List-mode SPECT reconstruction using empirical likelihood

by

Andre Lehovich

A Dissertation Submitted to the Faculty of the

Graduate Interdisciplinary Program
in Applied Mathematics

In Partial Fulfillment of the Requirements
For the Degree of

Doctor of Philosophy

In the Graduate College

The University of Arizona

2 0 0 5

The University of Arizona
Graduate College

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by Andre Lehovich
entitled *List-mode SPECT reconstruction using empirical likelihood*,
and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy.

Harrison H. Barrett	_____	Date: 5 April 2005
Eric Clarkson	_____	Date: 5 April 2005
Joseph Watkins	_____	Date: 5 April 2005
Matthew Kupinski	_____	Date: 5 April 2005

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Harrison H. Barrett	5 April 2005
Dissertation Director	Date

Statement by Author

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Signed: Andre Lehovich

Acknowledgments

I have greatly enjoyed my years at the Center for Gamma Ray Imaging. I have learned much of what I know about imaging science from Harry Barrett and Eric Clarkson. I would also like to thank the other members of my committee, Joe Watkins and Matt Kupinski, and my external reviewer, Lucas Parra, for their help with the dissertation.

I have also learned a tremendous amount from the other students in the lab. I'd especially like to thank Jean Chen and Bill Hunter for helping me with the MDRF experiments.

Data from small animals was collected by Zhonglin Liu and Gail Stevenson under their Institutional Animal Care and Use Committee (IACUC) approved protocols. The images shown here were produced by independently reprocessing their data.

This work was in part supported by NIH grants R37-EB000803 and P41-EB002035 and NSF VIGRE grant DMS-9771116.

Last, though not least, I am grateful to my parents, my sister, and the rest of the family for their patience and encouragement.

Table of Contents

Table of Contents—*Continued*

TABLE OF CONTENTS—*Continued*

List of Figures

List of Tables

Abstract

This dissertation investigates three topics related to image reconstruction from list-mode Anger camera data. Our main focus is the processing of photomultiplier-tube (PMT) signals directly into images.

First we look at the use of list-mode calibration data to reconstruct a non-parametric likelihood model relating the object to the data list. The reconstructed model can then be combined with list-mode object data to produce a maximum-likelihood (ML) reconstruction, an approach we call double list-mode reconstruction. This trades off reduced prior assumptions about the properties of the imaging system for greatly increased processing time and increased uncertainty in the reconstruction.

Second we use the list-mode expectation-maximization (EM) algorithm to reconstruct planar projection images directly from PMT data. Images reconstructed by EM are compared with images produced using the faster and more common technique of first producing ML position estimates, then histograming to form an image. A mathematical model of the human visual system, the channelized Hotelling observer, is used to compare the reconstructions by performing the Rayleigh task, a traditional measure of resolution. EM is found to produce higher resolution images than the histogram approach, suggesting that information is lost during the position estimation step.

Finally we investigate which linear parameters of an object are estimable, in other words may be estimated without bias from list-mode data. We extend the notion of a linear system operator, familiar from binned-mode systems, to list-mode systems, and show the estimable parameters are determined by the range of the adjoint of the system operator. As in the binned-mode case, the list-mode sensitivity functions define "natural pixels" with which to reconstruct the object.

Chapter 1

Introduction

Nuclear medicine is used to study physiological function by answering questions such as "Where in the body does a molecule accumulate?", "How fast is the uptake?", and "What is the washout rate?" By tracking molecules with known physiology, especially those which behave differently in healthy versus diseased tissue, one can also answer clinical questions such as "How much did the heart attack damage this patient's myocardium?", or "Has cancer spread to other parts of the body?" This chapter provides a brief introduction to nuclear-medicine imaging and the mathematical modeling of imaging systems. The reader desiring further depth is referred to Chandra [1992], Bushberg et al. [1994], and Barrett and Myers [2004].

1.1 The imaging chain

We begin, as has by now become traditional, with a discussion of the imaging chain sketched in figure 1.1. Our emphasis is on the imaging systems used in nuclear medicine, though a similar diagram can be made for any imaging modality.

The imaging process begins by choosing the object to be imaged. In our case this will be part of an animal or person, but in other contexts could mean a building, vehicle, or almost anything else. Almost all imaging systems are designed to image a variety of objects. For example, we expect medical systems to image many patients and both healthy and diseased tissue. In keeping with statistical parlance, we shall refer to the collection of objects we might wish to image as the population or ensemble.

In any imaging system, energy is transferred from the object to the imaging system. In nuclear medicine this energy takes the form of gamma-ray photons emitted

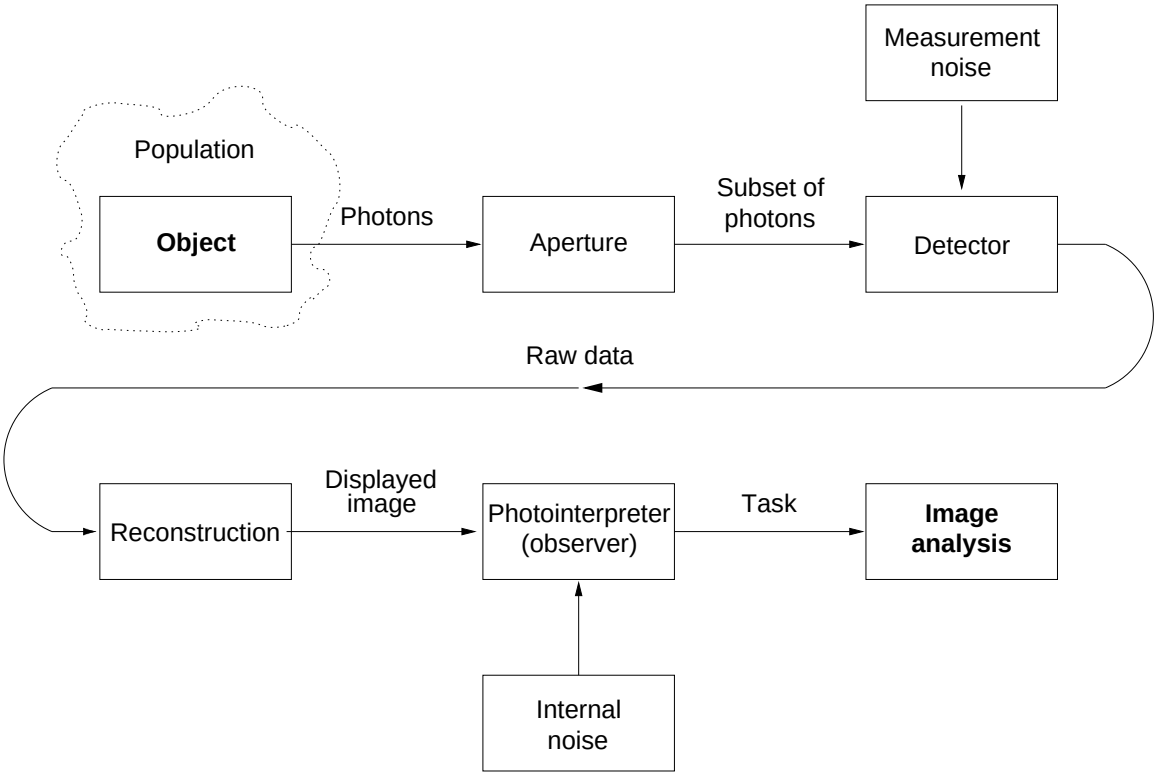

FIGURE 1.1. **Block diagram of the imaging chain for a nuclear-medicine system.** See discussion in section 1.1.

from inside the object; in other modalities it might be reflected energy. Gamma-rays are emitted in all directions, so an aperture is used to restrict which photons reach the detector. This has the effect of forming an image (really a fluence pattern) on the camera face. The detector then converts the incoming gamma rays into raw data, which are then transformed by a reconstruction algorithm into an image. (In 35 mm film systems the raw data correspond to the negative, and reconstruction is done by making a print.)

The image is then displayed to a photointerpreter, also called the observer, who extracts information from the image and produces the image analysis. The analysis may consist of a simple decision—"This patient may have cancer and needs followup" vs. "This patient does not have cancer"—or may have a quantitative component—"The liver contains three lesions bigger than 10 mm in diameter." We refer to the type of image analysis as the imaging task. The overall accuracy of analysis is called task performance. (Note a potential source of vocabulary confusion: in the remote-sensing community "tasking" refers to the scheduling of which objects the satellite will image and which images the photointerpreter will analyze.)

Computer-aided diagnosis (CAD) and other image-enhancement techniques are considered to be part of the reconstruction algorithm. The imaging chain is likely to include humans for the foreseeable future. Despite decades of artificial intelligence (AI) research, we still don't know how to create a machine observer for most imaging tasks. There are also legal considerations—in a medical context, most of the malpractice lawsuit risk is carried by the radiologist, not the imaging system manufacturer.

The remainder of this chapter will give more details about various parts of the nuclear-medicine imaging chain, with an eye towards studying the reconstruction process and its impact on task performance.

1.2 Nuclear medicine

Radiochemists can attach a radioactive atom, called the tracer, to almost any molecule of physiological interest. For popular molecules the pharmaceutical industry has created off-the-shelf kits to do the tagging. Because the tracer travels with the molecule we can determine where the tagged molecule is concentrating in the body by detecting which parts of the body are radioactive or "hot". The radiotracers we are interested in produce gamma rays as they decay. Many of these gamma rays escape the body. Detecting this emitted radiation is the basis of all nuclear-medicine imaging techniques. In imaging our goal is to use information from the detected radiation to produce a map of the radiotracer distribution in the body. If the radiotracer does not change the molecule's biochemistry, then this is equivalent to mapping the molecule's distribution. By looking for deviations from a "normal" map, radiologists can diagnose pathological conditions. Physiologists also use these maps for basic science.

For example, suppose we want to know if a cancer patient's tumor has spread to other parts of the body. By injecting a tracer-tagged substance that concentrates in tumors we can force them to become radioactive. (Note that the goal is to identify where the tumors are, not to damage the tumors with radiation. Isotopes used in imaging typically cause minimal tissue damage. In general, radiation oncology uses different isotopes than those used for imaging.) The simplest way to build up a map of radiotracer distribution is to scan a directionally-sensitive radiation detector, for example a collimated scintillation detector, over the patient, recording the activity at each location. Since tumors have high activity, they will stand out on the map. The rectilinear scanner, the first nuclear-medicine imaging system, works this way [Cassen et al., 1951]. Rectilinear scanners were in use at University Medical Center (UMC) until the late 1970s.

Building images up one pixel at a time is slow; most modern nuclear-medicine systems use gamma cameras that simultaneously record many pixels at one time.

The images produced by the rectilinear scanner or a gamma camera are planar projection images, so named because they show a projection of the radiotracer distribution onto the imaging plane. Nuclear-medicine projection images are similar to the more-familiar x-ray projection images used to check for broken bones. Examples of projection images from different angles are shown in figure 1.2.

Projection images are widely used, but for many imaging tasks tomographic[1] images are useful, since they show the activity located in a specific cross section. Typically projection images are directly measured by using a single gamma camera, whereas tomographic images are usually reconstructed[2] by processing projection data taken from many different angles. Figure 1.3 provides example tomographic images of a mouse. In projection images we decide which angles to look from; in tomographic imaging we decide through which planes to slice cross-sections. (The modular gamma cameras described below do not directly output projection images; rather the raw data needs to be reconstructed into a projection image.)

There are two types of tomographic imaging widely used in nuclear medicine: Single-Photon-Emission Computed Tomography (SPECT) and Positron-Emission Tomography (PET). PET will receive little attention in this dissertation, because it was not in use at the University of Arizona during my time there. PET had been used at UMC briefly during the mid-90s; as I was finishing, a new small-animal PET system was under development.

In human patients, nuclear medicine is most commonly used for clinical diagnostic tasks. Nuclear-medicine techniques are also starting to be used with small animals such as mice and rats. Here the primary interest is research, though doubtless these techniques will eventually also be used to treat pets. Since SPECT is done in vivo,

[1] From Greek $\tau\acute{o}\mu o\varsigma$ (slice or section) and $-\gamma\rho\alpha\phi o\varsigma$ (written) [OED 1989].

[2] The "re" in reconstructed is a vocabulary artifact due to the way mathematicians think about the imaging process. The radiotracer constructs the "true" cross-section image in the body, which physics transforms into the measured projection images. We then invert the projection process to reconstruct our estimate of the "true image".

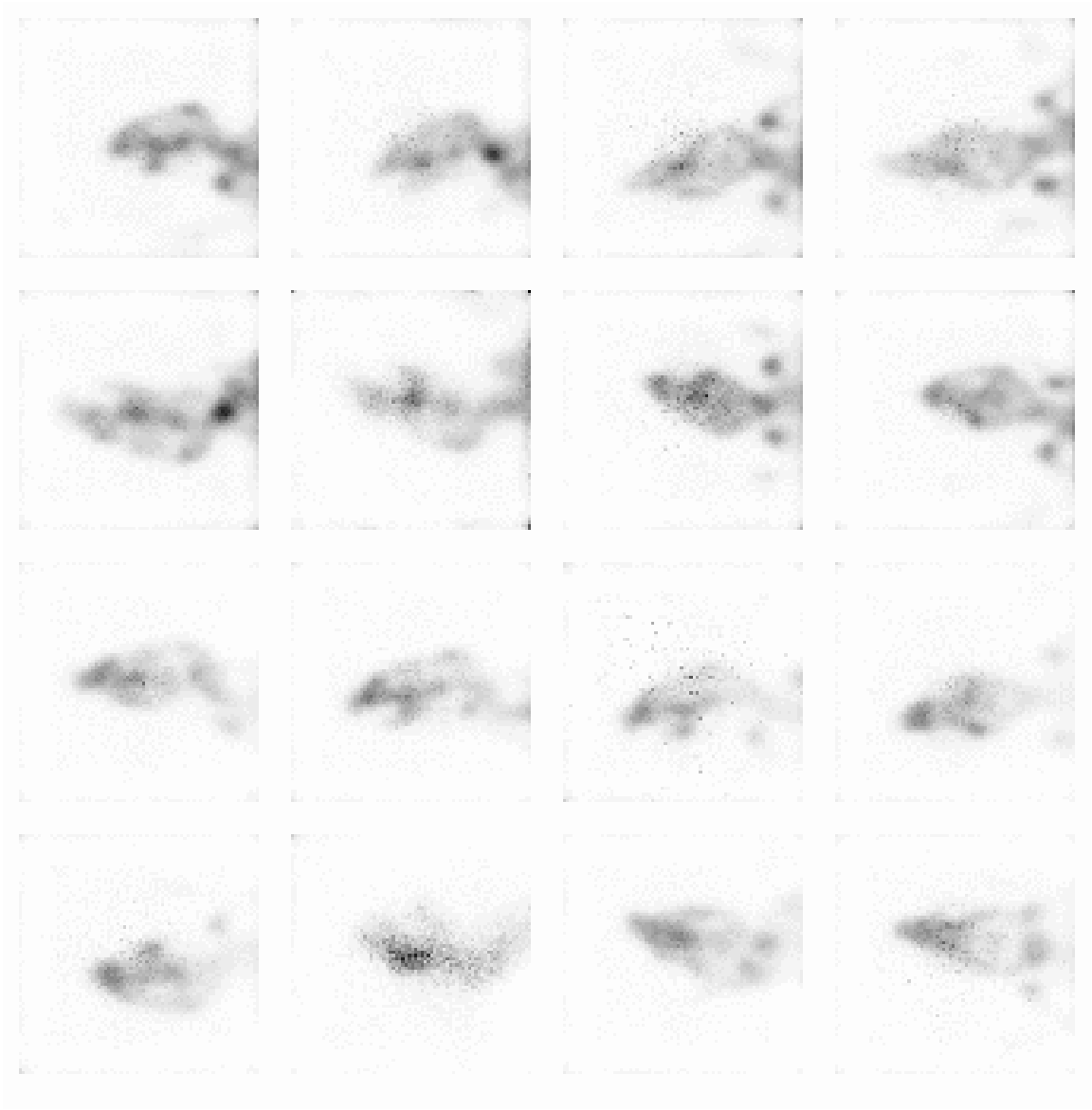

FIGURE 1.2. **Projection images from an MDP mouse bone-scan.** Shown are simultaneous views of the skull from all 16 FastSPECT-II cameras. As discussed in section 1.3, raw data from these cameras must be reconstructed to form the projection image. These images were reconstructed from raw-PMT data using two iterations of the 2-dimensional list-mode EM algorithm described in section 1.7.2.

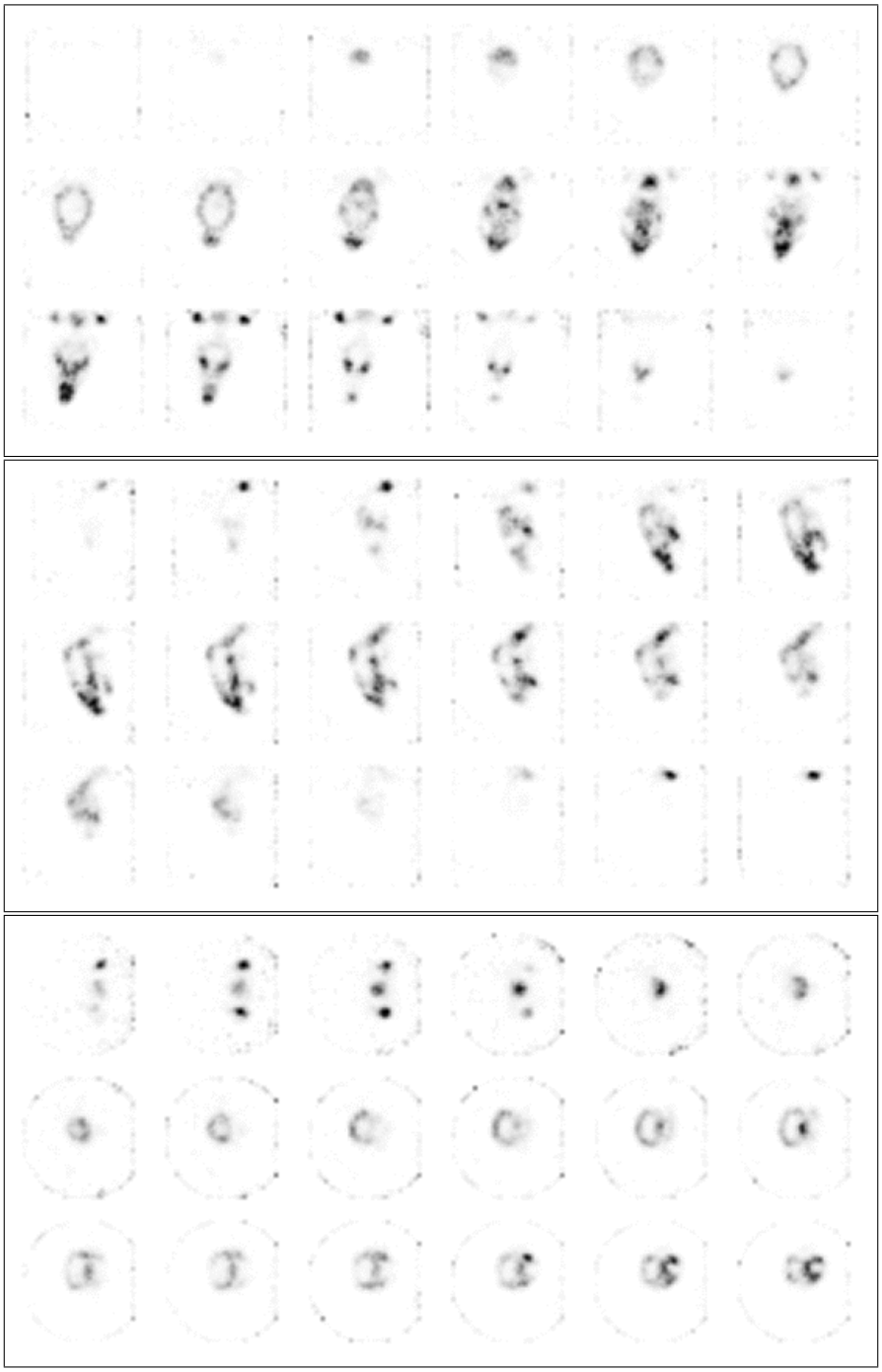

FIGURE 1.3. **Tomographic images of a mouse skull bone-scan.** Data from the same scan shown in figure 1.2 is reconstructed using 10 iterations of the 3-dimensional list-mode EM algorithm. Each sub-image is a different slice through the mouse; the top 18 are coronal slices (nose down), the middle are sagittal slices (nose down), and the bottom are transverse slices.

one can track an animal over days or weeks, monitoring the progress of a treatment or disease. Many other imaging techniques, in particular autoradiography, require sacrificing the animal.

Animal imaging can be used in preliminary drug trials for oncology. One can learn about a proposed drug's efficacy by comparing the size of mouse-grown xenograft tumors in the presence and absence of the drug. By tagging the drug with a radiotracer, we can determine to what extent it accumulates in the tumor and any other organs. Another important application is gene-expression imaging, which is used to determine where in the body a gene of interest is being expressed.

1.2.1 SPECT

The radioisotopes used in SPECT are gamma-ray emitters; upon decay the tracer atom produces a single high-energy photon. (PET uses different isotopes, as may autoradiography.) The most widely used SPECT isotope, and the one used for all of the images presented in this dissertation, is Technetium (^{99m}Tc) with a half-life of 6 hours and a gamma-ray energy of 140 keV. Technetium is a decay product of Molybdenum (^{99}Mo), from which it is easy to separate chemically in the form of pertechnetate ($^{99m}TcO_4^-$).

The pertechnetate ion is then attached to the molecule of interest, chosen for the clinical or research task at hand. For example in cardiac or tumor imaging one might label sestamibi which accumulates in mitochondria; for a bone scan one could label MDP (methylenediphosonate) which accumulates in areas of osteogenesis; or to detect stroke one would label HMPAO (hexamethylpropyleneamine oxime) which crosses the blood-brain barrier and accumulates in cerebral tissue.

Next the radiotracer-tagged molecule is injected into the patient. In humans the injection happens quickly, as anybody who has received a shot expects. Unfortunately a rapid injection is likely to kill a small animal, so in mice the process takes anywhere

from one to three minutes. To ensure a constant injection rate, the hypodermic plunger is often controlled by a machine. The molecule usually takes a while to concentrate in the organ of interest, so there may be a delay of between one-half and two hours after the injection and before commencement of imaging. In first-pass studies imaging is done during injection, so the molecule's first trip through the circulatory system can be tracked.

When projection images are desired, it typically suffices to collect data from a single angle. However producing tomographic images usually requires data collected from many angles. In most clinical systems a single gamma camera is rotated around the patient to produce projection images from different angles. To speed acquisition time some systems have two or three rotating cameras. Another option, used by research systems such as Arizona's FastSPECT and FastSPECT-II systems, is to have many fixed cameras simultaneously acquire images from multiple projections.

The tracer is continuously emitting gamma-rays as it undergoes radioactive decay. Very few of these photons are detected by the SPECT system. The system sensitivity, in other words the proportion of emitted photons detected by the SPECT imager, is typically on the order of 10^{-5}. FastSPECT-II has a sensitivity of 4×10^{-4} [Furenlid et al., 2004].

Most SPECT systems detect gamma rays using variations on the Anger camera [Anger, 1958]. Replacing Anger cameras with semiconductor detectors remains a hot area of research. Several systems already use semiconductor detectors, but this technology is not yet widely used in the clinic. All of the gamma-ray detectors we discuss below are variations on the original Anger design.

The Anger camera detects a gamma ray in a several-step process. (Some photons pass right through the camera without being detected.) First the incoming high-energy photon interacts with a scintillation crystal through Compton scatter or photoelectric absorption, freeing one or more high-energy electrons. (Pair production does not occur at the energy levels we are interested in.) By further interacting with

the crystal, the freed electrons produce a shower of lower-energy visible-light photons which spread through the crystal. If the initial interaction is a Compton scatter, the scattered photon may trigger a second shower at a different location in the crystal. Coupled to the crystal are a number of photomultiplier tubes (PMTs). The PMTs convert the visible-light photons into current which serves as input to the camera's detection-acquisition electronics. Ideally, the visible photons produced by one scintillation event reach many PMTs—this is one of the counter-intuitive cases where blur helps. In most commercial systems the camera electronics convert the PMT readings into an estimate of the camera-face location where the scintillation event occurred and an estimate of the energy of the incoming high-energy photon; we will discuss an alternative approach below.

1.3 Arizona modular gamma camera

The Anger cameras usually used in the clinic feature large crystals coupled to dozens of PMTs. This approach works well, but has the disadvantage that it is not possible to pack a number of cameras closely around the object. The modular gamma camera was developed to allow denser packing of the cameras and quick reconfiguration of the system geometry. Each module consists of a crystal and small number of PMTs. Two generations of camera, an old design developed for FastSPECT and a newer design for FastSPECT-II, are in use. Both are enhancements of the original Anger design.

The older model used in FastSPECT has a 100 mm $\times$ 100 mm $\times$ 5 mm crystal of sodium iodide doped with thallium, NaI(Tl), coupled to four PMTs by a 19 mm thick exit window [Milster et al., 1990]. The hygroscopic crystal is hermetically sealed to prevent degradation due to humidity. The sealed crystal and PMTs are housed in an aluminum can to shield them from external light and make the module easy to move. Output from the PMTs is digitized and processed to estimate the location and

energy of each scintillation event. A histogram of the locations is then used as the projection image.

The second-generation cameras used in FastSPECT-II have a 125 mm × 125 mm × 5 mm NaI(Tl) crystal connected with a 15 mm thick quartz light-guide to nine 1.5" diameter PMTs [Sain, 2001]. Figure 1.4 shows a sketch of this model of camera. Current output from the nine tubes is processed by custom electronics and transferred to a computer workstation in digital form. The raw data corresponding to one scintillation event comprise a 9-dimensional vector of 12-bit integers (11 data bits and one sign bit), optionally augmented with a time-stamp.

An example of the raw data is shown in figure 1.5. Each row shows the PMT outputs for one scintillation event. The columns show the output of one PMT for many events. A typical projection image will be reconstructed using thousands or millions of events.

1.4 FastSPECT-II

FastSPECT-II is a small-animal imaging system based on the second-generation 9-PMT modular camera [Furenlid et al., 2004]. In its usual configuration the system can handle mice, rats, and similarly sized animals. The system can also be reconfigured to handle rabbits and cats, but changing the aperture and recalibrating the cameras requires several days.

