Source Detection and Image Reconstruction with Position–Sensitive Gamma–Ray Detectors

by

Daniel J. Lingenfelter

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2012

Doctoral Committee:

Professor Jeffrey A. Fessler, Chair
Professor Clayton D. Scott
Professor Zhong He
Professor Honglak Lee

ACKNOWLEDGEMENTS

I would like to thank the faculty and students at the University of Michigan, my family, and my friends for everything they did to help me grow intellectually and as a person during my graduate school career.

My advisor Jeff Fessler gave me the guidance to succeed and the freedom to grow. I was reminded of his devotion to his students every time he carefully and swiftly proofread my draft papers, no matter how busy he was. His insightful ideas and suggestions helped to shape this dissertation into its current form. Jeff embodies the qualities of an excellent leader and mentor. His brilliance and joyful passion for his work inspired me to do my best, and I will carry that inspiration for the rest of my life.

Professor Clay Scott contributed to this dissertation and to my professional development. Upon entering the program, he recommended a book about how to make a career in research, which played a major role in clarifying my career goals. I am grateful that he shared his deep knowledge of statistics with me and also for his excellent suggestions that improved the clarity and completeness of this work.

I am grateful to Professor Zhong He for allowing me to collaborate with his lab and for not letting me forget the practical applications of this work. It was a pleasure to work with him and his students who were eager to teach me about radiation measurement and room temperature semiconductor systems. I am thankful for his advice on how to conduct experiments with real and simulated detectors.

I started working with Professor Honglak Lee when I gave my thesis proposal. It was excellent to have his perspective, which led to interesting new directions for future work. His insightful questions also helped me to improve the quality of this dissertation.

I would also like to thank Kurt Metzger, Adib Nashashibi, and Professor Fawwaz Ulaby for the discussions we had about graduate school while I was an undergraduate at the University of Michigan. I would not be where I am today without their advice and guidance.

I am grateful to the students and postdocs who helped me throughout my graduate school career. Chris Wahl, Weiyi Wang, and Jason Jaworski contributed ideas and insightful discussions. Andy Boucher and Willy Kaye helped with experimental data collection. Yuefeng Zhu shared his simulation software with me, allowing me to validate my ideas. I would also like to thank Sathish Ramani for our discussions about Augmented Lagrangian algorithms.

This dissertation would not have been possible without my parents Dave and Lora Lingenfelter, who worked hard so I could have the opportunities that I have today. I cannot thank them enough for their unconditional love and support. The times spent visiting and going on mountain adventures with my brother Tom Lingenfelter were energizing and healthy breaks from work. I would also like to thank my best friend Katie Ellison for her support.

This work was financially supported by the Electrical Engineering Systems department at the University of Michigan, the Domestic Nuclear Detection Office of the U.S. Department of Homeland Security, the MIT Lincoln Laboratory Fellowship, and the Innovative Signal Analysis Fellowship. I was also fortunate to spend the summer of 2010 at the U.S. Naval Research Laboratory.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF APPENDICES

Appendix

CHAPTER I

Introduction

Detection and identification of nuclear materials at ports of entry and in densely populated areas is a challenging but vitally important problem. In fiscal year 2009, the United States Customs and Border Protection agency processed a daily average of 989,689 passengers and pedestrians, 57,761 cargo containers, and 271,278 privately owned vehicles [13]. A system used to detect dangerous nuclear material at international borders must have a high detection probability to reliably detect actual threats, while having a small probability of false alarm to minimize delays for benign traffic. Some radiation detectors used at ports of entry use photon counting statistics alone to detect sources [66], while other, more sophisticated systems use both counting statistics and spectroscopy [67]. These detectors do not allow one to localize a radioactive source within a stationary object, and do not account for the spatial distribution of radioactive material when deciding whether or not a source is present.

The radiation measurement group at the University of Michigan is developing a position-sensitive Compton imaging system [81] that one can use to measure the spatial distribution and energy spectrum of gamma–ray emissions. We define a position–sensitive detector to be a detector that can infer information about the

direction of an incoming photon. Position–sensitive detectors have the potential to improve detection accuracy in settings such as border crossings by helping localize sources. Many ordinary materials, such as concrete, contain radioactive isotopes that contribute to the background radiation recorded by the detector. These materials, as well as cosmic radiation, are usually more broadly distributed in space than malicious radioactive objects. Position–sensitive detectors have the ability to use the *a priori* knowledge that malicious sources are spatially localized to increase detection performance [66].

A position–sensitive gamma–ray imaging system should provide information about radioactive sources in a given field of view. We consider two methods of accomplishing this: image reconstruction and source detection. The goal of image reconstruction is to render an image of the spatial gamma–ray emission density around the detector. The maximum likelihood (ML) method of image reconstruction [4] has been previously applied to gamma–ray imaging with 3D position–sensitive detectors [80], but the ML solution may not be representative of the true emission density with few recorded measurements. The ML estimate can exhibit poor behavior because it is the solution to an inverse problem that is typically ill–posed when the number of measurements is small.

Regularization is a method of incorporating *a priori* information into an ill–posed inverse problem to improve the condition of the system and generate a solution that is a compromise between fitting the recorded data and the being consistent with the *a priori* information [21]. Penalized likelihood problems are a type of regularized inverse problem where the cost function is based on a statistical likelihood. Penalized likelihood has been previously applied in medical emission tomography to enforce the *a priori* expectation that the images are smooth [16, 25]. In this work, we use the

method of penalized likelihood to generate images that fit gamma–ray measurements and the *a priori* knowledge that images should be sparse in space.

We found that reconstructed images, especially those reconstructed with a sparsity–promoting penalty function, are not necessarily the most useful for the task of detecting a radioactive source. Incorporating the expectation of sparsity into the reconstruction method can cause measurements of diffuse background emissions to appear as a point–source in the image, especially with few measurements. This shortcoming motivates our study of source detection performance with gamma–ray imaging systems.

Predicting the performance of methods for detecting a source in background with a position–sensitive gamma–ray measurement system is often difficult because the measurements are random variables with distributions that depend on nonlinear functions of the source intensity and position. Detection performance results reported in the literature are commonly generated by empirically performing the test for multiple trials with and without a source present, e.g., [66, 76]. Empirical methods for computing detection performance require a large amount of data and computation, and separate computations must be performed for each scan time to be considered. Furthermore, empirical computations do not give an intuitive formula to aid in understanding how the system design impacts the ability to detect a source in background.

We use the asymptotic distribution of maximum likelihood estimates to provide an efficient means of computing detection performance in terms of the receiver operating characteristic (ROC), which is the probability of detection as a function of the probability of false alarm. In the absence of model mismatch, the asymptotic distribution of the maximum likelihood estimates of source intensity is useful for

proving results about the utility of position and spectral information. We use the asymptotic distribution of maximum likelihood estimates to prove that imaging information can not decrease detection performance in the absence of model mismatch and provide a formula to quantify the performance increase in terms of the system response function.

Exact system models for some position–sensitive gamma–ray detectors, such as the pixelated CdZnTe used in this work, are often too computationally intensive to compute in practice, which motivates the use of approximate models. An approximate model for a 3D position–sensitive CdZnTe detector is given in [80]. Estimators derived from approximate models may not enjoy many of the desirable properties of the ML estimator, such as asymptotic unbiasedness, asymptotic efficiency, and asymptotic normality. Predicting asymptotic detection performance under the incorrect assumption that these properties are satisfied can lead to inaccurate, and sometimes overly optimistic results. We extend previous results on the asymptotic normality of maximum likelihood estimators under model mismatch, or quasi–maximum likelihood estimators (QMLE) [78], to the gamma–ray imaging case where the number of recorded measurements is Poisson. We apply the theoretical results to a simulated detector and a real CdZnTe system to show that the asymptotic normality of the QMLE is a useful tool for predicting detection performance with gamma–ray imaging systems.

1.1 Contributions

This work focuses on the problem of extracting useful information from the data recorded by position–sensitive gamma–ray detectors. Our contributions are primarily results about the performance of established algorithms applied to the gamma–ray

imaging and source detection problems.

We propose a penalized–likelihood algorithm using a penalty function based on the sum of logarithms of pixel values for reconstructing gamma–ray images when the images are known *a priori* to be sparse in the spatial domain. The proposed penalty function is discussed as a measure of sparsity in [10]. We compare the proposed algorithm to the more traditional ℓ_1 and ℓ_0 regularizers. Our results show that the ℓ_1 penalized algorithm reconstructs scaled versions of the maximum-likelihood (ML) solution, which does not improve the sparsity over the traditional ML estimate. We show that the ℓ_0 penalized solution is equivalent to the ML solution due to the singularity of the Poisson log-likelihood at zero. We demonstrate that the penalty based on the sum of logarithms produces sparser images than the ML solution. We evaluated these algorithms using experimental data from a position-sensitive Compton-imaging detector [30], where the spatial distribution of photon-emitters is known to be sparse.

For the problem of source detection, we prove Theorem IV.3, which states that in a known background, a uniform–sensitivity position–sensitive detector always has equal or better detection performance asymptotically, in terms of AUC, than a counting detector of equal sensitivity. Our analysis also provides an expression that quantifies how much position–sensitive capability increases the AUC. We give numerical examples to illustrate Theorem IV.3.

We also provide numerical examples demonstrating the utility of position–sensitive detectors of non–uniform sensitivity over counting detectors. Non–uniform detectors violate the assumptions of Theorem IV.3, but we numerically evaluate detection performance using the asymptotic method to compare the performance of various non–uniform detectors to the performance.

There are many choices for tests to detect the presence of a source. We give a numerical comparison of the performance of three different tests in known and unknown background.

Since exact system models are often difficult to compute for systems such as 3D Compton detectors, we propose a methodology for predicting detection performance of gamma–ray imaging systems when the system model is incorrect. This methodology is based on an extension of the asymptotic normality of the quasi maximum likelihood estimator (QMLE), which means a maximum likelihood estimator when the modeled likelihood is not correct. The asymptotic normality of QMLEs is proven in [2, 78], but only in the case of a nonrandom number of measurements. We extend the asymptotic normality of the QMLE for a nonrandom number of measurements in [78] to the gamma–ray source detection case where the number of measurements is Poisson.

Since source and background intensities are naturally nonnegative, we extend the results of [78] and [69] to prove Theorems V.5 and V.7, which show convergence in probability and distribution of the constrained QMLE under somewhat restrictive regularity conditions when the number of measurements is random and model mismatch is present. Since the regularity conditions for Theorem V.7 are often violated in practice, we propose an approximation for the distribution of the QMLE for finite scan times. Numerical results with a simulated detector show that the proposed approximation is reasonable for modest scan times.

We provide simulated numerical examples to illustrate the effects of model mismatch on gamma–ray imaging systems. We first evaluate the asymptotic performance of a simple detector that records a Poisson number of events, where each event has one Gaussian–distributed attribute. This example illustrates that model

mismatch can affect performance predictions significantly.

We evaluate the asymptotic performance of tests for detecting a point–source in a distributed background using an idealized position–sensitive Compton detector in a background of unknown intensity with constrained and unconstrained source intensity estimates. The joint probability distribution of photon interaction positions and deposited energies in position–sensitive Compton detectors is computationally expensive, motivating the use of approximate models. We use Monte Carlo simulation to evaluate the asymptotic detection performance of a parallel–plate detector employing the approximate model in [80].

To supplement the simulation results, we computed empirical and asymptotic performance predictions with real data from an 18 detector CdZnTe array for various source–to–background ratios with a Cs-137 source. Our results show that the asymptotic performance prediction method that accounts for model mismatch approximates the empirical performance well on a real system.

1.2 Outline

This thesis is organized as follows: Chapter II provides background information about gamma–ray imaging systems, detection theory, and estimation theory, Chapter III discusses image reconstruction algorithms and applications to position–sensitive CdZnTe detectors, Chapter IV covers asymptotic performance of tests for detecting a gamma–ray source in background and performance comparisons between detectors with and without position sensitivity, Chapter V develops theory for the asymptotic performance of tests for detecting a source in background in the presence of model mismatch and gives results for an application of the theory to position–sensitive gamma–ray detectors, and Chapter VII outlines proposed future work.

CHAPTER II

Background

We begin with an overview of gamma–ray measurement and Compton imaging. We then explain the classical maximum likelihood expectation maximization (MLEM) approach to image reconstruction and give an overview of its application to Compton imaging systems. We conclude the section with an overview of penalized–likelihood image reconstruction algorithms and statistical approaches to detection.

2.1 Gamma–ray measurement

Three prominent devices for gamma–ray measurement are gas detectors, scintillation detectors, and semiconductors [38]. Scintillation detectors work by converting incoming gamma–ray photons to visible light photons that are collected by some recording mechanism. Scintillation detectors typically have lower energy resolution than other types of detectors [38]. Scintillators have been use in medical imaging technologies such as positron emission tomography (PET) [61].

Gaseous detectors detect gamma–rays by measuring the charge in the atoms that are ionized by interacting gamma–rays. Gaseous detectors typically have lower sensitivity per unit volume than scintillators or semiconductors because gases are less dense than solids [38]. An example of a gaseous detector can be found in [37].

Semiconductor detectors sense the moving charge due to electrons freed from their

orbits by interacting high–energy photons. Semiconductors offer higher sensitivity than gaseous detectors due to their larger density, and can offer better energy resolution. For example, Germanium detectors can exhibit less than 0.5% full–width half maximum (FWHM) energy resolution at 662 keV [32]. Germanium detectors, however, must be cooled to achieve such fine energy resolution. At room temperature, semiconductor detectors made of CdZnTe have been demonstrated with energy resolution of 0.77% FWHM [83]. Position–sensitive CdZnTe detectors were first demonstrated in 1999 [30]. These detectors are made from a rectangular block of semiconducting CdZnTe. An array of pixellated anodes are fabricated on one surface of a CdZnTe crystal. Signals induced on pixel anodes are read out by a multichannel application specific integrated circuit, which provides information about position and energy deposition of each gamma–ray interaction. Room–temperature semiconductor detectors have also been made from Silicon with lower gamma–ray sensitivity. [63].

2.2 Imaging Methods

An imaging detector provides information about the direction of origin of an interacting photon. This directional information is useful because it allows one to gain information about the spatial distribution of radiation–emitting materials around the detector. Photon emission rates of gamma rays are typically small compared to the rest of the electromagnetic spectrum. A traditional technology for radioactive source detection is the Geiger counter, which detects an increase in the detected photon rate, disregarding the directions of origin. Imaging information could be helpful in the detection of radioactive sources because, intuitively, a small increase in the number of photons in a narrow direction would be more noticeable than a small increase in the count rate disregarding directional information.

2.2.1 Compton Imaging Fundamentals

When a high-energy particle interacts with matter, two of the possible outcomes are photoelectric absorption and Compton scattering [79]. In photoelectric absorption, all of the incoming particle's energy is transferred to an atomic electron. In Compton scattering, the incoming particle imparts some of its energy onto an electron, ejecting it from the atom and scattering the incoming photon by a certain angle.

One can use the locations of a pair of photon interactions in a body of matter, where the first interaction is a Compton–scatter event, to find a cone of possible directions of origin. If a particle with energy E_0 deposits energy E_1 at the first interaction, and undergoes photoelectric absorption at a second location depositing energy E_2, then the angle θ between the incident direction and the line connecting the two interaction locations must satisfy

$$\cos(\theta) = 1 - \frac{m_e c^2 E_1}{E_0 E_2}, \tag{2.1}$$

where m_e is the mass of an electron and c is the speed of light [79]. Using (2.1), one one can use the interaction locations and energy deposited to determine a cone of possible source locations where the axis is the line connecting the two events and the angle is determined by the calculated scatter angle. Figure 2.1, shows a two-interaction event and the associated cone.

2.2.2 Parallel-Plate Compton Imaging

A classical detector technology that exploits the physics of Compton scatter is a parallel-plate detector, which is described in [42] and [54]. Figure 2.2 shows a two-dimensional illustration of such a system.

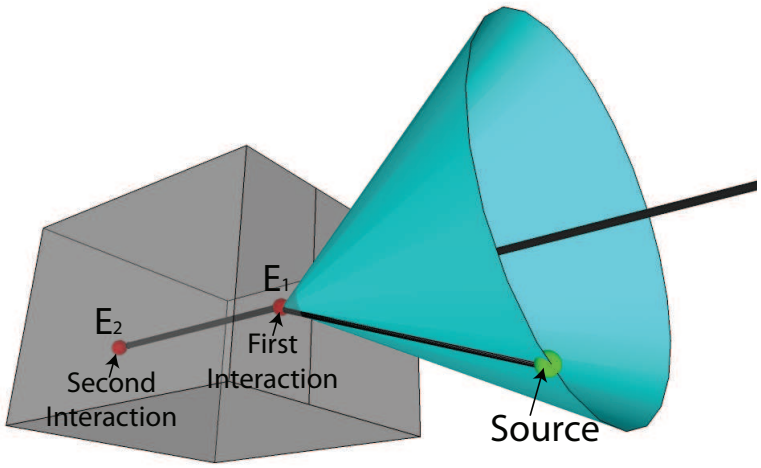


Figure 2.1: Cone of possible photon origin directions for a two-interaction event.

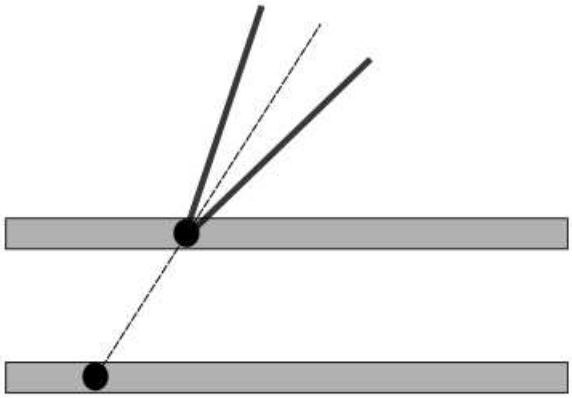


Figure 2.2: Parallel-plate Compton imaging system, interacting photon, and Compton cone

If an incoming photon scatters in the first plate and undergoes photoelectric absorption in the second, the Compton scatter formula yields a cone of possible directions of origin. To implement such a system, one does not need to determine the depth of interaction in the plate. However, if the plates are thin enough to ignore interaction depth, the escape probability is usually high, which leads to a low system sensitivity. There is a sensitivity-resolution tradeoff in the plate thickness because a thicker plate gives less precise interaction location information, but captures more photons. One means of increasing the detector sensitivity is to use one large block of

detector material instead of two plates. However, Compton imaging is only possible if the detector is depth–sensitive.

2.2.3 Position-Sensitive Compton Imaging

Position-sensitive Compton detectors made of the semiconductor CdZnTe, proposed in [30], are comprised of a bulk of CdZnTe and a biased capacitor whose plates lie on two of the opposite detector faces. The anode plate is pixelated and the cathode is not. When a photon interaction occurs, the voltage at the cathode changes almost immediately [79], and the voltage at the anode depends on the interaction depth and electron drift time. The time between a rise in the cathode signal and a rise in the anode signal allows one to determine the interaction depth. Also, the amplitude of the voltage signal at the anode allows one to determine the energy of the interacting photon because the amount of charge produced is proportional to the deposited energy. The signal induction in a biased semiconductor detector is described by the Shockley-Ramo Theorem, and are covered in detail in [29].

In reality, measurement uncertainty does not allow one to deterministically specify the cone of possible origin directions. The finite size of anode pixels leads to uncertainty in lateral interaction position, and electronic noise leads to uncertainty in interaction depth and deposited energy due to noise on the anode and cathode channels. The uncertainty in interaction position yields uncertainty in the cone axis and the uncertainty in the deposited energy yields uncertainty in the cone angle. Some image reconstruction and detection methods model this uncertainty to compute an estimate that is optimal in some sense.

When a photon scatters one or more times in the detector, the timing resolution of the system is insufficient to determine the order in which the interactions occurred. Sometimes one or more possible interaction sequences can be ruled out by using the

conservation of energy and (2.1). Choosing the best interaction sequence from the set of possible sequences is an active area of research and a comparison of two methods is given in [44].

2.2.4 Far–Field Approximation

In this work, we assume that the vertex of the cone of possible photon origin directions lies at the center of the detector. In reality, the vertex lies at the point of the first interaction, but for sources far from the detector, the error is negligible and the approximation greatly simplifies the geometric calculations. The far–field approximation applied to a source one meter from the detector will have an error in the cone axis of 0.6^o [44]. In detection problems, the far–field approximation allows one to treat the incoming photon directions as equal across the exposed detector surface. Most importantly, the far–field approximation allows us to geometrically determine the direction of origin without knowing the distance between the source and the detector, so we can pose imaging and detection problems in terms of direction of origin, disregarding the depth parameter.

2.2.5 Compton Imaging versus Collimation

An alternative approach to Compton imaging is to use a collimated detector. A collimated detector has thin strips of high-attenuating material that only allow photons from a particular direction to enter. Figure 2.3 illustrates a two-dimensional position-sensitive Compton detector and a collimated detector.

The "bin" in which a photon interacts in the collimated detector gives information about the direction of photon origin. The gray cone in Figure 2.3 illustrates the possible directions of origin for a measured photon. Note that this cone is solid and its half–angle is determined by the height of the collimator. Longer collimator

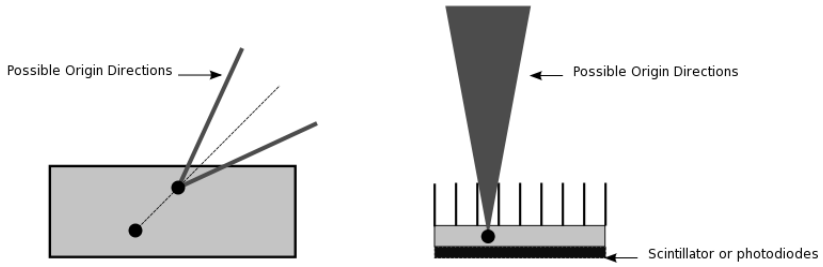


Figure 2.3: A position-sensitive Compton detector (left) and a collimated detector (right)

holes result in a narrower cone and the better the position resolution, but a longer collimator also blocks more photons, leading to lower sensitivity. This is another sensitivity-resolution tradeoff. In contrast, the position-sensitive Compton detector does not need a collimator to provide information about the photon direction of origin. Inference about the direction of photon origin from the recorded interaction positions is sometimes referred to as "electronic collimation." The cone of possible directions of origin is hollow for a position-sensitive Compton detector, whereas the cone for a collimated detector is solid.

Compton detectors that are not depth–sensitive typically consist of two parallel plates of scintillating material. Increasing the thickness of the scintillating plates increases the sensitivity, but degrades the interaction position resolution. However, increasing the thickness of the scintillating material in a collimated detector will not significantly degrade interaction position resolution because the depth of interaction is irrelevant. Inexpensive parallel–plate Compton detectors constructed from scintillating materials typically have a lower sensitivity than collimated scintillators for this reason.

2.3 Statistics of list–mode gamma–ray data

3D position–sensitive Compton detectors record the position and deposited energies of all photon interactions. The number of possible interaction coordinates and energies for an interacting photon is very large, so it is not practical to count the number of photons matching each possible interaction sequence [79]. Some imaging systems have a small set of possible attributes of interacting photons, and these systems typically operate in binned–mode [4], where the measurements are the numbers of photons matching each possible attribute. In this work we operate 3D position–sensitive detectors in list–mode [4], which means that the measurements are the recorded attributes (interaction positions and energies) of all interacting photons.

2.3.1 Recorded measurements

The measurements obtained from a gamma–ray detector consist a list of a random number of attribute vectors. Following the notation of [4], let $\tilde{\boldsymbol{r}} = (\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_J)$ be a list of J recorded attribute vectors from photon interaction events. Each element $\boldsymbol{r}_i$ of $\tilde{\boldsymbol{r}}$ is itself a vector of attributes describing the ith event. An example of a detector that one can describe with this model is a position–sensitive Compton detector. A Compton detector records a Poisson–distributed number J of gamma–ray photons. Each detected photon interacts one or more times inside the detector and the detector records these interaction locations and possibly other attributes such as deposited energy. The ith attribute vector $\boldsymbol{r}_i$ is comprised of these interaction positions and deposited energies.

In imaging problems where the number of observations has a Poisson distribution, the varying length of the list of recorded attributes for the observed events makes it more challenging to formulate a sample space. This complicates even conceptually

simple operations, such as expectation. In this section, we give an interpretation of the lists of recorded event attributes in which the sample space is clear and expectations are straightforward to compute. To rigorously define the sample space, it is helpful to view $\tilde{\boldsymbol{r}}$ as an infinitely long list of attributes rather than a list of finite, but random, length. The list of J recorded attributes is a finite–length collection of elements from $\tilde{\boldsymbol{r}}$.

We can view the randomness in the observations as a separate process whose outcome determines the number of observed events for a given scan time. We first define the random process governing the recorded attributes, and then define the process that determines the number of observed events and finally bring both processes together to derive the familiar density function for list–mode maximum–likelihood.

2.3.2 Process of recorded attributes

First, we define the random process governing the recorded attributes. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Consider the sequence of random independent and identically distributed (IID) event attributes $\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots$, where $\boldsymbol{r}_k \in \Omega$ for all $k \in \mathbb{N}$, with finite–dimensional distributions described by the probability measure

$$v_{n_1,\ldots,n_k}(H_1 \times H_2 \times \cdots \times H_k) = \mathsf{Pr}\left(\boldsymbol{r}_{n_1} \in H_1, \boldsymbol{r}_{n_2} \in H_2, \ldots, \boldsymbol{r}_{n_k} \in H_k\right),$$

on the measure space $(\Omega^k, \mathcal{F}^k)$, where $H_k \subseteq \Omega$. To continue, we assume independence of the recorded attributes.

Assumption II.1. *The attributes of distinct photon interaction events are IID.*

If Assumption II.1 is satisfied, we can write v as a product measure:

$$\begin{aligned} v_{n_1,\ldots,n_k}(H_1 \times H_2 \times \cdots \times H_k) &= \mathsf{Pr}\left(\boldsymbol{r}_{n_1} \in H_1\right) \mathsf{Pr}\left(\boldsymbol{r}_{n_2} \in H_2\right) \ldots \mathsf{Pr}\left(\boldsymbol{r}_{n_k} \in H_k\right) \\ &= v(H_1)v(H_2)\ldots v(H_k), \end{aligned} \tag{2.2}$$

where the probability measure v describes the attribute distribution of a single photon interaction event,

$$v(H) = \mathsf{Pr}\left(\boldsymbol{r}_k \in H\right). \tag{2.3}$$

The probability measure v does not depend on the sample index k by the assumption that the $\boldsymbol{r}_k$ are identically distributed. By the Radon–Nikodym theorem, there exists a function p such that

$$v(H) = \int_H p(x) d\mu(x),$$

for a measure μ such that v is absolutely continuous[1] with respect to μ. For example, if the $\boldsymbol{r}_k$ are discrete random variables, the counting measure would be an appropriate choice for μ. If the $\boldsymbol{r}_k$ are continuous random variables, then the Lebesgue measure would be an appropriate choice for μ.

One can show that $v_{n_1,\ldots,n_k}$ satisfies the hypothesis of the Kolmogorov extension theorem [8, p. 510], which states that there exists a probability measure P on the measure space $(\mathcal{R}^\infty, \mathcal{F}^\infty)$ such that the process$\{\boldsymbol{r}_k : k \in \mathbb{N}\}$ has $v_{n_1,\ldots,n_k}$ as its finite–dimensional distributions. The significance of this fact is that any finite collection of attributes has a well–defined distribution.

2.3.3 Event observation process

Assume that the process $\{\boldsymbol{r}_n\}$ is observed over a scan time τ. Let the random variable J_τ be the number of recorded events revealed after scan time τ, where

$$J_\tau \sim \text{Poisson}(\lambda_{\mathsf{s}}\tau),$$

and λ_{s} is a rate parameter in events per unit time. After scan time τ, the attributes $\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_{J_\tau}$ are revealed. Further assume that J_τ is independent of the process

[1] A σ-finite measure v is absolutely continuous with respect to the σ-finite measure μ on the measure space $(\Omega, \mathcal{F})$ if for each $A \in \mathcal{F}$, $v(A) = 0$ if $\mu(A) = 0$ [8, p. 442].

$\{\boldsymbol{r}_n\}$. Now, define the process of observed attributes $\boldsymbol{y}_n$, $n = 1, 2, \ldots$ by

$$\boldsymbol{y}_n \stackrel{\triangle}{=} (\boldsymbol{r}_n, d_n)\,,$$

where $d_n \in \{0, 1\}$ is equal to 1 if the event is observed and 0 otherwise.

Assume that the $\boldsymbol{y}_n$ are conditionally independent given J_τ for any integer $n \in \mathbb{N}$ with probability measure

$$\begin{aligned}&\mathsf{Pr}\left(\boldsymbol{y}_{n_1} \in H_1 \times D_1, \boldsymbol{y}_{n_2} \in H_2 \times D_2, \ldots, \boldsymbol{y}_{n_k} \in H_k \times D_k, J_\tau = J\right)\\ &= \prod_{j=1}^{k} \mathsf{Pr}\left(\boldsymbol{y}_{n_j} \in H_j \times D_j | J_\tau = J\right) \mathsf{Pr}\left(J_\tau = J\right),\end{aligned}$$

where

$$\mathsf{Pr}\left(\boldsymbol{y}_{n_j} \in H_j \times D_j | J_\tau = J\right) = \begin{cases} v(H_j)\mathsf{I}_{\{1\}}(D_j) & n_j \leq J \\ \mathsf{I}_{\{0\}}(D_j) & \text{else,} \end{cases}$$

and for sets A and B, the indicator function $\mathsf{I}_A(B)$ is defined by

$$\mathsf{I}_A(B) \stackrel{\triangle}{=} \begin{cases} 1 & A \subseteq B \\ 0 & \text{else.} \end{cases}$$

Let $c(x)$ be the counting measure. By the Radon–Nikodym theorem,

$$\mathsf{Pr}\left(\boldsymbol{y}_{n_j} \in H_j \times D_j | J_\tau = J\right) = \int_{H_j} \int_{D_j} \mathsf{q}(x, d | J; n_j) d\mu(x) dc(d),$$

where q is (except on a set of measure zero)

$$\mathsf{q}(x, d | J_\tau; n_j) = \begin{cases} 1 & d = 0, J_\tau > n_j \\ p(x) & d = 1, J_\tau \leq n_j \\ 0 & \text{else.} \end{cases}$$

Since J_τ is discrete,

$$\mathsf{Pr}\left(J_\tau \in \mathcal{J}\right) = \int_{\mathcal{J}} p_J(j) dc(j),$$

where

$$p_J(j) = e^{-\lambda_{\mathsf{s}}\tau}(\lambda_{\mathsf{s}}\tau)^j/j!$$

is the probability mass function of a Poisson random variable with mean $\lambda_{\mathsf{s}}\tau$.

2.3.4 Joint density of attributes and number of events

Let $n_1, \ldots, n_k$ be integers such that $n_1, n_2, \ldots, n_l \leq J_\tau < n_{l+1}, \ldots, n_k$. By Fubini's theorem [8, pp. 237-238], we can write the joint probability measure for a finite collection of attributes and J_τ by

$$\begin{aligned}
&\mathsf{Pr}\left(\boldsymbol{y}_{n_1} \in H_1 \times D_1, \boldsymbol{y}_{n_2} \in H_2 \times D_2, \ldots, \boldsymbol{y}_{n_k} \in H_k \times D_k, J_\tau \in \mathcal{J}\right) \\
&= \prod_{j=1}^{k} \mathsf{Pr}\left(\boldsymbol{y}_{n_j} \in H_j \times D_k | J_\tau\right) \mathsf{Pr}\left(J_\tau \in \mathcal{J}\right) \\
&= \int_{H_1 \times H_2 \times \cdots \times H_k \times \mathcal{J}} q(\boldsymbol{y}_1|j; n_1) \ldots q(\boldsymbol{y}_k|j; n_k) p_J(j) d\mu(\boldsymbol{y}_1) d\mu(\boldsymbol{y}_2) \ldots d\mu(\boldsymbol{y}_k) dc(j).
\end{aligned} \tag{2.4}$$

By definition, the integrand is the Radon-Nikodym density of the joint distribution of a finite collection of attributes and the number revealed J_τ with respect to the base measure $\mu^k \times c$.

By (2.4), we have that the joint Radon-Nikodym density $p(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_J, J_\tau)$ of the attributes $1, 2, \ldots, J_\tau$ and J_τ is, for some integer $N \in \mathbb{N}$,

$$\begin{aligned}
\mathsf{p}(\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_N, J_\tau) &= \prod_{j=1}^{N} q(\boldsymbol{y}_j|J_\tau; j) e^{-\lambda_{\mathsf{s}}\tau}(\lambda_{\mathsf{s}}\tau)^{J_\tau}/J_\tau! \\
&= \prod_{j=1}^{N} \mathsf{p}\left(\boldsymbol{r}_j\right) e^{-\lambda_{\mathsf{s}}\tau}(\lambda_{\mathsf{s}}\tau)^{J_\tau}/J_\tau! \quad \text{for} \quad N \leq J_\tau.
\end{aligned} \tag{2.5}$$

This joint Radon-Nikodym density is useful because one can replace the true attribute distribution p with a parameterized model distribution, in which case (2.5) is the likelihood.

2.4 System Model for 3D position–sensitive gamma–ray detector

The model for the probability distribution of recorded attributes in a position–sensitive gamma–ray detector is typically difficult to compute. In this section, we give a general model of $\mathsf{p}(\boldsymbol{r}|D; e_0, \boldsymbol{\phi})$ for two–interaction, full energy deposition events in a 3D detector, where D is the event that the gamma–ray interacts in the detector, $\boldsymbol{\phi}$ is the gamma-ray source position (assuming that the source is in the far–field), and e_0 is the source energy. This model is useful for likelihood-based image reconstruction and detection algorithms. This section gives an exact expression that one can approximate for implementation purposes.

2.4.1 The event D that a photon is detected

The letter D represents the event that a photon interacts with the detector. An event observed in a gamma–ray detector must have interacted if it was observed. Therefore, an accurate model for the a gamma–ray detector must condition the distribution of the observations on the event that the photon was recorded.

Consider the following one–dimensional illustrative example. Let a source of photons with unknown energy ϵ be located in the far–field so that its beam is parallel to a collimator on the face of a detector as shown in Figure 2.4. The detector has length d and linear attenuation coefficient $\mu(\epsilon)$, which is a function of energy. Assume that the detector only records the distance x from the collimated face of the detector when a photoelectric absorption event occurs. By the Beer-Lambert law [38], the probability density of an observation is

$$\mathsf{p}\left(x, D; \epsilon\right) = \mu(\epsilon) e^{-\mu(\epsilon) x} \quad 0 \leq x \leq d,$$

where the event D appears on the left–hand side of the semicolon because the event with a recorded interaction depth x between 0 and d implies that the photon was

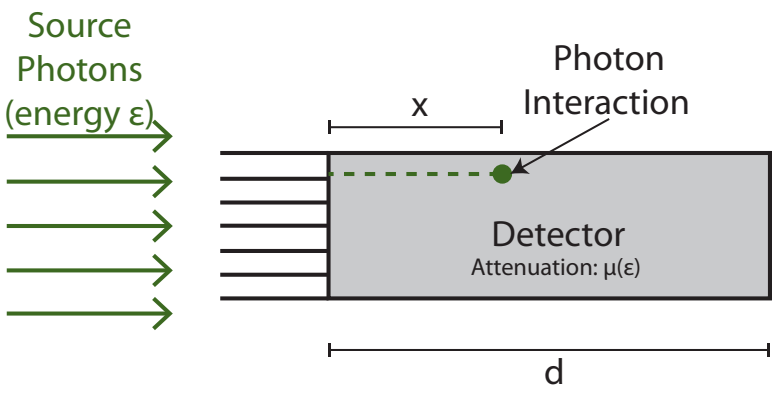


Figure 2.4: Illustration of a simple detector that records the depth of interaction from a collimated beam.

recorded. The probability that a photon is recorded is

$$\begin{aligned} \mathsf{p}\left(D;\epsilon\right) &= \int_0^d \mu(\epsilon) e^{-\mu(\epsilon)x} dx \\ &= 1 - e^{-\mu(\epsilon)d}. \end{aligned}$$

The conditional density of the observation x given that the photon was recorded is

$$\begin{aligned} \mathsf{p}\left(x|D;\epsilon\right) &= \mathsf{p}\left(x, D;\epsilon\right) / \mathsf{p}\left(D;\epsilon\right) \\ &= \frac{\mu(\epsilon) e^{-\mu(\epsilon)x}}{1 - e^{-\mu(\epsilon)d}}. \end{aligned}$$

Figure 2.5 shows an example of the source energy likelihood $\mathsf{p}\left(x, D;\epsilon\right)$ that is not conditioned on D and the likelihood $\mathsf{p}\left(x|D;\epsilon\right)$ that is conditioned on D for the detector described in Figure 2.4. The detector material is CdZnTe, the length $d = 2$ cm, and the interaction depth $x = 0.1$ cm. For this example, assume that the detector only records photoelectric absorption events. The likelihoods agree at low energies, where the probability that the photon is recorded is near one. At high energies, the linear mass attenuation coefficient $\mu(\epsilon)$ is much smaller, and the detector length is small compared to the mean–free–path $1/\mu(\epsilon)$. Figure 2.6 shows the attenuation coefficient $\mu(\epsilon)$ as a function of energy using the data in [5].

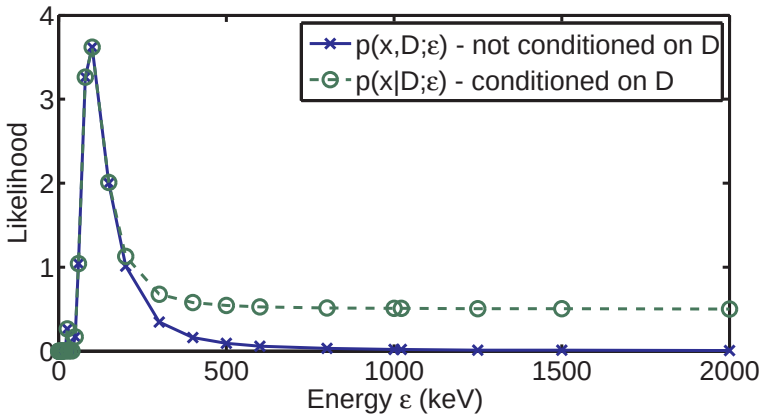


Figure 2.5: Likelihood of source energy with models conditioned on and not conditioned on D, the event that a photon was detected, for interaction depth $x = 0.1$ cm and detector length $d = 2$ cm.

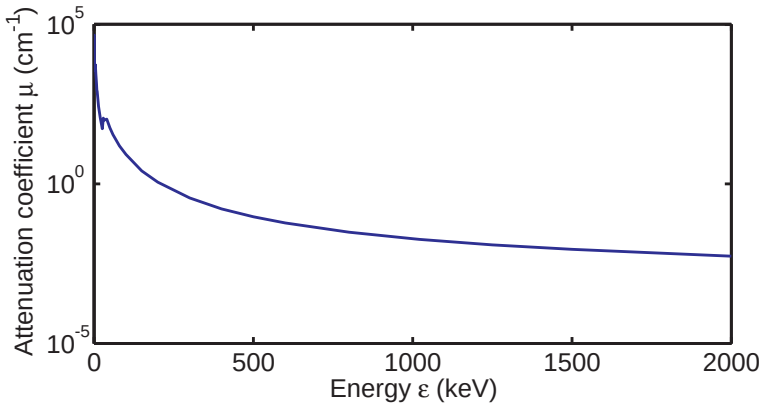


Figure 2.6: Linear attenuation coefficient for photoelectric absorption in CdZnTe.

Figure 2.7 shows the likelihoods as a function of energy for an observed interaction depth of 1.5 cm in a 2 cm detector. Intuitively, this interaction is likely to have a high energy because it traveled a large distance into the detector and the attenuation coefficient $\mu(\epsilon)$ is small for high energies. However, when the photon energy is large, there is a high probability that the photon will pass through the detector without interacting. Essentially, the likelihood that is not conditioned on D ignores the fact that the photon can pass though and not interact, so the likelihood is small at high energies. The likelihood conditioned on D is largest for high energies, consistent with intuition. This is because the likelihood conditioned on D uses the knowledge that

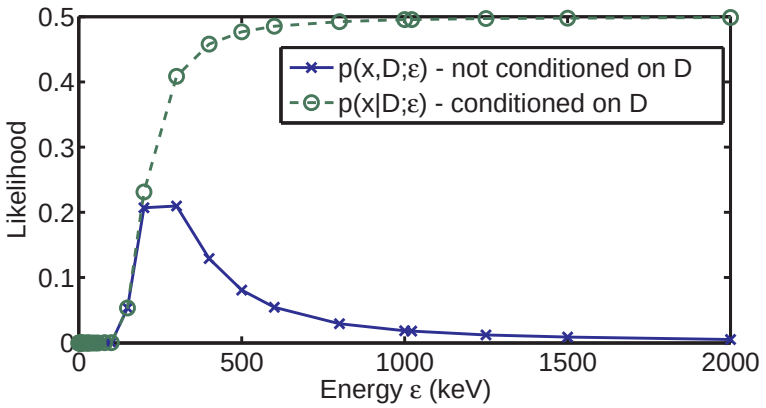


Figure 2.7: Likelihood of source energy with models conditioned on and not conditioned on D, the event that a photon was detected, for interaction depth $x = 1.5$ cm and detector length $d = 2$ cm.

the event was recorded.

This example shows that conditioning on the event that a photon was recorded is necessary to accurately model a gamma–ray measurement system.

2.4.2 General Model

We consider the case of a two–interaction full energy deposition event, where the first interaction is a Compton scatter event, and we denote its true position as $\boldsymbol{x}_1 \in \mathcal{X}$ where $\mathcal{X}$ is the set of all coordinates inside the detector. Let the true energy deposited in the first interaction be $e_1^t \in \mathcal{E}$. Since we are considering events with full energy deposition, the second interaction is a photoelectric absorption event. Let the true distance between the first and second interactions be $\rho \in P$ and the scatter direction $\psi \in \Psi$. The scatter direction ψ is not the Compton scatter angle, but rather it enumerates the possible scatter directions that are consistent with the Compton scatter angle. Figure 2.8 shows an example of a two–interaction event and the scatter angle ψ.

In 3D position–sensitive Compton detectors with pixelated anodes, the system records an anode pixel and depth for each interaction. Let $\boldsymbol{b}_1$ be the anode pixel

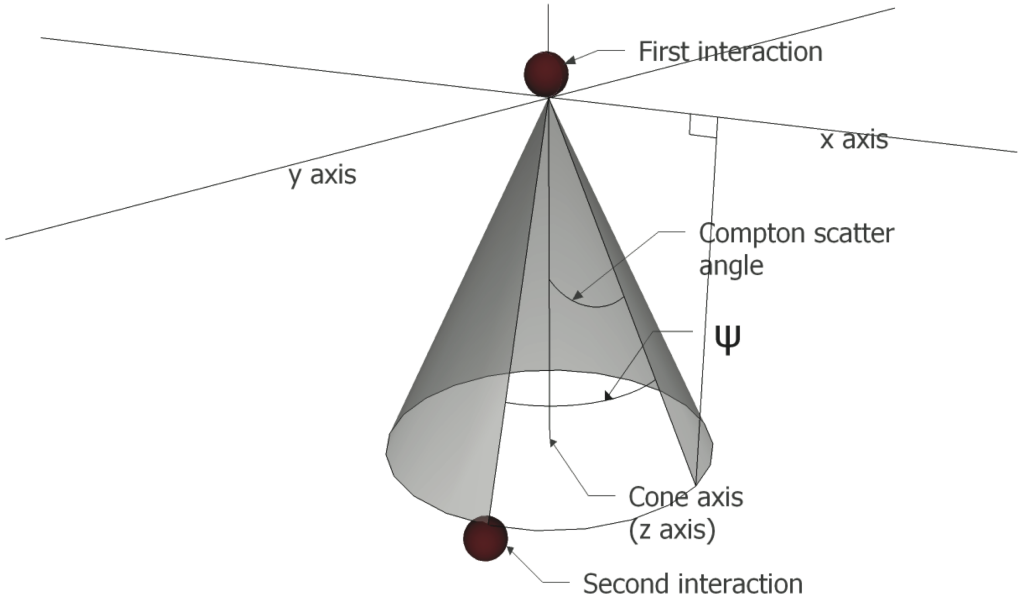


Figure 2.8: Illustration of the scatter direction angle ψ for a two–interaction Compton scatter event. The z-axis coincides with the incoming photon direction and the first interaction is at the origin.

number in which the first interaction is measured $\boldsymbol{b}_2$ be the anode pixel in which the second interaction is measured. Let z_1 and z_2 denote the interaction depth of the first and second interaction, respectively.

We separate the density function of the recorded attribute vector $\boldsymbol{r} = [\boldsymbol{b}_1, \boldsymbol{b}_2, z_1, z_2, e_1]$ into a component that depends on measurement uncertainty and another component that depends on physics. By total probability,

$$\mathsf{p}(\boldsymbol{r}|D; e_0, \boldsymbol{\phi}) = \int_{\mathcal{X}} \int_{\Psi} \int_{P} \int_{\mathcal{E}} \mathsf{p}(\boldsymbol{r}|\boldsymbol{x}_1, \psi, \rho, e_1^t, D; e_0, \boldsymbol{\phi}) \\ \mathsf{p}(\boldsymbol{x}_1, \psi, \rho, e_1^t|D; e_0, \boldsymbol{\phi}) de_1^t d\rho d\psi d\boldsymbol{x}_1.$$

2.4.3 Physics Modeling

In this section, we derive an expression for the portion of the model that corresponds to physics, namely $\mathsf{p}(\boldsymbol{x}_1, \psi, \rho, e_1^t|D; e_0, \boldsymbol{\phi})$. Let $D = D_1 \cap D_2$, where D_1 is the event that the first interaction is a recorded Compton scatter, and D_2 is the

event that the second interaction is a recorded photoelectric absorption. Using the definition of conditional probability and the chain rule,

$$\begin{aligned}\mathsf{p}(\boldsymbol{x}_1,\psi,\rho,e_1^t|D_1,D_2;e_0,\boldsymbol{\phi}) &= \mathsf{p}(\boldsymbol{x}_1,\psi,\rho,e_1^t,D_2|D_1;e_0,\boldsymbol{\phi})/\mathsf{p}(D_2|D_1;e_0,\boldsymbol{\phi})\\ &= \mathsf{p}(\rho|\psi,\boldsymbol{x}_1,e_1^t,D_1,D_2;e_0,\boldsymbol{\phi})\mathsf{p}(D_2|\psi,\boldsymbol{x}_1,e_1^t,D_1;e_0,\boldsymbol{\phi})\\ &\quad \mathsf{p}(\psi|\boldsymbol{x}_1,e_1^t,D_1;e_0,\boldsymbol{\phi})\mathsf{p}(e_1^t|\boldsymbol{x}_1,D_1;e_0,\boldsymbol{\phi})\mathsf{p}(\boldsymbol{x}_1|D_1;e_0,\boldsymbol{\phi})/\mathsf{p}(D_2|D_1;e_0,\boldsymbol{\phi}).\end{aligned} \tag{2.6}$$

The quantity $\mathsf{p}(D_2|D_1;e_0,\boldsymbol{\phi})$ is given by

$$\mathsf{p}(D_2|D_1;e_0,\boldsymbol{\phi}) = \int_{\mathcal{X}} \mathsf{p}(D_2|\boldsymbol{x}_1,D_1;e_0,\boldsymbol{\phi})\mathsf{p}(\boldsymbol{x}_1,|D_1;e_0,\boldsymbol{\phi})d\boldsymbol{x}_1, \tag{2.7}$$

where

$$\begin{aligned}&\mathsf{p}(D_2|\boldsymbol{x}_1,D_1;e_0,\boldsymbol{\phi})\\ &= \int_{\mathcal{E}}\int_{\Psi} \mathsf{p}(D_2|e_1^t,\psi,\boldsymbol{x}_1,D_1;e_0,\boldsymbol{\phi})\mathsf{p}(\psi|e_1^t,\boldsymbol{x}_1,D_1;e_0,\boldsymbol{\phi})\mathsf{p}(e_1^t|\boldsymbol{x}_1,D_1;e_0,\boldsymbol{\phi})d\psi,de_1^t.\end{aligned} \tag{2.8}$$

Since $\mathsf{p}(D_2|D_1;e_0,\boldsymbol{\phi})$ requires heavy computation, it may be beneficial to precompute the values and store them in a look-up-table to achieve shorter reconstruction times. Figure 2.9 shows an example of $\mathsf{p}(D_2|D_1;e_0,\boldsymbol{\phi})$ for a $2\times2\times1.5$ cm CdZnTe detector as a function of polar and azimuth angles for a source energy of 662 keV. This figure was generated by evaluating (2.7) and (2.8) numerically.

One can further simplify (2.6) as follows:

$$\begin{aligned}\mathsf{p}(\boldsymbol{x}_1,\psi,\rho,e_1^t|D_1,D_2;e_0,\boldsymbol{\phi}) &= \mathsf{p}(\rho,D_2|\psi,\boldsymbol{x}_1,e_1^t,D_1;e_0,\boldsymbol{\phi})\mathsf{p}(\psi|\boldsymbol{x}_1,e_1^t,D_1;e_0,\boldsymbol{\phi})\\ &\quad \mathsf{p}(e_1^t|\boldsymbol{x}_1,D_1;e_0,\boldsymbol{\phi})\frac{\mathsf{p}(\boldsymbol{x}_1,D_1;e_0,\boldsymbol{\phi})}{\mathsf{p}(D_1;e_0,\boldsymbol{\phi})}\left(\frac{\mathsf{p}(D_1,D_2;e_0,\boldsymbol{\phi})}{\mathsf{p}(D_1;e_0,\boldsymbol{\phi})}\right)^{-1}\\ &= \mathsf{p}(\rho,D_2|\psi,\boldsymbol{x}_1,e_1^t,D_1;e_0,\boldsymbol{\phi})\mathsf{p}(\psi|\boldsymbol{x}_1,e_1^t,D_1;e_0,\boldsymbol{\phi})\\ &\quad \mathsf{p}(e_1^t|\boldsymbol{x}_1,D_1;e_0,\boldsymbol{\phi})\mathsf{p}(\boldsymbol{x}_1,D_1;e_0,\boldsymbol{\phi})/\mathsf{p}(D_1,D_2;e_0,\boldsymbol{\phi}),\end{aligned} \tag{2.9}$$

where $\mathsf{p}(D_1,D_2;e_0,\boldsymbol{\phi})$ is the system sensitivity to two–interaction events where the second interaction is a photoelectric absorption.

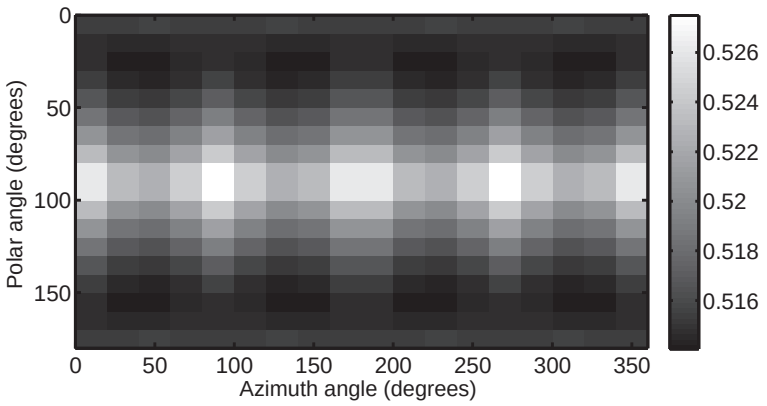


Figure 2.9: $\mathsf{p}(D_2|D_1; e_0, \boldsymbol{\phi})$ for a $2 \times 2 \times 1.5$ cm CdZnTe detector as a function of polar and azimuth angles for a source energy of 662 keV

Physical Assumptions

The expression in (2.6) does not show the independence and conditional independence that exist physically. Below are the simplifications that we can make because of independence:

- $\mathsf{p}(\psi|\boldsymbol{x}_1, e_1^t, D_1; e_0, \boldsymbol{\phi}) = \mathsf{p}\,(\psi|D_1)$,
- $\mathsf{p}(e_1^t|\boldsymbol{x}_1, D_1; e_0, \boldsymbol{\phi}) = \mathsf{p}(e_1^t|D_1; e_0)$.

Physical Quantities

Because we assume that the source is in the far field, we model the distribution of the first interaction location as

$$\mathsf{p}(\boldsymbol{x}_1, |D_1; e_0, \boldsymbol{\phi}) = \frac{\mu_c(e_0) e^{-\mu_c(e_0) d(\boldsymbol{x}_1; \boldsymbol{\phi})}}{\Omega(\boldsymbol{\phi}) \mathsf{p}(D_1; e_0, \boldsymbol{\phi})},$$

where $d(\boldsymbol{x}_1; \boldsymbol{\phi})$ is the distance from $\boldsymbol{x}_1$ to the edge of the detector along a line with direction $\boldsymbol{\phi}$ toward the source, $\Omega(\boldsymbol{\phi})$ is the surface area of the detector that is exposed to direct source photons, and $\mu_c(e_0)$ is the linear attenuation coefficient due to Compton scattering.

We use the Klein-Nishina formula to derive [38] $\mathsf{p}(e_1^t|D_1;e_0)$. The Klein-Nishina formula is given by

$$\frac{d\sigma}{d\Omega} = P(e_0,\theta)^2\left(P(e_0,\theta)+P(e_0,\theta)^{-1}-1+\cos^2(\theta)\right) \qquad 0<e_1^t<\frac{2e_0^2}{mc^2+2e_0}, \tag{2.10}$$

where m_e is the mass of an electron, c is the speed of light,

$$P(e_0,\theta)=\frac{1}{1+\frac{e_0}{mc^2}\left(1-\cos(\theta)\right)},$$

$\theta=\theta(e_1^t;e_0)$ is the Compton scatter angle given by

$$\theta(e_1^t;e_0)=\cos^{-1}\left(1-\frac{e_1^t m_e c^2}{e_0(e_0-e_1^t)}\right), \tag{2.11}$$

and $\frac{d\sigma}{d\Omega}$ is the differential cross section per unit volume, which is proportional to the expected number of scatters per solid angle around the atom. To compute $\mathsf{p}(e_1^t|D_1;e_0)$, we must first compute

$$\frac{d\sigma}{de_1^t}=\frac{d\sigma}{d\Omega}\frac{d\Omega}{de_1^t}. \tag{2.12}$$

Spherical geometry gives

$$d\Omega=\sin(\theta)d\theta d\phi. \tag{2.13}$$

Let $e_2=e_0-e_1^t$ and use (2.11) to compute $\frac{de_1^t}{d\theta}$:

$$\begin{aligned}\frac{de_1^t}{d\theta} &= \frac{d}{d\theta}(e_0-e_2)\\ &= \frac{d}{d\theta}\frac{e_0}{1+\frac{e_0}{m_ec^2}(1-\cos(\theta))}\\ &= \frac{\frac{e_0^2}{m_ec^2}\sin(\theta)}{\left(1+\frac{e_0}{m_ec^2}(1-\cos(\theta))\right)^2}\\ &= \frac{e_2^2}{m_ec^2}\sin(\theta). \end{aligned} \tag{2.14}$$

Substituting (2.13) and (2.14) into (2.12), we obtain:

$$\mathsf{p}(e_1^t|D_1;e_0) \propto \frac{d\sigma}{d\Omega}\frac{d\Omega}{d\theta}\frac{d\theta}{de_1^t} = \frac{2\pi m_e c^2}{(e_0-e_1^t)^2}\frac{d\sigma}{d\Omega}, \tag{2.15}$$

where $\frac{d\sigma}{d\Omega}$ is given by (2.10).

We model the scatter direction ψ as uniform on $[0, 2\pi]$:

$$\mathsf{p}\left(\psi|D_1\right) = \frac{1}{2\pi} \quad 0 \leq \psi \leq 2\pi. \tag{2.16}$$

The density of the distance between interactions is given by

$$\mathsf{p}(\rho|\psi,\boldsymbol{x}_1,e_1^t,D_1,D_2;e_0,\boldsymbol{\phi}) = \frac{\mu_p(e_0-e_1^t)e^{-\mu_p(e_0-e_1^t)\rho}}{1-e^{-\mu_p(e_0-e_1^t)l(\boldsymbol{x}_1,\boldsymbol{\phi},\psi,\theta)}} \quad 0 \leq \rho \leq l(\boldsymbol{x}_1,\boldsymbol{\phi},\psi,\theta), \tag{2.17}$$

where $l(\boldsymbol{x},\boldsymbol{\phi},\psi,\theta)$ is the Euclidean distance from the vector $\boldsymbol{x}$ to the edge of the detector along the path of the scattered photon with direction $\boldsymbol{\phi}$ before the first interaction and a Compton scatter characterized by θ and ψ, and $\mu_p(e)$ is the linear attenuation coefficient for photoelectric absorption at energy e.

The term $\mathsf{p}(D_2|\psi,\boldsymbol{x}_1,e_1^t,D_1;e_0,\boldsymbol{\phi})$ is given by the Beer-Lambert law using the distance from $\boldsymbol{x}_1$ to the edge of the detector in the direction of the scattered photon:

$$\mathsf{p}(D_2|\psi,\boldsymbol{x}_1,e_1^t,D_1;e_0,\boldsymbol{\phi}) = 1-e^{-\mu_p(e_0-e_1^t)l(\boldsymbol{x}_1,\boldsymbol{\phi},\psi,\theta)}. \tag{2.18}$$

In this case,

$$\begin{aligned}\mathsf{p}(\rho|\psi,\boldsymbol{x}_1,e_1^t,D_1,D_2;e_0,\boldsymbol{\phi})\mathsf{p}(D_2|\psi,\boldsymbol{x}_1,e_1^t,D_1;e_0,\boldsymbol{\phi}) &= \mathsf{p}(\rho,D_2|\psi,\boldsymbol{x}_1,e_1^t,D_1;e_0,\boldsymbol{\phi}) \\ &= \mu_p(e_0-e_1^t)e^{-\mu_p(e_0-e_1^t)\rho}.\end{aligned}$$

2.4.4 Measurement Uncertainty

Assuming that the uncertainties due to position and energy are conditionally independent given the unobserved exact positions and energy,

$$\begin{aligned}\mathsf{p}(\boldsymbol{b}_1,\boldsymbol{b}_2,e_1|\boldsymbol{x}_1,\psi,\rho,e_1^t,D;e_0,\boldsymbol{\phi}) &= \mathsf{p}\left(\boldsymbol{b}_1,\boldsymbol{b}_2|\psi,\rho,\boldsymbol{x}_1,D\right)\mathsf{p}(e_1|e_1^t,D;e_0,\boldsymbol{\phi}) \\ &= \mathsf{p}\left(\boldsymbol{b}_2|\psi,\rho,\boldsymbol{x}_1,D\right)\mathsf{p}\left(\boldsymbol{b}_1|\boldsymbol{x}_1,D\right)\mathsf{p}(e_1|e_1^t,D;e_0,\boldsymbol{\phi}).\end{aligned}$$

Compton detectors are typically pixelated, so the available measurements are the waveforms at the individual pixels. For simplicity, assume that the measured quantity is the pixel location where the interaction occurred. Let K denote the number of pixels and let $\mathcal{B}_k^l = \{\boldsymbol{x}_l : \boldsymbol{x}_l \text{ lies in pixel } k\}$ for $l = 1, 2$, and $k = 1, \ldots, K$. For example, in a Compton detector with 11×11 pixels, $k = 1, 2, \ldots, 121$, where the k enumerate the pixels. For the first interaction, let the random variable $\boldsymbol{b}_1$ be defined such that

$$\boldsymbol{b}_1 = \begin{cases} k & \boldsymbol{x}_1 \in \mathcal{B}_k^1 \\ 0 & \text{else.} \end{cases}$$

Similarly, define the random variable $\boldsymbol{b}_2$ describing the pixel containing the second interaction as

$$\boldsymbol{b}_2 = \begin{cases} k & \boldsymbol{x}_2 \in \mathcal{B}_k^2 \\ 0 & \text{else.} \end{cases}$$

Define the sets $\mathcal{R}_1$ and $\mathcal{R}_2$ as

$$\mathcal{R}_1 = \{\boldsymbol{b}_1 : \mathsf{p}(\boldsymbol{b}_1) > 0\}$$

$$\mathcal{R}_2 = \{\boldsymbol{b}_2 : \mathsf{p}(\boldsymbol{b}_2) > 0\}.$$

In the case where the exact interaction location is quantized,

$$\mathsf{p}(\boldsymbol{b}_1 = k | \boldsymbol{x}_1, D) = \begin{cases} 1 & \boldsymbol{x}_1 \in \mathcal{B}_k^1 \\ 0 & \text{else,} \end{cases} \tag{2.19}$$

and

$$\mathsf{p}(\boldsymbol{b}_2 | \psi, \rho, \boldsymbol{x}_1, D) = \begin{cases} 1 & \boldsymbol{x}_2(\boldsymbol{x}_1, \boldsymbol{\phi}, \theta, \psi) \in \mathcal{B}_k^2 \\ 0 & \text{else.} \end{cases} \tag{2.20}$$

A common model for the energy uncertainty $\mathsf{p}(e_1|e_1^t, D; e_0, \boldsymbol{\phi}) = \mathsf{p}\left(e_1|e_1^t\right)$ is the normal pdf with mean e_1^t and variance σ_e^2. Also, one could account for Doppler broadening in $\mathsf{p}\left(e_1|e_1^t\right)$.

