according to identities (5.5) and (5.6), the expected values of V and $V^2 - S$ vanish, the integral of $\bar{P}(y|x)$ over the sample space R_Y of the data then yields the integral of $P(y|\langle X\rangle)$ over the same space, which is one since the exact $P(y|x)$ is normalized at all values of X, including $X = \langle X\rangle$.

5.1.2 The MMSE estimator and the MMSE to second order

By writing $x = \delta x + \langle X\rangle$ in Eq. (2.4) and using the normalization of the posterior $P(x|y)$, followed by the Bayes relation, $P(x|y) = P(y|x)P(x)/P(y)$, we first express the MMSE estimator (2.4) as

$$X_M(y) = \langle X\rangle + \frac{1}{P(y)}\int_{R_X}(\delta x)P(y|x)P(x)dx \tag{5.9}$$

Using relation (5.2) to approximate $P(y|x)$ by $\bar{P}(y|x)$, substituting the second-order approximation (5.8) for $\bar{P}(y|x)$ inside the integral in Eq. (5.9), and then noting that the expected value of the linear deviation δX is zero for the MMSE estimator, we get the following expression correct to the second order in sensitivity:

$$X_M(y) = \langle X \rangle + \frac{P(y|\,\langle X \rangle)}{P(y)} \Big[V(y)\sigma_X^2 + \frac{1}{2}\big(V^2(y) - S(y)\big) s_X^{(3)} \Big], \tag{5.10}$$

where $\sigma_X^2 = \mathbb{E}[(\delta X)^2]$ and $s_X^{(3)} = \mathbb{E}[(\delta X)^3]$ are, respectively, the variance and skewness of the prior. For a prior that is symmetric about its mean, its skewness vanishes identically, and the MMSE estimator takes the following simple form correct to second order:

$$X_M(y) = \langle X \rangle + \frac{P(y|\,\langle X \rangle)}{P(y)} V(y)\sigma_X^2. \tag{5.11}$$

We shall be mostly concerned with symmetric priors and thus expect the expression (5.11), for the MMSE estimator, to be valid in the low-SNR regime. We now calculate the marginal data PD, $P(y)$, to second order by multiplying expression (5.8) by $P(x)$ and then integrating over the sample space R_X. Using the fact that the expected value of δX is identically zero, we obtain the result,

$$\bar{P}(y) = P(y|\,\langle X \rangle)\big(1 + \frac{1}{2}[V^2(y) - S(y)]\sigma_X^2\big). \tag{5.12}$$

Substituting this expression into Eq. (5.11) and keeping terms to second order only yields

$$X_M(y) = \langle X \rangle + \frac{1}{1 + (1/2)[V^2(y) - S(y)]\sigma_X^2} V(y)\sigma_X^2 \tag{5.13}$$

$$\approx \langle X \rangle + V(y)\sigma_X^2, \tag{5.14}$$

since terms of the form $V(V^2 - S)$ are clearly of higher order in sensitivity and must be omitted for consistency.

With this form for the MMSE estimator, we may now easily compute the MMSE in the low-SNR limit,

$$\begin{aligned}\mathcal{E}_M =&\mathbb{E}[\hat{X}_M - X]^2 &(5.15)\\ =&\mathbb{E}[\delta X]^2 - 2\sigma_X^2\mathbb{E}[(\delta X)V(Y)] + \sigma_X^4\mathbb{E}[V^2(Y)] &(5.16)\\ =&\sigma_X^2\big(1 - 2\mathbb{E}[(\delta X)V(Y)] + \sigma_X^2\mathbb{E}[V^2(Y)]\big). &(5.17)\end{aligned}$$

The middle term Eq. (5.17) may be simplified by performing the expectation in two steps - first over $P(y|x)$ and then over $P(x)$. The first of these expectations, when $P(y|x)$ is approximated by its second-order expression (5.8), evaluates to the second order as

$$\begin{aligned}\mathbb{E}_{Y|X}V(Y) =& \int_{R_Y} dyP(y|\,\langle X\rangle)V(y)\{1 + V(y)(\delta x)\\ &+ \frac{1}{2}[V^2(y) - S(y)](\delta x)^2\} &(5.18)\\ =&(\delta x)\mathbb{E}[V^2(Y)|\,\langle X\rangle] + \frac{1}{2}(\delta x)^2\mathbb{E}[(V^2(Y) - S(Y))V(Y)|\,\langle X\rangle], &(5.19)\end{aligned}$$

where we used the fact (5.5) that $\mathbb{E}[V(Y)|\,\langle X\rangle]$ vanishes identically. The final expectation of δX times this expression over the prior $P(x)$ then yields for the middle term in (5.17)

$$\begin{aligned}\mathbb{E}[(\delta X)V(Y)] =&\sigma_X^2\mathbb{E}[V^2(Y)|\,\langle X\rangle] + \frac{1}{2}s_X^{(3)}\mathbb{E}[V(Y)(V^2(Y) - S(Y))|\,\langle X\rangle] &(5.20)\\ =&\sigma_X^2\mathbb{E}[V^2(Y)], &(5.21)\end{aligned}$$

where the last term in the first equality has been dropped to keep all expressions self-consistent with our second-order approximation. For a symmetric PD function with zero skewness, that term vanishes identically, however, independent of any considerations of the order of the approximation. Using the result (5.21) in the MMSE expression (5.17) and Eq. (5.6) to replace $\mathbb{E}[V^2(Y)]$ by the FI, $J(Y|\,\langle X\rangle)$, we obtain the following simple form for the MMSE in the low-SNR, LSNR-MMSE, regime:

$$\mathcal{E}_M = \sigma_X^2[1 - \sigma_X^2 J(Y|\,\langle X\rangle)]. \tag{5.22}$$

5.2 The high SNR regime

Fisher information (FI) serves to provide a good local measure of sensitivity of statistical data to parameters to be estimated. Its inverse yields the CRB which, as we have noted earlier, is a lower bound on the variance of any unbiased estimator of the parameters being estimated. Bhattacharya, Chapman-Robbins, and Barankin lower bounds improve upon the CRB, whenever the latter is not attainable by *any* estimator. However, all such local estimation-fidelity measures fail dramatically at moderate to low values of SNR for which the information contained in the prior becomes increasingly important in controlling the MSE. A uniform prior illustrates rather dramatically the inadequacy of local bounds, since by their very construction they are unaffected by such a prior and thus unable to accommodate it. We shall now show how to accommodate arbitrary priors in a more global, Bayesian analysis. Specifically, we will use the Laplace method for evaluating integrals to obtain an approximate calculation of the MMSE in terms of the MAP estimator.

First, consider the joint PD to define a new variable u via the identity

$$\begin{aligned} P(x,y) =& P(\hat{X}_{\mathrm{MAP}}(y),y)\frac{P(x,y)}{P(\hat{X}_{\mathrm{MAP}}(y),y)} \\ =& P(\hat{X}_{\mathrm{MAP}}(y),y)\exp\Big\{-\frac{1}{2}\left\langle S(Y)\right\rangle [u-\hat{X}_{\mathrm{MAP}}(y)]^2\Big\}. \end{aligned} \tag{5.23}$$

Equivalently,

$$[u-\hat{X}_{\mathrm{MAP}}(y)]^2 = -\frac{2}{\langle S(Y)\rangle}\ln\left(\frac{P(x,y)}{P(\hat{X}_{\mathrm{MAP}}(y),y)}\right) \tag{5.24}$$

We have the definition of $S(y)$ from the following observation:

$$\begin{aligned} P(x,y) =& \exp[\ln P(x,y)] \\ \approx& \exp\Big\{\ln P(\hat{X}_{\mathrm{MAP}}(y),y) \\ & -\frac{1}{2}[x-\hat{X}_{\mathrm{MAP}}(y)]^2\big[-\frac{\partial^2}{\partial\hat{X}^2_{\mathrm{MAP}}(y)}\ln P(\hat{X}_{\mathrm{MAP}}(y),Y)\big]\Big\} \\ =& P(\hat{X}_{\mathrm{MAP}}(y),y)\exp\Big\{-\frac{1}{2}S(y)[x-\hat{X}_{\mathrm{MAP}}(y)]^2\Big\} \end{aligned} \tag{5.25}$$

where

$$S(y) = -\frac{\partial^2 \ln P(x,y)}{\partial x^2}\bigg|_{x=\hat{X}_{\rm MAP}(y)} \tag{5.26}$$

and the approximation in the second equation is valid at high values of the SNR. Note that we can informally derive the Bayesian central limit theorem [99] by noting that

$$P(y) = \int dx\, P(x,y) = \sqrt{\frac{2\pi}{S(y)}} P(\hat{X}_{\rm MAP}(y), y), \tag{5.27}$$

from which we can write the posterior PD as

$$P(x|y) = \sqrt{\frac{S(y)}{2\pi}} \exp\Big\{-\frac{S(y)}{2}[x - \hat{X}_{\rm MAP}(y)]^2\Big\}, \tag{5.28}$$

which is a Gaussian PD with mean $\hat{X}_{\rm MAP}(y)$. As a consequence of this the MMSEE is the same as the MAP estimator. At high SNR, since the data are highly sensitive to the parameter, $S(y)$ will be large which makes the variance of the posterior distribution about its mean small. As the SNR decreases, the posterior distribution departs from being a Gaussian, and higher order sensitivities start to become significant. Since $S(y)$ is a second-order sensitivity measure, we need to include higher order sensitivities to approximate the posterior distribution well at lower values of the SNR.

5.2.1 Higher order corrections

From relation (5.24) we can see that x is a function of both u and y. Keeping y fixed, let us expand x about $u = \hat{X}_{\rm MAP}(y)$ correct to the 4^{th} order:

$$\begin{aligned} x(u,y) = x[u = \hat{X}_{\rm MAP}(y), y] + \gamma[u - \hat{X}_{\rm MAP}(y)] + \epsilon[u - \hat{X}_{\rm MAP}(y)]^2 \\ + \eta[u - \hat{X}_{\rm MAP}(y)]^3 + \xi[u - \hat{X}_{\rm MAP}(y)]^4 \end{aligned} \tag{5.29}$$

where

$$\gamma = \frac{\partial x}{\partial u}\Big|_{u=\hat{X}_{\mathrm{MAP}}(y)},\ \epsilon = \frac{1}{2}\frac{\partial^2 x}{\partial u^2}\Big|_{u=\hat{X}_{\mathrm{MAP}}(y)}, \tag{5.30}$$

$$\eta = \frac{1}{6}\frac{\partial^3 x}{\partial u^3}\Big|_{u=\hat{X}_{\mathrm{MAP}}(y)},\ \xi = \frac{1}{24}\frac{\partial^4 x}{\partial u^4}\Big|_{u=\hat{X}_{\mathrm{MAP}}(y)}, \tag{5.31}$$

are the higher order corrections to an expansion of the posterior distribution when it departs from being Gaussian. We will now calculate the higher order corrections γ, ϵ, η and ξ. Note that as $u \to \hat{X}_{\mathrm{MAP}}(y)$, so does $x \to \hat{X}_{\mathrm{MAP}}(y)$.

Taking a partial derivative with respect to (w.r.t) u of both sides of the relation, 5.24, we have

$$u - \hat{X}_{\mathrm{MAP}}(y) = -\frac{1}{\langle S(Y)\rangle}\frac{\partial \ln P}{\partial x}\frac{\partial x}{\partial u}, \tag{5.32}$$

$$\Rightarrow \frac{\partial x}{\partial u} = -\frac{\langle S(Y)\rangle\,(u - \hat{X}_{\mathrm{MAP}}(y))}{\dfrac{\partial \ln P}{\partial x}}, \tag{5.33}$$

where $P(x, y)$ is shortened to P to make the expressions more readable. We will state the PD explicitly if needed. As $u \to \hat{X}_{\mathrm{MAP}}(y)$, the numerator $\to$ zero. Since $\hat{X}_{\mathrm{MAP}}(y)$ maximizes the posterior $P(x|y)$ and in turn the joint PD, $P(x, y)$, for a fixed y, the denominator also goes to 0 as $u \to \hat{X}_{\mathrm{MAP}}(y)$ at which $x \to \hat{X}_{\mathrm{MAP}}(y)$. The l'Hospital rule yields the proper limit of the ratio (5.33) as $u \to \hat{X}_{\mathrm{MAP}}(y)$:

$$\frac{\partial x}{\partial u}\Big|_{u=\hat{X}_{\mathrm{MAP}}(y)} = -\frac{\langle S(Y)\rangle}{\dfrac{\partial^2 \ln P(x,y)}{\partial x^2}\Big|_{x=\hat{X}_{\mathrm{MAP}}(y)}\dfrac{\partial x}{\partial u}\Big|_{u=\hat{X}_{\mathrm{MAP}}(y)}}. \tag{5.34}$$

Multiplying both sides by $\dfrac{\partial x}{\partial u}\Big|_{u=\hat{X}_{\mathrm{MAP}}(y)}$ yields

$$\left(\frac{\partial x}{\partial u}\Big|_{u=\hat{X}_{\mathrm{MAP}}}\right)^2 = \frac{\langle S(Y)\rangle}{S(y)}, \tag{5.35}$$

from which it follows that

$$\gamma = \sqrt{\frac{\langle S(Y)\rangle}{S(y)}}. \tag{5.36}$$

Taking partial derivative w.r.t u on both sides of Eq. (5.32),

$$1 = -\frac{1}{\langle S(Y)\rangle}\left[\frac{\partial^2 \ln P}{\partial x^2}\Big(\frac{\partial x}{\partial u}\Big)^2 + \frac{\partial \ln P}{\partial x}\frac{\partial^2 x}{\partial u^2}\right], \tag{5.37}$$

$$\Rightarrow \quad \frac{\partial^2 x}{\partial u^2} = -\frac{\langle S(Y)\rangle + \dfrac{\partial^2 \ln P}{\partial x^2}\Big(\dfrac{\partial x}{\partial u}\Big)^2}{\dfrac{\partial \ln P}{\partial x}}. \tag{5.38}$$