The animal lies on its stomach or back on a tray and is moved into the system down a central pipe. This is similar to the way a person enters a closed-bore MRI system, though on a smaller scale. Cables and tubes to monitor the animal and deliver anesthetics, heating, and the radiotracer injection all run down the central pipe.

The central pipe is surrounded by 16 cameras arranged in two parallel rings of 8. (See figure 1.6.) During imaging the cameras are fixed in the housing; the animal

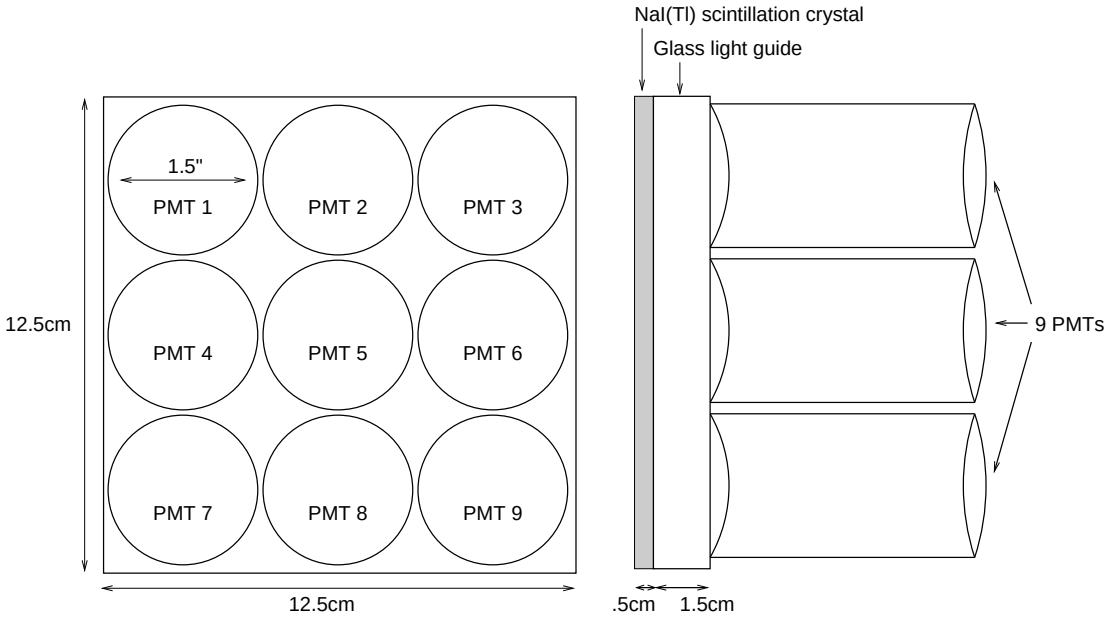

FIGURE 1.4. **Sketches of the 9-PMT modular gamma cameras used in FastSPECT-II.** On the left is an illustration of the camera face, on the right a side view. (Not to scale.)

PMT1	PMT2	PMT3	PMT4	PMT5	PMT6	PMT7	PMT8	PMT9
26	39	9	89	446	108	172	1463	460
51	60	24	250	302	64	599	440	73
41	69	47	261	659	109	164	926	275
1551	178	20	474	117	10	46	25	13
3	16	19	66	291	150	228	1241	357
13	0	1	61	107	31	66	499	843
26	41	18	41	105	41	143	1028	172
14	48	25	93	267	102	267	1481	419
9	145	159	58	653	492	7	176	204
27	11	17	44	78	110	82	855	497
25	27	32	50	151	132	88	1120	465

FIGURE 1.5. **Raw list-mode modular-camera data.** Shown are the digitized PMT voltages from 11 scintillation events collected during a mouse bone-scan. The optional time-stamp for each event is not shown.

must be moved to adjust the field of view. If data from more than 16 angles are desired, the tray must be tilted about the central axis. One drawback is that tilting the animal changes the direction of gravity, potentially slightly changing the shape of internal organs.

The central pipe is made of lead so it acts as shielding. In front of each camera a hole has been drilled in the pipe, allowing emitted gamma rays to reach each camera. Each hole is fitted with gold pinhole aperture which forms the fluence pattern on the camera face. The central pipe and pinholes may both be swapped out to change the system magnification or to use a coded aperture.

Positioning of the animal tray and data acquisition are controlled by an off-the-shelf personal computer running Windows NT® and LabVIEW® software. After imaging is complete, data files are transferred over the network to the computer(s) responsible for reconstruction.

1.5 List-mode data

Although Anger cameras detect individual scintillation events, traditionally the data are aggregated into a projection image, and information about each individual event is discarded. Though implementation details vary, the usual approach can be thought of as performing position estimation on each event, with the projection image provided by a two-dimensional histogram of the positions. For this reason we refer to this approach as histogram-mode or binned-mode acquisition. If photon energy has also been estimated a three-dimensional histogram may be used instead [Gagnon et al., 1989], or the energy estimate can be used to accept or reject the event.

Modern nuclear-medicine imaging systems typically also have the ability to preserve information about each detected scintillation event. Because the data consist of a list of events, this is called list-mode acquisition. FastSPECT-II's modular gamma cameras always record data in list mode; when projection images are desired they

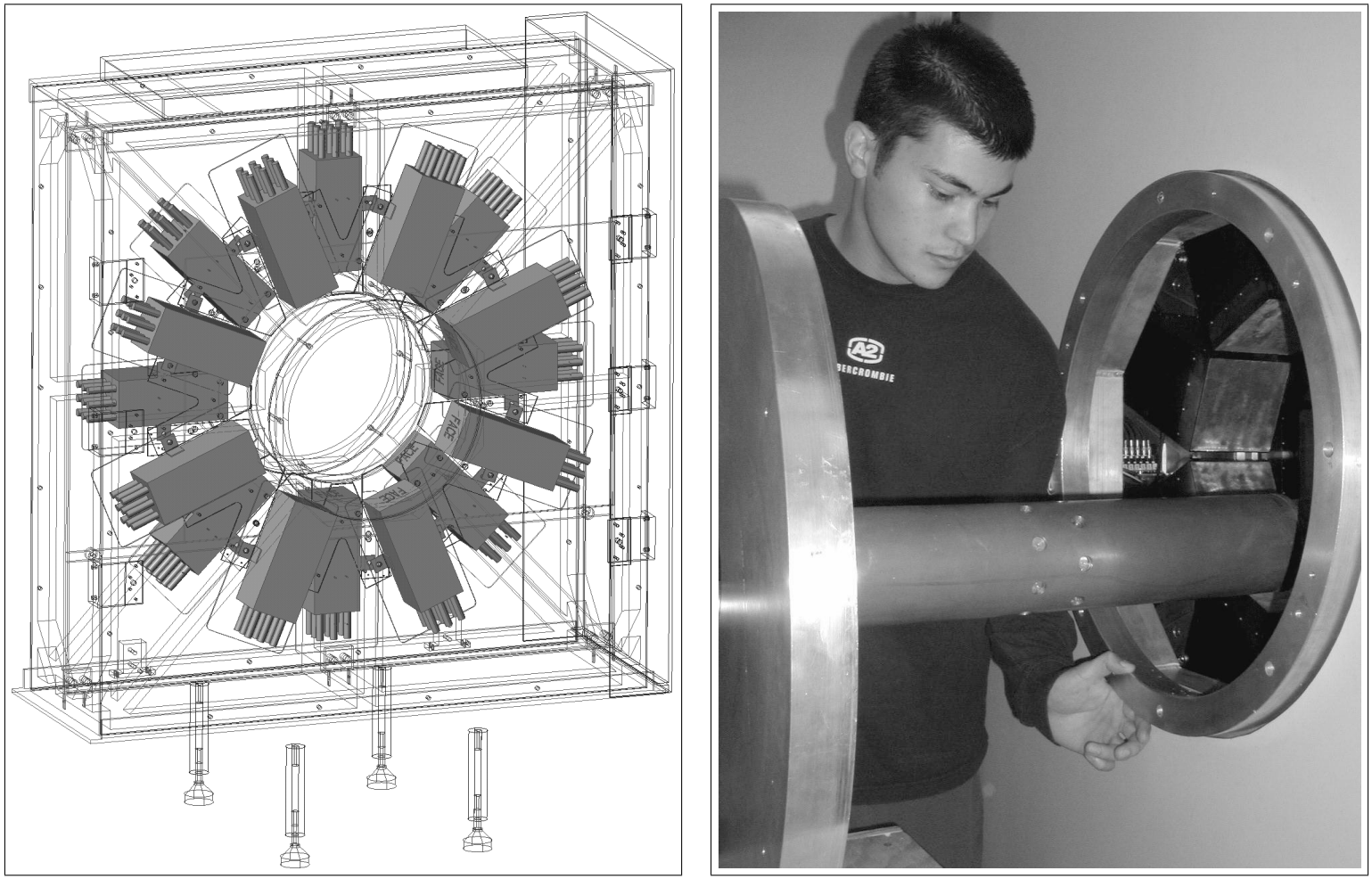

FIGURE 1.6. **FastSPECT-II.** On the left is a schematic image of FastSPECT-II, rendered from the original CAD design. The 16 cameras are solid, the housing and structure are wire-frame. On the right Lance Fesler shows the central lead pipe and gold pinhole collimators. Several modular cameras are visible inside the housing. (CAD by J. Butters, T. Lerdal, S. Lieber, and B. Stuckey. Photo by L. Furenlid.)

must first be reconstructed by processing the list. Several ways of doing so are compared in chapter 3.

Currently most imaging scientists think of list-mode SPECT imaging as recording the estimated position and energy of the scintillation event, perhaps augmented with a time stamp, because this is similar to how clinical PET systems work. However, as we saw in figure 1.5, a list entry can instead consist of the raw PMT readings. In dynamic studies the event's time stamp or stage in the cardiac cycle may also be included in the list. A list entry potentially includes anything one wishes to measure.

It is usual to produce tomographic images starting from projection images. For obvious reasons this is called projection-mode reconstruction. Even when the original raw data are list-mode, if projection images are generated during a preliminary step we shall call it projection-mode tomographic reconstruction. In section 1.7.2 we will discuss how to produce images directly from list-mode data; we refer to this as list-mode reconstruction.

We must distinguish between a list entry and how that entry is represented in the computer. For example, gamma rays detected by FastSPECT-II are recorded in a separate file for each camera. The camera ID is implicitly part of each event's list entry, even though it is stored in the file name, not inside the file together with the digitized PMT values and time stamps.

1.5.1 Advantages of list-mode data acquisition

Collecting data in list-mode offers several compelling advantages:

1. List-mode acquisition electronics are now easier to design, cheaper to build, and more flexible than hardware that produces binned-mode images. (Historically this has not always been the case.) Although designed for the 9-PMT modular camera, the FastSPECT-II electronics require no modifications to collect data from the older 4-PMT camera, from a position-sensitive PMT (PSPMT), or

from a modular camera designed to detect the higher-energy photons produced by PET tracers. Thus one may prefer list-mode acquisition even when doing projection-mode tomographic reconstruction.

2. If the projection image, or equivalently the list histogram, is sparse, then list-mode provides natural compression. This can occur with high resolution detectors and low count rates. In the case of FastSPECT-II the 9-dimensional histogram has $4096^9 = 2^{108} \approx 3.2 \times 10^{32}$ bins, so some form of compression is mandatory. (For comparison, a terabyte is $2^{40} \approx 1.1 \times 10^{12}$ bytes.) In high-resolution PET systems a sparse representation is also needed.

3. Because the list contains the full raw data, no information is lost. This allows us to compare different reconstruction algorithms on the same data. In particular we can compare different ways of producing projection images and study other data-reduction strategies.

4. For dynamic studies list-mode data has nanosecond time resolution; binned-mode systems require much longer to record an image. If EKG data are also recorded we get better cardiac-cycle resolution than with gated projection images. Other advantages of list-mode acquisition in dynamic imaging are described by Koss et al. [1997].

5. Lower-energy events due to a photon that underwent Compton scatter within the subject are usually discarded prior to generation of projection images, and therefore do not contribute to projection-mode reconstructions. Making use of scattered photons could improve image quality for a fixed radiotracer dose. How to extract information from these events in list-mode tomographic reconstruction is an ongoing area of research.

6. List-mode tomographic reconstruction may be faster than projection-mode reconstruction when we are in a low-count regime. In SPECT this situation is rare,

but it can occur in PET. (The list-mode tomographic reconstruction of SPECT data using the EM algorithm takes substantially longer than projection-mode reconstruction from the same data.)

1.6 Imaging-system modeling

In order to reconstruct a projection or tomographic image from the raw list-mode data, we need a model of the imaging system relating the object to the data and a reconstruction algorithm. Reconstructing a tomographic image is a special case of the more general problem of 3-dimensional or volume reconstruction. Because we are detecting the tracer's radioactive decay, a stochastic process, the model should be statistical in nature: multiple nuclear-medicine images of the same static object will *always* be different. This section provides a brief introduction to the mathematical models used for image and volume reconstruction. A more comprehensive treatment, including an extensive review of the mathematical background, is given by Barrett and Myers [2004].

The conventional wisdom in nuclear medicine is that photon-counting noise is the limiting factor on task performance, otherwise one would reduce the patient's radiation dose by lowering the amount of radioactive material injected. This is true for human imaging, but not necessarily in small-animal imaging where the long-term health effects of ionizing radiation are not a concern if the animal will be sacrificed after a few weeks. The amount of radiotracer injected in animal studies is proportionally much higher than in human imaging. It is common to give a mouse an amount of activity that is comparable to what would be given a human, even though there are several orders of magnitude difference in body mass. A good rule of thumb is that lab mice weigh 20–40 g [Suckow et al., 2001] and lab rats 150–500 g, though rats up to 1000 g are used in obesity studies [Krinke, 2000]. For comparison, 80,000 g is c. 176 lb.

First some notation: vectors will be written in lowercase boldface ($\boldsymbol{g}$)w, matrices in uppercase boldface ($\boldsymbol{H}$), and integral operators in uppercase calligraphy ($\mathcal{L}$). We represent the object being imaged, which might be the activity distribution in an animal or a human, by the non-negative $\mathbb{L}_2$ function $f(\boldsymbol{r}, t)$, the data list as a set of vectors $\{\boldsymbol{g}_i\}$, $i \in \{1, \cdots, N_{\text{list}}\}$, and the reconstructed object as an $\mathbb{L}_2$ function $\hat{f}(\boldsymbol{r}, t)$. (A useful mnemonic is that we are **g**iven g and want to **f**ind f.) A hat over the variable will always indicate estimate or reconstruction. The specific $\mathbb{L}_2$ function space will depend on whether we are interested in 2-dimensional projection or 3-dimensional volume imaging. Below we consider each case individually. Probabilities will always be indicated as a density $p()$, without the use of subscripts. In discrete cases this implies use of a Dirac delta function. Expected values will be indicated using angle brackets $\langle\,\rangle$. For mathematical convenience, and because it is may be dictated by the physics, we sometimes assume $f(\boldsymbol{r}, t) \in \mathbb{L}_2 \cap \mathbb{L}_1$, or that $f(\boldsymbol{r}, t)$ is a generalized function. (Some texts refer to generalized functions as Schwartz distributions, not to be confused with probability distributions.)

We relate the object $f(\boldsymbol{r}, t)$ to the data list $\{\boldsymbol{g}_i\}$ by using the likelihood model $p\left(\{\boldsymbol{g}_i\} \,|f(\boldsymbol{r}, t)\right)$. This is the conditional probability of recording list $\{\boldsymbol{g}_i\}$ given that $f(\boldsymbol{r}, t)$ is the object being imaged. Each $\boldsymbol{g}_i$ is a list entry recording the i^{th} scintillation event. For full generality we include time in $f(\boldsymbol{r}, t)$, although in many cases we will ignore time and assume a static object $f(\boldsymbol{r})$, a good assumption when imaging is quick compared to the radiotracer half-life and any relevant biological processes.

The physics of radioactive decay dictate that emitted photons are statistically independent, and the acquisition electronics are designed to reject the rare cases where the PMTs are activated by more than one gamma gray. As a result list entries are independent and identically distributed (i.i.d.). Therefore the list's probability can be formulated in terms of the individual events's probability,

$$p\left(\{\boldsymbol{g}_i\}, N_{\text{list}} | f(\boldsymbol{r}, t)\right) = p\left(N_{\text{list}} | f(\boldsymbol{r}, t)\right) \prod_{i=1}^{N_{\text{list}}} p\left(\boldsymbol{g}_i | f(\boldsymbol{r}, t)\right), \tag{1.1}$$

where $p(N_{\text{list}}|f(\boldsymbol{r},t))$ is the probability of recording a list with N_{list} entries. In SPECT the physics dictates that N_{list} follows a Poisson distribution, see (1.14). We shall discuss various ways to model $p\left(\boldsymbol{g}_i|f(\boldsymbol{r},t)\right)$ in chapter 2. For now note that $p\left(\boldsymbol{g}_i|f(\boldsymbol{r},t)\right)$ is a conditional probability, and so accounts for randomness due both to radioactive decay in the object and to noise in the measurement system. A common assumption is that $p\left(\boldsymbol{g}_i|f(\boldsymbol{r},t)\right)$ obeys amplified Poisson statistics, in other words that the digitized PMT output can be modeled as a Poisson random variable (RV) amplified by a gain process [Barrett and Myers, 2004, section 12.3.5].

A data list entry $\boldsymbol{g}_i$ from FastSPECT-II has at least ten dimensions, one for each PMT and one for the camera ID. The units of digitized PMT output are arbitrary, though they can be related to voltages. List entries $\boldsymbol{g}_i$ will have more dimensions if information such as EKG or a time stamp is included. Note that some components of $\boldsymbol{g}_i$, in particular the digitized output of the PMTs, may be negative.

At first blush, the problem of reconstructing $f(\boldsymbol{r},t)$ seems to be hideously underconstrained, since $\mathbb{L}_2$ is an infinite-dimensional Hilbert space, while each list entry $\boldsymbol{g}_i$ is finite dimensional and we only have a finite number N_{list} of them. In practice we regularize the problem by restricting ourselves to a finite-dimensional subspace of $\mathbb{L}_2$ determined by an orthogonal (typically orthonormal) set of basis functions $\Phi_j(\boldsymbol{r},t)$. The object $f\left(\boldsymbol{r},t\right)$ can be represented as

$$f\left(\boldsymbol{r},t\right) = \sum_{j=1}^{N_{\text{basis}}} \alpha_j \Phi_j(\boldsymbol{r},t) + \epsilon_{\text{trunc}}, \tag{1.2}$$

where

$$\alpha_j = \int_{\infty} \Phi_j(\boldsymbol{r},t) f(\boldsymbol{r},t) d\boldsymbol{r} dt. \tag{1.3}$$

We reconstruct the object by estimating the vector of coefficients. The modeling error ϵ_{trunc} introduced by using a finite number of basis elements can be made arbitrarily small by increasing the number of basis elements within the field of view. Put more

precisely, $\epsilon_{\text{trunc}} \in \mathbb{L}_2$ (or $\mathbb{L}_1 \cap \mathbb{L}_2$) and

$$\lim_{N_{\text{basis}} \to \infty} \|\epsilon_{\text{trunc}}\|_2 = 0. \tag{1.4}$$

It matters how this limit is taken. Merely expanding the field of view by adding new elements at the edge is not sufficient, rather one must add new elements which have support within the already established field of view.

Our estimate of the object, also known as the reconstruction, is

$$\hat{f}(\boldsymbol{r}, t) = \sum_{j=1}^{N_{\text{basis}}} \hat{\alpha}_j \Phi_j(\boldsymbol{r}, t). \tag{1.5}$$

We have deliberately chosen to use the same basis elements $\Phi(\boldsymbol{r}, t)$ to represent both the object and our reconstruction. There is no need to do so; some reconstruction algorithms choose different Hilbert (or sometimes Banach) spaces for $f(\boldsymbol{r}, t)$ and $\hat{f}(\boldsymbol{r}, t)$, which may make it impossible to use the same representation for both. However, using differing representations offers no advantage for the current discussion. The Rayleigh-task simulations in chapter 3 will use different basis elements for the object and reconstruction.

In choosing $\phi_j(\boldsymbol{r}, t)$ the reconstruction-algorithm designer has considerable latitude; some of the choices and tradeoffs are discussed by Lehovich et al. [2001]. We'll see several examples of basis functions below. Once we've chosen a representation, $\hat{\boldsymbol{\alpha}}$ contains all the information about the object available to the observer.

In static imaging problems, (1.2) simplifies to

$$f(\boldsymbol{r}) = \sum_{j=1}^{N_{\text{basis}}} \alpha_j \phi_j(\boldsymbol{r}) + \epsilon_{\text{trunc}}. \tag{1.6}$$

(We will continue to make the distinction between uppercase Φ, which is a function of location and time, and lowercase ϕ which is only a function of location.) In dynamic imaging it is common to use a separable representation

$$f(\boldsymbol{r}, t) = \sum_{j=1}^{N_{\text{basis}}} \alpha_j \phi_j(\boldsymbol{r}) \Psi_j(t) + \epsilon_{\text{trunc}}, \tag{1.7}$$

where $\Psi_j(t)$ is the time-activity curve for the spatial region defined by $\phi_j(\boldsymbol{r})$ representing the average radiotracer concentration within region $\phi_j(\boldsymbol{r})$ at time t. Note that $\Psi_j(t)$ can also be expanded as a sum,

$$\Psi_j(t) = \sum_k \beta_k \psi_k(t), \tag{1.8}$$

and therefore

$$f(\boldsymbol{r}, t) = \sum_j \sum_k \alpha_j \beta_k \phi_j(\boldsymbol{r}) \psi_k(t) + \epsilon_{\text{trunc}}. \tag{1.9}$$

A separable representation such as (1.9) is always possible because $f(\boldsymbol{r}, t) \in \mathbb{L}_2$, which has a countable basis [Royden, 1968]. The argument that $\alpha_j \beta_k$ form a countable set is essentially the same one used to show the rational numbers are a countable set, in other words they can be put in one-to-one correspondence with the integers. We need to show that the set of separable basis functions $\phi_i(\boldsymbol{r})\psi_j(t)$ is countable. The following diagram shows how to do so:

	ϕ_1	ϕ_2	ϕ_3	$\cdots$
ψ_1	Φ_1	Φ_2	Φ_4	$\cdots$
ψ_2	Φ_3	Φ_5	$\cdots$	
ψ_3	Φ_6	$\vdots$	$\ddots$	
$\vdots$	$\vdots$			

(1.10)

where $\Phi_1(\boldsymbol{r}, t) = \phi_1(\boldsymbol{r})\psi_1(t)$, $\Phi_2(\boldsymbol{r}, t) = \phi_2(\boldsymbol{r})\psi_1(t)$, $\Phi_3(\boldsymbol{r}, t) = \phi_1(\boldsymbol{r})\psi_2(t)$, and so on.

We are interested in using list-mode data to reconstruct both 2-dimensional projection images and 3-dimensional tomographic images. As the physical parameters being estimated are different, these two situations call for somewhat different Hilbert spaces and basis elements. Both cases are now considered.

1.6.1 2-dimensional modeling for projection images

If we want to produce projection images, then $f(\boldsymbol{r}, t)$ represents the gamma-ray flux on the camera face, $\boldsymbol{r} \in \mathbb{R}^2$, and $f(\boldsymbol{r}, t)$ has units of $\frac{\text{photons}}{\text{cm}^2\text{s}}$. The physics dictates that

$f(\boldsymbol{r}, t)$ is a non-negative function with compact spatial support and that spatially $f(\boldsymbol{r}, t)$ is in $\mathbb{L}_1 \cap \mathbb{L}_2$. Temporally $f(\boldsymbol{r}, t)$ is $\mathbb{L}_1 \cap \mathbb{L}_2$ because we image for a finite amount of time, and therefore will detect only a finite number of photons. (If we imaged for an infinite amount of time this might not be the case.) Because $f()$ represents a mean it need not be an integer-valued function.

We are often interested in the location of the emission event. In 2-dimensional imaging this is the location within the scintillation crystal where the incoming gamma ray undergoes a Compton or photoelectric interaction and emits a shower of visible-light photons.

The most common choice of a two-dimensional truncated basis uses square pixels for $\phi_j(\boldsymbol{r})$, as shown in figure 1.7. Rectangular pixels are also occasionally used. Many detectors are pixelated, making this a natural way to model them. With the notable exception of film, still widely used for image display even in digital imaging systems, most display devices are based on pixelated output. The coefficients α_j represent the average number of gamma rays hitting pixel j per second. A two-dimensional array is the natural data structure for a Cartesian grid of pixels.

In static imaging we are often interested in gamma-ray fluence, which has units photons/cm^2. In this case α_j is the average number of gamma rays hitting pixel j for the ensemble of repeated experiments.

1.6.2 3-dimensional modeling for tomographic images and volumes

If we want to reconstruct a 3-dimensional volume, or equivalently a set of tomographic images, then $\boldsymbol{r} \in \mathbb{R}^3$ and $f(\boldsymbol{r}, t)$ represents radiotracer concentration with units of mCi/cm^3. Because $f(\boldsymbol{r}, t)$ represents a concentration it must be non-negative. We can only fit finite-sized objects into our imaging system, so $f(\boldsymbol{r}, t)$ has compact support in space. A finite amount of radiotracer is injected and radioactive decay is exponential, so $f(\boldsymbol{r}, t)$ must be in $\mathbb{L}_1 \cap \mathbb{L}_2$ both spatially and temporally.