2.4.5 Entire Model

Assuming that the only position uncertainty is due to quantization, the distribution of recorded attributes $\boldsymbol{r} = (\boldsymbol{b}_1, \boldsymbol{b}_2, e_1)$ is given by

$$\begin{aligned}\mathsf{p}(\boldsymbol{b}_1, \boldsymbol{b}_2, e_1|D; e_0, \boldsymbol{\phi}) \;=\; & \int_{\mathcal{X}}\int_{\Psi}\int_{P}\int_{\mathcal{E}} \mathsf{p}\left(\boldsymbol{b}_2|\psi, \rho, \boldsymbol{x}_1, D\right)\mathsf{p}\left(\boldsymbol{b}_1|\boldsymbol{x}_1, D\right)\mathsf{p}\left(e_1|e_1^t\right) \\ & \mathsf{p}(\rho|\psi, \boldsymbol{x}_1, e_1^t, D_1, D_2; e_0, \boldsymbol{\phi})\mathsf{p}(D_2|\psi, \boldsymbol{x}_1, e_1^t, D_1; e_0, \boldsymbol{\phi}) \\ & \mathsf{p}\left(\psi|D_1\right)\mathsf{p}(e_1^t|D_1; e_0)\mathsf{p}(\boldsymbol{x}_1|D_1; e_0, \boldsymbol{\phi}) \\ & /\mathsf{p}(D_2|D_1; e_0, \boldsymbol{\phi}) de_1^t d\rho d\psi d\boldsymbol{x}_1.\end{aligned}$$

Changing the order of integration and substituting (2.16) and (2.19),

$$\begin{aligned}\mathsf{p}(\boldsymbol{b}_1, \boldsymbol{b}_2, e_1|D; e_0, \boldsymbol{\phi}) = & \frac{1}{2\pi\mathsf{p}(D_2|D_1; e_0, \boldsymbol{\phi})}\int_{\mathcal{E}} \mathsf{p}\left(e_1|e_1^t\right)\mathsf{p}(e_1^t|D_1; e_0) \\ & \int_{\mathcal{B}_k^1} \mathsf{p}(\boldsymbol{x}_1|D_1; e_0, \boldsymbol{\phi}) \int_{\Psi} \mathsf{p}(D_2|\psi, \boldsymbol{x}_1, e_1^t, D_1; e_0, \boldsymbol{\phi}) \\ & \int_{P} \mathsf{p}\left(\boldsymbol{b}_2|\psi, \rho, \boldsymbol{x}_1, D\right)\mathsf{p}(\rho|\psi, \boldsymbol{x}_1, e_1^t, D_1, D_2; e_0, \boldsymbol{\phi}) d\rho d\psi d\boldsymbol{x}_1 de_1^t. \qquad (2.21)\end{aligned}$$

It is not practical to evaluate (2.21) directly, and future work could explore approximations such as the saddle point approximation.

An alternative to (2.21) based on (2.9) is

$$\begin{aligned}\mathsf{p}(\boldsymbol{b}_1, \boldsymbol{b}_2, e_1|D; e_0, \boldsymbol{\phi}) = & \frac{1}{2\pi\mathsf{p}(D; e_0, \boldsymbol{\phi})}\int_{\mathcal{E}} \mathsf{p}\left(e_1|e_1^t\right)\mathsf{p}(e_1^t|D_1; e_0) \\ & \int_{\mathcal{B}_k^1} \mathsf{p}(\boldsymbol{x}_1, D_1; e_0, \boldsymbol{\phi}) \int_{\Psi} \\ & \int_{P} \mathsf{p}\left(\boldsymbol{b}_2|\psi, \rho, \boldsymbol{x}_1, D\right)\mathsf{p}(\rho, D_2|\psi, \boldsymbol{x}_1, e_1^t, D_1; e_0, \boldsymbol{\phi}) d\rho d\psi d\boldsymbol{x}_1 de_1^t. \qquad (2.22)\end{aligned}$$

2.4.6 Doppler Broadening

One source of energy uncertainty is due to the random momentum of the recoil electron before a photon strikes it. This phenomenon is known as Doppler broadening and can lead to significant uncertainty in the Compton scatter angle. An approximate model for the differential Compton cross section accounting for Doppler broadening is given in [56]

$$\begin{aligned}\frac{d^2\sigma}{d\Omega de_1} &= \frac{1}{2\pi}\mathsf{p}\left(D, e_1; e_0\right) \\ &= \frac{mr_0^2}{2e_0}\left(\frac{e_1}{e_0}+\frac{e_0}{e_1}-\sin^2(\theta(e_1^t;e_0))\right) \\ &\quad \frac{e_0-e_1}{\sqrt{e_0^2+(e_0-e_1)^2-2e_0(e_0-e_1)\cos(\theta(e_1^t;e_0))}}J(p_z),\end{aligned} \tag{2.23}$$

where $e_2 = e_0 - e_1$, Ω is the solid angle, r_0 is the radius of an electron, $e_1 + e_2 = e_0$, $\theta(e_1^t; e_0)$ is the Compton scatter angle given by (2.11), and J is the momentum distribution for electrons in the material, tabulated in [7], and the electron momentum p_z is given by

$$p_z = -mc\frac{e_1 - e_0(e_0-e_1)(1-\cos(\theta(e_1^t;e_0)))/mc^2}{\sqrt{e_0^2+(e_0-e_1)^2-2e_0e_2\cos(\theta(e_1^t;e_0))}}. \tag{2.24}$$

Since $J(\cdot)$ depends on the particular element and CZT is made up of three elements, the $J(\cdot)$ used in our calculations is the average of the distributions for Cd, Zn, and Te.

Figure 2.10 shows the system model with and without Doppler broadening for $e_0 = 662$ keV for an detector of infinitely small depth. The infinitely small depth assumption simplifies computation of (2.21) by only considering events that scatter parallel to the anode and cathode of the detector. The model with Doppler broadening is nonzero over a larger support than the model without Doppler broadening. Figure 2.11 shows the system model for for another event where the interactions are

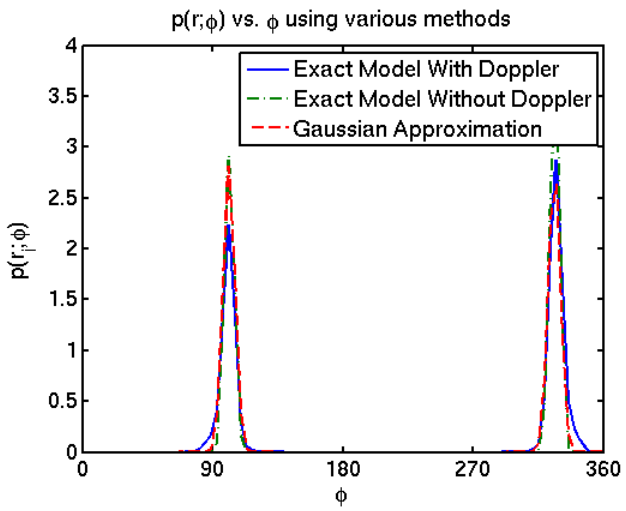


Figure 2.10: Likelihood using 2D model for a 20mm×20mm×15mm CdZnTe detector with 11×11 penalization for event with first interaction pixel (11,5), second interaction pixel (5,1), and deposited energy of 294.2 keV in the first interaction. The Gaussian approximation was calculated using [80].

much closer together. Here, the system model with and without Doppler broadening are nearly identical to the eye. The short distance between interactions leads to more uncertainty due to position quantization, and this uncertainty dominates for this event.

2.5 Estimation Methods

This section gives background on the theory of estimation, which is the process of inferring information about a true parameter through the realizations of a random variable whose distribution is governed by the parameter. We also outline iterative algorithms that are used when there is no closed form for an estimator.

2.5.1 Maximum Likelihood Estimation

In many statistical problems, a series of observations are known to come from a certain distribution, but this distribution may be governed by unknown, but nonrandom parameters. A common approach to this problem, called Maximum Likelihood (ML) estimation [34], is to find the parameter θ that maximizes the likelihood of the

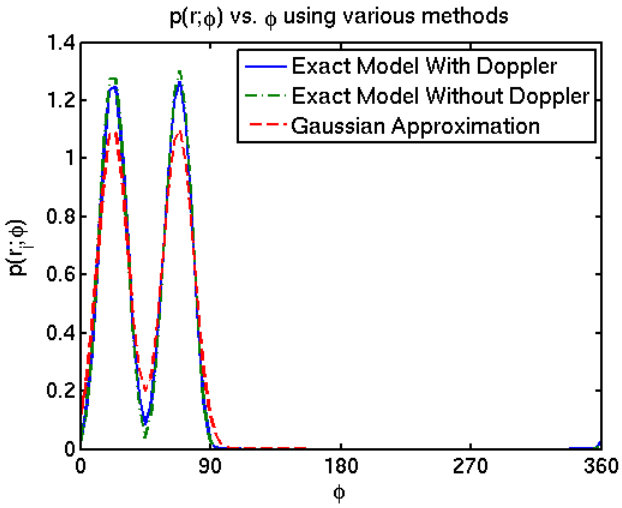


Figure 2.11: Likelihood using 2D model for a 20mm×20mm×15mm CdZnTe detector with 11×11 penalization for event with first interaction pixel (10,10), second interaction pixel (8,8), and deposited energy of 260.7 keV in the first interaction. The Gaussian approximation was calculated using [80].

observations $\boldsymbol{x}$, where the likelihood function is defined to be $f(\boldsymbol{x};\theta)$. Formally, the ML estimate is the solution to the optimization problem

$$\hat{\theta} = \arg\max_{\theta} f(\boldsymbol{x};\theta). \tag{2.25}$$

Since it is often easier to work with the natural logarithm of the likelihood, and the logarithm is a monotone transformation, the ML estimate is commonly expressed as

$$\hat{\theta} = \arg\max_{\theta} \ \log f(\boldsymbol{x};\theta). \tag{2.26}$$

In the literature, it is also common to pose (2.26) as the minimization problem

$$\hat{\theta} = \arg\min_{\theta} \ -\mathsf{L}(\theta), \tag{2.27}$$

where $\mathsf{L}(\theta) \stackrel{\triangle}{=} \log f(\boldsymbol{x};\theta)$. The ML estimator has many desirable properties including asymptotic normality, asymptotic efficiency, and asymptotic unbiasedness [34]. These desirable properties, however, only hold asymptotically, so the ML estimate may not be the most useful estimate in every application. This shortcoming is partially addressed by penalized–likelihood algorithms.

In the image reconstruction problem, the parameters to be estimated are the pixels in the image. In this case, finding the solution to (2.25) usually involves minimizing a coupled set of equations that can only be solved using iterative methods.

2.5.2 Penalized Likelihood Estimation

If one has *a priori* knowledge about a parameter to be estimated, one can penalize deviations in the estimate from the *a priori* expectations. These algorithms find the solution to the following penalized likelihood problem:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} -\mathsf{L}(\boldsymbol{x}) + \beta R(\boldsymbol{x}), \tag{2.28}$$

where $R(\boldsymbol{x})$ is a function known as a penalty or regularizer, and β is a parameter that determines how strictly the penalty is enforced. One can view the penalty $R(\boldsymbol{x})$ as the logarithm of a Bayesian prior distribution on the parameter. However, unlike in the Bayesian setting, $e^{-\beta R(\boldsymbol{x})}$ does not have to be a probability distribution.

Penalized likelihood estimation is common in image reconstruction because, *a priori*, medical and natural images have sparse gradients or are sparse in some basis or frame [12]. In these settings, one can formulate a penalty function that increases as the estimate deviates from this expectation. Penalized likelihood algorithms are especially appealing for Compton imaging because the images are known *a priori* to be sparse in the space domain.

2.5.3 Iterative Minimization of Cost Functions

In most image reconstruction problems, there is no analytical solution to (2.27) or (2.28), so one must resort to iterative methods to find a solution. Optimization transfer [40] is a general framework in which one can view all image reconstruction algorithms in this work. Suppose we have a cost function $\Psi(\boldsymbol{x})$ where $\boldsymbol{x} \in \mathbb{R}^n$. An optimization transfer method is one that minimizes $\Psi(\boldsymbol{x})$ by minimizing a series

of surrogate functions $Q^{(n)}\left(\boldsymbol{x};\boldsymbol{x}^{(n)}\right)$ that are easier to minimize. An optimization transfer algorithm iterates between minimizing the current surrogate function and finding a new surrogate function that allows the minimizer in the next iteration to be closer to the minimizer of $\Psi(\boldsymbol{x})$. All optimization transfer derivations in this work use surrogate functions that satisfy

$$\begin{aligned} Q^{(n)}\left(\boldsymbol{x};\boldsymbol{x}^{(n)}\right) &\geq \Psi(\boldsymbol{x}) \\ Q^{(n)}\left(\boldsymbol{x};\boldsymbol{x}^{(n)}\right) &= \Psi(\boldsymbol{x}^{(n)}). \end{aligned} \tag{2.29}$$

These conditions ensure that the sequence of estimates $\{\boldsymbol{x}^{(n)}\}$ is monotonically decreasing. Formally, if we denote the estimate at iteration n as $\boldsymbol{x}^{(n)}$, an optimization transfer algorithm follows the following procedure [24]:

$$\begin{aligned} &\text{S Step: choose a surrogate } Q^{(n)}\left(\boldsymbol{x};\boldsymbol{x}^{(n)}\right) \\ &\text{M Step: } \boldsymbol{x}^{(n+1)} = \arg\min_{\boldsymbol{x}} Q^{(n)}\left(\boldsymbol{x};\boldsymbol{x}^{(n)}\right). \end{aligned} \tag{2.30}$$

There are a number of tools commonly used to derive surrogate functions. One can derive the algorithms in this work using the optimization transfer framework by exploiting the convexity inequality, following the methodology of [15].

2.5.4 The EM Algorithm

Expectation Maximization (EM) algorithms, which were first proposed in [17], are a common approach for iteratively finding the ML estimate when no analytical solution is available. The EM algorithm has been previously applied to the image reconstruction problem for Compton Imaging [81]. EM algorithms are optimization transfer algorithms with a specific surrogate function. Before we give an expression for the surrogate, we must define the concept of "complete data." If the observed data is a random vector $\boldsymbol{Y}$, with realizations $\boldsymbol{y}$, from a random process governed by

parameter vector $\boldsymbol{\theta}$, then one chooses a complete data vector $\boldsymbol{Z}$, with realizations $\boldsymbol{z}$, such that

$$\mathsf{p}(\boldsymbol{y}, \boldsymbol{z}; \boldsymbol{\theta}) = \mathsf{p}(\boldsymbol{y}|\boldsymbol{z})\mathsf{p}(\boldsymbol{z}; \boldsymbol{\theta}), \tag{2.31}$$

and the likelihood based on the complete data $\mathsf{p}(\boldsymbol{z}; \boldsymbol{\theta})$ is easier to maximize than the likelihood based on the measured data $\mathsf{p}(\boldsymbol{y}; \boldsymbol{\theta})$. Using these definitions, the EM algorithm is an optimization transfer algorithm with surrogate function

$$Q^{(n)}\left(\boldsymbol{x}; \boldsymbol{x}^{(n)}\right) = \mathsf{E}\left[\log \mathsf{p}(\boldsymbol{z}; \boldsymbol{\theta}) | \boldsymbol{Y} = \boldsymbol{y}; \boldsymbol{\theta}^{(n)}\right]. \tag{2.32}$$

2.5.5 Cramer-Rao Lower Bound

Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ be IID random vectors with probability density or mass function $\mathsf{p}(\boldsymbol{x}; \boldsymbol{\theta})$ for some value of $\boldsymbol{\theta}$. The Cramer–Rao lower bound states that, under certain regularity conditions [34], any unbiased estimator $\hat{\boldsymbol{\theta}}$ of a parameter vector $\boldsymbol{\theta}$ satisfies

$$\mathsf{Cov}\left(\hat{\boldsymbol{\theta}}\right) - \mathsf{F}^{-1}(\boldsymbol{\theta}) \text{ is positive definite,}$$

where

$$\mathsf{F}(\boldsymbol{\theta}) \triangleq \mathsf{E}\left[\nabla_{\boldsymbol{\theta}} \log \mathsf{p}(\boldsymbol{x}; \boldsymbol{\theta}) \nabla_{\boldsymbol{\theta}}^T \log \mathsf{p}(\boldsymbol{x}; \boldsymbol{\theta})\right] \tag{2.33}$$

$$= -\mathsf{E}\left[\nabla_{\boldsymbol{\theta}}^2 \log \mathsf{p}(\boldsymbol{x}; \boldsymbol{\theta})\right], \tag{2.34}$$

is the Fisher information matrix.

2.6 Detection Theory

Detection is another approach to extracting useful information from a Compton imaging detector. In this section, we give an overview of detection methods and analysis. Suppose that a random vector of observations $\boldsymbol{x}$ has the density[2] $f(\boldsymbol{x}; \boldsymbol{\theta})$,

[2]Throughout, the word density precisely refers to the Radon-Nikodym density with respect to the appropriate base measure [8], which is general enough to include continuous, discrete, and mixed random variables.

and we want to decide which of two possible parameter values $\boldsymbol{\theta}$ governs the distribution that produced the observed random vector. The question is posed as a test between two "hypotheses" H_1 and H_0, where we define

$$H_1 : \ \boldsymbol{\theta} = \boldsymbol{\theta}_1$$

$$H_0 : \ \boldsymbol{\theta} = \boldsymbol{\theta}_0.$$

The probability of false alarm, or "level," is denoted α_0 and is the probability that a test decides that H_1 is true when H_0 is actually true. The probability of detection is the probability that the test decides that H_1 is true when H_1 is actually true. The Neyman-Pearson Lemma [35] states that the test with highest probability of detection for a constrained probability of false alarm has the form

$$\Lambda(\boldsymbol{x}) = \frac{f(\boldsymbol{x};\boldsymbol{\theta}_1)}{f(\boldsymbol{x};\boldsymbol{\theta}_0)} \underset{H_0}{\overset{H_1}{\gtrless}} \eta, \tag{2.35}$$

where η is chosen such that $\mathsf{Pr}\left(\Lambda(\boldsymbol{x}) > \eta | H_0\right) \leq \alpha_0$. A graph of the probability of detection versus the probability of false alarm is called the receiver operating characteristic (ROC) curve [74]. The area under this curve is called the AUC, which quantifies the ambiguity between H_1 and H_0. For a more thorough treatment of ROC curves and the AUC, see [28].

2.6.1 Test statistics for source detection

In the gamma–ray source detection problem, we would like to discern between two hypotheses:

$$\begin{aligned} H_0 &: \ \alpha = 0 \\ H_1 &: \ \alpha > 0, \end{aligned} \tag{2.36}$$

where α is a gamma–ray source intensity. There are several reasonable tests for this problem.

Generalized Likelihood Ratio Test (GLRT)

The GLRT is a common method of choosing between two hypotheses when one or more of the hypotheses depends on unknown parameters [35]. We can write the GLRT as

$$2\log\Lambda_{\mathrm{GLRT}} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma, \tag{2.37}$$

where the GLRT test statistic is

$$\Lambda_{\mathrm{GLRT}} \triangleq \frac{\max_{\boldsymbol{\theta}} \mathsf{p}(\tilde{\boldsymbol{r}};\boldsymbol{\theta},H_1)}{\max_{\boldsymbol{\theta}} \mathsf{p}(\tilde{\boldsymbol{r}};\boldsymbol{\theta},H_0)}. \tag{2.38}$$

To calibrate the test threshold γ and analyze the performance of the detector, one must determine the distribution of Λ_{GLRT}. If one uses a restricted ML estimator (MLE) motivated by the one–sided hypothesis in (2.36), the distribution of Λ_{GLRT} is complicated because under H_0, the parameter α lies on the boundary of the parameter space when the source intensity is constrained to be nonnegative [62]. To simplify the analysis of the GLRT and give intuition, instead of (2.36), we consider the two–sided test

$$\begin{aligned} H_0 &: \alpha = 0 \\ H_1 &: \alpha \neq 0, \end{aligned} \tag{2.39}$$

which is also the basis of the analysis in [52]. When the test is treated with the two–sided formulation in (2.39) and the source intensity is small, it is shown in [35, pp. 239-240] that the asymptotic distribution of the test statistic is given by

$$2\log\left(\Lambda_{\mathrm{GLRT}}\right) \sim \begin{cases} \chi_1^2\left(\eta\right), & H_1 \\ \chi_1^2\left(0\right), & H_0, \end{cases} \tag{2.40}$$

where $\chi_1^2(\eta)$ denotes the non–central chi-square distribution with one degree of freedom and noncentrality parameter η. Assuming $\mathsf{F}(\boldsymbol{\theta})$ is invertible, for the model (4.4), the noncentrality parameter is

$$\eta = \alpha^2 \left(\mathsf{F}^{-1}(\boldsymbol{\theta})_{[1,1]} \right)^{-1}. \tag{2.41}$$

This asymptotic framework is also used to analyze a test of whether one or two sources are present in [52].

Wald Test

The Wald test is a classical test for the composite hypothesis problem and is known to be approximately asymptotically equivalent to the GLRT for small source intensities [35, p. 188]. The Wald test for the source detection problem is given by

$$\Lambda_{\mathrm{W}} = \hat{\alpha}_{\mathrm{ML}}^2 \left(\mathsf{F}^{-1}\left(\hat{\boldsymbol{\theta}}\right)_{[1,1]} \right)^{-1}, \tag{2.42}$$

where $\hat{\boldsymbol{\theta}}$ is the ML estimate of $\boldsymbol{\theta}$. The Wald test is a non–monotonic function of $\hat{\alpha}_{\mathrm{ML}}$ when $\hat{\alpha}_{\mathrm{ML}}$ is the unrestricted MLE of α. Although the source intensity cannot be physically less than zero, one can find $\hat{\alpha}_{\mathrm{ML}}$ by maximizing the likelihood over the real line. Intuitively, a negative estimate of the source intensity is strong evidence for the null hypothesis. Either squaring or thresholding negative source intensity estimates at zero can reduce detection performance by reducing the separability of the test statistic distributions under the two hypotheses, especially for weak sources. Our numerical results do not include the Wald test since it is asymptotically equivalent to the GLRT.

Source Intensity Test (SIT)

Because the squaring operation in the Wald test statistic can degrade detection performance, we consider the following source intensity test statistic

$$\Lambda_{\mathrm{SIT}} \stackrel{\triangle}{=} \hat{\alpha}_{\mathrm{ML}}.$$

This test was also considered in the context of array processing [6]. By the asymptotic normality of the MLE [35, p. 240], we have that asymptotically,

$$\hat{\alpha}_{\mathrm{ML}} \sim \begin{cases} \mathcal{N}\left(0, \mathsf{F}^{-1}\left(\boldsymbol{\theta}_0\right)_{[1,1]}\right), & H_0 \\ \mathcal{N}\left(\alpha, \mathsf{F}^{-1}\left(\boldsymbol{\theta}\right)_{[1,1]}\right), & H_1, \end{cases} \tag{2.43}$$

where $\boldsymbol{\theta}_0$ is the parameter vector under the null hypothesis with $\alpha = 0$. The next section uses the asymptotic distributions in (2.40) and (2.43) to show that position–sensitive capability improves detection performance.

Mean Difference Test (MDT)

The mean difference test statistic is proposed in [51] and is equivalent to

$$\Lambda_{\mathrm{MDT}} \stackrel{\triangle}{=} J_{\mathrm{src}} - J_{\mathrm{back}}, \tag{2.44}$$

where J_{src} is the number of photons recorded on all source–exposed surfaces of the detector and J_{back} is the number of photons recorded on all surfaces not exposed to the source. When the source position is unknown, one must estimate which surfaces are exposed to the source. The asymptotics of the MDT are given in [51].

CHAPTER III

Image Reconstruction for Position–Sensitive Gamma–Ray Detectors [1]

Image reconstruction is one means of extracting meaningful information from gamma–ray measurements. The goal of image reconstruction is to estimate the gamma–ray emission rate as a function of space. The estimated emission density functions could be given to a detection algorithm or a human to detect or locate a radioactive source. In this chapter, we do not consider the detection or localization problem, but instead focus on image quality in terms of sparsity in the pixel basis.

Maximum likelihood (ML) estimation has been previously applied to the gamma–ray image reconstruction problem [80]. The ML problem is typically ill–conditioned when the number of recorded measurements is small, resulting in a nonunique solution. In this chapter, we propose a penalized likelihood algorithm that uses regularization to enforce the expectation that the image should be sparse in the spatial domain. For an example of penalized likelihood applied to the emission tomography problem with different prior knowledge, see [25]. We also examine other methods of regularization that work well with Gaussian data [11], but, as we will show, do not work well in the gamma–ray imaging case due to the Poisson nature of the measurements.

[1] The work in this chapter was previously published in [46]. We also investigated the use of an augmented Lagrangian method for emission tomography with a strict ℓ_1 finite–differencing prior [45].

We propose a regularization function based on the sums of the logarithms of the pixel values. This function was previously discussed in [10] as a measure of sparsity. To our knowledge, this function was not previously used as a regularizer for gamma–ray image reconstruction. We find that this regularizer results in thresholding behavior similar to the ℓ_1–regularized Gaussian penalized likelihood problem [18]. This thresholding behavior leads to sparse reconstructed images.

In this chapter, we develop a model for the likelihood in the image reconstruction problem and derive optimization transfer algorithms [40] for solving the unregularized and regularized problems. We also examine the behavior of the regularized and unregularized algorithms in the one–dimensional case to provide intuition about their behavior in the high dimensional coupled image reconstruction problem. We then demonstrate the maximum likelihood and proposed penalized likelihood algorithm on real data recorded with a 3D position–sensitive CdZnTe gamma–ray detector.

3.1 Statistical Model

We consider gamma–ray detectors operating in list–mode as discussed in §2.3. In this section, we will show that the list–mode model is equivalent to the more intuitive but memory–intensive binned–mode model. Let $\boldsymbol{x} = [x_1, x_2, \ldots, x_{n_p}]^T$ be the image of radiation emitters around the sphere discretized with n_p pixels. For generality, we assume that $\boldsymbol{x}$ is the emission density of interest and model the background with $\bar{\boldsymbol{r}} = [\bar{r}_1, \bar{r}_2, \ldots, \bar{r}_{n_d}]^T$, where $\bar{r}_i$ is the mean number of background photons matching the the ith possible measurement and n_d is the number of possible recorded measurements, or detector bins. In 4π Compton imaging, we model $\bar{r}_i$ as zero for all possible recorded events. We make this modeling choice because we wish to observe all radiation emitters, even those that may be considered background.

Let S be the event that a photon originated from the source $\boldsymbol{x}$, let B be the event that a photon originated from the background $\bar{\boldsymbol{r}}$, let E_j be the event that a photon originated from the jth image pixel, and let D be the event that a photon is recorded. Let $\boldsymbol{A}$ be a system matrix, where

$$a_{ij} \stackrel{\triangle}{=} \boldsymbol{A}_{[i,j]} = \mathsf{p}\left(\boldsymbol{r}_i, D | E_j, S\right). \tag{3.1}$$

Let s_j be the sensitivity of the detector to position bin j, which is the probability that a photon emitted from position bin j is recorded. This quantity is commonly found by simulation in the list-mode case [64] because it can be impractical to compute all rows in $\boldsymbol{A}$.

Let $\tilde{\boldsymbol{r}}$ the the list of recorded attributes for the interacting photons. The characteristics of $\tilde{\boldsymbol{r}}$ are described in detail in §2.3.1. The list–mode likelihood of the list of event attributes $\tilde{\boldsymbol{r}}$ is given by [4]

$$\mathsf{p}\left(\tilde{\boldsymbol{r}}; \boldsymbol{x}\right) = e^{\bar{J}(\boldsymbol{x})} \bar{J}(\boldsymbol{x})^J / J! \prod_{i=1}^{J} \mathsf{p}\left(\boldsymbol{r}_i | D; \boldsymbol{x}\right), \tag{3.2}$$

where $\bar{J}(\boldsymbol{x})$ is the mean number of recorded photons per scan, and

$$\bar{J}(\boldsymbol{x}) = \sum_{j=1}^{n_p} s_j x_j + \sum_{i=1}^{n_d} \bar{r}_i. \tag{3.3}$$

We show in Appendix A that the model in (3.2) is equivalent to the more intuitive binned–mode model:

$$y_i \sim \text{Poisson}([\boldsymbol{A}\boldsymbol{x}]_i + \bar{r}_i), \quad i = 1 \ldots n_d, \tag{3.4}$$

where y_i is the number of recorded photons in the ith discretized detector bin. The equivalence holds when $y_i = 1$ for $i = 1, 2, \ldots, J$ and zero otherwise. The log–likelihood, derived in Appendix A, is given by

$$\mathsf{L}\left(\boldsymbol{x}\right) \equiv \sum_{i=1}^{n_d} \left[- \sum_{j=1}^{n_p} a_{ij} x_j - \bar{r}_i + y_i \log \left(\sum_{j=1}^{n_p} a_{ij} x_j + \bar{r}_i \right) \right], \tag{3.5}$$

where $\equiv$ denotes equivalence because terms constant with respect to $\boldsymbol{x}$ are omitted.

3.2 Sparsity and Regularization Design

There are numerous choices of penalty functions for spatially–sparse reconstruction. We choose the ℓ_1 and ℓ_0 norms because of their popularity in Gaussian sparse reconstruction. Disappointed by their ineffectiveness with Poisson data, we investigated a third penalty function $R_l(\boldsymbol{x}) = \sum_{j=1}^{n_p} \log\left(\frac{x_j}{\delta} + 1\right)$. We found this penalty to be more effective at reconstructing sparse, meaningful images of true radiation source distributions. In summary, the three penalty functions that we consider in the following sections are:

$$\begin{aligned} R_1(\boldsymbol{x}) &= ||\boldsymbol{x}||_1, \\ R_0(\boldsymbol{x}) &= ||\boldsymbol{x}||_0, \\ R_l(\boldsymbol{x}) &= \sum_{j=1}^{n_p} \log\left(\frac{x_j}{\delta} + 1\right). \end{aligned}$$

3.3 Unregularized MLEM Algorithm

The well-known maximum likelihood expectation maximization (MLEM) algorithm iteratively finds the maximum-likelihood image based on the observed measurements [81], or equivalently, finds the solution to the optimization problem in (2.27), where the log-likelihood is given by (3.5). We follow [15] to show that the EM algorithm for Poisson data described in [26] is equivalent to an optimization transfer approach with the following surrogate function:

$$Q^{(n)}\left(\boldsymbol{x}; \boldsymbol{x}^{(n)}\right) = \sum_{j=1}^{n_p} Q_j^n\left(x_j; \boldsymbol{x}^{(n)}\right), \tag{3.6}$$

where

$$Q_j^n\left(x_j; \boldsymbol{x}^{(n)}\right) \triangleq \sum_{i=1}^{n_d} \left(\frac{a_{ij}\left(x_j^{(n)} + \gamma_j\right)}{\bar{y}_i^{(n)}}\right) \left[\frac{x_j + \gamma_j}{x_j^{(n)} + \gamma_j} \bar{y}_i^{(n)} - y_i \log\left(\frac{x_j + \gamma_j}{x_j^{(n)} + \gamma_j} \bar{y}_i^{(n)}\right)\right], \tag{3.7}$$

and γ_j, for $j = 1 \ldots n_p$, are nonnegative constants satisfying

$$\sum_{j=1}^{n_p} a_{ij}\gamma_j \leq \bar{r}_i \qquad i = 1, \ldots, n_d. \tag{3.8}$$

Typically, $\gamma_j = 0$, but using $\gamma_j > 0$ can accelerate convergence [26]. This MLEM algorithm has the following "expectation step" or E-step

$$e_j\left(\boldsymbol{x}\right) = \sum_{i=1}^{n_d} a_{ij} \frac{y_i}{\bar{y}_i\left(\boldsymbol{x}\right)}, \tag{3.9}$$

and a "maximization" step or M-step

$$x_j^{(n+1)} = \left[\frac{x_j^{(n)} + \gamma_j}{a_j} e_j\left(\boldsymbol{x}^{(n)}\right) - \gamma_j\right]_+ . \tag{3.10}$$

The operator $[\cdot]_+$, used in (3.10) is defined by

$$[u]_+ = \max(u, 0). \tag{3.11}$$

In the case where $\boldsymbol{A} = \operatorname{diag}\{a_i\}$, the cost function is separable and the problem becomes one of denoising rather than reconstruction. In this case, the ML estimate for the source intensity x_i from the measurement y_i is

$$\hat{x}_i = \frac{1}{a_i}\left[y_i - \bar{r}_i\right]_+ . \tag{3.12}$$

This ML estimator will be the basis of comparison with the regularized methods that follow.

3.3.1 Convergence

Reference [33] provides a set of sufficient conditions for convergence of an optimization transfer algorithm to a stationary point of the cost function. One can verify that the algorithm described by (3.9) and (3.10) satisfies conditions of feasibility of the algorithm (R1), agreement and continuity of first derivatives (R2), proper domain of surrogates (R3), connected surrogates (C2), and continuity of surrogate

generators in both arguments (C3) in [33], so by Theorem 4.1 of [33], a subsequence of the sequence of iterates $\{\boldsymbol{x}^{(n)}\}$ given by (3.9) and (3.10) converges to a stationary point of $-\mathsf{L}(\boldsymbol{x})$. For the case where $\gamma_j = 0$ for all $j = 1, 2, \ldots, n_p$, global convergence is shown in [33].

3.4 Penalized-Likelihood Algorithm for $R(\boldsymbol{x}) = ||\boldsymbol{x}||_1$

The ℓ_1 norm has been used previously as a measure of sparsity [10]. It is a convenient penalty function because it is convex and differentiable almost everywhere. This section shows that the influence of $||\cdot||_1$ for Poisson noise is quite different from the Gaussian case. The ℓ_1 regularized estimator solves the following optimization problem:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \Psi_1(\boldsymbol{x}), \;\; \Psi_1(\boldsymbol{x}) = -\mathsf{L}(\boldsymbol{x}) + \beta \left|\left|\boldsymbol{x}\right|\right|_1 . \tag{3.13}$$

We approach this problem using the optimization transfer approach [40], where one finds a separable surrogate function that satisfies (2.29). Extending (3.7), a valid surrogate for the cost function (3.13) is

$$Q_j^n\left(x_j; \boldsymbol{x}^{(n)}\right) = (x_j + \gamma_j)\, a_j - e_j\left(x_j^{(n)}\right)\left(x_j^{(n)} + \gamma_j\right) \log\left(x_j + \gamma_j\right) + \beta x_j, \tag{3.14}$$

which, minding the nonnegativity constraint on x_j, has the minimizer

$$x_j^{(n+1)} = \left[\frac{x_j^{(n)} + \gamma_j}{a_j + \beta} e_j\left(\boldsymbol{x}^{(n)}\right) - \gamma_j\right]_+ , \tag{3.15}$$

where $e_j(\boldsymbol{x})$ was defined by the E-step in (3.9). Note that (3.15) and (3.9) represent the M-step and E-step, respectively, of this penalized-likelihood algorithm. Also, (3.15) is identical to (3.10) except for a scaling by $a_j + \beta$ instead of a_j. This causes scaling of the iterates, but not thresholding behavior that is desirable in sparse reconstruction. We can gain insight by examining the behavior of ℓ_1 regularization

when $\boldsymbol{A} = \text{diag}\{a_i\}$, where the estimate reduces to

$$\hat{x}_i = \frac{[y_i - \bar{r}_i(1 + \beta/a_i)]_+}{a_i + \beta}. \tag{3.16}$$

When $\bar{r}_i = 0$, the estimate is just a scaled version of the observation, which is just shrinkage compared to the unregularized estimator in (3.12). Also, when $\bar{r}_i > 0$, the estimator introduces thresholding where the threshold becomes larger as β increases. However, this thresholding does not occur when one does not have knowledge of the background and $\bar{r}_i = 0$. This behavior differs from the case of Gaussian statistics where

$$\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{x} + \tilde{\boldsymbol{r}}, I). \tag{3.17}$$

The solution to the denoising problem in this case is

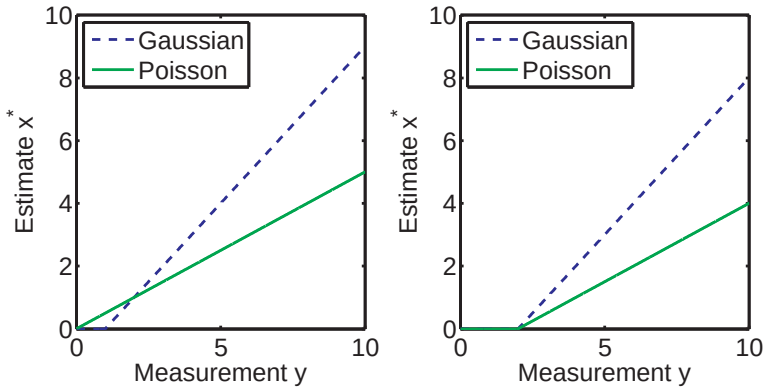
$$\hat{x}_i = \begin{cases} \frac{1}{a_i}\left(y_i - \bar{r}_i - \frac{\beta}{a_i}\right) & y_i > \bar{r}_i + \frac{\beta}{a_i} \\ 0 & \bar{r}_i - \frac{\beta}{a_i} \leq y_i \leq \bar{r}_i + \frac{\beta}{a_i} \\ \frac{1}{a_i}\left(y_i - \bar{r}_i + \frac{\beta}{a_i}\right) & y_i < \bar{r}_i - \frac{\beta}{a_i}, \end{cases} \tag{3.18}$$

which is a soft thresholding estimator. Figure 3.1 shows the estimators for the Poisson and Gaussian data models with various values of β and $\bar{r}_i$. Notice that when $\bar{r}_i = 0$, the estimator using the Poisson model does not perform thresholding. Also notice that when $\bar{r}_i \neq 0$, the estimator using the Poisson model scales the estimate by a factor less than one.

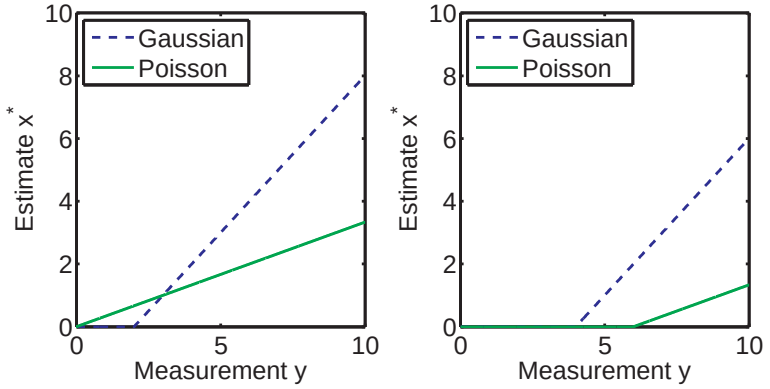
3.4.1 Non-Uniqueness

The following counterexample shows that the ℓ_1 regularized solution is not always unique for Poisson data. Let

$$A = \begin{bmatrix} 1 & 1 \end{bmatrix}, \qquad \boldsymbol{x} = \begin{bmatrix} x_1 \\ \\ x_2 \end{bmatrix}, \qquad \bar{r}_1 = 0.$$



(a) $\beta = 1$ and $r_i = 0$ (b) $\beta = 1$ and $r_i = 1$



(c) $\beta = 2$ and $r_i = 0$ (d) $\beta = 2$ and $r_i = 2$

Figure 3.1: Estimates $\hat{x}$ vs measurements y for denoising case with Poisson and Gaussian data models.

In this case, $y_1 \sim \text{Poisson}(x_1 + x_2)$. One can show that any pair of nonnegative values $\hat{x}_1$, $\hat{x}_2$ that satisfy

$$\hat{x}_1 + \hat{x}_2 = \frac{y_1}{1+\beta}$$

is a minimizer of the penalized-likelihood cost function Ψ_1 in (3.13). This non-uniqueness of the ℓ_1-regularized estimate is consistent with the Gaussian case.

3.4.2 Convergence

One can verify that the algorithm described by (3.9) and (3.15) satisfies conditions of feasibility of the algorithm (R1), agreement and continuity of first derivatives (R2), proper domain of surrogates (R3), connected surrogates (C2), and continuity

of surrogate generators in both arguments (C3) in [33], so by Theorem 4.1 of [33], a subsequence of the sequence of iterates $\{\boldsymbol{x}^{(n)}\}$ given by (3.9) and (3.10) converges to a stationary point of $\Psi_1(\boldsymbol{x})$.

3.5 ℓ_0 regularization for Poisson Data

The ℓ_0 "norm" measures sparsity in its purest sense by counting the number of nonzero elements. Unlike the ℓ_1 norm, it is nonconvex, not continuous, and uniform almost everywhere except zero. This section shows that ℓ_0 regularization can behave quite differently for Poisson data compared to the Gaussian case. The ℓ_0 regularized estimate is the solution to the following optimization problem:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \Psi_0(\boldsymbol{x}),\ \ \Psi_0(\boldsymbol{x}) = -\mathsf{L}\left(\boldsymbol{x}\right) + \beta \left|\left|\boldsymbol{x}\right|\right|_0, \tag{3.19}$$

where the log-likelihood term $\mathsf{L}\left(\boldsymbol{x}\right)$ is defined in (3.5). The solution to this problem is obtained by minimizing the non–convex cost function $\Psi(\boldsymbol{x}) = -\mathsf{L}\left(\boldsymbol{x}\right) + \beta \left|\left|\boldsymbol{x}\right|\right|_0$. Extending (3.7), we obtain the surrogate

$$Q_j^n\left(x_j; \boldsymbol{x}^{(n)}\right) \stackrel{\triangle}{=} \left(x_j + \gamma_j\right) a_j - e_j\left(x_j^{(n)}\right)\left(x_j^{(n)} + \gamma_j\right) \log\left(x_j + \gamma_j\right) + \beta I_{(0,\infty)}(x_j), \tag{3.20}$$

where $I_{(0,\infty)}(x_j)$ is the indicator function on the interval $(0, \infty)$. The M-step is given by

$$x_j^{(n+1)} = \arg\min_{x_j \geq 0} Q_j^n\left(x_j; \boldsymbol{x}^{(n)}\right). \tag{3.21}$$

If the minimizer $\tilde{x}_j^{(n+1)}$ is positive, it is found by setting the derivative equal to zero:

$$\frac{\partial}{\partial x_j} Q_j^n\left(x_j; \boldsymbol{x}^{(n)}\right) = e_j\left(\boldsymbol{x}^{(n)}\right) \frac{x_j^{(n)} + \gamma_j}{x_j + \gamma_j} - a_j = 0.$$

Solving and minding the non–negativity constraint, we obtain a candidate solution:

$$\tilde{x}_j^{(n+1)} = \left[\frac{x_j^{(n)} + \gamma_j}{a_j} e_j\left(\boldsymbol{x}^{(n)}\right) - \gamma_j\right]_+. \tag{3.22}$$

However, if $\gamma_j > 0$, the minimizer could either be at the location $\tilde{x}_j^{(n+1)}$ found on the interval $(0, \infty)$, or it could be at $x_j = 0$ because of the non–convexity of the surrogate. If $\tilde{x}_j^{(n+1)} > 0$, one should choose the minimizer $\hat{x}_j^{(n+1)} = 0$ if

$$Q_j^{(n)}\left(0; \boldsymbol{x}^{(n)}\right) < Q_j^{(n)}\left(\tilde{x}_j^{(n+1)}; \boldsymbol{x}^{(n)}\right)$$

$$e_j\left(\boldsymbol{x}^{(n)}\right)\left(x_j^{(n)} + \gamma_j\right)\log\left(\frac{\tilde{x}_j^{(n+1)} + \gamma_j}{\gamma_j}\right) - a_j\tilde{x}_j^{(n+1)} < \beta, \tag{3.23}$$

otherwise if $\tilde{x}_j^{(n+1)} = 0$, $\hat{x}_j^{(n+1)} = 0$. Combining (3.22) and (3.23), we obtain an EM algorithm with E-step given by (3.9) and M-step:

$$x_j^{(n+1)} = \begin{cases} \tilde{x}_j^{(n+1)} & \tilde{x}_j^{(n+1)} > 0 \ \text{ AND} \\ & e_j\left(\boldsymbol{x}^{(n)}\right)\left(x_j^{(n)} + \gamma_j\right)\log\left(\frac{\tilde{x}_j^{(n+1)} + \gamma_j}{\gamma_j}\right) - a_j\tilde{x}_j^{(n+1)} > \beta \\ 0 & \text{else.} \end{cases} \tag{3.24}$$

The behavior of the algorithm is different depending on the values chosen for $\bar{r}_i$ and γ_j. We treat the cases separately.

3.5.1 Convergence when $\bar{r}_i = 0$ for $i = 1, \ldots, n_d$

By (3.8), if $\bar{r}_i = 0$ for $i = 1, \ldots, n_d$, γ_j must be zero for all $j = 1 \ldots n_p$. When this is true, the algorithm reduces to the MLEM algorithm. The behavior of the algorithm depends on these tuning parameters because they shift the cost function so that the discontinuity in the ℓ_0 "norm" and the singularities of the log-likelihood function no longer overlap at $x_j = 0$ for $j = 1, \ldots, n_p$. The fact that the estimate is influenced by algorithm tuning parameters is troublesome, but the nonconvexity of the cost function allows the tuning parameters to influence the local minimum to which the algorithm converges. Since the algorithm is equivalent to MLEM in this case, the convergence result of §3.3.1 applies.

3.5.2 Convergence when $\bar{r}_i > 0$ for some i and $\gamma_j = 0$ for $j = 1, \ldots, n_p$

In this case, the nonzero $\bar{r}_i$ affects the $e_j\left(\boldsymbol{x}^{(n)}\right)$ term, but we always have $x_j^{(n)} = \tilde{x}_j^{(n+1)}$ because $\gamma_j = 0$. Thus the algorithm is equivalent to MLEM in this case as well.

3.5.3 Convergence $\bar{r}_i > 0$ for some i and $\gamma_j > 0$ for some j

In this case, the surrogates are not continuous at $x_j = 0$, so the regularity condition (R3) in [33] fails to hold, thus the algorithm is not guaranteed to converge to a stationary point of $\Psi_0(\boldsymbol{x})$.

3.5.4 ℓ_0 Regularization when $\boldsymbol{A}$ is diagonal

If $\boldsymbol{A} = \operatorname{diag}\{a_i\}$, the ℓ_0 regularized estimate is

$$\hat{x}_j = \begin{cases} \frac{1}{a_i}[y_i - \bar{r}_i]_+ & \kappa\left(y_i, \bar{r}_i\right) > \beta \\ 0 & \text{else,} \end{cases} \tag{3.25}$$

where $\kappa(u, v)$ denotes the Kullback-Leibler divergence [9] given by

$$\kappa(u, v) = \begin{cases} v & u = 0, v \geq 0 \\ u \log\left(\frac{u}{v}\right) - u + v & u > 0, v > 0 \\ \infty & u > 0, v = 0. \end{cases}$$

When $\bar{r}_i = 0$, (3.25) simplifies to (3.12) because $\kappa(y_i, 0) = \infty$ for all y_i. Again, in the case of ℓ_0 regularization, when $\bar{r}_i = 0$, there is no thresholding behavior. The ineffectiveness of ℓ_0 regularization for Poisson noise is different from that for Gaussian noise modeled by (3.17), where the denoising estimator is

$$\hat{x}_i = \begin{cases} \frac{y_i - \bar{r}_i}{a_i} & |y_i - \bar{r}_i| > \sqrt{2\beta} \\ 0 & \text{else,} \end{cases} \tag{3.26}$$

which is hard thresholding. The ineffectiveness of Poisson regularization in the ℓ_0 case is due to the singularity in the Poisson log-likelihood $-L(x)$ at $x = 0$. Figure 3.2 shows the Gaussian and Poisson likelihoods for a denoising problem where the single observation $y = 3$. Note that the Gaussian likelihood is well-behaved near $x = 0$, but the Poisson likelihood approaches infinity. This singularity causes the cost function to be arbitrarily large near $\boldsymbol{x} = 0$, regardless of the regularizer chosen.

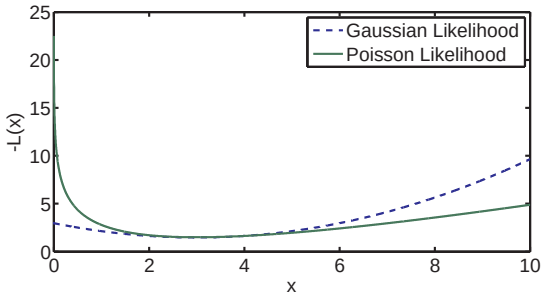


Figure 3.2: Gaussian and Poisson log-likelihoods vs. x for denoising case with measurement $y = 3$

3.6 log Regularization

We propose the use of the penalty function $R(\boldsymbol{x}) = \sum_{j=1}^{n_p} \log\left(\frac{x_j}{\delta} + 1\right)$ as an alternative to the ℓ_1 and ℓ_0 norms [10]. It is nonconvex and thus the minimizer is not guaranteed to be unique. However, it is a continuous function that is differentiable almost everywhere. It is also a better measure of sparsity than the ℓ_1 norm in the sense that it better approximates the true measure of sparsity, the ℓ_0 "norm." The minimization problem is:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}} \Psi_l(\boldsymbol{x}), \Psi_l(\boldsymbol{x}) = -\mathsf{L}\left(\boldsymbol{x}\right) + \beta \sum_{j=1}^{n_p} \log\left(\frac{x_j}{\delta} + 1\right). \tag{3.27}$$

We first examine the case where $\boldsymbol{A}$ is diagonal, i.e., $\boldsymbol{A} = \mathrm{diag}(\boldsymbol{a})$, for which the cost function is separable. This denoising case provides insight into the behavior of the

regularizer for a coupled cost function. The minimizer of the separable cost function is given by

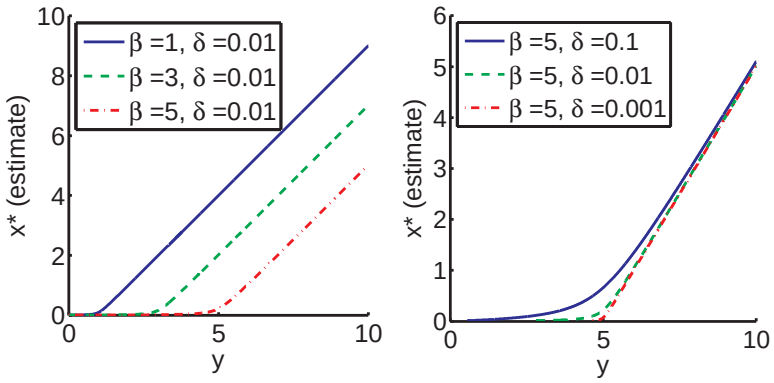
(3.28)

$$\hat{x}_i = \arg\min_{x_i \geq 0} \Psi_i\left(x_i\right), \;\; \Psi_i\left(x_i\right) = \sum_{i=1}^{n} a_i x_i + \bar{r}_i - y_i \log\left(a_i x_i + \bar{r}_i\right) + \beta \log\left(\frac{x_i}{\delta} + 1\right),$$

which has nonnegative minimizer

$$\begin{aligned} \hat{x}_i(y_i) = \frac{1}{2}\Bigg(& -\delta - \frac{1}{a_i}(\bar{r}_i + \beta - y_i) \\ & + \frac{1}{a_i}\sqrt{\delta^2 a_i^2 + 2\delta a_i(\bar{r}_i + \beta - y_i) + (\bar{r}_i + \beta - y_i)^2 + 4a_i\delta(y_i - \bar{r}_i) - 4\beta\bar{r}_i}\Bigg). \end{aligned} \tag{3.29}$$

Since it is difficult to visualize the behavior of the estimate with respect to δ and β, Figure 3.3 illustrates the estimator for various values of β and δ with $\bar{r}_i = 0$ and $a_i = 1$ for all i.



(a) $\beta = 1, 3, 5$, and $\delta = 0.01$ (b) $\beta = 5$, and $\delta = 0.001, 0.01, 0.1$

Figure 3.3: 1D penalized likelihood estimates using log regularization versus observations

3.6.1 Choosing δ

As $\delta \to 0$, the estimator $\hat{x}_i(y_i)$ in (3.6) converges pointwise to the following function (minding the non-negativity constraint):

$$\hat{x}_i(y_i)^{ideal} = \lim_{\delta \to 0} \hat{x}_i(y_i)$$

$$= \begin{cases} 0 & (\bar{r}_i + \beta - y_i)^2 - 4\beta\bar{r}_i < 0 \\ \frac{1}{2a_i}\left(y_i - \beta - \bar{r}_i + \sqrt{(\bar{r}_i - y_i + \beta)^2 - 4\beta\bar{r}_i}\right) & (\bar{r}_i + \beta - y_i)^2 - 4\beta\bar{r}_i \geq 0, \end{cases} \tag{3.30}$$

which is an approximation to hard thresholding. Figure 3.4 shows a graph of this function for a set of particular parameters. When $\bar{r}_i = 0$, which is of interest in 4π imaging, (3.30) simplifies to:

$$\hat{x}_i(y_i)^{ideal} = \frac{1}{2a_i}\left(y_i - \beta + |y_i - \beta|\right) = \frac{1}{a_i}\left[y_i - \beta\right]_+ . \tag{3.31}$$

Thus, when $\bar{r}_i = 0$, smaller δ results in better approximation of soft thresholding. However, the sharp curvature of $\hat{x}_i(y_i)$ when δ is small can slow convergence of the EM algorithm in the case where $\boldsymbol{A}$ is not diagonal. In our experiments, we found that $10 \leq \delta \leq 100$ worked well for the Compton imaging problem.

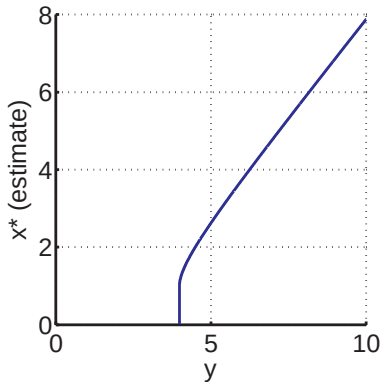


Figure 3.4: $\hat{x}_i(y_i)^{ideal}$ vs y_i for $\beta = 1, a_i = 1, \bar{r}_i = 1$

3.6.2 Choosing β

If $\bar{r}_i = 0$, the estimator approximates soft thresholding as shown in (3.31) and when $\bar{r}_i > 0$, the estimator approximates hard thresholding as shown in Figure 3.4, where the threshold is a function of β. In both cases, increasing β increases the threshold. The choice of β involves a trade-off between sensitivity to noise and failure to image low-intensity point sources.

3.6.3 EM Algorithm for log Regularization

To develop an EM algorithm for the cost function (3.27), we use the following separable surrogate functions to iteratively find the minimizer:

(3.32)

$$Q_j^n\left(x_j; \boldsymbol{x}^{(n)}\right) = \left(x_j + \gamma_j\right) a_j - e_j\left(x_j^{(n)}\right)\left(x_j^{(n)} + \gamma_j\right) \log\left(x_j + \gamma_j\right) + \beta \log\left(\frac{x_j}{\delta} + 1\right).$$

The E-step is given by $e_j\left(\boldsymbol{x}\right)$ in (3.9) and the M-step has the same form as (3.21). Differentiating (3.32) and zeroing yields the following quadratic equation that one can solve to compute $\boldsymbol{x}_j^{(n+1)}$:

$$\begin{aligned} 0 = \frac{a_j}{\delta} x_j^2 + \left(a_j + \delta^{-1}\left(\gamma_j a_j - e_j\left(\boldsymbol{x}^{(n)}\right)\left(x_j^{(n)} + \gamma_j\right) + \beta\right)\right) x_j \\ + a_j \gamma_j - e_j\left(\boldsymbol{x}^{(n)}\right)\left(x_j^{(n)} + \gamma_j\right) + \frac{\beta}{\delta}\gamma_j. \end{aligned} \tag{3.33}$$

The behavior of this algorithm is difficult to visualize in the reconstruction problem when $\boldsymbol{A}$ is not diagonal, but the separable analysis of the denoising problem gives some intuition. The algorithm will perform an approximation of soft thresholding when $\bar{r}_i = 0$ for $i = 1, \ldots, n_d$, which will lead to sparser reconstructed images than the ML estimate.

3.6.4 Convergence of EM Algorithm for log Regularization

One can verify that the algorithm described by (3.9) and the surrogate defined by (3.33) satisfy conditions of feasibility of the algorithm (R1), agreement and continuity of first derivatives (R2), proper domain of surrogates (R3), connected surrogates (C2), and continuity of surrogate generators in both arguments (C3) in [33], so by Theorem 4.1 of [33], a subsequence of the sequence of iterates $\{\boldsymbol{x}^{(n)}\}$ given by (3.9) and (3.10) converges to a stationary point of $\Psi_l(\boldsymbol{x})$. There may be multiple stationary points of $\Psi_l(\boldsymbol{x})$, so the value to which the estimates converge may not the the global minimizer.