As $u \to \hat{X}_{\mathrm{MAP}}(y)$, the numerator tends to $\langle S\rangle - S\dfrac{\langle S\rangle}{S} = 0$. The denominator also tends to zero as we saw earlier. Using the l'Hospital rule, as $u \to \hat{X}_{\mathrm{MAP}}(y)$, we have

$$\frac{\partial^2 x}{\partial u^2}\Big|_{u=\hat{X}_{\mathrm{MAP}}(y)} = -\frac{(\ln P)^{(3)}}{(\ln P)^{(2)}}\gamma^2 - 2\frac{\partial^2 x}{\partial u^2}\Big|_{u=\hat{X}_{\mathrm{MAP}}(y)} \tag{5.39}$$

$$\Rightarrow \epsilon = -\frac{1}{6}\frac{(\ln P)^{(3)}}{(\ln P)^{(2)}}\gamma^2, \tag{5.40}$$

where

$$(\ln P)^{(n)} \stackrel{\text{def}}{=} \frac{\partial^n \ln P(x,y)}{\partial \hat{x}^n}\Big|_{x=\hat{X}_{\mathrm{MAP}}(y)}. \tag{5.41}$$

To find η, we take a partial derivative w.r.t u of both sides of Eq. (5.37), and go through similar calculations as above to yield,

$$\eta = \left\{\frac{5}{12}\Big[\frac{(\ln P)^{(3)}}{(\ln P)^{(2)}}\Big]^2 - \frac{1}{4}\frac{(\ln P)^{(4)}}{(\ln P)^{(2)}}\right\}\frac{\gamma^3}{6}, \tag{5.42}$$

$$\xi = \left\{\frac{2}{9}\Big[\frac{(\ln P)^{(3)}}{(\ln P)^{(2)}}\Big]^3 - \frac{(\ln P)^{(4)}(\ln P)^{(3)}}{[(\ln P)^{(2)}]^2} - \frac{1}{5}\frac{(\ln P)^{(5)}}{(\ln P)^{(2)}}\right\}\frac{\gamma^4}{24}, \tag{5.43}$$

Using the expressions for γ, ϵ, η, and ξ we can express the partial derivative of x w.r.t u as,

$$\frac{\partial x}{\partial u} = \gamma + 2\epsilon[u - \hat{X}_{\mathrm{MAP}}(y)] + 3\eta[u - \hat{X}_{\mathrm{MAP}}(y)]^2 + 4\xi[u - \hat{X}_{\mathrm{MAP}}(y)]^3. \tag{5.44}$$

Since we are able to express x and $\dfrac{\partial x}{\partial u}$ in terms of the correction terms and u, we can transform the integrals from x to u, which will enable us to calculate the

probabilities and expectations. We first evaluate $P(y)$ as

$$\begin{aligned} P(y) &= \int dx P(x,y), \\ &= \int du \frac{\partial x}{\partial u} P(\hat{X}_{\mathrm{MAP}}(y), y) \exp\{-\frac{1}{2}\langle S(Y)\rangle [u - \hat{X}_{\mathrm{MAP}}(y)]^2\}, \\ &= P(\hat{X}_{\mathrm{MAP}}(y), y) \sqrt{\frac{2\pi}{\langle S(Y)\rangle}} \left(\gamma + \frac{3\eta}{\langle S(Y)\rangle}\right). \end{aligned} \tag{5.45}$$

The posterior PD is given by

$$P(x|y) = \frac{P(x,y)}{P(y)}, \tag{5.46}$$

$$= \sqrt{\frac{\langle S(Y)\rangle}{2\pi}} \left(\gamma + \frac{3\eta}{\langle S(Y)\rangle}\right)^{-1} \exp\{-\frac{1}{2}\langle S(Y)\rangle [u - \hat{X}_{\mathrm{MAP}}(y)]^2\}, \tag{5.47}$$

which is clearly not Gaussian. This shows that the expression for the posterior PD is moving away from the limiting asymptotic Gaussian form when higher order corrections are included. By writing $x = [x - \hat{X}_{\mathrm{MAP}}(y)] + \hat{X}_{\mathrm{MAP}}(y)$, we can now calculate the MMSE estimator as,

$$\hat{X}_M(y) = \int dx\, P(x|y) x, \tag{5.48}$$

$$= \hat{X}_{MAP(y)} + \int du\, \frac{\partial x}{\partial u} [x - \hat{X}_{\mathrm{MAP}}(y)] P(x|y), \tag{5.49}$$

$$= \hat{X}_{MAP(y)} + \frac{P(\hat{X}_{\mathrm{MAP}}(y), y)}{2P(y)} \int du \frac{\partial [x - \hat{X}_{\mathrm{MAP}}(y)]^2}{\partial u} \exp\{-\frac{1}{2}\langle S(Y)\rangle [u - \hat{X}_{\mathrm{MAP}}(y)]^2\}. \tag{5.50}$$

From relation (5.29), we have

$$\begin{aligned} [x - \hat{X}_{\mathrm{MAP}}(y)]^2 &= \gamma^2 (u - \hat{X})^2 + \epsilon^2 (u - \hat{X})^4 + 2\gamma\epsilon (u - \hat{X})^3 \\ &\quad + 2\gamma\eta (u - \hat{X})^4 + 2\gamma\xi (u - \hat{X})^5 + 2\epsilon\eta (u - \hat{X})^5 \end{aligned} \tag{5.51}$$

$$\begin{aligned} \Rightarrow \frac{\partial (x - \hat{X}_{MAP(y)})^2}{\partial u} &= 2\gamma^2 (u - \hat{X}) + 4\epsilon^2 (u - \hat{X})^3 + 6\gamma\epsilon (u - \hat{X})^2 \\ &\quad + 8\gamma\eta (u - \hat{X})^3 + 10(\gamma\xi + \epsilon\eta)(u - \hat{X})^4, \end{aligned} \tag{5.52}$$

where we abbreviated $\hat{X}_{MAP(y)}$ as $\hat{X}$ in the RHS of the equations to simplify the notation. Thus, from Eq. (5.50), we can express the difference between the MMSEE and the MAP estimator as

$$\hat{X}_M(y) - \hat{X}_{\mathrm{MAP}}(y) = \frac{P(\hat{X}_{\mathrm{MAP}}(y), y)}{2P(y)} \left\{ \left(\frac{6\gamma\epsilon}{\langle S(Y)\rangle} + \frac{30(\gamma\xi + \epsilon\eta)}{\langle S(Y)\rangle^2} \right) \sqrt{\frac{2\pi}{\langle S(Y)\rangle}} \right\} \tag{5.53}$$

$$= \left(\frac{3\gamma\epsilon}{\langle S(Y)\rangle} + \frac{15(\gamma\xi + \epsilon\eta)}{\langle S(Y)\rangle^2} \right) \left(\gamma + \frac{3\eta}{\langle S(Y)\rangle} \right)^{-1} \tag{5.54}$$

$$= \left(\frac{3\epsilon}{\langle S(Y)\rangle} + \frac{15(\gamma\xi + \epsilon\eta)}{\gamma \langle S(Y)\rangle^2} \right) \left(1 - \frac{3\eta}{\gamma \langle S(Y)\rangle} \right) \tag{5.55}$$

$$= \frac{3\epsilon}{\langle S(Y)\rangle} + \frac{6\epsilon\eta}{\gamma \langle S(Y)\rangle^2} + \frac{15\xi}{\langle S(Y)\rangle^2}, \tag{5.56}$$

where we collected all terms up to the order $\langle S(Y)\rangle^{-2}$. In arriving at (5.53) we used the fact that the fourth central moment of a Gaussian is three times its variance. Using this relation for the MMSEE, we can calculate the MMSE,

$$\mathcal{E}_M = \int\int dx\, dy\, P(x,y)[x - \hat{X}_M(y)]^2, \tag{5.57}$$

$$= \int dy\, P(y) \int dx\, P(x|y)[x - \hat{X}_M(y)]^2, \tag{5.58}$$

$$= \int dy\, P(y) \left(\gamma + \frac{3\eta}{\langle S(Y)\rangle} \right)^{-1} \int du \frac{\partial x}{\partial u} (x - \hat{X}_M(y))^2 \sqrt{\frac{\langle S(Y)\rangle}{2\pi}} \exp\{-\frac{1}{2} \langle S(Y)\rangle (u - \hat{X}_{\mathrm{MAP}}(y))^2\}, \tag{5.59}$$

where we used the expression (5.47) for the posterior PD in the third line. Using equations (5.29) and (5.56), we can calculate the integral over u by including terms up to the order of $\langle S(Y)\rangle^{-2}$ to give

$$\mathcal{E}_M = \int dy\, P(y) \left(\gamma + \frac{3\eta}{\langle S(Y)\rangle} \right)^{-1} \left\{ \frac{\gamma^3}{\langle S(Y)\rangle} + \frac{6\gamma\epsilon^2 + 15\gamma^2\eta}{\langle S(Y)\rangle^2} \right\}, \tag{5.60}$$

$$= \int dy\, P(y) \left(1 - \frac{3\eta}{\gamma \langle S(Y)\rangle} \right) \left\{ \frac{\gamma^2}{\langle S(Y)\rangle} + \frac{6\epsilon^2 + 15\gamma\eta}{\langle S(Y)\rangle^2} \right\}, \tag{5.61}$$

$$= \int dy P(y) \frac{1}{S(y)} + \int dy P(y) \frac{6\epsilon^2 + 12\gamma\eta}{\langle S(Y) \rangle^2}. \tag{5.62}$$

A detailed evaluation of $\mathcal{E}_M$ is given in Appendix C. Note that the fourth order correction ξ appears in the MMSE estimator, but not the MMSE itself. This is in agreement with the observation in (4.1) that the MSE of an estimator differs by a quadratic term. Hence, the higher order corrections from an estimator might not contribute to its MSE. Note that the Bayesian CRB is given by

$$-\left\{ \mathbb{E}\Big[\frac{\partial^2}{\partial x^2} \ln P(x,y) \Big] \right\}^{-1}, \tag{5.63}$$

where the expectation is over X and Y. The first term in the final expression for MMSE can be bounded below as

$$\int dy P(y) \frac{1}{S(y)} \geq \frac{1}{\langle S(Y) \rangle}, \tag{5.64}$$

which has a similar structure as the Bayesian CRB. The only difference is that a particular value of $X = \hat{X}_{\text{MAP}}(y)$ is being used in our calculations whereas the Bayesian CRB averages over the prior. For the high SNR case, we expect the two to be the same, and also equal to the deterministic CRB.

We will now evaluate the approximations developed in the previous sections to the MMSE at low and high SNR values for the TDE problem. We will consider both Gaussian and pseudo-Gaussian conditional data statistics.

5.3 TDE problem

5.3.1 The Gaussian channel

Consider the TDE problem with a Gaussian conditional data statistics (2.60),

$$P(\bar{y}|\tau) = \left(\frac{1}{\sqrt{2\pi\sigma_n^2}} \right)^{N_t} \exp\Big\{ -\frac{1}{2\sigma_n^2} \sum_{i=1}^{N_t} [y(t_i) - s(t_i - \tau)]^2 \Big\}, \tag{5.65}$$

First, we will calculate the low-SNR approximation, LSNR-MMSE, (5.22) for the TDE problem. To facilitate this calculation we need to first evaluate

$$J(\bar{Y}|\,\langle\tau\rangle) = -\mathbb{E}\Big[\frac{\partial^2 \ln P(\bar{Y}|\,\langle\tau\rangle)}{\partial\,\langle\tau\rangle^2}|\,\langle\tau\rangle\Big], \tag{5.66}$$

where

$$-\frac{\partial^2 \ln P(\bar{y}|\,\langle\tau\rangle)}{\partial\,\langle\tau\rangle^2} = \frac{1}{\sigma_n^2}\sum_{i=1}^{N_t}\Big\{[s(t_i-\langle\tau\rangle)-y_i]\frac{\partial^2 s(t_i-\langle\tau\rangle)}{\partial\,\langle\tau\rangle^2} + \Big(\frac{\partial s(t_i-\langle\tau\rangle)}{\partial\,\langle\tau\rangle}\Big)^2\Big\}. \tag{5.67}$$

Note that $P(y_i|\,\langle\tau\rangle)$ is Gaussian with mean $s(t_i-\langle\tau\rangle)$. So, the second order partial derivative term in the RHS of Eq. (5.67) vanishes when we calculate $J(\bar{Y}|\,\langle\tau\rangle)$. Therefore

$$J(\bar{Y}|\,\langle\tau\rangle) = \frac{1}{\sigma_n^2}\sum_{i=1}^{N_t}\Big(\frac{\partial s(t_i-\langle\tau\rangle)}{\partial\,\langle\tau\rangle}\Big)^2, \tag{5.68}$$

where we used the unit normalization of $P(\bar{Y}|\,\langle\tau\rangle)$. Substituting this $J(\bar{Y}|\,\langle\tau\rangle)$ into the expression (5.22) we can approximately calculate the LSNR-MMSE.

We will now calculate the high SNR approximation to the MMSE, given by Eq. (5.62). Even though this approximation is based on including higher order corrections, for this particular Gaussian channel case we will see that the first term of Eq. (5.62), which we will call HSNR-MMSE,

$$\text{HSNR-MMSE} = \mathbb{E}\Big[\frac{1}{S(Y)}\Big], \tag{5.69}$$

is sufficient to approximate the MMSE over a wide range of SNR values. From Eq. (5.26), we can calculate $S(y)$, by noting that the partial derivative of the uniform

prior w.r.t x is zero,

$$S(y) = -\frac{\partial^2 \ln P(\hat{\tau}_{\mathrm{MAP}}, \bar{y})}{\partial \hat{\tau}_{\mathrm{MAP}}^2} \tag{5.70}$$

$$= -\frac{\partial^2 \ln P(\bar{y}|\hat{\tau}_{\mathrm{MAP}})}{\partial \hat{\tau}_{\mathrm{MAP}}^2} \tag{5.71}$$

$$= \frac{1}{\sigma_n^2} \sum_{i=1}^{N_t} \left\{ [s(t_i - \hat{\tau}_{\mathrm{MAP}}) - y_i] \frac{\partial^2 s(t_i - \hat{\tau}_{\mathrm{MAP}})}{\partial \hat{\tau}_{\mathrm{MAP}}^2} + \Big(\frac{\partial s(t_i - \hat{\tau}_{\mathrm{MAP}})}{\partial \hat{\tau}_{\mathrm{MAP}}} \Big)^2 \right\}, \tag{5.72}$$

where $\hat{\tau}_{\mathrm{MAP}}$ is given by Eq. (2.64).