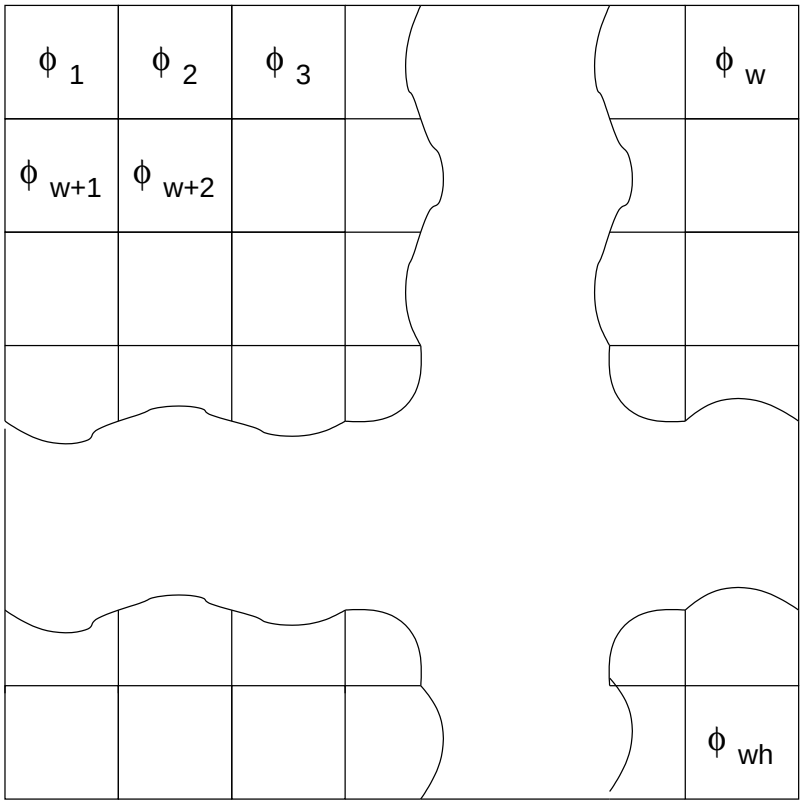

FIGURE 1.7. **Grid of square pixels used as a two-dimensional truncated basis.**

In 2-dimensional imaging the scintillation event location is within the camera crystal. However in 3-dimensional imaging the location of the emission event is inside the object. In SPECT we are interested in the location from which the gamma ray originated due to radioactive decay.

Voxels[3] arranged in a Cartesian grid are by far the most popular choice of 3-dimensional truncated-basis functions $\phi_j(\boldsymbol{r})$. A voxel is the characteristic function of a rectangular parallelepiped, usually a cube; it has value 1 inside the parallelepiped and 0 outside. For a cubic-voxel basis α_j can be thought of as the average number of gamma rays emitted per second from the cubic region of the subject specified by $\phi_j(\boldsymbol{r})$. (Strictly speaking one needs to include a conversion factor, because a millicurie (mCi) is 37,000,000 disintegrations per second and not all disintegrations produce a gamma ray.) As the natural extension of Cartesian pixel grids, voxels are also usually implemented as multi-dimensional arrays in reconstruction algorithms. Most 3-dimensional visualization software accepts voxelized input. It is trivial to extract tomographic slices along the cardinal axes from a voxelized data set.

Although voxels are more common, anatomical basis elements are also used in nuclear medicine. For example to estimate the time activity curve within the liver we might choose $\phi_1(\boldsymbol{r})$ as the organ's characteristic function, and $\phi_2(\boldsymbol{r})$ as the rest of the body. A truncated basis with only two elements is on the low end, but still useful in some problems.

1.7 Maximum likelihood (ML) image reconstruction

Once we have chosen basis functions $\phi_j(\boldsymbol{r})$ to represent the object and a likelihood model $p(\{\boldsymbol{g}_i\}|\boldsymbol{\alpha})$ relating the object to the data, we need an algorithm to turn the list of data $\{\boldsymbol{g}_i\}$ into a reconstructed vector of coefficients $\hat{\boldsymbol{\alpha}}$. An implicit assumption

[3]Voxel is a contraction of volume element not yet listed in any of the dictionaries I checked. I suspect the "x" carried over from pixel, an already lexicographed contraction of picture element [OED 1989].

in our modeling approach is that

$$p\left(\boldsymbol{g}|f(\boldsymbol{r})\right) \approx p\left(\boldsymbol{g}|\boldsymbol{\alpha}\right). \tag{1.11}$$

We will discuss such an algorithm in section 1.7.2, but first we look at what needs to be computed.

Our focus will be a static object $f(\boldsymbol{r})$. Much of the discussion in this section carries over to the dynamic case, but note that the EM algorithm presented below is only appropriate for static imaging. An algorithm intended for dynamic imaging would need to model temporal effects such as the exponential decay of the radiotracer. (Of course one could partition the dynamic data into different time slices and apply the static algorithm at each time, ignoring the other time slices, but then one doesn't have a dynamic algorithm.)

We begin by introducing several related conditional probabilities which, taken together, model the imaging system. For the reader's convenience a summary is provided in table 1.1. Recall from sections 1.6.1 and 1.6.2 that α_j represents a spatial average of $f(\boldsymbol{r})$ over the region $\phi_j(\boldsymbol{r})$; in the 2-dimensional case α_j is the average number of photons interacting with and passing through region $\phi_j(\boldsymbol{r})$ on the detector face, and in the 3-dimensional case α_j is the average number of photons emitted from region $\phi_j(\boldsymbol{r})$.

Not all gamma rays associated with $\phi_j(\boldsymbol{r})$ will be detected. The proportion of emitted photons that produce list entries is the sensitivity, s_j. One can also think of s_j as the probability that a photon associated with $\phi_j(\boldsymbol{r})$ is detected. In SPECT imaging the sensitivity is quite low. For FastSPECT-II, a system with exceptionally high sensitivity due to its many cameras, s_j on the order of 10^{-4}. FastSPECT-II loses most photons due to the system collimator; in clinical systems the small number of cameras means most photons are missed. On the other hand, the sensitivity of a modular gamma camera is in the range of 0.5–0.7. Of course the camera has a chance to detect only photons that make it through the collimator.

$\overline{N}_{\text{list}}$	Average number of list entries in a fixed time experiment.
α_j	Average number of gamma rays associated with $\phi_j(\boldsymbol{r})$. In the static case (1.3) reduces to $\alpha_j = \int_{\phi_j(\boldsymbol{r})} f(\boldsymbol{r})d\boldsymbol{r}$.
s_j	sensitivity to $\phi_j(\boldsymbol{r})$, i.e. the probability that a photon associated with $\phi_j(\boldsymbol{r})$ is detected.
$p\left(N_{\text{list}}\middle\vert\boldsymbol{\alpha}\right)$	Probability of recording a list with N_{list} entries from the object represented by $\boldsymbol{\alpha}$.
$p(\phi_j(\boldsymbol{r})\vert\boldsymbol{\alpha})$	Probability that a detected photon from the object represented by $\boldsymbol{\alpha}$ is associated with $\phi_j(\boldsymbol{r})$.
$p(\boldsymbol{g}_i\vert\boldsymbol{\alpha})$	probability that a detected photon from the object represented by $\boldsymbol{\alpha}$ produces data list entry $\boldsymbol{g}_i$.
$p\big(\boldsymbol{g}_i\big\vert\text{originated in }\phi_j(\boldsymbol{r})\big)$	probability that a detected event associated with the region $\phi_j(\boldsymbol{r})$ produces data list entry $\boldsymbol{g}_i$. In the 2-dimensional case the scintillation event occurs in $\phi_j(\boldsymbol{r})$; in the 3D case the gamma ray originates in $\phi_j(\boldsymbol{r})$.

TABLE 1.1. Conditional probabilities used in image reconstruction.

Note that

$$\overline{N}_{\text{list}} = \sum_{j=1}^{N_{\text{basis}}} \alpha_j s_j. \tag{1.12}$$

Because we know the total amount of activity injected, the time since injection, and the half-life of the radiotracer, we can easily calculate $\|f(\boldsymbol{r},t)\|_1$, and thus put a constraint on $\boldsymbol{\alpha}$ because, assuming no truncation error,

$$\int f(\boldsymbol{r})d\boldsymbol{r} = \sum_{j=1}^{N_{\text{basis}}} \alpha_j. \tag{1.13}$$

The physics of radioactive decay dictates that in SPECT the total number of detected photons follows a Poisson distribution

$$p(N_{\text{list}} = n|\overline{N}_{\text{list}}) = \frac{\overline{N}_{\text{list}}^n e^{-\overline{N}_{\text{list}}}}{n!}. \tag{1.14}$$

By definition, only detected photons produce data list entries. The probability that a detected photon associated with the object $\boldsymbol{\alpha}$ produces list entry $\boldsymbol{g}$ is

written $p(\boldsymbol{g}_i|\boldsymbol{\alpha})$. Similarly, the probability that a detected gamma ray associated with $\phi_j(\boldsymbol{r})$ produces list entry $\boldsymbol{g}$ is written $p\big(\boldsymbol{g}_i\big|\text{originated in}\,\phi_j(\boldsymbol{r})\big)$. We assume that $p\big(\boldsymbol{g}_i\big|\text{originated in}\,\phi_j(\boldsymbol{r})\big)$ and s_j are independent of $f(\boldsymbol{r})$, and $\boldsymbol{\alpha}$, which corresponds to an assumption of no scatter and no attenuation. This is a reasonable approximation for small animal imaging, but can lead to artifacts in human imaging.

If we choose basis elements that partition the object's support, then each gamma ray will be associated with one, and only one, region $\phi_j(\boldsymbol{r})$. Henceforth we shall assume this is the case. The probability that a detected photon from object $\boldsymbol{\alpha}$ is associated with region $\phi_j(\boldsymbol{r})$ is

$$p(\phi_j(\boldsymbol{r})|\boldsymbol{\alpha}) = \frac{\alpha_j s_j}{\sum_{n=1}^{N_{\text{basis}}} \alpha_n s_n} \tag{1.15}$$

$$= \frac{\alpha_j s_j}{\overline{N}_{\text{list}}}. \tag{1.16}$$

Since a photon is associated with exactly one region we have

$$p\left(\boldsymbol{g}_i|\boldsymbol{\alpha}\right) = \sum_j p\big(\boldsymbol{g}_i\big|\text{originated in}\,\phi_j(\boldsymbol{r})\big)\, p\left(\phi_j(\boldsymbol{r})|\boldsymbol{\alpha}\right). \tag{1.17}$$

Given an object represented by $\boldsymbol{\alpha}$, the model $p\left(\{\boldsymbol{g}_i\}\,|\boldsymbol{\alpha}\right)$ provides a probability density on $\{\boldsymbol{g}_i\}$. This is a forward model; it lets us predict what sort of data a known object may produce. But in image reconstruction we have the opposite situation: the data list $\{\boldsymbol{g}_i\}$ is fixed and we don't know the coefficients $\boldsymbol{\alpha}$. This is the inverse problem. The function

$$L\left(\hat{\boldsymbol{\alpha}}|\{\boldsymbol{g}_i\}\right) = p\left(\{\boldsymbol{g}_i\}|\hat{\boldsymbol{\alpha}}\right) \tag{1.18}$$

is called the likelihood function. Likelihood has a long history in statistics. List-mode likelihood was introduced by Snyder and Politte [1983], with further discussion provided by Barrett et al. [1997].

Note that $L\left(\hat{\boldsymbol{\alpha}}|\{\boldsymbol{g}_i\}\right)$ is not a probability density on reconstructions $\hat{\boldsymbol{\alpha}}$, hence the switch in notation. This is because we cannot rely on the likelihood to be normalized, in other words usually

$$\int_\infty \mathrm{d}\hat{\boldsymbol{\alpha}}\; L\left(\hat{\boldsymbol{\alpha}}|\{\boldsymbol{g}_i\}\right) \neq 1. \tag{1.19}$$

We have already noted that list entries are i.i.d. Therefore using (1.1) we can rewrite (1.18),

$$L\left(\hat{\boldsymbol{\alpha}}|\{\boldsymbol{g}_i\}\right) = p\left(N_{\text{list}}|\hat{\boldsymbol{\alpha}}\right) \prod_{i=1}^{N} p\left(\boldsymbol{g}_i|\hat{\boldsymbol{\alpha}}\right). \tag{1.20}$$

To avoid round-off error due to the small magnitudes of $p\left(\boldsymbol{g}_i|\hat{\boldsymbol{\alpha}}\right)$, we usually work with the log-likelihood

$$l\left(\hat{\boldsymbol{\alpha}}|\{\boldsymbol{g}_i\}\right) = \log\left(L(\hat{\boldsymbol{\alpha}}|\{\boldsymbol{g}_i\})\right) \tag{1.21}$$

$$= \log\left(p(N_{\text{list}}|\hat{\boldsymbol{\alpha}})\right) + \sum_{i=1}^{N} \log\left(p(\boldsymbol{g}_i|\hat{\boldsymbol{\alpha}})\right), \tag{1.22}$$

with the important side benefit that logarithms turn products into sums.

An obvious thing to compute is the vector of coefficients $\hat{\boldsymbol{\alpha}}_{\text{ML}}$ maximizing the likelihood $L\left(\hat{\boldsymbol{\alpha}}|\{\boldsymbol{g}_i\}\right)$:

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_{\text{ML}} &= \arg\max_{\boldsymbol{\alpha}} L\left(\boldsymbol{\alpha}|\{\boldsymbol{g}_i\}\right) \\ &= \arg\max_{\boldsymbol{\alpha}} l\left(\boldsymbol{\alpha}|\{\boldsymbol{g}_i\}\right). \end{aligned} \tag{1.23}$$

In statistics this is called the maximum-likelihood (ML) estimate, or in imaging the ML reconstruction.

1.7.1 Properties of ML estimation

ML estimators have many desirable theoretical and practical properties, some of which we list here. Further details and relevant proofs are given by Casella and Berger [1990] and Monahan [2001].

It usually works. Because of its excellent track record across many disciplines, especially for problems where a small number of parameters are estimated, ML is the most widely used estimation technique in applied statistics. As we shall see in section 1.7.5, image reconstruction is an exception.

Easy to calculate. Although in imaging the ML estimate cannot be explicitly calculated with a closed-form analytic solution, a variety of iterative algorithms converge to it. These include specialized algorithms such as the expectation-maximization (EM) algorithm described in the next section and general-purpose optimization algorithms such as steepest-gradient descent, classical Newton-Raphson, and quasi-Newton. Algorithms for both projection-mode and list-mode reconstruction are well known and tend to converge quickly. The intermediate result after only a few iterations, perhaps only one, is often usable.

Often the natural approach is ML. For many problems the "naive" approach taken by a smart person not trained in statistical theory turns out to give the ML estimate. For example, the obvious way to estimate the population mean of a random variable is by computing the sample mean, which turns out to be the ML estimate for Gaussian and Poisson random variables. (A proof for the Poisson case is given in section 2.2.1.) In the case of Gaussian noise the least-squares solution to an inconsistent system of linear equations can be shown to be the ML estimate. (Note that counter examples do exist. For example, the histogram algorithm used in chapter 3 to reconstruct list-mode PMT data into a projection image is a natural approach which does not give the ML solution.)

Invariance property. If $\hat{\boldsymbol{\alpha}}$ is the ML estimate of $\boldsymbol{\alpha}$ and $\tau()$ is a one-to-one function, then $\tau(\hat{\boldsymbol{\alpha}})$ is the ML estimate of $\tau(\boldsymbol{\alpha})$. A similar result holds even in the case where $\tau()$ is not one-to-one.

Asymptotically unbiased. In the limit of infinitely many samples (list entries), the mean of the ML estimate becomes almost-surely arbitrarily close to the true value.

Asymptotic efficiency In the limit of infinitely many list entries, the variances on the components of the ML estimate $\hat{\boldsymbol{\alpha}}$ are the smallest possible, i.e. they

approach the Cramér-Rao lower bound.

Asymptotic quadraticity. In the limit of an infinite number of list entries, the log-likelihood function is quadratic in a neighborhood around its maximum. This is nice because quadratic functions are by far the easiest to maximize computationally.

Asymptotic normality. In the limit of infinitely many list entries, the maximum-likelihood estimate has a normal distribution about the true value of the parameter.

One drawback of the ML approach is that it need not be stable: a small change in the data may lead to a large change in the estimate. This occurs when the likelihood has steep cliffs or discontinuities. In cases where the likelihood model is nearly flat it may be difficult to find the maximum. Fortunately, these situations appear not to occur in gamma-ray imaging; I don't know of any examples. We will see plots of $p\left(\boldsymbol{g}_i|\phi_j(\boldsymbol{r})\right)$ and $l\left(\hat{\boldsymbol{\alpha}}|\boldsymbol{g}_i\right)$ in chapter 2.

1.7.2 EM algorithm

The Expectation-Maximization (EM) algorithm is an iterative way of calculating an ML estimate. The algorithm has been around in one guise or another for over a century; for a history of the algorithm see the textbook by McLachlan and Krishnan [1997]. It was first named EM by Dempster et al. [1977], a paper which remains one of the canonical references. EM was first introduced into medical imaging by Metz and Pizer in an unpublished 1971 conference presentation. The first publications using EM for emission-tomography reconstruction were by Shepp and Vardi [1982] and independently by Lange and Carson [1984]. A readable derivation of the general algorithm as applied to any maximum-likelihood problem is found in Lange [1999].

The list-mode version of EM for tomographic reconstruction was first used by Snyder and Politte [1983], with a different derivation due to Parra and Barrett [1998].

The algorithm's name is due to the nature of its update iterations. First one embeds the problem space, called the "incomplete data", in a larger space, the "complete data", where the problem is easier to solve. In our case the data list is the incomplete data, and the list augmented with the true (but unknown) location of each emission event is the complete data [Parra and Barrett, 1998]. Since one doesn't know the location, one marginalizes it away by computing a conditional expected value. (This is the expectation step). Then one solves for the maximum-likelihood estimate in the complete space. (The maximization step). Alternating the two steps can be shown to converge a maximum-likelihood solution.

EM is numerically stable because the likelihood is monotonically increasing at each step of the algorithm. Uniqueness properties of the EM algorithm will be discussed in section 1.7.6.2.

1.7.3 Heuristic derivation of the list-mode EM update equation

We can intuitively derive the list-mode EM algorithm for a static object by starting with the obvious thing to calculate and proceeding in a straightforward fashion, hoping we wind up somewhere useful. As we have already noted, such ad-hoc derivations lead to the ML estimate a surprising amount of the time. An obvious reconstruction strategy is to "back project" each detected photon, in other words assign it to each region $\phi_j(\boldsymbol{r})$ in proportion to the probability of association, and then sum across list entries:

$$\text{uncorrected } \hat{\alpha}_j = \sum_{i=1}^{N_{\text{list}}} p\big(\text{originated in } \phi_j(\boldsymbol{r}) \big| \boldsymbol{g}_i\big). \tag{1.24}$$

We must also correct for the system sensitivity and normalize by the time of imaging:

$$\hat{\alpha}_j = \frac{1}{T} \sum_{i=1}^{N_{\text{list}}} \frac{p\big(\text{originated in } \phi_j(\boldsymbol{r}) \big| \boldsymbol{g}_i\big)}{s_j}. \tag{1.25}$$

Snyder and Politte [1983] refer to this as posterior-density weighting.

If one thinks only about the 2-dimensional case, it is tempting to insert a nonlinear position estimation step into (1.25), such that each event only contributes to the $\phi_j(\boldsymbol{r})$ it is most likely to have originated in. However in the 3-dimensional case this is less tempting. Each detected event could have originated inside a cone specified by the pinhole aperture. Any location we pick within this cone is likely to be quite far away from the actual position of the radioactive decay.

Using Bayes rule we know that

$$p\big(\text{originated in } \phi_j(\boldsymbol{r})\big|\boldsymbol{g}_i\big) = \frac{p\big(\boldsymbol{g}_i\big|\text{originated in } \phi_j(\boldsymbol{r})\big)p\left(\phi_j(\boldsymbol{r})\right)}{\text{normalization}}. \tag{1.26}$$

Notice that $p\left(\phi_j(\boldsymbol{r})\right)$ implicitly depends on $\boldsymbol{\alpha}$. So by substituting (1.16) we get

$$p\big(\text{originated in } \phi_j(\boldsymbol{r})\big|\boldsymbol{g}_i\big) = \frac{p\big(\boldsymbol{g}_i\big|\text{originated in } \phi_j(\boldsymbol{r})\big)\frac{\alpha_j s_j}{\sum_{n=1}^{N_{\text{basis}}}\alpha_n s_n}}{\text{normalization}} \tag{1.27}$$

Because we assume that each event uniquely originated in one region $\phi_j(\boldsymbol{r})$, in other words we ignore the possibility of a background event or multiple photons arriving simultaneously, we know that

$$1 = \sum_{j=1}^{N_{\text{basis}}} p\big(\text{originated in } \phi_j(\boldsymbol{r})\big|\boldsymbol{g}_i\big) \tag{1.28}$$

and therefore that

$$\text{normalization} = \sum_{j=1}^{N_{\text{basis}}} p\big(\boldsymbol{g}_i\big|\text{originated in } \phi_j(\boldsymbol{r})\big)\frac{\alpha_j s_j}{\sum_{n=1}^{N_{\text{basis}}}\alpha_n s_n} \tag{1.29}$$

(The FastSPECT-II acquisition electronics are designed to drop events where more than one photon was detected.)

Thus we have

$$\hat{\alpha}_j = \frac{1}{T}\sum_{i=1}^{N_{\text{list}}} \frac{p\big(\boldsymbol{g}_i\big|\text{originated in } \phi_j(\boldsymbol{r})\big)\alpha_j}{\sum_{k=1}^{N_{\text{basis}}} p\big(\boldsymbol{g}_i\big|\text{originated in } \phi_k(\boldsymbol{r})\big)\alpha_k s_k.} \tag{1.30}$$

This is a nonlinear equation with no known closed-form solution, even if we make generous assumptions about the nature of $p\big(\boldsymbol{g}_i\big|\text{originated in}\,\phi_j(\boldsymbol{r})\big)$ and are willing to substitute $\hat{\boldsymbol{\alpha}} \approx \boldsymbol{\alpha}$.

However (1.30) suggests a fixed-point iteration. An initial guess of a uniform object,

$$\forall j: \qquad \hat{\alpha}_j^{(0)} = 1, \tag{1.31}$$

is reasonable. We then iterate

$$\boxed{\hat{\alpha}_j^{(t+1)} = \frac{1}{T} \sum_{i=1}^{N_{\text{list}}} \frac{p\big(\boldsymbol{g}_i\big|\text{originated in}\,\phi_j(\boldsymbol{r})\big)\hat{\alpha}_j^{(t)}}{\sum_{k=1}^{N_{\text{basis}}} p\big(\boldsymbol{g}_i\big|\text{originated in}\,\phi_k(\boldsymbol{r})\big)\hat{\alpha}_k^{(t)} s_k}}. \tag{1.32}$$

The update equation enforces the invariance property

$$\sum_{j=1}^{N_{\text{basis}}} \alpha_j^{(t)} s_j = \frac{N_{\text{list}}}{T} \qquad \forall t{>}0. \tag{1.33}$$

When writing software implementing the list-mode EM algorithm, a good debugging check is to verify (1.33) holds.

The preceding argument is more than a bit hand-wavy. And yet, amazingly, as proved by Parra and Barrett [1998] iterating (1.32) does lead to an ML estimate when the number of detection events associated with $\phi_j(\boldsymbol{r})$ obeys Poisson statistics, as is the case in SPECT.

It is easy to see (1.32) does not define a contraction mapping. If it were a contraction map, the Banach fixed-point theorem would guarantee (1.33) converged to a unique fixed point regardless of starting point. If we choose as starting point $\boldsymbol{\alpha}^{(0)}$ such that $\alpha_j = 0$ for some j, then the algorithm is restricted to the hyperplane $\alpha_j = 0$ for all iterations. The only point in the intersection of all such hyperplanes is $\boldsymbol{\alpha} = 0$, so that would have to be the putative contraction point. However in practice the algorithm does not converge to $\boldsymbol{\alpha} = \mathbf{0}$; if it did we wouldn't use it. (The EM algorithm does converge, even though it is not a contraction map.)

1.7.4 List-mode EM in the special case of no detector blur

Chapter 3 compares two ways of producing projection images from list-mode data. One method is the list-mode EM algorithm, the other is ML position estimation on each event followed by histogramming. This section shows that for the special case of a gamma camera with no detector blur the two techniques produce the same answer.

Let us temporarily assume for all possible list entries $\boldsymbol{g}_i$ that the likelihood function $p\big(\boldsymbol{g}_i\big|\text{originated in } \phi_j(\boldsymbol{r})\big) = 1$ for exactly one j, and is 0 otherwise. Thus each list entry can be precisely assigned to exactly one region $\phi_j(\boldsymbol{r})$. One example of how this might occur is a gamma camera with no detector blur.

For this special case the list-mode update equation (1.32) converges after only one iteration to the fixed point

$$\hat{a}_j = \frac{N_j}{Ts_j}, \tag{1.34}$$

where N_j is the number of list entries associated with $\phi_j(\boldsymbol{r})$. Therefore, for the special case of no detector blur, the ML solution is to assign each event to exactly one region and count how many events are assigned to each region, in other words to histogram the event locations.

To see that this must occur, notice that for this special case the likelihood function $p\big(\boldsymbol{g}_i\big|\text{originated in } \phi_j(\boldsymbol{r})\big)$ acts much like the Kronecker δ function inside the sum. Because $p\big(\boldsymbol{g}_i\big|\text{originated in } \phi_j(\boldsymbol{r})\big)$ is nonzero at only one value of k, the denominator of (1.32) satisfies

$$\sum_{k=1}^{N_{\text{basis}}} p\big(\boldsymbol{g}_i\big|\text{originated in } \phi_k(\boldsymbol{r})\big)\hat{\alpha}_k^{(t)} s_k = \hat{\alpha}_{a_i}^{(t)} s_{a_i}, \tag{1.35}$$

where a_i identifies the region $\phi_{a_i}(\boldsymbol{r})$ associated with $\boldsymbol{g}_i$. The same thing happens to the numerator of (1.32) which becomes $\hat{\alpha}_{a_i}^{(t)}$. Summing over all list entries gives (1.34).

Outside the special case of no detector blur, estimating the locations of each event and then histograming does not lead to an ML estimate of $\boldsymbol{\alpha}$, not even when the position estimation is done using maximum likelihood. However it seems to be a

reasonable assumption that in cases of little detector blur the histogram of position estimates will be close to the ML solution. Chapter 3 compares the two approaches using data from a modular gamma camera.

1.7.5 Night skies and the myth of convergence

In practice we never run the EM algorithm to convergence for three reasons. First, the images are useful after only a few iterations, so there is no need to expend extra CPU time.

Second, although the ML estimate has many nice properties in the asymptotic limit of an infinitely long data list, in nuclear medicine we rarely have enough data for that to be relevant. In actual use, when EM is run to convergence the reconstructed images consist of a small number of point sources surrounded by a black background. This phenomenon is called the night-sky effect, by analogy with stars against a black-background. An example is shown in figure 1.8. Night-sky images are not considered to be especially useful for clinical tasks. Byrne [1993] proved that if we are reconstructing a discretized object from noisy data the projection-mode EM algorithm will converge to a unique night-sky reconstruction regardless of starting point. Eric Clarkson has shown that the same phenomenon occurs when reconstructing continuous objects; his results are still unpublished, for a summary see Barrett and Myers [2004, sec. 15.3.5]. However it is not clear how to extend Clarkson's "moment cone" argument to list-mode data.