3.7 Image Reconstruction Results

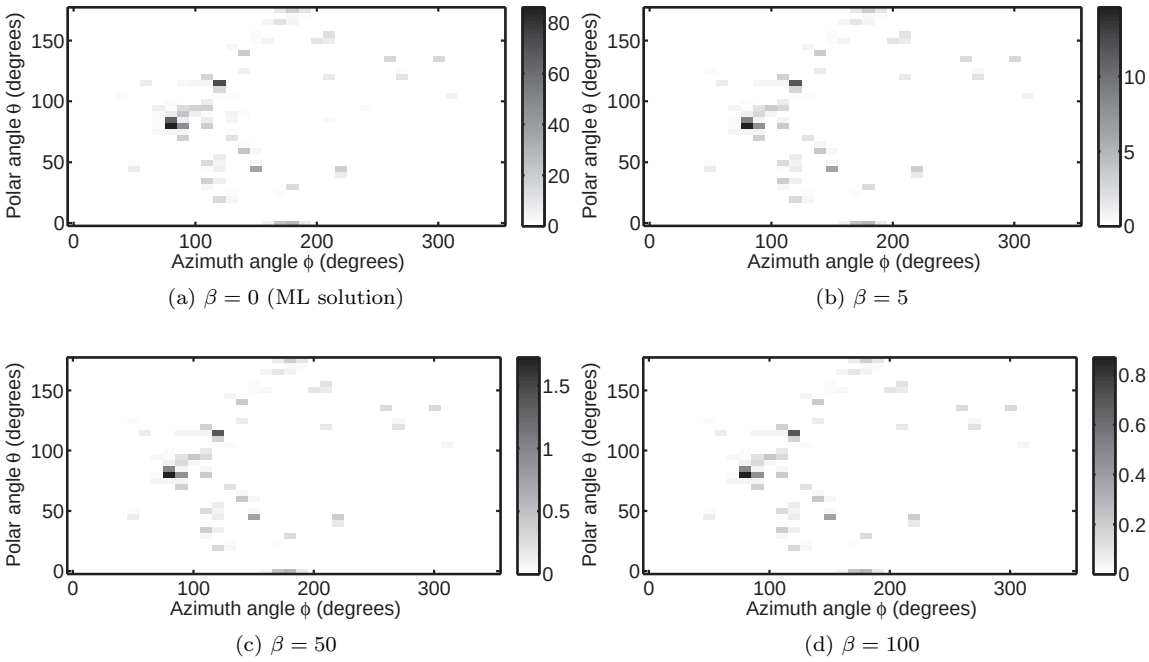
This section gives experimental results of the algorithms presented in this chapter for a list-mode position–sensitive imaging system. The scenes monitored by this system do not have a known background emission distribution, so we set the expected background contribution to each measurement, $\bar{r}_i$, to be zero. These images were produced from real laboratory measurements of a Cesium-137 source and natural background with a position-sensitive CdZnTe detector. The system model assumes that each photon deposits all of its energy in the detector. We used only 1000 recorded photons with two or more interactions to show how regularization can improve the clarity of the image when there are few recorded counts.

3.7.1 ℓ_0 Regularization

As mentioned in Sec. 3.5, if $\bar{r}_i = 0$ for all i, then $\gamma_j = 0$ for all j. We can see from (3.24) that the algorithm in this case simplifies to unregularized MLEM. The unregularized ML estimates are shown in the upper left of Figures 3.5 and 3.6.

3.7.2 ℓ_1 Regularization

Figure 3.5 shows reconstructed images using the ℓ_1–regularized algorithm. Note that the image corresponding to $\beta = 0$ is the ML estimate. Also note that the images appear identical to the eye except the pixel values are scaled by a constant. This is apparent from the color scales located to the right of each image. We investigated numerous other values of β and observed only pixel scaling. This behavior is consistent with the analysis for the separable system in (3.16).



(a) $\beta = 0$ (ML solution) (b) $\beta = 5$

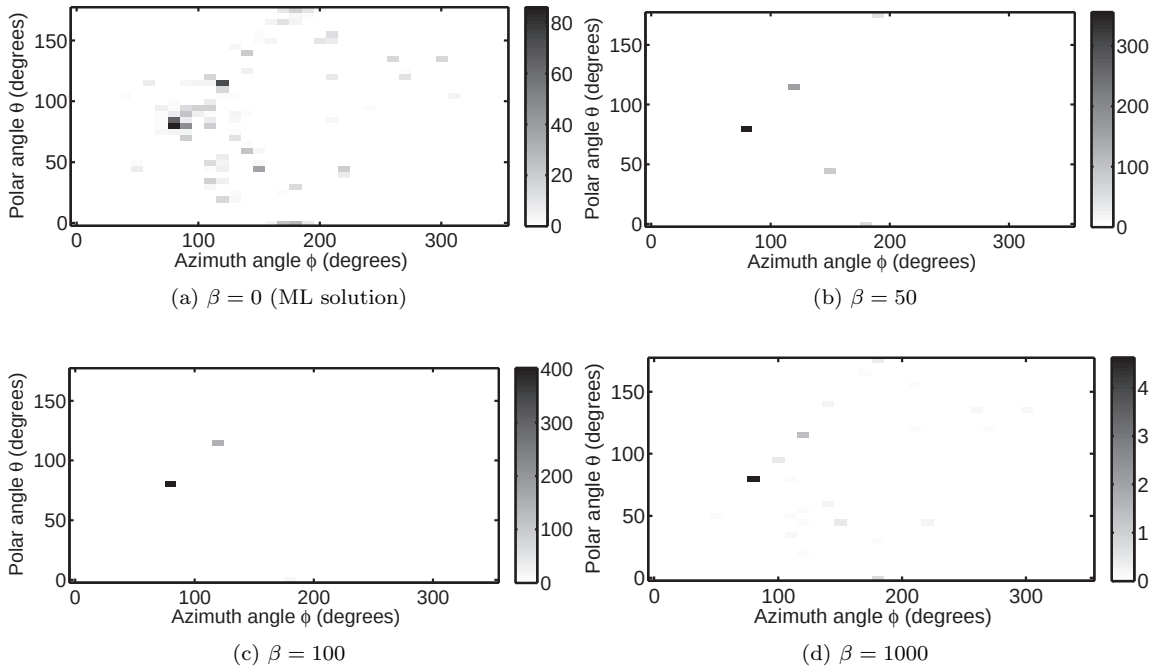
(c) $\beta = 50$ (d) $\beta = 100$

Figure 3.5: Reconstructed images using ℓ_1 regularization with various values of β. These images were reconstructed from 1000 recorded photons in a real CdZnTe detector with a Cs-137 source located at azimuth angle $\phi = 90^o$ and polar angle $\theta = 90^o$. The algorithm was run for 200 iterations.

3.7.3 log Regularization

Figure 3.6 shows four reconstructed images using the log regularization estimate for different values of β, where there is a Cs-137 source located at azimuth angle $\phi = 90^o$ and polar angle $\theta = 90^o$. We use a simplified system model where each recorded

event deposits all of its energy in the detector. This results in model mismatch because some interacting photons escape the detector after their final scattering interaction, but it significantly reduces the computational burden. Following the reasoning of Sec. 3.6.1, these images were reconstructed with the parameter $\delta = 10$. However, the choice of δ could impact the number of iterations needed to achieve convergence. The images with $\beta = 50$ in Figure 3.6b and $\beta = 100$ in Figure 3.6c



(a) $\beta = 0$ (ML solution)

(b) $\beta = 50$

(c) $\beta = 100$

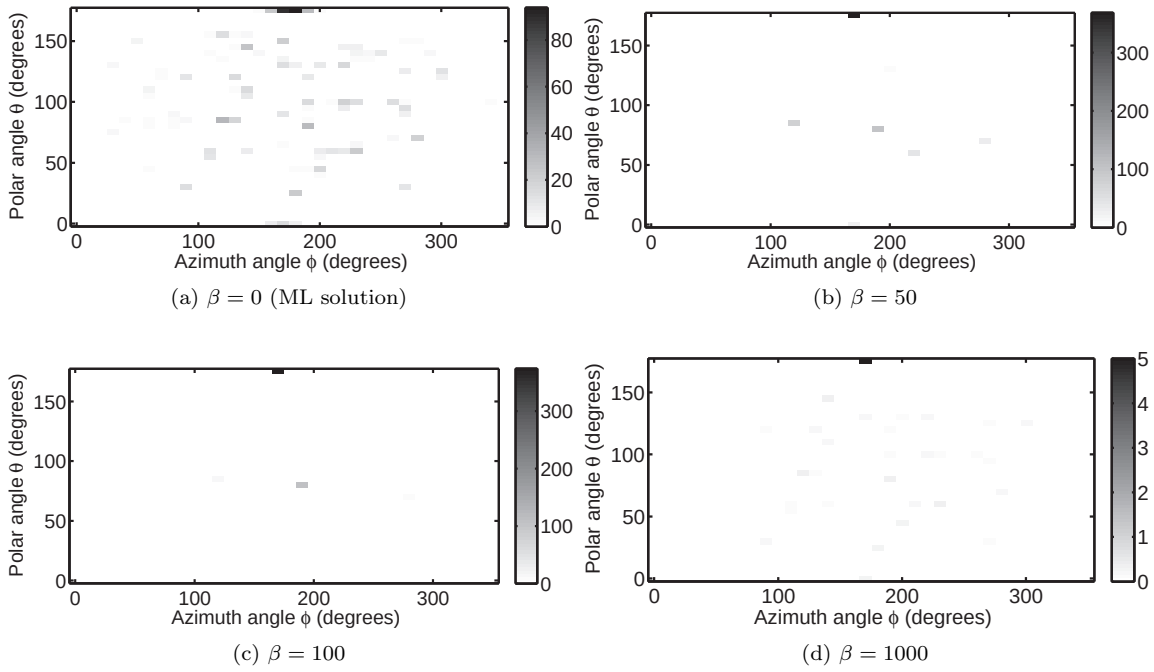
(d) $\beta = 1000$

Figure 3.6: Reconstructed images using log regularization with various values of β. These images were reconstructed from 1000 recorded photons in a real CdZnTe detector with a Cs-137 source located at azimuth angle $\phi = 90^{\circ}$ and polar angle $\theta = 90^{\circ}$. The algorithm was run for 200 iterations and the parameter $\delta = 10$.

are noticeably more sparse than the ML solution in Figure 3.6a. However, the image with $\beta = 1000$ has only small intensity values because the parameter β is too large. This shows that the log regularizer can result in images that are sparser than the ML solution with the correct parameter choice.

Figure 3.7 shows four reconstructed images for various values of the parameter β

where there is no source present and all measurements are due to natural background. The ML solution in Figure 3.7a has an intense region near the pole at the top of the



(a) $\beta = 0$ (ML solution)

(b) $\beta = 50$

(c) $\beta = 100$

(d) $\beta = 1000$

Figure 3.7: Reconstructed images using log regularization with various values of β. These images were reconstructed from 1000 recorded photons in a real CdZnTe detector in a room with concrete walls and no source present. The algorithm was run for 200 iterations and the parameter $\delta = 10$.

image. This is likely due to the inability of the detector to record photon interactions near the anode and model mismatch in the system sensitivity, which was calculated by Monte Carlo simulation. The log regularization method reconstructs few nonzero pixels when $\beta > 0$, even though the background is spatially diffuse. Although regularized image reconstruction is useful for locating sources, this example shows that the images may lead to erroneous conclusions when detecting a the presence of a source. This motivates our study of source detection in the next chapter.

3.8 Image Reconstruction Conclusion

The images shown in the previous section appear adequate to allow a human observer to locate a point–source of gamma–rays. However, all of the images contain artifacts that could possibly be misleading. In addition, experiments with real detectors have shown that measurements taken in the absence of a source can produce images that look like an image of a point–source. This phenomenon is due to the underdetermined nature of the problem, enforcing sparsity can force reconstruction of a sparse solution when the true spatial emission distribution is diffuse, and possibly model mismatch. In light of these observations, we will consider the problem of using a position–sensitive detector to detect the presence of a source and estimate its position and intensity. This formulation of the problem has the advantage of having fewer parameters, namely source intensity and position, assuming that the background is known. We investigate the source detection problem in the next chapter.

CHAPTER IV

Benefits of Position–Sensitive Detectors for Radioactive Source Detection[1]

This chapter addresses the question of whether or not position–sensitive detectors can perform better in terms of AUC than detectors without position–sensitive capability. We derive formulas for the asymptotic distributions of commonly used test statistics, and use these asymptotic distributions to determine the detection performance. We prove that the performance of a position–sensitive detector is asymptotically higher than that of a detector that merely counts photons when the detector sensitivity is uniform and the source position, source energy, and background intensity are known. We also use asymptotic performance approximations to explore the performance of simple detectors when the system response is known exactly.

Some detection algorithms are specific to certain gamma–ray imaging modalities, such as using a coded aperture system to detect a point–source [68]. A mean difference test (MDT) statistic was applied to scintillating arrays in [51]. In this work, we analyze the detection performance of two tests that are based on the likelihood: one is the generalized likelihood ratio test (GLRT), and the other is the ML estimate of the source intensity. We refer to the latter as the source intensity test (SIT) [48] .

When performing detection using the GLRT with a position–sensitive Compton

[1]The work in this chapter was previously published in [48] and [47].

imaging detector in a known background, a particular experiment found that position information did not significantly improve detection performance over merely counting received photons [75]. In light of the numerous modalities that provide information beyond the number of received counts, we seek to show theoretically how position–sensitive capability affects detection performance. The goals of this work are to explain previous empirical results, such as those contained in [75], and to help guide the design of future detectors.

The question of whether or not imaging capability improves detection performance was addressed in [84] in the context of coded–aperture imaging systems. Reference [84] shows that if the background intensity is unknown, imaging may improve SNR [84] and thus detection performance by providing a means to separate the otherwise indistinguishable source and background photons. The analysis in this work differs from that of [84] because we treat the problem from a detection task–based point of view [4]. We analyze the task of source detection using the asymptotic performance of various test statistics applied to detectors with and without position–sensitive capability. Our treatment of detection differs from that in [84] because SNR does not capture the additional information received on a per–photon basis by a position–sensitive detector.

To quantify detection performance, we calculate the receiver operating characteristic curve (ROC), which is the probability of detection as a function of the probability of false alarm. A desirable detector will have a high probability of detection for a low false alarm rate. We also quantify detection performance using the area under the receiver operating characteristic curve (AUC). This metric is independent of any particular threshold value and is a measure of the overall detectability [28]. Computation of the ROC requires expressions for the probability distribution of the

ML estimate of source intensity, which are often intractable for the complex system models used in gamma–ray imaging. Instead of attempting to compute the distribution of the ML estimates directly, we approximate them by their first and second moments. This is known as asymptotic approximation. Since ML estimates are asymptotically normal, the true distribution of the estimates will approach the asymptotic approximation as the number of received photons tends to infinity. The accuracy of the asymptotic approximation for finite sample sizes is related to how well the distribution of the ML estimates can be approximated by its first and second moments. We compare the asymptotic performance of the GLRT, SIT, and the MDT presented in [51] for detecting a gamma–ray point source in background with hypothetical gamma–ray imaging systems with no model mismatch.

We compare the AUCs of position–sensitive and counting detectors when the assumptions of Theorem IV.3 do not hold but the system model is exactly known. For the case of known background and nonuniform sensitivity, we numerically evaluate the asymptotic AUC of the GLRT and SIT for a simple position–sensitive detector of nonuniform sensitivity. Our results show that, unlike the uniform sensitivity case, a position–sensitive detector with nonuniform sensitivity can have a smaller AUC than a uniform–sensitivity counting detector of equal area for some source positions, particularly for a position–sensitive detector with poor sensitivity and poor position resolution. However, as sensitivity and resolution improve, the position–sensitive detector can have better detection performance. We use examples of various detectors to show how nonuniformity and other detector properties relate to detection performance, which could help in practical design problems.

4.1 Mathematical Background

4.1.1 Measurement Model

The model described in this section is general enough to describe any system that records a Poisson–distributed number of measurements or events, where the events are independent and each event is described by a vector of recorded attributes. This model accurately describes position–sensitive Compton detectors, coded aperture detectors, and scintillator arrays. The model is based on [4] and [57]. It assumes a fixed scan time, thus the number of recorded events is random.

Consistent with the notation in §2.3 , let the list of J recorded photon interaction event attributes be $\tilde{\boldsymbol{r}} = (\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_J)$. Each element $\boldsymbol{r}_i$ of $\tilde{\boldsymbol{r}}$ is itself a vector of attributes describing the ith event. An example of a detector that one can describe with this model is a position–sensitive Compton detector. A Compton detector records a Poisson–distributed number J of gamma–ray photons. Each detected photon interacts one or more times inside the detector and the detector records these interaction locations and possibly other attributes such as deposited energy. We assume that distinct attribute vectors are statistically independent, which is reasonable provided the count rates are low enough to avoid dead time effects [4].

In the detection problem considered here, the goal is to decide whether or not a point–source is present in an environment with some background. Denote the source intensity by $\alpha \in [0, \infty)$ with units of counts *emitted* per unit time. The probability distribution of recorded attribute vectors $\boldsymbol{r}$ for events that originate at the source may depend on parameters other than the source intensity, and we denote the vector of additional parameters by $\boldsymbol{\phi} \in \Phi$. In the 3D far–field when the source and background are of the same energy, an example for the set Φ is $[0, 2\pi] \times [0, \pi]$, which represents all possible source positions in terms of azimuthal and polar angle

in space. If a detector is energy–sensitive and the source and background energy spectra differ, the source energy could also be an element of $\boldsymbol{\phi}$.

We model the background as a linear combination of a finite number of fixed, known distributions. We parameterize the background by the rate λ_b in counts *recorded* per unit time and a vector of mixture coefficients $\boldsymbol{\beta}$, such that the probability distribution of recorded events given that they originated from background is a mixture of the distributions of recorded attributes given that they came from each object [57]. For this linear model, the background count rate, λ_b, is not a function of $\boldsymbol{\beta}$.

We define the vector $\boldsymbol{\theta}$ to be the vector of all unknown parameters. When the source and background intensities and position parameters are all unknown,

$$\boldsymbol{\theta} = (\alpha, \boldsymbol{\phi}, \lambda_b, \boldsymbol{\beta}). \tag{4.1}$$

A similar parameterization of a far–field point source is given in [51].

Let D be the event that a photon is recorded and G be the event that a photon passes through the detector. We define the sensitivity, which is the probability of recording a photon given that it came from direction $\boldsymbol{\phi}$, to be

$$s(\boldsymbol{\phi}) \stackrel{\triangle}{=} \mathsf{p}\,(D;\boldsymbol{\phi}) = \mathsf{p}(D|G;\boldsymbol{\phi})\mathsf{p}(G;\boldsymbol{\phi}), \tag{4.2}$$

where $\mathsf{p}\,(D|G;\boldsymbol{\phi})$ is the intrinsic sensitivity, which depends on the detector shape and attenuation, and $\mathsf{p}(G;\boldsymbol{\phi})$ is the geometric sensitivity, which depends on the fraction of emitted photons that pass through the detector [73, p. 65]. In 3D, the geometric sensitivity is the solid angle subtended by the detector in a spherical coordinate system centered at the source.

Let $\mathsf{p}(\boldsymbol{r}|D;\boldsymbol{\theta})$ denote the distribution of recorded attributes $\boldsymbol{r} \in \mathcal{R}$, where $\mathcal{R}$ is the set of all possible event attributes. Let $\mathsf{p}_{\mathrm{S}}(\boldsymbol{r}|D;\boldsymbol{\phi})$ and $\mathsf{p}_{\mathrm{B}}(\boldsymbol{r}|D;\boldsymbol{\beta})$ denote the

distributions of recorded event attributes $\boldsymbol{r}$ given that they are detected and come from the source and background, respectively. The overall distribution of recorded attributes (given that an event is detected) is a mixture of $\mathsf{p}_{\mathrm{S}}(\boldsymbol{r}|D;\boldsymbol{\phi})$ and $\mathsf{p}_{\mathrm{B}}(\boldsymbol{r}|D;\boldsymbol{\beta})$ given by [57]

$$\mathsf{p}(\boldsymbol{r}|D;\boldsymbol{\theta}) = \frac{\lambda_b \mathsf{p}_{\mathrm{B}}(\boldsymbol{r}|D;\boldsymbol{\beta}) + \alpha s(\boldsymbol{\phi})\mathsf{p}_{\mathrm{S}}(\boldsymbol{r}|D;\boldsymbol{\phi})}{\lambda_b + \alpha s(\boldsymbol{\phi})}. \tag{4.3}$$

As shown in [4], the likelihood of $\boldsymbol{\theta}$ is

$$\mathsf{p}(\tilde{\boldsymbol{r}};\boldsymbol{\theta}) = \prod_{i=1}^{J} \mathsf{p}(\boldsymbol{r}_i|D;\boldsymbol{\theta}) e^{-\tilde{J}(\boldsymbol{\theta})} \tilde{J}(\boldsymbol{\theta})^J / J!, \tag{4.4}$$

and the number of recorded photons obeys the Poisson distribution

$$J \sim \text{Poisson}\left(\tilde{J}(\boldsymbol{\theta})\right), \tag{4.5}$$

with mean given by

$$\tilde{J}(\boldsymbol{\theta}) \triangleq \mathsf{E}\left[J\right] = \tau\left(\lambda_b + \alpha s(\boldsymbol{\phi})\right), \tag{4.6}$$

where τ denotes the known measurement recording time.

4.1.2 Fisher Information Matrix

The Fisher information matrix, defined in (2.34), also appears in the asymptotic distributions of maximum–likelihood estimates. To facilitate the analysis of $\mathsf{F}\left(\boldsymbol{\theta}\right)$ and asymptotic detection performance, we define its block components as follows:

$$\mathsf{F}\left(\boldsymbol{\theta}\right) = \begin{bmatrix} \mathsf{F}_{[1,1]} & \mathsf{F}_{[2,1]}^T & \mathsf{F}_{[3,1]}^T & \mathsf{F}_{[4,1]}^T \\ \mathsf{F}_{[2,1]} & \mathsf{F}_{[2,2]} & \mathsf{F}_{[3,2]}^T & \mathsf{F}_{[4,2]}^T \\ \mathsf{F}_{[3,1]} & \mathsf{F}_{[3,2]} & \mathsf{F}_{[3,3]} & \mathsf{F}_{[4,3]}^T \\ \mathsf{F}_{[4,1]} & \mathsf{F}_{[4,2]} & \mathsf{F}_{[4,3]} & \mathsf{F}_{[4,4]} \end{bmatrix}, \tag{4.7}$$

where $\mathsf{F}_{[1,1]}$ is 1×1, $\mathsf{F}_{[2,1]}$ is $\dim(\boldsymbol{\phi}) \times 1$, $\mathsf{F}_{[2,2]}$ is $\dim(\boldsymbol{\phi}) \times \dim(\boldsymbol{\phi})$, $\mathsf{F}_{[3,1]}$ is 1×1, $\mathsf{F}_{[3,2]}$ is $1 \times \dim(\boldsymbol{\phi})$, $\mathsf{F}_{[3,3]}$ is 1×1, $\mathsf{F}_{[4,1]}$ is $\dim(\boldsymbol{\beta}) \times 1$, $\mathsf{F}_{[4,2]}$ is $\dim(\boldsymbol{\beta}) \times \dim(\boldsymbol{\phi})$, $\mathsf{F}_{[4,3]}$ is

$\dim(\boldsymbol{\beta}) \times 1$, and $\mathsf{F}_{[4,4]}$ is $\dim(\boldsymbol{\beta}) \times \dim(\boldsymbol{\beta})$. In source detection problems, the source position $\boldsymbol{\phi}$, the background intensity λ_b, and the background shape parameters $\boldsymbol{\beta}$ are nuisance parameters.

We can make the concept of a counting detector more concrete by the following definition:

Definition IV.1. A detector is a *counting* detector if and only if $\mathsf{p}_{\mathrm{S}}(\boldsymbol{r}|D;\boldsymbol{\phi}) = \mathsf{p}_{\mathrm{B}}(\boldsymbol{r}|D;\boldsymbol{\beta})$ almost everywhere[2] for all $\boldsymbol{\phi} \in \Phi$ and mixture coefficients $\boldsymbol{\beta}$.

Definition IV.1 says that in a counting detector, the distribution of event attributes is independent of whether or not the event originated from the source. Otherwise we call it a position–sensitive detector. By this definition, a detector that is energy–sensitive but not does not record interaction locations is not a counting detector. In such a detector, the recorded energies can produce some position information, so we treat spectrometers that do not record interaction locations similarly to spectrometers that do. The asymptotic expressions in this paper allow one to compare the asymptotic detection performance of particular spectrometers that do and do not record interaction locations. We refer to detectors that are not counting detectors as *position–sensitive* detectors, although a position–sensitive detector does not necessarily record interaction locations.

A counting detector is neither capable of estimating $\boldsymbol{\phi}$ nor distinguishing source and background events of the same energy because $\mathsf{p}_{\mathrm{S}}(\boldsymbol{r}|D;\boldsymbol{\phi})$ does not depend on $\boldsymbol{\phi}$ by Definition IV.1. Because of this, for the purposes of defining the Fisher information, we assume that λ_b, $\boldsymbol{\beta}$, and the value of $s(\boldsymbol{\phi})$ are known to a counting detector, so the Fisher information is a scalar in this case. Using the model in (4.4) and (4.5), the likelihood for the counting case is $\mathsf{p}\left(J;\boldsymbol{\theta}\right) = \tilde{J}(\boldsymbol{\theta})^J e^{-\tilde{J}(\boldsymbol{\theta})}/J!$, for which

[2]Throughout, "almost everywhere" means with respect to the distribution of $\boldsymbol{r}$ in (6.4)

the Fisher information is given by the scalar

$$\mathsf{F}_{\mathrm{c}}(\boldsymbol{\theta}) \triangleq \frac{\tau s^2(\boldsymbol{\phi})}{\lambda_b + \alpha s(\boldsymbol{\phi})}. \tag{4.8}$$

To help express the Fisher information matrix (4.7) for a position–sensitive detector, we first define the following functions of $\boldsymbol{r}$:

$$\begin{aligned} g_1(\boldsymbol{r}) &= \mathsf{p}_{\mathrm{S}}(\boldsymbol{r}|D;\boldsymbol{\phi}) \\ g_2(\boldsymbol{r}) &= \nabla_{\boldsymbol{\phi}}(s(\boldsymbol{\phi})\mathsf{p}_{\mathrm{S}}(\boldsymbol{r}|D;\boldsymbol{\phi})) \\ g_3(\boldsymbol{r}) &= \mathsf{p}_{\mathrm{B}}(\boldsymbol{r}|D;\boldsymbol{\beta}) \\ g_4(\boldsymbol{r}) &= \nabla_{\boldsymbol{\beta}}\mathsf{p}_{\mathrm{B}}(\boldsymbol{r}|D;\boldsymbol{\beta}), \end{aligned}$$

where $\nabla_{\boldsymbol{\beta}}$ is the column gradient with respect to $\boldsymbol{\beta}$. Note that $g_1, g_3 : \mathcal{R} \to \mathbb{R}$, $g_2 : \mathcal{R} \to \mathbb{R}^{\dim(\boldsymbol{\phi})}$, and $g_4 : \mathcal{R} \to \mathbb{R}^{\dim(\boldsymbol{\beta})}$. Using (6.4), (4.4), and (4.5), one can show that the Fisher Information for a position–sensitive detector is given by

$$\mathsf{F}\left(\boldsymbol{\theta}\right) = \mathsf{F}_{\mathrm{c}}(\boldsymbol{\theta}) \begin{bmatrix} K_{[1,1]} & \frac{\alpha K_{[2,1]}^T}{s(\boldsymbol{\phi})} & \frac{K_{[3,1]}}{s(\boldsymbol{\phi})} & \frac{\lambda_b K_{[4,1]}^T}{s(\boldsymbol{\phi})} \\ \frac{\alpha K_{[2,1]}}{s(\boldsymbol{\phi})} & \frac{\alpha^2 K_{[2,2]}}{s^2(\boldsymbol{\phi})} & \frac{\alpha K_{[3,2]}^T}{s(\boldsymbol{\phi})} & \frac{\alpha\lambda_b K_{[4,2]}^T}{s^2(\boldsymbol{\phi})} \\ \frac{K_{[3,1]}}{s(\boldsymbol{\phi})} & \frac{\alpha K_{[3,2]}}{s(\boldsymbol{\phi})} & \frac{K_{[3,3]}}{s^2(\boldsymbol{\phi})} & \frac{\lambda_b K_{[4,3]}^T}{s^2(\boldsymbol{\phi})} \\ \frac{\lambda_b K_{[4,1]}}{s(\boldsymbol{\phi})} & \frac{\alpha\lambda_b K_{[4,2]}}{s^2(\boldsymbol{\phi})} & \frac{\lambda_b K_{[4,3]}}{s^2(\boldsymbol{\phi})} & \frac{\lambda_b^2 K_{[4,4]}}{s^2(\boldsymbol{\phi})} \end{bmatrix}, \tag{4.9}$$

where

$$K_{[i,j]} \triangleq \mathsf{E}\left[\frac{g_i(\boldsymbol{r})g_j^T(\boldsymbol{r})}{\mathsf{p}^2(\boldsymbol{r}|D;\boldsymbol{\theta})}\right], \tag{4.10}$$

provided that the expectation and the gradient with respect to the parameters are interchangeable. Appendix B.2 gives sample derivations of the block Fisher information elements.

In the case where the background intensities are known, we remove the entries corresponding to the unknown background and the Fisher information for a position–

sensitive detector simplifies to:

$$\mathsf{F}(\boldsymbol{\theta}) = \mathsf{F}_{\mathrm{c}}(\boldsymbol{\theta}) \begin{bmatrix} K_{[1,1]} & \frac{\alpha}{s(\boldsymbol{\phi})} K_{[2,1]}^T \\ \frac{\alpha}{s(\boldsymbol{\phi})} K_{[2,1]} & \frac{\alpha^2}{s^2(\boldsymbol{\phi})} K_{[2,2]} \end{bmatrix}. \tag{4.11}$$

The elements of the Fisher information appear prominently in the detection analyses that follow.

4.2 Effect of Position–Sensitive Capability on Detection Performance

As illustrated in Figure 4.1, the AUC of the GLRT for (2.39) is a monotone function of the noncentrality parameter η in (2.41), so to show that position–sensitive capability improves AUC, it suffices to show that position–sensitive capability increases η. Furthermore, if position–sensitive capability increases η for all values of $\alpha \geq 0$, then position–sensitive capability will improve the performance of the SIT by reducing the variance of the test statistic under each hypothesis in (2.43).

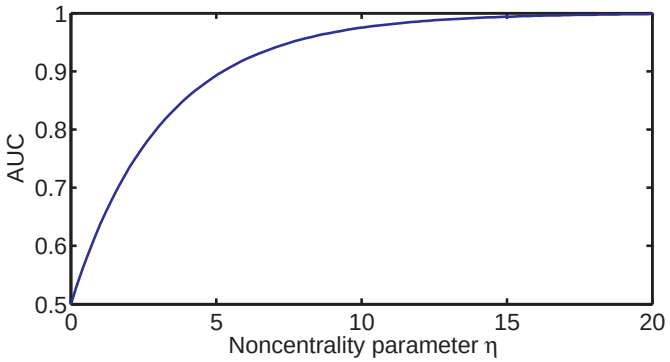


Figure 4.1: AUC of GLRT (2.38) versus noncentraltity parameter η of the asymptotic distribution of the GLRT.

4.2.1 Effect of Position–Sensitive Capability for a Uniform Sensitivity Detector with Known Background

We first define the concept of a uniform–sensitivity detector:

Definition IV.2. A detector has uniform–sensitivity if $s(\boldsymbol{\phi}) = s_0$ for all $\boldsymbol{\phi} \in \Phi$, where s_0 is a constant.

The main result of this section, given by Theorem IV.3, is that the detection performance of a uniform–sensitivity position–sensitive detector is greater than or equal to that of a uniform sensitivity counting detector. The AUC of the GLRT applied to a position–sensitive detector is greater than the AUC of the GLRT applied to a counting detector if the noncentrality parameter of the asymptotic distribution under H_1 in (2.41) is larger for a position–sensitive detector. As shown in [35, p. 232], the asymptotic distribution of the GLRT is most accurate for small source intensities because the derivation of the asymptotic distribution assumes that the log–likelihood evaluated at $\alpha = 0$ is approximately equal to the second order approximation of the log–likelihood about the value of the source intensity estimate $\hat{\alpha}_{\mathrm{ML}}$. For the SIT, a detector with position–sensitive capability performs better if the variances in (2.43) are smaller for the position–sensitive detector. The asymptotics of the SIT do not assume a small source intensity. The above discussion leads to the sufficient condition that position–sensitive capability improves the AUC of the GLRT and SIT for any $\boldsymbol{\theta} \in \boldsymbol{\Theta}$:

$$(\mathsf{F}(\boldsymbol{\theta})^{-1}_{[1,1]})^{-1} > \mathsf{F}_{\mathrm{c}}(\boldsymbol{\theta}). \tag{4.12}$$

Theorem IV.3. *For a uniform-sensitivity detector in a known background,* $(\mathsf{F}^{-1}(\boldsymbol{\theta})_{[1,1]})^{-1} \geq \mathsf{F}_{\mathrm{c}}(\boldsymbol{\theta})$*, i.e the reciprocal of the [1,1] component of the inverse Fisher Information Matrix (4.11) for a position–sensitive detector is greater than or equal to that of a counting detector (4.8). Therefore, the asymptotic AUC for a position–sensitive detector is greater than or equal to the asymptotic AUC of a counting detector in a known background when the GLRT with a small source intensity or the SIT is used.*

The proof, which is given in Appendix B.1, shows that the noncentrality parame-

ters of the GLRT for position–sensitive and counting detectors, η_{i} and η_{c}, respectively, obey

$$\frac{\eta_{\mathrm{i}}}{\eta_{\mathrm{c}}} = \frac{(\mathsf{F}^{-1}(\boldsymbol{\theta})_{[1,1]})^{-1}}{\mathsf{F}_{\mathrm{c}}(\boldsymbol{\theta})} \geq 1. \tag{4.13}$$

Furthermore, if $\mathsf{F}^{-1}(\boldsymbol{\theta})$ is diagonal,

$$\frac{\eta_{\mathrm{i}}}{\eta_{\mathrm{c}}} = K_{[1,1]}, \tag{4.14}$$

so position–sensitive capability increases the noncentrality parameter η in (2.41) by the factor $K_{[1,1]}$ in (4.9). In this case, the inequality in Theorem IV.3 becomes strict.

4.3 Setup for Numerical Calculations

In §4.4, we numerically evaluate the Fisher information (4.9) for 2D circular detectors of radius r and 2D square detectors with side length $2l$ to illustrate Theorem IV.3 and to explore the cases where Theorem IV.3 does not apply. The detectors in this section do not necessarily represent any particular detector technology and are used for illustrative purposes only. The simplified detectors have tractable models for the recorded attributes, facilitating accurate calculation of the Fisher information.

For simplicity, we assume that these detectors record only single photon interaction events and that the source and background energy spectra are identical. For each recorded event, the detector records the position of the interaction (x, y). The attribute vector $\boldsymbol{r}_i$ is the interaction position of the ith event. Each attribute vector $\boldsymbol{r}_i$ has length 2, so we compute the components of the Fisher information (5.40) numerically using Riemann approximation. This hypothetical system allows us to gain intuition on how detector nonuniformity and unknown background affect detection performance. This intuition will be useful when thinking about more realistic 3D detectors.

For this analysis, we assume that the point–source is in the far–field, so $\phi \in [0, 2\pi)$ denotes the source position in the 2D plane. The density of recorded attributes, $\mathsf{p}(x, y|D; \boldsymbol{\theta})$ depends on the source position ϕ, so the parameter vector $\boldsymbol{\phi}$ in (6.1) is equal to ϕ.

4.3.1 Single–Interaction Probabilities for Two-Dimensional Detectors

To analyze detection performance in a nonuniform sensitivity detector, we derive the interaction probabilities for single–interaction events in square and elliptical two–dimensional detectors. We present the formulas for interaction probabilities in this section on an abstract and general level so that they apply to all convex detectors.

Beer–Lambert law

Single–interaction events follow the Beer-Lambert law [3], which says that a beam of electromagnetic radiation with intensity I_0, after passing through a depth d of material with attenuation coefficient μ, is given by

$$I = I_0 e^{-\mu d}.$$

When the number of photons is large, $\frac{I}{I_0}$ will be approximately equal to $1 - \mathsf{Pr}\,(D; d)$, where $\mathsf{Pr}\,(D; d)$ is the probability that a photon interacts somewhere in the material prior to reaching the depth d. One can think of $\mathsf{Pr}\,(D; d)$ as a cumulative distribution function $F(d)$ of interaction depths in the material. Then, for a detector of thickness T, the probability density function for interaction depths for $0 \leq d \leq T$ is equal to

$$\mathsf{p}\,(d) = \frac{\partial}{\partial d} F(d) = \mu e^{-\mu d} \quad 0 \leq d \leq T.$$

This probabilistic interpretation of the Beer–Lambert law is the foundation of the analysis of single–interaction events.

Coordinate System

Instead of characterizing the interaction location by its (x, y) coordinates, we can define the 2×2 rotation matrix Q_ϕ that rotates a vector counterclockwise by ϕ. Then we can define a new basis for characterizing interaction locations

$$\boldsymbol{u} = Q_\phi \boldsymbol{e}_1$$

$$\boldsymbol{v} = Q_\phi \boldsymbol{e}_2.$$

We also define the interaction depth $d_i(u; v, \phi)$ to be the distance from the point $(u, v) = u\boldsymbol{u} + v\boldsymbol{v}$ to the edge of the detector along a ray with slope ϕ, and $d_t(v; \phi)$ to be the length of the line segment that lies inside the detector with slope ϕ, where each point on the line segment has v-coordinate v. If we define $l(\phi)$ to be the length of the longest line segment that intersects the detector and is perpendicular to any line with slope ϕ, the far field approximation allows us to say that $\mathsf{p}(v)$, which is the probability that a photon interacts with v-coordinate v is uniform across all possible v-coordinates inside the detector.

Interaction Probabilities

By the Beer–Lambert law,

$$\mathsf{p}(u|v, D) = \frac{\mu e^{-\mu d_i(u;v,\phi)}}{1 - e^{-\mu d_t(v;\phi)}}. \tag{4.15}$$

Also by the Beer–Lambert law,

$$\mathsf{p}(v|D) = \frac{\mathsf{p}(D|v)\mathsf{p}(v)}{\mathsf{Pr}(D)} \tag{4.16}$$

$$= \frac{(1 - e^{-\mu d_t(v;\phi)})\frac{1}{l(\phi)}}{\frac{1}{l(\phi)}\int_{l(\phi)/2}^{l(\phi)/2}(1 - e^{-\mu d_t(v;\phi)})dv} \tag{4.17}$$

$$= \frac{(1 - e^{-\mu d_t(v;\phi)})}{\int_{l(\phi)/2}^{l(\phi)/2}(1 - e^{-\mu d_t(v;\phi)})dv} \quad |v| \leq l(\phi)/2. \tag{4.18}$$

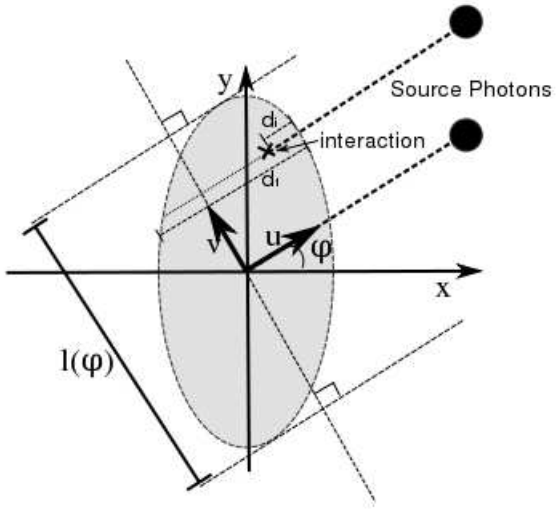


Figure 4.2: Diagram two-dimensional detector and basis change for interaction probability calculation

Figure 4.2 illustrates these quantities.

Now, by the definition of conditional probability,

$$\mathsf{p}(u, v|D) = \mathsf{p}(v|D)\mathsf{p}(u|v, D). \tag{4.19}$$

Substituting (4.15) into (4.19) and (4.18), we obtain

$$\mathsf{p}(u, v|D; \boldsymbol{\theta}) = \frac{\mu e^{-\mu d_i(u;v,\phi)}}{\int_{l(\phi)/2}^{l(\phi)/2}(1 - e^{-\mu d_t(v;\phi)})dv} \qquad (u, v) \in \mathcal{R}. \tag{4.20}$$

Noticing that

$$\int_{l(\phi)/2}^{l(\phi)/2}(1 - e^{-\mu d_t(v;\phi)})dv = l(\phi)\mathsf{Pr}\left(D\right), \tag{4.21}$$

we write (4.20) as

$$\mathsf{p}(u, v|D; \boldsymbol{\theta}) = \frac{1}{l(\phi)\mathsf{Pr}\left(D\right)}\mu e^{-\mu d_i(u;v,\phi)} \qquad (u, v) \in \mathcal{R}. \tag{4.22}$$

The quantities $d_i(u; v, \phi)$ and $d_t(v; \phi)$ can be computed geometrically for many different detector shapes. In the following sections, we visualize the detection performance of square and elliptical detectors.

It is difficult to define the intrinsic position resolution of these systems, so we report the angular uncertainty measured by the square root of the Cramer–Rao lower bound on the position estimate $\sqrt{(\mathsf{F}(\boldsymbol{\theta})^{-1})_{[2,2]}}$, where $\mathsf{F}(\boldsymbol{\theta})$ is defined in (4.9).

4.4 Numerical Results

We first consider uniform–sensitivity position–sensitive and counting detectors in a known background to illustrate the results of Theorem IV.3 and to explore under what conditions position information is most beneficial. Next, we consider square and elliptical position–sensitive detectors, which have nonuniform sensitivity, and compare their performance to uniform counting detectors of equal area. We then examine the performance of uniform–sensitivity detectors in unknown background. We exclude the case of a nonuniform sensitivity detector in an unknown background because its analysis requires many assumptions.

4.4.1 Uniform Sensitivity Detector with Known Background

A circular detector with radius r has inherent uniform sensitivity due to its circular symmetry. Figure 4.3a shows the AUC, the position–sensitive gain factor $K_{[1,1]}$ from (4.9), and the angular uncertainty as a function of attenuation–radius product μr for a circular detector. The quantity $\alpha\tau$ represents the expected number of photons emitted from the source during the scan and $\lambda_b\tau$ represents the expected number of background photons recorded. The expected number of source photons recorded is given by $\alpha\tau s_0$, where $s_0 = s(\boldsymbol{\phi})$ as defined in (4.2). We denote the sensitivity by s_0 to emphasize that the sensitivity of a circular detector is not a function of source position in the far field. In this section, we consider only source photons that pass through the detector, i.e., $\mathsf{p}(G;\boldsymbol{\phi}) = 1$ in (4.2). As guaranteed by Theorem IV.3, the AUC of the position–sensitive detector always exceeds that of the the counting

detector for each test.

The AUC of the SIT exceeds that of the GLRT for this experiment. This performance difference comes from the fact that the GLRT is asymptotically equivalent to the Wald test, which involves a square of the source activity estimate. When one uses an unrestricted MLE of the source intensity, this squaring operation reduces the separation between test statistic values under the two hypotheses.

Figure 4.3d shows s_0 as a function of μr to aid interpreting Figure 4.3a. For this experiment, the Fisher information matrix is diagonal, so $K_{[1,1]}$ is the multiplicative improvement in the noncentrality parameter of the asymptotic distribution of the GLRT under H_1, as expressed in (4.14). As μr increases, the angular uncertainty decreases and the difference in AUC between the detectors with and without position–sensitive information increases. Figures 4.3a and 4.3b illustrate that as the detector provides more precise position–sensitive information, the improvement in detection performance due to position–sensitive information increases *even when the background is known.*

Figure 4.4 shows the AUC for a circular uniform–sensitivity detector as a function of source intensity for a fixed background intensity. The AUC values for the position–sensitive and counting detectors differ the most for source intensities near the background intensity. For low source–to–background ratios, the source is difficult to detect with either detector, and when the source–to–background ratio is large, the source is so easily detected that position–sensitive capability provides little additional benefit. Again, the position–sensitive detector always has higher AUC than the counting detector as expressed in Theorem IV.3, and the SIT performs better than the GLRT.

The next section considers the detection performance of nonuniform–sensitivity

detectors, to which Theorem IV.3 does not apply, but which can be more practical to build than uniform sensitivity detectors.

4.4.2 Nonuniform Sensitivity Square Detector with Known Background

For implementation and manufacturing reasons, current position–sensitive Compton imaging detectors are often box–shaped [30]. We examined the performance of the GLRT and SIT applied to the 2D square detector in Figure 4.5 to gain insight into the performance of nonuniform sensitivity detectors.

In the case of a position–sensitive detector, the maximum–likelihood estimate of the source intensity depends on $s(\boldsymbol{\phi})$, but a counting detector, by definition, gives no information about $\boldsymbol{\phi}$. To compute the ML estimate of the source intensity with a counting detector, one would need to assume a particular value of $s(\boldsymbol{\phi})$ because $s(\boldsymbol{\phi})$ appears in the likelihood and is a function of the unobservable position $\boldsymbol{\phi}$. Substituting a fixed value for $s(\boldsymbol{\phi})$ will result in a likelihood model that does not match the true distribution, so the estimator based on that model can be biased. Since the asymptotic analysis considered in this work does not apply when the parameter estimators are biased, we compare the square position–sensitive detector to a circular uniform–sensitivity counting detector of equal area with radius $r = 2l/\sqrt{\pi}$. This comparison constrains the amount of detector material and explores whether the position–sensitive square or the counting circle has better detection performance using the SIT and GLRT.

Figure 4.6a shows the AUC of the square position–sensitive and the circular counting detectors as a function of source position for both the GLRT and SIT. For $\mu l = 0.5$, the square position–sensitive detector performs worse than the circular counting detector. However, the detector with $\mu l = 5$ outperforms the counting detector for all source positions. As in the uniform–sensitivity case, the SIT outper-

forms the GLRT.

To aid in visualization, we present the sensitivity of the square relative to that of a circular detector of equal area. To compute the relative sensitivity in Figure 4.6b, we normalized the geometric sensitivity of the square detector in (4.2) by that of a circular detector of the same area, so in this case,

$$\mathsf{p}(G|\boldsymbol{\phi}) = \frac{2l\left(|\cos\boldsymbol{\phi}| + |\sin\boldsymbol{\phi}|\right)}{\frac{4l}{\sqrt{\pi}}}.$$

By using the relative sensitivity instead of the absolute sensitivity, the source intensity α has units of counts impinging on the detector per unit time. This figure, along with Figure 4.6a, shows that the detection performance of the square is better when the sensitivity is larger, and that detection performance is better for source positions where the slope of the sensitivity curve is small. In practice, one could generate similar plots with the appropriate sensitivity for a particular application.

Figure 4.6c shows the angular uncertainty as a function of source position. The AUC in Figure 4.6a is largest when the angular uncertainty is largest near $\boldsymbol{\phi} = 45°$, but this is not contradictory because the sensitivity, shown in Figure 4.6b, is approximately uniform near $\boldsymbol{\phi} = 45°$. Because of this, the Fisher information matrix is approximately diagonal for source positions near $\boldsymbol{\phi} = 45°$, which means that the position nuisance parameter has little effect on the $[1, 1]$ component of $\mathsf{F}(\boldsymbol{\theta})$.

Whether a nonuniform sensitivity position–sensitive detector is better than a counting detector depends on the characteristics of the counting detector used for comparison. A position–sensitive detector provides information about the source position, whereas a counting detector does not. For some applications, the position information could outweigh a smaller AUC for some source positions.

4.4.3 Nonuniform Sensitivity Elliptical Detector with Known Background

We analyze the performance of an elliptical detector because we can see how detection performance varies as we vary its eccentricity, defined as the ratio of its major and minor axes. The eccentricity can be viewed as a measure of detector nonuniformity. These analyses were performed by evaluating the single–interaction probabilities on a grid of coordinates inside the detector, and using numerical integration to approximate the Fisher information as defined in (4.9). The ellipse is positioned so that its two axes are parallel to the x and y axes. Let a be half of the axis length in the x-direction, and let b be half of the axis length in the y direction. Figure 4.7 illustrates this setup.

Figure 4.8 shows the AUC vs. source angle ϕ for various elliptical eccentricities. As the eccentricity of the ellipse increases, the detection performance at its point of highest sensitivity ($\phi = 90^{\text{o}}$) increases, but at the expense of performance for other source positions. When $\mu = 2$, we see that the gain in performance at $\phi = 90^{\text{o}}$ and the degradation for other angles are less extreme. This is because as μ increases, the sensitivity becomes more uniform because the high material attenuation dominates the difference in cross–section length at different source positions.

4.4.4 Uniform Sensitivity Detector with Spatially–Uniform Unknown Background

We analyze the detection performance of a 2D uniform–sensitivity circular position–sensitive detector with attenuation μ and radius r with a point source in a spatially uniform background of *unknown* intensity λ_b, with the goal of examining how not knowing the background affects detection performance. We use the $[1, 1]$ element of the inverse of the 3×3 block Fisher information matrix in (4.9). We remove the 4th row and column of $\mathsf{F}(\boldsymbol{\theta})$ because there is no $\boldsymbol{\beta}$ to parameterize the background

mixture in this model. We then use (2.40) and (2.43) to quantify the AUC.

Figure 4.9 shows the AUC for a position–sensitive detector in a uniform background of known and unknown intensity as a function of the true background intensity. As the true background intensity increases, the difference between the AUCs of known and unknown background increases for each particular test statistic. As in the known background case, the SIT outperforms the GLRT. In Figure 4.9, we also show the performance of the MDT [51]. The MDT, as presented in [51] assumes a spherical array of detectors, and the detectors considered in this work record interactions in the interior of the detector. As the attenuation coefficient μ of these detectors increases, they behave more like a circular array of scintillators. To evaluate the performance of the MDT for finite–attenuation detectors, we computed the asymptotic mean number of counts recorded in the semicircles facing and opposite the source. The MDT performs better than the GLRT and worse than the SIT for low source–to–background ratios in this experiment. We assumed that the MDT knows the source position, so that the boundary between the two semicircles is not estimated. Although this is not a fair comparison, it gives the MDT the advantage.

Figure 4.10 shows the AUC for a position–sensitive detector in a spatially uniform background of known and unknown intensity as a function of the attenuation–radius product. As the μr product increases, the difference in AUC between the known and unknown background case decreases for the SIT and GLRT. Also, as μr increases, the SIT performs better than the MDT. A possible explanation for this is that the SIT considers the interaction location of each photon and the MDT considers only the semicircle in which the interaction occurred.

Notice that the difference in AUC between the known and unknown background cases for each test statistic decreases as μr increases. Recall from Figure 4.3c that the

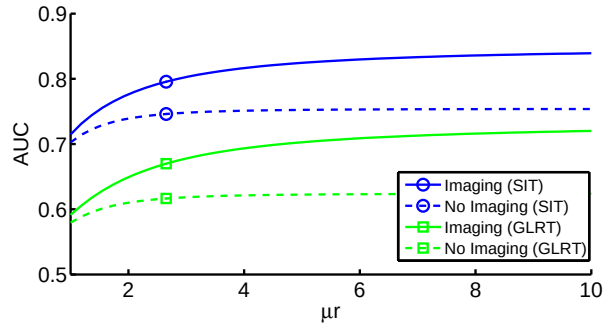
angular uncertainty decreases as μr increases. As the angular uncertainty decreases, the variance of the background intensity estimate decreases because the detector can more reliably distinguish source and background photons. One could recover the AUC "lost" by not knowing the background by increasing the scan time somewhat.

4.4.5 Position–Sensitive versus Counting Detectors with Unknown Background

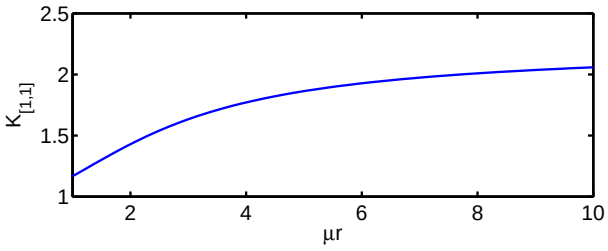
In practice, one can sometimes measure the background prior to screening for sources of interest. Also, an experienced operator of a counting detector could plausibly guess the background with some degree of uncertainty. The detection performance of such an operator depends on the accuracy or distribution of such guesses. As a hypothetical example, suppose that the operator of a counting detector applies the GLRT for a "known background" hypothesis test using a background rate $\check{\lambda}_b$ distributed according to the following gamma distribution [22, p. 291]:

$$\mathsf{p}\left(\check{\lambda}_b; \lambda_b, \xi\right) = \check{\lambda}_b^{\frac{\lambda_b}{\xi}-1} \frac{e^{-\check{\lambda}_b/\xi}}{\Gamma\left(\frac{\lambda_b}{\xi}\right)\xi^{\frac{\lambda_b}{\xi}}}, \tag{4.23}$$

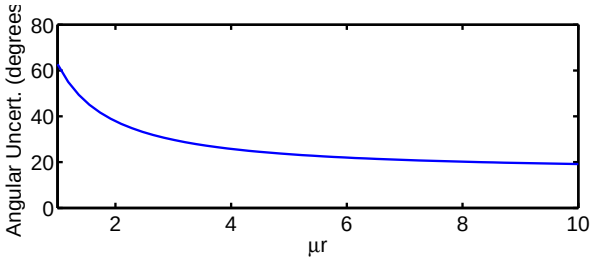
where ξ is a scale parameter that could represent the operator's accuracy and λ_b is the true background rate. Note that $\mathsf{E}\left[\check{\lambda}_b\right] = \lambda_b$ and $\mathsf{Var}\left(\check{\lambda}_b\right) = \lambda_b\xi$. If ξ is small, then the operator's guesses are narrowly distributed about the true background rate, and as ξ increases, the guesses are farther from the mean, on average. The gamma distribution is a reasonable model for operator uncertainty because of its nonnegativity, and there may be other models that are more accurate in practice. Figure 4.11 shows the AUC of the GLRT using an operator's guess, for various scale parameters ξ. Even when ξ is small, the position–sensitive detector with large μr still performs better. This is because as $\xi \to 0$ and $\mu r \to \infty$, the AUCs of both detectors approach the known background case, where Theorem IV.3 applies.



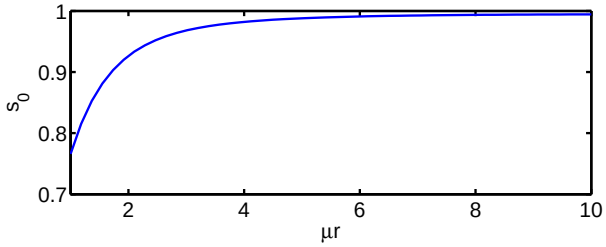
(a) AUC versus μr



(b) $K_{[1,1]}$ versus μr



(c) Angular Uncertainty versus μr



(d) s_0 versus μr

Figure 4.3: Various quantities for a circular uniform–sensitivity detector with $\alpha\tau = 10$, $\lambda_b\tau = 100$, and geometric sensitivity $\mathsf{p}(G|\phi) = 1$.

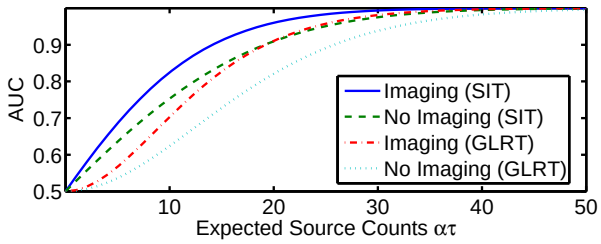


Figure 4.4: AUC versus expected source counts for a circular uniform–sensitivity detector with $\lambda_b \tau = 100$ and $\mu r = 5$.

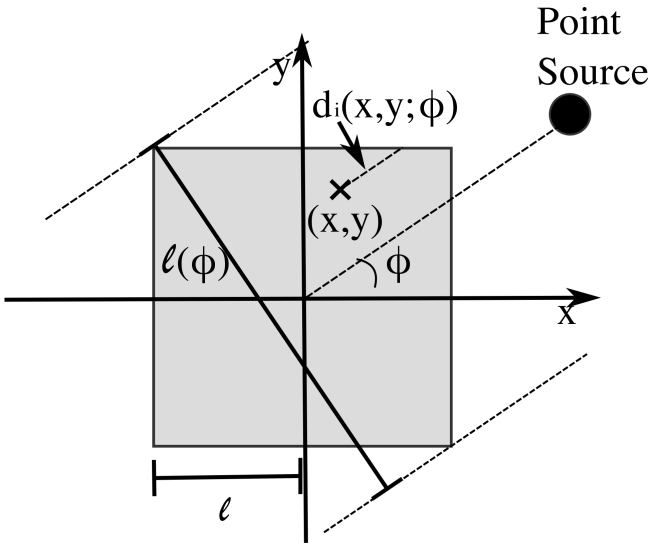
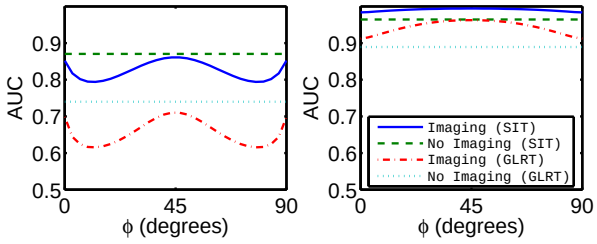
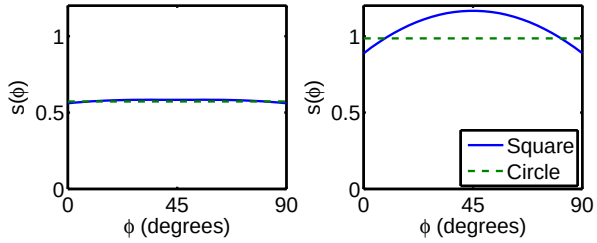


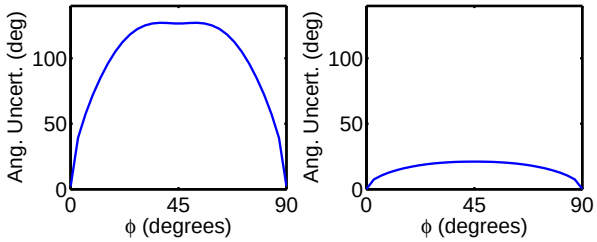
Figure 4.5: Diagram of 2D square detector



(a) AUC versus ϕ



(b) $s(\phi)$ versus ϕ



(c) Angular Uncertainty versus ϕ

Figure 4.6: AUC, relative sensitivity, and angular uncertainty for square position–sensitive and circular counting detectors of equal area ($r = 2l/\sqrt{\pi}$) vs. ϕ with $\mu l = 0.5$ (left) and $\mu l = 5$ (right), and $\alpha\tau = \lambda_b\tau = 10$.

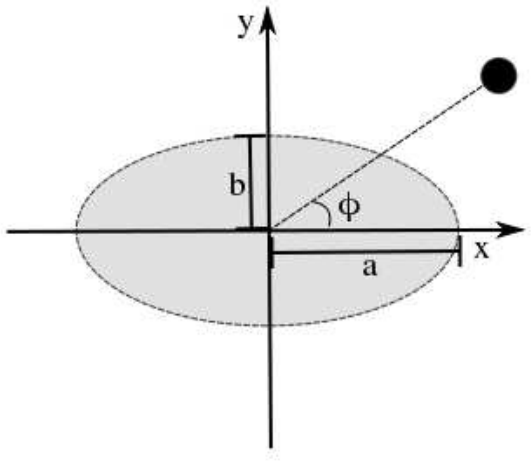


Figure 4.7: Diagram of elliptical detector

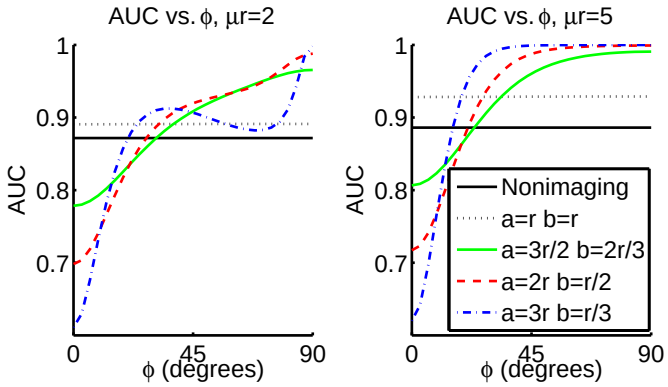


Figure 4.8: Detectability parameter λ vs. ϕ for elliptical detector with $\alpha = 1$, $\lambda_b = 1$, $\tau = 10$, and $\mu = 2$ (left) and $\mu = 5$ (right)

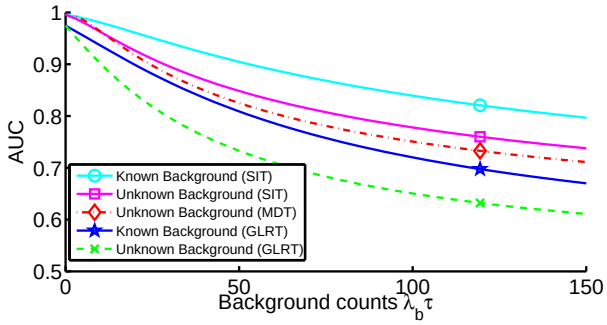


Figure 4.9: AUC vs. $\lambda_b\tau$ for spatially uniform background of known and unknown intensity background where $\mu r = 10$ and $\alpha\tau = 10$.