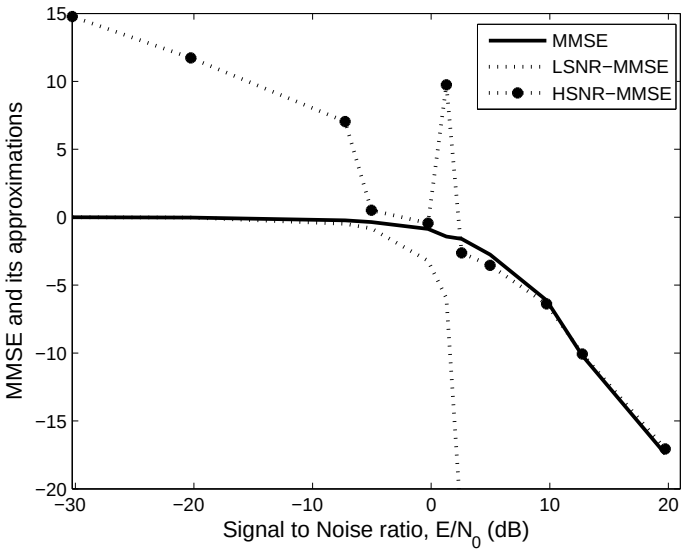

Figure 5.1: MMSE, and its approximations for the TDE problem with Gaussian conditional data statistics.

We calculated the HSNR-MMSE by simulating the MAP estimator. Figure 5.1 shows both the LSNR-MMSE and the HSNR-MMSE evaluated numerically. As we can see from Eq. (5.68), $J(\bar{Y}|\,\langle\tau\rangle)$ increases as the noise, represented by σ_n^2, decreases, and at some point the expression (5.22) even becomes negative. For such values of SNR, LSNR-MMSE is no longer a good approximation of the MMSE, as shown in

Fig. 5.1. On the other hand, the HSNR-MMSE seems to approximate the MMSE well even for lower values of the SNR.

5.3.2 Pseudo-Gaussian channel

In Sec. 4.6 we considered a pseudo-Gaussian channel, which approximates the joint statistics of Gaussian read-out noise and Poisson photon number distribution when imaging with photons. Now we will formulate the 1D-TDE problem with a pseudo-Gaussian conditional data statistics. The conditional PD of the data can be written as

$$P(\bar{y}|\tau) = \frac{1}{\prod_{i=1}^{N_t} \sqrt{2\pi v_i^2}} \exp\Big\{ -\sum_{i=1}^{N_t} \frac{[y(t_i) - s(t_i - \tau)\Delta t]^2}{2v_i^2} \Big\}, \tag{5.73}$$

where

$$v_i^2 = s(t_i - \tau)\Delta t + \sigma_n^2. \tag{5.74}$$

As we will see, the HSNR-MMSE is sufficient to approximate the MMSE even for the pseudo-Gaussian statistics. To calculate the LSNR-MMSE and HSNR-MMSE approximations to the MMSE, we need to first calculate $J(\bar{Y}|\tau)$, which depends on the second order partial derivative of the log-likelihood function w.r.t τ,

$$\ln P(\bar{y}|\tau) = -\frac{1}{2} \sum_{i}^{N_t} \Big\{ \ln(2\pi v_i^2 + \frac{[y_i - s(t_i - \tau)\Delta t]^2}{v_i^2}) \Big\}, \tag{5.75}$$

$$\Rightarrow \frac{\partial \ln P(\bar{y}|\tau)}{\partial \tau} = \frac{1}{2} \sum_{i=1}^{N_t} R \frac{\partial s(t_i - \tau)\Delta t}{\partial \tau}, \tag{5.76}$$

$$\Rightarrow \frac{\partial^2 \ln P(\bar{y}|\tau)}{\partial \tau^2} = \frac{1}{2} \sum_{i=1}^{N_t} \Big[R \frac{\partial^2 s(t_i - \tau)\Delta t}{\partial \tau^2} + Q\Big(\frac{\partial s(t_i - \tau)\Delta t}{\partial \tau} \Big)^2 \Big], \tag{5.77}$$

where

$$R = \frac{[s(t_i-\tau)\Delta t - y_i]^2 - 2v_i^2[s(t_i-\tau)\Delta t - y_i] - v_i^2}{v_i^4}, \tag{5.78}$$

$$Q = -\frac{2v_i^6 - v_i^4 - 4v_i^4[s(t_i-\tau)\Delta t - y_i] + 2v_i[s(t_i-\tau)\Delta t - y_i]^2}{v_i^8}. \tag{5.79}$$

We can now evaluate the LSNR-MMSE (5.22) by first calculating the expected value of the expression (5.77) evaluated at $\langle\tau\rangle$,

$$J(\bar{Y}|\langle\tau\rangle) = \mathbb{E}\left[-\frac{\partial^2 \ln P(\bar{y}|\langle\tau\rangle)}{\partial\langle\tau\rangle^2}\middle|\langle\tau\rangle\right] \tag{5.80}$$

$$= -\frac{1}{2}\sum_{i=1}^{N_t}\left[\mathbb{E}[R|\langle\tau\rangle]\frac{\partial^2 s(t_i-\tau)\Delta t}{\partial\tau^2} + \mathbb{E}[Q|\langle\tau\rangle]\Big(\frac{\partial s(t_i-\tau)\Delta t}{\partial\tau}\Big)^2\right]. \tag{5.81}$$

Note that $P(\bar{y}|\langle\tau\rangle)$ is a Gaussian PD with conditional mean $s(t_i-\tau)\Delta t$ and conditional variance v_i^2, so that

$$\mathbb{E}[R|\langle\tau\rangle] = 0 \tag{5.82}$$

$$\mathbb{E}[Q|\langle\tau\rangle] = -\frac{1+2v_i^2}{v_i^4}. \tag{5.83}$$

Thus, we have

$$J(\bar{Y}|\langle\tau\rangle) = \frac{1}{2}\sum_{i=1}^{N_t}\frac{1+2v_i^2}{v_i^4}\Big(\frac{\partial s(t_i-\tau)\Delta t}{\partial\tau}\Big)^2. \tag{5.84}$$

The LSNR-MMSE can then be calculated from Eq. (5.22). To calculate the HSNR-MMSE, we need to evaluate $S(y)$ (5.26). This is equal to negative of the expression (5.77) evaluated at the MAP estimator, $\tau = \hat{\tau}_{\rm MAP}$. The HSNR-MMSE (5.69) is then evaluated by simulating the MAP estimator.

The results are plotted in Fig. 5.2, which shows that the HSNR-MMSE involving the MAP estimator performs well for moderate to high values of SNR. Even though only the first term in Eq. (5.62) is included, the approximation to the MMSE is

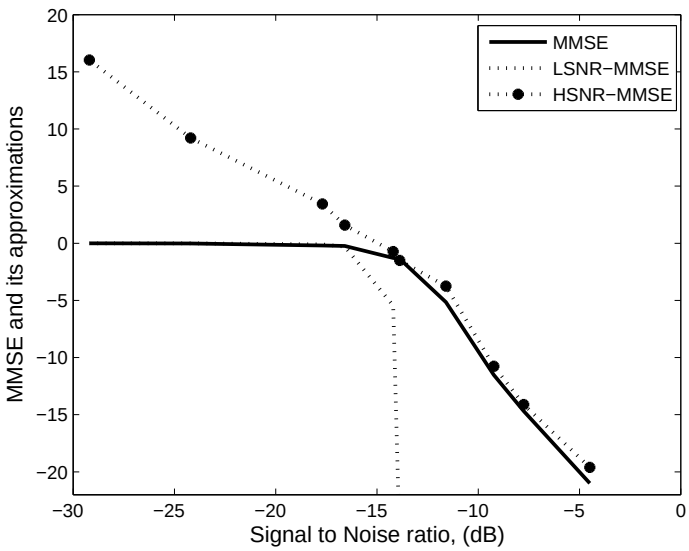

Figure 5.2: MMSE, and its approximations for the TDE problem with pseudo-Gaussian conditional data statistics.

remarkably tight. The LSNR-MMSE, which depends just on the second order sensitivity evaluated at the mean value of the prior, provides a loose approximation to the MMSE. This trend is observed both for the Gaussian and pseudo-Gaussian conditional data statistics. In the next section, we propose a new approximate expression for the MMSE which interpolates between the high SNR and the low SNR expressions for the MMSE, and is valid over a wide range of SNR values.

5.4 An approximate MMSE

The high SNR limit of the MMSE is $J(Y|x)$, as given by Eq. (5.7), which is obviously quite different from both the expressions (5.22) and (5.62). However, from the results of the previous sections, the approximate expressions, LSNR-MMSE (5.22) and HSNR-MMSE (5.69), have the necessary structure to approximate the MMSE in their regions of validity. Since the MAP estimator approximates the MMSE quite

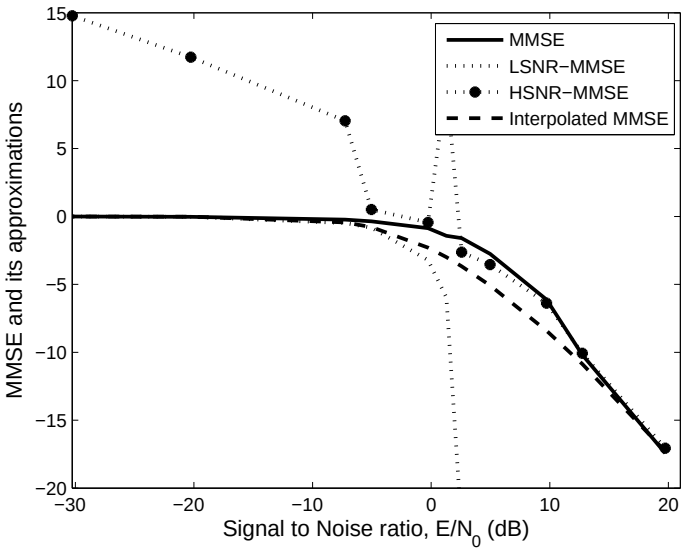

Figure 5.3: MMSE and its approximations for the TDE problem with Gaussian conditional data statistics.

well at high to moderate values of the SNR, and the MMSEE is the mean of the prior at extremely low values of the SNR, we may write down empirically an interpolated formula for the MMSE as follows:

$$\mathcal{E}_M = \frac{\sigma_X^2}{1 + \sigma_X^2 J(Y|\hat{X})}, \tag{5.85}$$

where $\hat{X}(Y)$ is our interpolating estimator

$$\hat{X}(Y) = \frac{[\mathrm{SNR}\hat{X}_{\mathrm{MAP}} + \mathbb{E}(X)]}{\mathrm{SNR} + 1}. \tag{5.86}$$

Note that the estimator (5.86) is exact for the fully Gaussian problem. The estimator, as desired, tends to the mean value of the prior at the low-SNR regime, $\sigma_X^2 J(Y|\hat{X}) << 1$, while it tends to the MAP estimator at the high-SNR regime, $\sigma_X^2 J(Y|\hat{X}) >> 1$. The FI term itself tends to $J(Y|\langle X \rangle)$ in the low-SNR regime and to $J(Y|\hat{X}_{\mathrm{MAP}})$ in the high-SNR regime. Thus, the interpolated MMSE (5.85), in the

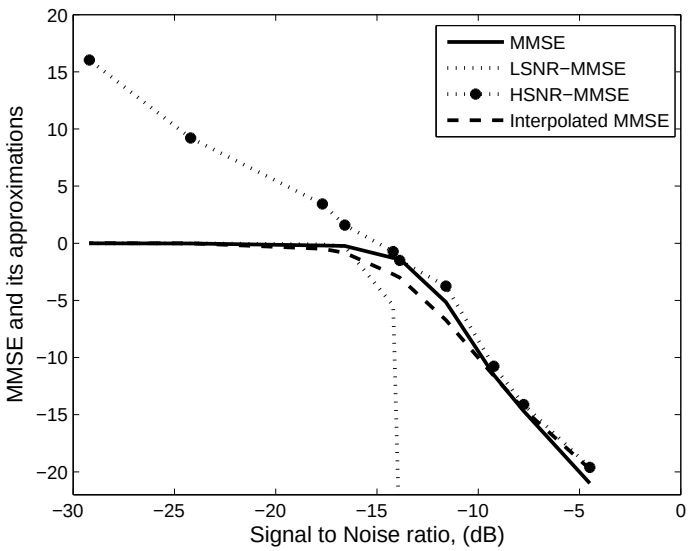

Figure 5.4: MMSE and its approximations for the TDE problem with pseudo-Gaussian conditional data statistics.

low-SNR regime, tends to

$$\frac{\sigma_X^2}{1+\sigma_X^2 J(Y|\,\langle X\rangle)} \approx \sigma_X^2[1-\sigma_X^2 J(Y|\,\langle X\rangle)], \tag{5.87}$$

which is the LSNR-MMSE (5.22), and in the high-SNR regime, tends to

$$\frac{\sigma_X^2}{1+\sigma_X^2 J(Y|\hat{X}_{\rm MAP})} \approx \frac{1}{J(Y|\hat{X}_{\rm MAP})}, \tag{5.88}$$

which is the CRB.

In Figs. 5.3 and 5.4, we plot, by means of dashed lines, this new interpolation formula for the Gaussian and pseudo-Gaussian conditional data statistics, respectively. The other curves in these figures are the same as those already presented in Figs. 5.1 and 5.2. The interpolation formula captures the behavior of the MMSE for all values of the SNR. Since expression (5.85) involves the data sensitivity function only to the lowest order, it may not be accurate in the intermediate-SNR regime where

$$\sigma_X^2 J(Y|\hat{X}) \sim 1 \tag{5.89}$$

and the MMSE is, in general, determined by a more detailed dependence of the data statistics $P(y|x)$ on the parameter X being estimated. This inadequacy of the interpolational formula is clearly seen in Figs. 5.3 and 5.4.

5.5 Summary

In this chapter, approximations to the MMSE were derived, at low values of the SNR using a perturbative calculation and at high values of the SNR using the Laplace method for the asymptotic evaluation of integrals. The results will be presented in [131] which is under preparation. The approximations were evaluated for Gaussian and pseudo-Gaussian conditional data statistics. Both the LSNR-MMSE and the HSNR-MMSE perform well in their respective regions of validity, with the HSNR-MMSE being remarkably tight in approximating the MMSE for a wide range of SNR values. An empirical interpolational formula for the MMSE was proposed, in terms of the mean of the prior and the MAP estimator, which captures the behavior of the MMSE at all values of the SNR. The interpolated formula, however, depends only on the second order sensitivity metric, the Fisher information, evaluated at the MAP estimator.