Third, when image quality for human observers is considered as a function of iteration number, it reaches a peak after a small number of iterations, and then decreases due to the night sky effect [Wilson, 1994; Abbey, 1998]. A precise definition of image quality will be given in section 1.9.

Nevertheless, output from a few iterations of the EM algorithm is often described as an ML estimate. Even outside of the ML framework, the specific version of the

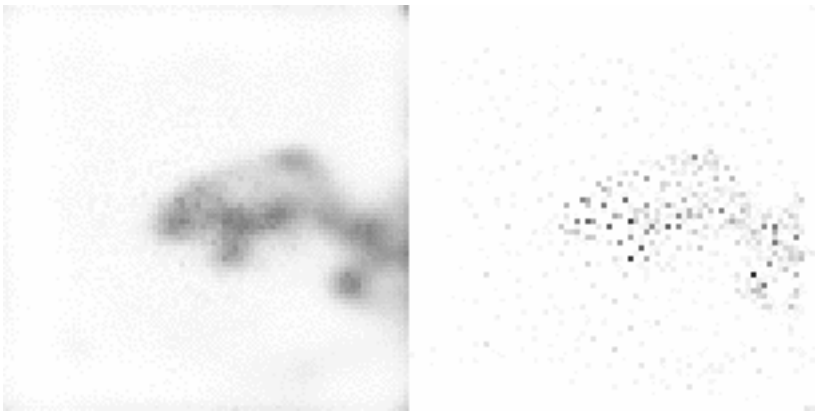

FIGURE 1.8. **Example of a night-sky reconstruction.** Data is from the same mouse bonescan shown in figures 1.2 and 1.3. On the left is a projection image after two iterations of the 2-dimensional list-mode EM algorithm. On the right is 100 iterations.

EM algorithm for Poisson data has a useful property for imaging: the multiplicative updates of (1.32) ensure that $\hat{f}(\boldsymbol{r})$ is non-negative if the starting point is non-negative. Since we know that physically $f(\boldsymbol{r}, t)$ must be non-negative, this is valuable prior knowledge to include in the reconstruction. A history of the early uses of positivity in image reconstruction is given by Frieden [1981].

1.7.6 Uniqueness of maximum-likelihood solutions

For a given likelihood model $p\left(\boldsymbol{g}_i | f(\boldsymbol{r})\right)$, it is natural to wonder about the uniqueness of maximum-likelihood solutions. Given a list of data $\{\boldsymbol{g}_i\}$, can the EM algorithm converge to more than one fixed point? In the asymptotic case of infinite data, $\lim N_{\text{list}} \rightarrow \infty$, is there more than one $f(\boldsymbol{r})$ consistent with the data? If there are multiple solutions can we put any bounds on their differences?

We must worry about three potential sources of non-uniqueness: basis truncation, EM starting point, and the identifiability of $\boldsymbol{\alpha}$ from $p\left(\{\boldsymbol{g}_i\} | \boldsymbol{\alpha}\right)$. (As will be discussed in section 1.7.6.3, the parameter is identifiable if different values of $\boldsymbol{\alpha}$ are guaranteed to produce different probability laws.)

1.7.6.1 Non-uniqueness due to basis truncation Having decided in (1.2) to restrict ourselves to a finite dimensional subspace of $\mathbb{L}_2$ we've given up any hope of uniquely reconstructing $f(\boldsymbol{r})$, as there are infinitely many $\mathbb{L}_2$ functions that share identical $\alpha_1, \cdots, \alpha_{N_{\text{basis}}}$ but differ for higher-order basis elements. Although the truncation error can be made smaller by increasing the number of basis elements, we usually will not have the strong prior knowledge required to guarantee it is zero. Furthermore, increasing the number of basis elements will also increase the uncertainty in our reconstruction; the error bars about each coefficient will get bigger. Therefore we must pick basis elements which capture important aspects of the object. This begs the question: how do we measure importance? Our answer is to use the techniques for measuring reconstruction algorithm quality which will be introduced in section 1.9.

If we choose non-negative basis elements $\phi_j(\boldsymbol{r}) \geq 0$, and also restrict ourselves to non-negative coefficients $\alpha_j \geq 0$, then we can leverage our knowledge that $f(\boldsymbol{r}) \geq 0$, to derive a bound on the truncation error:

$$\|\epsilon_{\text{trunc}}\|_1 \leq \|f(\boldsymbol{r})\|_1 + \left\| \sum_j \alpha_j \phi_j(\boldsymbol{r}) \right\|_1 \tag{1.36}$$

The proof is fairly simple. By definition

$$\epsilon_{\text{trunc}} = f(\boldsymbol{r}) - \sum_j \alpha_j \phi_j(\boldsymbol{r}). \tag{1.37}$$

First let us consider what happens if there exists another non-negative function $f_0(\boldsymbol{r})$ which has the same set of coefficients, i.e.

$$\int_\infty f_0(\boldsymbol{r}) \phi_j(\boldsymbol{r})\, d\boldsymbol{r} = \alpha_j \quad \forall j. \tag{1.38}$$

Note that

$$f_\alpha(\boldsymbol{r}) = \sum_j \alpha_j \phi_j(\boldsymbol{r}) \tag{1.39}$$

is one example of such a function, and

$$\epsilon_{\text{trunc}} = f(\boldsymbol{r}) - f_\alpha(\boldsymbol{r}). \tag{1.40}$$

Therefore if we can derive a bound on $\|f(\boldsymbol{r}) - f_0(\boldsymbol{r})\|_1$, we will also have a bound on $\|\epsilon_{\text{trunc}}\|_1$.

Although $f(\boldsymbol{r})$ and $f_0(\boldsymbol{r})$ are both non-negative, their difference need not be. Let

$$n(\boldsymbol{r}) \;\; = f(\boldsymbol{r}) - f_0(\boldsymbol{r}) \;\; = n_+(\boldsymbol{r}) - n_-(\boldsymbol{r}), \tag{1.41}$$

where $n_+(\boldsymbol{r})$ is the positive portion of $n(\boldsymbol{r})$ and $n_-(\boldsymbol{r})$ is the negative portion. Because $f(\boldsymbol{r})$ and $f_0(\boldsymbol{r})$ are non-negative,

$$n_+(\boldsymbol{r}) \leq f(\boldsymbol{r}) \tag{1.42}$$

therefore

$$\|n_+(\boldsymbol{r})\|_1 \leq \|f(\boldsymbol{r})\|_1 . \tag{1.43}$$

Similarly,

$$n_-(\boldsymbol{r}) \leq f_0(\boldsymbol{r}) \tag{1.44}$$

$$\|n_-(\boldsymbol{r})\|_1 \leq \|f_0(\boldsymbol{r})\|_1 . \tag{1.45}$$

Combining (1.43) and (1.45) yields the bound (1.36). (This argument is in the same spirit as the positivity results of Clarkson and Barrett [1997, 1998].) Unfortunately this bound is fairly large, and equality can be achieved by choosing $f(\boldsymbol{r})$ to be a sum of Dirac delta functions.

1.7.6.2 Non-uniqueness due to EM starting point The EM algorithm is guaranteed to converge only to a point where the gradient of the likelihood function is zero, not its global maximum. Therefore if the likelihood has multiple maxima it is possible for EM to get stuck at a local maximum, in which case our estimate would depend on the initial guess used to start iteration. In section 1.7.3 we saw that choosing a starting point with $\alpha_j = 0$ for some j forever restricted the updates to that hyperplane. For this reason the starting point is always chosen such that $\alpha_j > 0 \, \forall j$. (A typical choice

is $\alpha_j = 1$.) The projection-mode EM algorithm is known to have a unique fixed-point, provided one starts somewhere in the positive orthant [Byrne, 1993]. Parra and Barrett [1998] showed that list-mode EM converges to a unique fixed point when the likelihood function is strictly concave; it is not obvious (at least to this author) that in practice list-mode likelihood functions are concave.

Much of the analysis of the EM algorithm has been done in a general context, without assuming the Poisson properties yielding the specific update in (1.32). Such an analysis only guarantees convergence to a local critical point of the likelihood, not to the global maximum, unless the likelihood is strictly concave. It is theoretically possible for the general version of the EM algorithm to become stuck at a saddle point (see Lange [1999] for an example), though in practice this doesn't occur. For the imaging applications we are interested in, the likelihood seems to be a well-behaved function, as we shall see in chapter 2.

1.7.6.3 Non-uniqueness due to a non-identifiable parameter Do there exist $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2$ such that $\boldsymbol{\alpha}_1 \neq \boldsymbol{\alpha}_2$ yet $p(\boldsymbol{g}|\boldsymbol{\alpha}_1) = p(\boldsymbol{g}|\boldsymbol{\alpha}_2)\ \forall \boldsymbol{g}$? If so then, even in the asymptotic limit of infinite data, there is no way to distinguish between $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$, and there is no unique ML estimate. Conversely, if we know

$$p(\boldsymbol{g}|\boldsymbol{\alpha}_1) \neq p(\boldsymbol{g}|\boldsymbol{\alpha}_2) \tag{1.46}$$

$$\text{and } p(N_{\text{list}}|\boldsymbol{\alpha}_1) \neq p(N_{\text{list}}|\boldsymbol{\alpha}_2) \tag{1.47}$$

$$\forall \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \; s.t. \; \boldsymbol{\alpha}_1 \neq \boldsymbol{\alpha}_2 \tag{1.48}$$

then, in the asymptotic limit of infinitely many list entries, we can distinguish between any two objects. We refer to this as the parameter being identifiable. In the next section we will derive conditions for identifiability and examine what happens in its absence.

Harris [1964] considered the question of distinguishing between a single-star and a double-star. He showed that if the parameter specifying the number of stars in

the likelihood model is identifiable, then in the limit of infinitely many photons one can distinguish between arbitrarily close stars. Furthermore, in this special case it is reasonable to believe that this parameter of the likelihood is identifiable even for real-world imaging systems. Note that the availability of an infinite number of list entries is essential for this conclusion. Our real-life experience that distinguishing between single- and double-stars is difficult is largely due to the paucity of photons available to us. In chapter 3 we shall consider this matter in more detail.

1.8 Null and Measurement space decomposition in list-mode systems

This section reviews the null/measurement decomposition of object space familiar from the traditional linear-systems view of imaging and extends the idea to list-mode systems. The binned-mode version of the decomposition is thoroughly discussed by Barrett and Myers [2004, chapter 1]; the list-mode extension is novel.[4] The discussion in this section focuses on a discretized object $\boldsymbol{\alpha}$. A similar analysis for a continuum object $f(\boldsymbol{r}) \in \mathbb{L}_2$ is found in chapter 4. The two analyses are equivalent only for the for the special case where all ϕ_j define estimable parameters. (The definition of an estimable parameter will be given in chapter 4.)

The fundamental equations of linear noise-free imaging are

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{\alpha} \tag{1.49}$$

for imaging systems viewing discretized objects and

$$\boldsymbol{g} = \mathcal{H}f(\boldsymbol{r}) \tag{1.50}$$

for systems viewing continuum objects. These are the same $\boldsymbol{\alpha}$ and $f(\boldsymbol{r})$ as in (1.2) and (1.1) respectively, albeit for a static object. Here $\boldsymbol{g}$ represents the entire binned-mode data set, while in the equations we saw earlier $\boldsymbol{g}_i$ was just one list entry.

[4] Eric Clarkson provided valuable help with this section.

Because $\boldsymbol{H}$ and $\mathcal{H}$ are linear operators they partition the object space into two orthogonal subspaces: the null space and its complement the measurement space. Any object can be written as the sum of measurement and null components:

$$\boldsymbol{\alpha} = \boldsymbol{\alpha}_{\text{meas}} + \boldsymbol{\alpha}_{\text{null}}, \tag{1.51}$$

$$f(\boldsymbol{r}) = f_{\text{meas}}(\boldsymbol{r}) + f_{\text{null}}(\boldsymbol{r}). \tag{1.52}$$

The measurement components pass through the system, while the null-space components are mapped to zero by $\boldsymbol{H}$ or $\mathcal{H}$, and are therefore invisible to the system:

$$\boldsymbol{H}\boldsymbol{\alpha}_{\text{null}} = 0 \tag{1.53}$$

$$\mathcal{H}f_{\text{null}}(\boldsymbol{r}) = 0. \tag{1.54}$$

Thus one cannot distinguish between $f(\boldsymbol{r})$ and $f_{\text{meas}}(\boldsymbol{r})$, even in the limit of no measurement noise. One can construct many objects consistent with the noise-free data $\boldsymbol{g}$ by adding arbitrary null functions to $f_{\text{meas}}(\boldsymbol{r})$.

A similar decomposition of object space exists for list-mode systems. Here we focus on the discretized vector case. Two objects $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ are indistinguishable by a list-mode system, even in the limit of an infinitely long data list, if

$$p(\boldsymbol{g}_i|\boldsymbol{\alpha}_1) = p(\boldsymbol{g}_i|\boldsymbol{\alpha}_2) \quad \forall i \tag{1.55}$$

$$\text{and} \quad p(\overline{N}_{\text{list}}|\boldsymbol{\alpha}_1) = p(\overline{N}_{\text{list}}|\boldsymbol{\alpha}_2). \tag{1.56}$$

When seeking null functions we are seeking $\boldsymbol{\alpha}_1 \neq \boldsymbol{\alpha}_2$ which satisfy (1.55) and (1.56).

Using (1.16) and (1.17), we see that

$$p\left(\boldsymbol{g}_i|\boldsymbol{\alpha}_1\right) = \sum_j p\big(\boldsymbol{g}_i\big|\text{originated in } \phi_j(\boldsymbol{r})\big)p(\phi_j(\boldsymbol{r})|\boldsymbol{\alpha}) \tag{1.57}$$

$$= \sum_j p\big(\boldsymbol{g}_i\big|\text{originated in } \phi_j(\boldsymbol{r})\big)\frac{\alpha_{1j}s_j}{\overline{N}_{\text{list}}}$$

$$= \sum_k p\big(\boldsymbol{g}_i\big|\text{originated in } \phi_j(\boldsymbol{r})\big)\frac{\alpha_{2k}s_k}{\overline{N}_{\text{list}}} \tag{1.58}$$

$$= p\left(\boldsymbol{g}_i|\boldsymbol{\alpha}_2\right)$$

must hold for such $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$. (The double subscript on $\boldsymbol{\alpha}$ is used to indicate both which component of the vector and whether we have object 1 or 2.) Therefore

$$\sum_j p\big(\boldsymbol{g}_i\big|\text{originated in}\,\phi_j(\boldsymbol{r})\big)\alpha_{1j}s_j = \sum_k p\big(\boldsymbol{g}_i\big|\text{originated in}\,\phi_j(\boldsymbol{r})\big)\alpha_{1k}s_k. \tag{1.59}$$

Define the kernel $h_j(\boldsymbol{g}_i)$ of the system operator to be

$$h_j(\boldsymbol{g}_i) = p\big(\boldsymbol{g}_i\big|\text{originated in}\,\phi_j(\boldsymbol{r})\big)s_j. \tag{1.60}$$

In (1.49) the system operator H, a matrix, is a discrete-to-discrete operator; in other words it maps a vector $\boldsymbol{f}$ to another vector $\boldsymbol{g}$. In the list-mode case the system operator maps a discrete vector $\boldsymbol{\alpha}$ to a continuum function of possible list entries $\boldsymbol{g}_i$. (Similar discrete-to-continuum operators are found in the analysis of consistency conditions for the continuum-to-discrete system of (1.52).)

Regrouping (1.59) gives

$$\sum_j h_j(\boldsymbol{g}_i)\left[\alpha_{1j} - \alpha_{2j}\right] = 0. \tag{1.61}$$

In other words, $\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2$ is in the null space of the list-mode imaging system.

If the functions $h_j(\boldsymbol{g}_i)$ are linearly independent then the system has no null functions other than the trivial $\boldsymbol{\alpha} = 0$, and we are guaranteed that different values of $\boldsymbol{\alpha}$ produce different probability laws. In practice showing linear independence may not be hard! For example, if we assume $p\big(\boldsymbol{g}_i\big|\text{originated in}\,\phi_j(\boldsymbol{r})\big)$ is a Gaussian, then having different a mean μ_j for each region $\phi_j(\boldsymbol{r})$ would be sufficient. Note that if we are fitting the parameters hidden in $p\big(\boldsymbol{g}_i\big|\text{originated in}\,\phi_j(\boldsymbol{r})\big)$ from calibration data then we will only be able to make conclusions about the null functions after calibration is complete.

As in the binned-mode case, null functions need not be physically realizable. In particular, although the physics constrains $\alpha_j \geq 0\,\forall j$ for physical objects, we allow α_j to go negative for null functions. The additional constraint of (1.56) places a limit

on how far negative α_j may go. By (1.12) we know that

$$\sum_{j=1}^{N_{\text{basis}}} \alpha_j s_j > 0. \tag{1.62}$$

So, just as in the traditional case of a linear binned-mode system, we can decompose the object $\boldsymbol{\alpha}$ into null and measurement components $\boldsymbol{\alpha}_{\text{null}}$ and $\boldsymbol{\alpha}_{\text{meas}}$. Because $\boldsymbol{\alpha}_{\text{null}}$ is invisible to the imaging system, we can only make inferences about $\boldsymbol{\alpha}_{\text{meas}}$. All properties of the binned-mode decomposition, for example bounds on $\boldsymbol{\alpha}_{\text{null}}$ because we have a positivity constraint $\alpha_j \geq 0$, carry over to the list-mode case.

As with the binned-mode null/measurement decomposition, which considers the optimistic limit of no measurement noise, this analysis considers only what happens in the limit of infinite list entries. It therefore places a boundary on what is achievable in the best-case, but doesn't tell us anything about how much worse things get in the presence of noise and modeling error. For that we need the techniques discussed in the next section.

1.9 Quality assessment of reconstruction algorithms

Asserting as we did in section 1.7.5 that the true maximum-likelihood reconstruction, for example obtained by running the EM algorithm to convergence, produces less useful images than stopping after only a few iterations begs the question: how do we know there is a difference in image quality, and how can we quantitatively measure that difference? This section provides a brief introduction to the quality-assessment of reconstruction algorithms. Chapter 3 will use these ideas to compare several reconstruction algorithms.

The word "useful" in the previous paragraph hints at our approach. The reconstruction algorithm is not, alas, itself of interest to many people. Nor is the physician or physiologist interpreting the images especially interested in the reconstructed image for its own sake. Rather, as discussed in section 1.1, they want to extract information

from the image to perform some clinical or scientific task. So to claim one reconstruction algorithm is better than another, or equivalently to say it produces better quality images, is really to say that it helps the photointerpreter more accurately perform the task at hand.

Relating algorithm quality to task performance may mean there is no "best" algorithm; different tasks may have their own optimal algorithms. For example, in many medical imaging situations it is common to produce an initial scout image, used to position the patient and select which anatomical region to focus on. Rapid reconstruction is crucial for this task; the person operating the system wants real-time feedback. For this task reconstruction time is an important part of algorithm quality assessment. On the other hand, it may be acceptable for the radiologist to wait an hour after imaging, if the extra time enables a slower reconstruction algorithm yielding a more accurate diagnosis. Just as different imaging modalities are chosen for different clinical tasks, we may choose different reconstruction algorithms for different jobs.

The sorts of clinical and scientific tasks we are interested in can be divided into several categories:

Quantitative estimation tasks where the goal is to produce a single number. How many mCi of activity are in the liver? What is the volume of this tumor? What is the rate at which the kidney filters the radiotracer out of the bloodstream?

Detection tasks Does this patient have a tumor? The objective of a detection task is to decide whether or not a specific signal is present in the image. In medical contexts the signal is usually a lesion or tumor. It could also be a bone fracture or arterial narrowing. In military contexts the signal may be a tank or an airplane.

Discrimination and classification tasks Is this lesion malignant or benign? Are we seeing one large or two small lesions? The goal is to assign the image into

one of several classes. The special case where we discriminate between only two classes is especially important—note that detection tasks can also be thought of as binary discrimination tasks.

Expository tasks for which the objective is to produce an image for use in teaching or research publication. Unlike the other tasks, where we are concerned with answering questions about a specific patient, in exposition we often have access to images from many patients and must choose one patient that illustrates a point. Quantitative quality metrics would require us to worry about the impact of this image on student test scores or the researcher's tenure committee, both of which are far afield from our main topic.

Aesthetic tasks in which the object is to produce an image with artistic value. We shall not further consider expository or aesthetic tasks, as quality assessment is fundamentally different than for the other tasks.

Because we are interested in the quality of the reconstruction algorithm, not in the quality of a specific reconstructed image, *the assessment must be statistical in nature.* Most imaging systems are used with many patients, so the quality assessment must consider that ensemble.

The reader is cautioned that "image quality" is a dangerous term, because different communities have radically differing opinions on what it means. The biggest dichotomy is between those who are interested, as we are, in the ensemble behavior of an imaging system and those who have a specific image in hand and want to know the quality of *that image.* When using an ensemble approach it is prudent to emphasize that one is interested in performance of the entire imaging system across a range of objects, not just in one image.

1.9.1 Ensemble mean error

In principle, the ensemble mean error (EMSE)

$$\left\langle \left\langle \left\| f(\boldsymbol{r},t) - \hat{f}(\boldsymbol{r},t) \right\| \right\rangle_{\boldsymbol{g}|f} \right\rangle_{f} \tag{1.63}$$

seems like a good way to measure reconstruction algorithm quality. Because we have specified that $f \in \mathbb{L}_2$, the 2-norm

$$\left\| f(\boldsymbol{r},t) - \hat{f}(\boldsymbol{r},t) \right\|^2 = \iint \left(f(\boldsymbol{r},t) - \hat{f}(\boldsymbol{r},t) \right)^2 d\boldsymbol{r}dt, \tag{1.64}$$

also known as the mean square error (MSE), seems the most natural way of measuring reconstruction error for one image. Averaging over all objects that may be placed into the imaging system and over all possible noise realizations leads to (1.63). EMSE suffers from several serious flaws that make it inappropriate to use for assessment of reconstruction-algorithm quality.

First, any kind of mean error requires both $f(\boldsymbol{r},t)$ and $\hat{f}(\boldsymbol{r},t)$ to be in the same Hilbert space. This is not a problem for us, as our reconstruction $\hat{\boldsymbol{\alpha}}$ can be mapped into the same space as $f(\boldsymbol{r},t)$ using (1.2). However many reconstruction algorithms do not ensure that $f(\boldsymbol{r},t)$ and $\hat{f}(\boldsymbol{r},t)$ are commensurable; various ways to proceed in that case are discussed by Barrett and Myers [2004, section 13.3.2]. (The Rayleigh task simulations we will see in chapter 3 have the object in a different space than the reconstructed image.)

Second, EMSE can be difficult to calculate in practice. Usually Monte Carlo methods are the only way to estimate the expected value in (1.63). Therefore we need to be able to draw many samples $f_k(\boldsymbol{r},t)$ from the ensemble of objects. Doing so for realistic objects is not trivial, although usable techniques now exist. Even worse, we need to know the true $f_k(\boldsymbol{r},t)$ to compute the norm; merely knowing the true $\boldsymbol{\alpha}$ is not enough, as ϵ_{trunc} also contributes to (1.63). Realistically this means we can only use EMSE in simulation studies with a discretized object or in carefully controlled phantom studies.

Third, any type of square-error penalizes large deviations more severely than small ones. If one is serious about minimizing EMSE, it is desirable to deliberately introduce small errors in a systematic fashion such that the likelihood of a large error is reduced. This is Stein's famous paradox; see Efron and Morris [1977] for a readable introduction.

Fourth, there is growing evidence that EMSE is a poor predictor of observer performance on photointerpretation tasks. Miguel Eckstein and colleagues have studied the effect of the specialized reconstruction algorithms used in the field of lossy image compression on tasks related to the diagnoses made in coronary angiography. In this case $f(\boldsymbol{r})$, the digital image before compression, is known exactly, as is the reconstructed (a.k.a. decompressed) image $\hat{f}(\boldsymbol{r})$, so (1.63) is easy to calculate. They have found that EMSE does not accurately predict performance on a lesion detection task, for details see e.g. Zhang et al. [2004a,b].

Fifth, the 2-norm is not the only possible norm. In particular the 1-norm is also natural, since the physics of nuclear medicine restrict us to $f(\boldsymbol{r},t) \in \mathbb{L}_1 \cap \mathbb{L}_2$. The ∞-norm also seems reasonable. The relationship between either of these norms and photointerpretation task performance has not been carefully studied, and may not exist.

The various forms of ensemble error are most appropriate when we are considering a quantitative estimation task. In this task we aren't interested in accurately estimating the full $f(\boldsymbol{r},t)$, but just some functional $\Theta[f(\boldsymbol{r},t)]$. A linear functional can be written as

$$\Theta[f(\boldsymbol{r},t)] = \iint f(\boldsymbol{r},t)W(\boldsymbol{r},t)d\boldsymbol{r}dt, \tag{1.65}$$

for some weighting function $W(\boldsymbol{r},t)$. Nonlinear functionals are also possible.

Some possible FOMs in this case are

$$\left\langle \left\langle \left| \Theta\left[f(\boldsymbol{r},t)\right] - \Theta\left[\hat{f}(\boldsymbol{r},t)\right] \right|^2 \right\rangle_{\boldsymbol{g}|f} \right\rangle_f \tag{1.66}$$

$$\left\langle \left\langle \left| \Theta\left[f(\boldsymbol{r},t)\right] - \Theta\left[\hat{f}(\boldsymbol{r},t)\right] \right| \right\rangle_{\boldsymbol{g}|f} \right\rangle_f \tag{1.67}$$

$$\max_{g,f} \left| \Theta\left[f(\boldsymbol{r},t)\right] - \Theta\left[\hat{f}(\boldsymbol{r},t)\right] \right|. \tag{1.68}$$

It is not obvious that these ensemble mean functional errors are closely correlated with the ensemble mean errors such as (1.63). To the best of my knowledge the question has not been studied. (Note that if Θ is not an estimable parameter then even in the absence of measurement noise these expected values are not zero, see chapter 4.)