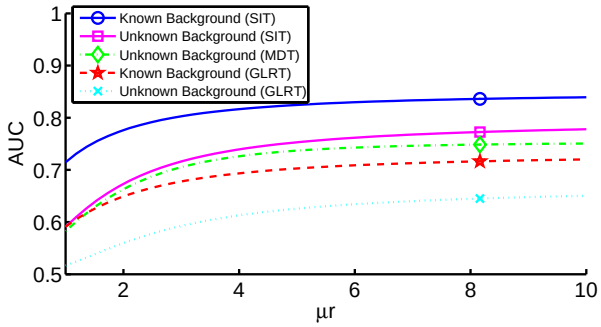


Figure 4.10: AUC for spatially uniform background of known and unknown intensity versus attenuation–radius product for $\alpha\tau = 10, \lambda_b\tau = 100$.

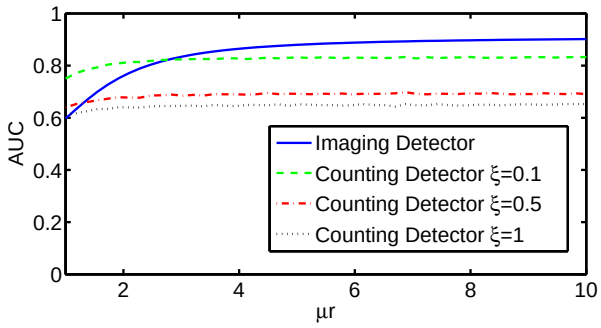


Figure 4.11: AUC vs. μr for position–sensitive and counting detectors, where the counting detector uses a "guessed" background rate in the "known background" GLRT. $\alpha\tau = \lambda_b\tau = 10$.

CHAPTER V

Asymptotic Source Detection Performance Under Model Mismatch[1]

In Chapter IV, we used asymptotic approximations to prove that position–sensitive detectors perform better in terms of AUC for the task of detecting a source in background than photon–counting detectors. We also used the asymptotic approximations to study the detection performance of simple non–uniform detectors in various environments. These detectors were easy to model because we assumed that the detectors are two–dimensional and that photons interact only once. More practical gamma–ray imaging systems, such as pixelated CdZnTe [30], are difficult to model exactly [80] due to the nonlinear physics of Compton scattering and the non–Gaussian uncertainties introduced by position quantization and Doppler broadening.

In this chapter, we extend the theory of Chapter IV to account for model mismatch introduced by approximations to the system response. We first study the asymptotic normality of estimates obtained by maximizing a possibly incorrect, or misspecified, likelihood. Our theory is general enough to accommodate regularized likelihoods as well. We then employ the asymptotic results to approximate the distribution of the source intensity test (SIT). We provide numerical results showing that the proposed approximation method reasonably predicts the empirical performance.

[1]This work was partially published in [49] with results from a more practical and complex simulated system.

Asymptotics of estimators and likelihood ratios have been studied previously. Estimators derived from misspecified models are known as quasi–maximum–likelihood estimators (QMLE) [78]. For a nonrandom number of independent and identically distributed (IID) measurements, a QMLE derived from a misspecified model is asymptotically normal provided that certain regularity conditions are satisfied [78]. A QMLE can also be viewed as an M-estimate. The asymptotic normality and convergence of M-estimates is discussed in [70] and [31]. In section 5.2, we extend [78] to cases where the number of recorded measurements is a Poisson random variable. This extension is needed for gamma–ray imaging systems.

Theorems 2.2 and 3.2 of [78] show convergence and asymptotic normality, respectively, of the QMLE for a nonrandom number of measurements. In section 5.2, we extend [78] to cases where the number of recorded measurements is a Poisson random variable. This extension is needed for gamma–ray imaging systems. We also provide expressions for the asymptotic distribution of the source intensity test (SIT) [48] and generalized likelihood ratio test (GLRT) [35] that one can evaluate in practice using Monte Carlo methods. One can use these asymptotic distributions to compute the receiver operating characteristic curves (ROC) for these tests more efficiently than using empirical methods.

Radiation sources have nonnegative intensities, and we examine how enforcing the nonnegativity constraint on the source intensity estimate affects detection performance. One can choose not to enforce the nonnegativity constraint because the sum of the source and background intensities could be nonnegative if one of the individual estimates is negative. For example, if the background rate is misspecified to be higher than the true background rate, a negative source intensity estimate may have a higher likelihood. We prove Theorem V.7 which gives the asymptotic

distribution of the constrained QMLE for the gamma–ray imaging problem.

We show that the asymptotic distribution of the source intensity estimates are useful for detection performance prediction with a series of experiments with a simulated parallel–plate Compton detector. We consider cases where the background and source intensities are unknown, and the parameter estimates are constrained and unconstrained. Our results show that the proposed method of performance prediction based on asymptotics is reasonably accurate in most cases. We also provide examples where the model mismatch is severe and the proposed method gives poor performance predictions.

5.1 Background

To characterize asymptotic detection performance, we must first define the true distribution of recorded events and reasonable models for it. The true distribution is governed by physics but its exact form is typically unknown or difficult to compute in practice. We also describe a model distribution for a list of recorded events, which is often an approximation of the true distribution. We describe the model in general terms to allow adaptation to different types of gamma–ray detectors and perhaps other applications such as list–mode positron emission tomography (PET) scans [60].

5.1.1 True Distribution of Recorded Events

We assume that the true distribution of recorded events follows the list–mode model of [4] for Poisson measurements. During a fixed–duration scan, a gamma–ray imaging system records attribute vectors, such as interaction locations within the detector and deposited energy, for each photon interaction event. Let the list of attribute vectors recorded by the system be $\tilde{\boldsymbol{r}} = (\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_J)$. The random number

of measurements is $J \sim \text{Poisson}(\bar{J})$, where $\bar{J} = \lambda_{\mathsf{s}}\tau$, λ_{s} is the rate of recorded events in counts per unit time, and τ is the deterministic scan time chosen by the user. Each $\boldsymbol{r}_j \in \mathcal{R}$ where $\mathcal{R}$ is the set of all possible event attribute vectors. For example, in an energy and position–sensitive detector, $\boldsymbol{r}_j$ is a vector of the coordinates and energies of all interactions of the jth photon.

Let $\mathsf{p}(\boldsymbol{r})$ denote the true density[2] of the recorded attributes for a single interacting photon. In general, we do not parameterize this true distribution because it represents the true physical process. Provided that the count rate is low enough to avoid dead–time effects [4], the true probability density of the list of event attributes $\tilde{\boldsymbol{r}}$ is given by

$$\mathsf{p}(\tilde{\boldsymbol{r}}) = e^{-\bar{J}}\bar{J}^J/J! \prod_{j=1}^{J} \mathsf{p}(\boldsymbol{r}_j)\,. \tag{5.1}$$

5.1.2 Measurement Model

The true distribution of the observations in (5.1) is not always known exactly, so we now give an observation model. Since the mean number of emissions $\bar{J}$ is unknown, we model it by $\tilde{J}(\boldsymbol{\theta})$, where $\tilde{J}(\boldsymbol{\theta}) = \tilde{\lambda}(\boldsymbol{\theta})\tau$, $\tilde{\lambda}(\boldsymbol{\theta})$ is the modeled photon emission rate in counts per unit time, and $\boldsymbol{\theta}$ is a parameter vector that lies in the set $\Theta \subset \mathbb{R}^d$ for some integer d. Examples of parameters of interest in gamma–ray imaging are source intensity, source position, and source energy. Let $\tilde{\mathsf{p}}(\boldsymbol{r};\boldsymbol{\theta})$ denote the modeled attribute density, which should approximate the true density $\mathsf{p}(\boldsymbol{r})$. Under the above assumptions, we model the likelihood of the observations as follows [4, 48]:

$$\tilde{\mathsf{p}}(\tilde{\boldsymbol{r}};\boldsymbol{\theta}) \stackrel{\triangle}{=} e^{-\tau\tilde{\lambda}(\boldsymbol{\theta})}[\tau\tilde{\lambda}(\boldsymbol{\theta})]^J/J! \prod_{j=1}^{J} \tilde{\mathsf{p}}(\boldsymbol{r}_j;\boldsymbol{\theta}). \tag{5.2}$$

[2]Throughout, density refers to the probability density function if the attributes are continuous random variables, or the probability mass function if the attributes are discrete random variables. For more general or mixed distributions, density refers to the Radon-Nikodym density [8] with respect to an appropriate base measure.

Throughout, we use superscript "$\sim$" to denote functions or distributions that are part of the model and might differ from the true underlying functions or distributions that they represent.

We define model mismatch in Definition V.1 [78].

Definition V.1. Model mismatch exists if for all $\boldsymbol{\theta} \in \Theta$ such that $\tilde{\lambda}(\boldsymbol{\theta}) = \lambda_{\mathsf{s}}$, there exists an $\boldsymbol{r} \in \mathcal{R}$ such that $\tilde{\mathsf{p}}(\boldsymbol{r}; \boldsymbol{\theta}) \neq \mathsf{p}(\boldsymbol{r})$.

This definition means that there is no parameter in the parameter space such that the modeled and true distributions and count rates are identical. The modeled likelihood may have model mismatch as a result of regularization, and one can apply this theory to modeled likelihoods that include a regularization term.

5.1.3 Estimator definition

The quasi maximum likelihood (QML) estimate of the parameter vector $\boldsymbol{\theta}$ is given by

$$\tilde{\boldsymbol{\theta}}_\tau \stackrel{\triangle}{=} \arg\max_{\boldsymbol{\theta} \in \Theta} \; \log \tilde{\mathsf{p}}(\tilde{\boldsymbol{r}}; \boldsymbol{\theta}), \tag{5.3}$$

assuming a solution exists and is unique. The subscript τ is included to emphasize that the estimate is a function of a list of event attributes acquired during a scan of duration τ. In the next section, we examine the asymptotic properties of $\tilde{\boldsymbol{\theta}}_\tau$ as $\tau \to \infty$, i.e., as one records more events by increasing the scan time. Section 5.4 shows that the asymptotic approximation can accurately characterize the distribution of $\tilde{\boldsymbol{\theta}}_\tau$, even for a finite scan time.

5.2 Properties of the QMLE

We now describe the asymptotic behavior of the QMLE in (6.6). The results, stated in Theorems V.2 and V.4, are extensions of the results of [78] to the case

where the number of measurements is Poisson. The results in [78] apply only to the case where the number of measurements is not random, and therefore are not directly applicable to gamma–ray detection systems. The regularity conditions required for Theorems V.2 and V.4 are similar to those required for the results of [78], except we add conditions on the modeled count rate $\tilde{\lambda}(\boldsymbol{\theta})$ to ensure that the modeled likelihood in (6.3) is well–defined.

Let $\nabla_{\boldsymbol{\theta}}$ denote the $d \times 1$ column gradient with respect to $\boldsymbol{\theta}$, and let $\nabla^2_{\boldsymbol{\theta}}$ denote the $d \times d$ Hessian with respect to $\boldsymbol{\theta}$. By the definition in (6.6), assuming that the log–likelihood is differentiable and the maximizer is in the interior of the parameter space, the QMLE $\tilde{\boldsymbol{\theta}}_\tau$ satisfies

$$\nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}\left(\tilde{\boldsymbol{r}}; \boldsymbol{\theta}\right)\big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}_\tau} = \mathbf{0}. \tag{5.4}$$

We also define

$$\begin{aligned} \tilde{f}\left(\boldsymbol{\theta}\right) &\stackrel{\triangle}{=} \frac{1}{\tau} \mathsf{E}\left[\log \tilde{\mathsf{p}}\left(\tilde{\boldsymbol{r}}; \boldsymbol{\theta}\right)\right] \\ &= \frac{1}{\tau} \int_{\mathcal{R}} \log \tilde{\mathsf{p}}\left(\tilde{\boldsymbol{r}}; \boldsymbol{\theta}\right) \mathsf{p}\left(\tilde{\boldsymbol{r}}\right) d\tilde{\boldsymbol{r}}, \end{aligned} \tag{5.5}$$

and

$$\tilde{g}\left(\boldsymbol{\theta}\right) = \mathsf{E}\left[\nabla_{\boldsymbol{\theta}} \log \mathsf{p}\left(\tilde{\boldsymbol{r}}; \boldsymbol{\theta}\right)\right], \tag{5.6}$$

where the expectation is with respect to the true distribution $\mathsf{p}\left(\tilde{\boldsymbol{r}}\right)$. Even though the true distribution is unknown, we assume one can compute expectations by Monte Carlo integration with samples obtained from it.

5.2.1 Convergence

Let the asymptotic limit $\tilde{\boldsymbol{\mu}} \in \Theta$ be defined by

$$\tilde{\boldsymbol{\mu}} \stackrel{\triangle}{=} \arg\max_{\boldsymbol{\theta} \in \Theta} \; \tilde{f}\left(\boldsymbol{\theta}\right), \tag{5.7}$$

assuming that such a solution exists and is unique. Under suitable conditions, $\tilde{\boldsymbol{\mu}}$ is the limit of the sequence of estimates $\tilde{\boldsymbol{\theta}}_\tau$ as $\tau \to \infty$. Appendix C.1 describes one sufficient set of regularity conditions that guarantees existence, uniqueness, and convergence. These conditions are similar to those in [78], with extensions to the case of a random number of measurements. The regularity conditions apply only to the model, with the exception of the assumption that the attributes of different photon interactions are IID. The regularity conditions ensure that the log–likelihood function exists and that the QMLE exists and is unique. We discuss verification of these conditions in Appendix C.7. The convergence of the QMLE is stated in Theorem V.2, which extends Theorem 2.2 of [78] to the case of a Poisson number of measurements.

Theorem V.2. *Under suitable regularity conditions,* $\tilde{\boldsymbol{\theta}}_\tau \overset{\text{a.s.}}{\to} \tilde{\boldsymbol{\mu}}$ *as* $\tau \to \infty$.

A useful result for interpretation of $\tilde{\boldsymbol{\mu}}$ is given in Corollary V.3. The additional regularity conditions required for Corollary V.3 place restrictions on the true distribution. The proof and sufficient regularity conditions are given in Appendix C.1.

Corollary V.3. *Under suitable regularity conditions,* $\tilde{\boldsymbol{\mu}}$ *minimizes the Kullback-Leibler divergence [39] of the true distribution with respect to the modeled distribution.*

This Theorem and Corollary are shown in [78] for the case of J nonrandom. In the absence of model mismatch, the true distribution (5.1) and modeled distributions (6.3) are equal for some $\boldsymbol{\theta} \in \Theta$ such that $\lambda_{\mathsf{s}} = \tilde{\lambda}(\boldsymbol{\theta}_{\text{true}})$. Let $\boldsymbol{\theta}_{\text{true}}$ be the parameter value, assuming a unique solution, under which this equality is achieved. Then, in the absence of model mismatch, $\tilde{\boldsymbol{\mu}} = \boldsymbol{\theta}_{\text{true}}$ by Corollary V.3, and Theorem V.2 reduces to the usual asymptotic consistency of ML estimates [34] in the absence of model mismatch.

5.2.2 Asymptotic normality

A QMLE may also be asymptotically normal if certain conditions on the model and true distributions are met. Theorem V.4 is an extension of Theorem 3.2 of [78] to the case of a Poisson number of measurements. The regularity conditions ensure that the expectations in Theorem V.4 are well–defined. The proof is given in Appendix C.2.

Theorem V.4. *Under suitable regularity conditions, asymptotically as* $\tau \to \infty$*,*

$$\sqrt{\tau}\left(\tilde{\boldsymbol{\theta}}_\tau - \tilde{\boldsymbol{\mu}}\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \Sigma(\tilde{\boldsymbol{\mu}})), \tag{5.8}$$

where

$$\Sigma(\boldsymbol{\theta}) = \tilde{\mathsf{H}}^{-1}(\boldsymbol{\theta})\,\tilde{\mathsf{G}}(\boldsymbol{\theta})\,\tilde{\mathsf{H}}^{-1}(\boldsymbol{\theta}), \tag{5.9}$$

$$\begin{aligned}\tilde{\mathsf{G}}(\boldsymbol{\theta}) \triangleq \lambda_{\mathsf{s}} \mathsf{E}\Big[&\left(\nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}(\boldsymbol{r};\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}} \log \tilde{\lambda}(\boldsymbol{\theta})\right) \\ &\left(\nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}(\boldsymbol{r};\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}} \log \tilde{\lambda}(\boldsymbol{\theta})\right)^T\Big],\end{aligned} \tag{5.10}$$

$$\tilde{\mathsf{H}}(\boldsymbol{\theta}) \triangleq -\lambda_{\mathsf{s}} \nabla^2_{\boldsymbol{\theta}} \log \tilde{\lambda}(\boldsymbol{\theta}) + \nabla^2_{\boldsymbol{\theta}} \tilde{\lambda}(\boldsymbol{\theta}) - \lambda_{\mathsf{s}} \mathsf{E}\left[\nabla^2_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}(\boldsymbol{r};\boldsymbol{\theta})\right], \tag{5.11}$$

and expectations are with respect to the true distribution.

This theorem differs from Theorem 3.2 of [78] because convergence is shown as the scan time τ, rather than the number of measurements, goes to infinity. Our proof in Appendix C.2 avoids dividing the scan into intervals and naturally shows convergence as scan time goes to infinity.

In the absence of model mismatch, $\tilde{\mathsf{G}}(\boldsymbol{\theta})$ and $\tilde{\mathsf{H}}(\boldsymbol{\theta})$ are equal to the time-normalized Fisher information matrix $\mathsf{F}(\boldsymbol{\theta})$ [78], where

$$\mathsf{F}(\boldsymbol{\theta}) \triangleq -\frac{1}{\tau}\mathsf{E}\left[\nabla^2_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}(\tilde{\boldsymbol{r}};\boldsymbol{\theta})\right].$$

In this special case, the covariance in (6.11) simplifies to $\Sigma(\boldsymbol{\theta}) = \mathsf{F}^{-1}(\boldsymbol{\theta})$. Thus, Theorem V.4 reduces to the classic asymptotic normality and asymptotic efficiency of ML estimation [34] in the absence of model mismatch.

5.2.3 Asymptotics for Constrained Estimators

Theorems V.2 and V.4 give results for unconstrained estimates, but it is often desirable to enforce constraints on the estimates in practice. In the case of gamma–ray imaging, the source intensity parameter is physically nonnegative, and it may be desirable to enforce this constraint by computing a constrained QMLE. In this section, we extend Theorems V.2 and V.4 to the case where one or more of the parameters in the parameter vector $\boldsymbol{\theta}$ are constrained and the unconstrained asymptotic limit $\tilde{\boldsymbol{\mu}}$ lies in or on the boundary of the constraint set. When the unconstrained asymptotic limit $\tilde{\boldsymbol{\mu}}$ lies outside the closure of the constraint set, the sequence of constrained parameter estimates converges in distribution to a degenerate random variable, and we use a quadratic approximation of the log–likelihood to approximate the distribution of parameter estimates for finite scan times.

Let $\Omega \subset \Theta$ be the constrained parameter set, which is the set in which the constrained QMLE must lie. The constrained QMLE is defined by

$$\tilde{\boldsymbol{\theta}}_{\tau}^{+} \stackrel{\triangle}{=} \arg\max_{\boldsymbol{\theta} \in \Omega} \; \log \tilde{\mathsf{p}}\left(\tilde{\boldsymbol{r}}; \boldsymbol{\theta}\right). \tag{5.12}$$

Also let the constrained asymptotic limit $\tilde{\boldsymbol{\mu}}^{+}$ be

$$\tilde{\boldsymbol{\mu}}^{+} \stackrel{\triangle}{=} \arg\max_{\boldsymbol{\theta} \in \Omega} \; \tilde{f}\left(\boldsymbol{\theta}\right). \tag{5.13}$$

Theorem V.5 shows that the constrained QMLE $\tilde{\boldsymbol{\theta}}_{\tau}^{+}$ converges to the asymptotic limit $\tilde{\boldsymbol{\mu}}^{+}$. The proof is given in Appendix C.3.

Theorem V.5. *Under suitable regularity conditions,* $\tilde{\boldsymbol{\theta}}_{\tau}^{+} \stackrel{p}{\rightarrow} \tilde{\boldsymbol{\mu}}^{+}$ *as* $\tau \rightarrow \infty$.

The constrained asymptotic limit $\tilde{\boldsymbol{\mu}}^+$ can lie in the interior or on the boundary of the constrained parameter set. The unconstrained asymptotic mean $\tilde{\boldsymbol{\mu}}$ can either lie in the interior of the constrained parameter set Ω, on the boundary, or outside the constrained parameter set. The asymptotic distribution of the constrained parameter estimates $\tilde{\boldsymbol{\theta}}_\tau^+$ depends on which of these three cases applies to the unconstrained asymptotic limit $\tilde{\boldsymbol{\mu}}$. When the asymptotic limit $\tilde{\boldsymbol{\mu}}^+$ is in the interior of the constrained parameter set Ω, the unconstrained asymptotic limit $\tilde{\boldsymbol{\mu}}$ is equal to the constrained asymptotic limit $\tilde{\boldsymbol{\mu}}^+$, and asymptotic distribution of the parameter estimates is the same as in the unconstrained case.

Figure 5.1 shows an example of an expected log–likelihood $\tilde{f}\,(\theta)$, defined in (6.7), for a scalar parameter. In this example the parameter θ is constrained to be nonnegative. The unconstrained asymptotic limit $\tilde{\mu}$ is defined as the global maximizer of $\tilde{f}\,(\theta)$ in (6.8) and is positive in this case. The constrained asymptotic limit $\tilde{\mu}^+$, defined in (5.13), is the maximizer over the nonnegative real line, which is equal to the global maximizer. The asymptotic distribution depends on the behavior of the log–likelihood near the constrained asymptotic limit $\tilde{\mu}^+$, which is not affected by imposing the nonnegativity constraint.

Figure 5.2 shows an example where the constrained and unconstrained maximizers of the expected log–likelihood are equal and on the boundary of the constrained parameter set at zero. The nonnegativity constraint affects the log–likelihood in any neighborhood around the maximizer, and this changes the asymptotic distribution. We consider this case in the next section.

The third case occurs when the global maximizer of the expected log–likelihood is outside the constrained parameter set. Figure 5.3 shows an example scalar expected log–likelihood where the global maximizer is negative and the maximizer on

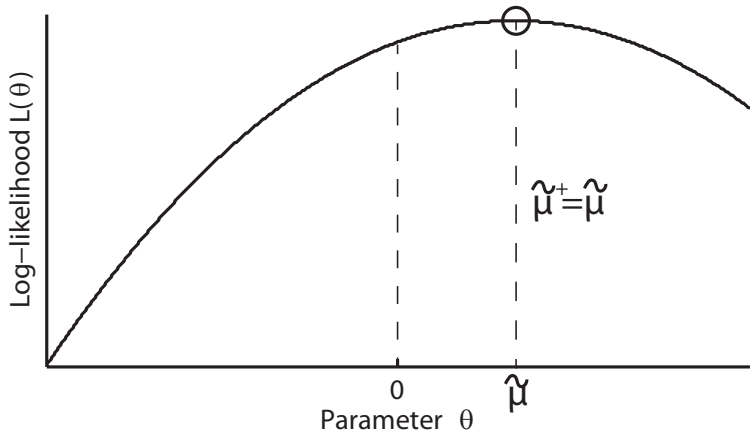


Figure 5.1: An example expected log–likelihood where the global maximizer is equal to the maximizer on the nonnegative real line.

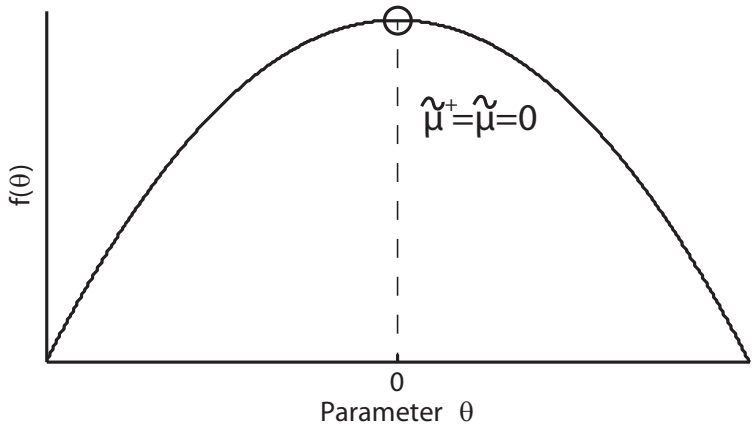


Figure 5.2: An example expected log–likelihood where the asymptotic limit is on the boundary of the constrained parameter set, which is the nonnegative real line.

the nonnegative real line is zero. In the unconstrained case, we used an asymptotic convergence in distribution result to approximate the distribution of estimates for finite scan times. The gradient of the expected log–likelihood is nonzero at the constrained asymptotic limit $\tilde{\mu}^+$, which causes the estimates to converge in distribution to a degenerate random variable. Because this asymptotic distribution is not useful, we propose a finite–sample approximation based on a quadratic approximation of the log–likelihood.

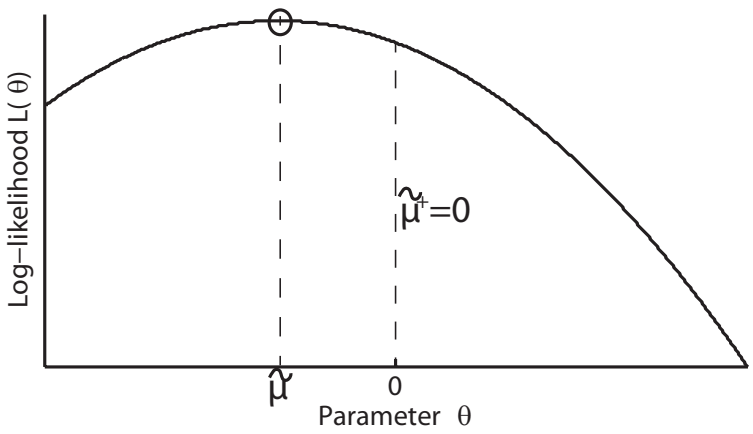


Figure 5.3: An example expected log–likelihood where the unconstrained maximizer is outside the constrained parameter set and the maximizer on the nonnegative real line is zero

Unconstrained asymptotic limit on boundary of constrained parameter set

We first consider the case where the unconstrained asymptotic limit $\tilde{\boldsymbol{\mu}}$ lies on the boundary of the constrained parameter set Ω, which implies that $\tilde{g}\left(\tilde{\boldsymbol{\mu}}^{+}\right) = 0$. In the presence of model mismatch, it may be rare that the unconstrained asymptotic limit will fall exactly on the boundary of the constrained parameter set. An example demonstrating why this is rare is included in Appendix C.5. We include the asymptotic distribution in this case although it may not be a common one. The sequence of parameter estimates $\tilde{\boldsymbol{\theta}}_{\tau}^{+}$ converges in distribution by an application of the results of [2] to the case of a Poisson number of measurements.

To prove asymptotic normality, we follow [2, 69] and approximate the feasible set Ω with a cone. Let C_{Ω} be the cone approximation of Ω, defined by

Definition V.6. The cone C_{Ω} with vertex $\tilde{\boldsymbol{\mu}}^{+}$ is a cone approximation for set Ω at $\tilde{\boldsymbol{\mu}}^{+}$ if

1. $\inf_{\boldsymbol{x} \in C_{\Omega}} ||\boldsymbol{x} - \boldsymbol{y}_n|| = o\left(||\boldsymbol{y}_n - \tilde{\boldsymbol{\mu}}^{+}||\right)$ for all sequences $\{\boldsymbol{y}_n\} \subset \Omega$ such that $\lim_{n \to \infty} \boldsymbol{y}_n = \tilde{\boldsymbol{\mu}}^{+}$.

2. $\inf_{\boldsymbol{y}\in\Omega} ||\boldsymbol{x}_n - \boldsymbol{y}|| = o(||\boldsymbol{x}_n - \tilde{\boldsymbol{\mu}}^+||)$ for all sequences $\{\boldsymbol{x}_n\} \subset C_\Omega$ such that $\lim_{n\to\infty} \boldsymbol{x}_n = \tilde{\boldsymbol{\mu}}^+$.

Figure 5.4 illustrates a 2D example of a nonnegatively constrained parameter set and its cone approximation. Figure 5.5 shows two sequences: one that satisfies

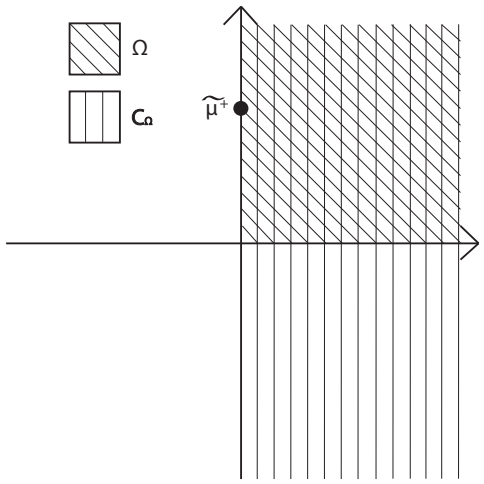


Figure 5.4: An example illustration of a constrained parameter set Ω and its cone approximation C_Ω.

condition 1 of Definition V.6 and one that does not. Condition 2 is satisfied if $\Omega \subset \Theta$.

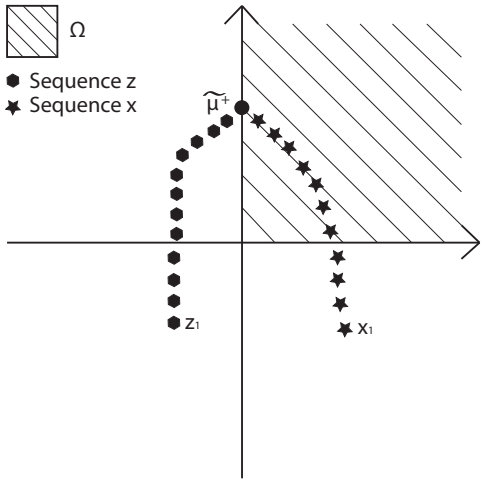


Figure 5.5: An example illustration of sequences that do and do not satisfy condition 1 of Definition V.6.

Theorem V.7 shows that the QMLE converges in distribution to a truncated Gaussian distribution. The proof is based on [2] and [69], which give similar results for a nonrandom number of measurements, and is included in Appendix C.4. Our proof essentially shows convergence in distribution of the normalized gradient of the log–likelihood, convergence in probability of the normalized Hessian of the log–likelihood, and that these convergence results imply that the result follows by Theorem 3 of [2].

Theorem V.7. *Let the random vector* $\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \Sigma(\tilde{\boldsymbol{\mu}}^+))$*, where* $\Sigma(\boldsymbol{\theta})$ *is defined in (6.11).*

Let F *denote the distribution of* $\tilde{w}$*, where*

$$\tilde{w} \stackrel{\triangle}{=} \arg \min_{w \in C_\Omega - \tilde{\boldsymbol{\mu}}^+} ||\boldsymbol{\zeta} - w||_{\tilde{\mathsf{H}}(\tilde{\boldsymbol{\mu}}^+)}.$$

If $\tilde{g}(\tilde{\boldsymbol{\mu}}^+) = \mathbf{0}$ *and suitable regularity conditions are satisfied, then*

$$\sqrt{\tau}\left(\tilde{\boldsymbol{\theta}}_\tau^+ - \tilde{\boldsymbol{\mu}}^+\right) \stackrel{d}{\to} F \text{ as } \tau \to \infty.$$

The result of Theorem V.7 is different from the results of [2] and [69] because Theorem V.7 allows for a *random* number of measurements.

Figure 5.6 shows an example of a nonnegativity constraint set Ω in one dimension, and Figure 5.7 shows the set $\Omega - \mu$, for some real number μ.

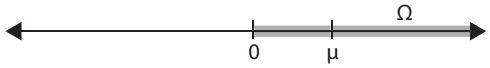


Figure 5.6: An illustration of the set Ω, which is the nonnegative real line.

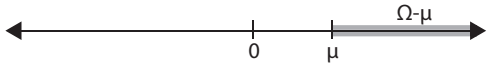


Figure 5.7: An illustration of the nonnegative real line shifted by the real number μ.

Theorem V.7 shows that the distribution of the constrained QMLE is a projection of the asymptotic normal distribution of the unconstrained QMLE onto the

constrained parameter set Ω. In the case of a single constrained parameter, the distribution of the constrained QMLE $\tilde{\boldsymbol{\theta}}_\tau^+$ is a simple truncation of the unconstrained asymptotic distribution given by Theorem V.4. However, when multiple parameters are constrained, the distribution is more difficult to compute.

Asymptotic limit outside constrained parameter set

In gamma–ray imaging, the unconstrained asymptotic limit $\tilde{\boldsymbol{\mu}}$ may have a negative element and be outside the nonnegative orthant. In this case, $\tilde{g}\left(\tilde{\boldsymbol{\mu}}^+\right)$ may be nonzero and the conditions for Theorem V.7 are not satisfied. However, we follow [2,69] and propose the following local quadratic approximation of the log–likelihood at the *constrained* asymptotic limit $\tilde{\boldsymbol{\mu}}^+$:

$$\begin{aligned}
&\frac{2}{\tau}\log\tilde{\mathsf{p}}\left(\tilde{\boldsymbol{r}};\tilde{\boldsymbol{\theta}}_\tau^+\right) \approx \mathsf{L}_{\mathsf{Q}}\left(\tilde{\boldsymbol{\theta}}_\tau^+\right) \\
&\quad\stackrel{\triangle}{=}\ \frac{2}{\tau}\log\tilde{\mathsf{p}}\left(\tilde{\boldsymbol{r}};\tilde{\boldsymbol{\mu}}^+\right) + \frac{2}{\tau}\nabla_{\boldsymbol{\theta}}^T\log\tilde{\mathsf{p}}\left(\tilde{\boldsymbol{r}};\tilde{\boldsymbol{\mu}}^+\right)\left(\tilde{\boldsymbol{\theta}}_\tau^+ - \tilde{\boldsymbol{\mu}}^+\right) \\
&\qquad + \frac{1}{\tau}\left(\tilde{\boldsymbol{\theta}}_\tau^+ - \tilde{\boldsymbol{\mu}}^+\right)^T\nabla_{\boldsymbol{\theta}}^2\log\tilde{\mathsf{p}}\left(\tilde{\boldsymbol{r}};\tilde{\boldsymbol{\mu}}^+\right)\left(\tilde{\boldsymbol{\theta}}_\tau^+ - \tilde{\boldsymbol{\mu}}^+\right). \qquad (5.14)
\end{aligned}$$

Let $\boldsymbol{z}\left(\tilde{\boldsymbol{r}};\tilde{\boldsymbol{\mu}}^+\right)$ be the normalized score function given by

$$\boldsymbol{z}\left(\tilde{\boldsymbol{r}};\tilde{\boldsymbol{\mu}}^+\right) \stackrel{\triangle}{=} \frac{1}{\tau}\tilde{\mathsf{H}}^{-1}\left(\tilde{\boldsymbol{\mu}}^+\right)\nabla_{\boldsymbol{\theta}}\log\tilde{\mathsf{p}}\left(\tilde{\boldsymbol{r}};\boldsymbol{\theta}\right)\big|_{\tilde{\boldsymbol{\mu}}^+}, \qquad (5.15)$$

where the translated parameter estimate is given by

$$w(\tilde{\boldsymbol{\theta}}_\tau,\tilde{\boldsymbol{\mu}}^+) \stackrel{\triangle}{=} \left(\tilde{\boldsymbol{\theta}}_\tau^+ - \tilde{\boldsymbol{\mu}}^+\right). \qquad (5.16)$$

The maximizer of the likelihood in (5.14) is equivalent to the minimizer of

$$\begin{aligned}
\mathsf{L}_{\mathsf{Q}}\left(\tilde{\boldsymbol{\theta}}_\tau^+\right) &\equiv -\boldsymbol{z}^T\left(\tilde{\boldsymbol{r}};\tilde{\boldsymbol{\mu}}^+\right)\tilde{\mathsf{H}}\left(\tilde{\boldsymbol{\mu}}^+\right)\boldsymbol{z}\left(\tilde{\boldsymbol{r}};\tilde{\boldsymbol{\mu}}^+\right) + 2\boldsymbol{z}^T\left(\tilde{\boldsymbol{r}};\tilde{\boldsymbol{\mu}}^+\right)\tilde{\mathsf{H}}\left(\tilde{\boldsymbol{\mu}}^+\right)\left(\tilde{\boldsymbol{\theta}}_\tau^+ - \tilde{\boldsymbol{\mu}}^+\right) \\
&\quad + \left(\tilde{\boldsymbol{\theta}}_\tau^+ - \tilde{\boldsymbol{\mu}}^+\right)^T\frac{1}{\tau}\nabla_{\boldsymbol{\theta}}^2\log\tilde{\mathsf{p}}\left(\tilde{\boldsymbol{r}};\tilde{\boldsymbol{\mu}}^+\right)\left(\tilde{\boldsymbol{\theta}}_\tau^+ - \tilde{\boldsymbol{\mu}}^+\right), \\
&\stackrel{p}{\to} -\left(\boldsymbol{z}\left(\tilde{\boldsymbol{r}};\tilde{\boldsymbol{\mu}}^+\right) - \left(\tilde{\boldsymbol{\theta}}_\tau^+ - \tilde{\boldsymbol{\mu}}^+\right)\right)^T\tilde{\mathsf{H}}\left(\tilde{\boldsymbol{\mu}}^+\right)\left(\boldsymbol{z}\left(\tilde{\boldsymbol{r}};\tilde{\boldsymbol{\mu}}^+\right) - \left(\tilde{\boldsymbol{\theta}}_\tau^+ - \tilde{\boldsymbol{\mu}}^+\right)\right) \\
&\equiv -\left(\boldsymbol{z}\left(\tilde{\boldsymbol{r}};\tilde{\boldsymbol{\mu}}^+\right) - w(\tilde{\boldsymbol{\theta}}_\tau,\tilde{\boldsymbol{\mu}}^+)\right)^T\tilde{\mathsf{H}}\left(\tilde{\boldsymbol{\mu}}^+\right)\left(\boldsymbol{z}\left(\tilde{\boldsymbol{r}};\tilde{\boldsymbol{\mu}}^+\right) - w(\tilde{\boldsymbol{\theta}}_\tau,\tilde{\boldsymbol{\mu}}^+)\right). \qquad (5.17)
\end{aligned}$$

By Lemma C.27, we can approximate the distribution of $\boldsymbol{z}\left(\tilde{\boldsymbol{r}};\tilde{\boldsymbol{\mu}}^{+}\right)$ by

(5.18) $\boldsymbol{z}\left(\tilde{\boldsymbol{r}};\tilde{\boldsymbol{\mu}}^{+}\right)$ is approximately distributed as $\mathcal{N}\left(\tilde{\mathsf{H}}^{-1}\left(\tilde{\boldsymbol{\mu}}^{+}\right)\tilde{g}\left(\tilde{\boldsymbol{\mu}}^{+}\right),\frac{1}{\tau}\Sigma(\tilde{\boldsymbol{\mu}}^{+})\right)$.

By (5.17), w is equal to a QMLE for the mean of $\boldsymbol{z}\left(\tilde{\boldsymbol{r}};\tilde{\boldsymbol{\mu}}^{+}\right)$ with a possibly incorrect covariance matrix. The formula for the QMLE $\tilde{w}$ for $w(\tilde{\boldsymbol{\theta}}_{\tau},\tilde{\boldsymbol{\mu}}^{+})$ is given by

$$\tilde{w}=\arg\min_{w\in C_{\Omega}-\tilde{\boldsymbol{\mu}}^{+}}\left(\boldsymbol{z}\left(\tilde{\boldsymbol{r}};\tilde{\boldsymbol{\mu}}^{+}\right)-w\right)^{T}\tilde{\mathsf{H}}\left(\tilde{\boldsymbol{\mu}}^{+}\right)\left(\boldsymbol{z}\left(\tilde{\boldsymbol{r}};\tilde{\boldsymbol{\mu}}^{+}\right)-w\right). \tag{5.19}$$

Note that the fact that C_{Ω} is a cone and the constraint $w\in C_{\Omega}-\tilde{\boldsymbol{\mu}}^{+}$ implies that $\tilde{\boldsymbol{\theta}}_{\tau}^{+}\in C_{\Omega}$.

The distribution of $\tilde{w}=\tilde{\boldsymbol{\theta}}_{\tau}^{+}-\tilde{\boldsymbol{\mu}}^{+}$ in (5.19) is equal to the distribution of $\boldsymbol{z}\left(\tilde{\boldsymbol{r}};\tilde{\boldsymbol{\mu}}^{+}\right)$ in (5.18) projected onto C_{Ω}, so it is reasonable to approximate the distribution of $\tilde{\boldsymbol{\theta}}_{\tau}^{+}-\tilde{\boldsymbol{\mu}}^{+}$ with the distribution of $\tilde{w}$.

5.2.4 Using Asymptotic Distributions for Approximation

It may be challenging to verify the regularity conditions for Theorems 1 and 2 listed in the Appendix. Our numerical results show that Theorems V.2 and V.4 are useful as practical approximations for gamma–ray detection problems. The conditions listed in the Appendix are sufficient, but not necessary. A discussion of how one could verify the regularity conditions is given in Appendix C.7.

The asymptotic mean $\tilde{\boldsymbol{\mu}}$ in (6.8) and covariance $\Sigma(\tilde{\boldsymbol{\mu}})$ in (5.8) both depend on the true attribute distribution $\mathsf{p}\left(\boldsymbol{r}\right)$. We now present a simple example of an estimation problem with model mismatch.

Example: Simple Poisson Intensity Estimation

Assume that a system records a random number of interacting photons J, where we model the number of recorded photons as a Poisson random variable whose mean

$\tilde{J}(\boldsymbol{\theta}) = \tau(\alpha + \lambda_b)$ is the sum of source and background photons:

$$\tilde{\mathsf{p}}\left(J;\alpha\right) = e^{-\tau(\alpha+\lambda_b)}\left(\tau(\alpha+\lambda_b)\right)^J/J!,$$

where $\alpha \in \Omega \stackrel{\triangle}{=} [0,\infty)$ is the source intensity and λ_b is the background intensity, which is assumed known. The model is equivalent to the list–mode model in (6.3) with the parameter vector $\boldsymbol{\theta} = \alpha$, count rate

$$\tilde{\lambda}(\boldsymbol{\theta}) = \alpha + \lambda_b, \tag{5.20}$$

and empty recorded attributes $\boldsymbol{r}$. The recorded attributes are empty because this system records no information beyond the number of interacting photons. Let ϵ represent the empty attribute. Then for this photon–counting system,

$$\tilde{\mathsf{p}}\left(\boldsymbol{r};\alpha\right) = \begin{cases} 1 & \boldsymbol{r} = \epsilon \\ & \text{for all } \alpha \in [0,\infty), \\ 0 & \text{else} \end{cases}$$

which is constant with respect to the source intensity parameter α. We also assume that the true distribution is also poisson with mean $\tau\lambda_{\mathsf{s}}$:

$$\mathsf{p}\left(J\right) = e^{-\tau\lambda_{\mathsf{s}}}\left(\tau\lambda_{\mathsf{s}}\right)^J/J!.$$

The true density of J is equivalent to the list–mode density in (5.1) with attribute vectors that are almost surely empty.

The log–likelihood, neglecting terms constant with respect to α is

$$\mathsf{L}\left(\alpha\right) = -\tau(\alpha+\lambda_b) + J\log(\alpha+\lambda_b), \tag{5.21}$$

with first derivative

$$\frac{d}{d\alpha}\mathsf{L}\left(\alpha\right) = \frac{J}{\alpha+\lambda_b} - \tau. \tag{5.22}$$

Solving (5.22) with respect to the source intensity parameter α gives the constrained QMLE for source intensity $\tilde{\alpha}_\tau^+$:

$$\tilde{\alpha}_\tau^+ = \left[\frac{J}{\tau} - \lambda_b\right]_+, \tag{5.23}$$

where $[x]_+$ is the larger of x and 0 for a real number x. The expected value of the gradient of the log–likelihood with respect to the true distribution is

$$\tilde{g}(\alpha) = \mathsf{E}\left[\frac{d}{d\alpha}\mathsf{L}(\alpha)\right] = \frac{\tau\lambda_s}{\alpha + \lambda_b} - \tau,$$

which has a unique zero at the unconstrained asymptotic limit $\tilde{\mu}$, where

$$\tilde{\mu} = \lambda_s - \lambda_b.$$

The constrained asymptotic limit is the maximizer over the constrained parameter set $[0, \infty)$ is

$$\tilde{\mu}^+ = [\lambda_s - \lambda_b]_+ . \tag{5.24}$$

By (6.12) and (5.20), the first derivative matrix for this problem is

$$\tilde{\mathsf{G}}(\alpha) = \frac{\lambda_s}{(\alpha + \lambda_b)^2}.$$

Similarly, by (6.13) and (5.20),

$$\tilde{\mathsf{H}}(\alpha) = \tilde{\mathsf{G}}(\alpha) = \frac{\lambda_s}{(\alpha + \lambda_b)^2}.$$

By (6.11), the asymptotic covariance is

$$\Sigma(\alpha) = \tilde{\mathsf{H}}^{-1}(\alpha)\,\tilde{\mathsf{G}}(\alpha)\,\tilde{\mathsf{H}}^{-1}(\alpha) = \frac{(\alpha + \lambda_b)^2}{\lambda_s}. \tag{5.25}$$

Let the true recorded count rate $\lambda_s = 0.9$ and let the modeled background count rate $\lambda_b = 1$. In this case, the unconstrained asymptotic mean $\tilde{\mu} = -0.1$ which is

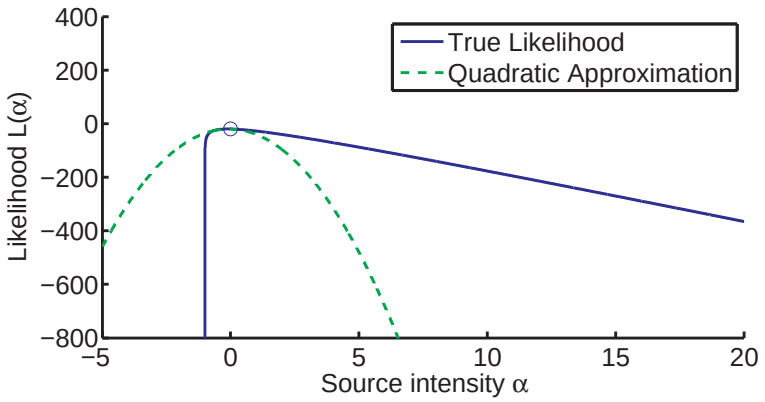


Figure 5.8: Poisson log–likelihood and second–order Taylor series expansion about $\tilde{\mu}^+ = 0$ for scan time $\tau = 20$, true source intensity $\lambda_s = 0.9$, and modeled background intensity $\lambda_b = 1$

outside the constrained parameter set $[0, \infty)$. To approximate the distribution of $\tilde{\alpha}_\tau^+$, we use (5.18) and (5.19) to obtain the approximation

(5.26) $\qquad \tilde{\alpha}_\tau^+$ is approximately distributed as $[z]_+$,

where the random variable z is has the law

$$z \sim \mathcal{N}\left(\tilde{\mathsf{H}}^{-1}\left(\tilde{\mu}^+\right)\tilde{g}\left(\tilde{\mu}^+\right), \frac{1}{\tau}\Sigma(\tilde{\mu}^+)\right),$$

and $\tilde{\mu}^+ = 0$ by (5.24).

The derivation of (5.19) is based on a quadratic approximation that follows the derivation in [2]. Figure 5.8 shows the true log–likelihood in (5.21) and its second–order Taylor series expansion about the constrained asymptotic limit $\tilde{\mu}^+ = 0$ for scan time $\tau = 20$, true source intensity $\lambda_s = 0.9$, and modeled background intensity $\lambda_b = 1$. Asymptotics are based on quadratic approximations of the log–likelihood, and their accuracy depends on how well the true likelihood is approximated by a quadratic function in the neighborhood of the constrained asymptotic limit $\tilde{\mu}^+$. The Poisson log–likelihood in Figure 5.21 has extremely high curvature near $\alpha = -\lambda_b = -1$. This large curvature makes quadratic approximations to the likelihood inaccurate

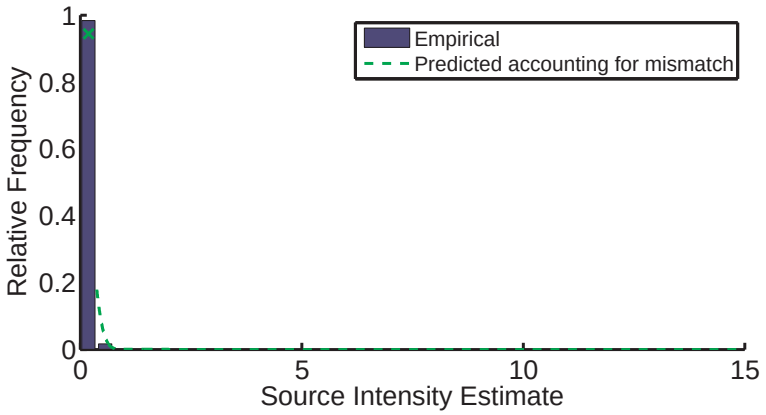


Figure 5.9: Histogram of 10000 empirical estimates and the distribution predicted by asymptotics for scan time $\tau = 20$, true source intensity $\lambda_s = 0.9$, and modeled background intensity $\lambda_b = 1$

when the modeled count rate $\tilde{\lambda}(\boldsymbol{\theta})$ is small. Increasing the scan time will improve the asymptotic approximation because as the scan time increases, the estimates are more likely to be close to $\tilde{\mu}^+$ where the quadratic approximation is more accurate.

Figure 5.9 shows a histogram of 10000 empirical estimates of the source intensity $\tilde{\alpha}_\tau^+$ and the distribution predicted by asymptotics in (5.26) for scan time $\tau = 20$, true source intensity $\lambda_s = 0.9$, and modeled background intensity $\lambda_b = 1$. Although the quadratic approximation to the likelihood is poor in Figure 5.8, the scan time $\tau = 20$ is large enough to obtain an accurate approximation. The empirical and predicted distributions agree reasonably well.

We now consider the case where the modeled background intensity $\lambda_b = 1$ and the true count rate $\lambda_s = 11$. In this case,

$$\tilde{\mu}^+ = \tilde{\mu} = 10, \tag{5.27}$$

which is inside the constrained parameter set $[0, \infty)$. Since the constrained aysmptotic limit $\tilde{\mu}^+$ is in the interior of the parameter space, we use Theorem V.4 to obtain the approximation

$$\tilde{\alpha}_\tau^+ \sim \mathcal{N}\left(\tilde{\mu}^+, \frac{1}{\tau}\Sigma(\tilde{\mu}^+)\right).$$

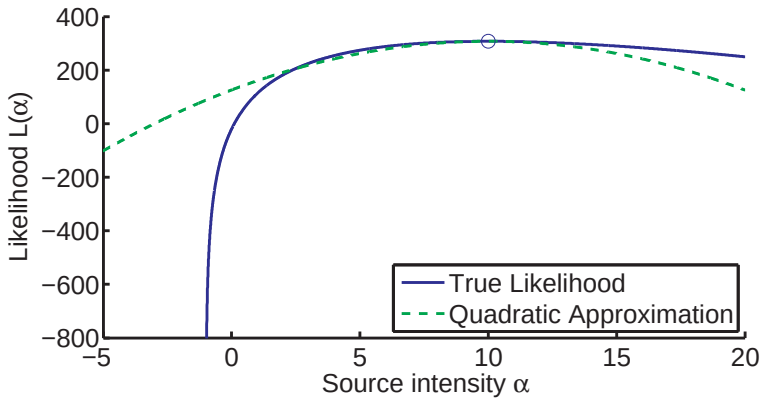


Figure 5.10: Poisson log–likelihood and second–order Taylor series expansion about $\tilde{\mu}^{+} = 10$ for scan time $\tau = 20$, true source intensity $\lambda_{\mathsf{s}} = 11$, and modeled background intensity $\lambda_b = 1$

Figure 5.10 shows the Poisson log–likelihood and the second order Taylor series expansion at $\tilde{\mu}^{+} = 10$. The curvature of the log–likelihood is much lower than in the case where $\tilde{\lambda}(\boldsymbol{\theta}) = 0.9$, so the quadratic approximation is more accurate for a larger range of source intensity parameter values. The low curvature and good approximation will make the asymptotic approximation more accurate for smaller scan times than in the casw where $\tilde{\lambda}(\boldsymbol{\theta}) = 0.9$. The lower curvature of the log–likelihood for higher values of the source intensity estimate means that the asymptotic approximations should be more accurate if the background intensity is assumed known and underestimated. However, this will cause model mismatch that could degrade detection performance.

Figure 5.11 shows a histogram of empirical source intensity estimates and the distribution predicted by asymptotics. The distribution predicted by asymptotics matches the empirical distribution very well.

We assumed in this example that the true distribution was known. In some gamma–ray imaging problems, the exact form of $\mathsf{p}(\boldsymbol{r})$ is computationally intractable, but it is relatively easy to compute expectations (6.12) and (6.13) with respect to $\mathsf{p}(\boldsymbol{r})$ by Monte Carlo methods [53]. The next section uses Theorems V.2 and V.4

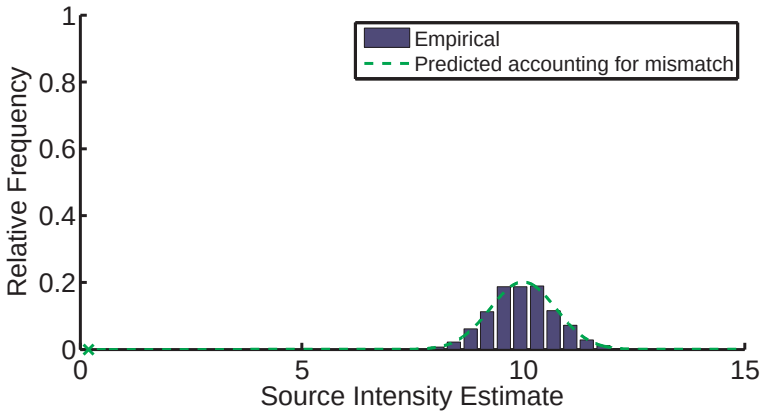


Figure 5.11: Histogram of 10000 empirical estimates and the distribution predicted by asymptotics for scan time $\tau = 20$, true source intensity $\lambda_s = 11$, and modeled background intensity $\lambda_b = 1$

to develop specialized expressions for the asymptotic distribution of test statistics applied to gamma–ray source detection problems.

5.3 Asymptotics for Gamma–Ray Source Detection

The true and modeled distributions in Section 5.1 are more general than required for gamma–ray source detection problems. This section develops a model and a true distribution for the specific task of source detection, and uses the model to develop asymptotic distributions of useful test statistics for detecting a source in background.

5.3.1 True Distribution

In §5.1, we made no assumption about the form of the true distribution of recorded attributes $\mathsf{p}(\boldsymbol{r})$. In gamma–ray source detection problems, the measurements are often generated by a mixture of source and background emission processes. We assume that the true distribution follows the mixture

$$\mathsf{p}(\boldsymbol{r}) = \frac{\alpha^{\mathsf{t}} s^{\mathsf{t}} \mathsf{p}_{\mathrm{S}}(\boldsymbol{r}) + \lambda_b^{\mathsf{t}} \mathsf{p}_{\mathrm{B}}(\boldsymbol{r})}{\alpha^{\mathsf{t}} s^{\mathsf{t}} + \lambda_b^{\mathsf{t}}}, \tag{5.28}$$

where $\mathsf{p}_{\mathrm{S}}(\boldsymbol{r})$ is the density of recorded attributes given that they originated from the source, $\mathsf{p}_{\mathrm{B}}(\boldsymbol{r})$ is the density of recorded attributes given that they originated from

the background, α^{t} is the nonnegative true mean number of *emitted* source counts per unit time, λ_b^{t} is the nonnegative true mean number of background counts *recorded* per unit time, and s^{t} is the true sensitivity, or probability that a photon emitted from the source is recorded. We adopt a mixture model because the recorded attributes of events that originate from the source have a different distribution than the attributes of recorded events that originate from the background. One can derive (5.28) using the law of total probability. The nonnegative sensitivity s^{t} is a function of the spatial relationship between the source and detector. The quantities α^{t} and λ_b^{t} are usually not known in practice, but are needed for simulation. One can estimate s^{t} by simulation for a particular detector geometry, source position, and energy spectrum. The total mean number of received counts obeys the relation $\lambda_{\mathsf{s}} = \alpha^{\mathsf{t}} s^{\mathsf{t}} + \lambda_b^{\mathsf{t}}$. The density $\mathsf{p}(x)$ is a probability density function by the nonnegativity of α^{t}, s^{t}, and λ_b^{t}.

We formulate source detection as the hypothesis testing problem:

$$\begin{aligned} H_1 &: \alpha^{\mathsf{t}} > 0 \\ H_0 &: \alpha^{\mathsf{t}} = 0. \end{aligned}$$

5.3.2 Model Distribution

We consider a modeled attribute distribution similar to that of [48] for detecting a source in background. The parameters characterizing the source are the intensity α with units of counts *emitted* per unit time and source position[3] $\boldsymbol{\phi} \in \Phi$. In the 3D far–field with a known source energy, the set Φ could be $[0, 2\pi] \times [0, \pi]$, representing all possible azimuth and polar angles on a sphere. We parameterize the background intensity by the background count rate λ_b with units of gamma–ray counts *recorded* per unit time. Let $\boldsymbol{\theta}$ be the vector of all parameters, where $\boldsymbol{\theta}$ lies in the d-dimensional

[3] $\boldsymbol{\phi}$ could also denote a vector containing both spatial position and energy

parameter space Θ. In what follows, we assume that $\boldsymbol{\theta}$ takes the form:

$$\boldsymbol{\theta} = (\alpha, \boldsymbol{\phi}). \tag{5.29}$$

Let the modeled sensitivity $\tilde{s}(\boldsymbol{\phi})$ approximate the probability that a photon emitted from a source positioned at $\boldsymbol{\phi}$ is recorded. We model the rate of recorded photons by

$$\tilde{\lambda}(\boldsymbol{\theta}) \stackrel{\triangle}{=} \lambda_b + \alpha\tilde{s}(\boldsymbol{\phi}). \tag{5.30}$$

We adopt the model in (6.2) because the total rate of recorded photons is the sum of the recorded count rate due to background λ_b and the recorded count rate due to the source $\alpha\tilde{s}(\boldsymbol{\phi})$.

Let $\tilde{\mathsf{p}}_{\mathrm{S}}(\boldsymbol{r};\boldsymbol{\phi})$ denote the modeled distribution of a recorded attribute vector x given it originated from a source at positon $\boldsymbol{\phi}$ and let $\tilde{\mathsf{p}}_{\mathrm{B}}(\boldsymbol{r})$ denote the modeled distribution of a recorded attribute vector x given that it originated from the background. Note that $\tilde{\mathsf{p}}_{\mathrm{S}}(\boldsymbol{r};\boldsymbol{\phi})$ depends only on the source position, and $\tilde{\mathsf{p}}_{\mathrm{B}}(\boldsymbol{r})$ does not depend on any of the parameters in (6.1).