Chapter 6

Conclusions and Future Directions

In this dissertation, the Bayesian parameter estimation problem was considered anew, with our main focus being on deriving tighter performance bounds and optimal estimators. The fundamental lower bound on the MSE, namely the MMSE, and the corresponding optimum estimator, namely the MMSEE, are difficult to compute numerically, as they depend on the explicit evaluation of the posterior PD. In view of this computational complexity, *ad-hoc* estimators like the MLE and the MAP estimators are traditionally used. With the help of illustrative problems for which we can compute the MMSE and its bounds with relative ease, it was shown that the MAP estimator is not optimal in the MSE sense, particularly at moderate to low SNR values. It is for this reason, exacerbated by the inability of the MAP estimator to handle bias at low values of the SNR, the MAP estimation error departs from the MMSE as the SNR decreases. Even though for many problems, the MAP estimator and the MMSE estimator can be shown to be quite close [21] in the high SNR regime, they depart significantly from each other as we move into the moderate to low SNR regime.

To evaluate the performance of sub-optimal estimators, tight, computationally

efficient bounds on the MMSE are sought, which avoid any explicit calculation of the posterior distribution. Two popular lower bounds on the MMSE are the Ziv-Zakai bound (ZZB) and the Weiss-Weinsetin bound (WWB). Both the ZZB and the WWB provide such a formulation avoiding the posterior distribution while still making use of the prior information. Both these bounds can be applied to arbitrary prior distributions and to vector parameters. An attempt was made to relate the two bounds in [41], but there are no other results relating the two bounds analytically. These bounds were compared numerically for various problems of interest [94], and it is generally observed that the WWB is tighter in the low SNR region while the ZZB does well in the asymptotic region providing a better estimate of the threshold region [109]. Nevertheless, these two families of lower bounds are difficult to compute [70,94] and their tightness to the MMSE has not been established analytically. Moreover, optimal estimators achieving these bounds, when they are achievable, or even their achievability are not determined by their general theory.

As shown in this dissertation, these two lower bounds are not particularly tight at low to moderate values of the SNR. A particular case where the ZZB performs poorly is presented which illustrates the claim of sub-optimality of the ZZB. An extension to the ZZB using an M-ary hypothesis formulation is presented. Since the M-ary formulation, in general, is more comprehensive than the binary one, this extension is expected to improve the ZZB. The ZZB was also extended to discrete priors. A practical result for the ZZB was presented through an approximation to the MPE which involves a free parameter. Utilizing a Markov-Chain Monte Carlo (MCMC) approach, the integrals in this approximated ZZB can be evaluated tightly. The advantage of this approximation is that there is no need to calculate the different regions in the binary hypothesis testing problem.

An important result in this dissertation was deriving an upper bound on the MMSE. The so derived upper bound is numerically efficient to evaluate, and works

by a removal of the Bayesian bias from an estimator. Since the MSE of any estimator differs from the MMSE by a quadratic deviation of the estimator from the MMSE estimator, a variational approach that minimizes the upper bound over suitably chosen families of test estimators can provide an accurate approximation for the MMSE. We have considered a number of families of test estimators in this dissertation, and compared their performances in terms of the tightness of the optimal upper bounds on the MMSE that they provide. A generalization of upper bound for vector parameters was also presented.

The upper bound was considered for the polynomial class of estimators as well as for a linear class based on the MAP estimator. The MAP estimation error was shown, by a simple linear modification of the estimator, to saturate to values close to the MMSE in the moderate to low regions of SNR by a simple application of the upper bound. The variational upper bound involving the MAP estimator thus ties both the high SNR and the low SNR ends of the MSE to the value of the MMSE in those regions, and thus tightly bounds the MMSE for all values of the SNR. The polynomial class of estimators is of interest mainly for two reasons. One is that the optimization over the variational parameters can be performed analytically and the other is that the optimum values of the MSE and the optimum estimators can be found with minimal computational cost.

For certain problems, the polynomial class of estimators was shown to provide remarkably tight MMSE bounds that are almost indistinguishable from the MMSE over the entire range of the SNR values. The polynomial class of estimators was shown to be inadequate, however, for problems where the signal depends in a highly non-linear fashion on the data. A piecewise quasi-linear estimator, which is more useful in such situations, was presented. It accounts for any such non-linearities through certain data-dependent weighting functions. These weighting functions allow the variational parameters of the linear estimator to evolve from one subregion to the

next in the support space of the prior while accounting for the non-linear dependency of the signal on the prior over the full support of the prior.

The upper bounds were applied to both imaging and non-imaging problems. Our results showed that the piecewise quasi-linear estimator based upper bound and the MAP-VB tightly bound the MMSE for all values of SNR, and outperform the existing lower bounds on the MMSE, namely the ZZB and the WWB, in terms of their tightness. The MAP-VB was used to analyze and compare the performance of a newly developed rotating-PSF imager with respect to that of a conventional imager in localizing a point source beyond the diffraction limit. Based on our results, it was shown that the RPSF imager is quite robust to large changes of defocus, encoding defocus in PSF rotation, and thus providing a practical method for 3D snap-shot imaging.

Asymptotic expressions were constructed for the MMSE estimator and the MMSE for high and low values of the SNR. In the high-SNR region, since the MAP estimator is known to achieve the CRB and the MMSE, the MMSE estimator was expanded around the MAP estimator in this region. In the low-SNR region, where data add little information, the mean of the prior seems to be a suitable point about which to MMSE estimator was perturbatively approximated. Using this formalism a different kind of interpolational approximation to the MMSE was proposed, which captures the behavior of the MMSE for all values of the SNR.

A number of topics related to the upper bound require further research. We list the most important of these below.

- *Iteratively calculating the upper bound*

  By introducing a non-linear bias, one that is possibly dependent on the parameter, the upper bound can be evaluated iteratively. Such an iterative scheme can be implemented by performing bias-subtraction and upper bound minimization

sequentially in each iteration.

- *The sampled PQE*

  The calculation of PQE-VB involves inverting a matrix that involves a quadratic scaling in the product of number of data samples times the number of prior intervals. For high dimensional estimation this will pose a computational challenge. For such problems, more sophisticated methods can be formulated which rely on sampling from a limited number of smaller subregions into which the larger support of the prior may be regarded as divided.

- *Comparative studies of the PQE-VB and the MAP-VB*

  The MAP-VB when used in an expectation-maximization (EM) formulation, might provide a viable approach for iterative algorithms. However, such algorithms can be computationally inefficient. For active systems, an estimator that can give the estimate in a single shot greatly improves the temporal resolution, potentially providing real-time performance. In such cases, the PQE presents a powerful formalism to address the temporal resolution. A comparative study can be done with regard to the computational overhead for evaluating the MAP-VB versus the upper bound of the PQE.

- *Comprehensive 3D performance analysis of the RPSF*

  In this dissertation, we demonstrated the robustness of a rotating-PSF based imager in terms of the MMSE for localizing a point source to sub-diffractive errors in the transverse plane. The axial localization problem too can be analyzed along with a full 3D performance analysis of longitudinal and transverse resolution of the RPSF, based on the MMSE.

- *Mutual information and the MMSE*

  A direct relationship exists between mutual information and the MMSE for certain Gaussian problems [65]. Using tight approximations to the MMSE

and thus bounding mutual information evaluations is of much importance to characterize fidelity of various communication systems. In general, for all the problems that involve an accurate evaluation of the MMSE, the upper bound presents a tight, and numerically viable approximation.

Appendices

Appendix A

Poisson channel and exponential prior

A.1 $P(y)$ for the single-pixel problem

We will evaluate data PD given in (2.39),

$$P(y) = \int_0^\infty \frac{dx}{\langle X \rangle} \frac{(ax+b)^y}{y!} \exp[-(ax+b)] \exp(-x/\langle X \rangle) \tag{A.1}$$

$$= \frac{1}{y! \langle X \rangle} \int_0^\infty (ax+b)^y \exp[-(ax + \frac{x}{\langle X \rangle} + b)] \tag{A.2}$$

$$= \frac{1}{y! a \langle X \rangle} \exp\left(\frac{b}{a \langle X \rangle}\right) \int_b^\infty du\, u^y \exp\left[-u \frac{1 + a \langle X \rangle}{a \langle X \rangle}\right] \tag{A.3}$$

$$= \frac{1}{y!} \frac{(a \langle X \rangle)^y}{(1 + a \langle X \rangle)^{y+1}} \exp\left(\frac{b}{a \langle X \rangle}\right) \Gamma\left[y+1, \frac{b(1 + a \langle X \rangle)}{a \langle X \rangle}\right], \tag{A.4}$$

where

$$\Gamma(y+1, z) = \int_z^\infty dx\, x^y \exp(-x). \tag{A.5}$$

To arrive at (A.3), a change of integration variable $u = ax + b$ was used. To arrive at (A.4), a change of variable $\frac{1+a\langle X\rangle}{a\langle X\rangle}u = x$ was used.

Note that $P(y)$ can be calculated iteratively from the following relation,

$$\Gamma(y+1,z) = \int_z^\infty dx\, x^y \exp(-x), \tag{A.6}$$

$$= -x^y \exp(-x)\big|_z^\infty + y\int_z^\infty dx\, x^y \exp(-x), \tag{A.7}$$

$$= z^y \exp(-z) + y\Gamma(y,z). \tag{A.8}$$

A.2 Two-pixel problem

The conditional data statistics for the two-pixel problem are given by,

$$P(y_1,y_2|x_1,x_2) = P(y_1|x_1,x_2)P(y_2|x_1,x_2), \tag{A.9}$$

$$P(y_1|x_1,x_2) = \frac{(\alpha_1 x_1 + \beta_1 x_2 + \gamma_1)^{y_1}}{y_1!}\exp[-(\alpha_1 x_1 + \beta_1 x_2 + \gamma_1)], \quad y_1 = 0,1,\ldots, \tag{A.10}$$

$$P(y_2|x_1,x_2) = \frac{(\alpha_2 x_1 + \beta_2 x_2 + \gamma_2)^{y_2}}{y_2!}\exp[-(\alpha_2 x_1 + \beta_2 x_2 + \gamma_2)], \quad y_2 = 0,1,\ldots. \tag{A.11}$$

We assume X_1 and X_2 are independent and exponentially distributed random variables,

$$P_{X_i}(x) = \begin{cases} \frac{1}{\langle X_i\rangle}\exp(-x/\langle X_i\rangle) & \text{for } x > 0 \\ 0 & \text{otherwise,} \end{cases} \tag{A.12}$$

for $i = 1,2$ and $P_X(x_1,x_2) = P_{X_1}(x_1)P_{X_2}(x_2)$. We can now express the MMSE estimators, $\hat{X}_{1M}$ ans $\hat{X}_{2M}$ as

$$\hat{X}_{1M} = \int_0^\infty\int_0^\infty dx_1\, dx_2\, x_1 P(x_1)P(x_2)P(y_1|x_1,x_2)P(y_2|x_1,x_2) \tag{A.13}$$

$$\hat{X}_{2M} = \int_0^\infty\int_0^\infty dx_1\, dx_2\, x_2 P(x_1)P(x_2)P(y_1|x_1,x_2)P(y_2|x_1,x_2) \tag{A.14}$$

Appendix A. Poisson channel and exponential prior

Let us first consider the evaluation of $\hat{X}_{1M}$. By expressing x_1 as a linear combination of $\alpha_1 x_1 + \beta_1 x_2 + \gamma_1$ and $\alpha_2 x_1 + \beta_2 x_2 + \gamma_2$, we can express $\hat{X}_{1M}$ in terms of $P(y_1, y_2)$,

$$x_1 = k_1(\alpha_1 x_1 + \beta_1 x_2 + \gamma_1) + l_1(\alpha_2 x_1 + \beta_2 x_2 + \gamma_2) + m_1, \tag{A.15}$$

where

$$k_1 = \frac{\beta_2}{\alpha_1\beta_2 - \alpha_2\beta_1} \tag{A.16}$$

$$l_1 = -\frac{\beta_1}{\alpha_1\beta_2 - \alpha_2\beta_1} \tag{A.17}$$

$$m_1 = -k_1\gamma_1 - l_1\gamma_2 \tag{A.18}$$

We can then write down $\hat{X}_{1M}$ as

$$\hat{X}_{1M} = \frac{K_1(y_1, y_2)}{P(y_1, y_2)}, \tag{A.19}$$

where

$$K_1(y_1, y_2) = k_1(y+1)P(y_1+1, y_2) + l_1(y_2+1)P(y_1, y_2+1) + m_1 P(y_1, y_2). \tag{A.20}$$

Similarly, we can express $\hat{X}_{2M}$ as

$$\hat{X}_{2M} = \frac{K_2(y_1, y_2)}{P(y_1, y_2)}, \tag{A.21}$$

where

$$K_2(y_1, y_2) = k_2(y+1)P(y_1+1, y_2) + l_2(y_2+1)P(y_1, y_2+1) + m_2 P(y_1, y_2), \tag{A.22}$$

where

$$k_2 = -\frac{\alpha_1}{\alpha_1\beta_2 - \alpha_2\beta_1} \tag{A.23}$$

$$l_2 = \frac{\alpha_2}{\alpha_1\beta_2 - \alpha_2\beta_1} \tag{A.24}$$

$$m_2 = -k_2\gamma_1 - l_2\gamma_2 \tag{A.25}$$

The MMSE can be written as

$$\mathcal{E}_M = \langle X_1^2 \rangle + \langle X_2^2 \rangle - \langle \hat{X}_1^2 \rangle - \langle \hat{X}_2^2 \rangle \,, \tag{A.26}$$

$$= \langle X_1^2 \rangle + \langle X_2^2 \rangle - \sum_{y_1, y_2} \frac{K_1^2(y_1, y_2) + K_2^2(y_1, y_2)}{P(y_1, y_2)}. \tag{A.27}$$

Thus, we can see that if we can evaluate $P(y_1, y_2)$ iteratively, both the MMSEE and the MMSE can be calculated. Note that we used $\mathbf{R}$ to be an identity matrix in defining the MMSE (4.18) for multi-variate parameters.