1.9.2 ROC analysis for detection and discrimination tasks

In detection and discrimination tasks we reduce the entire image to a single bit of information, a tremendous dimensionality reduction. By examining how accurately a human photointerpreter estimates this bit, in other words how accurately the image observer performs the task, we can quantitatively measure the quality of a reconstruction algorithm *for this task and observer*. This section gives only a brief introduction to quality assessment of binary decision tasks. For further details see Green and Swets [1966] or Barrett and Myers [2004]. A nice non-mathematical introduction to the same ideas, targeted towards public-policy makers, is given by Swets et al. [2000].

To make the ideas concrete, we shall always use the vocabulary of signal detection theory, even though any binary discrimination task can be analyzed using this framework. Therefore the two hypothesis we must choose between are H_a: signal absent, and H_p: signal present. (The decision not to label one of the choices H_0 is deliberate. Here we are choosing between two competing and well-defined alternatives. In classical hypothesis testing one "fails to reject" the null hypothesis, but never "accepts"

it.) We shall use D_a and D_p to indicate the observer's decision of signal absent or present, respectively. The observer is constrained to these two choices—saying "not sure" is not an option.

Percentage of correct signal detections is the most obvious candidate for a quality metric, but it is an inappropriate measure of quality because of its dependency on signal prevalence. Consider a hypothetical screening program for a rare form of cancer, found only once in 10,000 patients. An observer who always picks D_a (no cancer) without even looking at the image will be 99.99% accurate, and yet never detect a single case of cancer.

Instead we should consider the conditional probability of correct diagnosis given the patient's true status. There are several equivalent ways to do so. In keeping with convention in the imaging community, the parameters we shall consider are probability of detecting a signal that is actually present, $p(D_p|H_p)$, and the false-alarm rate, or probability of detecting a signal when no signal is present, $p(D_p|H_a)$. Because the observer only has two choices, other parameters of interest are directly related to these, for example the probability of missing a present signal

$$p(D_a|H_p) = 1 - p(D_p|H_p). \tag{1.69}$$

Table 1.2 relates these terms to those used in other communities.

Deterministic binary-discrimination algorithms can always be modeled as a two-step process. First, the input data are processed, perhaps nonlinearly, to compute a single scalar called the test statistic. Second, the test statistic is compared with a threshold. If the test statistic is greater than or equal to the threshold the decision is D_p, signal present. Otherwise the decision is D_a, signal absent.

A deterministic algorithm is one that always produces the same answer for a given data set; this eliminates random decision rules. Note that human photointerpreters are not deterministic—people will interpret the same image differently at different times. We shall sweep this under the rug. To rigorously handle it one would also

false-alarm rate	$p(D_p\|H_a)$, the probability of deciding the signal is present when in actuality it is absent.
probability of detection	$p(D_p\|H_p)$, the probability of correctly deciding the signal is present.
TPF	True-Positive Fraction, $p(D_p\|H_p)$, another way of saying probability of detection.
FPF	False-Positive Fraction, $p(D_p\|H_a)$, another way of saying false-alarm rate.
FNF	False-Negative Fraction, $p(D_a\|H_p) = 1 - p(D_p\|H_p)$.
TNF	True-Negative Fraction, $p(D_a\|H_a) = 1 - p(D_p\|H_a)$.
sensitivity	$p(D_p\|H_p)$, how MDs pronounce probability of detection.
specificity	$p(D_a\|H_a) = 1 - p(D_p\|H_a)$, another way of saying TNF.
prevalence	$p(H_p)$, the probability the signal is present.
Type-I error rate	$p(D_p\|H_a)$, how statisticians pronounce false-alarm rate.
Type-II error rate	$p(D_a\|H_p) = 1 - p(D_p\|H_p)$, how statisticians say FNF.

TABLE 1.2. Various terms used in ROC analysis, and equivalent vocabulary used in other disciplines.

need consider the ensemble of repeated readings of the same image in addition to the ensembles of images and readers.

Associated with the test statistic t is a signal to noise ratio

$$\mathrm{SNR}_t = \frac{\langle t \rangle_{H_p} - \langle t \rangle_{H_a}}{\sqrt{\frac{\sigma^2_{H_a} + \sigma^2_{H_p}}{2}}}. \tag{1.70}$$

SNR_t is often used as a FOM for the test statistic.

Each value of the decision threshold has associated with it a probability of detection $p(D_p|H_p)$ and a false-alarm rate $p(D_p|H_a)$. Increasing the threshold will lower both the probability of detection and the false-alarm rate. The aforementioned hypothetical observer who always says "signal absent" is using a threshold of $+\infty$.

Figure 1.9 shows the probability of detection plotted against false-alarm rate, where both are parameterized by the decision threshold. This graph is called the Receiver Operating Characteristic (ROC) curve. ROC analysis originated during World War II in the radar processing community and has since been applied to

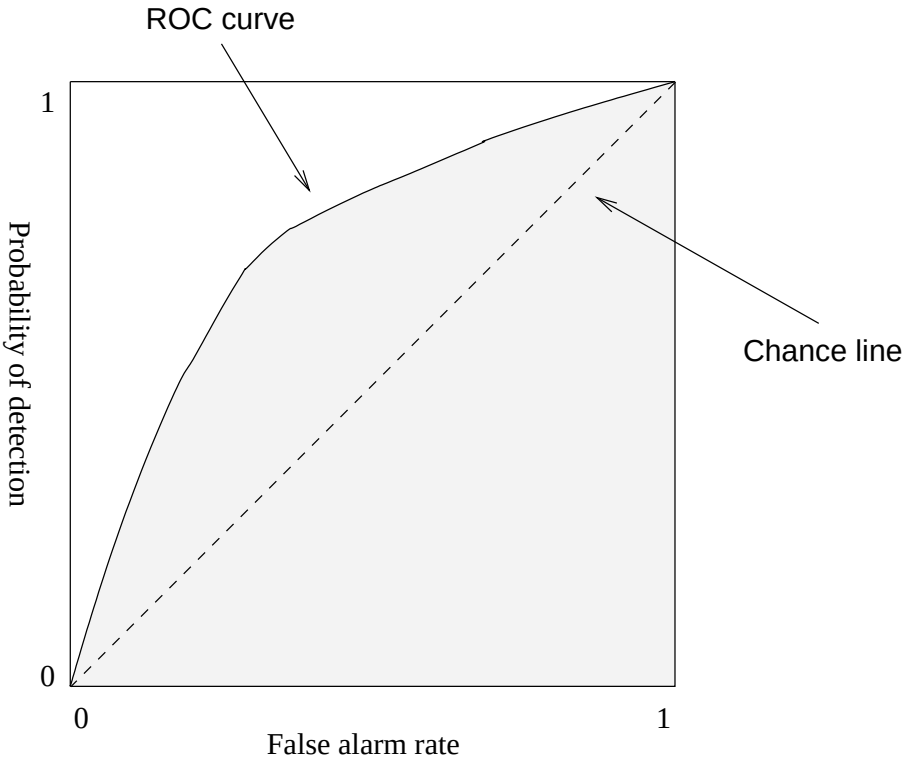

FIGURE 1.9. **Sketch of an ROC curve.** The decision threshold parameterizes the curve, with a threshold of $-\infty$ corresponding to the upper right corner and $+\infty$ to the lower left. The area under the curve (AUC) is shaded gray. The diagonal dashed line corresponds to the random observer described in section 1.11.3.

photointerpretation tasks in many domains.

Reconstruction algorithms may be compared by plotting ROC curves for both algorithms on the same set of axes. For a given false-alarm rate, the highest curve corresponds to the best algorithm. There is no a-priori reason why ROC curves for two different imaging systems may not cross; indeed in practice this does occur. Thus in different regimes of acceptable false-alarm rate different algorithms may be better. It is the combination of reconstruction algorithm and observer (or decision algorithm) that is being compared. Changing the observer could change which algorithm is better. Ideally, when comparing reconstruction algorithms, one should average over the performance of several human observers. If that is too expensive, model observers designed to mimic human performance may be used.

There is considerable evidence that thresholds vary under various circumstances. For example, some radiologists are more aggressive than others, i.e. they operate at different points on the ROC curve. Prior knowledge about a patient may also affect the threshold; chest x-rays of a patient with a forty-year history of smoking two packs a day will be read differently those of a twenty year old who never smoked. Thresholds may also drift over time. A good example can be found in mammography. As the risks associated with needle biopsy have decreased, radiologists have referred more patients for biopsy, increasing both the probability of detection and the false-alarm rate.

It is often desirable to reduce the information in an ROC curve down to a single number. There are many ways of doing so, but the most widely used is to compute the area under the curve (AUC). Better algorithms will have higher AUCs. A nice feature of AUC is that it can be estimated without computing the full ROC curve by doing a two-alternative forced choice (2-AFC) experiment. In a 2-AFC study the observer is shown many pairs of images, in which exactly one of the images contains a signal. The observer must choose which of the pair contains the signal. The proportion of correct decisions can be shown to be an unbiased estimator of the AUC. The AUC

is also related to the SNR_t of (1.70):

$$\mathrm{AUC} \approx \frac{1}{2} + \frac{1}{2}\,\mathrm{erf}\left(\frac{\mathrm{SNR}_t}{2}\right). \tag{1.71}$$

The equation is exact when the test statistic is normally distributed under both hypothesis.

1.10 Model tasks

Ideally, an imaging system should be evaluated using actual clinical data. However, doing so is difficult and expensive, prohibitively so for those doing basic research on system design. Running a new clinical trial each time one changes a system parameter is impossible. As a result the community has developed a variety of model tasks, which serve as proxies for actual clinical tasks.

A good model task has several conflicting desiderata:

- The model task should be amenable to theoretical analysis. We would like closed-form solutions for all relevant probabilities and the observer's optimal strategy.

- There should be a clear gold standard. One of the major difficulties in assessment of clinical task performance is knowing the true state of the patient. Often the imaging system under consideration is the best diagnostic we have; as a result assessing its accuracy is difficult. Model tasks bypass this problem.

- The task should be easy to simulate. We want to generate a large number of sample images which can be presented to the observer(s) while measuring task performance.

- The task should be realistic and similar to actual clinical tasks. The holy grail is a model task that passes the Turing test: can a trained photointerpreter tell whether she is reading a simulated image generated for a model task or an

actual image from the clinic? (Fortunately we can learn a lot from model tasks which fail this test.)

- We would like to know the strategies that observers use to perform the model task. This is often possible if the task can be related to psychophysical experiments.

1.10.1 Signal known exactly (SKE) tasks

The most common type of model task is called the Signal-Known-Exactly (SKE) paradigm. Other binary decision tasks, for example the Rayleigh task which we shall discuss in detail in chapter 3, can also be studied in the same framework. In an SKE task the observer must decide whether or not a known signal is present in the image. All relevant parameters of the signal are known to the observer, including location, size, shape, orientation, and so on. By eliminating search from the model task we consciously trade realism for the other desiderata. (But note that one way to model search for a known object is to perform an SKE task at each possible location!)

In the SKE paradigm we assume that the object can be decomposed into background and signal components,

$$f(\boldsymbol{r}) = b(\boldsymbol{r}) + s(\boldsymbol{r}). \tag{1.72}$$

Medical examples of a signal include tumors, bone fractures, and strokes; the background is the surrounding tissue. Although the signal in an SKE task is, by definition, known exactly, the background need not be. Choice of background can have a large impact on task performance, as shown by Myers et al. [1990]. For example, detecting a faint signal on a flat background is much easier than detecting the same signal on a complicated texture. If the reconstruction algorithm has tunable parameters, say the number of EM iterations, changing the nature of background is likely to affect the optimal choice of parameter value.

The special case where the background is known exactly (BKE) is widely used. In an SKE/BKE model task the signal contrast is usually low and the various sources of measurement noise high. Another, more realistic, possibility is the case where the background is random but known statistically (BKS). The lumpy background developed by Rolland and Barrett [1992] and used in section 3.2.2 and is an example.

1.11 Model observers

In section 1.9 we observed that measuring the task performance of human photointerpreters is the preferred way of assessing the quality of a reconstruction algorithm. Doing experiments with actual human observers presents a variety of logistical hurdles: first one must get permission from the Human Subjects Committee, then one must gather several observers and have them perform the task on a collection of images. This is time consuming and requires a budget to recompense the observers. (Even in the most optimistic case, a certain amount of beer and pizza is required.) A variety of model observers have been invented to avoid these problems. A model observer is an algorithm used to predict human performance or to analyze the theoretical bounds on task performance. Model observers may also be used to perform the task in place of, or in assistance to, a human observer. This section describes several of the most widely used observers, including the ideal observer, the Hotelling observer and the channelized-Hotelling observer (CHO).

It is important to note that we are primarily interested in using the model observer to predict human task performance, not to actually perform the task. The channelized-Hotelling observer we will use in chapter 3 to assess reconstruction algorithm quality is deliberately suboptimal, and more accurate algorithms are known. If we wanted a machine to perform the task we would use one of these better algorithms!

1.11.1 Ideal observer

The first model observer we shall consider is called the Bayesian ideal observer, or simply the ideal observer. It is of considerable theoretical interest, despite being a poor model of human performance on many tasks. The ideal observer is almost omniscient—it fully understands all of the relevant probability laws: The ideal observer knows the prior probability $p\left(f(\boldsymbol{r})\right)$ of placing an object into the system. And it knows all about the imaging system's blur and measurement noise, in other words it knows $p\left(\{\boldsymbol{g}_i\}|f(\boldsymbol{r})\right)$. It doesn't even have to worry about modeling error! The only thing the ideal observer doesn't know is the object's true state. The ideal observer is therefore able to compute the likelihood ratio

$$t_{\text{ideal}} = \frac{p\left(\{\boldsymbol{g}_i\}|\, H_p\right)}{p\left(\{\boldsymbol{g}_i\}|\, H_a\right)}, \tag{1.73}$$

and thus make the most accurate inference possible given then constraints of the imaging system. It provides an upper bound on the performance of humans or any other observer. Computing is often easier with the log of the likelihood ratio:

$$\lambda_{\text{ideal}} = \log\left(t_{\text{ideal}}\right) = \log\left[p(\boldsymbol{g}_i|H_p)\right] - \log\left[p(\boldsymbol{g}_i|H_a)\right] \tag{1.74}$$

In practice we are able to calculate the ideal observer's performance only for simple model tasks, although there has recently been progress in estimating performance on more complicated tasks. We shall mostly ignore the ideal observer, as it is prohibitively expensive to calculate for the model tasks we have in mind.

Furthermore, the ideal observer is of little use in assessing image reconstruction algorithms, because it acts on the raw data. In the case of modular gamma-camera data, the ideal observer considers a list of 9-dimensional PMT data, not the much higher-dimensional reconstructed image. (The ideal observer is useful for hardware designers, as it measures of how much information about the task passes through the imaging system, and thus gives an upper bound on reconstruction algorithm performance.)

1.11.2 List-mode ideal observer

Barrett et al. [1997] considered the list-mode ideal observer for an SKE task. They show for fixed-time imaging the log of the likelihood ratio is

$$\lambda_{\text{ideal}} = \overline{N_{\text{list|a}}} - \overline{N_{\text{list|p}}} + \sum_{j=1}^{N_{\text{list}}} log \left[\frac{\overline{h_p}(\boldsymbol{g}_j)}{\overline{h_a}(\boldsymbol{g}_j)} \right], \tag{1.75}$$

where

$$\overline{h_i}(\boldsymbol{g}_j) = \overline{N_{\text{list|i}}} \quad p\left(\boldsymbol{g}_j | H_i\right). \tag{1.76}$$

They also show for list-mode data the ideal observer's SNR (1.70) on an SKE/BKE task is:

$$SNR^2 = \frac{\left\{ \int d\boldsymbol{g} \left[\overline{h_p}(\boldsymbol{g}) - \overline{h_a}(\boldsymbol{g}) \right] \log \left[\frac{\overline{h_p}(\boldsymbol{g})}{\overline{h_a}(\boldsymbol{g})} \right] \right\}^2}{\frac{1}{2} \int d\boldsymbol{g} \left[\overline{h_a}(\boldsymbol{g}) + \overline{h_p}(\boldsymbol{g}) \right] \log^2 \left[\frac{\overline{h_p}(\boldsymbol{g})}{\overline{h_a}(\boldsymbol{g})} \right]}. \tag{1.77}$$

1.11.3 Random observer

The random observer ignores the data and performs the signal detection task solely by chance. More precisely, it performs a Bernoulli trial with the desired probability of detection p_d and false-alarm rate of $1 - p_d$. The random observer's ROC curve is a diagonal line, as shown in figure 1.9, with an AUC of 0.5.

The random observer is of theoretical interest because flipping a coin should provide a lower bound on the task performance of any observer. After all, any observer whose ROC curve goes below the chance line could just invert its decision. It is sobering to note that photointerpreters performing worse than the random observer are well documented. [Swets et al., 2000]

1.11.4 Hotelling observer

The Hotelling observer performs a binary-decision task by computing a linear functional of the data and comparing it with a threshold. For some simple model tasks

it can be shown that the Hotelling observer is the ideal observer, though in most cases the ideal observer is nonlinear. The Hotelling observer's appeal lies in its ease of calculation (at least when compared with the ideal observer) and amenity to theoretical analysis. It is often used as a proxy for the ideal observer. This section gives the major results; detailed theory of the Hotelling observer, including extension to non-binary classification tasks, is developed by Barrett and Myers [2004].

The Hotelling observer can be expressed simply in vector notation:

$$t_{\text{Hot}} = \boldsymbol{w}^t \boldsymbol{x}, \tag{1.78}$$

where $\boldsymbol{x}$ represents the data, ^{t} means matrix transpose, $\boldsymbol{w}$ is a template vector of weights, and t_{Hot} is the scalar test statistic. The switch in notation from $\boldsymbol{g}$ to $\boldsymbol{x}$ is deliberate, because the Hotelling observer is used in various ways. When applied to a reconstruction, $\boldsymbol{x}$ is the reconstructed image. Below we shall consider a single data-list entry to be $\boldsymbol{x}$. In section 1.11.6 $\boldsymbol{x}$ represents channel outputs. In all three cases the test statistic t_{Hot} is a linear function of the $\boldsymbol{x}$.

The weighting template incorporates the first and second moments of $\boldsymbol{x}$:

$$\boldsymbol{w} = \boldsymbol{K}_{\boldsymbol{x}}^{-1} \Delta\overline{\boldsymbol{x}}, \tag{1.79}$$

where $\boldsymbol{K}_{\boldsymbol{x}}^{-1}$ is the average covariance matrix of the data under both hypotheses

$$\boldsymbol{K}_{\boldsymbol{x}} = \frac{1}{2}\left(\boldsymbol{K}_{\boldsymbol{x}|H_a} + \boldsymbol{K}_{\boldsymbol{x}|H_p}\right), \tag{1.80}$$

and $\Delta\overline{\boldsymbol{x}}$ is the difference of the mean vectors under both hypotheses

$$\Delta\overline{\boldsymbol{x}} = \langle \boldsymbol{x} \rangle_{H_p} - \langle \boldsymbol{x} \rangle_{H_a} \tag{1.81}$$

One should use the true population values for the means and covariances, but in practice we sometimes estimate them from a training sample.

The Hotelling observer's SNR, as defined in (1.70), is

$$\text{SNR}_{\text{Hot}}^2 = \Delta\overline{\boldsymbol{x}}^t \boldsymbol{K}_{\boldsymbol{x}}^{-1} \Delta\overline{\boldsymbol{x}} \tag{1.82}$$

$$= \text{trace}\left[\boldsymbol{K}_{\boldsymbol{x}}^{-1} \Delta\overline{\boldsymbol{x}} \Delta\overline{\boldsymbol{x}}^t\right]. \tag{1.83}$$

The SNR is also sometimes called the Hotelling trace because of (1.83).

In the case where $\boldsymbol{x}$ represents an image it can be difficult to estimate the covariance of the image $\boldsymbol{K}_{\boldsymbol{x}}$. If $\boldsymbol{x}$ represents an $n \times n$ image it has n^2 pixels and $\boldsymbol{K}_{\boldsymbol{x}}$ has n^4 elements. Directly estimating the covariance would require a huge library of training samples, $O(n^4)$ images for each hypothesis. If $n = 64$ we'd already be into the tens of millions. There are several ways to avoid this dimensionality problem. The channelized-Hotelling observer, presented in section 1.11.6, is one. Other approaches are outlined in Barrett et al. [2001].

Note that even for a low-dimensional $\boldsymbol{x}$, a large number of training samples may be needed to accurately estimate SNR_{Hot} if the statistics are non-Gaussian. This was the case for the list-mode Hotelling observer calculations in section 3.4.1, where $\boldsymbol{x}$ consisted of the 9 PMT values for one scintillation event from a non-trivial object

1.11.5 Naive list-mode Hotelling observer

When list entries are i.i.d., as is the case for the systems we are interested in, the naive list-mode Hotelling observer is easy to compute. (The adjective naive is used to distinguish this Hotelling observer from the other list-mode Hotelling observer introduced in section 4.4.) First consider the Hotelling observer acting on a single list entry,

$$\boldsymbol{w}_{\boldsymbol{g}_i} = \boldsymbol{K}_{\boldsymbol{g}_i}^{-1} \overline{\Delta \boldsymbol{g}_i}, \tag{1.84}$$

which yields

$$SNR^2_{\boldsymbol{g}_i} = \text{trace}\left[\boldsymbol{K}_{\boldsymbol{g}_i}^{-1} \overline{\Delta \boldsymbol{g}_i}\, \overline{\Delta \boldsymbol{g}_i}^{t}\right]. \tag{1.85}$$

If we have a list with N_{list} entries it can be written as a single vector

$$g_{\text{list}} = \begin{bmatrix} \boldsymbol{g}_1 \\ \vdots \\ \boldsymbol{g}_{N_{\text{list}}} \end{bmatrix} \tag{1.86}$$

by concatenating the list entries. The covariance matrix of the entire list will be block diagonal

$$\boldsymbol{K}_{\text{list}} = \begin{bmatrix} \boldsymbol{K}_{\boldsymbol{g}_i} & & & 0 \\ & \boldsymbol{K}_{\boldsymbol{g}_i} & & \\ & & \ddots & \\ 0 & & & \boldsymbol{K}_{\boldsymbol{g}_i} \end{bmatrix}. \tag{1.87}$$

As a result the Hotelling trace of the whole list is easy to calculate:

$$SNR^2_{\text{list}} = N_{\text{list}} SNR^2_{\boldsymbol{g}_i} \tag{1.88}$$

In other words the list-mode Hotelling observer's SNR increases as $\sqrt{N_{\text{list}}}$, a result which should not be surprising. It is not clear how useful a figure of merit (1.88) is. Because the list entries have non-Gaussian statistics, the ideal list-mode observer will be highly nonlinear. The list-mode Hotelling observer thus serves as an easy-to-compute lower bound on the performance of the list-mode ideal observer.

1.11.6 Channelized-Hotelling observer

The channelized-Hotelling observer (CHO) was introduced by Myers and Barrett [1987]. The CHO reduces dimensionality by filtering the reconstructed image with a bank of channels. The output of one channel is usually computed by taking the inner product of a reconstructed image and a template mask

$$c_i = \boldsymbol{m}_i^\dagger \boldsymbol{x}. \tag{1.89}$$

Equivalently the output of all channels can be written as the result of a matrix multiplication

$$\boldsymbol{c} = \boldsymbol{M}\boldsymbol{x}. \tag{1.90}$$

(In the pattern-recognition literature $\boldsymbol{c}$ is called a feature vector.) We focus on linear channels for simplicity and because in practice they produce useful results. There is no reason one couldn't use nonlinear channels.

The precise choice of channels will depend on the type of observer we are modeling with the CHO. If the CHO is being used to approximate the Hotelling observer acting on the full image, one should use sufficient channels, whereas when approximating human observer performance it is more appropriate to mimic the human visual system with anthropomorphic channels. We focus on the latter case, but for completeness briefly discuss the former.

1.11.6.1 Sufficient channels Sufficient channels do not lose any information during the dimensionality reduction from image space to channel space. The ideal or Hotelling observer acting on the output of a set of sufficient channels computes the same decision as the un-channelized-ideal or un-channelized-Hotelling observer respectively. In the language of statisticians, the sufficient channels are a sufficient statistic for the full image.

In practice we don't know how to rigorously prove a set of channels is sufficient for any non-trivial task. However it may be possible to create approximately-sufficient channels, in other words produce channels that do a good job of approximating the unchannelized observer. Gallas and Barrett [2003] provide one example of how one might do so, using the CHO to estimate the true-Hotelling observer. Gallas [2003] suggests one might be able to use training data to search for approximately sufficient channels.

Others have used the phrase "efficient channels" to describe the same concept. My objection to the word "efficient" is that it triggers principal-components intuition in those who are hearing the concept for the first time. Too often they latch onto the principal-components idea and never figure out what is really happening.

1.11.6.2 Anthropomorphic channels On the other hand, if the CHO is being used as a predictor of human performance, we should use anthropomorphic channels that mimic known properties of the visual system. This is the case we are interested in. There is considerable physiological and psychological evidence that the human visual

system employs some sort of channel mechanism. A review of the literature is given by Barrett and Myers [2004, sec. 14.2.1].

There are several ways to construct anthropomorphic channels. For a rotationally-symmetric task Abbey [1998] had good success with channels constructed from the difference of Gaussians. Because in chapter 3 we shall be interested in a task with an obvious orientation axis, we focus on the Gabor channels introduced by Watson [1983]. Watson showed the Gabor channels accurately predict human performance in a variety of psychophysical experiments. More recently the set of 80 Gabor channels described here has been successfully used in a variety of medical-imaging contexts [Eckstein et al., 1998, 1999; Shimozaki et al., 2003; Zhang et al., 2004a,b].

Each channel template consists of a Gaussian envelope function multiplied by an oriented sinusoidal grating:

$$c(x,y) = \exp\left[-4(\ln 2)\frac{x^2+y^2}{w^2}\right]\cos\left[2\pi f\left(x\cos\theta + y\sin\theta\right)+\beta\right], \tag{1.91}$$

where f is grating's spatial frequency, θ is one of eight equally-spaced orientations, β is the phase, and the width of the envelope w is related to the frequency by

$$w = \frac{6\ln 2}{\pi f}. \tag{1.92}$$

The values used for the other parameters are:

$$f \in \{1,2,4,8,16\}, \tag{1.93}$$

$$\theta \in \left\{0, \frac{\pi}{8}, \frac{\pi}{4}, \frac{3\pi}{8}, \frac{\pi}{2}, \frac{5\pi}{8}, \frac{3\pi}{4}, \frac{7\pi}{8}\right\} \tag{1.94}$$

$$\beta \in \left\{0, \frac{\pi}{2}\right\}. \tag{1.95}$$

Figure 1.10 shows the eighty resulting channel templates.