We model the overall distribution of recorded attributes as a mixture of $\tilde{\mathsf{p}}_{\mathrm{S}}(\boldsymbol{r};\boldsymbol{\phi})$ and $\tilde{\mathsf{p}}_{\mathrm{B}}(\boldsymbol{r})$ given by

$$\tilde{\mathsf{p}}\left(\boldsymbol{r};\boldsymbol{\theta}\right) = \frac{\alpha\tilde{s}(\boldsymbol{\phi})\tilde{\mathsf{p}}_{\mathrm{S}}(\boldsymbol{r};\boldsymbol{\phi}) + \lambda_b\tilde{\mathsf{p}}_{\mathrm{B}}(\boldsymbol{r})}{\alpha\tilde{s}(\boldsymbol{\phi}) + \lambda_b}. \tag{5.31}$$

As in (5.28), $\tilde{\mathsf{p}}\left(\boldsymbol{r};\boldsymbol{\theta}\right)$ in (6.4) is a valid probability density function when α, $\tilde{s}(\boldsymbol{\phi})$, and λ_b are nonnegative.

5.3.3 Asymptotic Performance of Source Intensity Test Under Model Mismatch

The source intensity test (SIT) [48] for detecting the presence of a radiation source of unknown intensity α is given by

$$\tilde{\alpha}_\tau \underset{H_0}{\overset{H_1}{\gtrless}} \gamma, \tag{5.32}$$

where $\tilde{\alpha}_\tau$ is the QMLE for α. By Theorem V.4, assuming that the regularity conditions are satisfied,

$$\sqrt{\tau}\left(\tilde{\alpha}_\tau - \tilde{\mu}_{[1]}\right) \xrightarrow{d} \mathcal{N}\left(0, \Sigma(\tilde{\boldsymbol{\mu}})_{[1,1]}\right), \tag{5.33}$$

as $\tau \to \infty$, where $\Sigma(\boldsymbol{\theta})$ is defined in (6.11), $[1,1]$ denotes the $[1,1]$ component of the matrix, and $\tilde{\mu}_{[1]}$ is the asymptotic mean of $\tilde{\alpha}_\tau$ defined by the first element of $\tilde{\boldsymbol{\mu}}$ in (6.8). Define $\tilde{\boldsymbol{\mu}}^{(1)}$ to be the solution to (6.8) under H_1 and $\tilde{\boldsymbol{\mu}}^{(0)}$ to be the solution to (6.8) under H_0. Thus, $\tilde{\alpha}_\tau$ is a consistent estimator of $\tilde{\mu}_{[1]}$ under H_1 and a consistent estimator of $\tilde{\mu}^{(0)}_{[1]}$ under H_0. Because of model mismatch, $\tilde{\mu}_{[1]}$ may not be equal to the true intensity and $\tilde{\mu}^{(0)}_{[1]}$ may not be zero.

We calculate $\tilde{\mu}^{(1)}_{[1]}$ and $\tilde{\mu}^{(0)}_{[1]}$ by simulating or recording a large number of observations from the true distribution under H_1 and H_0 and solving for the QMLE using (5.4) under both hypotheses. Although obtaining accurate estimates of $\tilde{\mu}^{(1)}_{[1]}$ and $\tilde{\mu}^{(0)}_{[1]}$ requires many observations, we found that more are required to generate an ROC empirically with high accuracy. We calculate $\Sigma(\tilde{\boldsymbol{\mu}}^{(1)})$, and $\Sigma(\tilde{\boldsymbol{\mu}}^{(0)})$ using (6.11) and evaluate the expectations in (6.12) and (6.13) by Monte Carlo integration with data sampled from the true distribution. One can use these values with (5.33) to predict the corresponding asymptotic ROC curve.

5.3.4 Asymptotic Distribution of GLRT Under Model Mismatch

The generalized likelihood ratio test (GLRT) is another common detection method given by [35]

$$\Lambda_{\text{GLRT}} \triangleq \frac{\tilde{\mathsf{p}}(\tilde{\boldsymbol{r}}; \tilde{\boldsymbol{\theta}}_\tau)}{\arg\max_{\boldsymbol{\theta}, \alpha=0} \tilde{\mathsf{p}}(\tilde{\boldsymbol{r}}; \boldsymbol{\theta})} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma. \tag{5.34}$$

By a similar argument to that in [78] for a nonrandom number of measurements, asymptotically as $\tau \to \infty$,

$$2 \log \Lambda_{\mathrm{GLRT}} \approx \tilde{\alpha}_\tau^2 \left(\tilde{\mathsf{H}}^{-1} \left(\tilde{\boldsymbol{\theta}}_\tau \right)_{[1,1]} \right)^{-1}. \tag{5.35}$$

The expression in (5.35) resembles the Wald test [35], which is a variance–normalized statistic of the ML estimates.

Combining Theorem V.4 with (5.35) and performing algebraic manipulations, we have that under model mismatch for the GLRT with a scalar parameter and τ large,

$$2 \log \Lambda_{\mathrm{GLRT}} \sim \begin{cases} a(\tilde{\boldsymbol{\mu}}^{(1)}) \chi_1^2 \left(\eta(\tilde{\boldsymbol{\mu}}^{(1)}) \right), & \text{under } H_1 \\ a(\tilde{\boldsymbol{\mu}}^{(0)}) \chi_1^2 \left(\eta(\tilde{\boldsymbol{\mu}}^{(0)}) \right), & \text{under } H_0, \end{cases} \tag{5.36}$$

where $\chi_k^2(\eta)$ denotes a noncentral chi–square random variable with noncentrality parameter η and k degrees of freedom [35], and

$$\eta(\boldsymbol{\theta}) \stackrel{\triangle}{=} \tau \alpha^2 \Sigma^{-1}(\boldsymbol{\theta})_{[1,1]}, \tag{5.37}$$

$$a(\boldsymbol{\theta}) \stackrel{\triangle}{=} \left(\Sigma^{-1}(\boldsymbol{\theta})_{[1,1]} \tilde{\mathsf{H}}^{-1}(\boldsymbol{\theta})_{[1,1]} \right)^{-1}. \tag{5.38}$$

The result in (5.36) generalizes the result from [35] that the GLRT is asymptotically noncentral chi–square under H_1 and central chi–square under H_0. The multiplicative factors $a(\tilde{\boldsymbol{\mu}}^{(1)})$ and $a(\tilde{\boldsymbol{\mu}}^{(0)})$ equal unity in the absence of model mismatch.

One can use (5.36) to compute the asymptotic area under the ROC curve (AUC) of the GLRT using Monte Carlo simulation to estimate $\tilde{\boldsymbol{\mu}}^{(1)}$, $\tilde{\boldsymbol{\mu}}^{(0)}$, $\Sigma(\tilde{\boldsymbol{\mu}}^{(1)})$, and $\Sigma(\tilde{\boldsymbol{\mu}}^{(0)})$. Note that $a(\boldsymbol{\theta})$ in (5.38) and $\eta(\boldsymbol{\theta})$ in (5.37) are functions of a generic parameter vector $\boldsymbol{\theta}$. When α appears on the right–hand side of (5.37), it represents the first element of $\boldsymbol{\theta}$, which corresponds to the source intensity parameter α. One can use the distributions in (5.36) to compute the ROC and AUC.

5.3.5 Asymptotic Distributions for Gamma–Ray Imaging

The covariance matrix in (5.8) has a specific form when the recorded attributes are distributed according to the mixture model in (6.4) and the parameter vector has the form (6.1). In this section, we derive expressions for $\tilde{\mathsf{G}}\left(\boldsymbol{\theta}\right)$ and $\tilde{\mathsf{H}}\left(\boldsymbol{\theta}\right)$ in (6.11). Let

$$\tilde{\mathsf{F}}\left(\boldsymbol{\theta}\right) = \tilde{\mathsf{F}}_{\mathrm{c}}(\boldsymbol{\theta}) \begin{bmatrix} K(\boldsymbol{\theta})_{[1,1]} & \frac{\alpha K(\boldsymbol{\theta})_{[2,1]}^{T}}{\tilde{s}(\boldsymbol{\phi})} \\ \frac{\alpha K(\boldsymbol{\theta})_{[2,1]}}{\tilde{s}(\boldsymbol{\phi})} & \frac{\alpha^2 K(\boldsymbol{\theta})_{[2,2]}}{\tilde{s}^2(\boldsymbol{\phi})} \end{bmatrix}, \tag{5.39}$$

where

$$K(\boldsymbol{\theta})_{[i,j]} \stackrel{\triangle}{=} \mathsf{E}\left[\frac{\tilde{q}_i(\boldsymbol{r};\boldsymbol{\theta})\tilde{q}_j^{T}(\boldsymbol{r};\boldsymbol{\theta})}{\tilde{\mathsf{p}}^2(\boldsymbol{r};\boldsymbol{\theta})}\right], \tag{5.40}$$

$$\begin{aligned} \tilde{q}_1(x;\boldsymbol{\theta}) &\stackrel{\triangle}{=} \tilde{\mathsf{p}}_{\mathrm{S}}(\boldsymbol{r};\boldsymbol{\phi}), \\ \tilde{q}_2(x;\boldsymbol{\theta}) &\stackrel{\triangle}{=} \nabla_{\boldsymbol{\phi}}(\tilde{s}(\boldsymbol{\phi})\tilde{\mathsf{p}}_{\mathrm{S}}(\boldsymbol{r};\boldsymbol{\phi})), \end{aligned} \tag{5.41}$$

and

$$\tilde{\mathsf{F}}_{\mathrm{c}}(\boldsymbol{\theta}) \stackrel{\triangle}{=} \frac{\lambda_{\mathsf{s}} s^2(\boldsymbol{\phi})}{(\alpha\tilde{s}(\boldsymbol{\phi}) + \lambda_b)^2} \tag{5.42}$$

is the scalar Fisher information for a counting detector. In the absence of model mismatch, $\tilde{\mathsf{F}}\left(\boldsymbol{\theta}\right)$ would simplify to the time–normalized Fisher information matrix $\mathsf{F}\left(\boldsymbol{\theta}\right)$ for the model (6.3). Let

$$\mathsf{B}(\boldsymbol{\theta}) \stackrel{\triangle}{=} \begin{bmatrix} 0 & 0 \\ 0 & \alpha w(\boldsymbol{\theta}) \end{bmatrix}, \tag{5.43}$$

where

$$w(\boldsymbol{\theta}) \stackrel{\triangle}{=} \nabla_{\boldsymbol{\phi}}^2 \tilde{s}(\boldsymbol{\phi}) \left(\frac{\lambda_{\mathsf{s}}}{\tilde{\lambda}(\boldsymbol{\theta})} - 1\right).$$

The matrix $\mathsf{B}(\boldsymbol{\theta})$ depends on the degree of model mismatch and is zero when $\tilde{\lambda}(\boldsymbol{\theta}) = \lambda_{\mathsf{s}}$. By algebra, one can show that for gamma–ray imaging detectors,

$$\tilde{\mathsf{G}}\left(\boldsymbol{\theta}\right) = \tilde{\mathsf{F}}\left(\boldsymbol{\theta}\right), \tag{5.44}$$

and

$$\tilde{\mathsf{H}}\left(\boldsymbol{\theta}\right) = \tilde{\mathsf{F}}\left(\boldsymbol{\theta}\right) - \mathsf{B}(\boldsymbol{\theta}). \tag{5.45}$$

The expression in (5.45) shows that model mismatch can decrease the Fisher information. This decrease results in increased variance of the parameter estimates $\tilde{\boldsymbol{\theta}}_{\tau}$.

In the next section, we evaluate these expressions at $\boldsymbol{\theta} = \tilde{\boldsymbol{\mu}}$ to explore the detection performance of various detectors.

5.4 Numerical Results

We applied the asymptotic distributions of test statistics derived in the previous section to detection performance prediction for gamma–ray measurement systems. We considered a hypothetical one–dimensional detector and a 3D position–sensitive Compton detector as example applications of the theory and to show that the theory accurately predicts empirical performance. Throughout, we refer to the performance predicted using the asymptotic theory as "predicted performance." We used the SIT (6.9) in our experiments because of its simplicity and superior performance in the absence of model mismatch compared to the GLRT as shown experimentally in [48].

We used the asymptotic distributions to create receiver operating characteristic (ROC) curves that one can use to set the test threshold and to estimate the probability of detection for a given false alarm rate. The ROC curve is a plot of the probability of detection as a function of the probability of false alarm. Each point

on the ROC is a pair of probability of false alarm and probability of detection values resulting from a particular choice of the threshold γ in (6.9) for the SIT or (5.34) for the GLRT. The ROC curve allows one to choose a threshold value γ that results in the desired false alarm rate.

We also state our results in terms of area under the ROC curve (AUC) [28], which is a measure of detector performance. The AUC ranges from 0.5 to 1, where 0.5 is the poorest performance and 1 is the best performance. The AUC is a measure of overall detectability that does not depend on the threshold γ in (6.9) or (5.34).

Throughout this section, we compare predicted performance accounting for and ignoring model mismatch. The purpose of this comparison is to illustrate the pitfalls of applying the asymptotic formulas that do not account for model mismatch in [48] to problems where model mismatch exists.

In the absence of model mismatch, there exists a parameter vector $\boldsymbol{\theta}_{\text{true}}$ such that $\lambda_{\mathsf{s}} = \tilde{\lambda}(\boldsymbol{\theta}_{\text{true}})$ and $\tilde{\mathsf{p}}(\boldsymbol{r}; \boldsymbol{\theta}_{\text{true}}) = \mathsf{p}(\boldsymbol{r})$ for all $\boldsymbol{r} \in \mathcal{R}$. In this case, the theory from [48] states that the maximum likelihood estimator is a consistent estimator of $\boldsymbol{\theta}_{\text{true}}$. When applying the theory in [48] in this section, we used the incorrect assumption that the estimates are normally distributed with mean $\boldsymbol{\theta}_{\text{true}}$ rather than $\tilde{\boldsymbol{\mu}}$. This assumption is incorrect in the presence of model mismatch, and is one factor that contributes to the failure of the theory from [48] in the results that follow. In the absence of model mismatch, $\Sigma(\boldsymbol{\theta}) = \tilde{\mathsf{G}}^{-1}(\boldsymbol{\theta}) = \tilde{\mathsf{H}}^{-1}(\boldsymbol{\theta})$, so we make the substitution $\Sigma(\boldsymbol{\theta}) = \tilde{\mathsf{G}}^{-1}(\boldsymbol{\theta})$ to capture the naïve use of the methods in [48].

To compute the predicted performance, we sampled event attributes from the true distribution. We generated one list of events with a source present and another without a source present. We used these lists to compute the asymptotic limits $\tilde{\boldsymbol{\mu}}^{(1)}$ and $\tilde{\boldsymbol{\mu}}^{(0)}$ by numerically solving (6.6). This is justified by the convergence

result of Theorem V.2, provided that the number of events in the list is sufficiently large. Although the list must be large enough to obtain an accurate estimate of the asymptotic limit, the asymptotic method often requires fewer simulated events to compute the ROC than empirical methods. We also used the lists of recorded events to evaluate (6.12) and (6.13) by Monte Carlo integration to obtain $\Sigma(\tilde{\boldsymbol{\mu}}^{(1)})$ and $\Sigma(\tilde{\boldsymbol{\mu}}^{(0)})$. In the experiments that follow, we used an equal number of background events in the two lists. Let N_{back} be the number of sampled background events in each scan. The list of recorded events without a source present consists of N_{back} events and the list of recorded events with a source present consists of N_{back} background events and $N_{\text{back}}\alpha^{\mathsf{t}}s^{\mathsf{t}}/\lambda_b^{\mathsf{t}}$ source events. We use N_{pred} to report the total number of recorded events used to predict the detection performance using approximations based on asymptotics, where

$$N_{\text{pred}} = N_{\text{back}}(2 + \alpha^{\mathsf{t}}s^{\mathsf{t}}/\lambda_b^{\mathsf{t}}), \tag{5.46}$$

and the $\alpha^{\mathsf{t}}s^{\mathsf{t}}/\lambda_b^{\mathsf{t}}$ term is due to the source events in the list that contains both source and background events. We use the relation in (5.46) because N_{pred} is the expected number of total measurements used to compute the ROC using the proposed method.

We compared the performance predicted by asymptotics to the empirical performance to show the accuracy of the proposed method. To calculate the empirical performance, we simulated a large number of scans with and without the source present. To simulate a scan, we first drew the number of emitted source counts J_{src} from a Poisson distribution with mean $\tau\alpha^{\mathsf{t}}$. We then simulated J_{src} emitted source photons, recording the attributes of each photon that interacts with the detector. To simulate the background events for the scan, we drew the number of background counts J_{back} from a Poisson distribution with mean $\tau\lambda_b^{\mathsf{t}}$ and simulated background photons until we record the attributes for J_{back} interactions. For each scan, we com-

puted the QMLE $\tilde{\alpha}$ by solving (6.6).

We computed the empirical ROC and AUC using the source intensity estimates for all simulated scans with and without a source. For each value of the threshold in (6.9), the fraction of source intensity estimates above that threshold when a source is present is the probability of detection and the fraction of source intensity estimates above that threshold when a source is absent is the probability of false alarm. We computed the empirical AUC using the Wilcoxon-Mann-Whitney statistic [82]. Let N_{scans} be the number of scans used to compute the ROC. We simulated N_{scans} scans with a source present and N_{scans} scans without a source present. Let N_{emp} be the mean number of recorded sample events needed to complete all $2N_{\text{scans}}$ scans. We report the mean number of events used to compute an empirical ROC or AUC as N_{emp}, where

$$N_{\text{emp}} = N_{\text{scans}}\tau(2\lambda_b^{\mathsf{t}} + \alpha^{\mathsf{t}}s^{\mathsf{t}}). \tag{5.47}$$

5.4.1 Source Intensity Test for Simple Gaussian Detector

Setup

We first examine the effects of model mismatch using a hypothetical detector where the event attributes $\boldsymbol{r}$ obey the mixture model (5.28), where $\mathsf{p}_{\text{S}}(\boldsymbol{r}) = \mathcal{N}(m_{\text{s}}, \sigma_{\text{s}}^2)$, and $\mathsf{p}_{\text{B}}(\boldsymbol{r}) = \mathcal{N}(m_{\text{b}}, \sigma_{\text{b}}^2)$. Although the detector response in this example may not represent a physical system, the tractability of the true attribute distribution allows one to compare tests using misspecified models with tests using the correct model. Throughout this experiment, $\tau = 20$, $\alpha^{\mathsf{t}} = 1$, $\lambda_b^{\mathsf{t}} = 5$, and $s^{\mathsf{t}} = 1$.

The modeled distributions of the source and background attributes are $\tilde{\mathsf{p}}_{\text{S}}(\boldsymbol{r}; \boldsymbol{\phi}) = \mathcal{N}(\tilde{m}_{\text{s}}, \tilde{\sigma}_{\text{s}}^2)$ and $\tilde{\mathsf{p}}_{\text{B}}(\boldsymbol{r}) = \mathcal{N}(\tilde{m}_{\text{b}}, \tilde{\sigma}_{\text{b}}^2)$. In this simple example, we define the source position to be $\boldsymbol{\phi} = m_{\text{s}}$, which is assumed to be known. Model mismatch occurs when the modeled and true distributions have different means or standard deviations. We

used $N_{\text{emp}} = N_{\text{asym}} = 2.2 \times 10^5$ events to compute both the asymptotic and empirical performance.

Results

We first consider the effect of mismatch in the modeled variance. Figure 5.12 shows the AUC of the SIT as a function of the modeled attribute standard deviation $\tilde{\sigma}^2 = \tilde{\sigma}_{\text{s}}^2 = \tilde{\sigma}_{\text{b}}^2$, where the true standard deviations and means are $\sigma_{\text{s}}^2 = \sigma_{\text{b}}^2 = 1$, $m_{\text{b}} = \tilde{m}_{\text{b}} = 0$, and $m_{\text{s}} = \tilde{m}_{\text{s}} = 1$. The AUC accounting for mismatch is highest when the modeled attribute standard deviation equals the true standard deviation of one. This represents a correctly specified model for this problem. For small values of $\tilde{\sigma}$, the asymptotic AUC is overly optimistic when model mismatch is ignored. As $\tilde{\sigma}$ becomes large, the AUC approaches that of a counting detector. The agreement between the asymptotic AUC accounting for model mismatch and the empirical AUC demonstrates the accuracy of the proposed method in this scenario.

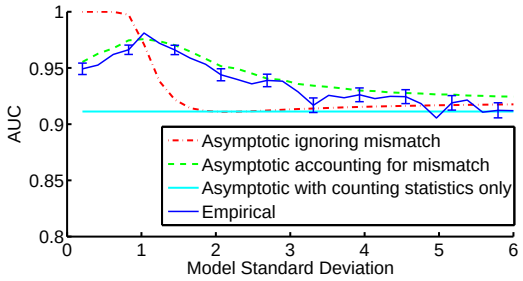


Figure 5.12: AUC versus modeled standard deviation $\tilde{\sigma}$ for true detector response $\sigma_{\text{s}}^2 = \sigma_{\text{b}}^2 = 1$, $m_{\text{b}} = \tilde{m}_{\text{b}} = 0$, and $m_{\text{s}} = \tilde{m}_{\text{s}} = 1$, $\tau = 20$, $\alpha^{\text{t}} = 1$, $s^{\text{t}} = 1$, and $\lambda_b^{\text{t}} = 5$. Error bars denote standard error.

Next, we considered mismatch in the source mean. Figure 5.13 shows the asymptotic AUC of the SIT using the Gaussian detector response defined above as a function of the modeled attribute mean $\tilde{m}_{\text{s}}$. The maximum predicted AUC accounting for model mismatch in Figure 5.13 is achieved when the modeled and true means are equal at $\tilde{m}_{\text{s}} = 1$. As in the case of standard deviation mismatch in Figure 5.12, the

asymptotic AUC accounting for mismatch closely agrees with the empirical AUC.

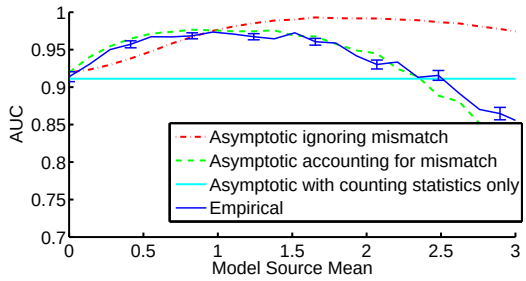


Figure 5.13: AUC versus modeled source mean $\tilde{m}_{\mathrm{s}}$ for the true detector response $\sigma_{\mathrm{s}}^2 = \sigma_{\mathrm{b}}^2 = \tilde{\sigma}_{\mathrm{s}}^2 = \tilde{\sigma}_{\mathrm{b}}^2 = 1$, $m_{\mathrm{b}} = 0$, and $m_{\mathrm{s}} = 1$, $\tau = 20$, $\alpha^{\mathrm{t}} = 1$, $s^{\mathrm{t}} = 1$, and $\lambda_b^{\mathrm{t}} = 5$. Error bars denote standard error.

This example shows that mismatch between the attribute variances σ_{s}^2 and σ_{b}^2 or the attribute means m_{s} and m_{b} can lead to highly inaccurate detection performance predictions when the mismatch is ignored. For large model mismatch, the actual detection performance can even be worse in terms of AUC than than the performance of a test that uses only counting statistics, in contrast to the optimistic predictions that ignore mismatch. The agreement between AUC computed using the proposed asymptotic method and the empirical AUC shows that the proposed method accurately predicts true detection performance for this system.

5.4.2 Source Intensity Test for Compton Detector - Monoenergetic Source and Background

Setup

We used the expressions derived in §5.3.5 to evaluate the AUC for a Compton imaging detector when the source position $\boldsymbol{\phi}$ and background spatial distribution are assumed known, but the background intensity λ_b is unknown. The detector is an idealized parallel–plate Compton camera [55] illustrated in Figure 5.14. We assumed that the position uncertainty and attenuation due to the finite thickness of the detector plates are negligible. The detector records positions and energies of interacting photons that undergo Compton scatter in the first, or scattering detector,

and either Compton scatter or photoelectric absorption in the second, or absorption detector. The photoelectric and Compton cross sections for this simulation are equal to that of CdZnTe. The absorption detector is infinitely large, the scatter detector is 1cm by 1cm, and the plates are separated by a distance $d = 0.1$ cm. The interaction positions in both detectors are quantized into 0.1 cm $\times$ 0.1 cm pixels. We also added Gaussian noise with zero mean and standard deviation of 1 keV to the recorded energies to simulate the effect of electronic noise on the recorded energy.

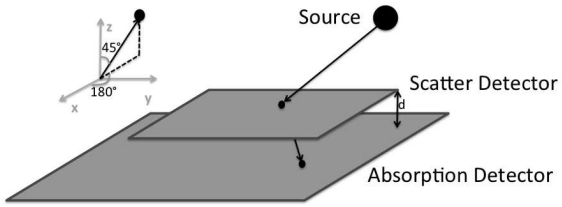


Figure 5.14: Illustration of parallel–plate Compton detector used for simulation.

The source is a 662 keV point–source located at an azimuth angle of 180 degrees and an elevation angle of 45 degrees in the far–field. For this experiment, let $\boldsymbol{\phi}$ be the azimuth and elevation angles of the source. We estimated the true sensitivity s^{t} for the chosen source position by simulating 10^4 emitted source photons and divided the number of recorded photons by the number of simulated photons. This is justified because one can show that in the absence of background, $\bar{J}/\tau$ converges in probability to $\lambda_{\mathsf{s}} = \alpha^{\mathsf{t}} s^{\mathsf{t}}$ as $\tau \to \infty$. In this experiment, we assumed that every photon strikes the scatter detector, but only some interacting photons result in two–interaction events. We found the true sensitivity under this assumption to be $s^{\mathsf{t}} = 0.44$. We model the sensitivity by its true value.

The true background is monoenergetic with the same energy as the source. Although a monoenergetic background is not realistic, it is useful to study this case

where energy alone does not differentiate between source and background photons. Let $\boldsymbol{\phi} = (\theta, \phi)$ be the angular position about the detector in spherical coordinates with inclination angle θ and azimuth angle ϕ. When the background is spatially uniform in the hemisphere above the detector, the origin direction of a photon emitted from the background obeys the density

$$f(\theta, \phi) = \frac{1}{2\pi} \sin\theta \quad \theta \in [0, \pi/2] \quad \phi \in [0, 2\pi].$$

We model the background as spatially uniform, but simulate uniform and nonuniform background emission densities. The difference between the modeled and true background is a source of model mismatch.

In this example, the attribute vector $\boldsymbol{r}$ for an interacting photon is a 6×1 vector containing the x-coordinate, y-coordinate, and recorded energy. When the source is in the far–field, the x and y-coordinates of the interaction in the scatter (top) detector are uniformly distributed, and the measured values are quantized to 0.1cm because the detector is pixelated. The energy deposited in each detector follows the Klein-Nishina formula [38]. The coordinates of the interaction in the second detector are a function of the interaction location in the first detector and the Compton scattering angle of the first interaction. The second interaction is also quantized to 0.1cm. This true distribution [80] is computationally difficult to compute, so we use the model proposed in [80], which provides a Gaussian approximation to the uncertainties induced by detector pixelization.

The goal of this section is to show that the proposed performance prediction method accounting for model mismatch agrees with the empirical performance. A predicted ROC that is lies above the the empirical ROC is not desirable because such a predicted performance is an overly optimistic assessment of the system's capability. Reducing model mismatch and improving the statistics of the data will improve the

empirical ROC. The most desirable scenario is one where the empirical ROC is high and the predicted ROC agrees with it.

Results for Source Located in Intense Background Region

We first simulate a nonuniform background to demonstrate the accuracy of the asymptotic predictions in the presence of model mismatch. We draw the origin direction of each background photon from the density

$$f(\theta,\phi) = \frac{\sin(2\phi+\pi/2)+1.5}{3\pi}\sin\theta, \quad \theta\in[0,\pi/2], \quad \phi\in[0,2\pi]. \tag{5.48}$$

Figure 5.15 shows the probability density function for the azimuth angle of an incoming background photon. The probability density is uniform in the polar direction.

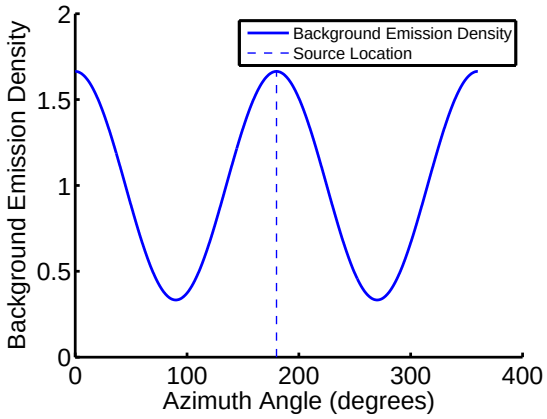


Figure 5.15: Probability density of azimuth angle of incoming background photon corresponding to (5.48).

Figure 5.16 shows the ROC for the true mean number of emitted source counts $\alpha^{\mathrm{t}}\tau = 60$ and known true mean number of recorded background counts $\lambda_b^{\mathrm{t}}\tau = 300$ with unconstrained source intensity estimates. Source intensity is naturally nonnegative, and one can choose whether or not to enforce that natural constraint on

the intensity estimates. Intuitively, a negative source intensity is strong evidence against the presence of a source, but we also found that asymptotic approximations are less accurate when the source intensity estimate is negative with high probability. Although nonphysical, there is no mathematical contradiction when the source intensity estimate is negative because the modeled likelihood in (6.3) is well-defined if α and λ_b are chosen such that $\tilde{\mathsf{p}}\left(\boldsymbol{r};\boldsymbol{\theta}\right)$ is nonnegative for any observed event $\boldsymbol{r}$, even it α is negative.

We used $N_{\text{pred}} = 208,782$ events to compute the predicted ROC and $N_{\text{emp}} = 313,174$ events to compute the empirical ROC. The predicted ROC accounting for model mismatch is much closer to the empirical ROC than the predicted ROC ignoring model mismatch. Even with this small number of source counts per scan, the asymptotic approximation appears to be reasonable. This result shows that the proposed asymptotic analysis gives a reasonable approximation of the true ROC in this case.

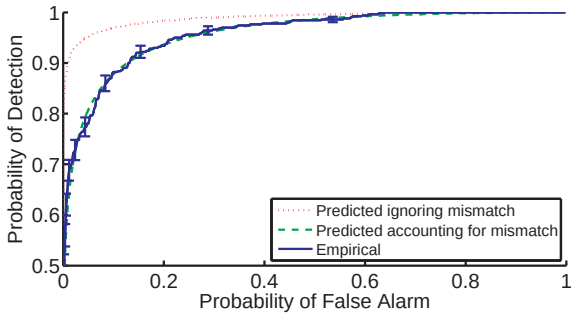


Figure 5.16: Empirical and predicted ROC curves of the SIT with $\alpha^{\mathsf{t}}\tau = 60$ counts, $\lambda_b^{\mathsf{t}}\tau = 300$ counts, unconstrained source intensity estimates, and the source positioned in the most intense background region. Error bars denote standard error.

Figure 5.17 shows the ROC for the same detection problem as Figure 5.16, except the source and background intensity estimates are constrained to be nonnegative. In this scenario, the ROC curves for the SIT with constrained and unconstrained

estimates are nearly identical.

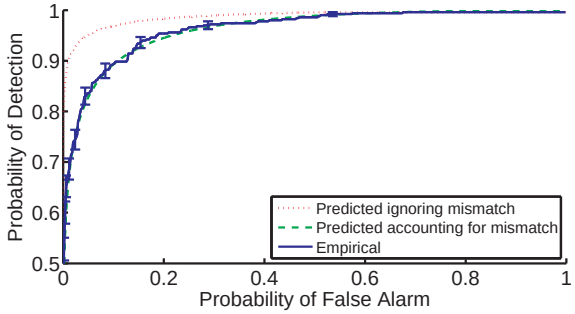
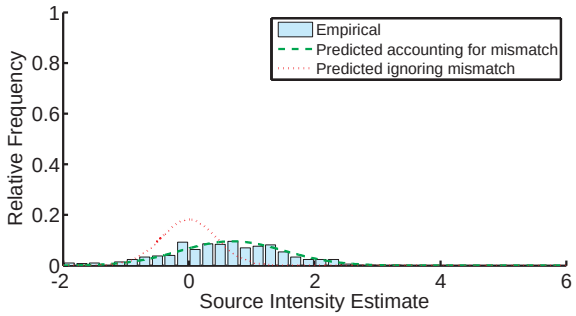


Figure 5.17: Empirical and predicted ROC curves of the SIT with $\alpha^{\mathrm{t}}\tau = 60$ counts, $\lambda_b^{\mathrm{t}}\tau = 300$ counts, constrained source intensity estimates, and the source placed in the most intense background region. Error bars denote standard error.

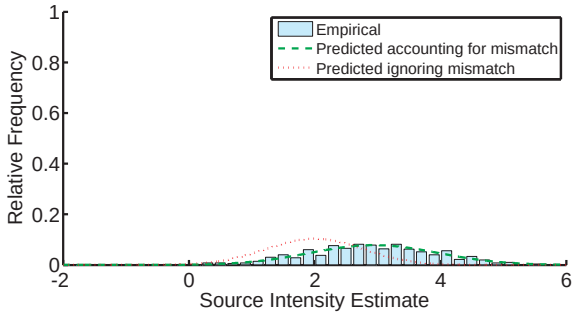
The asymptotic distributions of the source intensity estimates used to calculate the ROC in Figure 5.16 are also reasonably accurate approximations of the empirical distributions. Figure 5.18 shows histograms of the empirical source intensity estimates and the asymptotic probability density functions (we applied scaling to the density functions to make the units consistent with the histogram). The asymptotic distributions that account for mismatch fit the empirical distributions reasonably well.

Figure 5.19 shows histograms of the source intensity estimates used to compute the constrained ROC curve in Figure 5.17. The histogram with a source present in Figure 5.19 shows that there are few estimates that are equal to zero, thus the nonnegativity constraint does not affect the distribution of the intensity estimates under H_1. Thus, for any threshold value greater than zero, the probability of false alarm and probability of detection is not affected much by the nonnegativity constraint.

We also examined the agreement between the predicted AUC and the empirical AUC as a function of scan time. We expect the agreement between the empirical and predicted AUC to improve as scan time increases because Theorems V.2 and V.4

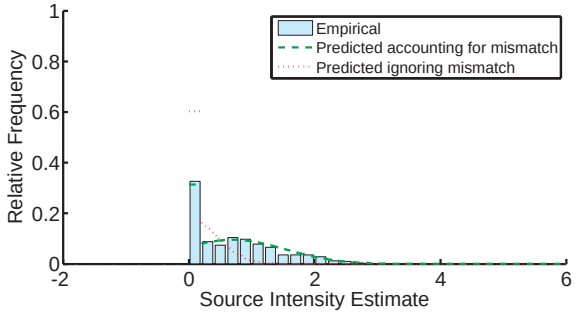


(a) Source Absent (H_0)

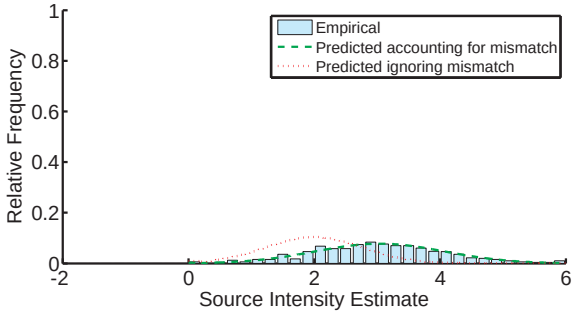


(b) Source Present (H_1)

Figure 5.18: Histograms and scaled asymptotic probability density functions of source intensity estimates where $\alpha^{\mathrm{t}}\tau = 60$ counts, $\lambda_b^{\mathrm{t}}\tau = 300$ counts, and the source is in the most intense background region.



(a) Source Absent (H_0)



(b) Source Present (H_1)

Figure 5.19: Histograms and scaled asymptotic probability density functions of source intensity estimates where $\alpha^{\mathrm{t}}\tau = 60$ counts, $\lambda_b^{\mathrm{t}}\tau = 300$ counts, constrained source intensity estimates, and the source in the most intense background region

give convergence results as scan time goes to infinity. Figure 5.20 shows the empirical and asymptotic AUC versus scan time for source intensity $\alpha^{\mathsf{t}} = 2$ counts/sec and background intensity $\lambda_b^{\mathsf{t}} = 10$ counts/sec. The numbers of events used to compute the asymptotic and empirical AUC are given by (5.46) and (5.47), respectively, where $N_{\text{back}} = 10,000$ counts and $N_{\text{scans}} = 500$ counts. In this example, the predicted AUC approximates the empirical AUC well even for small scan times. In both Figures 5.16 and 5.20, the predicted performance using the proposed expressions that account for model mismatch are much closer to the empirical performance than the predictions that ignore model mismatch.

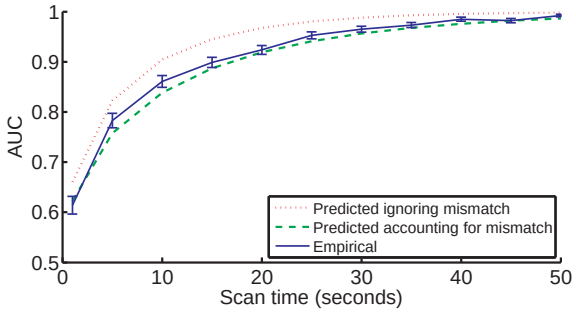


Figure 5.20: Empirical and asymptotic AUC of SIT vs. scan time τ for $\alpha^{\mathsf{t}} = 2$ count/sec, $\lambda_b^{\mathsf{t}} = 10$ counts/sec, and unconstrained source intensity estimates with the source located in the most intense background region. Error bars denote standard error.

Figure 5.21 shows the AUC as a function of scan time when the source and background intensities are constrained. The AUC values are similar to the values computed using unconstrained estimates in Figure 5.20 because the source intensity estimates are nonnegative with high probability under H_1 for the scan times in the figure.

Results for Source Located in Weak Background Region

We now examine the detection performance when the source is located in a weak region of a nonuniform background and the background is assumed to be spatially

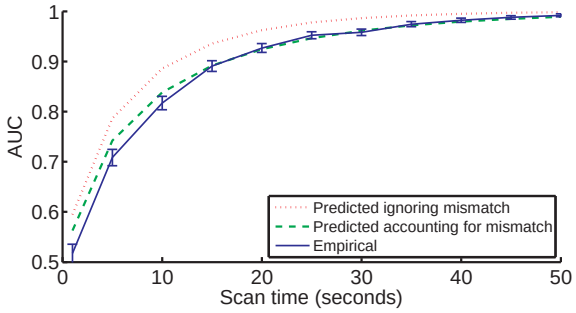


Figure 5.21: Empirical and asymptotic AUC of SIT vs. scan time τ for $\alpha^{\mathrm{t}} = 2$ count/sec, $\lambda_b^{\mathrm{t}} = 10$ counts/sec, and constrained source intensity estimates with the source located in the most intense background region. Error bars denote standard error.

uniform. We draw the origin direction of each background photon from the density

$$f(\theta,\phi) \;=\; \frac{\sin(2\phi-\pi/2)-1.5}{3\pi}\sin\theta, \tag{5.49}$$
$$\theta\in[0,\pi/2],\quad \phi\in[0,2\pi].$$

Figure 5.22 shows the probability density function for the azimuth angle of an incoming background photon.

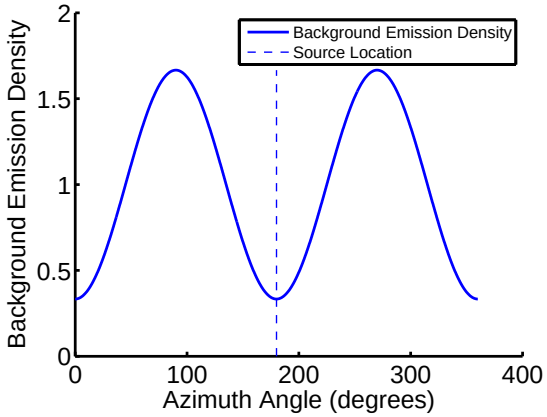


Figure 5.22: Probability density of the azimuth direction of an emitted background photon where the source is located at the weakest point.

Figure 5.23 shows the ROC resulting from the SIT with unconstrained source intensity estimates with a modeled spatially uniform background and a true background obeying the azimuthal density in (5.49). The asymptotic prediction in this

case is very poor because the unconstrained asymptotic limit $\tilde{\boldsymbol{\mu}}$ is much less than zero, and the log–likelihood has a high curvature in this regime. The agreement between the empirical performance and the performance predicted ignoring model mismatch is likely a coincidence because the ROC agrees, but the distributions of source intensity estimates do not.

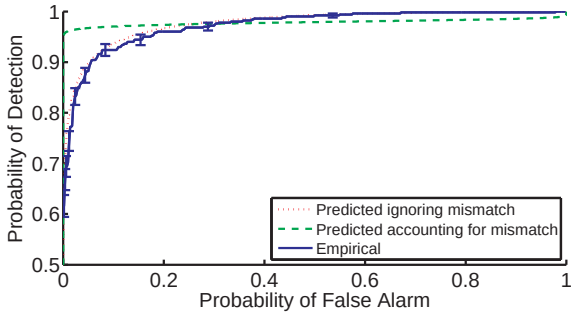


Figure 5.23: Empirical and predicted ROC curves of the SIT with $\alpha^{\mathrm{t}}\tau = 60$ counts, $\lambda_b^{\mathrm{t}}\tau = 300$ counts, unconstrained source intensity estimates, and the source positioned in the least intense background region. Error bars denote standard error.

Figure 5.24 shows the threshold γ for the SIT in (6.9) as a function of the probability of false alarm. The disagreement between the thresholds in this figure shows that although the ROCs agree in Figure 5.23, the predicted thresholds do not. In this case, application of the predictions that ignore model mismatch will result in an improper threshold choice for the desired false alarm rate.

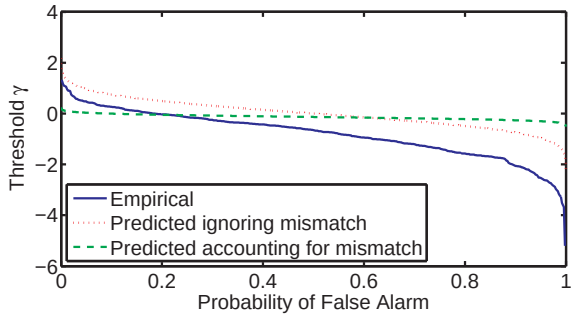


Figure 5.24: Threshold γ as a function of the probability of false alarm with the source positioned in the least background region.

Figure 5.25 shows the threshold γ for the SIT in (6.9) as a function of the probability of detection. As in Figure 5.24, the agreement is poor between the thresholds from the predictions that ignore mismatch and the empirical predictions. The thresholds from the predictions that account for mismatch are closer to the empirical thresholds because the asymptotic approximation to the SIT is better with a source present in this case.

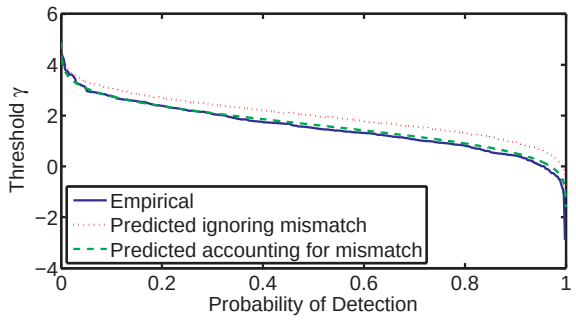


Figure 5.25: Threshold γ as a function of the probability of detection with the source positioned in the least background region.

Figure 5.26 shows the ROC resulting from the SIT with constrained source intensity estimates. As in the unconstrained case in Figure 5.23, the agreement between the predicted and empirical ROC is poor.

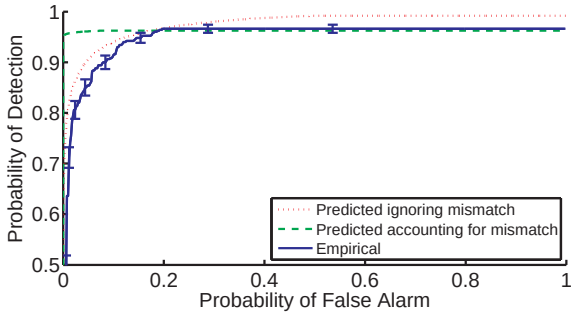
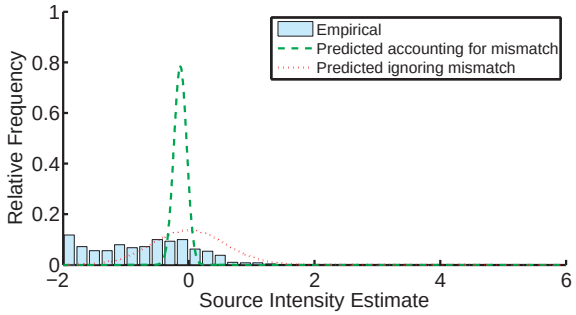
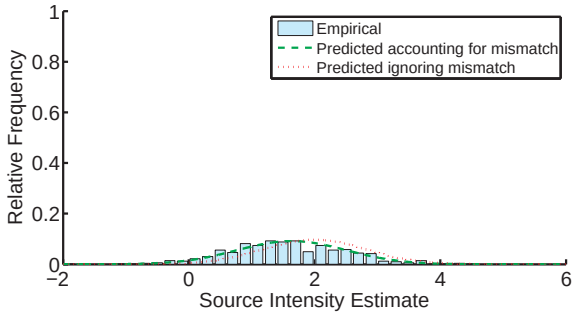


Figure 5.26: Empirical and predicted ROC curves of the SIT with $\alpha^{\mathrm{t}}\tau = 60$ counts, $\lambda_b^{\mathrm{t}}\tau = 300$ counts, constrained source intensity estimates, and the source positioned in the least intense background region. Error bars denote standard error.

Figure 5.27 shows the histograms of the unconstrained source intensity estimates



(a) Source Absent (H_0)



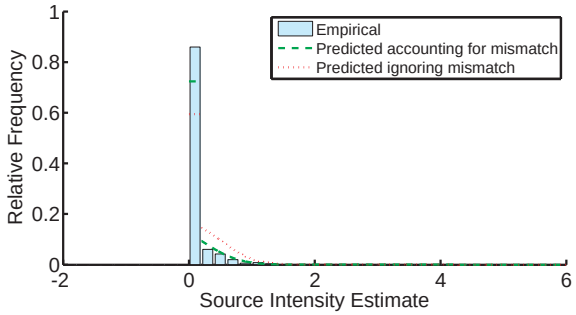
(b) Source Present (H_1)

Figure 5.27: Histograms and scaled asymptotic probability density functions of source intensity estimates where $\alpha^{\mathsf{t}}\tau = 60$ counts, $\lambda_b^{\mathsf{t}}\tau = 300$ counts, unconstrained source intensity estimates, and the source in the least intense background region.

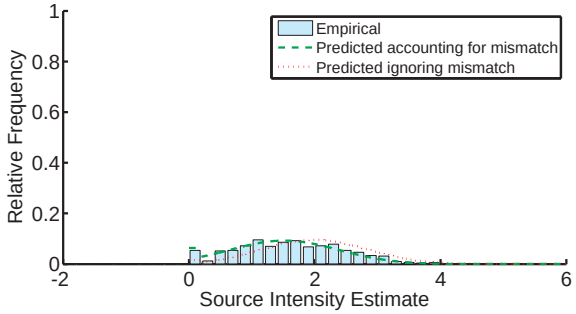
used to produce the ROC in Figure 5.23. The histogram under H_1 in Figure 5.27a shows that large negative source intensity estimates occur frequently, contrary to the asymptotic prediction. Note that the predicted distribution ignoring model mismatch does not appear to fit the empirical distribution better than the predicted distribution accounting for model mismatch.

Figure 5.28 shows the histogram of the constrained source intensity estimates used to compute the ROC in Figure 5.26. As in the unconstrained case in Figure 5.27, the predicted distributions poorly approximate the empirical distribution.

Figure 5.29 shows the AUC as a function of scan time when the source is located in the weak background region and the source intensity estimates are unconstrained. The predicted AUC is more accurate than the predicted ROC in Figure 5.23 for large scan times.



(a) Source Absent (H_0)



(b) Source Present (H_1)

Figure 5.28: Histograms and scaled asymptotic probability density functions of source intensity estimates where $\alpha^{\mathsf{t}}\tau = 60$ counts, $\lambda_b^{\mathsf{t}}\tau = 300$ counts, constrained source intensity estimates, and the source in the least intense background region.

Figure 5.30 shows the AUC as a function of scan time when the source is located in the weak background region and source intensity estimates are constrained. The predicted performance consistently overestimates the empirical performance, and the approximation is especially poor for small scan times. The predicted performance with constrained source intensity estimates approximates the empirical performance more poorly than when the source intensity estimates are unconstrained in Figure 5.29.

Results for Source Located in Moderate Background Region

Thus far, we examined detection performance when the source is located at the most intense and least intense point of a sinusoidal background and the background is modeled as uniform. In this section, we study the detection performance when the source is located between the least and most intense points of the background. The

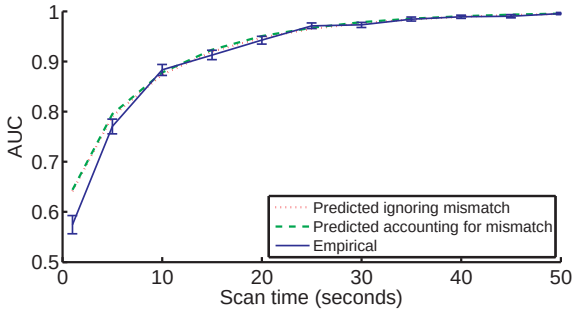


Figure 5.29: Empirical and asymptotic AUC of SIT vs. scan time τ for $\alpha^{\mathsf{t}} = 2$ count/sec, $\lambda_b^{\mathsf{t}} = 10$ counts/sec, and unconstrained source intensity estimates with the source located in the least intense background region. Error bars denote standard error.

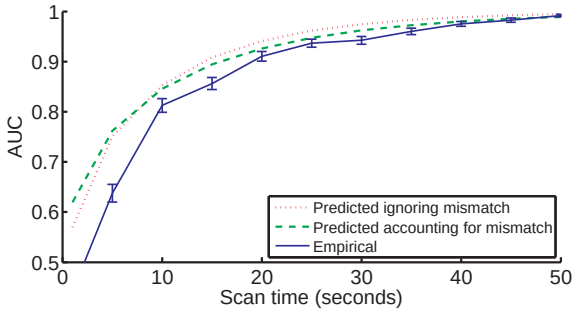


Figure 5.30: Empirical and asymptotic AUC of SIT vs. scan time τ for $\alpha^{\mathsf{t}} = 2$ count/sec, $\lambda_b^{\mathsf{t}} = 10$ counts/sec, and constrained source intensity estimates with the source located in the least intense background region. Error bars denote standard error.

true distribution of the angle of incoming recorded background photons is

$$f(\theta, \phi) = \frac{\sin(2\phi) - 1.5}{3\pi} \sin\theta, \tag{5.50}$$
$$\theta \in [0, \pi/2], \quad \phi \in [0, 2\pi].$$

Figure 5.31 shows the probability density function for the azimuth angle of a recorded background photon with the source location represented by the dashed vertical line. We model the spatial background density as uniform, which introduces model mismatch.

Figure 5.32 shows the ROC when the true background distribution obeys (5.50). The empirical and predicted ROC accounting for mismatch agree well, and the ROC predicted ignoring model mismatch overstates the true ROC. The agreement between

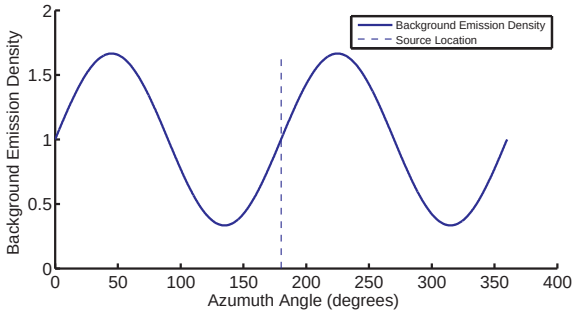


Figure 5.31: Probability density of the azimuth direction of an emitted background photon where the source is located between the most and least intense points.

the predicted and empirical ROC is much better than when the source is located in the least intense background region in Figure 5.23

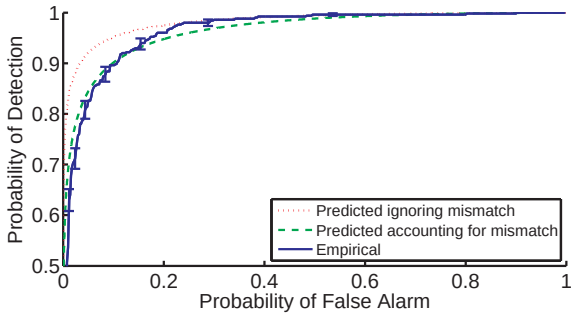


Figure 5.32: Empirical and predicted ROC curves of the SIT with $\alpha^{\mathrm{t}}\tau = 60$ counts, $\lambda_b^{\mathrm{t}}\tau = 300$ counts, unconstrained source intensity estimates, and the source positioned between the least and most intense background regions. Error bars denote standard error.

Figure 5.33 shows the ROC under the same conditions, except the source intensity estimates are constrained. The predicted ROC accounting for mismatch more accurately represents the empirical performance than the predicted ROC ignoring model mismatch. The agreement between the predicted ROC and the empirical ROC is not as good as in the unconstrained case in Figure 5.32.

Figure 5.34 shows histograms of the unconstrained source intensity estimates used to produce the ROC in Figure 5.32. The unconstrained asymptotic limit $\tilde{\boldsymbol{\mu}}$, which appears as the mean of the distribution of the source intensity estimates, is positive

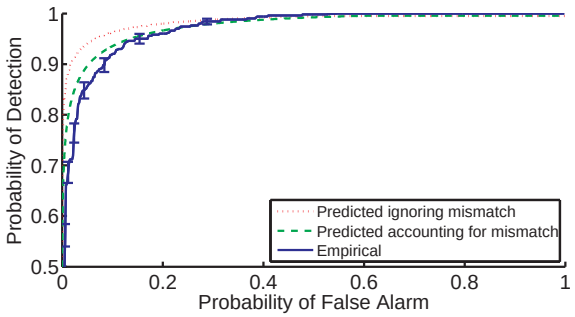


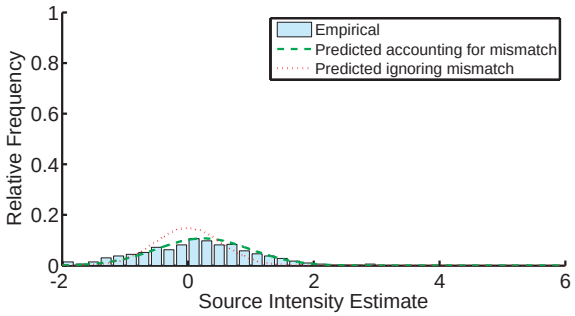
Figure 5.33: Empirical and predicted ROC curves of the SIT with $\alpha^{\mathrm{t}}\tau = 60$ counts, $\lambda_b^{\mathrm{t}}\tau = 300$ counts, unconstrained source intensity estimates, and the source positioned between the least and most intense background regions. Error bars denote standard error.

under H_1 and H_0. The estimates fit the Gaussian approximation much better than those in Figure 5.27, where the unconstrained asymptotic limit is negative under H_0. The predicted histogram that accounts for mismatch fits the empirical distribution much better than the predicted histogram that ignores model mismatch.

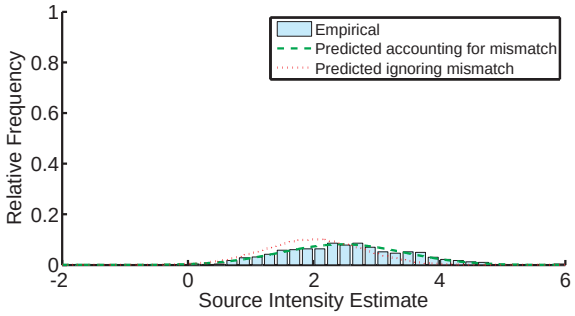
Figure 5.35 shows histograms of the source intensity estimates with the nonnegativity constraint enforced. As in the unconstrained case in Figure 5.34, the predicted distribution accounting for model mismatch fits the empirical distribution better than the predicted distribution that ignores model mismatch.

Figure 5.36 shows the AUC versus scan time when the source is located between the least and most intense background regions. The predicted AUC accounting for mismatch approximates the empirical AUC well, and the predicted AUC ignoring model mismatch overstates the empirical AUC.

Figure 5.37 shows the AUC under the same conditions as Figure 5.36, except the source intensity estimates are constrained. The agreement between the empirical and predicted AUC is not as good for small scan times as in the unconstrained case in Figure 5.36.

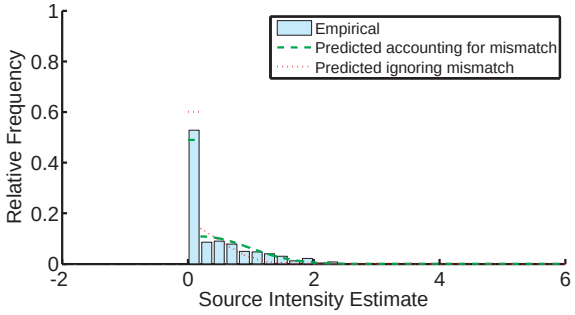


(a) Source Absent (H_0)

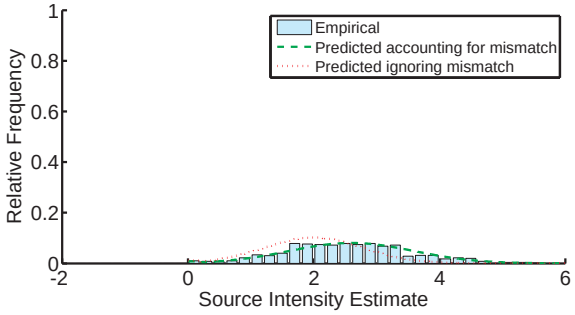


(b) Source Present (H_1)

Figure 5.34: Histograms and scaled asymptotic probability density functions of source intensity estimates where $\alpha^{\mathrm{t}}\tau = 60$ counts, $\lambda_b^{\mathrm{t}}\tau = 300$ counts, unconstrained source intensity estimates, and the source is between the least and most intense background regions.



(a) Source Absent (H_0)



(b) Source Present (H_1)

Figure 5.35: Histograms and scaled asymptotic probability density functions of source intensity estimates where $\alpha^{\mathrm{t}}\tau = 60$ counts, $\lambda_b^{\mathrm{t}}\tau = 300$ counts, constrained source intensity estimates, and the source is between the least and most intense background regions.

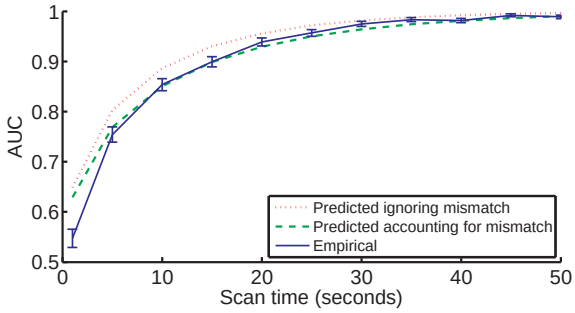


Figure 5.36: Empirical and asymptotic AUC of SIT vs. scan time τ for $\alpha^{\mathrm{t}} = 2$ count/sec, $\lambda_b^{\mathrm{t}} = 10$ counts/sec, and unconstrained source intensity estimates with the source located between the least and most intense background regions. Error bars denote standard error.