We will now consider the case when α is the on-axis signal strength, β is the cross-talk signal strength, and γ is a uniform bias. That is $\alpha_1 = \beta_2 = \alpha$, $\alpha_2 = \beta_1 = \beta$ and $\gamma_1 = \gamma_2 = \gamma$. To write down $P(y_1, y_2)$ iteratively, we will first do a change of variables,

$$\phi_1 = \alpha x_1 + \beta x_2, \tag{A.28}$$

$$\phi_2 = \beta x_1 + \alpha x_2. \tag{A.29}$$

This will transform the prior PD $P_X(x_1, x_2)$ to $P_\phi(\phi_1, \phi_2)$,

$$P_\phi(\phi_1, \phi_2) = \frac{1}{\alpha^2 - \beta^2} P_X\left(\frac{\alpha\phi_1 - \beta\phi_2}{\alpha^2 - \beta^2}, \frac{\alpha\phi_2 - \beta\phi_1}{\alpha^2 - \beta^2} \right) \tag{A.30}$$

$$= \frac{1}{\alpha^2 - \beta^2} \frac{1}{\langle X_1 \rangle \langle X_2 \rangle} \exp\left(-\frac{\phi_1}{\langle \phi_1 \rangle} - \frac{\phi_2}{\langle \phi_2 \rangle} \right), \tag{A.31}$$

where

$$\langle \phi_1 \rangle = (\alpha^2 - \beta^2) \left(\frac{\alpha}{\langle X_1 \rangle} - \frac{\beta}{\langle X_2 \rangle} \right)^{-1}, \tag{A.32}$$

$$\langle \phi_2 \rangle = (\alpha^2 - \beta^2) \left(\frac{\alpha}{\langle X_2 \rangle} - \frac{\beta}{\langle X_1 \rangle} \right)^{-1}. \tag{A.33}$$

The region of support for (ϕ_1, ϕ_2) is

$$R_\phi = \{(\phi_1, \phi_2) | \frac{\beta}{\alpha}\phi_2 \leq \phi_1 \leq \frac{\alpha}{\beta}\phi_2\}, \tag{A.34}$$

as shown in Fig. A.1.

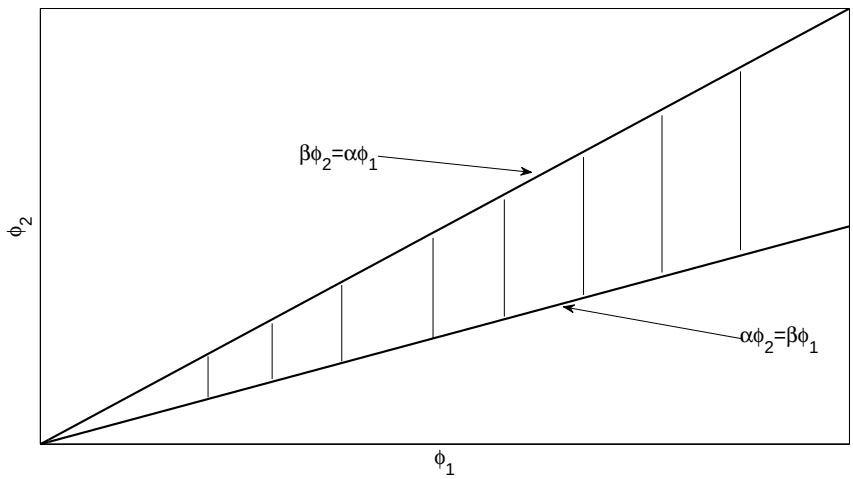

Figure A.1: Region of support for (ϕ_1, ϕ_2) is the shaded region between the lines $\alpha\phi_1 = \beta\phi_2$ and $\alpha\phi_2 = \beta\phi_1$.

We can now express the joint PD $P(y_1, y_2)$ as,

$$P(y_1, y_2) = \iint\limits_{R_\phi} d\phi_1\, d\phi_2 \frac{(\phi_1+\gamma)^{y_1}}{y_1!} \exp(-\phi_1-\gamma) \frac{(\phi_2+\gamma)^{y_2}}{y_2!}$$

$$\exp(-\phi_2-\gamma) P_\phi(\phi_1, \phi_2), \tag{A.35}$$

$$= \frac{1}{\alpha^2-\beta^2} \frac{1}{\langle\phi_1\rangle \langle\phi_2\rangle} \int_0^\infty d\phi_2\, \frac{(\phi_2+\gamma)^{y_2}}{y_2!} \exp\left(-\phi_2 - \frac{\phi_2}{\langle\phi_2\rangle} - \gamma\right)$$

$$\int_{\frac{\beta}{\alpha}\phi_2}^{\frac{\alpha}{\beta}\phi_2} d\phi_1\, \frac{(\phi_1+\gamma)^{y_1}}{y_1!} \exp\left(-\phi_1 - \frac{\phi_1}{\langle\phi_1\rangle} - \gamma\right), \tag{A.36}$$

$$= \frac{n}{y_1! y_2!} \int_0^\infty d\phi_2\, (\phi_2+\gamma)^{y_2} \exp\big(-v_2\phi_2\big)$$

$$\int_{\frac{\beta}{\alpha}\phi_2}^{\frac{\alpha}{\beta}\phi_2} d\phi_1\, (\phi_1+\gamma)^{y_1} \exp\big(-v_1\phi_1\big), \tag{A.37}$$

where $v_i = 1 + \frac{1}{\langle\phi_i\rangle}$, for $i = 1, 2$ and where $n = \exp(-2\gamma)[\langle\phi_1\rangle \langle\phi_2\rangle (\alpha^2-\beta^2)^{-1}]$.

$P(y_1+1, y_2)$ is given by,

$$P(y_1+1, y_2) = \frac{n}{(y_1+1)! y_2!} \int_0^\infty d\phi_2\, (\phi_2+\gamma)^{y_2} \exp\big(-v_2\phi_2\big)$$

$$\int_{\frac{\beta}{\alpha}\phi_2}^{\frac{\alpha}{\beta}\phi_2} d\phi_1 \, (\phi_1+\gamma)^{y_1+1} \exp\bigl(-v_1\phi_1\bigr). \tag{A.38}$$

Consider the integral over ϕ_1, where we will use integration by parts to express the integral over ϕ_1 as sum of two terms, one of which is $P(y_1, y_2)$. Specifically,

$$\begin{aligned}\int_{\frac{\beta}{\alpha}\phi_2}^{\frac{\alpha}{\beta}\phi_2} d\phi_1 \, (\phi_1+\gamma)^{y_1+1} \exp\bigl(-v_1\phi_1\bigr) = \frac{-(\phi_1+\gamma)^{y_1+1}\exp\bigl(-v_1\phi_1\bigr)}{v_1}\Bigg|_{\frac{\beta}{\alpha}\phi_2}^{\frac{\alpha}{\beta}\phi_2} \\ + \int_{\frac{\beta}{\alpha}\phi_2}^{\frac{\alpha}{\beta}\phi_2} d\phi_1 \, (y_1+1)(\phi_1+\gamma)^{y_1}\frac{\exp\bigl(-v_1\phi_1\bigr)}{v_1}.\end{aligned} \tag{A.39}$$

Note that the first term of the RHS is entirely a function of ϕ_2 and the second term contributes to $P(y_1, y_2)$ yielding,

$$P(y_1+1, y_2) = \frac{1}{v_1}[P(y_1, y_2) + I_1], \tag{A.40}$$

where

$$\begin{aligned}I_1 = \frac{n}{(y_1+1)!y_2!}\Bigg\{\int_0^\infty d\phi_2(\phi_2+\gamma)^{y_2}(\frac{\beta}{\alpha}\phi_2+\gamma)^{y_1+1}\exp(-v_2\phi_2-\frac{\beta}{\alpha}v_1\phi_2) \\ - \int_0^\infty d\phi_2(\phi_2+\gamma)^{y_2}(\frac{\alpha}{\beta}\phi_2+\gamma)^{y_1+1}\exp(-v_2\phi_2-\frac{\alpha}{\beta}v_1\phi_2)\Bigg\}.\end{aligned} \tag{A.41}$$

Similarly, we can express $P(y_1, y_2+1)$ as,

$$P(y_1, y_2+1) = \frac{1}{v_2}[P(y_1, y_2) + I_2], \tag{A.42}$$

where

$$\begin{aligned}I_2 = \frac{n}{y_1!(y_2+1)!}\Bigg\{\int_0^\infty d\phi_1(\phi_1+\gamma)^{y_1}(\frac{\beta}{\alpha}\phi_1+\gamma)^{y_2+1}\exp(-v_1\phi_1-\frac{\beta}{\alpha}v_2\phi_1) \\ - \int_0^\infty d\phi_1(\phi_1+\gamma)^{y_1}(\frac{\alpha}{\beta}\phi_1+\gamma)^{y_2+1}\exp(-v_1\phi_1-\frac{\alpha}{\beta}v_2\phi_1)\Bigg\}.\end{aligned} \tag{A.43}$$

Using the iterative expressions for $P(y_1+1, y_2)$ and $P(y_1, y_2+1)$, we can evaluate $K_1(y_1, y_2)$ and $K_2(y_1, y_2)$ from (A.20) (A.20) and (A.22). Then the MMSE can be evaluated from (A.26). Note that, the integrals I_1 and I_2 will be evaluated numerically.

On the other hand, a quadratic upper bound can be formulated which can be evaluated analytically with far less computational complexity. Note that the MMSE (A.26) can be written as

$$\mathcal{E}_M = \mathcal{E}_{1M} + \mathcal{E}_{2M}, \tag{A.44}$$

which is a sum of the MMSEs for the two sensors pixels. Thus, the corresponding upper bound can be found from (4.26),

$$\mathcal{E}_{UB} = \mathcal{E}_{1UB} + \mathcal{E}_{2UB}, \tag{A.45}$$

$$= \mathbb{E}\big[(\delta\hat{X}_1 - \delta X_1)^2\big] + \mathbb{E}\big[(\delta\hat{X}_2 - \delta X_2)^2\big]. \tag{A.46}$$

Let the quadratic estimator be written as

$$\hat{X}_k = \bar{b}'_{(k)}\bar{u}, \text{ for } k = 1, 2 \tag{A.47}$$

where $\bar{u}$ is the vector $[y_1\ y_2\ y_1^2\ y_1y_2\ y_2^2]$ and the elements of $\bar{b}_{(k)}$ are the variational parameters over which will minimize the upper bound. Note that the minimization can be performed independently over each of the terms in (A.46). This amounts to finding the optimal $\bar{b}_{(1)}$ and $\bar{b}_{(2)}$ independently. This can be done along the same lines from (4.28) to (4.30). After finding the optimal $\bar{b}_{(1)}$ and $\bar{b}_{(2)}$ and substituting them back into the upper bound we arrive at,

$$\mathcal{E}_{1UB} = \sigma^2_{X_1} - \bar{v}^T_{(1)}\mathbf{M}^{-1}\bar{v}_{(1)}, \tag{A.48}$$

$$\mathcal{E}_{2UB} = \sigma^2_{X_2} - \bar{v}^T_{(2)}\mathbf{M}^{-1}\bar{v}_{(2)} \tag{A.49}$$

where $\mathbf{M}$ a 5×5 symmetric matrix with elements, $M_{ij} = \mathbb{E}[\delta u_i \delta u_j]$, $i, j = 1, \ldots, 5$, and $\bar{v}_{(k)}$ a 5×1 vector of elements $\mathbb{E}[\delta X_k\, \delta\bar{u}]$.

To evaluate the upper bound, we need to calculate terms like $\langle y_1^i \rangle$, $\langle y_1^i y_2^j \rangle$, $\langle X_k y_1^i \rangle$, and $\langle X_k y_2^i \rangle$ for $k = 1, 2$. All of them can be computed analytically very easily in terms of $\langle X_1 \rangle$, $\langle X_2 \rangle$, α, β, and γ. We will show an outline of the calculation of one of the terms, e.g., $\mathbb{E}[Y_1^2 Y_2]$,

$$\mathbb{E}[Y_1^2 Y_2] = \mathbb{E}_X \big[\mathbb{E}_{Y|X}[Y_1^2 Y_2] \big], \tag{A.50}$$

$$= \mathbb{E}_X \Big[\big[(\alpha X_1 + \beta X_2 + \gamma)^2 + (\alpha X_1 + \beta X_2 + \gamma) \big] (\beta X_1 + \alpha X_2 + \gamma) \Big], \tag{A.51}$$

where we used the expressions for the moments of a Poisson channel in terms of its mean. $\mathbb{E}[Y_1^2 Y_2]$ can then be written in terms of $\langle X_k^i \rangle$ and $\langle X_1^i X_2^j \rangle$. The latter can be written as $\langle X_1^i \rangle \langle X_2^j \rangle$ since X_1 and X_2 are independent of each other. For an exponential random variable X the higher order moments can be written in terms of its mean $\langle X^i \rangle = i! \langle X \rangle$. Similarly, all other expectations can be expressed in terms of the moments of X_1 and X_2, which are trivial to compute.

Hence, we can evaluate the upper bound for a quadratic test estimator, which has minimal computational complexity as there is no explicit evaluation of $P(y_1, y_2)$.