1.11.6.3 Internal noise The channelized-Hotelling observer often overestimates human performance, even when using anthropomorphic channels. To lower the CHO's performance into the human range, and to model observer uncertainty, the channel

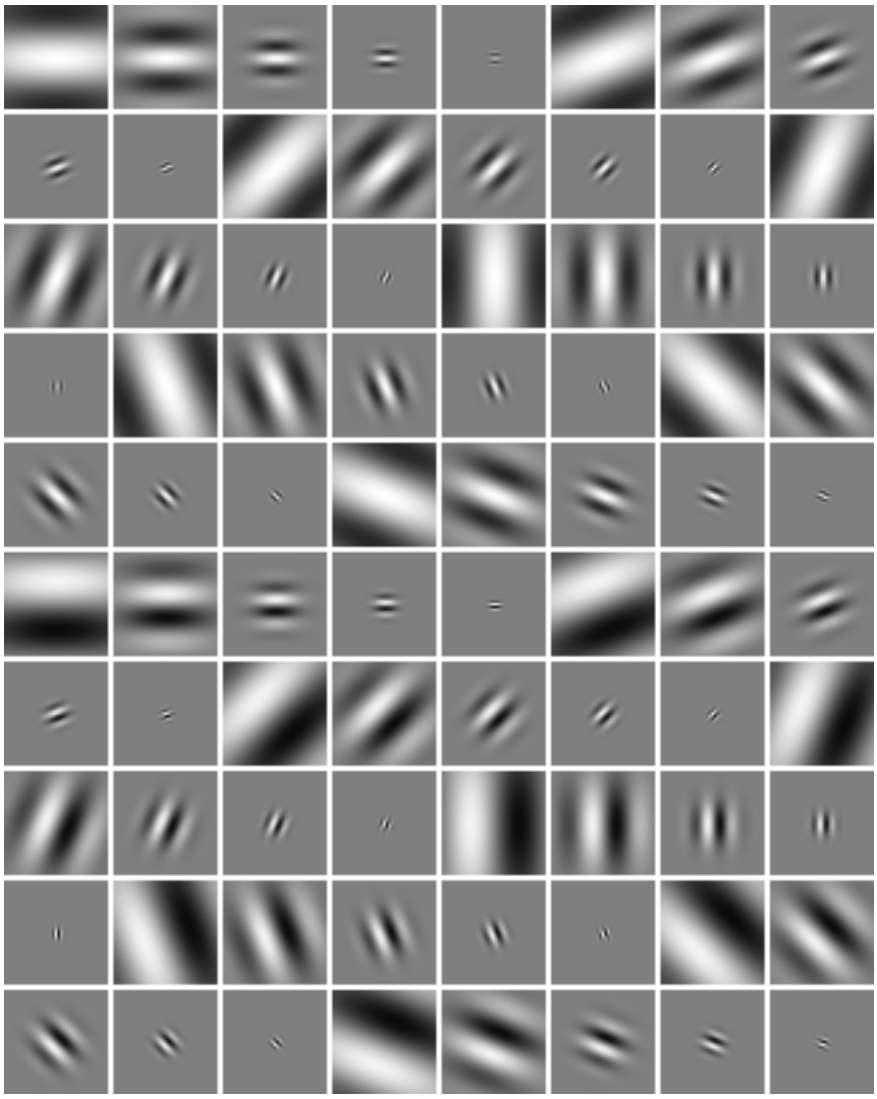

FIGURE 1.10. **The Gabor-channel templates.** Shown are the eighty templates used by the anthropomorphic channelized-Hotelling observer. Template values range from -1.0 (black) to 1.0 (white). The gray background visible in many channels has value 0.0.

output is often corrupted with noise. Because this noise is found only in the observer it is called internal noise. Abbey and Bochud [2000] show that randomizing the output of each channel with Guassian noise is equivalent to adding a single random variable to the test statistic. Therefore the CHO with internal noise is

$$t_{\mathrm{CHO}} = \Delta\overline{\boldsymbol{c}}^{t} K_{\mathrm{CHO}}^{-1} \boldsymbol{c} + \epsilon, \tag{1.96}$$

where ϵ is a zero-mean Gaussian random variable. The variance σ^2 of ϵ is a tunable parameter.

1.12 Preview of coming chapters

Chapter 2 discusses how to calculate $p\big(\boldsymbol{g}_i \big| \text{originated in } \phi_j(\boldsymbol{r})\big)$ for a known imaging system, paying special attention to the case where extensive calibration data are available. This chapter introduces double list-mode reconstruction, a technique where list-mode calibration data are used to reconstruct the likelihood model and then combine it with list-mode object data to produce a maximum-likelihood reconstruction of the object.

Chapter 3 looks at the problem of reconstructing projection images from list-mode PMT data. The image quality of two algorithms is compared by an anthropomorphic channelized-Hotelling observer performing the Rayleigh task, a traditional measure of system resolution. One algorithm reconstructs the image by histograming ML position estimates, while the other uses the list-mode EM algorithm to produce a fluence estimate.

Chapter 4, originally written as a stand-alone article, considers the question: which parameters of an object can be accurately estimated from list-mode data? In the process the notion of a list-mode system operator is developed, as is the list-mode Hotelling observer. The estimability analysis also suggests an approach to doing list-mode reconstruction.

Finally chapter 5 gives the obligatory summary and directions for future work. For the reader's convenience appendix A provides a list of abbreviations and acronyms used in the text.

Chapter 2

Empirical list-mode likelihood models

This chapter describes several techniques for representing and computing the likelihood $p\big(\boldsymbol{g}_i\big|\text{originated in}\,\phi_j(\boldsymbol{r})\big)$ that a detected scintillation event $\boldsymbol{g}_i$ is associated with the region defined by $\phi_j(\boldsymbol{r})$. Although our emphasis will be on the modular gamma cameras found in FastSPECT-II, the same techniques can be applied to any list-mode gamma-ray imaging system. As previously discussed in section 1.6, the likelihood model for a single event is all we need in order to be able to compute the likelihood of the full data list, provided we ignore scatter and attenuation. Special attention is paid to the case where extensive calibration data, specifically samples from the likelihood function, are available. Because these models are heavily based on empirical data, we shall refer to them as empirical likelihood.[1] The concept of double list-mode reconstruction is introduced in section 2.5. Double list-mode is a nonparametric approach where list-mode calibration data are combined with list-mode data from an object to produce a maximum-likelihood reconstruction of the object.

In practice one may want to combine several of the approaches mentioned below, for example one could use nonparametric data-driven techniques for some components of $\boldsymbol{g}$ and analytic parametric models for others. Or one might use different models for different imaging tasks. For example, one is likely to use a quick-and-dirty approach to provide real-time feedback to the imaging-system operator, but a slower method to produce images used for research and publication.

[1] This use of the phrase "empirical likelihood" is original to the author; as far as I can tell the statistical community does not have any existing vocabulary for this concept. Instead of *empirical*, another possible word would have been *observed*, by way of analogy with the observed Fisher information matrix computed by Parra and Barrett [1998].

2.1 Available calibration data

A general imaging-system model has free parameters, allowing it to describe many potential physical systems. This section describes the various forms of calibration data available for tuning the model. This is chosen as the starting point because plots of the calibration data such as histograms and scattergrams should suggest what sorts of models are reasonable. Further details on the calibration process are given by ?.

2.1.1 Monte Carlo simulated data

Monte Carlo simulation of the measurement process is widely used as a form of calibration. Monte Carlo techniques are the gold-standard method of generating simulated data because they can model all sources of randomness in the measurement process. Simulation results are often used to select or tune the likelihood models used for image reconstruction. Despite the difficulties in coding a simulation which accurately models an imager, in particular one which correctly captures details of the collimator geometry and flaws such as hot or dead spots in the detector, this approach is more prevalent than physically measuring extensive calibration data. (PET systems may be easier to simulate than SPECT because they don't have a collimator. Commercial systems often have expensive steps in the manufacturing process to prevent the "personality quirks" found in the much cheaper modular cameras, and thus also may be easier to model in simulation.) For systems where it is impossible to measure calibration data, in particular proposed systems which haven't yet been built, Monte Carlo simulation may be the only possible way to proceed.

2.1.2 MDRF data for gamma camera calibration

To assign each detected event to a location on the camera face (position estimation), or to produce projection images using the 2-dimensional list-mode EM algorithm,

we need a likelihood model $p\big(\boldsymbol{g}_i\big|\text{originated in}\,\phi_j(\boldsymbol{r})\big)$ which relates a data list entry $\boldsymbol{g}_i$ to each region $\phi_j(\boldsymbol{r})$ on the camera face. This model is usually called the Mean Detector-Response Function (MDRF), even when it includes higher-order statistics. Each modular camera has a unique MDRF due to variations in the scintillation crystals, PMT response, and gain process. Therefore each camera must be individually calibrated.

The MDRF may drift as the scintillation crystal degrades with age. On shorter time scales, response may be fluctuate with temperature. Ideally the likelihood model should account for all of these effects, but in practice they tend to be ignored. By permanently mounting the cameras and carefully controlling climate in the imaging lab, one can get away with recalibrating the MDRF model only every year or so.

FastSPECT-II MDRF calibration is performed by a robot arm, as shown in figure 2.1. To gain acess to the camera face during MDRF calibration, the collimator housing must be removed from the system. Performing the MDRF calibration for all 16 cameras is labor intensive, because the robot must be manually positioned over each camera. Once in position, the robot scans a collimated-beam source across the camera face, stopping at preselected points on a Cartesian grid. A raster scan is used, because this allows more accurate positioning of the robot's stepper motors than would a slightly faster boustrophedonic scan. At each grid point a number of events are recorded.

Each calibration event can be thought of as a sample from the likelihood function, $p\big(\boldsymbol{g}_i\big|\text{originated in}\,\phi_j(\boldsymbol{r})\big)$, where j indexes the calibration grid points, and $\phi_j(\boldsymbol{r})$ is determined by the spread of the calibration beam. The calibration source is designed so that

$$\phi_j(\boldsymbol{r}) \approx \delta(\boldsymbol{r} - \boldsymbol{r}_j) \tag{2.1}$$

is a good approximation. Another reasonable assumption is that

$$\phi_j(\boldsymbol{r}) \approx \frac{1}{S}\,\text{rect}\left(\frac{\boldsymbol{r} - \boldsymbol{r}_j}{S}\right), \tag{2.2}$$

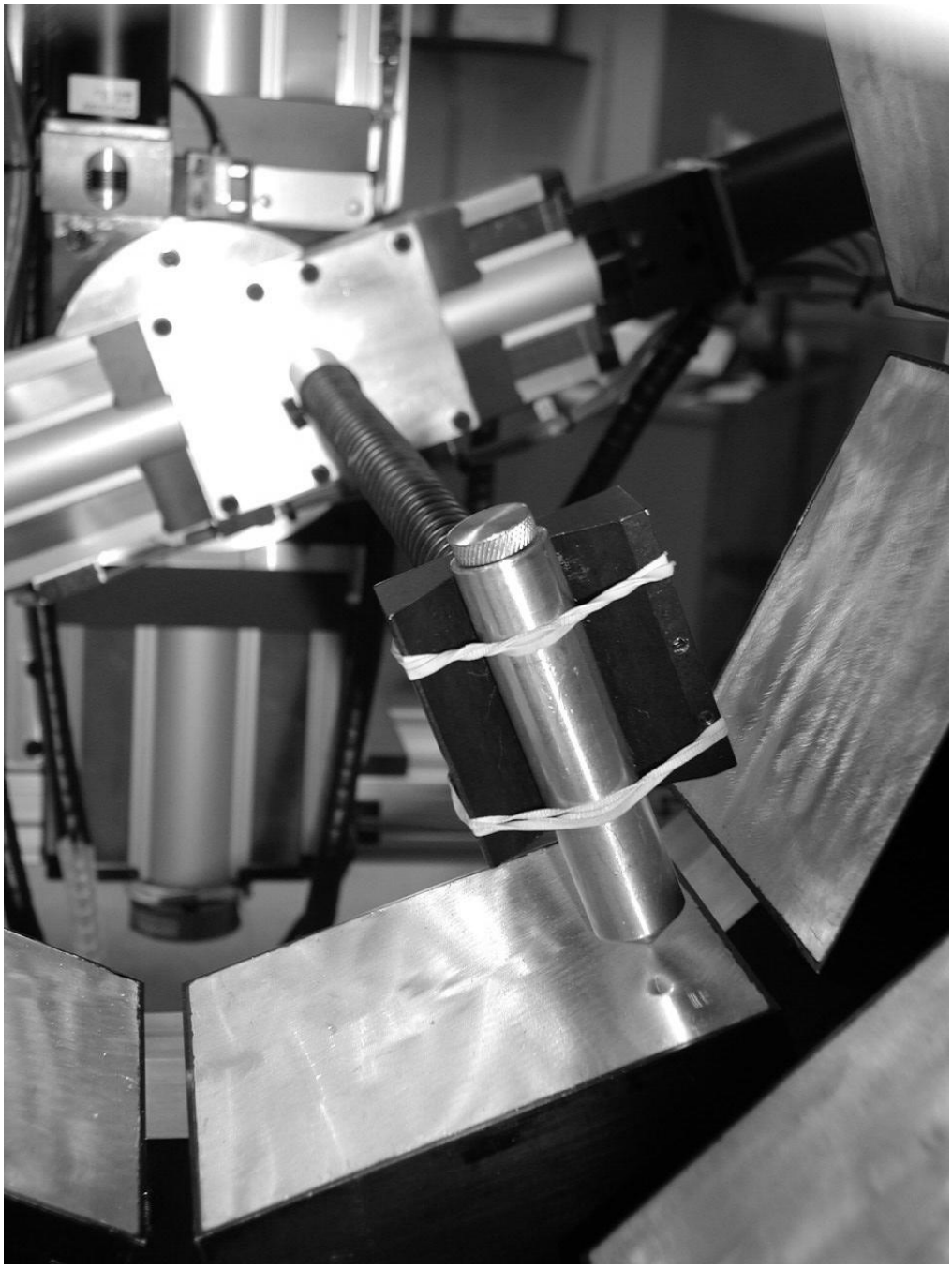

FIGURE 2.1. **Robot arm measuring the MDRF.** This robot is used to automatically acquire MDRF calibration data for one modular camera. Note the "space-age" rubber-band technology used to attach the pencil-beam collimator onto the arm. (Photo by J. Chen.)

where S is the grid spacing. This assumption corresponds to the square pixel basis illustrated in figure 1.7.

The MDRF calibration source is created by placing a syringe cap filled with activity at one end of a cylindrical metal holder. A 0.84mm opening at the other end of the cylinder allows the calibration beam to exit. The cylinder is lined with lead to prevent the enclosed activity from radiating in other directions. The opening is held approximately 1.5mm from the aluminum camera face. The collimated beam has a divergence of 0.056 radians (3.2°), which results in a beam width within the scintillation crystal of between 0.84 mm and 0.853 mm. A typical calibration grid spacing is 1.5 mm, which results in a 78×78 grid. Also typical is a count rate of c. 30 times the background. We usually record approximately 5000 events per grid point, for a total list-mode calibration file size of c. 523 MiB[2] per camera.

It is preferable to have the source above each grid location for a pre-specified time, as this guarantees Poisson statistics and makes various forms of estimation a bit easier. Calibration may also be done for a fixed number of counts at each location, but then the probability models for certain types of estimation are slightly less pleasant.

Figure 2.2 shows histograms of MDRF calibration data for one grid point. For this figure the MDRF source was located at calibration point (12,12), near the corner of the camera of PMT3. The pattern seen here is similar to that at other locations. In particular note that the tube output shows a Gaussian-like peak and may have a fat left tail. The fat tail, presumably caused by escaped Compton scatter, is strongest in the tube closest to the calibration point, in this case PMT 3, less pronounced on neighboring tubes, and not visible on further-away tubes. Peak PMT output has the highest value close to the source, and decreases with increasing distance from the

[2]MiB is short for mebibyte, or 2^{20} bytes. It is not just out of pedantry that I write the new binary prefixes instead of the more widely used, though technically incorrect, decimal prefixes. I often use units-aware calculators, for example the GNU/Linux command `units(1)`, to predict how long a file transfer or computation will take. These calculators distinguish between the two types of prefix, and give different answers depending on which is used.

tube.

2.1.3 PSF data for system calibration

As in the 2-dimensional case, to produce 3-dimensional and tomographic reconstructions we need a likelihood model relating a data list entry $\boldsymbol{g}$ to volume regions $\phi_j(\boldsymbol{r})$ within the imaging system. This model is usually called the Point Spread Function (PSF), even though it may account for nonlinear blurring and may be shift variant. The PSF includes measurement properties of the gamma cameras as well as effects due to the collimator geometry. Therefore calibration must be done each time the system geometry is changed, for example by swapping the collimator pinholes, moving the cameras, or replacing a camera. Because removing the collimator may inadvertently change the geometry and is necessary for MDRF measurement, PSF calibration must be done at least as often as MDRF calibration. Fortunately the PSF calibration runs unattended once started, so it is less labor-intensive than MDRF calibration. Depending on how subsequent processing is done, the PSF model may be used to partially correct for errors in the MDRF model due to drift. Therefore it may be desirable to measure the PSF more often than the MDRF.

PSF calibration of FastSPECT-II is performed by a robot arm moving a source to locations on a 3-dimensional Cartesian grid within the imaging system, and then recording number of events at each location. As in the 2-dimensional case, the source defines $\phi_j(\boldsymbol{r})$ and we are effectively sampling from the likelihood function. The source is created by saturating small (50 micron diameter) plastic resin chromatography beads with pertechnetate solution. The beads are then heated to remove excess water, allowing more pertechnetate to be added. Finally epoxy is used to glue the beads together and mount them on glass capillary. The final result is smaller than a half millimeter in diameter, and therefore approximately a point source:

$$\phi_j(\boldsymbol{r}) \approx \delta(\boldsymbol{r} - \boldsymbol{r}_j). \tag{2.3}$$

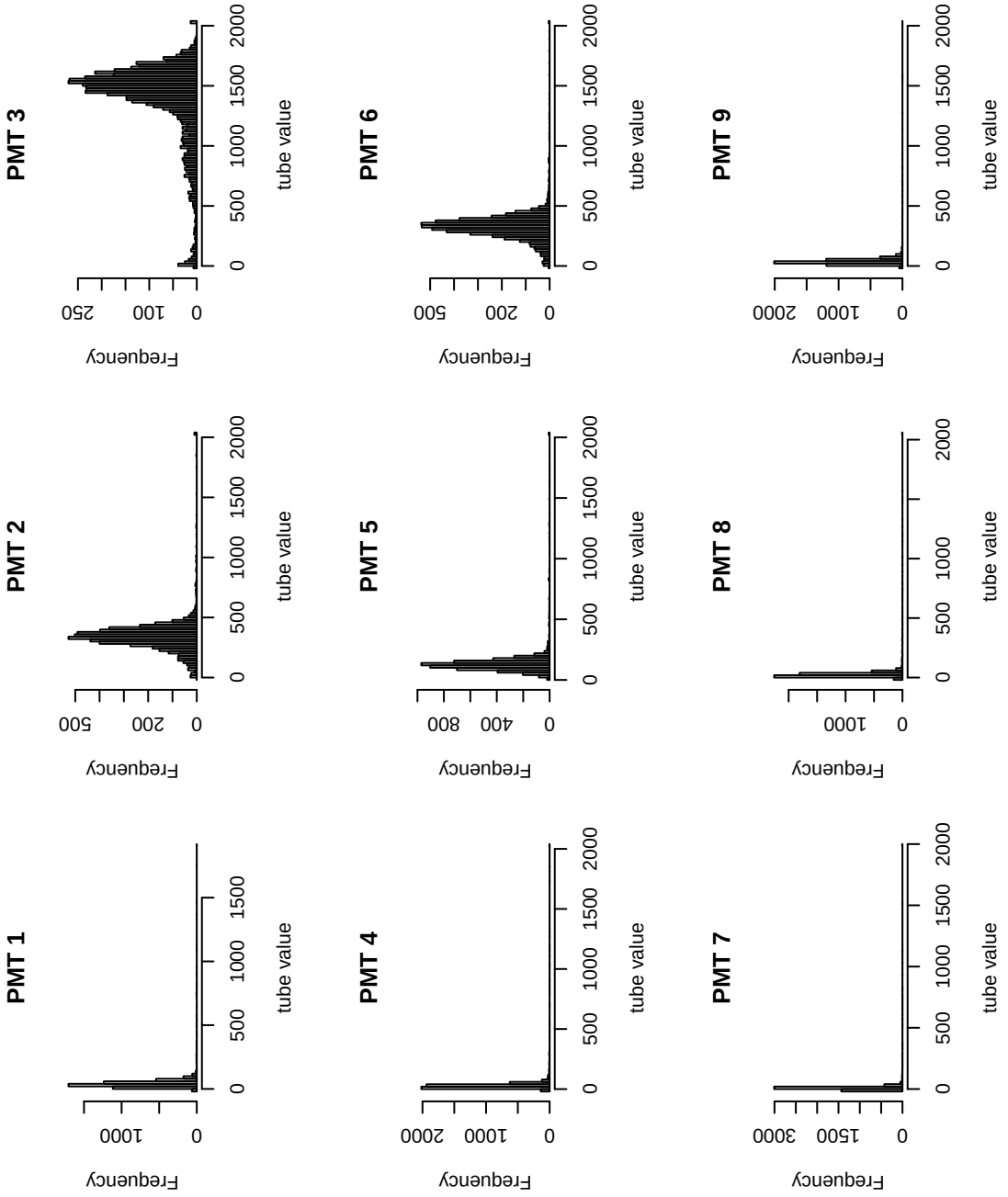

FIGURE 2.2. **Histograms of MDRF events.** The calibration point was (12,12), located in the corner of the camera near the center of PMT 3. Note the frequency axes is different for each tube. A scattergram of the same data are shown in figure 2.4.

After preparation, the capillary is attached to the robot arm and moved around the calibration grid.

The PSFs for all cameras in FastSPECT-II are simultaneously calibrated, therefore measurements are always done for a fixed time at each grid point. As calibration proceeds, the time at each location must be increased to compensate for radioactive decay of the PSF source. Fully sampling along the 1 mm × 1 mm × 1 mm grid we usually use requires about 24 hours (four half lives of $^{99\mathrm{m}}$Tc). The grid points at which the PSF is sampled can be embedded into a 41 × 41 × 41 Cartesian grid. In planes perpendicular to FastSPECT-II's central cylinder only points inside the inscribed circle are sampled. The list-mode PSF data are about 900 MiB per camera.

Histograms of PSF calibration data for one camera at one grid location are shown in figure 2.3. Several differences from the MDRF data shown in figure 2.2 are worth noting. First, there are fewer calibration samples. PSF data typically have fewer events per camera per grid point than MDRF data, and the variance in the number of events is higher; for this grid point the number of samples per camera varies from 357 events on camera 14 to 4051 events on camera 6. Second, the left tails are more pronounced. Third, the data appear to be bimodal. In this plot the bimodality is especially pronounced for tube 9, and also visible in tubes 5 and 8.

2.2 Parametric models

Among those who do maximum-likelihood reconstruction, a known analytic form is almost always assumed for $p\big(\boldsymbol{g}_i \big| \text{originated in } \phi_j(\boldsymbol{r})\big)$. (ML reconstructions are still not widely used in the clinic. Variations on Anger arithmetic remain the dominant technique for 2-dimensional position estimation more than fifty years after their invention. Barrett and Swindell [1981] provide a good introduction to the basic ideas. For tomographic reconstruction, filtered backprojection and its cousins remain the incumbents.) Common assumptions used to model statistics of the measurement

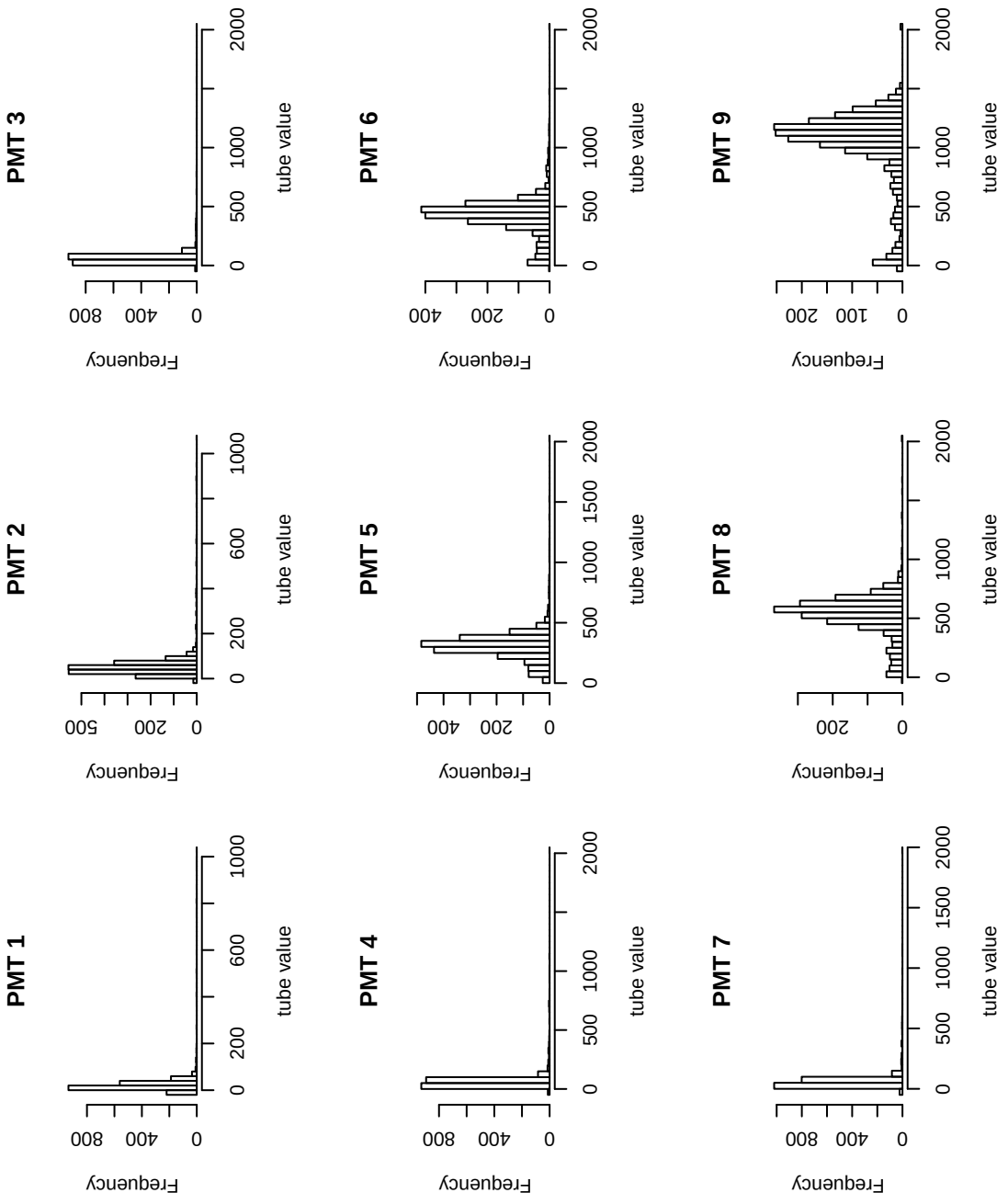

FIGURE 2.3. **Histograms of PSF events.** The calibration data was collected from point (1000,1000,1000), near the center of the imager. Shown are histograms for each PMT in camera 0.

process include conditional Poisson, where the mean PMT output μ_{tj} of tube t varies according to grid point j, conditional amplified Poisson, similar to conditional Poisson but with a gain process, and conditional Gaussian, where the mean and variance μ_{tj} and σ_{tj}^2 vary with grid point j. This section considers several possible parametric models for the likelihood and looks at how to pick the parameters.