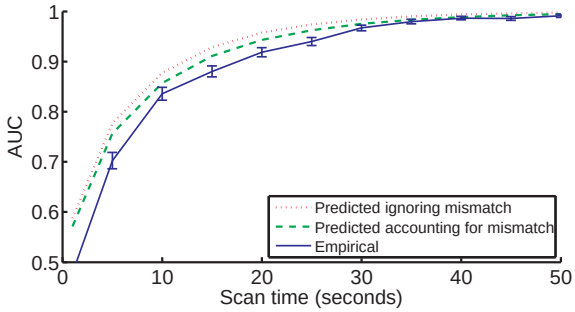


Figure 5.37: Empirical and asymptotic AUC of SIT vs. scan time τ for $\alpha^{\mathrm{t}} = 2$ count/sec, $\lambda_b^{\mathrm{t}} = 10$ counts/sec, and constrained source intensity estimates with the source located between the least and most intense background regions. Error bars denote standard error.

Results for Source in Uniform Background

In the previous sections, we examined detection performance when the background is modeled incorrectly, inducing model mismatch. In this section, both the true and modeled background spatial distributions are uniform. The probability density for the azimuth angle of a recorded background photon is given by

$$f(\theta, \phi) = \frac{1}{2\pi} \sin\theta, \quad \theta \in [0, \pi/2], \quad \phi \in [0, 2\pi]. \tag{5.51}$$

Figure 5.38 illustrates the density of the azimuth angle of recorded background photons.

Figure 5.39 shows the predicted and empirical ROC when the background is truly

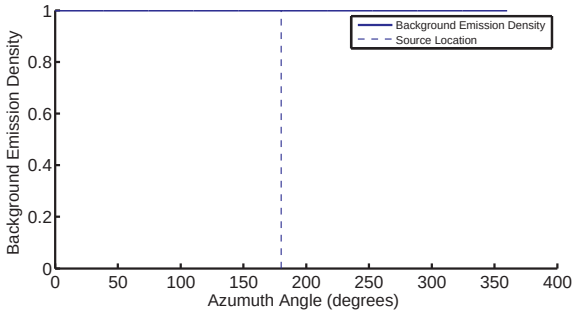


Figure 5.38: Probability density of the azimuth direction of an emitted background photon where the true background is spatially uniform.

spatially uniform and the background model is correct. The predicted ROC accurately approximates the empirical ROC. The predicted ROC accounting for model mismatch is very close to the predicted ROC ignoring model mismatch. This is because when the background is modeled correctly, there is little model mismatch. This shows that the approximations to the system response used in this work and first stated in [80] have little effect on detection performance.

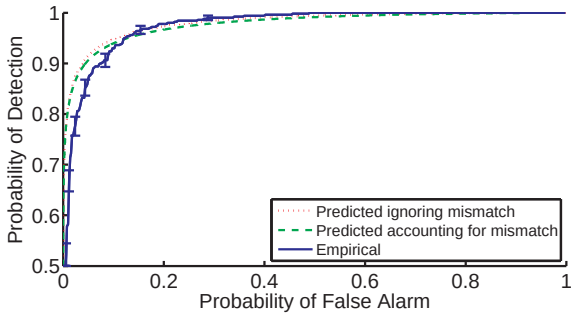


Figure 5.39: Empirical and predicted ROC curves of the SIT with $\alpha^{\mathrm{t}}\tau = 60$ counts, $\lambda_b^{\mathrm{t}}\tau = 300$ counts, unconstrained source intensity estimates, and modeled and true backgrounds are spatially uniform. Error bars denote standard error.

Figure 5.40 shows the ROC under the same conditions, except the source intensity estimates are constrained. The agreement between the predicted and empirical ROC is similar to the agreement in the unconstrained case in Figure 5.39.

Figure 5.41 shows histograms of the unconstrained source intensity estimates used

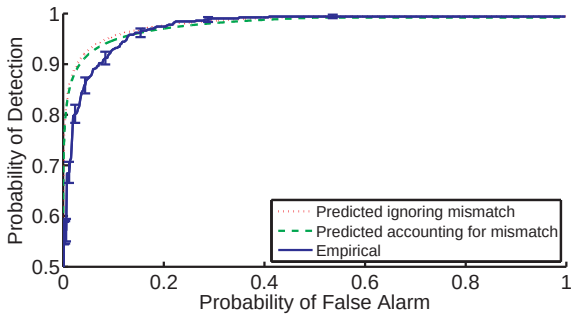


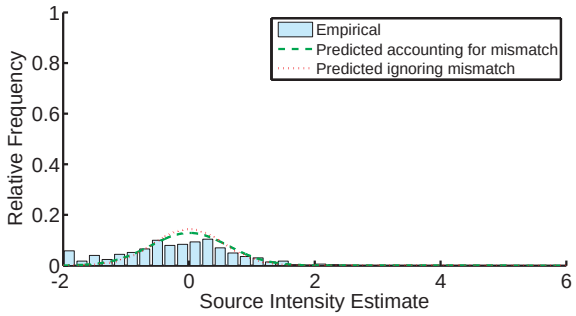
Figure 5.40: Empirical and predicted ROC curves of the SIT with $\alpha^{\mathrm{t}}\tau = 60$ counts, $\lambda_b^{\mathrm{t}}\tau = 300$ counts, unconstrained source intensity estimates, and modeled and true backgrounds are spatially uniform. Error bars denote standard error.

to compute the ROC in Figure 5.39. The empirical estimates under H_0 in Figure 5.41a shows that there are more estimates less than -2 than predicted. This is likely because the asymptotic limit $\tilde{\boldsymbol{\mu}}$, which appears as the mean of the predicted distribution, is near zero, and the log–likelihood has a large curvature for negative source intensity values. The large curvature of the log–likelihood causes the asymptotic approximations, which rely on a local quadratic approximation to the log–likelihood, to be less accurate.

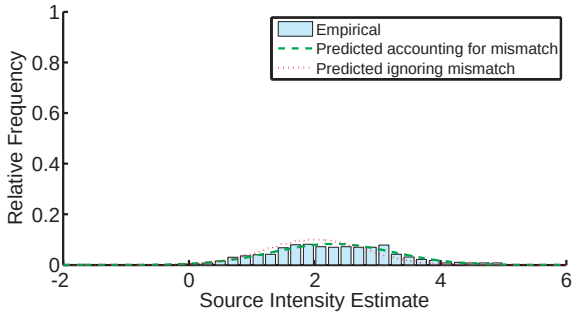
Figure 5.42 shows histograms of the source intensity estimates under the nonnegativity constraint. Enforcing the nonnegativity constraint causes the disagreement in the left tail of the distributions to be less important because all negative intensity estimates are truncated to zero.

Figure 5.43 shows the AUC as a function of scan time when the true background is spatially uniform and the background model is correct. The predicted AUC agrees well with the empirical AUC, and the agreement improves as scan time increases.

Figure 5.44 shows the AUC under the same conditions as in Figure 5.43, except the source intensity estimates are constrained. The agreement between the predicted and empirical AUC is slightly worse than it is in the unconstrained case for small

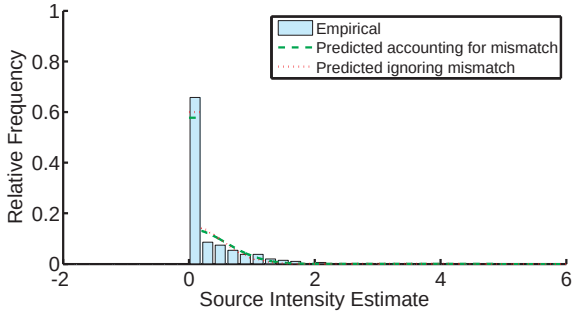


(a) Source Absent (H_0)

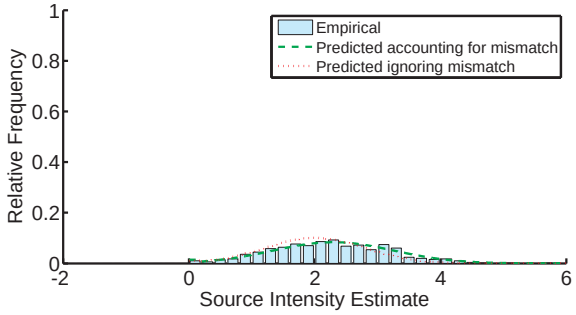


(b) Source Present (H_1)

Figure 5.41: Histograms and scaled asymptotic probability density functions of source intensity estimates where $\alpha^{\mathrm{t}}\tau = 60$ counts, $\lambda_b^{\mathrm{t}}\tau = 300$ counts, unconstrained source intensity estimates, and the true and modeled backgrounds are spatially uniform.



(a) Source Absent (H_0)



(b) Source Present (H_1)

Figure 5.42: Histograms and scaled asymptotic probability density functions of source intensity estimates where $\alpha^{\mathrm{t}}\tau = 60$ counts, $\lambda_b^{\mathrm{t}}\tau = 300$ counts, constrained source intensity estimates, and the true and modeled backgrounds are spatially uniform.

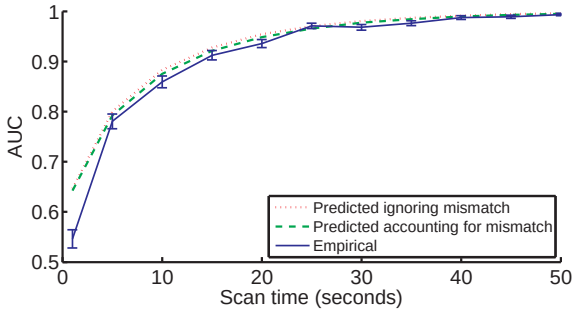


Figure 5.43: Empirical and asymptotic AUC of SIT vs. scan time τ for $\alpha^{\mathrm{t}} = 2$ count/sec, $\lambda_b^{\mathrm{t}} = 10$ counts/sec, and unconstrained source intensity estimates when the true background is spatially uniform and the background model is correct. Error bars denote standard error.

scan times, but the difference in agreement is small.

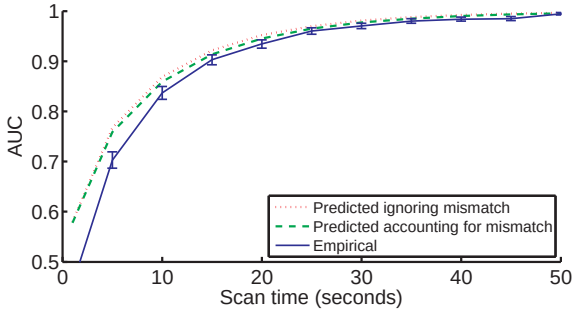


Figure 5.44: Empirical and asymptotic AUC of SIT vs. scan time τ for $\alpha^{\mathrm{t}} = 2$ count/sec, $\lambda_b^{\mathrm{t}} = 10$ counts/sec, and constrained source intensity estimates when the true background is spatially uniform and the background model is correct. Error bars denote standard error.

Accuracy of Asymptotic Predictions

To support the claim that the proposed asymptotic method for predicting the ROC and AUC requires less data than empirical methods, we computed various measures of error and uncertainty of the predicted and empirical AUC with various numbers of events sampled from the true distribution. We modeled the background as uniform, but unless stated otherwise, the true background was spatially distributed according to (5.48), where the background is a sinusoid with two periods in the azimuth direction and the source is located at a peak of the sinusoid. Recall that the number of events used to compute the ROC asymptotically N_{pred}, given by (5.46),

is the number of events used to simulate a list of events from the background and another list of events containing source and background events. The number of events N_{emp} used to compute the ROC empirically is given by (5.47).

We computed all measures of error and uncertainty in this section using the jackknife method [20] with 1000 trials. We drew N_{pred} sampled events without replacement from a pool of $10N_{\text{pred}}$ for each trial. We computed the true AUC empirically using 1,021,956 sampled events.

We examined the root mean square error (RMSE) [59, p. 261] of the AUC predictions as a measure of average deviation from the true AUC. Figure 5.45 shows the RMSE of the empirical and predicted AUC as a function of the number of photon interaction events $N_{\text{pred}} = N_{\text{emp}}$ sampled from the true distribution. Figure 5.45 shows that the predicted AUC using the proposed asymptotic method is more reliable in terms of RMSE than the empirical AUC, especially for a modest number of samples. The RMSE of the predicted AUC not accounting for mismatch is higher than the RMSE of the predicted AUC accounting for mismatch for large numbers of events because of the bias in the predictions that ignore model mismatch. This bias is caused by the false assumption that the model is correctly specified.

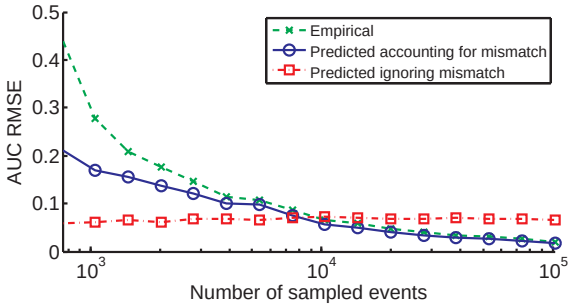


Figure 5.45: RMSE of the empirical and predicted AUC vs. the number of samples from the true distribution where $\alpha\tau = 20$ counts and $\lambda_b^{\text{t}}\tau = 200$ counts. The true AUC was computed empirically using 1,021,956 sampled events and the background shape is misspecified.

Figure 5.46 shows the RMSE versus the number of sampled events when the

background is correctly specified. In this case, the RMSE of the predictions ignoring mismatch is lower because the method that ignores mismatch does not estimate the asymptotic mean $\tilde{\boldsymbol{\mu}}$. This example shows that if the user knows that the model is correctly specified, the predictions ignoring model mismatch will be more accurate on average than predictions that account for model mismatch.

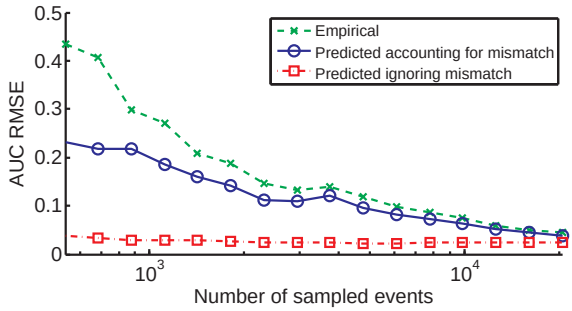
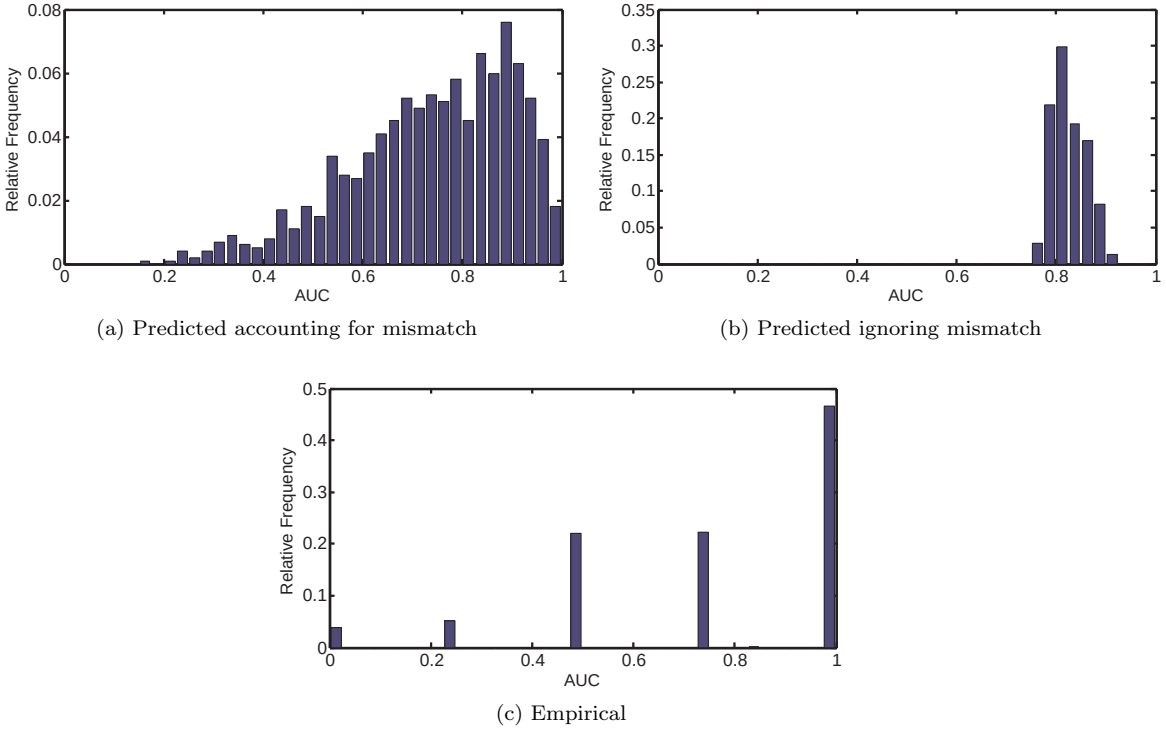


Figure 5.46: RMSE of the empirical and predicted AUC vs. the number of samples from the true distribution where $\alpha\tau = 20$ counts and $\lambda_b^{\mathrm{t}}\tau = 200$ counts. The true AUC was computed empirically using 1,021,956 sampled events and the background shape is correctly specified.

RMSE is the second moment of the error, and is most meaningful when the higher moments are small, or the error is approximately Gaussian. We found that the error was not approximately Gaussian for small numbers of sampled events. Figure 5.47 shows histograms of the predicted and empirical AUC for 1049 sampled events. The distribution is asymmetrical about the true AUC of 0.81 for both prediction methods. A distribution that is asymmetrical about its mean has a non–negligible third moment, and RMSE only measures the second moment of the error. The empirical histogram in Figure 5.47c appears to be multimodal because of the small number of events used in the computation.

Figure 5.48 shows histograms of the empirical and predicted AUC for 3881 events. The distributions in Figure 5.48 appear more Gaussian than the distributions corresponding to predictions with 1262 events in Figure 5.47. The predicted AUC in



(a) Predicted accounting for mismatch

(b) Predicted ignoring mismatch

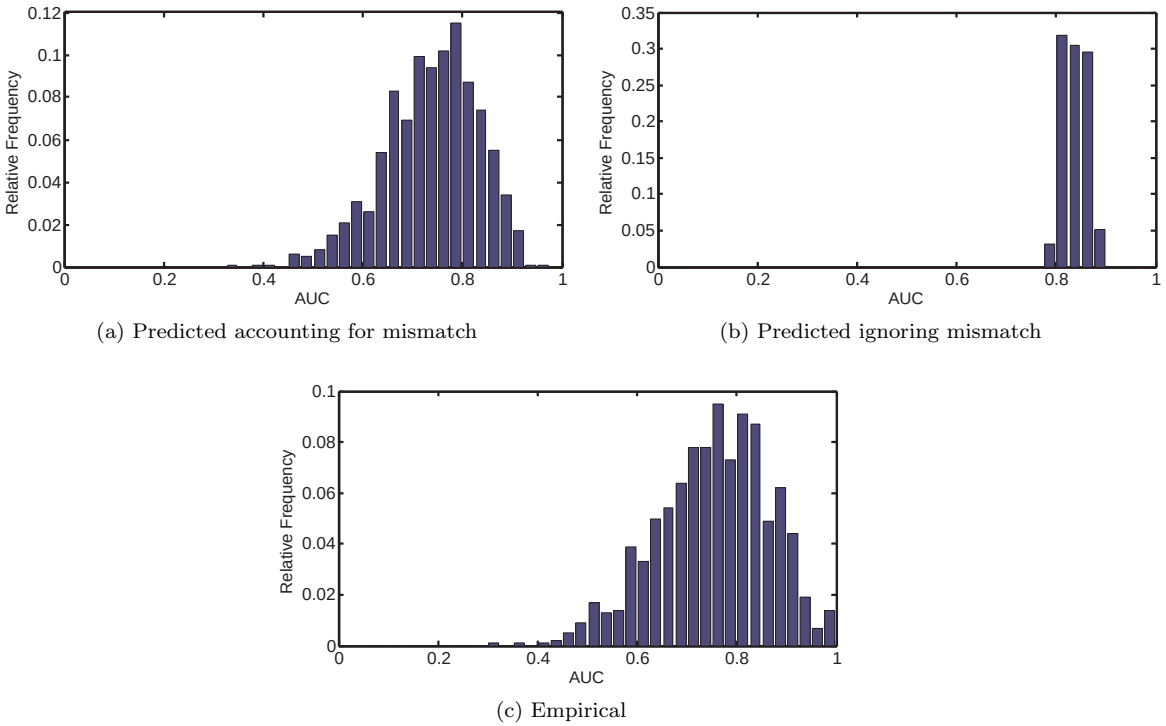
(c) Empirical

Figure 5.47: Histograms of the predicted and empirical AUC using 1049 recorded events. The AUC values were generated with 1000 jackknife trials, where the AUC was computed with 1049 events drawn without replacement from a pool of 10490 events for each trial.

Figure 5.48b appears to be more tightly distributed about the true AUC of 0.81 than the empirical AUC in Figure 5.48c, which is consistent with the observation of lower RMSE of the predicted AUC in Figure 5.45.

Figure 5.49 shows histograms of the empirical and predicted AUC for 53129 events. The distributions appear to fit a Gaussian distribution and have lower variance than the distributions with fewer events in Figures 5.48 and 5.47. The predicted AUC in Figure 5.49b is slightly more concentrated about the true AUC of 0.81 than the empirical AUC in Figure 5.49c.

Since the error of the AUC prediction is not well–approximated by its first and second moments with a small number of sampled events, we chose to examine the probability that the error exceeds 5% of the true value. For each number of sampled



(a) Predicted accounting for mismatch

(b) Predicted ignoring mismatch

(c) Empirical

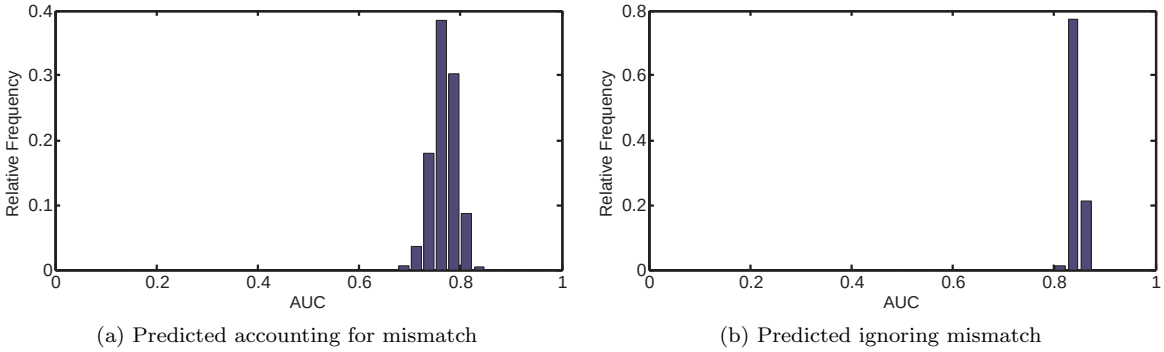
Figure 5.48: Histograms of the predicted and empirical AUC using 3881 recorded events. The AUC values were generated with 1000 jackknife trials, where the AUC was computed with 3881 events drawn without replacement from a pool of 38810 events for each trial.

events, we estimate the probability of error

$$p_e \triangleq \mathsf{Pr}\left(\frac{|\mathrm{AUC}_{\mathrm{measured}} - \mathrm{AUC}_{\mathrm{true}}|}{\mathrm{AUC}_{\mathrm{true}}}\right) < 0.05,$$

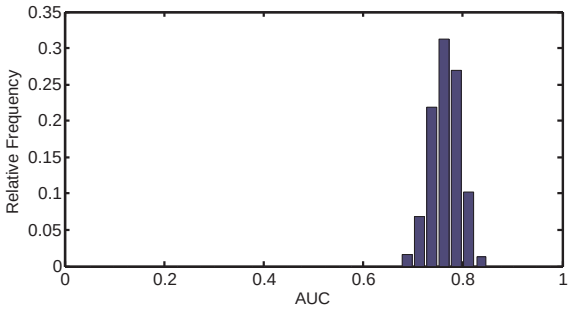
where $\mathrm{AUC}_{\mathrm{measured}}$ is the empirical or predicted AUC and $\mathrm{AUC}_{\mathrm{true}}$ is the true AUC. Figure 5.50 shows estimated value of the probability of error p_e as a function of the number of sampled events when the background shape is misspecified. Figure 5.50 shows that the predicted method results in a lower probability of error than the empirical method, but the difference is not as dramatic as the difference in the RMSE in Figure 5.45.

Figure 5.51 shows the probability of error when the background is correctly specified. As in the RMSE case, the predicted AUC ignoring model mismatch has a lower probability of error for a fixed number of events.



(a) Predicted accounting for mismatch

(b) Predicted ignoring mismatch



(c) Empirical

Figure 5.49: Histograms of the predicted and empirical AUC using 53129 recorded events. The AUC values were generated with 1000 jackknife trials, where the AUC was computed with 53129 events drawn without replacement from a pool of 531290 events for each trial.

The RMSE and probability of error characterize the distribution of the error of the AUC predictions, which depends on an estimate of the true AUC. We now consider measures of uncertainty in the AUC predictions that depend on the uncertainty in the predictions themselves rather than the deviation from the truth. We assume that the true background is spatially uniform and the model is nonuniform and given by (5.48). Figure 5.52 shows the standard deviation of the predicted and empirical AUC as a function of the number of sampled events $N_{\text{pred}} = N_{\text{emp}}$. The predicted AUC has a lower standard deviation than the empirical AUC, especially for a small number of sampled events.

We also considered the interquartile spread of the predicted and empirical AUC in Figure 5.53 since standard deviation only depends on the first and second moments of

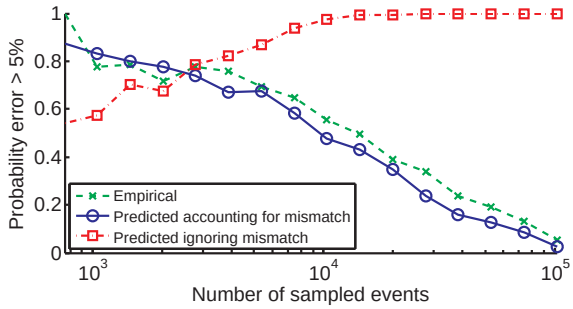


Figure 5.50: Estimated value of p_e, the probability that the error between the predicted and true AUC exceeds 5 vs. the number of samples from the true distribution where $\alpha\tau = 20$ counts and $\lambda_b^t\tau = 200$ counts. The true AUC was computed empirically using 1,021,956 sampled events and the background is misspecified.

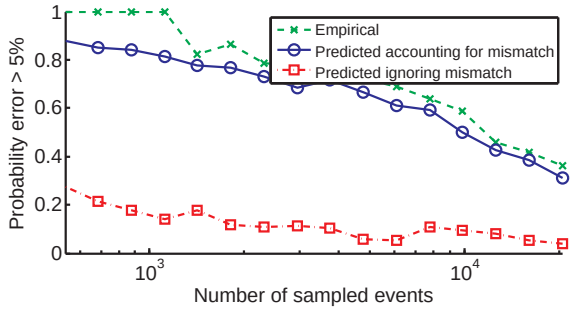


Figure 5.51: Estimated value of p_e, the probability that the error between the predicted and true AUC exceeds 5 vs. the number of samples from the true distribution where $\alpha\tau = 20$ counts and $\lambda_b^t\tau = 200$ counts. The true AUC was computed empirically using 1,021,956 sampled events and the background shape is correctly specified

the distribution. The interquartile spread of the predicted AUC is consistently lower than the interquartile spread of the empirical AUC, and the benefit of the proposed prediction method is more pronounced for smaller numbers of sampled events.

The experiments in this section show the utility of asymptotic approximations for AUC prediction with gamma–ray detectors. We considered measures of error and uncertainty in the predictions themselves. The proposed prediction method outperformed empirical AUC computation with every measure of error or uncertainty that we considered.

The benefits of asymptotic approximations for AUC prediction are not limited to

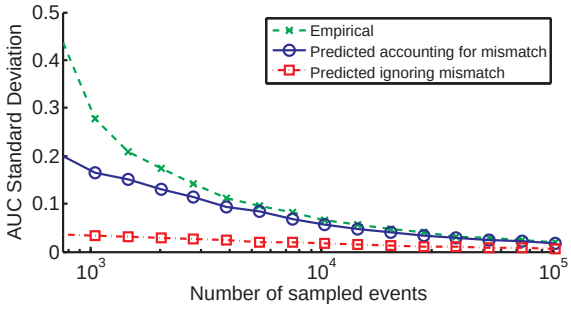


Figure 5.52: Standard deviation of predicted and empirical AUC vs. the number of samples from the true distribution where $\alpha\tau = 20$ counts and $\lambda_b^{\mathrm{t}}\tau = 200$ counts.

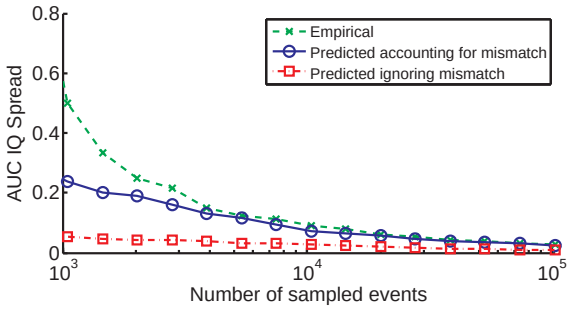


Figure 5.53: Interquartile spread of predicted and empirical AUC vs. the number of samples from the true distribution where $\alpha\tau = 20$ counts and $\lambda_b^{\mathrm{t}}\tau = 200$ counts.

lower error and uncertainty. We also observed in our experiments that the asymptotic predictions required less computation time than the empirical methods. Also, the values of the asymptotic limit in (6.8) and variance in (6.12) and (6.13) are useful for determining the degree of model mismatch. For instance, if the asymptotic covariance matrices $\tilde{\mathsf{G}}(\boldsymbol{\theta})$ and $\tilde{\mathsf{H}}(\boldsymbol{\theta})$ in Theorem V.4 are identical and the asymptotic limit $\tilde{\boldsymbol{\mu}}$ is close to the value one would expect from one's knowledge of the detector and environment, one can conclude that there is little model mismatch. Furthermore, one can use (5.39) to verify that the asymptotic covariance matrices scale linearly with the scan time τ, so one only needs to compute the computationally costly Monte Carlo integrations in (5.40) once to obtain performance predictions for multiple scan times. The empirical AUC computation method does not scale conveniently with

the scan time τ, so one would need to perform a complete computation for each scan time considered.

CHAPTER VI

Application of Asymptotic Detection Performance Prediction to Real Systems

In Chapters IV and V, we developed asymptotic approximations for the distributions of likelihood–based estimators and used those approximations to predict detection performance. We showed that these approximations are reasonable with simulated and idealistic detectors. The simulation results do not demonstrate that the proposed method can accurately predict the performance of real detectors. In this chapter, we show that the performance prediction method that accounts for model mismatch developed in Chapter V can accurately predict source detection performance with a real Compton imaging system.

In addition to demonstrating the practical utility of the asymptotic performance predictions of Chapter V, this work serves as an example application of the asymptotic performance prediction method to practical problems. We use the asymptotic method to predict the probability of detection as a function of scan time with a fixed false–alarm rate for various source–to–background ratios. These examples demonstrate how the asymptotic performance prediction method can be applied to evaluate the performance of real detectors in the field.

Previous work has focused on computing the detection performance in terms of the ROC empirically, e.g., [66, 76]. We showed in Chapter V that the asymptotic

performance prediction method yields more accurate AUC predictions in terms of mean–square error than empirical methods, especially when few measurements are available. To our knowledge, this work is the first to apply the asymptotic performance prediction method to characterize the performance of the source detection task with a real gamma–ray imaging system.

Demonstrating the accuracy of the asymptotic performance prediction method with real data is significant because there is more model mismatch than in the simulated case. For example, Doppler broadening [38] is not simulated in Chapter V. Room–temperature pixellated semiconductor detectors, including the detectors used in this work [30], have an area near the anode where interacting photons are not detected. To simplify computation, the model used in this work, based on [80], does not account for this non–ideal detector behavior. Furthermore, crystal defects can cause errors in the measured interaction positions that are not accounted for by the model. Our results show that the asymptotic prediction method is reasonably accurate in the scenarios considered despite the system response approximation and failure to account for all non–ideal detector behavior.

The contributions of this work are: (i) to show that the asymptotic performance prediction method developed in Chapter V gives reasonable predictions with a real system, (ii) to illustrate practical uses of this method, and (iii) to provide representative detection performance figures for a real CdZnTe gamma–ray imaging system. This chapter is organized as follows: §6.1 describes the experimental setup, §6.2 shows predicted and empirical performance of the detector for various tasks, and §6.3 gives our conclusions.

6.1 Methods

We recorded gamma–ray interaction data with a Compton imaging system consisting of an 18 detector CdZnTe array similar to the system described in [83]. We obtained list–mode measurements of the natural background in a room with concrete walls, and measurements in the same position with a Cs-137 source located 1.83 meters from the front of the detector. We use the events obtained from these measurements to evaluate the source detection performance of the system.

6.1.1 Measurement Model

There are many aspects of the gamma–ray source detection problem that one can model. The system model and sensitivity are necessary for the likelihood–based detection methods used in this work. We also model the background spatial and energy distributions because this improves detection performance when the modeling is reasonably accurate. Background modeling is beneficial in applications where the detector and environment are stationary.

Model Parameters

There are several parameters that characterize the gamma–ray source detection problem. We characterize the source by its intensity α with units of counts *emitted* per unit time and position[1] $\boldsymbol{\phi} \in \Phi$. In the 3D far–field with a known source energy, the set Φ could be $[0, 2\pi] \times [0, \pi]$, representing all possible azimuth and polar angles on a sphere. We parameterize the background intensity by the background count rate λ_b with units of gamma–ray counts *recorded* per unit time. We assume that the background spectrum is known. Let $\boldsymbol{\theta}$ be the vector of all parameters, where $\boldsymbol{\theta}$ lies in the d-dimensional parameter space Θ. In what follows, we assume that $\boldsymbol{\theta}$ takes

[1] $\boldsymbol{\phi}$ could also denote a vector containing both spatial position and energy

the form:

$$\boldsymbol{\theta} = (\alpha, \boldsymbol{\phi}, \lambda_b), \tag{6.1}$$

for which $d = 3$. Throughout this work, we assume that the source position in space and energy $\boldsymbol{\phi}$ is known, and λ_b can be either known or unknown. Let the modeled sensitivity $\tilde{s}(\boldsymbol{\phi})$ approximate the probability that a photon emitted from a source positioned at $\boldsymbol{\phi}$ is recorded. We model the total rate of recorded photons by the sum of the rates of recorded source and background photons

$$\tilde{\lambda}(\boldsymbol{\theta}) \stackrel{\triangle}{=} \lambda_b + \alpha \tilde{s}(\boldsymbol{\phi}). \tag{6.2}$$

System Model

We use the model given in in [50] to describe the system used in this work. Let $\boldsymbol{r}$ be a vector of recorded attributes associated with a single photon interaction. In a position–sensitive Compton detector, the attribute vector $\boldsymbol{r}$ contains the interaction positions and deposited energies for a single interacting photon. In fixed–time mode, the number of recorded photons J is reasonably modeled as a random variable, where $J \sim \text{Poisson}(\tilde{J}(\boldsymbol{\theta}))$. The mean number of recorded photons $\tilde{J}(\boldsymbol{\theta})$ is given by

$$\tilde{J}(\boldsymbol{\theta}) \stackrel{\triangle}{=} \tilde{\lambda}(\boldsymbol{\theta})\tau,$$

where τ is the scan time. Let $\tilde{\boldsymbol{r}} = [\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_J]$ be a list of the recorded attributes for all interacting photons during a fixed–time scan. By the statistics of list–mode data [4], a reasonable model for the list of recorded attributes $\tilde{\boldsymbol{r}}$ is

$$\tilde{\mathsf{p}}\left(\tilde{\boldsymbol{r}}; \boldsymbol{\theta}\right) \stackrel{\triangle}{=} e^{-\tau\tilde{\lambda}(\boldsymbol{\theta})}[\tau\tilde{\lambda}(\boldsymbol{\theta})]^J / J! \prod_{j=1}^{J} \tilde{\mathsf{p}}(\boldsymbol{r}_j; \boldsymbol{\theta}). \tag{6.3}$$

We model the probability density of the individual recorded attributes $\tilde{\mathsf{p}}\left(\boldsymbol{r}; \boldsymbol{\theta}\right)$ using the approximate model in [80], which makes approximations to achieve com-

putational efficiency. We use this approximate model because it results in reasonable detection performance and is much faster than computing the true density exactly.

Let $\tilde{\mathsf{p}}_{\mathrm{S}}(\boldsymbol{r};\boldsymbol{\phi})$ denote the modeled density of a recorded attribute vector $\boldsymbol{r}$ given it originated from a source at position $\boldsymbol{\phi}$ and let $\tilde{\mathsf{p}}_{\mathrm{B}}(\boldsymbol{r})$ denote the modeled density of a recorded attribute vector $\boldsymbol{r}$ given that it originated from the background. Note that $\tilde{\mathsf{p}}_{\mathrm{S}}(\boldsymbol{r};\boldsymbol{\phi})$ depends only on the source position, and $\tilde{\mathsf{p}}_{\mathrm{B}}(\boldsymbol{r})$ does not depend on any of the parameters in (6.1).

We model the overall distribution of recorded attributes as a mixture of $\tilde{\mathsf{p}}_{\mathrm{S}}(\boldsymbol{r};\boldsymbol{\phi})$ and $\tilde{\mathsf{p}}_{\mathrm{B}}(\boldsymbol{r})$ given by

$$\tilde{\mathsf{p}}\left(\boldsymbol{r};\boldsymbol{\theta}\right) = \frac{\alpha\tilde{s}(\boldsymbol{\phi})\tilde{\mathsf{p}}_{\mathrm{S}}(\boldsymbol{r};\boldsymbol{\phi}) + \lambda_b\tilde{\mathsf{p}}_{\mathrm{B}}(\boldsymbol{r})}{\alpha\tilde{s}(\boldsymbol{\phi}) + \lambda_b}. \tag{6.4}$$

Sensitivity Model

We computed the sensitivity model $\tilde{s}(\boldsymbol{\phi})$ by simulating the detector system in a uniform background using GEANT4 [1]. We used simple back projection [79] to reconstruct the sensitivity as a function of position and energy. We normalized the sensitivity so the sensitivity to the source at its true position and energy is one. This method of computing the sensitivity is approximate and it too may be a source of model mismatch.

Background distribution

The simulated results in [50] assumed a monoenergetic source and background. The natural radiation background is polyenergetic, so we discretize the energy spectrum into 80 uniformly–spaced bins from 0keV to 2000 keV. We assume that the source is monoenergetic and that its energy is known. We assume that the shape of the background spectrum is known, and perform detection with and without the assumption that its intensity is known. We examine the performance difference between

a uniform spectral model and a spectral model based an independent measurement of the same environment.

We measured the natural background using 10,000 recorded background photons with two or more interactions. We used an expectation maximization (EM) algorithm [17] to reconstruct the emission density as a function of energy. Figure 6.1 shows the reconstructed background spectrum.

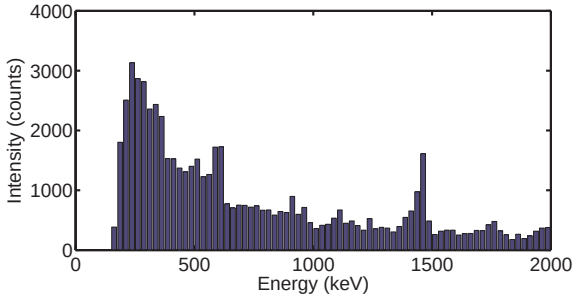


Figure 6.1: Intensity versus energy reconstructed from 10,000 recorded events from the natural background.

Let $\tilde{\mathsf{p}}\,(\boldsymbol{\phi}|B)$ be the probability of the incoming position and energy $\boldsymbol{\phi}$ of a background photon. The source detection model assumes that the background emission density is uniform in space, but varying in energy. Let $f(\boldsymbol{\phi})$ be the measured intensity of the background. We have that [4]

$$\tilde{\mathsf{p}}\,(\boldsymbol{\phi}|B) = \frac{f(\boldsymbol{\phi})\tilde{s}(\boldsymbol{\phi})}{\int_{\Phi} f(\boldsymbol{\phi})\tilde{s}(\boldsymbol{\phi})d\boldsymbol{\phi}}. \tag{6.5}$$

We evaluate (6.5) numerically to compute $\tilde{\mathsf{p}}\,(\boldsymbol{\phi}|B)$. Figure 6.2 shows the modeled probability density of recorded energy given that the photon originated from background $\tilde{\mathsf{p}}\,(\boldsymbol{\phi}|B)$ computed using the measured background intensity from Figure 6.1 and (6.5).

Figure 6.3 shows the modeled probability of the recorded energy given that it originated from the background under the assumption that the recorded energies are uniform. Use of this model will introduce model mismatch.

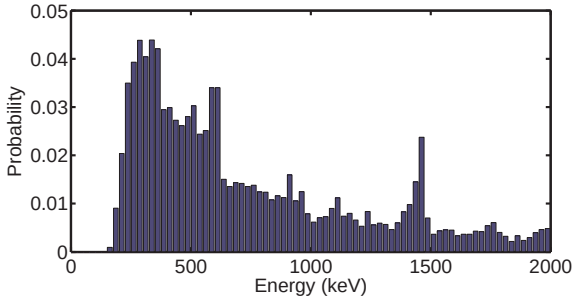


Figure 6.2: Estimated probability density function for the incident energy of recorded background photons. This density was estimated using 10,000 recorded events from the natural background.

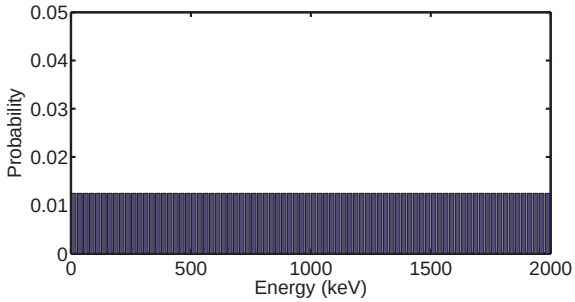


Figure 6.3: Uniform incident background energy model.

6.1.2 Detection Methods

We analyze the performance of the source intensity test (SIT) [48] applied to the gamma–ray source detection problem. The SIT is based on the quasi maximum–likelihood (QML) estimate for the source intensity. A QML estimator is equivalent to the ML estimator if the modeled distribution of the observations is equal to the true distribution. The QML estimate for the parameter vector $\boldsymbol{\theta}$ is defined as [78]

$$\tilde{\boldsymbol{\theta}}_\tau \stackrel{\triangle}{=} \arg\max_{\boldsymbol{\theta}\in\Theta} \log \tilde{\mathsf{p}}(\tilde{\boldsymbol{r}};\boldsymbol{\theta}). \tag{6.6}$$

We use the SIT because experiments showed that its performance was superior to the generalized likelihood ratio test (GLRT) when applied to simple systems [48].

In the absence of model mismatch, the parameter estimate vector $\tilde{\boldsymbol{\theta}}_\tau$ converges in probability to the true parameter values as the scan time goes to infinity. However,

when model mismatch is present, the estimates may converge to some value that is not the true parameter. For example, when there is no source present, the true source intensity is zero, but model mismatch may cause the source intensity estimate to converge to some nonzero value. To precisely define the value to which the parameter estimates converge, we first define the expected log–likelihood by

$$\tilde{f}\left(\boldsymbol{\theta}\right) \triangleq \mathsf{E}\left[\log \tilde{\mathsf{p}}\left(\tilde{\boldsymbol{r}};\boldsymbol{\theta}\right)\right] = \int_{\mathcal{R}} \log \tilde{\mathsf{p}}\left(\tilde{\boldsymbol{r}};\boldsymbol{\theta}\right) \mathsf{p}\left(\tilde{\boldsymbol{r}}\right) d\tilde{\boldsymbol{r}}, \tag{6.7}$$

where the expectation is with respect to the true distribution and $\mathcal{R}$ is the set of all lists of recorded attributes. The parameter estimates converge to the asymptotic mean, which is given by

$$\tilde{\boldsymbol{\mu}} \triangleq \arg\max_{\boldsymbol{\theta}\in\Theta} \tilde{f}\left(\boldsymbol{\theta}\right). \tag{6.8}$$

The asymptotic mean is an important component of the performance prediction method.

Let α^{t} be the true source intensity which is unknown in practice. The objective of the SIT is to determine whether a source is present (H_1) or a source is absent (H_0), i.e.,

$$\begin{aligned} H_1 &: \alpha^{\mathsf{t}} > 0 \\ H_0 &: \alpha^{\mathsf{t}} = 0. \end{aligned}$$

The source intensity test (SIT) [48] for detecting the presence of a radiation source of unknown intensity α is given by

$$\tilde{\alpha}_{\tau} \underset{H_0}{\overset{H_1}{\gtrless}} \gamma, \tag{6.9}$$

where $\tilde{\alpha}_{\tau}$ is the QMLE for α, which is the first element of $\tilde{\boldsymbol{\theta}}_{\tau}$, and γ is a threshold chosen by the user to obtain the desired false alarm rate. The user postulates that

a source is present, or H_1 is true, when the source intensity estimate $\tilde{\alpha}_\tau$ is greater than the threshold γ. The distribution of $\tilde{\alpha}_\tau$ determines the threshold value that satisfies the desired false alarm rate, but the distribution of $\tilde{\alpha}_\tau$ is intractable in the gamma–ray imaging problem.

6.1.3 Performance Measure

We state our results in terms of the probability of detection as a function of scan time in contrast to previous works that state performance prediction [48, 50, 76] in terms of ROC and the area under the ROC curve (AUC). We choose to fix a false alarm probability and examine how the probability of detection varies as a function of scan time because the probability of detection is arguably more important to the practitioner than the AUC.

6.1.4 Performance Prediction

We predict the ROC by approximating the distribution of $\tilde{\alpha}_\tau$, justified by Theorems 1 and 2 of [50], by

$$\tilde{\alpha}_\tau \overset{\text{approx}}{\sim} \mathcal{N}\left(\tilde{\mu}_{[1]}, \frac{1}{\tau}\Sigma(\tilde{\boldsymbol{\mu}})_{[1,1]}\right), \tag{6.10}$$

where $\tilde{\mu}_{[1]}$ is the first element of $\tilde{\boldsymbol{\mu}}$, which corresponds to the asymptotic mean of the source intensity. The covariance is

$$\Sigma(\boldsymbol{\theta}) \triangleq \tilde{\mathsf{H}}^{-1}(\boldsymbol{\theta})\,\tilde{\mathsf{G}}(\boldsymbol{\theta})\,\tilde{\mathsf{H}}^{-1}(\boldsymbol{\theta}), \tag{6.11}$$

where [50]

$$\begin{aligned}\tilde{\mathsf{G}}(\boldsymbol{\theta}) \triangleq \lambda_{\mathsf{s}}\mathsf{E}\Bigg[&\Big(\nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}(\boldsymbol{r};\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}} \log \tilde{\lambda}(\boldsymbol{\theta})\Big) \\ &\Big(\nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}(\boldsymbol{r};\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}} \log \tilde{\lambda}(\boldsymbol{\theta})\Big)^T\Bigg],\end{aligned} \tag{6.12}$$

and

$$\tilde{\mathsf{H}}\left(\boldsymbol{\theta}\right) \triangleq -\lambda_{\mathsf{s}} \nabla_{\boldsymbol{\theta}}^{2} \log \tilde{\lambda}(\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}}^{2} \tilde{\lambda}(\boldsymbol{\theta}) - \lambda_{\mathsf{s}} \mathsf{E}\left[\nabla_{\boldsymbol{\theta}}^{2} \log \tilde{\mathsf{p}}\left(\boldsymbol{r}; \boldsymbol{\theta}\right)\right], \tag{6.13}$$

$\nabla_{\boldsymbol{\theta}}$ is the column gradient with respect to $\boldsymbol{\theta}$, $\nabla_{\boldsymbol{\theta}}^{2}$ is the Hessian with respect to $\boldsymbol{\theta}$, λ_{s} is the true recorded count rate, and expectations are with respect to the true distribution. If the source position $\boldsymbol{\phi}$ and background intensity λ_b are known, then $\Sigma(\boldsymbol{\theta})$, $\tilde{\mathsf{G}}\left(\boldsymbol{\theta}\right)$, and $\tilde{\mathsf{H}}\left(\boldsymbol{\theta}\right)$ are scalar, otherwise they are matrices.

6.1.5 Procedure for computing predicted performance

We used the following procedure to apply the asymptotic approximation in (6.10) for detection performance prediction.

1. Obtain N recorded events with a source placed d meters from the detector at an azimuth angle ϕ_s and polar angle θ_s in a coordinate system centered at the detector, and $(\phi_s = 0°, \theta_s = 90°)$ is the vector pointing from the front of the detector.

2. Evaluate the asymptotic mean $\tilde{\mu}_{[1]}$ by solving (6.6) using the recorded events.

3. Evaluate the asymptotic covariance $\Sigma(\tilde{\boldsymbol{\mu}})$ using (6.11), (6.12), and (6.13). Use Monte Carlo integration with the recorded data to evaluate the expectations, e.g., for i.i.d. random variables X_i, $i = 1, 2, \ldots, N$ with density functions $p(x)$ and a function $f(x)$,
$$\mathsf{E}\left[f(x)\right] \approx \frac{1}{N} \sum_{i=1}^{N} f(X_i).$$

4. Compute the true count rate λ_{s} by averaging the count rate over two hours.

5. Obtain N recorded events with the detector in the same position as step 1 but without a source present.

6. Repeat steps 2-4 in the absence of a source.

7. Use the computed asymptotic means and covariances to compute the approximate distribution of $\tilde{\alpha}_\tau$ with and without a source present.

8. Use the approximate distributions to compute the probabilities of false alarm and detection.

In this work, we use $N = 10,000$, $d = 1.83$m, $\phi_s = 0°$, and $\theta_s = 90°$.

6.1.6 Source Intensity Variation

We recorded data with and without a source present to predict detection performance. We achieved the desired source–to–background ratio by combining events from the measurements with and without a source. For example, the measurements of a Cs-137 source placed 1.83 meters from the detector contain approximately 48% source events and 52% background events. We observed the recorded background count rate λ_b to be approximately 9114 counts per minute using a measurement of 90.7 hours. The observed count rate of the measurement containing both source and background events is 18984 counts per minute, which we obtained from a measurement of 121 minutes. We combined events from the lists obtained with and without a source explore a range of source–to–background ratios.

6.1.7 Conventional Method for Empirical Calculations

We compared the predicted performance to the empirical performance in terms of probability of detection, or equivalently, ROC. This method is similar to the method used in [76], [50], and [48]. We used the following procedure compute the empirical performance:

1. Repeat N_e times with a source present and N_e times without a source present:

(a) Draw the number of recorded source counts from a Poisson distribution with mean $\tau\alpha^{\mathsf{t}}s^{\mathsf{t}}$, where α^{t} is the true source intensity, s^{t} is the true sensitivity, and $\tau\alpha^{\mathsf{t}}s^{\mathsf{t}}$ is the mean number of received counts from the source.

(b) Draw the number of recorded background counts from a Poisson distribution with mean equal to the mean number of recorded background counts $\tau\lambda_b$.

(c) Generate a list of events that contains $\tau\alpha^{\mathsf{t}}s^{\mathsf{t}}$ source counts and $\tau\lambda_b$ background counts using the recorded events in the presence and absence of a source. Combine the events from the measurements with and without a source to achieve the proper mean number of source counts.

(d) Solve (6.6) using the list generated in the previous step.

2. Use the empirical source intensity estimates obtained in step 4 to compute the empirical probabilities of detection and false alarm.

We computed the empirical performance by emulating $N_e = 100$ scans with a source present and $N_e = 100$ scans without a source present.

This empirical calculation method requires the emulation of 200 scans for each point on the graph of probability of detection versus scan time. In contrast, the asymptotic prediction method based on (6.10) requires one computation with approximately 20,000 recorded events for all scan times because the asymptotic mean is invariant to scan time and the asymptotic covariance in (6.11) scales as the inverse of the scan time. Thus, computing detection performance as a function of scan time with the asymptotic prediction method requires much less computation than computing the performance empirically.

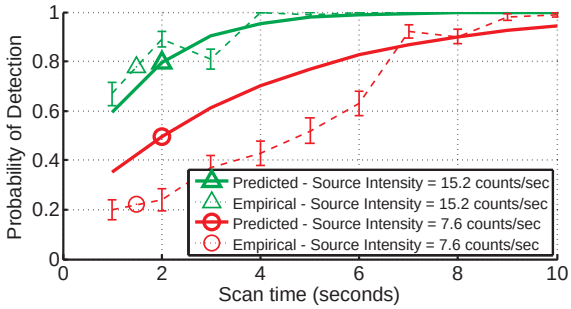
6.2 Results

We computed the probability of detection as a function of scan time when the source position, source energy, and background spectrum are assumed known. We examined the case of a known and unknown background intensity and investigated how inaccuracies in the background spectral model affect detection performance. We found that the performance predicted using (6.10) agreed well with the empirical performance in the known background case. In the unknown background case, the agreement between the theoretical and empirical predictions is not as good as in the known background case.

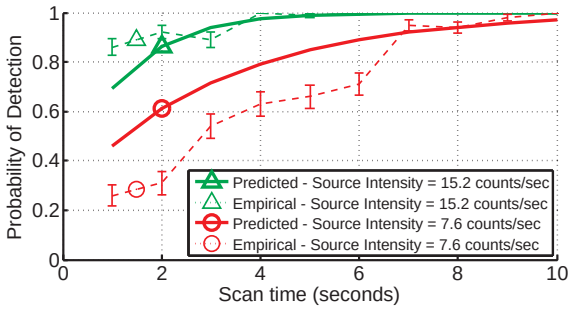
6.2.1 Cs-137 with measured background spectrum and known background intensity

We first examined the problem of detecting a Cs-137 source when the background spectrum is modeled by the measured background spectrum in Figure 6.2, and the background intensity is assumed known. Figure 6.4 shows the probability of detection as a function of scan time for probability of false alarms of 5% and 10% with source intensities of 7.6 counts per second and 15.2 counts per second. We use large false alarm rates so the system can achieve a probability of detection near one in less than ten seconds. One could apply this theory to obtain predicted performance for lower false alarm rates. The background intensity is 152 counts per second. The agreement between the empirical and predicted probability of detection is better with the higher source intensity for both false alarm rates. This is likely due to the fact that the Gaussian approximation for the distribution of the source intensity estimate (6.10) improves as the number of recorded counts increases [50].

Figure 6.5 shows the probability of detection versus scan time for source intensities of 7.6 and 15.2 counts per second with a false alarm rate of 10%. The three



(a) Probability of false alarm: 5%



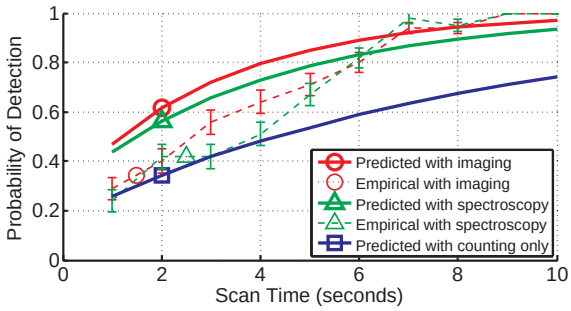
(b) Probability of false alarm: 10%

Figure 6.4: Probability of detection versus scan time for detecting a Cs-137 source in a natural background with intensity 152 counts per second using an 18 detector CdZnTe array with various false alarm rates and source intensities. The background shape and intensity are assumed known and the background shape is modeled using a prior spectral measurement.

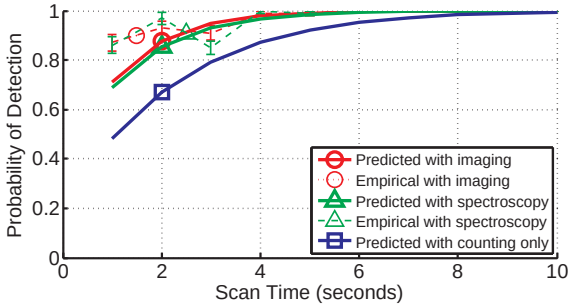
curves in each graph correspond to the performance achieved when using all imaging and spectral information, spectral information only, and the number of recorded counts only. The detector used in this experiment records positions and energies of all interacting photons, and the performance using all imaging information takes all of this information into account. We obtained the performance using spectral information by only parameterizing the source energy and not its position. We eliminated the parameterization with respect to position by numerically integrating the likelihood with respect to position over the sphere surrounding the detector. The spectral information case essentially disregards the information from the recorded positions in the detector. We computed the performance using counting only by considering the Gaussian approximation to the distributions of the source intensity estimates obtained using only the number of recorded counts. The counting case essentially disregards all information except the number of recorded counts.

Figure 6.5 shows that using imaging information results in the best detection performance with both source intensities. The increase in probability of detection over the spectral and counting cases is more when the source intensity is lower. For both source intensities, the performance difference between the case of spectral information and imaging information is small. The estimation problem is much simpler when only spectral information is used. In this experiment, there are 324 position bins and 80 energy bins, so the imaging case requires 324 times more computation. There may be practical applications where the slight decrease in performance is justified by the reduced computation when the source position and energy are known.

As in Figure 6.4, the agreement between the predicted and empirical performance is better for the higher source intensity for small scan times. As scan time increases, the empirical and predicted performance agree well.

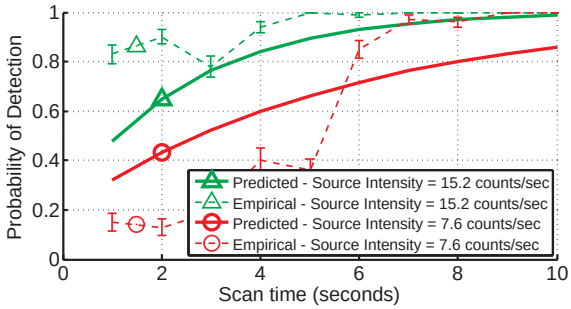


(a) Source intensity: 7.6 counts/sec

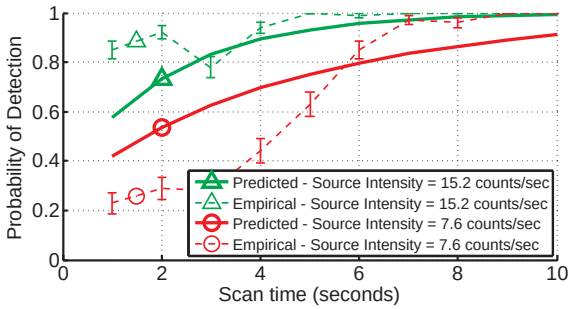


(b) Source intensity: 15.2 counts/sec

Figure 6.5: Probability of detection versus scan time for detecting a Cs-137 source in a natural background with intensity 152 counts per second at a false alarm rate of 10% using an 18 detector CdZnTe array with imaging and spectral information, spectral information only, and counting information only. The background shape and intensity are assumed known and the background shape is modeled using a prior spectral measurement.



(a) Probability of false alarm: 5%



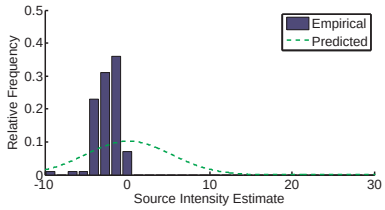
(b) Probability of false alarm: 10%

Figure 6.6: Probability of detection versus scan time for detecting a Cs-137 source in a natural background with intensity 152 counts per second using an 18 detector CdZnTe array with various false alarm rates and source intensities. The background shape is assumed known and is spectral measurement, but the background intensity is assumed unknown.