Appendix B

The various expectations in the optimal coefficients of the PQE

The vector and matrices in (4.60) can be calculated by first evaluating the following expectations, $\mathbb{E}[e_\mu]$, $\mathbb{E}[y_i e_\mu]$, $\mathbb{E}[\tau e_\mu]$, $\mathbb{E}[\tau y_i e_\mu]$, $\mathbb{E}[e_\mu e_\nu]$, $\mathbb{E}[y_i e_\mu e_\nu]$, $\mathbb{E}[y_i y_j e_\mu e_\nu]$. The conditional data statistics are Gaussian and given by Eq. (2.60), reproduced here

$$P(\bar{y}|\tau) = \left(\frac{1}{\sqrt{2\pi\sigma_n^2}}\right)^{N_t} \exp\left(-\frac{\|\bar{y} - \bar{s}(\tau)\|^2}{2\sigma_n^2}\right), \tag{B.1}$$

where

$$\|\bar{y} - \bar{s}(\tau)\|^2 = \sum_{i=1}^{N_t} \big[y_i - s_i(\tau)\big]^2, \tag{B.2}$$

$y_i = y(t_i)$, and $s_i(\tau) = s(t_i - \tau)$. We considered the following weighting function, as defined in Eq. (4.47):

$$e_\mu(\bar{y}) = \exp\left(-\frac{\|\bar{y} - \bar{s}(\tau_\mu)\|^2}{2w_\mu^2}\right). \tag{B.3}$$

Now, we will evaluate each of the expected values.

$$\mathbb{E}[e_\mu] = \left(\frac{1}{\sqrt{2\pi\sigma_n^2}}\right)^{N_t} \int d\tau P(\tau) \int d\bar{y} \exp\left(-\frac{\|\bar{y} - \bar{s}(\tau)\|^2}{2\sigma_n^2} - \frac{\|\bar{y} - \bar{s}(\tau_\mu)\|^2}{2w_\mu}\right), \tag{B.4}$$

where

$$\int d\bar{y} = \int \ldots \int dy_1 \ldots dy_{N_t}, \tag{B.5}$$

is an N_t dimensional multivariate integral. Note that we can complete squares for $\bar{y}$ as

$$\frac{\|\bar{y} - \bar{s}(\tau)\|^2}{2\sigma_n^2} + \frac{\|\bar{y} - \bar{s}(\tau_\mu)\|^2}{2w_\mu^2} = \frac{\|\bar{y} - \bar{m}[\tau, \bar{s}(\tau_\mu), w_\mu]\|^2}{2\sigma^2(w_\mu)} + \frac{\|\bar{s}(\tau) - \bar{s}(\tau_\mu)\|^2}{2(w_\mu^2 + \sigma_n^2)}, \tag{B.6}$$

where

$$\bar{m}[\tau, \bar{r}, w] = \frac{\bar{s}(\tau)w^2 + \bar{r}\sigma_n^2}{w^2 + \sigma_n^2} \text{ and} \tag{B.7}$$

$$\sigma^2(w) = \frac{\sigma_n^2 w^2}{\sigma_n^2 + w^2}, \tag{B.8}$$

for some vector $\bar{r}$ of N_t element vector and real number w. Thus, $\mathbb{E}[e_\mu]$ can be written as

$$\begin{aligned}\mathbb{E}[e_\mu] = \mathbb{E}_\tau \Bigg[& \exp\left(-\frac{\|\bar{s}(\tau) - \bar{s}(\tau_\mu)\|^2}{2(w_\mu^2 + \sigma_n^2)}\right) \\ & \left(\frac{1}{\sqrt{2\pi\sigma_n^2}}\right)^{N_t} \int d\bar{y} \exp\left(-\frac{\|\bar{y} - \bar{m}[\tau, \bar{s}(\tau_\mu), w_\mu]\|^2}{2\sigma^2(w_\mu)}\right)\Bigg] \end{aligned} \tag{B.9}$$

$$= \mathbb{E}_\tau \big[f[\tau, \bar{s}(\tau_\mu), w_\mu] \big], \tag{B.10}$$

where

$$f[\tau, \bar{r}, w] = \exp\left(-\frac{\|\bar{s}(\tau) - \bar{r}\|^2}{2(w^2 + \sigma_n^2)}\right)\left(\frac{\sigma(w)}{\sigma_n}\right)^{N_t}, \tag{B.11}$$

for some vector $\bar{r}$ of N_t element vector and real number w. We used the unit normalization of the integral in Eq. (B.9) after multiplying and dividing by $[2\pi\sigma^2(w_\mu)]^{N_t/2}$. Note that $\mathbb{E}[\tau e_\mu]$ involves the same kind of calculation, and can be evaluated as,

$$\mathbb{E}[\tau e_\mu] = \mathbb{E}_\tau \big[\tau f[\tau, \bar{s}(\tau_\mu), w_\mu] \big], \tag{B.12}$$

Appendix B. The various expectations in the optimal coefficients of the PQE

Evaluating $\mathbb{E}[y_i e_\mu]$,

$$\mathbb{E}[y_i e_\mu] = \mathbb{E}_\tau\big[\mathbb{E}_{\bar{y}|\tau}[y_i e_\mu]\big] \tag{B.13}$$

$$= \mathbb{E}_\tau\Bigg[\exp\bigg(-\frac{\|\bar{s}(\tau) - \bar{s}(\tau_\mu)\|^2}{2(w_\mu^2 + \sigma_n^2)}\bigg) \bigg(\frac{1}{\sqrt{2\pi\sigma_n^2}}\bigg)^{N_t} \int d\bar{y}\, y_i \exp\bigg(-\frac{\|\bar{y} - \bar{m}[\tau, \bar{s}(\tau_\mu), w_\mu]\|^2}{2\sigma^2(w_\mu)}\bigg)\Bigg], \tag{B.14}$$

$$= \mathbb{E}_\tau\Bigg[\frac{f[\tau, \bar{s}(\tau_\mu), w_\mu]}{\sqrt{2\pi\sigma^2(w_\mu)}} \int dy_i\, y_i \exp\bigg(-\frac{\|\bar{y} - \bar{m}[\tau, \bar{s}(\tau_\mu), w_\mu]\|^2}{2\sigma^2(w_\mu)}\bigg)\Bigg], \tag{B.15}$$

$$= \mathbb{E}_\tau\big[f[\tau, \bar{s}(\tau_\mu), w_\mu] m_i[\tau, \bar{s}(\tau_\mu), w_\mu]\big] \tag{B.16}$$

$$= \frac{h_i[\tau, \bar{s}(\tau_\mu), w_\mu] w_\mu^2 + \mathbb{E}[e_\mu] s_i(\tau_\mu)\sigma_n^2}{w_\mu^2 + \sigma_n^2}, \tag{B.17}$$

where

$$h_i[\tau, \bar{r}, w] = \mathbb{E}_\tau\big[s_i(\tau) f[\tau, \bar{r}, w]\big]. \tag{B.18}$$

We can evaluate $\mathbb{E}[\tau y_i e_\mu]$ in a similar fashion, to give

$$\mathbb{E}[\tau y_i e_\mu] = \frac{\eta_i[\tau, \bar{s}(\tau_\mu), w_\mu] w_\mu^2 + \mathbb{E}[\tau e_\mu] s_i(\tau_\mu)\sigma_n^2}{w_\mu^2 + \sigma_n^2}, \tag{B.19}$$

where

$$\eta_i[\tau, \bar{r}, w] = \mathbb{E}_\tau\big[\tau s_i(\tau) f[\tau, \bar{r}, w]\big]. \tag{B.20}$$

To evaluate the remaining expectations $\mathbb{E}[e_\mu e_\nu]$, $\mathbb{E}[y_i e_\mu e_\nu]$, and $\mathbb{E}[y_i y_j e_\mu e_\nu]$, note that

$$\frac{\|\bar{y} - \bar{s}(\tau_\mu)\|^2}{2w_\mu^2} + \frac{\|\bar{y} - \bar{s}(\tau_\nu)\|^2}{2w_\nu} = \frac{\|\bar{y} - \bar{\tilde{s}}\|^2}{2w_{\mu\nu}^2} + \frac{\|\bar{s}(\tau_\mu) - \bar{s}(\tau_\nu)\|^2}{2(w_\mu^2 + w_\nu^2)}, \tag{B.21}$$

where

$$\tilde{s}_i = \frac{s_i(\tau_\mu) w_\nu^2 + s_i(\tau_\nu) w_\nu^2}{w_\mu^2 + w_\nu^2} \text{ and} \tag{B.22}$$

$$w_{\mu\nu}^2 = \frac{w_\mu^2 w_\nu^2}{w_\mu^2 + w_\nu^2}. \tag{B.23}$$

Appendix B. The various expectations in the optimal coefficients of the PQE

We can now express $\mathbb{E}[e_\mu e_\nu]$ as

$$\mathbb{E}[e_\mu e_\nu] = \mathbb{E}\left[\exp\left(-\frac{\|\bar{y}-\bar{s}(\tau_\mu)\|^2}{2w_\mu^2} - \frac{\|\bar{y}-\bar{s}(\tau_\nu)\|^2}{2w_\nu^2}\right)\right] \tag{B.24}$$

$$= \mathbb{E}\left[\exp\left(-\frac{\|\bar{y}-\bar{\bar{s}}\|^2}{2w_{\mu\nu}^2} - \frac{\|\bar{s}(\tau_\mu)-\bar{s}(\tau_\nu)\|^2}{2(w_\mu^2+w_\nu^2)}\right)\right] \tag{B.25}$$

$$= d\,\mathbb{E}\big[e_{\mu\nu}\big], \tag{B.26}$$

where

$$e_{\mu\nu} = \exp\left(-\frac{\|\bar{y}-\bar{\bar{s}}\|^2}{2w_{\mu\nu}^2}\right), \tag{B.27}$$

$$d = \exp\left(-\frac{\|\bar{s}(\tau_\mu)-\bar{s}(\tau_\nu)\|^2}{2(w_\mu^2+w_\nu^2)}\right). \tag{B.28}$$

We can evaluate $\mathbb{E}[e_{\mu\nu}]$ by following similar steps from (B.4) to (B.10) to yield

$$\mathbb{E}[e_{\mu\nu}] = \mathbb{E}_\tau[f[\tau, \bar{\bar{s}}, w_{\mu\nu}]], \tag{B.29}$$

from which we can evaluate $\mathbb{E}[e_\mu e_\nu]$ using Eq. (B.26).

To evaluate $\mathbb{E}[y_i e_\mu e_\nu]$, , we first observe that

$$\mathbb{E}[y_i e_\mu e_\nu] = d\,\mathbb{E}\big[y_i e_{\mu\nu}\big]. \tag{B.30}$$

The evaluation of $\mathbb{E}\big[y_i e_{\mu\nu}\big]$ follows similar steps from (B.13) to (B.17) to yield

$$\mathbb{E}[y_i e_{\mu\nu}] = \frac{h_i[\tau, \bar{\bar{s}}, w_{\mu\nu}]w_{\mu\nu}^2 + \mathbb{E}[e_{\mu\nu}]\tilde{s}_i\sigma_n^2}{w_{\mu\nu}^2+\sigma_n^2}, \tag{B.31}$$

from which we can evaluate $\mathbb{E}[y_i e_\mu e_\nu]$ using Eq. (B.30).

To evaluate $\mathbb{E}[y_i y_j e_\mu e_\nu]$ we will first express it as

$$\mathbb{E}[y_i y_j e_\mu e_\nu] = d\,\mathbb{E}\big[y_i y_j e_{\mu\nu}\big]. \tag{B.32}$$

Note that $\mathbb{E}\big[y_i y_j e_{\mu\nu}\big]$ involves a joint conditional expectation of $y_i y_j$. For all $i \neq j$ the conditional expectation is a product of individual expectations. For $i = j$, the

conditional expectation can be written as a sum of variance and conditional mean square. We can then write $\mathbb{E}\big[y_i y_j e_{\mu\nu}\big]$ as

$$\mathbb{E}\big[y_i y_j e_{\mu\nu}\big] = \mathbb{E}_\tau\big[\mathbb{E}_{\bar{y}|\tau}[y_i y_j e_{\mu\nu}]\big] \tag{B.33}$$

$$= \mathbb{E}_\tau\bigg[f[\tau,\bar{\tilde{s}},w_{\mu\nu}]\Big\{m_i[\tau,\bar{\tilde{s}},w_{\mu\nu}]m_j[\tau,\bar{\tilde{s}},w_{\mu\nu}]+\delta_{i,j}\sigma^2(w_{\mu\nu})\Big\}\bigg] \tag{B.34}$$

From (B.7) we can write down

$$\bar{m}[\tau,\bar{\tilde{s}},w_{\mu\nu}]\bar{m}[\tau,\bar{\tilde{s}},w_{\mu\nu}] = \frac{\bar{s}(\tau)\bar{s}'(\tau)w_{\mu\nu}^4 + w_{\mu\nu}^2\sigma_n^2(\bar{s}\bar{\tilde{s}}'+\bar{\tilde{s}}\bar{s}') + \sigma_n^4\bar{\tilde{s}}\bar{\tilde{s}}'}{w_{\mu\nu}^2+\sigma_n^2} \tag{B.35}$$

Thus, we can write $\mathbb{E}\big[y_i y_j e_{\mu\nu}\big]$ as

$$\mathbb{E}\big[y_i y_j e_{\mu\nu}\big] = \frac{F_{ij}w_{\mu\nu}^4 + w_{\mu\nu}^2\sigma_n^2(h_i\tilde{s}_j+\tilde{s}_i h_j)+\sigma_n^4\tilde{s}_i\tilde{s}_j\mathbb{E}[e_{\mu\nu}]}{w_{\mu\nu}^2+\sigma_n^2} + \delta_{ij}\sigma^2(w_{\mu\nu})\mathbb{E}[e_{\mu\nu}], \tag{B.36}$$

where

$$F_{ij}[\tau,\bar{r},w] = \mathbb{E}_\tau\big[s_i(\tau)s_j(\tau)f[\tau,\bar{r},w]\big], \tag{B.37}$$

and the arguments for calculating h_j, h_j, and F_{ij} are the same, namely $[\tau,\bar{\tilde{s}},w_{\mu\nu}]$. We can then evaluate $\mathbb{E}[y_i y_j e_\mu e_\nu]$ from Eq. (B.32).