Often a central-limit argument or knowledge of the measurement physics dictates a particular family of models. The biggest drawback of parametric models is difficulty in accurately modeling the system when we do not have such a prior reason to pick a known family. This drawback is usually outweighed by the many advantages of parametric models. For example, they provide a tremendous data reduction from the calibration data, are computationally efficient, and are amenable to theoretical analysis.

There are two basic ways of choosing the parameters. The first, and by far more widespread, is to model the collimator geometry and detector physics. The second, used here in Arizona but almost nowhere else, is to make extensive calibration measurements of the sort discussed in sections 2.1.2 and 2.1.3, and then fit the parameters at each grid point. This is the approach we shall focus on. (As in many other areas, there is an expense tradeoff here. Commercial manufacturers work hard to produce hardware for which simple approaches such as Anger arithmetic can be used to estimate the parameters. The modular cameras are much cheaper, but require us to use more sophisticated calibration and software.)

2.2.1 Conditional Poisson with lookup

The simplest likelihood model is conditional Poisson. This model can be used to describe the number of photoelectrons produced in the detector by a scintillation event, but is too simple to describe the output of most PMTs. (For an in depth discussion of PMT physics see Barrett and Myers [2004].) Note that we are modelling

the output of one detector element in response to one event as Poisson. It may also be the case that the number of events obeys Poisson statistics, but that is a not the issue of interest here.

The Poisson parameter is a function of the region associated with the scintillation event:

$$p\big(g_t\big|\text{originated in}\,\phi_j(\boldsymbol{r})\big) = \frac{\mu_{tj}^{g_t} e^{-\mu_{tj}}}{g_t!}, \tag{2.4}$$

$$p\big(\boldsymbol{g}\big|\text{originated in}\,\phi_j(\boldsymbol{r})\big) = \prod_{t=1}^{T} p\big(g_t\big|\text{originated in}\,\phi_j(\boldsymbol{r})\big). \tag{2.5}$$

The parameter μ_{tj} is the mean value (and, as this is Poisson, also the variance) of photocathode t's output for a scintillation event associated with region $\phi_j(\boldsymbol{r})$.

The mean must be estimated from the calibration data. Just as there are many image reconstruction algorithms, there are also many possible ways to choose $\hat{\mu}_{tj}$. The maximum likelihood approach is a good candidate for the reasons given in section 1.7.1. For the Poisson model, the ML estimate is obtained by computing the mean value of the calibration data at each point:

$$\hat{\mu}_{tj} = \frac{1}{N_{\text{list}}} \sum_{l=1}^{N_{\text{list}}} g_{tjl}. \tag{2.6}$$

Here is a short proof that the data (sample) mean is the ML estimate of the population mean:

$$p\left(\{g_{tjl}\}|\mu_{tj}\right) = \prod_{l=1}^{N_{\text{list}}} p\left(g_{tjl}|\mu_{tj}\right) \tag{2.7}$$

$$= \prod_{l=1}^{N_{\text{list}}} \frac{\mu_{tj}^{g_{tjl}} e^{-\mu_{tj}}}{g_{tjl}!}. \tag{2.8}$$

The log-likelihood is

$$l\left(\mu_{tj}|\{g_{tjl}\}\right) = \sum_{l=1}^{N_{\text{list}}} \log\left(\frac{\mu_{tj}^{g_{tjl}} e^{-\mu_{tj}}}{g_{tjl}!}\right) \tag{2.9}$$

$$= \sum_{l=1}^{N_{\text{list}}} g_{tjl} \log(\mu_{tj}) - \mu_{tj} - \log(g_{tjl}!). \tag{2.10}$$

Because logarithm is a monotonic function, the maximum of the log-likelihood is also the maximum of the likelihood. To find the maximum we compute the derivative,

$$\frac{dl}{d\mu_{tj}} = \sum_{i=1}^{N_{\text{list}}} \frac{g_{tjl}}{\mu_{tj}} - 1, \tag{2.11}$$

set it to zero,

$$0 = -N_{\text{list}} + \frac{1}{\mu_{tj}} \sum_{l=1}^{N_{\text{list}}} g_{tjl}, \tag{2.12}$$

and solve:

$$\hat{\mu}_{tj} = \frac{1}{N_{\text{list}}} \sum_{l=1}^{N_{\text{list}}} g_{tjl}. \tag{2.13}$$

To confirm this is a maximum we compute the second derivative,

$$\frac{d^2 l}{d\mu_{tj}^2} = -\mu_{tj}^{-2} \sum_{l=1}^{N_{\text{list}}} g_{tjl} \tag{2.14}$$

$$< 0. \tag{2.15}$$

(The Poisson assumption implies $g_{tjl} > 0$.)

In practice, we wouldn't want to use maximum-likelihood to estimate the Poisson parameter from the calibration data shown in figures 2.2 and 2.3. The fat left tail of each tube's histogram, caused by escaped Compton scatters, would pull the mean data value to the left of the photopeak, resulting in an underestimate of the photopeak mean. One possible approach is to mask off the tail before computing the mean, giving a Bayesian estimate of the parameter. Another is to use log-matched filtering to estimate the parameter [Aarsvold et al., 1988].

2.2.2 Transformed Poisson

In practice we will not find that PMT output obeys Poisson statistics. While the Poisson model does a good job of describing the number of photoelectrons produced by the PMT's cathode, the subsequent gain process violates the Poisson assumptions. In some systems, in particular the original modular camera electronics used in

FastSPECT-I, the Poisson statistics are also broken by applying a square root transform (see section 2.3.1.1). Therefore a more sophisticated approach than computing a simple mean is called for.

Aarsvold et al. [1988] explain how to handle transformed Poisson variables using log-matched filtering. Note that their approach relies on a Poisson number of counts, and thus depends on having fixed-time calibration data.

Advantages of the Poisson or transformed-Poisson approaches include computational efficiency during reconstruction and a compact representation of the model in memory. We need only store the mean value, a 4 byte float, per detector element per calibration location.

2.2.3 Conditional Gaussian with lookup

A conditional Gaussian model can also be used for the likelihood. Unlike the Poisson model, the Gaussian model can handle correlation between the different PMTs. Furthermore, because of the central limit theorem, a Gaussian does a good job of approximating a Poisson or amplified Poisson if the mean is bigger than about 10. For real-world PMTs the Gaussian model is better than the Poisson, because it allows the variance to be different from the mean.

The Gaussian model requires estimation of the mean and covariance of the PMT values at each grid point. This uses more parameters than the Poisson model, but the increased size of the parameter lookup table is not a problem for modern computers. Of greater concern is the increased uncertainty inherent in jointly estimating more parameters from the calibration data. For a modular camera with T tubes, the Gaussian model requires T parameters to describe the means and an additional $T(T+1)/2$ parameters to describe the covariance matrix. (If we assume the tubes are statistically independent the covariance matrix becomes diagonal and we just need to estimate T variances; see section 2.4).

As in the Poisson case, the ML estimate of a Gaussian's mean is the sample mean (2.6). The ML estimate of the PMT covariance is the sample covariance matrix K, where

$$\hat{K}_{ij} = \frac{1}{N_{\text{list}}} \sum_{l=1}^{N_{\text{list}}} (g_{li} - \overline{g}_{li})(g_{lj} - \overline{g}_{lj}). \tag{2.16}$$

2.3 Nonparametric models

Because our calibration data consists of a large number of samples from the likelihood, it should be possible to somehow directly use the calibration data in reconstruction without assuming a specific parametric family for the likelihood. Statisticians refer to such approaches as *nonparametric estimation* for obvious reasons. The problem of nonparametric likelihood estimation falls within the branch of statistics called density estimation. The next few sections introduce several different nonparametric ways of estimating the likelihood. In all cases we take advantage of our knowledge that, for a fixed $\phi_j(\boldsymbol{r})$, the likelihood $p\big(\boldsymbol{g}_i \big| \text{originated in } \phi_j(\boldsymbol{r})\big)$ is a probability density. Therefore we can estimate the likelihood by estimating these densities at each grid point. Further details on nonparametric density estimation can be found in the readable introduction by Silverman [1986] and the text by Scott [1992]. The theory of density estimation in $\mathbb{L}_1$—as opposed to the more common assumption of $\mathbb{L}_1 \cap \mathbb{L}_2$—is developed by Devroye [1987].

2.3.1 Histograms

The first type of nonparametric density estimate to consider is the histogram. For a rigorous discussion of the theory and a brief history see the text by Scott [1992]. The basic idea of histograming, doubtless already known to the reader, is to partition the data domain into a series of bins and then count how many data samples fall into each bin. If the bin sizes are all equal, the typical case, though perhaps not the

optimal case, then the only free parameters for the modeler to tune are bin size, or equivalently the number of bins, and the center points of each bin. Apart from their conceptual simplicity, histograms are appealing as a density estimator because they are easy to efficiently implement in software through use of a lookup table.

While choosing the number of bins one encounters the familiar resolution versus noise tradeoff. A large number of bins means that sharp changes in the density can be more accurately captured, but also decreases the number of counts in each bin, thereby increasing the variance (noise) of the estimate. For the case of a 1-dimensional histogram, Sturges's rule, derived from error analysis based on a normality assumption, is to divide the sample range into equal sized bins with

$$N_{\text{bins}} = 1 + \log_2 N_{\text{list}}. \tag{2.17}$$

If we have $N_{\text{list}} = 5000$, a typical number of calibration events per camera for one grid point of the MDRF or PSF, then Sturges's rule suggests $N_{\text{bins}} = 13$. But the FastSPECT-II A/D hardware has $2^{12} = 4096$ possible output values per PMT. So histograming this way is a huge reduction in dynamic range! This remains the case even if one uses a different rule of thumb to choose the bins. (Sturges's rule is generally considered to be inappropriate for skewed data. Scott [1992, equation 3.42] suggests using

$$N_{\text{bins}} = \sqrt[3]{2N_{\text{list}}}, \tag{2.18}$$

which works out to 22 bins.)

So far we have only considered univariate histograms. Yet the modular gamma cameras have 9 PMTs. If we used only 13 bins for each dimension, we would still have $13^9 \approx 10^{10} \approx 2^{34}$ bins per grid location. Unless we used a sparse representation, storing the histograms for all PSF grid points would put us outside the memory range not only of today's computers but also those in the foreseeable future, even after Moore's Law [Moore, 1965] is taken into account. Even if we used a sparse representation there would be problems. With only c. 5000 samples almost all of the bins would be

empty, leading to an underestimate of the probability of those values and an overestimate for the non-zero bins. The problems with high dimensional histograms (in this case high means bigger than approximately two) make them inappropriate for modeling the number of PMTs found in a modular camera.

When modeling the output of a single PMT, there are two options for handling the gap between the hardware's 12 bit resolution and the much smaller number of bins in a histogram. The first is to use adaptive histograms where the range of the histogram and the bin sizes are adaptively set using the calibration data. Advice on how to best do this is given by Scott [1992].

2.3.1.1 Variance stabilizing transform The second option, which is easier to program and works very well in practice, is to apply a square-root transformation to each tube's measured value

$$g'_t = \text{sign}(g_t)\sqrt{|g_t|}, \tag{2.19}$$

and then use uniform bins to histogram g'_t. This reduces the range of the output from [-2047, 2048] to [c. -45.25, c. 45.26]. If the outputs of the detectors are assumed Poisson distributed then the transformed values will have an approximately constant variance for all values of the mean. (This is a fishy assumption for the modular camera, as Poisson random variables don't go negative, but digitized PMT outputs do, nor do Poisson random variables model the sorts of gain process seen in PMTs. In any case, if you believe a variation on the Poisson assumption applies, you're better off with a parametric model, and you don't need to mess around with histogram or other nonparametric density estimators.) Because the variance of the square root transform output is approximately constant, it is sometimes called a variance-stabilizing transform [Bartlett, 1936; Anscombe, 1948].

In practice, we do see negative output from the FastSPECT-II modular-camera electronics, but the values are only slightly less than zero. The specific approach used in my reconstruction code is to clip all negative values to zero and only transform

positive values, rounding down to the nearest integer:

$$g'_t = \left\lfloor \sqrt{\text{step}(g_t)\,|g_t|} \right\rfloor, \tag{2.20}$$

which results in 46 histogram bins per PMT per grid point. At any given grid location many of the bins contain zero events, but ignoring this sparseness makes the code much simpler.

2.3.2 Kernel density estimation

Suppose we want to estimate a 1-dimensional probability density $p(x)$ without knowing its parametric form, and are given only a set of N calibration samples c_j from $p(x)$. (We will extend the idea to higher dimensions in a bit.) The kernel density estimate $\hat{p}(\boldsymbol{x})$ is

$$\hat{p}(x) = \frac{1}{Nh} \sum_{j=1}^{N} k\left(\frac{x - c_j}{h}\right) \tag{2.21}$$

$$= \frac{1}{N} \sum_{j=1}^{N} k_h\,(x - c_j) \tag{2.22}$$

$$= \frac{1}{N} \sum_{j=1}^{N} \delta\,(x - c_j) \circledast k_h(x), \tag{2.23}$$

where $k(x)$ is the kernel function, h is a tunable width or smoothing parameter, $\delta(x)$ is the Dirac delta, and $\circledast$ indicates convolution. The modeler may choose the kernel function $k(x)$ and the smoothing parameter h.

Notice if $k(x)$ is itself a probability density, then $\hat{p}(x)$ is guaranteed to also be a density. (Aesthetically speaking, it is nice for an estimate of a density to actually be a probability density. However we have already tipped our hand in section 1.9. The quality of a density estimate is intimately related to overall task performance, and if an estimate which is not a true density performs better than one which is, so be it.) Likewise, $\hat{p}(x)$ has the same differentiability properties as $k(x)$. Different choices

for $k(x)$ have been extensively studied; summaries of what is known are given by Silverman [1986] and Scott [1992]. Low-order symmetric polynomials such as

$$k(x) = \frac{3}{4}(1 - x^2)\,\mathrm{rect}(x), \tag{2.24}$$

and Gaussians

$$k(x) = \frac{1}{\sqrt{2\pi h^2}} \exp\left(\frac{-x^2}{2h^2}\right) \tag{2.25}$$

are the most common choices of kernel function.

Surprisingly, error analysis within an $\mathbb{L}_2$ context has shown that the choice of smoothing parameter h is more important than choice of kernel. When using Gaussian kernels, Scott [1992, equation 6.17] suggests choosing the smoothing parameter to be

$$h = \left(\frac{4}{3}\right)^{1/5} \sigma N^{-1/5} \tag{2.26}$$

$$\approx 1.06\hat{\sigma} N^{-1/5}. \tag{2.27}$$

This rule of thumb originates in a Taylor-series analysis of the density estimator's ensemble mean square error.

Of course, our calibration data are multivariate, so (2.21) doesn't apply directly. The obvious extension is to use a multivariate kernel,

$$\hat{p}(\boldsymbol{x}) = \frac{1}{Nh} \sum_{j=1}^{N} K\left(\frac{\boldsymbol{x} - \boldsymbol{c}_j}{h}\right). \tag{2.28}$$

The easiest way to create a multivariate kernel is to use a product kernel:

$$K(\boldsymbol{x}) = \prod_{t=1}^{T} k(x_t - c_{j_t}), \tag{2.29}$$

where the t subscript indicates component (for example, the PMT), and the j subscript indicates calibration list entry.

Amazingly, very many density estimators, including histograms and even parametric families, can be shoe-horned into a generalization of the kernel-density framework. For example, suppose we wish to parametrically estimate a Gaussian density

with known variance $\sigma^2 = 1$. The parameter of interest is the sample mean $\overline{x}$. Let us pick as the kernel function

$$k(x, x_i) = \frac{1 + (x_i - \overline{x})(x - \overline{x})}{\sqrt{2\pi}} e^{-(x-\overline{x})^2/2}. \tag{2.30}$$

Then the density estimate is

$$\hat{p}(x) = \frac{1}{N} \sum_{i=1}^{N} k(x, x_i) \tag{2.31}$$

$$= \frac{1}{N} \sum_{i=1}^{N} \frac{1 + (x_i - \overline{x})(x - \overline{x})}{\sqrt{2\pi}} e^{-(x-\overline{x})^2/2} \tag{2.32}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-(x-\overline{x})^2/2}. \tag{2.33}$$

Scott [1992] gives precise conditions for which types of density estimator can be written in this way, and also explains under which circumstances a density estimator is nonparametric. Because of this result, error analysis developed in the context of kernel estimators can often be applied to other density estimators.

2.3.3 The curse of dimensionality

Kernel density estimators have not successfully been used in situations where $d > 5$. In low dimensions the tails of a probability distribution only contain a small percent of the cumulative density, whereas in high dimensions the majority of mass is in the tails. Thus in low dimensions ignoring the tails doesn't significantly affect the accuracy of the density estimator. But in high dimensions one cannot ignore the tail. In practice, we cannot measure enough calibration samples to fill the probability space, which prevents the kernel density estimator from doing a good job. Even if one somehow did measure enough samples, practical computational issues would become a problem.

This is an example of the phenomenon known as the "curse of dimensionality". The phrase was coined by Bellman [1962] to describe how certain problems become

exponentially more difficult with increasing dimension. A more recent discussion of the curse is given by Donoho [2000], who also looks at classes of problem that become easier as the dimension increases. Scott [1992] devotes several chapters to the curse and its impact on density estimation.

A good way to develop intuition about the curse is to consider the relationship between the d-dimensional unit hypercube and its inscribed hypersphere. In one dimension, the two are both the unit line segment, and therefore both have length 1. In two dimensions, the unit square has area one and the inscribed circle has area $\pi/4 \approx 0.79$. In three dimensions the unit cube has volume one and the unit sphere has volume $\pi/6 \approx 0.52$. In d-dimensions the unit hypercube has hypervolume 1, and the inscribed hypersphere has hypervolume $\frac{2^{-d}\pi^{d/2}}{\Gamma(\frac{d}{2}+1)}$, where $\Gamma()$ is Euler's gamma function. As the number of dimensions increases, more and more of the volume is concentrated in the corners of the unit cube. Similarly, in a many-dimensional Gaussian much of the mass is contained in the tails of the density. Already in 9 dimensions approximately half the mass of a Gaussian is found in the tails, see Scott [1992] for the derivation.

2.3.4 Density estimation using the k-nearest-neighbor estimator

The kernel estimator is not the only type of nonparametric density estimator in widespread use. Another class of estimator, the k-nearest-neighbor (k-NN) estimator, has been successful, especially in high-dimensional classification and pattern-recognition problems. (Note that position estimation of a scintillation event is often done by assigning the event to a grid point, and therefore can be thought of as a classification problem.) Although still subject to the curse of dimensionality, the k-NN technique has a much better track record than kernel density estimators when the number of dimensions is greater than four or five. The k-NN technique was introduced by Loftsgaarden and Quesenberry [1965] and is discussed in depth by Silverman [1986].

The technique is simple; the probability density is estimated using distance from the sample points:

$$\hat{p}(\boldsymbol{x}) = \frac{1}{\text{dist}(\boldsymbol{x}, \boldsymbol{x}_k)}, \tag{2.34}$$

where dist is the Euclidean distance and $\boldsymbol{x}_k$ is the k-th nearest neighbor to $\boldsymbol{x}$. The parameter k controls smoothing; its role is analogous to that of h in kernel density estimation. Silverman [1986] gives the rule of thumb

$$k = n^{4/(d+4)}, \tag{2.35}$$

where n is the number of calibration samples and d is the number of dimensions. For the modular gamma camera we have $n \approx 5000$ and $d = 9$, so this rule suggests we should use $k = 14$. The easiest (though not quickest) way to implement k-NN is to compute the distance from $\boldsymbol{x}$ to all calibration samples $\boldsymbol{x}_i$, sort, and then pick the kth value. It is important to note that $\hat{p}(\boldsymbol{x})$ *is not* a true density estimate, in particular it is not properly normalized, for $\int_\infty \hat{p}(\boldsymbol{x})\, d\boldsymbol{x} = \infty$.

2.3.4.1 Computational issues with k-NN estimation There are several ways to speed up k-NN density estimation over the obvious naive implementation. The simplest speedup is to avoid sorting the distances. A fast sort algorithm such as Quicksort has on average $O(N \log N)$ time complexity. But if one only wants to select the kth largest value while ignoring the other values, the time can be reduced to $O(N)$ on average by using a variation on Quicksort. For the k-select algorithm pseudocode and time-complexity analysis see Cormen et al. [1990]. My experience with this trick is that it speeds up k-NN position estimation by about 50% over sorting.

Another approach avoids computing all the distances by using a branch-and-bound algorithm, as described by Fukunaga and Narendra [1975]. The idea is to build a tree structure out of the calibration samples, where each node of the tree stores a centroid and radius of all the points below it. Reports in the literature indicate a speed increase over Quicksort of between 50% and 90% using this algorithm. I didn't

bother to implement it because, as will be discussed in section 2.5, we would have needed to reduce the computation time by at least 99.99% for k-NN to be of practical use in image reconstruction; furthermore the much faster independent histograms model introduced in section 2.4.1 gives useful results.

A third option, really a simplification of the branch-and-bound approach, is to set the density estimate to zero when $\boldsymbol{x}$ is far away from the cloud of calibration points. For example, one can compute the distance from $\boldsymbol{x}$ to the centroid of the calibration cloud, and set $p(\boldsymbol{x}) = 0$ if the distance is greater than w times the radius of the cloud. I implemented this for non-parametric maximum-likelihood position estimation of scintillation events using MDRF data and was disappointed with the speedup. In order to avoid changing the ML position estimate, w had to be set large enough that the speedup was less than 10%.

2.4 Independence assumptions

We have already noted that each list entry is statistically independent from the other list entries. It is often reasonable to assume that within a single list entry the different components are also statistically independent. We have already seen an example of this in section 2.2.1with the Poisson assumption for detector output. A related example can be found in the modular gamma cameras, where the nature of gain in the PMT-amplification process dictates that tube outputs for a photo-peak event should be independent. (Since not all events are photo-peak, a two-dimensional scattergram of the PMTs shows correlation. See figure 2.4.)

Even if the components of a list entry are not truly independent, well, we can always just pretend they are and charge ahead, hoping things will turn out fine. The independence assumption

$$p\left(\boldsymbol{g}|\phi_j(\boldsymbol{r})\right) = \prod_t p\left(g_t|\phi_j(\boldsymbol{r})\right) \tag{2.36}$$

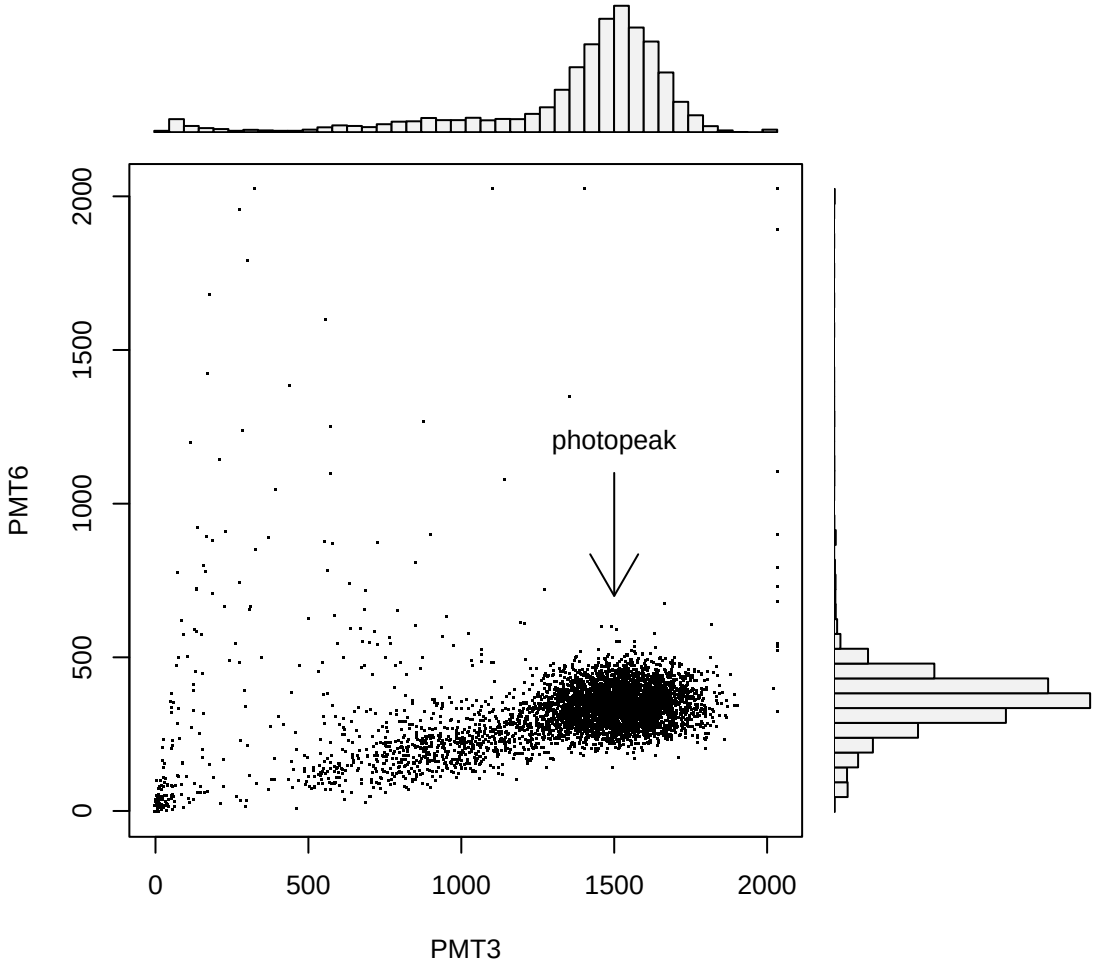

FIGURE 2.4. **Scattergram and marginal histograms of MDRF data for two neighboring PMTs.** The MDRF source was located at grid point (26,26) near the center of tube 3. Photopeak events appear uncorrelated, but the tail of non-photopeak events shows correlation. Histograms of the same data for all tubes are shown in figure 2.2.

allows us to model each component of the data separately. Although this assumption can be, and often is, used with parametric models, it is especially useful for nonparametric approaches, as it allows us to avoid the curse of dimensionality by using a histogram or other one-dimensional density estimator for each tube.