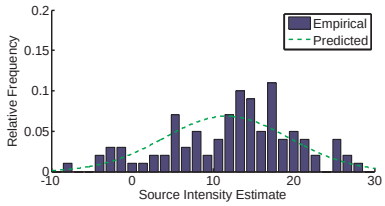
6.2.2 Cs-137 with measured background spectrum and unknown background intensity

We also examined the problem of detecting a Cs-137 source in background when the background spectrum is modeled by the measured background spectrum in Figure 6.2, but the background intensity is unknown and estimated. Figure 6.6 shows the probability of detection as a function of scan time with a background intensity of 152 counts per second. The predicted and empirical performance values are lower than in the known background case of Figure 6.4, which is expected since we introduced a nuisance parameter.

The agreement between the predicted and empirical probability of detection is poorer in the unknown background case than in the known background case. Figure 6.7 shows histograms of the empirical source intensity estimates with and without



(a) Source Absent (H_0)

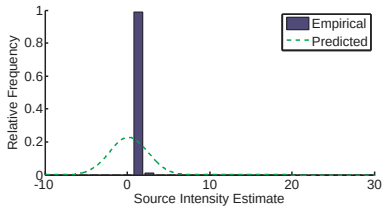


(b) Source Present (H_1)

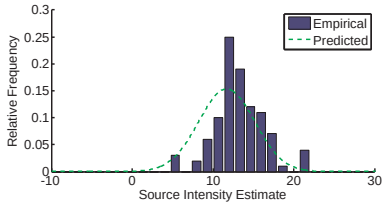
Figure 6.7: Histograms of source intensity estimates and predicted distributions for scan time $\tau = 2$ and source intensity $\alpha^{\mathrm{t}} = 15.2$

a source present when the scan time $\tau = 2$ and the source intensity is 15.2 counts per second. The empirical performance is based on the empirical histogram and the predicted performance is based on the predicted distribution. The histograms in Figure 6.7 do not agree well with the predicted distributions because the scan time is too small for the predicted distribution, to be accurate. The accuracy of the predicted distribution increases as scan time increases because it is based on an asymptotic approximation [50].

Figure 6.8 shows empirical histograms and predicted distributions when the scan time $\tau = 10$. The distribution agrees better under H_1 than it did with scan time $\tau = 2$, but the histogram under H_1 shows that the empirical distribution has a lower variance than predicted, causing the empirical performance to be better than the predicted performance. The poor agreement between the empirical and predicted distributions is likely because the predicted distribution is less accurate for small scan times when the background intensity is unknown. Our results only use 100 scans with and without a source, and one may obtain better agreement by using a



(a) Source Absent (H_0)



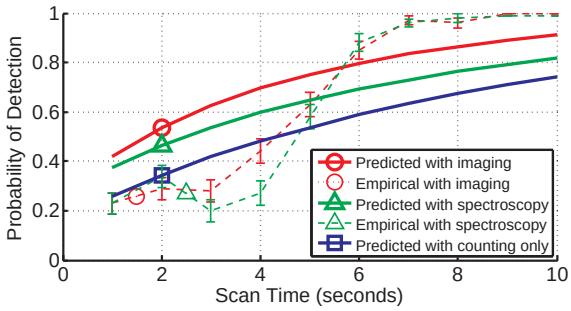
(b) Source Present (H_1)

Figure 6.8: Histograms of source intensity estimates and predicted distributions for scan time $\tau = 10$ and source intensity $\alpha^{\mathrm{t}} = 15.2$

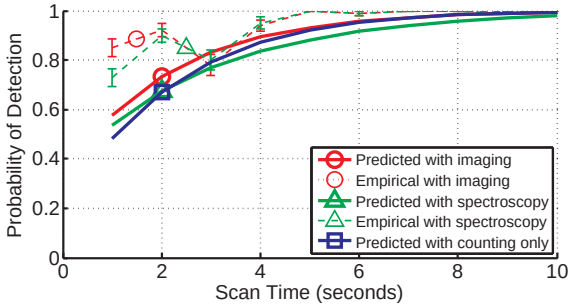
larger number of trial scans.

Figure 6.9 shows the probability of detection as a function of scan time for a probability of false alarm of 10% with position, spectral, and counting information. In the unknown background case, the predicted performance using counting statistics assumes a known background intensity because the source and background intensities cannot be estimated otherwise. The performance using imaging and spectral information in an *unknown* background is better than the performance of the counting method, even though the counting method assumes a *known* background.

The difference in performance between the imaging, spectroscopy, and counting cases is more pronounced with the weaker source in Figure 6.9a. Even when the source intensity is 15.2 counts per second in Figure 6.9b, the performance with imaging information in an unknown background is still better than the performance using counting statistics only in a known background. However, the performance using spectroscopy only is worse than the performance using counting statistics only when the source–to–background ratio is high.



(a) Source intensity: 7.6 counts/sec



(b) Source intensity: 15.2 counts/sec

Figure 6.9: Probability of detection versus scan time for detecting a Cs-137 source in a natural background with intensity 152 counts per second at a false alarm rate of 10% using an 18 detector CdZnTe array with imaging and spectral information, spectral information only, and counting information only. The background shape is assumed known and is modeled using a prior spectral measurement, but the background intensity is assumed unknown.

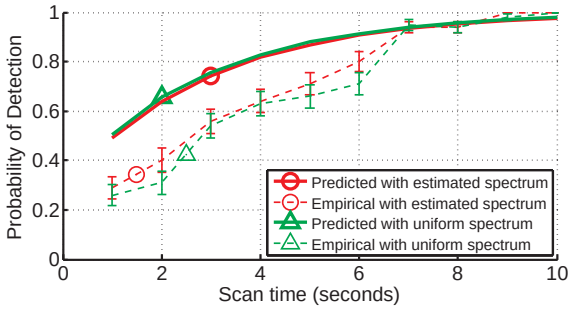


Figure 6.10: Probability of detection vs. scan time with an estimated background spectral model and a uniform background spectral model. The background intensity is known, the source intensity is 7.6 counts per second, the background intensity is 152 counts per second, and the probability of false alarm is 10%

6.2.3 Effect of incorrect modeled background spectrum

We also explored the robustness of the SIT to changes in the modeled background spectrum. The background spectrum may change over time or if the detector is moved to a new location, making it difficult to acquire a background spectrum estimate that will be accurate at all times. Figure 6.11 shows the probability of detection as a function of scan time when the background spectrum is modeled using the estimated spectrum in Figure 6.2 and the uniform spectrum in Figure 6.3 when the background intensity is known.

Figure 6.11 shows that the predicted performance does not change significantly when the background spectral model is changed. This is likely because the source energy is known to be 662 keV.

Figure 6.11 shows the probability of detection as a function of scan time for the estimated and uniform background models with an unknown background intensity. The difference between the predicted performance curves is also negligible in this case. However, the empirical predictions are better when the estimated spectral model is used and the scan time is small, causing the predicted performance to be a poor approximation.

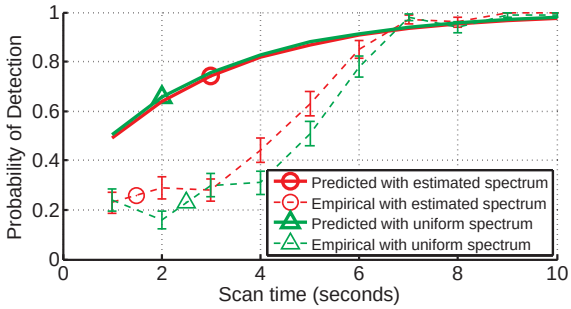


Figure 6.11: Probability of detection vs. scan time with an estimated background spectral model and a uniform background spectral model. The background intensity is unknown, the source intensity is 7.6 counts per second, the background intensity is 152 counts per second, and the probability of false alarm is 10%

The comparison between the estimated and uniform spectral models shows that the SIT is robust to the background spectral model when the source position and isotope are known.

6.3 Conclusions for Real Data Analysis

In this chapter, we applied the asymptotic detection performance prediction method developed in Chapter V to performance prediction of a real system with real recorded data. Our results showed that the asymptotic prediction method accurately predicts detection performance for modest scan times. The prediction accuracy was higher when the background intensity is known, which is in contrast to the simulation results in Chapter V, where the prediction method worked nearly well when the background intensity was unknown.

We also found that the array of 18 CdZnTe detectors outperformed a counting detector of equal sensitivity *even when the background intensity is unknown.* We showed that the imaging information, or parameterization of the source direction relative to the detector, improved detection performance more when the source intensity is low.

This chapter serves as an example application of the theory of Chapter V to

performance characterization of real systems. Our results show that the asymptotic performance prediction method gives reasonably accurate performance predictions that one can use to determine sensor placement, configuration, or viability.

CHAPTER VII

Summary Future Work

7.1 Summary

This work summarizes our research on image reconstruction and source detection with position–sensitive gamma–ray detectors. We investigated and proposed new methods for image reconstruction and source detection performance prediction. These contributions address the need for algorithms for detecting and locating radiation sources with position–sensitive gamma–ray detectors.

We first examined a penalized–likelihood approach to image reconstruction with position–sensitive detectors. We investigated penalty functions that encourage sparsity in the pixel basis because radiation sources are typically small in space. We found that the standard penalty functions used with Gaussian data do not yield useful solutions with well–modeled list–mode data from gamma–ray detectors. We proposed a penalty function based on the sum of the logarithms of the pixel values that proved to yield images that are sparse in the pixel basis. We also found that these images are difficult to use for detecting a source, which motivated our study of detection performance.

We used asymptotic analysis, which employs statistical approximations when the scan time is large, to quantify the benefit of position–sensitive capability and approx-

imate detection performance. To compare position–sensitive detectors to their less–costly photon–counting competitors, we used asymptotics to prove that position–sensitive capability always improves detection performance when the scan time is large and the detector sensitivity is uniform. When sensitivity is nonuniform, we provided expressions that allow the practitioner to compare the performance of two detectors.

We also used asymptotic analysis to provide a method for detection performance prediction that is more computationally efficient than empirical performance prediction. Our method accounts for model mismatch because position–sensitive detectors are typically difficult to model accurately and in a computationally efficient manner. The method is based on an extension [78] to the case of a Poisson number of measurements. We showed that the asymptotic prediction method reasonably agrees with empirical calculations with simulated data. We also showed that the asymptotic prediction method gives reasonable performance predictions for a real system using real recorded data.

7.2 Future Work

This section outlines some ideas that we did not have time to explore and may prove fruitful in future work.

7.2.1 Unknown source position

The experiments in this work assume that the source position is known. The theory developed in Chapters IV and V is general enough to accommodate the unknown position case when the parameter space is continuous. The parameter space is naturally continuous in the gamma–ray imaging problem, but it is typically coarsely discrete in practice due to computational limitations. Future work would investigate

the accuracy of performance prediction when derivatives are taken numerically in the discrete case and when the derivatives of the system response function with respect to position are calculated analytically.

7.2.2 Unknown isotope

The experiments in this work assume that the source isotope is known. The theory developed in this work is not general enough to account for a discrete parameter space. The unknown isotope case is different from the unknown position case because the set of possible isotopes is not a set of samples from a corresponding continuous space. The asymptotic distribution of the isotope intensity is uninteresting because the isotope estimate converges in distribution to a degenerate random variable. Future work would investigate another method of approximating the discrete distribution of the isotope estimates for finite sample sizes.

7.2.3 Sensor networks

The experiments in this work focused on systems where the sensors are concentrated in one location, but the theory is general enough to analyze distributed networks of sensors. In this case the recorded attributes could include a sensor's identifier and geographic coordinates in addition to any photon interaction attributes recorded by the sensor. Future work would use the asymptotic performance prediction method to show how detection performance scales as the number of sensors increases and how performance varies as a function of sensor position.

7.2.4 Moving source or detector

The experiments in this work assume that the detector is stationary, but the theory is general enough to analyze scenarios where the source or detector is moving. In this case, the recorded attributes would include the detector and source position,

assuming the source position is estimated by a separate system, such as an optical camera. Future work would examine how detection performance varies as a function of the relative speed between the source and the detector.

7.2.5 Additional sources of model mismatch

We used an incorrect background shape model as a source of model mismatch in our simulated results. Model mismatch is present in the real data results because of the simple background model and approximation of the detector response. Future work will explore other sources of model mismatch, such as mismatch in the source shape, or a source position that is assumed known but is incorrect.

7.2.6 Improvements to system model

Computational constraints prevent use of the exact system model for two–interaction events in 3D position–sensitive semiconductor detectors in (2.21). The model proposed in [80] has been shown to be useful for image reconstruction and source detection [75].

In §2.4.3, the exact model is derived by conditioning on the event the photon interacts twice in the detector. The model in [80] does not condition upon this event, making it an incomplete model. Future work will extend the model in [80] to condition on the event that the photon interacts twice in the detector by calculating $\mathsf{p}(D_2|D_1; e_0, \boldsymbol{\phi})$ in (2.7) numerically or by simulation. This calculation could be performed on a grid of source positions and source energies. The values of $\mathsf{p}(D_2|D_1; e_0, \boldsymbol{\phi})$ do not depend on the attributes of a single event, so one can compute it once for a particular detector. The goal of this investigation would be to show that conditioning on the event that a two–interaction event is detected when computing the attribute density for two–interaction events increases detection performance by reducing model

mismatch.

7.2.7 Machine Learning Methods for Source Detection

Likelihood–based methods for source detection with 3D position–sensitive semiconductor gamma–ray detectors do not perform optimally due to model mismatch. An alternative approach to source detection is to employ classification methods from machine learning to decide whether or not a source is present. Machine learning methods use training data to formulate an optimal decision boundary between data with and without a source, whereas the model determines the decision boundary in likelihood–based methods. Machine learning classification algorithms can also be faster than likelihood–based algorithms because machine learning algorithms do not compute the often nonlinear and computationally intensive likelihood function.

There are several challenges to applying machine learning classification algorithms to the gamma–ray imaging problem. The background energy spectrum is often unknown in practice, leading to considerable variability in data recorded in the absence of a source. The background spectrum can vary with both location and time. The sources of interest are also highly variable. There are many types of material that one may wish to detect, each with different energy spectra, size, position relative to the detector, and intensity.

Extracting useful features from the recorded data that are invariant to nuisance parameters such as source position is a challenging open problem. Transforming the list of recorded interaction positions and energies into a reconstructed image involves highly nonlinear operations that are based on physics. Attempting to classify data based solely on the list of recorded positions and energies would be challenging. Future work would investigate feature extraction from images reconstructed using backprojection. Reconstructing simple images from the data greatly reduces the

dimensionality of the data for modest amounts of recorded events, which is advantageous for feature extraction and classification.

A preliminary study on application of the support vector machine (SVM) to backprojected images recorded with a 3D position–sensitive CdZnTe detector led us to consider the angular power spectral coefficients as features because they are invariant to the source position. The angular power spectrum is given by

$$g(l) = \sum_{m=l}^{l} |f(l,m)|^2 \quad l = 0, 1, \ldots, L, \tag{7.1}$$

where $f(l,m)$ is the spherical harmonic basis function of degree l and order m [19]. We chose the angular power spectrum as a feature vector because it is invariant to rotation [36], and thus invariant to source position.

To show that angular power coefficients are useful features, we created histograms of the angular power coefficients with and without a source present using 2000 test images with a source present and 2000 images with background only. We generated data with random source intensities from 10 to 100 counts per scan, uniformly random source positions, and random background intensities of 10 to 100 counts per scan. The detector was a simulated single 2cm $\times$ 2cm $\times$ 1.5cm 3D position–sensitive CdZnTe detector with 11×11 pixels on the anode. The simulation was performed using GEANT4 [1]. The source energy was 662 keV and the background had the measured soil spectrum given in [65]. The images were sampled uniformly in azimuth and polar angles to produce images of 36×36 pixels. We used the model in [80] to compute the backprojection operator.

Perhaps the simplest feature of each image is the total number of photons recorded during the scan that produced the image. Figure 7.1 shows a histogram of the total number of counts in images with and without a source present.

Figure 7.2 shows histograms of angular power coefficients $g(l)$ for $l = 0, 1, 2, 3$.

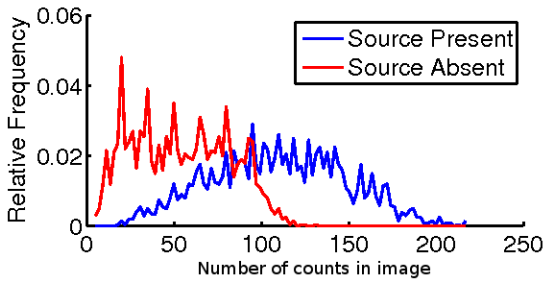


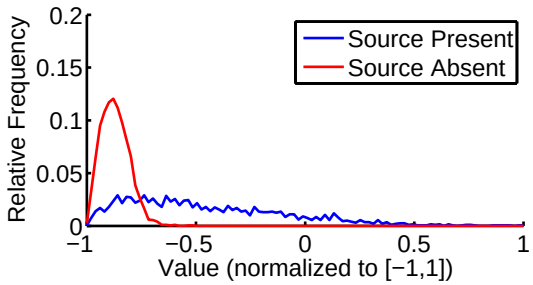
Figure 7.1: Histogram of total counts received using 2000 realizations with and without a source present. The source and background intensities are randomly chosen from a uniform distribution on [10,100], and the source positions are uniformly distributed on the sphere.

The data is still not separable with these features, but there is less overlap between the histograms with and without a source for $l = 2$ and 3. However, there is no guarantee that these features are optimal. Future work would apply the method of [43] to determine features in backprojected images that give better classification performance. The first step would be to learn features from backprojected images because their reconstruction involves fewer nonlinear operations than likelihood–based methods.

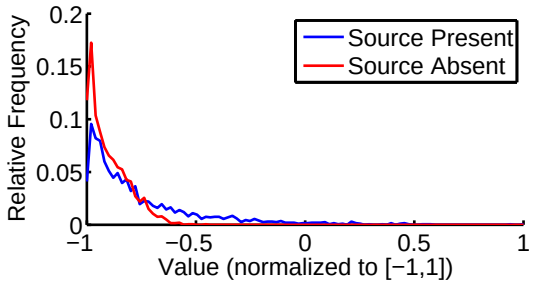
7.2.8 Confidence measures for reconstructed images

We found in Chapter III that reconstructed images of radiation emission density do not give clear information about the presence or absence of a source, especiallly with regularization. Perhaps the images would be more useful from a source detection standpoint if one could compute a confidence in the accuracy of the reconstructed image.

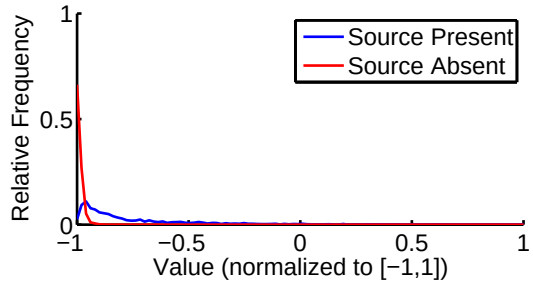
Bootstrap resampling methods for estimating the mean and standard deviation of maximum–likelihood PET images are propsed in [14] and [27]. These methods are based on resampling the list–mode data and reconstructing and image from each resampled list. A similar approach may be useful for variance prediction of maximum likelihood and penalized likelihood Compton images.



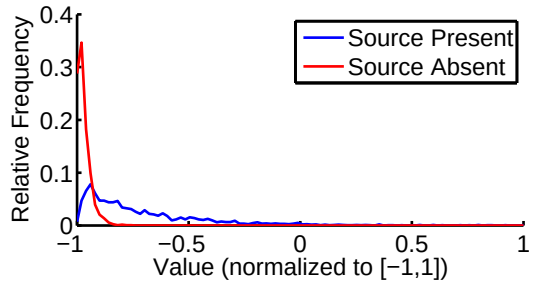
(a) $g(0)$



(b) $g(1)$



(c) $g(2)$



(d) $g(3)$

Figure 7.2: Histograms of angular power coefficients using 2000 realizations with and without a source present. The source and background intensities are randomly chosen from a uniform distribution on [10,100], and the source positions are uniformly distributed on the sphere.

A review of several resampling methods for variance PET imaging is given in [41]. The study considers resampling of list mode data similar to [14] and [27], but it also considers resampling of the sinogram. In Compton imaging, the number of possible recorded measurements is extremely large, so a sinogram would be difficult to compute and store.

7.2.9 Use of reconstructed images for source detection

We found in Chapter III that the images reconstructed using log regularization with and without a source were indistinguishable. However, if one increases the regularization parameter β enough, the entire reconstructed image will be near zero. Future work would investigate the use of cross-validation to select a regularization parameter β that would give the optimal images for source detection. A possible test statistic would be the maximum intensity in the image.

APPENDICES

APPENDIX A

Equivalence of Imaging Likelihoods

To show the equivalence between the list–mode and binned–mode likelihood for image reconstruction, we first compute the probability mass function of the attribute vector for a single recorded photon. In this Appendix, we assume that the detector measurements are discrete, which is a valid assumption in practice due to machine quantization.

We assume a list–mode detector that records a random number of photons $J \sim \text{Poisson}(\bar{J}(\boldsymbol{x}))$ and that there are n_d elements in the discrete measurement space. When viewed in binned mode, the detector has n_d measurement bins.

By the law of total probability,

$$\begin{aligned}
\mathsf{p}\left(\boldsymbol{r}_i|D;\boldsymbol{x}\right) &= \mathsf{p}\left(\boldsymbol{r}_i|S,D;\boldsymbol{x}\right)\mathsf{Pr}\left(S|D\right) + \mathsf{p}\left(\boldsymbol{r}_i|B,S;\boldsymbol{x}\right)\mathsf{Pr}\left(B|D\right) \\
&= \sum_{j=1}^{n_p} \mathsf{p}\left(\boldsymbol{r}_i|E_j,S,D\right)\mathsf{p}\left(E_j|S,D;\boldsymbol{x}\right)\mathsf{Pr}\left(S|D\right) + \mathsf{p}\left(\boldsymbol{r}_i|B,D;\boldsymbol{x}\right)\mathsf{Pr}\left(B|D\right) \\
&= \sum_{j=1}^{n_p} \frac{\mathsf{p}\left(\boldsymbol{r}_i,D|E_j,S\right)}{\mathsf{p}\left(D|E_j,S\right)}\mathsf{p}\left(E_j|S,D;\boldsymbol{x}\right)\mathsf{Pr}\left(S|D\right) + \mathsf{p}\left(\boldsymbol{r}_i|B,D;\boldsymbol{x}\right)\mathsf{Pr}\left(B|D\right). \qquad (1.1)
\end{aligned}$$

By a counting argument and (3.3), we have

$$\mathsf{Pr}\left(S|D\right) = \frac{\sum_{j=1}^{n_p} s_j x_j}{\bar{J}(\boldsymbol{x})}, \qquad (1.2)$$

and

$$\mathsf{Pr}\left(B|D\right) = \frac{\sum_{i=1}^{n_d} \bar{r}_i}{\bar{J}(\boldsymbol{x})}. \qquad (1.3)$$

Note that the sum in (1.3) is over all possible detector measurement bins. Also by counting,

$$\mathsf{p}\left(E_j|S,D;\boldsymbol{x}\right)=\frac{s_j x_j}{\sum_{j=1}^{n_p} s_j x_j}, \tag{1.4}$$

and

$$\mathsf{p}\left(\boldsymbol{r}_i|B,D;\boldsymbol{x}\right)=\frac{\bar{r}_i}{\sum_{i=1}^{n_d}\bar{r}_i}. \tag{1.5}$$

By the definition of the sensitivity s_j,

$$\mathsf{p}\left(D|E_j,S\right)=s_j. \tag{1.6}$$

Substituting (3.1) and (1.2)–(1.5) into (1.1),

$$\mathsf{p}\left(\boldsymbol{r}_i|D;\boldsymbol{x}\right)=\left(\sum_{j=1}^{n_p} a_{ij}x_j+\bar{r}_i\right)/\bar{J}(\boldsymbol{x}). \tag{1.7}$$

The index i enumerates all possible recorded attributes in the attribute space. In list–mode, let $i(k)$ be the index i corresponding to the kth observed event, where $k=1,2,\ldots,J$. Substituting the density of a single attribute in (1.7) into the likelihood (3.2)and taking the natural logarithm,

$$\begin{aligned}
\mathsf{L}\left(\boldsymbol{x}\right) &= \log\mathsf{p}\left(\tilde{\boldsymbol{r}};\boldsymbol{x}\right)=J\log\bar{J}(\boldsymbol{x})-\bar{J}(\boldsymbol{x})-\log J! \\
&\quad+\sum_{k=1}^{J}\log\left(\left(\sum_{j=1}^{n_p}a_{i(k)j}x_j+\bar{r}_{i(k)}\right)/\bar{J}(\boldsymbol{x})\right) \\
&= -\bar{J}(\boldsymbol{x})-\log J!+\sum_{k=1}^{J}\log\left(\sum_{j=1}^{n_p}a_{i(k)j}x_j+\bar{r}_{i(k)}\right) \\
&\equiv -\sum_{j=1}^{n_p}s_j x_j-\sum_{i=1}^{n_d}\bar{r}_i+\sum_{k=1}^{J}\log\left(\sum_{j=1}^{n_p}a_{i(k)j}x_j+\bar{r}_{i(k)}\right) \\
&= \sum_{i=1}^{n_d}\left[-\sum_{j=1}^{n_p}a_{ij}x_j-\bar{r}_i+y_i\log\left(\sum_{j=1}^{n_p}a_{ij}x_j+\bar{r}_i\right)\right], \qquad (1.8)
\end{aligned}$$

where y_i is the number of photons having attributes corresponding to the ith detector measurement bin if the detector were operated in binned–mode, and

$$\sum_{i=1}^{n_d} a_{ij} = \sum_{i=1}^{n_d} \mathsf{p}\left(\boldsymbol{r}_i, D|E_j, S\right) = \mathsf{p}\left(D|E_j, s\right) = s_j.$$

In a typical list–mode case, the number of detector bins n_d is very large, so y_i is either 0 or 1 with high probability. The equivalence still holds if $y_i > 1$. By inspection, the likelihood in (1.8) is the log–likelihood corresponding to the model

$$y_i \sim \text{Poisson}\left([\boldsymbol{A}\boldsymbol{x}]_i + \bar{r}_i\right).$$

APPENDIX B

Asymptotics not Accounting for Model Mismatch

B.1 Proof of Theorem IV.3

Proof. We first show the inequality (4.12). By the block matrix inversion formula applied to the the Fisher information matrix (4.9),

$$(\mathsf{F}^{-1}(\boldsymbol{\theta})_{[1,1]})^{-1} = \mathsf{F}_{\mathrm{c}}(\boldsymbol{\theta}) \left(K_{[1,1]} - K_{[2,1]}^T K_{[2,2]}^{-1} K_{[2,1]} \right), \tag{2.1}$$

thus, to show (4.12), it suffices to show that

$$\left((K_{[1,1]} - 1) - K_{[2,1]}^T K_{[2,2]}^{-1} K_{[2,1]} \right) > 0. \tag{2.2}$$

To simplify notation, we introduce the following shorthand: $\mathsf{p} = \mathsf{p}(\boldsymbol{r}|D;\boldsymbol{\theta})$ and $\mathsf{p}_{\mathrm{S}} = \mathsf{p}_{\mathrm{S}}(\boldsymbol{r}|D;\boldsymbol{\phi})$. Since the sensitivity is uniform, let $s(\boldsymbol{\phi}) = s_0$ for all $\boldsymbol{\phi} \in \Phi$. Now,

$$\begin{aligned}
K_{[1,1]} - 1 - K_{[2,1]}^T K_{[2,2]}^{-1} K_{[2,1]} &= K_{[1,1]} - 1 - 2K_{[2,1]}^T K_{[2,2]}^{-1} K_{[2,1]} + K_{[2,1]}^T K_{[2,2]}^{-1} K_{[2,1]} \\
&= (K_{[1,1]} - 1) - \mathsf{E}\left[\frac{\mathsf{p}_{\mathrm{S}} \nabla_{\boldsymbol{\phi}}^T (s_0 \mathsf{p}_{\mathrm{S}})}{\mathsf{p}^2} \right] K_{[2,2]}^{-1} K_{[2,1]} \\
&\quad - K_{[2,1]}^T K_{[2,2]}^{-1} \mathsf{E}\left[\frac{\mathsf{p}_{\mathrm{S}} \nabla_{\boldsymbol{\phi}} (s_0 \mathsf{p}_{\mathrm{S}})}{\mathsf{p}^2} \right] \\
&\quad + K_{[2,1]}^T K_{[2,2]}^{-1} \mathsf{E}\left[\frac{\nabla_{\boldsymbol{\phi}} (s_0 \mathsf{p}_{\mathrm{S}}) \nabla_{\boldsymbol{\phi}}^T (s_0 \mathsf{p}_{\mathrm{S}})}{\mathsf{p}^2} \right] K_{[2,2]}^{-1} K_{[2,1]}.
\end{aligned} \tag{2.3}$$

Using the fact that

$$\begin{aligned} K_{[1,1]} - 1 &= \mathsf{E}\left[\frac{\mathsf{p}_{\mathrm{S}}^2}{\mathsf{p}^2}\right] - 2\mathsf{E}\left[\frac{\mathsf{p}_{\mathrm{S}}}{\mathsf{p}}\right] + \mathsf{E}\left[\frac{\mathsf{p}^2}{\mathsf{p}^2}\right] \\ &= \mathsf{E}\left[\frac{(\mathsf{p}_{\mathrm{S}} - \mathsf{p})^2}{\mathsf{p}^2}\right], \end{aligned}$$

we can rewrite (2.3) as

$$\begin{aligned} K_{[1,1]} - 1 - K_{[2,1]}^T K_{[2,2]}^{-1} K_{[2,1]} &= \mathsf{E}\left[\frac{(\mathsf{p}_{\mathrm{S}} - \mathsf{p})^2}{\mathsf{p}^2}\right] \\ &\quad -\mathsf{E}\left[\frac{(\mathsf{p}_{\mathrm{S}} - \mathsf{p})\nabla_{\boldsymbol{\phi}}^T(s_0\mathsf{p}_{\mathrm{S}})}{\mathsf{p}^2} K_{[2,2]}^{-1} K_{[2,1]}\right] \\ &\quad -\mathsf{E}\left[\left(\frac{\nabla_{\boldsymbol{\phi}}^T(s_0\mathsf{p}_{\mathrm{S}})}{\mathsf{p}} K_{[2,2]}^{-1} K_{[2,1]}\right)^T \frac{\mathsf{p}_{\mathrm{S}} - \mathsf{p}}{\mathsf{p}}\right] \\ &\quad +\mathsf{E}\left[\left(\frac{\nabla_{\boldsymbol{\phi}}^T(s_0\mathsf{p}_{\mathrm{S}})}{\mathsf{p}} K_{[2,2]}^{-1} K_{[2,1]}\right)^T \right. \\ &\quad \left. \left(\frac{\nabla_{\boldsymbol{\phi}}^T(s_0\mathsf{p}_{\mathrm{S}})}{\mathsf{p}} K_{[2,2]}^{-1} K_{[2,1]}\right)\right] \qquad (2.4) \\ &= \mathsf{E}\left[\left(\frac{\mathsf{p}_{\mathrm{S}} - \mathsf{p}}{\mathsf{p}} - \frac{\nabla_{\boldsymbol{\phi}}^T(s_0\mathsf{p}_{\mathrm{S}})}{\mathsf{p}} K_{[2,2]}^{-1} K_{[2,1]}\right)^2\right] \qquad (2.5) \\ &\geq 0. \qquad (2.6) \end{aligned}$$

Equality holds when $\mathsf{p}_{\mathrm{S}}(\boldsymbol{r}|D;\boldsymbol{\phi}) = \mathsf{p}(\boldsymbol{r}|D;\boldsymbol{\theta})$ because in this case,

$$\mathsf{E}\left[\left(\frac{\mathsf{p}_{\mathrm{S}} - \mathsf{p}}{\mathsf{p}}\right)^2\right] = \mathsf{E}\left[\left(\frac{\mathsf{p}_{\mathrm{S}} - \mathsf{p}_{\mathrm{S}}}{\mathsf{p}_{\mathrm{S}}}\right)^2\right] = 0, \qquad (2.7)$$

and

$$K_{[2,1]} = \mathsf{E}\left[\frac{\nabla_{\boldsymbol{\phi}}(s_0\mathsf{p}_{\mathrm{S}})}{\mathsf{p}_{\mathrm{S}}}\right] = s_0 \int_{\mathcal{R}} \nabla \mathsf{p}_{\mathrm{S}}(\boldsymbol{r}|D;\boldsymbol{\phi}) d\boldsymbol{r} = \mathbf{0}, \qquad (2.8)$$

where $\mathbf{0}$ is the zero vector, so equality is attained in (2.2). Note that (2.4) follows from (2.3) by the fact that when sensitivity is uniform,

$$\mathsf{E}\left[\frac{\mathsf{p}_{\mathrm{S}}\nabla_{\boldsymbol{\phi}}(s_0\mathsf{p}_{\mathrm{S}})}{\mathsf{p}^2}\right] = \mathsf{E}\left[\frac{(\mathsf{p}_{\mathrm{S}} - \mathsf{p})\nabla_{\boldsymbol{\phi}}(s_0\mathsf{p}_{\mathrm{S}})}{\mathsf{p}^2}\right], \qquad (2.9)$$

because

$$\mathsf{E}\left[\frac{\mathsf{p}\nabla_{\boldsymbol{\phi}}^{T}\left(s_0\mathsf{p}_{\mathrm{S}}\right)}{\mathsf{p}^2}\right] = s_0\int_{\mathcal{R}}\nabla_{\boldsymbol{\phi}}^{T}\mathsf{p}_{\mathrm{S}}(\boldsymbol{r}|D;\boldsymbol{\phi})d\boldsymbol{r} = \mathbf{0}^{T}. \tag{2.10}$$

□

B.2 Sample Derivations of Fisher Information

B.2.1 Derivation of $K_{[1,1]}$

The $[1,1]$ component of the Fisher information is given by

$$\mathsf{F}(\boldsymbol{\theta})_{[1,1]} = -\mathsf{E}\left[\frac{\partial^2}{\partial\alpha^2}\mathsf{L}\left(\boldsymbol{\theta}\right)\right],$$

where $\mathsf{L}\left(\boldsymbol{\theta}\right)$ is the log–likelihood, which is the logarithm of the likelihood in (4.4). Substituting

$$\frac{\partial^2}{\partial\alpha^2}\mathsf{L}\left(\boldsymbol{\theta}\right) = -\sum_{i=1}^{J}\frac{\left(s(\boldsymbol{\phi})\mathsf{p}_{\mathrm{S}}(\boldsymbol{r}_i|D;\boldsymbol{\phi})\right)^2}{\left(\lambda_b\mathsf{p}_{\mathrm{B}}(\boldsymbol{r}_i|D;\boldsymbol{\beta}) + \alpha s(\boldsymbol{\phi})\mathsf{p}_{\mathrm{S}}(\boldsymbol{r}_i|D;\boldsymbol{\phi})\right)^2},$$

and using (6.4) with iterated expectation yields

$$\begin{aligned}
\mathsf{F}(\boldsymbol{\theta})_{[1,1]} &= -\frac{s^2(\boldsymbol{\phi})}{(\lambda_b+\alpha s(\boldsymbol{\phi}))^2}\mathsf{E}\left[\sum_{j=1}^{J}\frac{\mathsf{p}_{\mathrm{S}}^2(\boldsymbol{r}_i|D;\boldsymbol{\phi})}{\mathsf{p}^2\left(\boldsymbol{r}_i|D;\boldsymbol{\theta}\right)}\right] \\
&= \frac{s^2(\boldsymbol{\phi})}{(\lambda_b+\alpha s(\boldsymbol{\phi}))^2}\mathsf{E}\left[J\right]\mathsf{E}\left[\frac{\mathsf{p}_{\mathrm{S}}^2(\boldsymbol{r}|D;\boldsymbol{\phi})}{\mathsf{p}^2(\boldsymbol{r}|D;\boldsymbol{\theta})}\right] \\
&= \frac{\tau s^2(\boldsymbol{\phi})}{\lambda_b+\alpha s(\boldsymbol{\phi})}\mathsf{E}\left[\frac{\mathsf{p}_{\mathrm{S}}^2(\boldsymbol{r}|D;\boldsymbol{\phi})}{\mathsf{p}^2(\boldsymbol{r}|D;\boldsymbol{\theta})}\right] \\
&= \frac{\tau s^2(\boldsymbol{\phi})}{\lambda_b+\alpha s(\boldsymbol{\phi})}K_{[1,1]} = \mathsf{F}_{\mathrm{c}}(\boldsymbol{\theta})K_{[1,1]}.
\end{aligned}$$

B.2.2 Derivation of $K_{[2,1]}$

The $[2,1]$ component of the Fisher information is given by

$$\mathsf{F}(\boldsymbol{\theta})_{[2,1]} = -\mathsf{E}\left[\nabla_{\boldsymbol{\phi}}\frac{\partial}{\partial\alpha}\mathsf{L}\left(\boldsymbol{\theta}\right)\right].$$

Using the model in (4.4),

$$\begin{aligned}
\mathsf{F}(\boldsymbol{\theta})_{[2,1]} &= \tau\nabla_{\boldsymbol{\phi}} s(\boldsymbol{\phi}) - \frac{\mathsf{E}\left[J\right]}{\lambda_b + \alpha s(\boldsymbol{\phi})}\mathsf{E}\left[\frac{\nabla_{\boldsymbol{\phi}}(s(\boldsymbol{\phi})\mathsf{p}_{\mathrm{S}}(\boldsymbol{r}|D;\boldsymbol{\phi}))}{\mathsf{p}(\boldsymbol{r}|D;\boldsymbol{\theta})}\right] \\
&\quad + \frac{\mathsf{E}\left[J\right]\alpha s(\boldsymbol{\phi})}{(\lambda_b + \alpha s(\boldsymbol{\phi}))^2}\mathsf{E}\left[\frac{\mathsf{p}_{\mathrm{S}}(\boldsymbol{r}|D;\boldsymbol{\phi})\nabla_{\boldsymbol{\phi}}(s(\boldsymbol{\phi})\mathsf{p}_{\mathrm{S}}(\boldsymbol{r}|D;\boldsymbol{\phi}))}{\mathsf{p}^2(\boldsymbol{r}|D;\boldsymbol{\theta})}\right] \\
&= \tau\nabla_{\boldsymbol{\phi}} s(\boldsymbol{\phi}) - \tau\mathsf{E}\left[\frac{\nabla_{\boldsymbol{\phi}}(s(\boldsymbol{\phi})\mathsf{p}_{\mathrm{S}}(\boldsymbol{r}|D;\boldsymbol{\phi}))}{\mathsf{p}(\boldsymbol{r}|D;\boldsymbol{\theta})}\right] \\
&\quad + \frac{\tau\alpha s(\boldsymbol{\phi})}{\lambda_b + \alpha s(\boldsymbol{\phi})}\mathsf{E}\left[\frac{\mathsf{p}_{\mathrm{S}}(\boldsymbol{r}|D;\boldsymbol{\phi})\nabla_{\boldsymbol{\phi}}(s(\boldsymbol{\phi})\mathsf{p}_{\mathrm{S}}(\boldsymbol{r}|D;\boldsymbol{\phi}))}{\mathsf{p}^2(\boldsymbol{r}|D;\boldsymbol{\theta})}\right] \\
&= \frac{\tau\alpha s(\boldsymbol{\phi})}{\lambda_b + \alpha s(\boldsymbol{\phi})}\mathsf{E}\left[\frac{\mathsf{p}_{\mathrm{S}}(\boldsymbol{r}|D;\boldsymbol{\phi})\nabla_{\boldsymbol{\phi}}(s(\boldsymbol{\phi})\mathsf{p}_{\mathrm{S}}(\boldsymbol{r}|D;\boldsymbol{\phi}))}{\mathsf{p}^2(\boldsymbol{r}|D;\boldsymbol{\theta})}\right] \\
&= \frac{\alpha\mathsf{F}_{\mathrm{c}}(\boldsymbol{\theta})}{s(\boldsymbol{\phi})}K_{[2,1]},
\end{aligned}$$

where the third step follows because

$$\begin{aligned}
\mathsf{E}\left[\frac{\nabla_{\boldsymbol{\phi}}(s(\boldsymbol{\phi})\mathsf{p}_{\mathrm{S}}(\boldsymbol{r}|D;\boldsymbol{\phi}))}{\mathsf{p}(\boldsymbol{r}|D;\boldsymbol{\theta})}\right] &= \int_{\mathcal{R}} \nabla_{\boldsymbol{\phi}} s(\boldsymbol{\phi})\mathsf{p}_{\mathrm{S}}(\boldsymbol{r}|D;\boldsymbol{\phi})d\boldsymbol{r} \\
= \nabla_{\boldsymbol{\phi}} s(\boldsymbol{\phi})\int_{\mathcal{R}} \mathsf{p}_{\mathrm{S}}(\boldsymbol{r}|D;\boldsymbol{\phi})d\boldsymbol{r} &= \nabla_{\boldsymbol{\phi}} s(\boldsymbol{\phi}),
\end{aligned}$$

by the assumption that integration and differentiation are interchangeable.

The other terms in (4.9) have similar derivations.

APPENDIX C

Asymptotics Accounting for Model Mismatch

C.1 Proof of Theorem V.2 - Consistency of the QMLE

C.1.1 Existence of the QMLE

Let the probability measure $v(H) \stackrel{\triangle}{=} \mathsf{Pr}\left(\boldsymbol{r} \in H\right)$ be the true distribution of a single recorded attribute. Since v is typically unknown, we model it with the probability measure $\tilde{v}_{\boldsymbol{\theta}}$. We follow [78] and make the following assumptions:

Assumption C.1. *The attributes of distinct photon interaction events are IID.*

Assumption C.2. *Assume that $\tilde{v}_{\boldsymbol{\theta}}$ has a Radon-Nikodym density with respect to the measure μ.*

By Assumption C.2, there exists a function $\tilde{\mathsf{p}}(\boldsymbol{r};\boldsymbol{\theta})$ such that

$$\tilde{v}_{\boldsymbol{\theta}}(H) = \int_H \tilde{\mathsf{p}}(\boldsymbol{r};\boldsymbol{\theta}) d\mu(\boldsymbol{r}).$$

The true mean number of recorded photons is typically unknown, so we model it as

$$J_\tau \sim \text{Poisson}(\tilde{\lambda}(\boldsymbol{\theta})\tau).$$

One can show that Assumptions C.1 and C.2 imply that the Radon-Nikodym density of the modeled distribution of J_τ observed attributes and J_τ is given by (6.3). The

Radon-Nikodym density in (6.3) is the likelihood of $\boldsymbol{\theta}$ given the number of recorded attributes and their values. The same formula is given in [4].

We have now shown that the modeled likelihood function exists, but we have yet to show the existence of a maximizer over the parameter space $\boldsymbol{\theta}$. To show existence, we introduce three more assumptions:

Assumption C.3. $\tilde{\mathsf{p}}(\boldsymbol{r};\boldsymbol{\theta})$ *is continuous in* $\boldsymbol{\theta}$

Assumption C.4. $\tilde{\lambda}(\boldsymbol{\theta})$ *is continuous in* $\boldsymbol{\theta}$

Assumption C.5. Θ *is a compact subset of* $\mathbb{R}^d$

Assumptions C.3, C.4, and C.5, along with the fact that the Radon-Nikodym density of the modeled distribution exists implies that

$$\tilde{\boldsymbol{\theta}}_\tau \stackrel{\triangle}{=} \arg\max_{\boldsymbol{\theta}\in\Theta} \tilde{\mathsf{p}}(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots, \boldsymbol{r}_J, J; \boldsymbol{\theta})$$

exists.

C.1.2 Strong Consistency of QMLE

A set of suitable regularity conditions for Theorem V.2 are:

Assumption C.6. $|\tilde{\mathsf{p}}(\boldsymbol{r};\boldsymbol{\theta})| < m(\boldsymbol{r})$ *for all* $\boldsymbol{\theta} \in \Theta$ *for some* $m(\boldsymbol{r})$ *that is integrable with respect to the true attribute distribution, i.e.,* $\int_{\mathcal{R}} m(\boldsymbol{r})\mathsf{p}(\boldsymbol{r})\, d\boldsymbol{r}$ *is finite*

Assumption C.7. *The solution to (6.6) is unique.*

In this proof, the QMLE is assumed to exist and Assumptions C.6-C.7 are assumed satisfied.

Proof. We use the result for strong consistency under model mismatch with a nonrandom number of measurements in [78] to prove strong consistency when the number of measurements in Poisson.

We divide the scan of length τ into n intervals of unit duration. We can assume without loss of generality that τ is an integer because our choice of scan interval duration is arbitrary. Treating the scan of duration $n = \tau$ as n independent scans of unit duration leads to the following restatement of the log–likelihood:

$$\begin{aligned} \log \tilde{\mathsf{p}}(\tilde{\boldsymbol{r}}; \boldsymbol{\theta}) \quad &= \quad \sum_{k=1}^{n} \Bigg(J_k \log(\tilde{\lambda}(\boldsymbol{\theta})) - \tilde{\lambda}(\boldsymbol{\theta}) - \log(J_k!) + \\ &\qquad \sum_{j=1}^{J_k} \log \tilde{\mathsf{p}}(\boldsymbol{r}_{kj}; \boldsymbol{\theta}) \Bigg), \end{aligned} \tag{3.1}$$

where $J_k \sim \text{Poisson}(\lambda_{\mathsf{s}})$, $J_\tau = \sum_{k=1}^{n} J_k$ by the summation property of independent Poisson random variables [59, p. 196], $\boldsymbol{r}_{kj}$ is the attribute vector of the jth event to occur during the kth scan. Since the modeled mean number of total received counts is $\tilde{J}(\boldsymbol{\theta})$, the modeled mean number of counts per scan interval is $\tilde{J}(\boldsymbol{\theta})/\tau = \tilde{\lambda}(\boldsymbol{\theta})$ by the summation property of independent Poisson random variables.

Assumption C.6 implies that

$$\begin{aligned} \log \tilde{\mathsf{p}}(\tilde{\boldsymbol{r}}; \boldsymbol{\theta}) \quad &= \quad J_\tau \log(\tilde{\lambda}(\boldsymbol{\theta})) - \tau \tilde{\lambda}(\boldsymbol{\theta}) \\ &\qquad - \log(J_\tau!) + \sum_{j=1}^{J_\tau} \log \tilde{\mathsf{p}}\left(\boldsymbol{r}_j; \boldsymbol{\theta}\right) \\ &\leq \quad J_\tau \log(\tilde{\lambda}(\boldsymbol{\theta})) - \tau \tilde{\lambda}(\boldsymbol{\theta}) \\ &\qquad - \log(J_\tau!) + \sum_{j=1}^{J_k} m(\boldsymbol{r}_j) \\ &\triangleq \quad f(\tilde{\boldsymbol{r}}), \end{aligned}$$

so $\log \tilde{\mathsf{p}}(\tilde{\boldsymbol{r}}; \boldsymbol{\theta}) < f(\tilde{\boldsymbol{r}})$ for all $\boldsymbol{\theta} \in \Theta$ and $f(\tilde{\boldsymbol{r}})$ is integrable with respect to the true distribution because it is the sum of integrable functions. The above argument and the uniqueness of the minimizer of (6.6) provided by Assumption C.7 imply that Assumption A3 of [78] is satisfied.

Strong convergence of $\tilde{\boldsymbol{\theta}}_\tau$ to $\tilde{\boldsymbol{\mu}}$ follows by Theorem 2.1 of [77].

To show that $\tilde{\boldsymbol{\mu}}$ is the member of the parameter space that minimizes the Kullback-Leibler divergence between the model and the true distributions, we make the following additional assumptions:

Assumption C.8. *The probability measure v characterizing the true distribution is absolutely continuous with respect to the base measure μ, with Radon-Nikodym density $\mathsf{p}(\boldsymbol{r})$ for $\boldsymbol{r} \in \mathcal{R}$.*

Assumption C.9. $\mathsf{E}[\log \mathsf{p}(\boldsymbol{r})]$*, the logarithm of the true density of a single attribute, exists and is finite.*

Assumption C.8 guarantees that the true distribution of the list of recorded attributes a Radon-Nikodym density. Assumption C.9 implies that $\mathsf{E}[\log \mathsf{p}(\tilde{\boldsymbol{r}})]$, the logarithm the density of a list of attributes, exists by the linearity of expectation. Thus, the Kullback-Leibler divergence $\mathsf{E}[\log(\mathsf{p}(\tilde{\boldsymbol{r}})/\tilde{\mathsf{p}}(\tilde{\boldsymbol{r}};\boldsymbol{\theta}))]$ exists.

The existence of the QMLE and Assumptions C.6-C.9 satisfy the conditions of Theorem 2.2 of [78], which gives strong convergence of the estimates to the element $\tilde{\boldsymbol{\mu}} \in \Theta$ that minimizes $\mathsf{E}[\log(\mathsf{p}(\tilde{\boldsymbol{r}})/\tilde{\mathsf{p}}(\tilde{\boldsymbol{r}};\boldsymbol{\theta}))]$. □

C.2 Proof of Theorem V.4 - Asymptotic Normality of QMLE

We make the following assumptions to guarantee the asymptotic normality of the QMLE:

Assumption C.10. $\log \tilde{\mathsf{p}}(\boldsymbol{r};\boldsymbol{\theta})$ *is continuously differentiable in $\boldsymbol{\theta}$ for each $\boldsymbol{r} \in \mathcal{R}$*

Assumption C.11. $\log \tilde{\lambda}(\boldsymbol{\theta})$ *and $\tilde{\lambda}(\boldsymbol{\theta})$ are continuously differentiable in $\boldsymbol{\theta}$*

Assumption C.12. $\nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}(\boldsymbol{r};\boldsymbol{\theta})$ *and* $\nabla^2_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}(\boldsymbol{r};\boldsymbol{\theta})$ *are measurable in $\boldsymbol{r}$ for each $\boldsymbol{\theta} \in \Theta$.*

Assumption C.13. *The elements of*

$$\left(\nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}\left(\boldsymbol{r};\boldsymbol{\theta}\right) + \nabla_{\boldsymbol{\theta}} \log \tilde{\lambda}(\boldsymbol{\theta})\right)\left(\nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}\left(\boldsymbol{r};\boldsymbol{\theta}\right) + \nabla_{\boldsymbol{\theta}} \log \tilde{\lambda}(\boldsymbol{\theta})\right)^{T}$$

are dominated by functions integrable with respect to the true distribution.

Assumption C.14. *The elements of* $\nabla^2_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}\left(\boldsymbol{r};\boldsymbol{\theta}\right)$ *are dominated by functions integrable with respect to the true distribution.*

Assumption C.15. $\tilde{\boldsymbol{\mu}}$ *is an interior point of* Θ

Assumption C.16. $\tilde{\mathsf{G}}\left(\tilde{\boldsymbol{\mu}}\right)$ *is nonsingular*

Assumption C.17. $\tilde{\mathsf{H}}\left(\boldsymbol{\theta}\right)$ *is invertible in an open neighborhood around* $\tilde{\boldsymbol{\mu}}$

Assumption C.18. $\tilde{\boldsymbol{\theta}}_{\tau} \xrightarrow{p} \tilde{\boldsymbol{\mu}}$

Proof. Let the gradient of the log–likelihood be

$$\begin{aligned}\tilde{g}_{\tau}\left(\boldsymbol{\theta}\right) \stackrel{\triangle}{=} \nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}\left(\tilde{\boldsymbol{r}};\boldsymbol{\theta}\right) &= J_{\tau}\nabla_{\boldsymbol{\theta}} \log\left(\tilde{\lambda}(\boldsymbol{\theta})\right) - \nabla_{\boldsymbol{\theta}}\left(\tilde{\lambda}(\boldsymbol{\theta})\right) + \sum_{j=1}^{J_{\tau}} \nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}\left(\boldsymbol{r}_j;\boldsymbol{\theta}\right) \\ &= J_{\tau}\nabla_{\boldsymbol{\theta}} \log\left(\tilde{\lambda}\left(\boldsymbol{\theta}\right)\right) - \tau\nabla_{\boldsymbol{\theta}}\tilde{\lambda}\left(\boldsymbol{\theta}\right) + \sum_{j=1}^{J_{\tau}} \nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}\left(\boldsymbol{r}_j;\boldsymbol{\theta}\right),\end{aligned}$$

and let the Hessian of the log–likelihood be

$$\begin{aligned}\tilde{\mathsf{H}}_{\tau}\left(\boldsymbol{\theta}\right) \stackrel{\triangle}{=} \nabla^2_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}\left(\tilde{\boldsymbol{r}};\boldsymbol{\theta}\right) &= J_{\tau}\nabla^2_{\boldsymbol{\theta}} \log(\tilde{\lambda}(\boldsymbol{\theta})) - \nabla^2_{\boldsymbol{\theta}}(\tilde{\lambda}(\boldsymbol{\theta})) + \sum_{j=1}^{J_{\tau}} \nabla^2_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}\left(\boldsymbol{r}_j;\boldsymbol{\theta}\right) \\ &= J_{\tau}\nabla^2_{\boldsymbol{\theta}} \log \tilde{\lambda}\left(\boldsymbol{\theta}\right) - \tau\nabla^2_{\boldsymbol{\theta}}\tilde{\lambda}\left(\boldsymbol{\theta}\right) + \sum_{j=1}^{J_{\tau}} \nabla^2_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}\left(\boldsymbol{r}_j;\boldsymbol{\theta}\right),\end{aligned}$$

which both exist by Assumptions C.10, C.11, and C.12. By the mean value theorem,

$$\tilde{g}_{\tau}\left(\tilde{\boldsymbol{\theta}}_n\right) - \tilde{g}_{\tau}\left(\tilde{\boldsymbol{\mu}}\right) = \tilde{\mathsf{H}}_{\tau}\left(\check{\boldsymbol{\theta}}_{\tau}\right)\left(\tilde{\boldsymbol{\theta}}_{\tau} - \tilde{\boldsymbol{\mu}}\right),$$

for some $\check{\boldsymbol{\theta}}_{\tau}$ such that $\check{\boldsymbol{\theta}}_{\tau}$ lies on the line segment between $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\theta}}_{\tau}$. Since $\tilde{g}_{\tau}\left(\tilde{\boldsymbol{\theta}}_n\right) = 0$ by the definition of the QMLE,

$$\sqrt{\tau}(\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\mu}}) = \left(-\frac{1}{\tau}\tilde{\mathsf{H}}_{\tau}\left(\check{\boldsymbol{\theta}}_{\tau}\right)\right)^{-1}\left(\frac{1}{\sqrt{\tau}}\tilde{g}_{\tau}\left(\tilde{\boldsymbol{\mu}}\right)\right). \tag{3.2}$$

From (6.8), we have

$$\mathsf{E}\left[J_\tau \nabla_{\boldsymbol{\theta}} \log\left(\tilde{\lambda}\left(\boldsymbol{\theta}\right)\right) - \tau \nabla_{\boldsymbol{\theta}} \tilde{\lambda}\left(\boldsymbol{\theta}\right) + \sum_{j=1}^{J_\tau} \nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}\left(\boldsymbol{r}_j; \boldsymbol{\theta}\right)\right] = 0$$

$$\mathsf{E}\left[\sum_{j=1}^{J_\tau}\left(\nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}\left(\boldsymbol{r}_j; \boldsymbol{\theta}\right) + \nabla_{\boldsymbol{\theta}} \log\left(\tilde{\lambda}\left(\boldsymbol{\theta}\right)\right)\right)\right] = \tau \nabla_{\boldsymbol{\theta}} \tilde{\lambda}\left(\boldsymbol{\theta}\right)$$

$$\mathsf{E}\left[\nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}\left(\boldsymbol{r}_j; \boldsymbol{\theta}\right)\right] + \nabla_{\boldsymbol{\theta}} \log\left(\tilde{\lambda}\left(\boldsymbol{\theta}\right)\right) = \nabla_{\boldsymbol{\theta}} \tilde{\lambda}\left(\boldsymbol{\theta}\right) / \lambda_{\mathsf{s}}. \tag{3.3}$$

By (3.3) and Theorem C.31,

$$\frac{1}{\sqrt{\tau}} \tilde{g}_\tau\left(\tilde{\boldsymbol{\mu}}\right) \xrightarrow{d} \mathcal{N}(0, \mathsf{G}(\tilde{\boldsymbol{\mu}})) \text{ as } \tau \to \infty,$$

where

$$\mathsf{G}(\boldsymbol{\theta}) \stackrel{\triangle}{=} \lambda_{\mathsf{s}} \mathsf{E}\left[\left(\nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}\left(\boldsymbol{r}; \boldsymbol{\theta}\right) + \nabla_{\boldsymbol{\theta}} \log \tilde{\lambda}\left(\boldsymbol{\theta}\right)\right)\left(\nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}\left(\boldsymbol{r}; \boldsymbol{\theta}\right) + \nabla_{\boldsymbol{\theta}} \log \tilde{\lambda}\left(\boldsymbol{\theta}\right)\right)^T\right], \tag{3.4}$$

which exists by Assumption C.13. By an extension of the weak law of large numbers to Poisson random sums and by Assumptions C.15 and C.18, we have that

$$\frac{1}{\tau} \tilde{\mathsf{H}}_\tau\left(\check{\boldsymbol{\theta}}_\tau\right) \xrightarrow{p} -\mathsf{H}\left(\tilde{\boldsymbol{\mu}}\right) \quad \text{as } \tau \to \infty,$$

where

$$\mathsf{H}(\boldsymbol{\theta}) \stackrel{\triangle}{=} -\lambda_{\mathsf{s}} \nabla^2_{\boldsymbol{\theta}} \log \tilde{\lambda}\left(\boldsymbol{\theta}\right) + \nabla^2_{\boldsymbol{\theta}} \tilde{\lambda}\left(\boldsymbol{\theta}\right) - \lambda_{\mathsf{s}} \mathsf{E}\left[\nabla^2_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}\left(\boldsymbol{r}; \boldsymbol{\theta}\right)\right], \tag{3.5}$$

which exists by Assumption C.14. Combining (3.2), (3.5), and (3.4), and using Slutsky's theorem [23, p. 39],

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \tilde{\boldsymbol{\mu}}) \xrightarrow{d} \mathcal{N}\left(0, \Sigma(\boldsymbol{\theta})\right),$$

where

$$\Sigma(\boldsymbol{\theta}) = \mathsf{H}^{-1}\left(\tilde{\boldsymbol{\mu}}\right) \mathsf{G}\left(\tilde{\boldsymbol{\mu}}\right) \mathsf{H}^{-1}\left(\tilde{\boldsymbol{\mu}}\right),$$

which exists and is nonsingular by Assumptions C.16 and C.17. □

C.3 Proof of Theorem V.5 - Convergence of the constrained QMLE

We state the regularity conditions that are required for Theorem V.5 to hold we provide a proof.

Assumption C.19. Ω *is compact.*

Assumption C.20. *The solution to (5.13) is unique.*

Theorem V.5 requires that Assumptions C.1–C.6, Assumption C.19, and Assumption C.20 hold. We now provide the proof of Theorem V.5.

Proof. We can separate the list of recorded measurements $\tilde{\boldsymbol{r}}$ into n independent random vectors as shown in the proof of Theorem V.2 in Appendix C.1.2. The result follows from the Theorem 1 of [69], with $\tilde{\boldsymbol{\mu}}^+$ substituted for the "true" parameter θ_0 defined in [69]. This result is also given without proof in [2].

□

C.4 Proof of Theorem V.7 - Asymptotic distribution of the constrained QMLE

We first state the additional assumptions necessary for the theorem to hold.