Appendix C

MMSE approximation in the high-SNR regime

We have the following relation, from an expansion of x about $u = \hat{X}_{\mathrm{MAP}}(y)$ correct to the 4^{th} order (5.29):

$$x(u,y) = x[u = \hat{X}_{\mathrm{MAP}}(y), y] + \gamma[u - \hat{X}_{\mathrm{MAP}}(y)] + \epsilon[u - \hat{X}_{\mathrm{MAP}}(y)]^2 + \eta[u - \hat{X}_{\mathrm{MAP}}(y)]^3 + \xi[u - \hat{X}_{\mathrm{MAP}}(y)]^4 \tag{C.1}$$

where γ, ϵ, η and ξ are the higher order corrections to an expansion of the posterior distribution when it departs from being Gaussian. They are given by equations (5.36), (5.40), (5.42) and (5.43), respectively. We can then write the partial derivative of x w.r.t u as,

$$\frac{\partial x}{\partial u} = \gamma + 2\epsilon[u - \hat{X}_{\mathrm{MAP}}(y)] + 3\eta[u - \hat{X}_{\mathrm{MAP}}(y)]^2 + 4\eta[u - \hat{X}_{\mathrm{MAP}}(y)]^3. \tag{C.2}$$

We will start from the expression for the high-SNR approximation to the MMSE

as given in Eq. (5.59),

$$\mathcal{E}_M = \int dy\, P(y)\left(\gamma + \frac{3\eta}{\langle S(Y)\rangle}\right)^{-1} \int du \frac{\partial x}{\partial u}(x-\hat{X}_M)^2 \sqrt{\frac{\langle S(Y)\rangle}{2\pi}}\exp\Big\{-\frac{1}{2}\langle S(Y)\rangle\,[u-\hat{X}_{\rm MAP}(y)]^2\Big\} \tag{C.3}$$

The MMSE estimator is expressed in terms of the MAP estimator and the higher order correction in (5.56) as

$$\hat{X}_M(y) - \hat{X}_{\rm MAP}(y) = \frac{3\epsilon}{\langle S(Y)\rangle} + \frac{6\epsilon\eta}{\gamma\langle S(Y)\rangle^2} + \frac{15\xi}{\langle S(Y)\rangle^2}, \tag{C.4}$$

$$\stackrel{\rm def}{=} f(y) \tag{C.5}$$

We can express the product $(x-\hat{X}_M(y))^2\partial x/\partial u$ as

$$\frac{\partial x}{\partial u}(x-\hat{X}_M(y))^2 = [\gamma + 2\epsilon(u-\hat{X}) + 3\eta(u-\hat{X})^2 + 4\xi(u-\hat{X})^3][x-\hat{X}-f(y)]^2, \tag{C.6}$$

$$\begin{aligned} &= [\gamma + 2\epsilon(u-\hat{X}) + 3\eta(u-\hat{X})^2 + 4\xi(u-\hat{X})^3] \\ &\quad\times [\gamma(u-\hat{X}) + \epsilon(u-\hat{X})^2 + \eta(u-\hat{X})^3 + \xi(u-\hat{X})^4 - f(y)]^2, \end{aligned} \tag{C.7}$$

$$\begin{aligned} &= [\gamma + 2\epsilon(u-\hat{X}) + 3\eta(u-\hat{X})^2 + 4\xi(u-\hat{X})^3] \\ &\quad\times \big[\gamma^2(u-\hat{X})^2 + \epsilon^2(u-\hat{X})^4 + f^2(y) \\ &\quad + 2\gamma\epsilon(u-\hat{X})^3 + 2\gamma\eta(u-\hat{X})^4 - 2\gamma(u-\hat{X})f(y) \\ &\quad - 2\epsilon(u-\hat{X})^2 f(y) - 2\eta(u-\hat{X})^3 f(y) - 2\xi(u-\hat{X})^4 f(y)\big], \end{aligned} \tag{C.8}$$

where $\hat{X}$ is used to denote the MAP estimator. Note that only the even powers of $(u-\hat{X})$ contribute to the integral over u in Eq. (C.3). The even powers of $(u-\hat{X})$ in the product $(x-\hat{X}_M)^2\partial x/\partial u$ are,

$$\begin{aligned} \text{Even Powers} = \gamma\big[&\gamma^2(u-\hat{X})^2 + \epsilon^2(u-\hat{X})^4 + f^2(y) + 2\gamma\eta(u-\hat{X})^4 \\ &- 2\epsilon(u-\hat{X})^2 f(y) - 2\xi(u-\hat{X})^4 f(y)\big] \end{aligned}$$

$$
\begin{aligned}
&+ 2\epsilon\big[2\gamma\epsilon(u-\hat{X})^4 - 2\gamma(u-\hat{X})^2 f(y) - 2\eta(u-\hat{X})^4 f(y)\big] \\
&+ 3\eta\big[\gamma^2(u-\hat{X})^4 - 2\epsilon(u-\hat{X})^4 f(y) + (u-\hat{X})^2 f^2(y)\big] \\
&- 8\gamma\xi(u-\hat{X})^4 f(y), \qquad \text{(C.9)}
\end{aligned}
$$

where we discarded powers of $(u-\hat{X})$ greater than 4. Only these terms contribute up to $\langle S(Y)\rangle^{-2}$. Making this approximation and noting that the fourth central moment of a Gaussian is three times its variance, we can finally evaluate the integral over u in (C.3) to yield,

$$
\begin{aligned}
\mathcal{E}_M = \int dy\, P(y) &\left(\gamma + \frac{3\eta}{\langle S(Y)\rangle}\right)^{-1} \left\{ \frac{\gamma^3}{\langle S\rangle} + \frac{3\gamma\epsilon^2}{\langle S\rangle^2} + \frac{6\gamma^2\eta}{\langle S\rangle^2} + \gamma f^2 - \frac{2\gamma\epsilon f}{\langle S\rangle} \right. \\
&\left. - \frac{6\gamma\xi f}{\langle S\rangle^2} + \frac{12\gamma\epsilon^2}{\langle S\rangle^2} - \frac{4\gamma\epsilon f}{\langle S\rangle} - \frac{12\epsilon\eta f}{\langle S\rangle^2} + \frac{9\gamma^2\eta}{\langle S\rangle^2} + \frac{3\eta f^2}{\langle S\rangle} - \frac{18\epsilon\eta f}{\langle S\rangle^2} - \frac{24\gamma\xi f}{\langle S\rangle^2} \right\}
\end{aligned} \qquad \text{(C.10)}
$$

Note that $f^2 = 9\epsilon^2/\langle S\rangle^2$ as we are including only orders up to $\langle S\rangle^{-2}$. Therefore, after some simple algebraic manipulations, we can write the high-SNR approximation to the MMSE as

$$
\begin{aligned}
\mathcal{E}_M &= \int dy\, P(y) \left(\gamma + \frac{3\eta}{\langle S\rangle}\right)^{-1} \left\{ \frac{\gamma^3}{\langle S\rangle} + \frac{6\gamma\epsilon^2 + 15\gamma^2\eta}{\langle S\rangle^2} \right\} && \text{(C.11)} \\
&= \int dy\, P(y) \frac{1}{\gamma}\left(1 - \frac{3\eta}{\gamma\langle S\rangle}\right) \left\{ \frac{\gamma^3}{\langle S\rangle} + \frac{6\gamma\epsilon^2 + 15\gamma^2\eta}{\langle S\rangle^2} \right\} && \text{(C.12)} \\
&= \int dy\, P(y) \left\{ \frac{\gamma^2}{\langle S\rangle} + \frac{6\epsilon^2 + 12\gamma\eta}{\langle S\rangle^2} \right\} && \text{(C.13)} \\
&= \int dy\, P(y) \left\{ \frac{1}{S} + \frac{6\epsilon^2 + 12\gamma\eta}{\langle S\rangle^2} \right\}, && \text{(C.14)}
\end{aligned}
$$

which is the final expression (5.62) for the high-SNR approximation to the MMSE.

Appendix D

References

References

[1] K. L. Bell, Y. Steinberg, Y. Ephraim, and H. L. Van Trees. Extended ziv-zakai lower bound for vector parameter estimation. *IEEE Transactions on Information Theory*, 43(2):624–637, march 1997.

[2] G. C. Carter. *Time Delay Estimation*. PhD thesis, University of Connecticut, 1976.

[3] D.J. Torrieri. Statistical theory of passive location systems. *Aerospace and Electronic Systems, IEEE Transactions on*, AES-20(2):183–198, March 1984.

[4] G. C. Carter. Time delay estimation for passive sonar signal processing. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 29(3):463–470, Jun 1981.

[5] R. Kumar and S. Prasad. PSF rotation with changing defocus and applications to 3D imaging for space situational awareness. In *Technical Papers*. AMOS Conference, 2013.

[6] R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A*, 222(594-604):309–368, 1922.

[7] R. A. Fisher. Theory of Statistical Estimation. *Mathematical Proceedings of the Cambridge Philosophical Society*, 22:700–725, 7 1925.

[8] H. Cramer. A contribution to the theory of statistical estimation. *Scandinavian Actuarial Journal*, 1946(1):85–94, 1946.

[9] C. R. Rao. Information and the accuracy attainable in the estimation of statistical parameters. *Bulletin of Cal. Math. Soc.*, 37(3):81–91, 1945.

[10] C. R. Rao. Minimum variance and the estimation of several parameters. *Mathematical Proceedings of the Cambridge Philosophical Society*, 43:280–283, 4 1947.

[11] P. K. Bhattacharya. On some analogous of the amount of information and their use in statistical estimation. *Shankya*, 8(3):1–14, 201–218, 315–328, 1947.

[12] E. W. Barankin. Locally best unbiased estimates. *The Annals of Mathematical Statistics*, 20(4):477–501, 1949.

[13] D. G. Chapman and H. Robbins. Minimum variance estimation without regularity assumptions. *The Annals of Mathematical Statistics*, 22(4):581–586, 12 1951.

[14] J. Kiefer. On minimum variance estimators. *The Annals of Mathematical Statistics*, 23(4):627–629, 12 1952.

[15] J. M. Hammersley. On estimating restricted parameters. *Journal of the Royal Statistical Society. Series B (Methodological)*, 12(2):192–240, 1950.

[16] J. S. Abel. A bound on mean-square-estimate error. *Information Theory, IEEE Transactions on*, 39(5):1675–1680, Sep 1993.

[17] R. McAulay and L. P. Seidman. A useful form of the barankin lower bound and its application to ppm threshold analysis. *Information Theory, IEEE Transactions on*, 15(2):273–279, Mar 1969.

[18] R. McAulay and E. M. Hofstetter. Barankin bounds on parameter estimation. *Information Theory, IEEE Transactions on*, 17(6):669–676, Nov 1971.

[19] P. Forster and P. Larzabal. On lower bounds for deterministic parameter estimation. In *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, volume 2, pages II–1137–II–1140, May 2002.

[20] E. Chaumette, J. Galy, A. Quinlan, and P. Larzabal. A new barankin bound approximation for the prediction of the threshold region performance of maximum likelihood estimators. *IEEE Transactions on Signal Processing*, 56(11), November 2008.

[21] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part 1.* New York: Wiley, 1968.

[22] D. Rife and R. R. Boorstyn. Single tone parameter estimation from discrete-time observations. *Information Theory, IEEE Transactions on*, 20(5):591–598, Sep 1974.

[23] F. Athley. Threshold region performance of maximum likelihood direction of arrival estimators. *Signal Processing, IEEE Transactions on*, 53(4):1359–1373, April 2005.

[24] K. Todros and J. Tabrikian. General classes of performance lower bounds for parameter estimation-part i: Non-bayesian bounds for unbiased estimators. *Information Theory, IEEE Transactions on*, 56(10):5045–5063, Oct 2010.

[25] A. Quinlan, E. Chaumette, and P. Larzabal. A direct method to generate approximations of the barankin bound. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 3, pages III–808–III–811, May 2006.

[26] J. Ziv and M. Zakai. Some lower bounds on signal parameter estimation. *Information Theory, IEEE Transactions on*, 15(3):386–391, May 1969.

[27] B. Efron. Biased versus unbiased estimation. *Advances in Mathematics*, 16(3):259 – 277, 1975.

[28] G. Demoment. Image reconstruction and restoration: overview of common estimation structures and problems. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 37(12):2024–2036, Dec 1989.

[29] A. E. Hoerl and R. W. Kennard. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67, 1970.

[30] J. D. Gorman and A. O. Hero. Lower bounds for parametric estimation with constraints. *Information Theory, IEEE Transactions on*, 36(6):1285–1301, Nov 1990.

[31] A. O. Hero. A cramer-rao type lower bound for essentially unbiased parameter estimation. In *Lincoln Lab., Mass. Inst. Technol.*, 1992.

[32] A. O. Hero, J. A. Fessler, and M. Usman. Exploring estimator bias-variance tradeoffs using the uniform cr bound. *Signal Processing, IEEE Transactions on*, 44(8):2026–2041, Aug 1996.

[33] Y. C. Eldar. Minimum variance in biased estimation: bounds and asymptotically optimal estimators. *Signal Processing, IEEE Transactions on*, 52(7):1915–1930, July 2004.

[34] K. Todros and J. Tabrikian. A new lower bound on the mean-square error of biased estimators. In *Electrical and Electronics Engineers in Israel, 2008. IEEEI 2008. IEEE 25th Convention of*, pages 745–749, Dec 2008.

[35] Y. C. Eldar. Uniformly improving the cramer-rao bound and maximum-likelihood estimation. *Signal Processing, IEEE Transactions on*, 54(8):2943–2956, Aug 2006.

[36] S. Kay and Y. C. Eldar. Rethinking biased estimation [lecture notes]. *Signal Processing Magazine, IEEE*, 25(3):133–136, May 2008.

[37] L. P. Seidman. Performance limitations and error calculations for parameter estimation. *Proceedings of the IEEE*, 58(5):644–652, May 1970.

[38] D. Chazan, M. Zakai, and J. Ziv. Improved lower bounds on signal parameter estimation. *Information Theory, IEEE Transactions on*, 21(1):90–93, Jan 1975.

[39] S. Bellini and G. Tartara. Bounds on error in signal parameter estimation. *Communications, IEEE Transactions on*, 22(3):340–342, Mar 1974.

[40] V. A. Kotelnikov. *The Theory of Optimum Noise Immunity.* McGraw-Hill, 1959.

[41] K. L. Bell. *Performance Bounds in Parameter Estimation with Application to Bearing Estimation.* PhD thesis, George Mason University, 1995.

[42] L. D. Brown and R. C. Liu. Bounds on the bayes and minimax risk for signal parameter estimation. *Information Theory, IEEE Transactions on*, 39(4):1386–1394, Jul 1993.

[43] A. J. Weiss and E. Weinstein. A lower bound on the mean-square error in random parameter estimation (corresp.). *Information Theory, IEEE Transactions on*, 31(5):680–682, Sep 1985.

[44] B. Bobrovsky and M. Zakai. A lower bound on the estimation error for certain diffusion processes. *Information Theory, IEEE Transactions on*, 22(1):45–52, Jan 1976.

[45] E. Weinstein and A. J. Weiss. A general class of lower bounds in parameter estimation. *IEEE Transactions on Information Theory*, 34(2), March 1988.

[46] A. Renaux, P. Forster, P. Larzabal, C.D. Richmond, and A. Nehorai. A fresh look at the bayesian bounds of the weiss-weinstein family. *Signal Processing, IEEE Transactions on*, 56(11):5334–5352, Nov 2008.

[47] K. Todros and J. Tabrikian. General classes of performance lower bounds for parameter estimation part ii: Bayesian bounds. *Information Theory, IEEE Transactions on*, 56(10):5064–5082, Oct 2010.

[48] Z. Ben-Haim and Y. C. Eldar. A lower bound on the bayesian mse based on the optimal bias function. *Information Theory, IEEE Transactions on*, 55(11):5179–5196, Nov 2009.

[49] H. L. Van Trees and K. L. Bell. *Bayesian Bounds for Parameter Estimation and Nonlinear Filtering/Tracking*. Wiley-IEEE Press, 2007.

[50] S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 03 1951.

[51] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc.*, 35:99–109, 1943.

[52] P. A. Devijver. On a new class of bounds on bayes risk in multihypothesis pattern recognition. *Computers, IEEE Transactions on*, C-23(1):70–80, Jan 1974.

[53] S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *J. Royal Stat. Soc.*, 25:131 –142, 1966.

[54] T. Routtenberg and J. Tabrikian. A general class of lower bounds on the probability of error in multiple hypothesis testing. In *Electrical and Electronics Engineers in Israel, 2008. IEEEI 2008. IEEE 25th Convention of*, pages 750–754, Dec 2008.

[55] T. Kailath. The divergence and bhattacharyya distance measures in signal selection. *Communication Technology, IEEE Transactions on*, 15(1):52–60, February 1967.

[56] H. V. Poor and J. B. Thomas. Applications of ali-silvey distance measures in the design generalized quantizers for binary decision systems. *Communications, IEEE Transactions on*, 25(9):893–900, Sep 1977.

[57] T. Routtenberg and J. Tabrikian. General classes of bayesian lower bounds for outage error probability and mse. In *Statistical Signal Processing, 2009. SSP '09. IEEE/SP 15th Workshop on*, pages 69–72, Aug 2009.

[58] C. E. Shannon. Mutual information and minimum mean-square error in gaussian channels. *The Bell System Technical Journal*, 27:379423, 623656, July, October 1948.

[59] R. M. Fano. *Transmission of Information*. Wiley, 1961.

[60] M. Feder and N. Merhav. Relations between entropy and error probability. *Information Theory, IEEE Transactions on*, 40(1):259–266, Jan 1994.

[61] M. E. Hellman and J. Raviv. Probability of error, equivocation, and the chernoff bound. *Information Theory, IEEE Transactions on*, 16(4):368–372, Jul 1970.

[62] J. Seidler. Bounds on the mean-square error and the quality of domain decisions based on mutual information. *Information Theory, IEEE Transactions on*, 17(6):655–665, Nov 1971.

[63] N. Santhi and A. Vardy. On an improvement over rnyi's equivocation bound. In *Proceedings of the 44th Annual Allerton Conference on Communication, Control, and Computing (Curran, 2006)*, page 11181124, 2006.

[64] J. Golic. On the relationship between the information measures and the bayes probability of error. *Information Theory, IEEE Transactions on*, 33(5):681–693, Sep 1987.

[65] D. Guo, S. Shamai, and S. Verdu. Mutual information and minimum mean-square error in gaussian channels. *Information Theory, IEEE Transactions on*, 51(4):1261–1282, April 2005.

[66] D. Guo, S. Verdu, and S. Shamai. Mutual information and conditional mean estimation in poisson channels. In *Information Theory Workshop, 2004. IEEE*, pages 265–270, Oct 2004.

[67] S. Prasad. Certain Relations between mutual information and fidelity of statistical estimation. *ArXiv e-prints*, October 2010.

[68] L. P. Seidman. An upper bound on average estimation error in nonlinear systems. *Information Theory, IEEE Transactions on*, 14(2):243–250, Mar 1968.

[69] R. Hawkes and J. Moore. An upper bound on the mean-square error for bayesian parameter estimators (corresp.). *Information Theory, IEEE Transactions on*, 22(5):610–615, Sep 1976.

[70] J. T. Flam, S. Chatterjee, K. Kansanen, and T. Ekman. On mmse estimation: A linear model under gaussian mixture statistics. *Signal Processing, IEEE Transactions on*, 60(7):3840–3845, 2012.

[71] G. J. McLachlan and K. E. Basford. *Mixture models: Inference and Applications to Clustering*. New York: Marcel Dekker, 1989.

[72] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero, R. L. Moses, and N. S. Correal. Locating the nodes: cooperative localization in wireless sensor networks. *Signal Processing Magazine, IEEE*, 22(4):54–69, July 2005.

[73] B. Huang, M. Bates, and X. Zhuang. Super-resolution fluorescence microscopy. *Annual Review of Biochemistry*, 78(1):993–1016, 2009. PMID: 19489737.

[74] T. Quan, P. Li, F. Long, S. Zeng, Q. Luo, P. N. Hedde, G. U. Nienhaus, and Z. Huang. Ultra-fast, high-precision image analysis for localization-based super resolution microscopy. *Opt. Express*, 18(11):11867–11876, May 2010.

[75] A. J. Berglund, M. D. McMahon, J. J. McClelland, and J. A. Liddle. Fast, bias-free algorithm for trackingsingle particles with variable size andshape. *Opt. Express*, 16(18):14064–14075, Sep 2008.

[76] B. Rieger and S. Stallinga. The lateral and axial localization uncertainty in super-resolution light microscopy. *ChemPhysChem*, 15(4):664670, Sep 2014.

[77] A. R. Small and R. Parthasarathy. Superresolution localization methods. *Annual Review of Physical Chemistry*, 65(1):107–125, 2014. PMID: 24274701.

[78] A. V. Abraham, S. Ram, J. Chao, E. S. Ward, and R. J. Ober. Quantitative study of single molecule location estimation techniques. *Optics Express*, 17(26):23352–23373, Dec 2009.

[79] S. Prasad. New error bounds for M-testing and estimation of source location with subdiffractive error. *J. Opt. Soc. Am. A*, 29(3):354–366, Mar 2012.

[80] S. Prasad and S. Narravula. Applications of Shannon information and statistical estimation theory to inverse problems in imaging. In *Imaging and Applied Optics*. Optical Society of America, 2011.

[81] C. W. Helstrom. Detection and resolution of incoherent objects by a background-limited optical system. *J. Opt. Soc. Am.*, 59(2):164–175, Feb 1969.

[82] M. Shahram and P. Milanfar. Imaging below the diffraction limit: a statistical analysis. *Image Processing, IEEE Transactions on*, 13(5):677–689, May 2004.

[83] C. C. Leang and D. H. Johnson. On the asymptotics of m-hypothesis bayesian detection. *Information Theory, IEEE Transactions on*, 43(1):280–282, Jan 1997.

[84] S. Prasad. Asymptotics of Bayesian error probability and rotating-PSF-based source super-localization in three dimensions. *Optics Express*, submitted.

[85] S. Prasad. Asymptotics of Bayesian error probability and 2D pair superresolution. *Optics Express*, submitted.

[86] Y. Wang, T. Quan, S. Zeng, and Z. Huang. Palmer: a method capable of parallel localization of multiple emitters for high-density localization microscopy. *Opt. Express*, 20(14):16039–16049, Jul 2012.

[87] T. Quan, H. Zhu, X. Liu, Y. Liu, J. Ding, S. Zeng, and Z. Huang. High-density localization of active molecules using structured sparse model and bayesian information criterion. *Opt. Express*, 19(18):16963–16974, Aug 2011.

[88] S. Cox, E. Rosten, J. Monypenny, T. Jovanovic-Talisman, D. T. Burnette, J. Lippincott-Schwartz, G. E Jones, and R. Heintzmann. Bayesian localization microscopy reveals nanoscale podosome dynamics. *Nature Methods*, 9(2):195–200, 2012. PMID: 19489737.

[89] C. Knapp and G. C. Carter. The generalized correlation method for estimation of time delay. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 24(4):320–327, Aug 1976.

[90] J. C. Hassab and R. Boucher. Optimum estimation of time delay by a generalized correlator. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 27(4):373–380, Aug 1979.

[91] Y. T. Chan and K. C. Ho. A simple and efficient estimator for hyperbolic location. *Signal Processing, IEEE Transactions on*, 42(8):1905–1915, Aug 1994.

[92] A. J. Weiss and E. Weinstein. Fundamental limitations in passive time delay estimation–part i: Narrow-band systems. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 31(2):472–486, Apr 1983.

[93] E. Weinstein and A. J. Weiss. Fundamental limitations in passive time-delay estimation–part ii: Wide-band systems. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 32(5):1064–1078, Oct 1984.

[94] B. M. Sadler and R. J. Kozick. A survey of time delay estimation performance bounds. In *Sensor Array and Multichannel Processing, 2006. Fourth IEEE Workshop on*, pages 282–288, July 2006.

[95] W. H. Richardson. Bayesian-based iterative method of image restoration. *J. Opt. Soc. Am.*, 62(1):55–59, Jan 1972.

[96] L. B. Lucy. An iterative technique for the rectification of observed distributions. *The Astronomical Journal*, 79:745, 1974.

[97] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, 39(1):1–38, 1977.

[98] G.J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. John Wiley & Sons, New York, 1997.

[99] A. M. Walker. On the asymptotic behaviour of posterior distributions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 31(1):pp. 80–88, 1969.

[100] S L Braunstein. How large a sample is needed for the maximum likelihood estimator to be approximately gaussian? *Journal of Physics A: Mathematical and General*, 25(13):3813, 1992.

[101] E. Weinstein. Relations between belini-tartara, chazan-zakai-ziv, and wax-ziv lower bounds. *Information Theory, IEEE Transactions on*, 34(2):342–343, Mar 1988.

[102] E. Cinlar. *Introduction to Stochastic Processes*. Englewood Cliffs, NJ: Prentice-Hall, 1975.

[103] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *The Annals of Mathematical Statistics*, 23(4):493–507, 12 1952.

[104] R. J. Kenefic. A bayesian approach to time delay estimation. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 29(3):611–614, Jun 1981.

[105] B. M. Sadler, N. Liu, and Z. Xu. Ziv-zakai bounds on time delay estimation in unknown convolutive random channels. *Signal Processing, IEEE Transactions on*, 58(5):2729–2745, May 2010.

[106] G. Sun, Jie Chen, Wei Guo, and K.J.R. Liu. Signal processing techniques in network-aided positioning: a survey of state-of-the-art positioning designs. *Signal Processing Magazine, IEEE*, 22(4):12–23, July 2005.

[107] F. Seco, A.R. Jimenez, C. Prieto, J. Roa, and K. Koutsou. A survey of mathematical methods for indoor localization. In *Intelligent Signal Processing, 2009. WISP 2009. IEEE International Symposium on*, pages 9–14, Aug 2009.

[108] H. Stark and J. W. Woods, editors. *Probability, Random Processes, and Estimation Theory for Engineers*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1986.

[109] H. Nguyen and H. L. Van Trees. Comparison of performance bounds for doa estimation. In *Statistical Signal and Array Processing., IEEE Seventh SP Workshop on*, pages 313–316, Jun 1994.

[110] S. R. Narravula and S. Prasad. Extensions to the Ziv-Zakai Lower Bound. *in preparation*.

[111] Y. C. Eldar. Mse bounds with affine bias dominating the cramer-rao bound. *Signal Processing, IEEE Transactions on*, 56(8):3824–3836, Aug 2008.

[112] S. M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[113] A. Greengard, Y. Y. Schechner, and R. Piestun. Depth from diffracted rotation. *Opt. Lett.*, 31(2):181–183, Jan 2006.

[114] S. R. P. Pavani and R. Piestun. High-efficiency rotating point spread functions. *Opt. Express*, 16(5):3484–3489, Mar 2008.

[115] M. D. Lew, S. F. Lee, M. Badieirostami, and W. E. Moerner. Corkscrew point spread function for far-field three-dimensional nanoscale localization of pointlike objects. *Opt. Lett.*, 36(2):202–204, Jan 2011.

[116] S. Prasad. Rotating point spread function via pupil-phase engineering. *Opt. Lett.*, 38(4):585–587, Feb 2013.

[117] S. R. P. Pavani, A. Greengard, and R. Piestun. Three-dimensional localization with nanometer accuracy using a detector-limited double-helix point spread function system. *Applied Physics Letters*, 95(2):–, 2009.

[118] A. Gahlmann, J. L. Ptacin, G. Grover, S. Quirin, A. R. S. von Diezmann, M. K. Lee, M. P. Backlund, L. Shapiro, R. Piestun, and W. E. Moerner. Quantitative multicolor subdiffraction imaging of bacterial protein ultrastructures in three dimensions. *Nano Letters*, 13(3):987–993, 2013.

[119] C. Roider, A. Jesacher, S. Bernet, and M. Ritsch-Marte. Axial super-localisation using rotating point spread functions shaped by polarisation-dependent phase modulation. *Opt. Express*, 22(4):4029–4037, Feb 2014.

[120] G. Grover, S. R. P. Pavani, and R. Piestun. Performance limits on three-dimensional particle localization in photon-limited microscopy. *Opt. Lett.*, 35(19):3306–3308, Oct 2010.

[121] S. Prasad, R. Kumar, and S. Narravula. Angular momentum, PSF rotation, and 3D source localization: A statistical performance analysis. In *Imaging and Applied Optics*. Optical Society of America, 2014.

[122] S. R. Narravula and S. Prasad. An upper bound on the MMSE and application to the time delay problem. *in preparation*.

[123] S. R. Narravula and S. Prasad. A new piecewise quasi-linear estimator. *in preparation*.

[124] S. Prasad, R. Kumar, and S. Narravula. PSF rotation with defocus and applications to 3D imaging. In *Quantitative Bio-Imaging*. Quantitative Bio-Imaging, 2014.

[125] C. F. Chen. On asymptotic normality of limiting density functions with bayesian implications. *Journal of the Royal Statistical Society. Series B (Methodological)*, 47(3):pp. 540–546, 1985.

[126] L. Tierney and J. B. Kadane. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81(393):pp. 82–86, 1986.

[127] R. E. Kass, L. Tierney, and J. B. Kadane. The validity of posterior expansions based on laplace's method. In *Bayesian and Likelihood Methods in Statistics and Econometrics. Geisser, S., Hodges, J. S., Press, S. J., and Zellner, A., editors*. Elsevier Science Publisher, March 1990.

[128] Y. Miyata. Fully exponential laplace approximations using asymptotic modes. *Journal of the American Statistical Association*, 99(468):pp. 1037–1049, 2004.

[129] L. Tierney, R. E. Kass, and J. B. Kadane. Fully exponential laplace approximations to expectations and variances of nonpositive functions. *Journal of the American Statistical Association*, 84(407):710–716, 1989.

[130] George F. Carrier, Max Krook, and Carl E. Pearson. *Functions of a Complex Variable: Theory and Technique*. Society for Industrial and Applied Mathematics, 2005.

[131] S. Prasad and S. R. Narravula. Low-SNR and high-SNR approximations for the MMSE and the MMSE estimator. *in preparation*.