Assumption C.21. $\tilde{\mathsf{G}}\left(\tilde{\boldsymbol{\mu}}^+\right)$ *is nonsingular*

Assumption C.22. $\tilde{g}\left(\tilde{\boldsymbol{\mu}}^+\right) = 0$

Assumption C.23. $\tilde{\mathsf{H}}\left(\boldsymbol{\theta}\right)$ *is invertible in an open neighborhood of* $\tilde{\boldsymbol{\mu}}^+$ *that is a subset of* $\boldsymbol{\theta}$

Assumption C.24. C_Ω *is a cone approximation for the set* Ω *at* $\tilde{\boldsymbol{\mu}}^+$

Assumption C.25. Ω *is convex*

Assumption C.26. $\tilde{\boldsymbol{\theta}}_\tau^+ \xrightarrow{p} \tilde{\boldsymbol{\mu}}^+$

The theorem holds if Assumptions C.10-C.14 and C.21-C.26 hold.

Proof. The proof follows from Theorem 3 of [2]. Assumptions C.10, C.11, and C.12 imply Assumption 2* of [2], Assymption C.26 implies Assumption 4 of [2], Assymption C.24 implies Assumption 5 of [2], and Assymption C.25 implies Assumption 6 of [2].

Assumptions C.13 and C.21 imply that $\tilde{\mathsf{G}}\left(\tilde{\boldsymbol{\mu}}^{+}\right)$ exists. By Lemma C.27 and Assymption C.22, we have that

$$\frac{1}{\sqrt{\tau}} \nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}\left(\tilde{\boldsymbol{r}}; \boldsymbol{\theta}\right)\big|_{\tilde{\boldsymbol{\mu}}^{+}} = \tilde{g}_{\tau}\left(\tilde{\boldsymbol{\mu}}^{+}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \tilde{\mathsf{G}}\left(\tilde{\boldsymbol{\mu}}^{+}\right)\right). \tag{3.6}$$

Note that if Assymption C.22 is not satisfied, the random variable $\tilde{g}_{\tau}\left(\tilde{\boldsymbol{\mu}}^{+}\right)$ does not converge in distribution to anything by Lemma C.27.

By an extension of the weak law of large numbers to Poisson random sums and by Assumptions C.15 and C.18, we have that

$$\frac{1}{\tau}\tilde{\mathsf{H}}_{\tau}\left(\tilde{\boldsymbol{\theta}}_{\tau}\right) \xrightarrow{p} -\mathsf{H}\left(\tilde{\boldsymbol{\mu}}^{+}\right) \quad \text{as } \tau \to \infty. \tag{3.7}$$

The results in (3.6) and (3.7) imply that Assumption 3 of [2] is satisfied

Assumptions C.14 and C.23 imply that $\tilde{\mathsf{H}}^{-1}\left(\tilde{\boldsymbol{\mu}}^{+}\right)$ exists. By, the properties of normal random variables, and Slutsky's Theorem,

$$\boldsymbol{\zeta} \stackrel{\triangle}{=} \tilde{\mathsf{H}}^{-1}\left(\tilde{\boldsymbol{\mu}}^{+}\right) \tilde{g}_{\tau}\left(\tilde{\boldsymbol{\mu}}^{+}\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \Sigma(\tilde{\boldsymbol{\mu}}^{+})\right).$$

The result follows by Theorem 3 of [2].

□

Lemma C.27. *Let $\tilde{\mathsf{p}}\left(\tilde{\boldsymbol{r}}; \boldsymbol{\theta}\right)$ be the model for a list of recorded attributes defined in (6.3) and let $\tilde{\boldsymbol{\mu}} \in \Omega \subset \Theta$. Then,*

$$\sqrt{\tau}\left(\frac{1}{\tau}\nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}\left(\tilde{\boldsymbol{r}}; \boldsymbol{\theta}\right)\Big|_{\tilde{\boldsymbol{\mu}}} - \tilde{g}\left(\tilde{\boldsymbol{\mu}}\right)\right) \xrightarrow{d} \mathcal{N}\left(0, \tilde{\mathsf{G}}\left(\tilde{\boldsymbol{\mu}}\right)\right) \quad \textit{as } \tau \to \infty,$$

where $\tilde{g}(\boldsymbol{\theta})$ *is defined in (6.7) and* $\tilde{\mathsf{G}}(\boldsymbol{\theta})$ *is defined in (6.12).*

Proof. To simplify notation, let

$$\nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}(\tilde{\boldsymbol{r}}; \boldsymbol{\theta}) \stackrel{\triangle}{=} \nabla_{\boldsymbol{\theta}'} \log \tilde{\mathsf{p}}\left(\tilde{\boldsymbol{r}}; \boldsymbol{\theta}'\right)\Big|_{\boldsymbol{\theta}'=\boldsymbol{\theta}}.$$

By algebra and (6.3),

$$\begin{aligned}\sqrt{\tau}\left(\frac{1}{\tau}\nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}(\tilde{\boldsymbol{r}}; \boldsymbol{\theta}) - \tilde{g}(\boldsymbol{\theta})\right) &= \sqrt{\tau}\Bigg(- \nabla_{\boldsymbol{\theta}}\tilde{\lambda}(\boldsymbol{\theta}) - \tilde{g}(\boldsymbol{\theta}) \\ &\quad + \frac{1}{\tau}\sum_{j=1}^{J}\left(\nabla_{\boldsymbol{\theta}} \log \tilde{\lambda}(\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}(\boldsymbol{r}_j; \boldsymbol{\theta})\right)\Bigg) \\ &= \frac{1}{\sqrt{\tau}}\Bigg(- \tau\left(\tilde{g}(\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}}\tilde{\lambda}(\boldsymbol{\theta})\right) \\ &\quad + \sum_{j=1}^{J}\left(\nabla_{\boldsymbol{\theta}} \log \tilde{\lambda}(\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}(\tilde{\boldsymbol{r}}_j; \boldsymbol{\theta})\right)\Bigg). \qquad (3.8)\end{aligned}$$

We first compute the expected value of the term inside the sum over j in (3.8). By the definition of $\tilde{g}(\boldsymbol{\theta})$ in (5.6),

$$\tilde{g}(\boldsymbol{\theta}) = -\nabla_{\boldsymbol{\theta}}\tilde{\lambda}(\boldsymbol{\theta}) + \lambda_{\mathsf{s}}\left(\nabla_{\boldsymbol{\theta}}\tilde{\lambda}(\boldsymbol{\theta}) + \mathsf{E}\left[\nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}(\boldsymbol{r}; \boldsymbol{\theta})\right]\right).$$

By algebra and the linearity of expectation,

$$\mathsf{E}\left[\nabla_{\boldsymbol{\theta}}\tilde{\lambda}(\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}} \log \tilde{\mathsf{p}}(\boldsymbol{r}; \boldsymbol{\theta})\right] = \frac{\tilde{g}(\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}}\tilde{\lambda}(\boldsymbol{\theta})}{\lambda_{\mathsf{s}}}. \qquad (3.9)$$

We must also compute the covariance of the term inside the sum over j in (3.8) before applying the asymptotic normality result. Let $\mathsf{Cov}_{\boldsymbol{\theta}}\{\cdot\}$ denote the covariance

where $\boldsymbol{\theta}$ is a nonrandom parameter.

$$\begin{aligned}
\Sigma^{'}(\boldsymbol{\theta}) &\triangleq \mathsf{Cov}_{\boldsymbol{\theta}}\left\{\nabla_{\boldsymbol{\theta}}\tilde{\lambda}(\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}}\log\tilde{\mathsf{p}}\left(\boldsymbol{r};\boldsymbol{\theta}\right)\right\} \\
&= \mathsf{E}\Bigg[\left(\nabla_{\boldsymbol{\theta}}\tilde{\lambda}(\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}}\log\tilde{\mathsf{p}}\left(\boldsymbol{r};\boldsymbol{\theta}\right) - \frac{\tilde{g}\left(\boldsymbol{\theta}\right) + \nabla_{\boldsymbol{\theta}}\tilde{\lambda}(\boldsymbol{\theta})}{\lambda_{\mathsf{s}}}\right) \\
&\quad \left(\nabla_{\boldsymbol{\theta}}\tilde{\lambda}(\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}}\log\tilde{\mathsf{p}}\left(\boldsymbol{r};\boldsymbol{\theta}\right) - \frac{\tilde{g}\left(\boldsymbol{\theta}\right) + \nabla_{\boldsymbol{\theta}}\tilde{\lambda}(\boldsymbol{\theta})}{\lambda_{\mathsf{s}}}\right)^{T}\Bigg] \\
&= \frac{1}{\lambda_{\mathsf{s}}}\tilde{\mathsf{G}}\left(\boldsymbol{\theta}\right) - \mathsf{E}\left[\left(\nabla_{\boldsymbol{\theta}}\tilde{\lambda}(\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}}\log\tilde{\mathsf{p}}\left(\boldsymbol{r};\boldsymbol{\theta}\right)\right)\left(\frac{\tilde{g}\left(\boldsymbol{\theta}\right) + \nabla_{\boldsymbol{\theta}}\tilde{\lambda}(\boldsymbol{\theta})}{\lambda_{\mathsf{s}}}\right)^{T}\right] \\
&\quad -\mathsf{E}\left[\left(\frac{\tilde{g}\left(\boldsymbol{\theta}\right) + \nabla_{\boldsymbol{\theta}}\tilde{\lambda}(\boldsymbol{\theta})}{\lambda_{\mathsf{s}}}\right)\left(\nabla_{\boldsymbol{\theta}}\tilde{\lambda}(\boldsymbol{\theta}) + \nabla_{\boldsymbol{\theta}}\log\tilde{\mathsf{p}}\left(\boldsymbol{r};\boldsymbol{\theta}\right)\right)^{T}\right] \\
&\quad +\mathsf{E}\left[\left(\frac{\tilde{g}\left(\boldsymbol{\theta}\right) + \nabla_{\boldsymbol{\theta}}\tilde{\lambda}(\boldsymbol{\theta})}{\lambda_{\mathsf{s}}}\right)\left(\frac{\tilde{g}\left(\boldsymbol{\theta}\right) + \nabla_{\boldsymbol{\theta}}\tilde{\lambda}(\boldsymbol{\theta})}{\lambda_{\mathsf{s}}}\right)^{T}\right]. \qquad (3.10)
\end{aligned}$$

Substituting (3.9) into (3.10),

$$\Sigma^{'}(\boldsymbol{\theta}) = \frac{1}{\lambda_{\mathsf{s}}}\tilde{\mathsf{G}}\left(\boldsymbol{\theta}\right) - \left(\frac{\tilde{g}\left(\boldsymbol{\theta}\right) + \nabla_{\boldsymbol{\theta}}\tilde{\lambda}(\boldsymbol{\theta})}{\lambda_{\mathsf{s}}}\right)\left(\frac{\tilde{g}\left(\boldsymbol{\theta}\right) + \nabla_{\boldsymbol{\theta}}\tilde{\lambda}(\boldsymbol{\theta})}{\lambda_{\mathsf{s}}}\right)^{T}. \qquad (3.11)$$

By (3.8), (3.9), (3.11), and Theorem C.31,

$$\sqrt{\tau}\left(\frac{1}{\tau}\nabla_{\boldsymbol{\theta}}\log\tilde{\mathsf{p}}\left(\tilde{\boldsymbol{r}};\boldsymbol{\theta}\right) - \tilde{g}\left(\boldsymbol{\theta}\right)\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \tilde{\mathsf{G}}\left(\boldsymbol{\theta}\right)\right).$$

The result follows by evaluating the distribution at $\boldsymbol{\theta} = \tilde{\boldsymbol{\mu}}$. □

C.5 Example of asymptotic mean on parameter space boundary

Let A be an $m \times n$ matrix such that $A^T A$ is invertible. Let $\tilde{\boldsymbol{y}}$ be a list of observations where

$$\tilde{\boldsymbol{y}} \triangleq \left[\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_N\right],$$

N is the nonrandom number of measurements, and $\boldsymbol{y}_i$ is an $n_d \times 1$ vector of observations for $i = 1, 2, \ldots, N$. We assume the following parametric model for the measurements:

$$\boldsymbol{y}_i \overset{\text{model}}{\sim} \mathcal{N}\left(A\boldsymbol{\theta}, \sigma^2 I\right) \quad i = 1, 2, \ldots, N, \qquad (3.12)$$

where $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_{n_p}]$ is an $n_p \times 1$ vector of unknown parameters to be estimated, I is the identity matrix, and $\sigma^2 > 0$ is the variance of the measurements. We will enforce the constraint that each element of $\boldsymbol{\theta}$ is nonnegative, i.e.,

$$\theta_j \geq 0 \quad j = 1, 2, \ldots, n.$$

Assume that the true distribution of the measurements is also Gaussian with mean $\boldsymbol{m}$:

$$\boldsymbol{y}_i \overset{\text{true}}{\sim} \mathcal{N}\left(\boldsymbol{m}, \sigma^2 I\right).$$

An imaging system that captures a transform A of a natural image is an example of a system where the model (3.12) is plausible. The transform must be either invertible or overcomplete to satisfy the assumption that $A^T A$ is invertible. Fourier and orthonormal or overcomplete wavelet transforms are example transforms that satisfy the requirements for A. A natural image is inherently nonnegative because light intensities cannot be negative. In this example, the parameter vector $\boldsymbol{\theta}$ would be a discretized version of the image, and $\tilde{\boldsymbol{y}}$ would be a series of recorded transform coefficients recorded at times $t(i),\ i = 1, 2, \ldots, N$.

The log–likelihood, ignoring terms constant with respect to $\boldsymbol{\theta}$, is

$$\mathsf{L}\left(\boldsymbol{\theta}\right) = -\frac{1}{2\sigma^2} \sum_{i=1}^{N} ||\boldsymbol{y}_i - A\boldsymbol{\theta}||_2^2,$$

with expected value

$$\mathsf{E}\left[\mathsf{L}\left(\boldsymbol{\theta}\right)\right] = \tilde{f}\left(\boldsymbol{\theta}\right) = -\frac{N}{2\sigma^2} \mathsf{E}\left[\left(\boldsymbol{y}_i - A\boldsymbol{\theta}\right)^T \left(\boldsymbol{y}_i - A\boldsymbol{\theta}\right)\right] \tag{3.13}$$

$$= -\frac{N}{2\sigma^2}\left(n_d\sigma^2 + \boldsymbol{m}^T\boldsymbol{m} - \boldsymbol{\theta}^T A^T\boldsymbol{m} - \boldsymbol{m}^T A\boldsymbol{\theta} + \boldsymbol{\theta}^T A^T A\boldsymbol{\theta}\right) \tag{3.14}$$

$$\equiv -\frac{N}{2\sigma^2} ||\boldsymbol{m} - A\boldsymbol{\theta}||_2^2. \tag{3.15}$$

By the definition of $\tilde{\boldsymbol{\mu}}^+$ in (5.13) and the expected value of the log–likelihood in (3.15), the constrained asymptotic limit $\tilde{\boldsymbol{\mu}}^+$ is given by the constrained least squares

problem:

$$\tilde{\boldsymbol{\mu}}^{+} = \arg \min_{\theta_i \geq 0 \; i=1,\ldots,n_p} ||\boldsymbol{m} - A\boldsymbol{\theta}||_2 \,. \tag{3.16}$$

The expected value of the gradient of the log–likelihood, ignoring constant terms, is

$$\mathsf{E}\left[\nabla_{\boldsymbol{\theta}} \mathsf{L}\left(\boldsymbol{\theta}\right)\right] = \tilde{g}\left(\boldsymbol{\theta}\right) = \frac{N}{\sigma^2} A^T \left(\boldsymbol{m} - A\boldsymbol{\theta}\right), \tag{3.17}$$

which is only equal to $\mathbf{0}$ when the asymptotic mean $\tilde{\boldsymbol{\mu}}$ satisfies the normal equations

$$A^T \boldsymbol{m} - A^T A \tilde{\boldsymbol{\mu}}^{+} = \mathbf{0},$$

or, equivalently, $\tilde{\boldsymbol{\mu}}^{+}$ is equal to the global minimizer of the unconstrained least–squares problem

$$\tilde{\boldsymbol{\mu}}^{+} = \tilde{\boldsymbol{\mu}} \stackrel{\triangle}{=} \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^m} ||\boldsymbol{m} - A\boldsymbol{\theta}||_2 \,.$$

The condition $\tilde{\boldsymbol{\mu}}^{+} = \tilde{\boldsymbol{\mu}}$ only holds if the unconstrained asymptotic limit $\tilde{\boldsymbol{\mu}}^{+}$ on the boundary or in the interior of the nonnegative orthant. When model mismatch is present, small perturbations in the model can change the value of $\tilde{\boldsymbol{\mu}}$, so if $\tilde{\boldsymbol{\mu}}$ lies exactly on the boundary of the nonnegative orthant, a small perturbation in the model, which in this example is the matrix A, will cause the asymptotic mean to be either in the interior or outside the nonnegative orthant. When $\tilde{\boldsymbol{\mu}}$ is outside the nonnegative orthant, Assumption C.22 of Theorem V.7 does not hold, so the convergence in distribution result does not hold. When $\tilde{\boldsymbol{\mu}}^{+}$ is in the interior of the nonnegative orthant, the unconstrained asymptotic normality theorem (Theorem V.4) applies.

C.6 Additional Convergence Results for Poisson Sums

This section gives basic convergence results for Poisson sums that are useful in proving Theorem V.4. Lemma C.28 and its proof appear in [8].

Lemma C.28. *Let $z_1, \ldots, z_m$ and $w_1, \ldots, w_m$ be complex numbers of modulus at most 1. Then,*

$$\left| \prod_{i=1}^{m} z_i - \prod_{i=1}^{m} w_i \right| \leq \sum_{i=1}^{m} |z_i - w_i| .$$

Theorem C.29. *Let $x_1, x_2, \ldots$ be an IID sequence of random variables with $\mathsf{E}[x_k] = m$ and $\mathsf{Var}(x_k) = \sigma^2$ for $k = 1, 2, \ldots$, and let $J_\tau \sim \text{Poisson}(\lambda\tau)$. Then*

$$S_\tau \stackrel{\triangle}{=} \frac{1}{\sigma\sqrt{\tau}} \sum_{i=1}^{J_\tau} (x_i - m) \stackrel{d}{\to} N(0, \lambda) \quad \text{as} \quad \tau \to \infty,$$

where $\stackrel{d}{\to}$ denotes convergence in distribution.

Proof. By the continuity theorem [8, p. 359], it suffices to show that $\psi_\tau(t) \stackrel{\triangle}{=} \mathsf{E}\left[e^{itS_\tau}\right]$ converges pointwise to $\psi(t) \stackrel{\triangle}{=} e^{-t^2\lambda}$, the characteristic function of a Gaussian random vector with mean zero and variance λ.

Let $y_i \stackrel{\triangle}{=} x_i - m$ and $\psi_y(t)$ be the characteristic function of y_i.

Simplifying the characteristic function of S_τ,

$$\begin{aligned}
\psi_\tau(t) &= \mathsf{E}\left[e^{\frac{it}{\sigma\sqrt{\tau}} \sum_{i=1}^{J_\tau} y_i}\right] \\
&= \mathsf{E}\left[\prod_{i=1}^{J_\tau} e^{\frac{it}{\sigma\sqrt{\tau}} y_i}\right] \\
&= \mathsf{E}\left[\mathsf{E}\left[\prod_{i=1}^{J_\tau} e^{\frac{it}{\sigma\sqrt{\tau}} y_i} \,\middle|\, J_\tau\right]\right] \\
&= \mathsf{E}\left[\psi_y^{J_\tau}\left(\frac{t}{\sigma\sqrt{\tau}}\right)\right] \quad \text{by independence.}
\end{aligned}$$

Since the y_i are zero mean, the Taylor series expansion of the characteristic func-

tion y_i gives

$$\begin{aligned}\psi_y\left(\frac{t}{\sigma\sqrt{\tau}}\right) &= \mathsf{E}\left[e^{i\frac{t}{\sigma\sqrt{\tau}}y_i}\right] \\ &= 1 + i\frac{t}{\sigma\sqrt{\tau}}\mathsf{E}\left[y_i\right] + \frac{i^2t^2}{2\sigma^2\tau}\mathsf{E}\left[y_i^2\right] + \frac{1}{\tau^{3/2}}R(t) \\ &= 1 + 0 - \frac{\sigma^2 t}{2\sigma^2\tau} + \frac{1}{\tau^{3/2}}R(t) \\ &= 1 - \frac{t^2}{2\tau} + \frac{1}{\tau^{3/2}}R(t), \qquad (3.18)\end{aligned}$$

where

$$R(t) \triangleq \int_0^t \frac{i^3 s^3 \mathsf{E}\left[y_i^3\right]}{\sigma^3 3!}\left(t-s\right)^2 ds.$$

Now we will show that $\psi_\tau(t) = \mathsf{E}\left[\psi_y^{J_\tau}\left(\frac{t}{\sigma\sqrt{\tau}}\right)\right]$ converges pointwise to $\mathsf{E}\left[\left(1-\frac{t^2}{2\tau}\right)^{J_\tau}\right]$ as $\tau \to \infty$. For $\tau > t^2/4$,

$$\begin{aligned}&\left|\mathsf{E}\left[\psi_y^{J_\tau}(t)\left(\frac{t}{\sigma\sqrt{\tau}}\right)\right] - \mathsf{E}\left[\left(1-\frac{t^2}{2\tau}\right)^{J_\tau}\right]\right| \\ &\leq \mathsf{E}\left[\left|\psi_y^{J_\tau}(t)\left(\frac{t}{\sigma\sqrt{\tau}}\right) - \left(1-\frac{t^2}{2\tau}\right)^{J_\tau}\right|\right] \\ &\leq \mathsf{E}\left[J_\tau\left|\psi_y(t)\left(\frac{t}{\sigma\sqrt{\tau}}\right) - \left(1-\frac{t^2}{2\tau}\right)\right|\right] \quad \text{lemma C.28} \\ &= \tau\lambda\left|\psi_y(t)\left(\frac{t}{\sigma\sqrt{\tau}}\right) - \left(1-\frac{t^2}{2\tau}\right)\right| \\ &= \tau^{-1/2}\lambda R(t) \quad \text{by (3.18)} \\ &\to 0 \quad \text{as} \quad \tau \to \infty.\end{aligned}$$

This shows that $\mathsf{E}\left[\psi_\tau^{J_\tau}\left(\frac{t}{\sigma\sqrt{\tau}}\right)\right]$ converges pointwise to $\mathsf{E}\left[\left(1-\frac{t^2}{2\tau}\right)^{J_\tau}\right]$ for each t as

$\tau \to \infty$. Evaluating the expected value of the binomial term, we have

$$\begin{aligned}
\mathsf{E}\left[\left(1-\frac{t^2}{2\tau}\right)^{J_\tau}\right] &= \sum_{k=1}^{\infty}\left(1-\frac{t^2}{2\tau}\right)^k e^{-\lambda\tau}(\lambda\tau)^k/k! \\
&= e^{-\lambda\tau}\sum_{k=1}^{\infty}\left(\lambda\tau-\frac{\lambda t^2}{2}\right)^k/k! \\
&= e^{-\lambda\tau}e^{\lambda\tau-\frac{\lambda t^2}{2}} \\
&= e^{-\frac{\lambda t^2}{2}},
\end{aligned}$$

which is the characteristic function of a Gaussian random variable with mean 0 and variance λ. □

We now generalize Theorem C.29 to the case of IID random vectors. The proof is similar to the case of a nonrandom number of measurements considered in [8].

Theorem C.30. *Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots \in \mathbb{R}^d$ be an IID sequence of random vectors with $\mathsf{E}[\boldsymbol{x}_k] = \boldsymbol{m}$ and $\mathsf{E}\left[(\boldsymbol{x}_k-\boldsymbol{m})(\boldsymbol{x}_k-\boldsymbol{m})^T\right] = \Sigma$ for $k = 1, 2, \ldots,$ and let $J_\tau \sim$* Poisson$(\lambda\tau)$. *Then*

$$S_\tau \stackrel{\triangle}{=} \frac{1}{\sqrt{\tau}}\sum_{i=1}^{J_\tau}(\boldsymbol{x}_i-\boldsymbol{m}) \stackrel{d}{\to} N(0,\lambda\Sigma) \quad \text{as} \quad \tau\to\infty$$

Proof. Let $\boldsymbol{y}_i \stackrel{\triangle}{=} \boldsymbol{x}_i - \boldsymbol{m}$ and let $S_\tau \stackrel{\triangle}{=} \frac{1}{\sqrt{\tau}}\sum_{i=1}^{J_\tau}\boldsymbol{y}_i$. Let $t \in \mathbb{R}^d$ be arbitrary. The characteristic function of $t'S_\tau$ is

$$\psi_{t'S\tau}(u) = \mathsf{E}\left[e^{iu\frac{1}{\sqrt{\tau}}\sum_{i=1}^{J_\tau}t'\boldsymbol{y}_i}\right]$$

Because $t'S_\tau$ is a scalar random variable, we have by Theorem C.29 that

$$\psi_{t'S\tau}(u) \to e^{u^2t'\lambda\Sigma t/2} \ \text{ pointwise as } \ \tau\to\infty \ \text{ for each } u\in\mathbb{R}.$$

Putting $u = 1$, we have

$$\mathsf{E}\left[e^{\frac{i}{\sqrt{\tau}}\sum_{i=1}^{J_\tau}t'\boldsymbol{y}_i}\right] \to e^{-t'\lambda\Sigma t/2},$$

which, by the multivariate Levy continuity theorem [72, p. 56], gives the desired result. □

We extend the result of Theorem C.30 to the case where the mean is outside the sum.

Theorem C.31. *Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots \in \mathbb{R}^d$ be an IID sequence of random vectors with $\mathsf{E}[\boldsymbol{x}_k] = \boldsymbol{m}$ and $\mathsf{E}\left[(\boldsymbol{x}_k - \boldsymbol{m})(\boldsymbol{x}_k - \boldsymbol{m})^T\right] = \Sigma$ for $k = 1, 2, \ldots,$ and let $J_\tau \sim$ Poisson$(\lambda\tau)$. Then*

$$S_\tau \stackrel{\triangle}{=} \frac{1}{\sqrt{\tau}} \left(\sum_{i=1}^{J_\tau} \boldsymbol{x}_i - \lambda\tau\boldsymbol{m} \right) \stackrel{d}{\to} N(0, \lambda(\Sigma + \boldsymbol{m}\boldsymbol{m}^T)) \quad \text{as} \quad \tau \to \infty$$

Proof.

$$\begin{aligned} \frac{1}{\sqrt{\tau}} \left(\sum_{i=1}^{J_\tau} \boldsymbol{x}_i - \lambda\tau\boldsymbol{m} \right) &= \frac{1}{\sqrt{\tau}} \left(\sum_{i=1}^{J_\tau} \boldsymbol{x}_i - \boldsymbol{m} \right) + \frac{1}{\sqrt{\tau}} \left(J_\tau \boldsymbol{m} - \lambda\tau\boldsymbol{m} \right) \\ &= \frac{1}{\sqrt{\tau}} \left(\sum_{i=1}^{J_\tau} \boldsymbol{x}_i - \boldsymbol{m} \right) + \frac{\boldsymbol{m}}{\sqrt{\tau}} \left(J_\tau - \lambda\tau \right) \\ &\stackrel{d}{\to} \mathcal{N}(0, \lambda(\Sigma + \boldsymbol{m}\boldsymbol{m}^T)), \end{aligned}$$

by Theorem C.30, Theorem C.32, and the properties of normal random variables. □

Theorem C.32. *Let $J_\tau \sim$ Poisson$(\lambda\tau)$. Then,*

$$\frac{1}{\sqrt{\tau}} \left(J_\tau - \lambda\tau \right) \stackrel{d}{\to} \mathcal{N}(0, \lambda) \quad \textit{as} \quad \tau \to \infty.$$

Proof. Let $S_\tau \stackrel{\triangle}{=} \frac{1}{\tau}(J_\tau - \lambda\tau)$. By the Levy continuity theorem [8, p. 359], it suffices to show that

$$\psi_\tau(t) \stackrel{\triangle}{=} \mathsf{E}\left[e^{itS_\tau}\right] \to e^{-\lambda t^2/2} \quad \text{pointwise as } \tau \to \infty.$$

Evaluating $\psi_\tau(t)$,

$$\psi_\tau(t) = \mathsf{E}\left[\exp\left(\frac{it}{\sqrt{\tau}}(J_\tau - \lambda\tau)\right)\right] = \mathsf{E}\left[\exp\left(\frac{it}{\sqrt{\tau}} J_\tau\right)\right] e^{-it\lambda\sqrt{\tau}}.$$

The left–hand expectation is the characteristic function of a Poisson random variable[1] with mean $\lambda\tau$ [58, p. 118] evaluated at $t/\sqrt{\tau}$. Thus,

$$\psi_\tau(t) = \exp\left(\lambda\tau\left(e^{\frac{it}{\sqrt{\tau}}} - 1\right) - it\lambda\sqrt{\tau}\right).$$

Expanding the Taylor series of the nested exponential,

$$\begin{aligned}
\psi_\tau(t) &= \exp\left(\lambda\tau\left(1 + \frac{it}{\sqrt{\tau}} - \frac{t^2}{2\tau} + \sum_{k=3}^{\infty}\frac{(it\tau^{-1/2})^k}{k!} - 1\right) - it\lambda\sqrt{\tau}\right) \\
&= \exp\left(-\frac{\lambda t^2}{2} + \lambda\sum_{k=3}^{\infty}\frac{(it)^k\tau^{1-k/2}}{k!}\right) \\
&\to e^{-\lambda t^2/2} \quad \text{pointwise as} \quad \tau\to\infty.
\end{aligned}$$

□

Theorem C.33. *Let $J_\tau \sim \text{Poisson}(\lambda\tau)$, and let $\boldsymbol{x}_j \in \mathbb{R}^d, j = 1, 2, \ldots$ be IID random vectors with mean $\boldsymbol{\mu}$ and finite covariance. Then,*

$$\frac{1}{\tau}\sum_{j=1}^{J_\tau}\boldsymbol{x}_j \xrightarrow{p} \lambda\boldsymbol{\mu}.$$

Proof. Let $\boldsymbol{x}_j = [x_{j1}, x_{j2}, \ldots, x_{jd}]$. By Theorem 11.9 of [71, p.345], it suffices to show that

$$\frac{1}{\tau}\sum_{j=1}^{J_\tau}x_{ji} \xrightarrow{p} \lambda\mu_i,$$

for $i = 1, 2, \ldots, d$. Let $\epsilon > 0$ be arbitrary. Then,

$$\begin{aligned}
\mathsf{Pr}\left(\left|\frac{1}{\tau}\sum_{j=1}^{J_\tau}x_{ji} - \mu_i\right| < \epsilon\right) &\le \frac{\mathsf{Var}\left(\frac{1}{\tau}\sum_{j=1}^{J_\tau}x_{ji}\right)}{\epsilon^2} \quad \text{Chebyschev inequality} \\
&= \frac{\frac{1}{\tau^2}\mathsf{Var}\left(\sum_{j=1}^{J_\tau}x_{ji}\right)}{\epsilon^2} \\
&= \frac{\tau\lambda\mathsf{Var}\left(x_{ji}\right)}{\tau^2\epsilon^2} \quad \text{by iterated expectation} \qquad (3.19) \\
&= \frac{\lambda\mathsf{Var}\left(x_{ji}\right)}{\tau\epsilon^2} \to 0 \quad \text{as} \quad \tau\to\infty. \qquad (3.20)
\end{aligned}$$

□

[1] If $y \sim \text{Poisson}(\lambda)$, then $\mathsf{E}\left[e^{ity}\right] = \exp\left(\lambda\left(e^{it} - 1\right)\right)$

Lemma C.34. *If* $J_\tau \sim \text{Poisson}(\tau\lambda)$*, then* $\frac{J_\tau}{\tau} \xrightarrow{p} \lambda$ *as* $\tau \to \infty$*.*

Proof. By the Chebyschev inequality, for any $\epsilon > 0$,

$$\begin{aligned}
\mathsf{Pr}\left(\left|\frac{J_\tau}{\tau} - \lambda\right| \leq \epsilon\right) &\leq \frac{\mathsf{Var}\left(J_\tau/\tau\right)}{\epsilon^2} \\
&= \frac{\lambda\tau}{\tau^2\epsilon^2} \\
&= \frac{\lambda}{\tau\epsilon^2} \\
\to 0 \text{ as } \tau \to \infty.
\end{aligned}$$

□

Lemma C.35. *Let* $f(\boldsymbol{x}; \check{\boldsymbol{\theta}})$ *be continuous at* $\check{\boldsymbol{\theta}} = \tilde{\boldsymbol{\mu}}$ *and uniformly continuous in* $\boldsymbol{x}$*, and let* $J_\tau \sim \text{Poisson}(\lambda\tau)$*. If* $\frac{1}{\tau}\sum_{i=1}^{J_\tau} f(\boldsymbol{x}_i; \tilde{\boldsymbol{\mu}}) \xrightarrow{p} \lambda\mathsf{E}\left[f(\boldsymbol{x}; \tilde{\boldsymbol{\mu}})\right]$ *and* $\tilde{\boldsymbol{\theta}}_\tau \xrightarrow{p} \tilde{\boldsymbol{\mu}}$ *then* $\frac{1}{\tau}\sum_{i=1}^{J_\tau} f(\boldsymbol{x}_i; \tilde{\boldsymbol{\theta}}_\tau) \xrightarrow{p} \lambda\mathsf{E}\left[f(\boldsymbol{x}; \tilde{\boldsymbol{\mu}})\right]$

Proof.

$$\begin{aligned}
\frac{1}{\tau}\sum_{i=1}^{J_\tau} f(\boldsymbol{x}_i; \tilde{\boldsymbol{\theta}}_\tau) - \lambda\mathsf{E}\left[f(\boldsymbol{x}; \tilde{\boldsymbol{\mu}})\right] &= \frac{1}{\tau}\sum_{i=1}^{J_\tau} f(\boldsymbol{x}_i; \tilde{\boldsymbol{\theta}}_\tau) - \frac{1}{\tau}\sum_{i=1}^{J_\tau} f(\boldsymbol{x}_i; \tilde{\boldsymbol{\mu}}) \\
&\quad + \frac{1}{\tau}\sum_{i=1}^{J_\tau} f(\boldsymbol{x}_i; \tilde{\boldsymbol{\mu}}) - \lambda\mathsf{E}\left[f(\boldsymbol{x}; \tilde{\boldsymbol{\mu}})\right] \\
&= \frac{1}{\tau}\sum_{i=1}^{J_\tau} \left(f(\boldsymbol{x}_i; \tilde{\boldsymbol{\theta}}_\tau) - f(\boldsymbol{x}_i; \tilde{\boldsymbol{\mu}})\right) \\
&\quad + \frac{1}{\tau}\sum_{i=1}^{J_\tau} f(\boldsymbol{x}_i; \tilde{\boldsymbol{\mu}}) - \lambda\mathsf{E}\left[f(\boldsymbol{x}; \tilde{\boldsymbol{\mu}})\right].
\end{aligned}$$

By the continuous mapping theorem, $\tilde{\boldsymbol{\theta}}_\tau \xrightarrow{p} \tilde{\boldsymbol{\mu}}$ implies that $f(\boldsymbol{x}_i; \tilde{\boldsymbol{\theta}}_\tau) - f(\boldsymbol{x}_i; \tilde{\boldsymbol{\mu}}) \xrightarrow{p} 0$. Since the individual terms converge in probability to 0 and f is continuous, by the continuous mapping theorem,

$$\frac{1}{\tau}\sum_{i=1}^{J_\tau} \left(f(\boldsymbol{x}_i; \tilde{\boldsymbol{\theta}}_\tau) - f(\boldsymbol{x}_i; \tilde{\boldsymbol{\mu}})\right) \xrightarrow{p} 0 \text{ as } n \to \infty. \tag{3.21}$$

By the hypothesis of the theorem,

$$\frac{1}{\tau}\sum_{i=1}^{J_\tau} f(\boldsymbol{x}_i;\tilde{\boldsymbol{\mu}}) - \lambda \mathsf{E}\left[f(\boldsymbol{x};\tilde{\boldsymbol{\mu}})\right] \xrightarrow{p} 0. \tag{3.22}$$

By (3.21) and (3.22), we have that

$$\frac{1}{\tau}\sum_{i=1}^{J_\tau} f(\boldsymbol{x}_i;\tilde{\boldsymbol{\theta}}_\tau) - \lambda \mathsf{E}\left[f(\boldsymbol{x};\tilde{\boldsymbol{\mu}})\right] \xrightarrow{p} 0,$$

which gives the desired result. □

C.7 Verification of Regularity Conditions

This appendix describes the methods by which one could verify Assumptions 1-7 needed for Theorem V.2 and Assumptions 10-18 required for Theorem V.4 using the example system in §5.4.

The recorded attributes are IID because of the chosen system setup, and one can verify that Assumptions 2-4, 6, and 10-12 are satisfied. Assumption 5 is satisfied if one places an upper bound on the set of possible source and background intensity estimates. Assumption 7 is satisfied in all but degenerate cases where there are multiple sources of exactly the same intensity, which are of little practical interest. Assumptions 13 and 14 are difficult to verify directly without an expression for the true distribution of recorded attributes. However, we expect that the true distribution is well–behaved so that the continuous gradient and Hessian of the log–likelihood are integrable with respect to the true distribution. One can verify Assumptions 15-17 for a particular case after computing $\tilde{\boldsymbol{\mu}}$, $\tilde{\mathsf{G}}(\boldsymbol{\theta})$, and $\tilde{\mathsf{H}}(\boldsymbol{\theta})$. Assumption 18 follows from Theorem V.2.

C.8 Implementation of Algorithms

The algorithms proposed in this thesis were implemented in a software package called UMImaging, which is maintained by Professor Zhong He's group. The software

includes maximum likelihood image reconstruction and multiple source detection methods. I added the log–regularization image reconstruction method in Chapter III and the detection performance prediction method in Chapter V.

The software can compute detection performance predictions with known or unknown background intensity and known or unknown source position, but the background spectrum and source isotope must be known. The calculations are performed assuming that the background is spatially uniform for each energy. I created a user manual for the performance prediction software and it is located in the UMImaging repository.

C.9 Sample Calculations of $\tilde{\mathsf{G}}(\tilde{\boldsymbol{\mu}})$ and $\tilde{\mathsf{H}}(\tilde{\boldsymbol{\mu}})$

We provide calculations for the $[1,1]$ elements of $\tilde{\mathsf{G}}(\tilde{\boldsymbol{\mu}})$ and $\tilde{\mathsf{H}}(\tilde{\boldsymbol{\mu}})$ in (5.44) and (5.45). Calculations of the other elements of these matrices follows from similar calculation.

C.9.1 Calculation of $\tilde{\mathsf{G}}(\boldsymbol{\theta})_{[1,1]}$

From the definition of $\tilde{\lambda}(\boldsymbol{\theta})$ in (6.2),

$$\frac{\partial}{\partial\alpha}\tilde{\lambda}(\boldsymbol{\theta}) = \frac{\tilde{s}(\boldsymbol{\phi})}{\alpha\tilde{s}(\boldsymbol{\phi}) + \lambda_b}. \tag{3.23}$$

From the mixture–model definition of $\tilde{\mathsf{p}}(\boldsymbol{r};\boldsymbol{\theta})$ in (6.4),

$$\frac{\partial}{\partial\alpha}\tilde{\mathsf{p}}(\boldsymbol{r};\boldsymbol{\theta}) = \frac{\tilde{s}(\boldsymbol{\phi})\tilde{\mathsf{p}}_{\mathrm{S}}(\boldsymbol{r};\boldsymbol{\phi})}{\alpha\tilde{s}(\boldsymbol{\phi})\tilde{\mathsf{p}}_{\mathrm{S}}(\boldsymbol{r};\boldsymbol{\phi}) + \lambda_b\tilde{\mathsf{p}}_{\mathrm{B}}(\boldsymbol{r})} - \frac{\tilde{s}(\boldsymbol{\phi})}{\alpha\tilde{s}(\boldsymbol{\phi}) + \lambda_b}. \tag{3.24}$$

Substituting (3.23) and (3.24) into the definition of $\tilde{\mathsf{G}}(\tilde{\boldsymbol{\mu}})$ in (6.12),

$$\begin{aligned}\tilde{\mathsf{G}}(\boldsymbol{\theta}) &= \lambda_{\mathsf{s}}\mathsf{E}\left[\frac{\tilde{s}^2(\boldsymbol{\phi})\tilde{\mathsf{p}}_{\mathrm{S}}^2(\boldsymbol{r};\boldsymbol{\phi})}{\left(\alpha\tilde{s}(\boldsymbol{\phi})\tilde{\mathsf{p}}_{\mathrm{S}}(\boldsymbol{r};\boldsymbol{\phi}) + \lambda_b\tilde{\mathsf{p}}_{\mathrm{B}}(\boldsymbol{r})\right)^2}\right] \\ &= \frac{\lambda_{\mathsf{s}}\tilde{s}^2(\boldsymbol{\phi})}{\tilde{\lambda}^2(\boldsymbol{\theta})}\mathsf{E}\left[\frac{\tilde{\mathsf{p}}_{\mathrm{S}}^2(\boldsymbol{r};\boldsymbol{\phi})}{\tilde{\mathsf{p}}^2(\boldsymbol{r};\boldsymbol{\theta})}\right].\end{aligned} \tag{3.25}$$

C.9.2 Calculation of $\tilde{\mathsf{H}}(\boldsymbol{\theta})_{[1,1]}$

Taking derivatives with respect to α:

$$\frac{\partial^2}{\partial\alpha^2}\log\tilde{\lambda}(\boldsymbol{\theta}) = -\frac{\tilde{s}^2(\boldsymbol{\phi})}{(\alpha\tilde{s}^2(\boldsymbol{\phi})+\lambda_b)}, \tag{3.26}$$

$$\frac{\partial^2}{\partial\alpha^2}\tilde{\lambda}(\boldsymbol{\theta}) = 0, \tag{3.27}$$

and

$$\frac{\partial^2}{\partial\alpha^2}\log\tilde{\mathsf{p}}\left(\boldsymbol{r};\boldsymbol{\theta}\right) = \frac{-\tilde{s}^2(\boldsymbol{\phi})\tilde{\mathsf{p}}_{\mathrm{S}}^2(\boldsymbol{r};\boldsymbol{\phi})}{(\alpha\tilde{s}(\boldsymbol{\phi})\tilde{\mathsf{p}}_{\mathrm{S}}(\boldsymbol{r};\boldsymbol{\phi})+\lambda_b\tilde{\mathsf{p}}_{\mathrm{B}}(\boldsymbol{r}))^2} + \frac{\tilde{s}^2(\boldsymbol{\phi})}{(\alpha\tilde{s}^2(\boldsymbol{\phi})+\lambda_b)^2}. \tag{3.28}$$

Substituting (3.26), (3.27), and (3.28) into (6.13), we obtain

$$\begin{aligned}\tilde{\mathsf{H}}\left(\boldsymbol{\theta}\right) &= \lambda_{\mathsf{s}}\mathsf{E}\left[\frac{\tilde{s}^2(\boldsymbol{\phi})\tilde{\mathsf{p}}_{\mathrm{S}}^2(\boldsymbol{r};\boldsymbol{\phi})}{\left(\alpha\tilde{s}(\boldsymbol{\phi})\tilde{\mathsf{p}}_{\mathrm{S}}(\boldsymbol{r};\boldsymbol{\phi})+\lambda_b\tilde{\mathsf{p}}_{\mathrm{B}}(\boldsymbol{r})\right)^2}\right] \\ &= \frac{\lambda_{\mathsf{s}}\tilde{s}^2(\boldsymbol{\phi})}{\tilde{\lambda}^2(\boldsymbol{\theta})}\mathsf{E}\left[\frac{\tilde{\mathsf{p}}_{\mathrm{S}}^2(\boldsymbol{r};\boldsymbol{\phi})}{\tilde{\mathsf{p}}^2(\boldsymbol{r};\boldsymbol{\theta})}\right].\end{aligned}$$

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce Dubois, M. Asai, G. Barrand, R. Capra, S. Chauvie, R. Chytracek, G. A. P. Cirrone, G. Cooperman, G. Cosmo, G. Cuttone, G. G. Daquino, M. Donszelmann, M. Dressel, G. Folger, F. Foppiano, J. Generowicz, V. Grichine, S. Guatelli, P. Gumplinger, A. Heikkinen, I. Hrivnacova, A. Howard, S. Incerti, V. Ivanchenko, T. Johnson, F. Jones, T. Koi, R. Kokoulin, M. Kossov, H. Kurashige, V. Lara, S. Larsson, F. Lei, O. Link, F. Longo, M. Maire, A. Mantero, B. Mascialino, I. McLaren, P. Mendez Lorenzo, K. Minamimoto, K. Murakami, P. Nieminen, L. Pandola, S. Parlati, L. Peralta, J. Perl, A. Pfeiffer, M. G. Pia, A. Ribon, P. Rodrigues, G. Russo, S. Sadilov, G. Santin, T. Sasaki, D. Smith, N. Starkov, S. Tanaka, E. Tcherniaev, B. Tome, A. Trindade, P. Truscott, L. Urban, M. Verderi, A. Walkden, J. P. Wellisch, D. C. Williams, D. Wright, and H. Yoshida. Geant4 developments and applications. *IEEE Trans. Nucl. Sci.*, 53(1):270–278, February 2006.

[2] Donald W. K. Andrews. Estimation When a Parameter is on a Boundary. *Econometrica*, 67(6):1341–1383, 1999.

[3] Peter Atkins and Julio De Paula. *Elements of Physical Chemistry*. Oxford University Press, March 2006.

[4] H. H. Barrett, T. White, and L. C. Parra. List-mode likelihood. *J. Opt. Soc. Am. A*, 14(1):2914–23, November 1997.

[5] M. J. Berger, J. H. Hubbell, S. M. Seltzer, J. Chang, J. S. Coursey, R. Sukumar, D. S. Zucker, and Olsen. XCOM: Photon cross sections database. *NIST Standard Reference Database 8 (XGAM)*, 2010.

[6] R. E. Bethel and K. L. Bell. Maximum likelihood approach to joint array detection/estimation. *Aerospace and Electronic Systems, IEEE Transactions on*, 40(3):1060–1072, September 2004.

[7] F. Biggs. Hartree-fock compton profiles for the elements. *Atomic Data and Nuclear Data Tables*, 16(3):201–309, September 1975.

[8] R. Billingsley. *Probability and measure*. Wiley, New York, 2 edition, 1986.

[9] C. Byrne. Sequential unconstrained minimization algorithms for constrained optimization. *Inverse Prob.*, 24(1):015013, February 2008.

[10] E. J. Candes, M. B. Wakin, and S. Boyd. Enhancing sparsity by reweighted l1 minimization, 2007.

[11] E. J. Candes, M. B. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted l1 minimization. *Journal of Fourier Analysis and Applications*, 14(5):877–905, December 2008.

[12] J-H. Chang, J. M. M. Anderson, and J. R. Votaw. Regularized image reconstruction algorithms for positron emission tomography. *IEEE Trans. Med. Imag.*, 23(9):1165–75, September 2004.

[13] U.S. Customs and Border Patrol. On a typical day in fiscal year 2009, cbp. `http://www.cbp.gov/xp/cgov/about/accomplish/fy09_typical_day.xml`.

[14] M. Dahlbom. Estimation of image noise in PET using the bootstrap method. *IEEE Trans. Nuc. Sci.*, 49(5):2062–6, October 2002.

[15] A. R. De Pierro. On the relation between the ISRA and the EM algorithm for positron emission tomography. *IEEE Trans. Med. Imag.*, 12(2):328–33, June 1993.

[16] A. R. De Pierro. A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography. *IEEE Trans. Med. Imag.*, 14(1):132–7, March 1995.

[17] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Stat. Soc. Ser. B*, 39(1):1–38, 1977.

[18] D. L. Donoho. De-noising by soft-thresholding. *IEEE Trans. Info. Theory*, 41(3):613–27, May 1995.

[19] J. Driscoll and D. Healy. Computing fourier transforms and convolutions on the 2-sphere. *Advances in Applied Mathematics*, 15(2):202–250, June 1994.

[20] B. Efron. *The jackknife, the bootstrap and other resampling plans.* SIAM, Philadelphia, 1982.

[21] H. W. Engl. Regularization methods for the stable solution of inverse problems. *Surveys on Mathematics for Industry*, 3:71–143, 1993.

[22] D. Feldman and M. Fox. *Probability: the mathematics of uncertainty.* Marcel Dekker, New York, 1991.

[23] T. S. Ferguson. *A Course in Large Sample Theory.* Chapman and Hall, 1996.

[24] J. A. Fessler. *Image reconstruction: Algorithms and analysis.* 2006. Book in preparation.

[25] J. A. Fessler and H. Erdogan. A paraboloidal surrogates algorithm for convergent penalized-likelihood emission image reconstruction. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, volume 2, pages 1132–5, 1998.

[26] J. A. Fessler and A. O. Hero. Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms. *IEEE Trans. Im. Proc.*, 4(10):1417–29, October 1995.

[27] C. J. Groiselle and S. J. Glick. Using the bootstrap method to evaluate image noise for investigation of axial collimation in hybrid PET. *IEEE Trans. Nuc. Sci.*, 52(1):95–101, February 2005.

[28] J. A. Hanley and B. J. McNeil. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143(1):29–36, April 1982.

[29] Z. He. Review of the shockley-ramo theorem and its application in semiconductor gamma-ray detectors. *Nuclear Instruments and Methods in Physics Research A*, 463:250–267, May 2001.

[30] Z. He, W. Li, G. F. Knoll, D. K. Wehe, J. Berry, and C. M. Stahle. 3-D position sensitive CdZnTe gamma-ray spectrometers. *Nucl. Instr. Meth. Phys. Res. A*, pages 173–178, February 1999.

[31] P. J. Huber. *Robust statistics.* Wiley, New York, 1981.

[32] S. E. Inderhees, B. F. Phlips, R. A. Kroeger, W. N. Johnson, R. L. Kinzer, J. D. Kurfess, B. Graham, and N. Gehrels. Spectroscopy, imaging and compton-scatter polarimetry with a germanium strip detector. *IEEE Transactions on Nuclear Science*, 43(3):1467–1471, June 1996.

[33] M. W. Jacobson and J. A. Fessler. An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms. *IEEE Trans. Im. Proc.*, 16(10):2411–22, October 2007.

[34] S. M. Kay. *Fundamentals of statistical signal processing: Estimation theory.* Prentice-Hall, New York, 1993.

[35] S. M. Kay. *Fundamentals of statistical signal processing: Detection theory.* Prentice-Hall, New York, 1998.

[36] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz. Rotation invariant spherical harmonic representation of 3d shape descriptors. In *SGP '03: Proceedings of the 2003 Eurographics/ACM SIGGRAPH symposium on Geometry processing*, pages 156–164, Aire-la-Ville, Switzerland, Switzerland, 2003. Eurographics Association.

[37] S. D. Kiff, Z. He, and G. C. Tepper. Improving spectroscopic performance of a coplanar-anode high-pressure xenon gamma-ray spectrometer. *IEEE Transactions on Nuclear Science*, 54(4):1263–1270, August 2007.

[38] G. F. Knoll. *Radiation detection and measurement.* Wiley, New York, 4 edition, 2010.

[39] S. Kullback and R. A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, March 1951.

[40] K. Lange, D. R. Hunter, and I. Yang. Optimization transfer using surrogate objective functions. *J. Computational and Graphical Stat.*, 9(1):1–20, March 2000.

[41] C. Lartizien, J. B. Aubin, and I. Buvat. Comparison of bootstrap resampling methods for 3-D pet imaging. *IEEE Trans. Med. Imag.*, 29(7):1442–54, 2010.

[42] J. W. Leblanc, N. H. Clinthorne, C. H. Hua, E. Nygard, W. L. Rogers, D. K. Wehe, P. Weilhammer, and S. J. Wilderman. Experimental results from the c-sprint prototype compton camera. *Nuclear Science, IEEE Transactions on*, 46(3):201–204, 1999.

[43] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In *ICML '09: Proceedings of the 26th Annual International Conference on Machine Learning*, pages 609–616, New York, NY, USA, 2009. ACM.

[44] C. Lehner. *4-pi Compton imaging using a single 3-d position sensitive CdZnTe detector.* PhD thesis, Univ. of Michigan, Ann Arbor, MI, 48109-2122, Ann Arbor, MI., 2004.

[45] D. J. Lingenfelter and J. A. Fessler. Augmented lagrangian methods for penalized likelihood reconstruction in emission tomography. In *IEEE Medical Imaging Conference Record.* IEEE, November 2010.

[46] D. J. Lingenfelter, J. A. Fessler, and Z. He. Sparsity regularization for image reconstruction with Poisson data. In *Proc. SPIE 7246, Electronic Imaging 2008: Computational Imaging VII*, page 72460F, 2009.

[47] D. J. Lingenfelter, J. A. Fessler, C. D. Scott, and Z. He. Benefits of position-sensitive detectors for source detection with known background. In *IEEE Nuclear Science Symposium Conference Record*, pages 636–640. IEEE, October 2009.

[48] D. J. Lingenfelter, J. A. Fessler, C. D. Scott, and Z. He. Benefits of position-sensitive detectors for radioactive source detection. *IEEE Trans. Sig. Proc.*, 58(9):4473–83, September 2010.

[49] D. J. Lingenfelter, J. A. Fessler, C. D. Scott, and Z. He. Predicting ROC curves for source detection under model mismatch. In *IEEE Nuclear Science Symposium Conference Record*, November 2010.

[50] D. J. Lingenfelter, J. A. Fessler, C. D. Scott, and Z. He. Asymptotic source detection performance of gamma-ray imaging systems under model mismatch. *IEEE Trans. Sig. Proc.*, 59(11):5141–5151, November 2011.

[51] Z. Liu and A. Nehorai. Detection of particle sources with directional detector arrays and a mean-difference test. *Signal Processing, IEEE Transactions on*, 53(12):4472–4484, 2005.

[52] Z. Liu and A. Nehorai. Statistical angular resolution limit for point sources. *Signal Processing, IEEE Transactions on*, 55(11):5521–5527, 2007.

[53] D. J. C. MacKay. Introduction to Monte Carlo methods. In M. I. Jordan, editor, *Learning in Graphical Models*, pages 175–204. Kluwer, Dordrencht, 1998.

[54] J. B. Martin, G. F. Knoll, D. K. Wehe, N. Dogan, V. Jordanov, N. Petrick, and M. Singh. A ring compton scatter camera for imaging medium energy gamma rays. *Nuclear Science, IEEE Transactions on*, 40(4):972–978, 1993.

[55] V. Maxim, M. Frandeş, and R. Prost. Analytical inversion of the compton transform using the full set of available projections. *Inverse Problems*, 25(9):095001+, September 2009.

[56] C. E. Ordonez, A. Bolozdynya, and W. Chang. Doppler broadening of energy spectra in compton cameras. In *Nuclear Science Symposium, 1997. IEEE*, volume 2, pages 1361–1365 vol.2, 1997.

[57] F. O'Sullivan. Imaging radiotracer model parameters in PET: a mixture analysis approach. *IEEE Trans. Med. Imag.*, 12(3):399–412, September 1993.

[58] A. Papoulis. *Probability, random variables, and stochastic processes*. McGraw-Hill, New York, 2 edition, 1984.

[59] A. Papoulis and S. U. Pillai. *Probability, random variables, and stochastic processes*. McGraw-Hill, 2002.

[60] L. C. Parra and H. H. Barrett. List-mode likelihood - EM algorithm and noise estimation demonstrated on 2D-PET. *IEEE Trans. Med. Imag.*, 17(2):228–35, April 1998.

[61] A. E. Perkins, G. Muehllehner, S. Surti, and J. S. Karp. Performance measurements of a pixelated nai(t1) pet scanner. *IEEE Transactions on Nuclear Science*, 50(3):373–377, June 2003.

[62] R. Protassov, D. A. van Dyk, A. Connors, V. L. Kashyap, and A. Siemiginowska. Statistics: Handle with care, detecting multiple model components with the likelihood ratio test. *Astrophys. J.*, 571:545–559, Jan 2002.

[63] D. Protic, E. L. Hull, T. Krings, and K. Vetter. Large-volume si(li) orthogonal-strip detectors for compton-effect-based instruments. *IEEE Transactions on Nuclear Science*, 52(6):3181–3185, December 2005.

[64] J. Qi and R. H. Huesman. Propagation of errors from the sensitivity image in list mode reconstruction. *IEEE Trans. Med. Imag.*, 23(9):1094–9, September 2004.

[65] S. M. Robinson, W. R. Kaye, J. E. Schweppe, and E. R. Siciliano. Optimal background attenuation for fielded radiation detection systems. *IEEE Transactions on Nuclear Science*, 54(4):1279–1284, August 2007.

[66] R. C. Runkle, T. M. Mercier, K. K. Anderson, and D. K. Carlson. Point source detection and characterization for vehicle radiation portal monitors. *IEEE Trans. Nucl. Sci.*, 52(6):3020–3025, 2005.

[67] R. C. Runkle, M. F. Tardiff, K. K. Anderson, D. K. Carlson, and L. E. Smith. Analysis of spectroscopic radiation portal monitor data using principal components analysis. *IEEE Transactions on Nuclear Science*, 53(3):1418–1423, June 2006.

[68] P. C. Schaich, G. A. Clark, S. K. Sengupta, and K. P. Ziock. Automatic image analysis for detecting and quantifying gamma-ray sources in coded-aperture images. *Nuclear Science, IEEE Transactions on*, 43(4):2419–2426, 1996.

[69] S. G. Self and K. Y. Liang. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398):605–610, 1987.

[70] R. J. Serfling. *Approximation theorems of mathematical statistics*. Wiley, New York, 1980.

[71] T. A. Severini. *Elements of distribution theory*. Cambridge University Press, New York, NY, 2005.

[72] J. Shao. *Mathematical Statistics*. Springer Science, New York, NY, 2003.

[73] S. Vallabhajosula. *Molecular Imaging: Radiopharmaceuticals for PET and SPECT*. Springer-Verlag, Berlin Heidelberg, 2009.

[74] H. L. Van Trees. *Detection, estimation, and modulation theory*. Wiley, New York, 1968.

[75] C. G. Wahl and Z. He. Sensitivity of gamma-ray source detection using 3D-position-sensitive semiconductor detectors. In *Nuclear Science Symposium Conference Record, 2008. NSS '08. IEEE*, pages 3334–3338, 2008.

[76] C. G. Wahl and Zhong He. Point-source detection using energy and imaging information from 3D-position-sensitive semiconductor detectors. In *2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC)*, pages 1069–1073, October 2009.

[77] H. White. Consequences and detection of misspecified nonlinear regression models. *Journal of the American Statistical Association*, 76(374):419–433, 1981.

[78] H. White. Maximum likelihood estimation of misspecified models. *Econometrica*, 50(1):1–25, 1982.

[79] D. Xu. *Gamma-ray imaging and polarization measure using 3-D position-sensitive CdZnTe detectors*. PhD thesis, University of Michigan, 2006.

[80] D. Xu and Z. He. Gamma-ray energy-imaging integrated spectral deconvolution. *Nucl. Instr. Meth. Phys. Res. A.*, 574(1):98–109, April 2007.

[81] D. Xu, Z. He, C. E. Lehner, and F. Zhang. 4-pi Compton imaging with single 3D position-sensitive CdZnTe detector. In *Proc. SPIE 5540, Hard X-Ray and Gamma-Ray Detector Physics VI*, pages 144–55, 2004.

[82] L. Yan, R. Dodier, M. C. Mozer, and R. Wolniewicz. Optimizing Classifier Performance via an Approximation to the Wilcoxon-Mann-Whitney Statistic. In *Proceedings of the Twentieth International Conference on Machine Learning*, August 2003.

[83] F. Zhang, Z. He, and C. E. Seifert. A prototype three-dimensional position sensitive CdZnTe detector array. *IEEE Trans. Nucl. Sci.*, 54(4):843–848, August 2007.

[84] K. P. Ziock, W. W. Craig, L. Fabris, R. C. Lanza, S. Gallagher, B. K. P. Horn, and N. W. Madden. Large area imaging detector for long-range, passive detection of fissile material. *Nuclear Science, IEEE Transactions on*, 51(5):2238–2244, 2004.