2.4.1 Likelihood model for modular gamma cameras using independent histograms

I have had success coupling the independence assumption (2.36) with the square-root transform (2.20) and then using histograms to model the likelihood of PMT output. (Here success means images reconstructed using this model look good and are similar to those produced using other models. No task-based study was done to verify this subjective impression.) I have used this approach calibrated with MDRF data for position estimation from a single event and fluence estimation from a list of events. I have also used it calibrated with PSF data for tomographic reconstruction directly from list-mode PMT data. With these assumptions the nonparametric PSF likelihood model for all 16 modular cameras in FastSPECT-II requires only 871 MiB,[3] and therefore fits comfortably into the limitations of a 32-bit address space. Using these assumptions, computing the probability a list event originated in region $\phi_j(\boldsymbol{r})$ is quick and easy to implement, as it merely involves looking up each tube value in the histogram and then multiplying the results.

It would also be possible to use a covariance-diagonalizing transform such as principal components analysis (PCA) or independent components analysis (ICA) [Comon, 1994] to preprocess the PSF data. Since the data being reconstructed would also need the transform applied, the coordinated change must be stored, which would add at least 210 MiB to the PSF likelihood model. I did not pursue this idea in part because the independence assumption was working well without any decorrelation, and in

[3]The lab purchased its first server with more than 3 GiB of RAM shortly before I finished. The majority of our compute nodes had only 512 MiB, with some having up to 2 GiB.

part because storing the additional information would force the reconstruction program to use more than the 1 GiB of RAM found in many of our compute nodes, resulting in a significant increase in how long it would take to process a data list. Because photopeak events already produce uncorrelated PMT data (see figure 2.4), a decorrelating step would focus on the tail, and therefore might actually degrade the likelihood estimate for photopeak events.)

2.5 Double list-mode reconstruction

A drawback of parametric likelihood models is that they are different for each type of imaging system. Thus a modeling a PET system requires a different parametric family than one would use to model a SPECT system. Furthermore, choosing the right family often requires considerable understanding of the system's measurement physics; poor approximations and invalid assumptions made during modeling will introduce errors and artifacts in the reconstructed images. The big appeal of nonparametric models is that they allow us to avoid explicitly modeling the system. Ideally we would be able to use only the calibration data to formulate the likelihood model, without imposing any other prior knowledge on the model. Such an approach can be useful even with systems for which we already have parametric models, as we may be able to use the nonparametric model to validate the assumptions and approximations behind the parametric model.

Using either a kernel density estimator or a k-nearest-neighbor estimator, it is possible to formulate the likelihood model directly from list-mode calibration data without imposing any assumptions about the system. Because this sort of nonparametric likelihood model involves comparing each data list entry with each calibration list entry, we refer to this approach as *double list-mode reconstruction*.

I have implemented a version of double-list-mode reconstruction by using the k-NN estimator calibrated with MDRF data to produce maximum-likelihood position

estimates. After position estimation, the reconstructed projection image is produced by histograming the position estimates.

A likelihood plot for the position of one list entry is shown in figure 2.5. Notice that although the likelihood function is nicely peaked, the tails are quite noisy, and even the peak is bumpy. This means non-brute-force search algorithms such as steepest-ascent or conjugate-gradient will have trouble finding the max unless started quite close to the correct location. In practice, something similar to exhaustive search is required. (Using an external model derived from prior knowledge about the camera to choose the search starting point would not be in the spirit of non-parametric estimation. Once one builds in assumptions one is no longer doing pure double list-mode reconstruction.)

Several other points are worth making about my implementation. First, double-list-mode works—it is possible to do image reconstruction without any model of the imaging system other than extensive list-mode calibration data.

Second, double list-mode is very slow. The k-NN approach requires 7.5 seconds *per data list entry* to do ML position estimation. (Average of 1000 events. Timing was done on host gamma-33, which has a 2.1 GHz AMD Athlon MP2600+, the fastest CPU type currently in our compute nodes. Most of our CPUs are slower.) Given that a typical data set for animal imaging has on the order of one-hundred thousand events per camera, using double list-mode would require 208 hours of CPU time just to reconstruct one projection image. Of course FastSPECT-II has sixteen cameras, and each imaging session usually produces several data sets. Even with the parallel-processing resources available, this is too slow for any practical use. To make k-NN practical we would need speedups of several orders of magnitude, far more than any of the available algorithm tricks would allow.

Third, double list-mode, or any other non-parametric technique, will have less statistical power than an accurate parametric model. The noisy tails in figure 2.5 are a symptom of this.

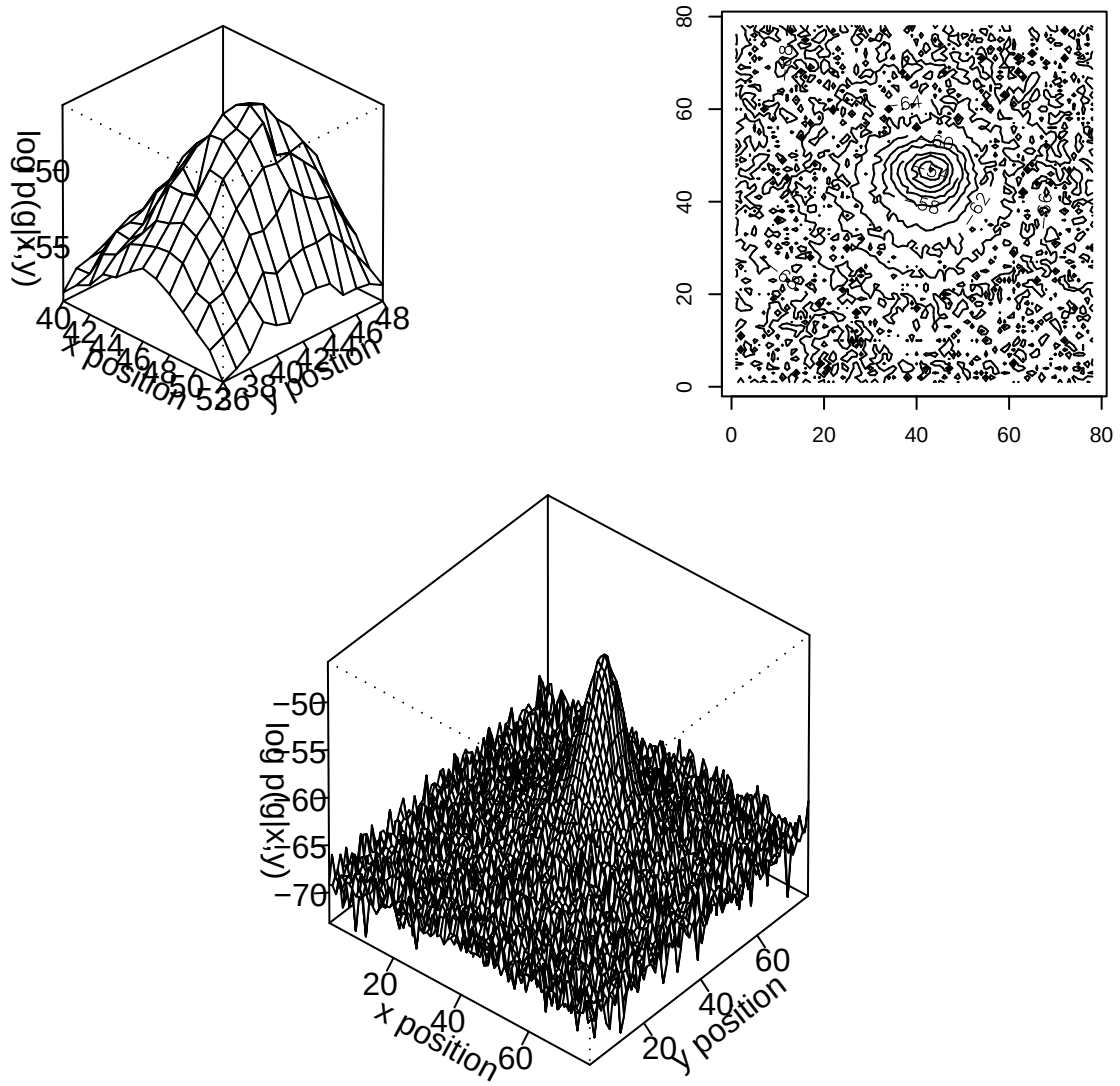

FIGURE 2.5. **Nonparametric log-likelihood plot for one event.** The log-likelihood was computed using the double-list-mode (k-NN) technique. On the upper left is a close-up of the maximum-likelihood location, and on the upper right is a contour plot. The maximum-likelihood position is (42, 46) with a log-likelihood of -45.78.

Fourth, double list-mode reconstruction only takes account of properties present in the data. The calibration techniques described in sections 2.1.2 and 2.1.3 do not measure the effects of attenuation and scatter, both of which may be easier to incorporate into a parametric model.

Given these practical drawbacks, at this point double list-mode reconstruction seems to be a solution in search of a problem. The ideal imaging system for double list-mode would be difficult or impossible to model parametrically, would have extensive list-mode calibration data available, and yet would produce only a limited number of list entries when imaging an object.

Chapter 3

Image-quality comparison of two reconstruction algorithms

This chapter compares two different algorithms for producing projection images from list-mode modular-gamma-camera data. Both algorithms start with a list of PMT readings and use a likelihood model to output a reconstructed planar projection. The image quality of each algorithm, more precisely the quality of the images it reconstructs, is measured using an anthropomorphic channelized-Hotelling observer (CHO) as a proxy for human observers. The Rayleigh task, a traditional measure of image system resolution, is used for quality comparison of the two algorithms.

Although the raw PMT data from the modular camera have 9 dimensions, we usually produce images from a reduced-dimensionality list consisting of 2-dimensional position estimates for reasons of computational efficiency. A similar dimensionality reduction is found in almost all nuclear-medicine detectors. An important question is: do we lose any information by working with the reduced dimension data, and if so does the loss have practical implications? The simulations in this chapter partially answer this question by doing observer studies using both types of list.

The first algorithm of interest is a conventional two-step process. It begins by processing all list entries, using maximum likelihood (ML) to estimate the position of each scintillation event. The likelihood model used for this study is the independence model introduced in section 2.4.1, calibrated using mean detector response function (MDRF) data. After position estimation, the locations are binned into a two-dimensional histogram, which is displayed to the observer as the reconstructed projection image. (Because each event is assigned to its ML position estimate, this method is often called ML reconstruction. However, as discussed in section 1.7.4,

this approach *does not* produce an ML estimate of the projection-image $\boldsymbol{f}$, except in the limiting case of no detector blur.) In the position estimation step each individual event is assigned to exactly one region on the camera face. Variations on this approach are the standard methods of producing planar-projection images from the modular gamma cameras. We refer to this technique as histogram reconstruction.

The other algorithm is the list-mode expectation-maximization (EM) algorithm described in section 1.7.2. EM uses the same likelihood model as the position-estimation approach, but is able to partially attribute each event to multiple regions on the camera face. We shall refer to the EM method of reconstruction as fluence estimation.

Marcotte [1993] was the first to use EM to produce projection images from modular camera data. Her primary focus was post-processing histogram reconstructions with binned-mode EM, thereby removing blur introduced by incorrectly estimating the positions of events. She also considered using binned-mode EM to process histograms of raw PMT output, but did not pursue the idea because of computational difficulties.

3.1 Subjective image-quality comparison

The first thing to do when assessing the quality of a reconstruction algorithm is to eyeball some images it has produced, preferably images of known objects. Such a subjective image-quality comparison is hardly rigorous, yet may suggest differences we can verify using task-based metrics. Figure 3.1 shows a planar-projection mouse bone-scan reconstructed using both algorithms.[1] The EM image appears to have slightly higher resolution, most noticeable in the skull region and in the shoulder.

This is suggestive, but hardly conclusive. Further examples are needed. Figure 3.2 shows reconstructions of a shadow mask image, again using both algorithms.[2] The EM algorithm is iterative, so results after various numbers of iterations are shown.

[1] Mouse data provided by Zhonglin Liu.

[2] Shadow-mask data provided by Jean Chen.

The mask consists of a lead plate, into which holes have been drilled to form the image. The mask is placed flush against the camera face, and is illuminated by a point source located several meters from the camera. Notice that at 100 iterations it appears as though some individual holes in the mask are visible. Running further iterations doesn't change the image; the algorithm has essentially converged by this point.

3.1.1 Pseudo point-spread function

Because the point spread function (PSF) of a linear imaging system contains all information about the system resolution, it has become traditional to include images of a point source in any discussion about resolution. Even though the modular cameras have nonlinear response, and both reconstruction algorithms under consideration are nonlinear, looking at a point source may still be useful. Figure 3.3 shows images of a point source reconstructed using both algorithms. The PMT data was produced by aiming a pencil-collimated beam at one spot on the camera face. (The collimator was the same one used for MDRF calibration. To produce these point-source reconstructions a special high-resolution 0.5 mm $\times$ 0.5 mm MDRF was used to calibrate the likelihood model.) More strongly than the mouse and shadow-mask images shown above, the point images suggest that the EM fluence-estimation approach is producing higher-resolution images than the histogram of position estimates.

It is important to note that *the images in figure 3.3 are not point spread functions (PSFs) in the usual linear-systems theory sense!* The PSF is an inherently linear concept and may not be relevant for nonlinear imaging systems. It is not correct to think of the modular camera as convolving or blurring an object with these "PSFs". The mapping from scintillation event to PMT value, the ML position estimation, and the EM algorithm are all highly nonlinear, therefore it is likely that moving the source to a different location on the camera face, or changing the number of list

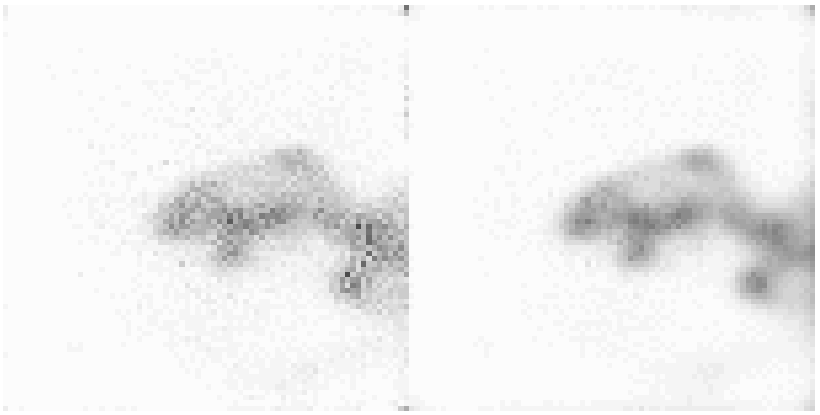

FIGURE 3.1. **Subjective comparison of histogram reconstruction and EM fluence estimation.** The image on the left is the histogram of maximum-likelihood position estimates from a mouse bone scan. On the right is a fluence estimate from the same data set, produced using 2 iterations of the 2-dimensional list-mode EM algorithm.

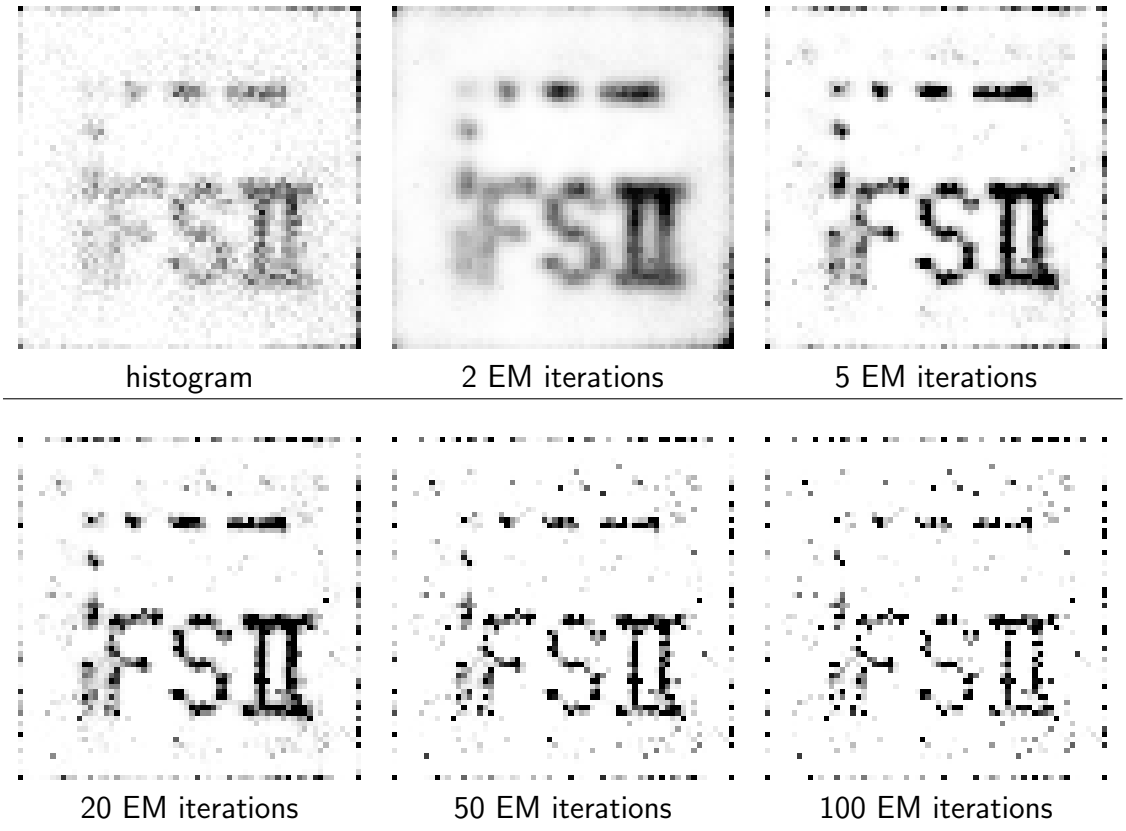

FIGURE 3.2. **Shadow-mask image reconstructed from raw PMT data using the histogram and fluence-estimation approaches.**

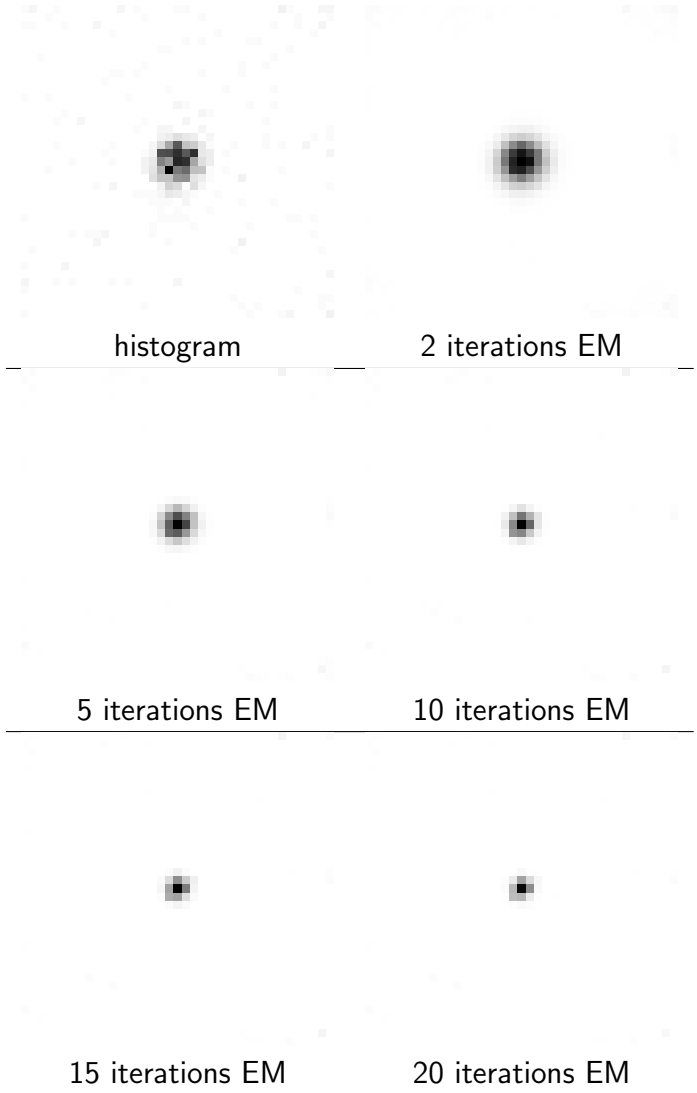

FIGURE 3.3. **Pseudo point-spread function (PSF), created by reconstructing a point source using the histogram and fluence-estimation approaches.** This is not a true PSF, because the imaging system and reconstruction algorithms are nonlinear. Each sub-image has been separately normalized, so gray-scale levels are not comparable. This figure is at higher magnification than the preceding bone-scan and shadow-mask examples.

entries, would result in a different PSF. As a result it is questionable how indicative of system performance these images are. Wilson [1994] showed that when analyzing the EM algorithm the linear approach is appropriate only for small perturbations.

The apparent resolution of the EM approach shown in figure 3.3 is suspicious for two other reasons. First, the background is set to zero. Due to EM's multiplicative update equation (1.32), zero is a fixed-point for each coefficient $\hat{\alpha}_j$, and therefore is the easiest value for EM to estimate.

Second, as discussed in section 1.7.5, the EM algorithm is notorious for producing "night-sky" reconstructions, in which the reconstruction consists of a small number of point objects. How can we be sure that EM converged to a point-like image because it has higher resolution, and not because that is what it always does? For these reasons, reconstruction of a point source is not a fair way to compare EM with other approaches.

3.2 Rayleigh task

As discussed in section 1.9, the best way to compare the resolution of nonlinear reconstruction algorithms is to compare observer performance on a resolution-sensitive task. The task chosen here is a variation of the classical Rayleigh task, in which the observer must decide whether the image contains one or two objects.

3.2.1 Literature review

The task is named after Lord Rayleigh [1879, 1880], who suggested it as the way to determine the resolution of spectroscopes:

> "As the power of a telescope is measured by the closeness of the double stars which it can resolve, so the power of a spectroscope ought to be measured by the closeness of the closest double lines in the spectrum which it is competent to resolve. In this sense it is possible for one instrument

> to be more powerful than a second in one part of the spectrum, while in another part of the spectrum the second instrument is more powerful than the first." [Rayleigh, 1879, p. 269]

Rayleigh was not the first person to think along these lines. He cites an earlier study of double stars by Dawes [1867], and notes that Foucault also studied the question of resolving-power. In fact the idea is much older. Since ancient times a test of eyesight has been to distinguish between Mizar (ζ *Urasae Majoris*, magnitude 2.3) and Alcor (80 *Ursae Majoris*, magnitude 4.0), a pair of stars in the handle of the Big Dipper[3] [Chartrand and Tirion, 1991]. (Because of the large difference in intensity magnitude, distinguishing between the two stars is a test of both resolving power and sensitivity to brightness.)

Harris [1964] studied the Rayleigh task from the viewpoint of statistical-decision theory, calculating how the ideal observer processes detector output. He formulated the problem as an SKE task, and showed that given a sufficient number of photons the ideal observer can distinguish between a single point source and two arbitrarily-close point sources. The ideal observers accuracy increases with the number of photons. The Rayleigh task differs from linear measures of resolution such as the PSF full-width half-maximum (FWHM) by also considering noise level. Rayleigh task performance explicitly depends on the number of detected photons.

The Rayleigh task was introduced into medical imaging as a task for image-quality assessment by Wagner et al. [1981], who suggested it as a way of comparing different coded-aperture designs. Together with their collaborators they published a series of papers on the topic: Hanson and Myers [1991a,b] were the first to use the Rayleigh task to compare reconstruction algorithms and regularization techniques. Later Myers et al. [1993] showed that model-observer performance on the Rayleigh task can be used

[3]Mizar and Alcor are not a binary-star system; they merely appear close together from our vantage point. In 1650 the Jesuit astronomer Giovanni Battista Riccioli discovered Mizar is a multiple-star system, the first found by telescope. Mizar is now considered to be a quadruple star [Encyclopædia Britannica Online].

to predict human performance on images reconstructed by different reconstruction algorithms. Myers et al. [1990] used the Rayleigh task to compare different aperture designs, and showed that precise specification of the task—in particular the choice of background—has a significant impact on performance. Wagner et al. [1996] used the Rayleigh task to illustrate why Bayesian reconstruction approaches should consider knowledge of the task as a form of prior knowledge.

Lehner et al. [2004] were the first, to my knowledge, to use the Rayleigh task to compare two list-mode reconstruction algorithms. They reconstructed the same two-point-source data using two algorithms and noted that only one of the algorithms produced a two-point image. They used only a single data set from one object; no attempt was made to produce ROC curves or vary the background.

3.2.2 Precise description of the Rayleigh tasks

As noted by Rayleigh, the precise task definition affects measured performance ratings:

> "A double *line* is therefore probably more easily resolvable than a double *point*; but the difference is not great." [Rayleigh, 1879, p. 264]

In the version of the Rayleigh task used here, the observer must distinguish between two Gaussian blobs, also called a dumbbell, and a single Gaussian convolved with a 1-dimensional rect function, also called a Gaussian bar. All three Gaussians have standard deviation $\sigma = 2$. All units of width and length are in object pixels, divide by two to get millimeters. The pair of Gaussians are separated by specified length l. The length and amplitude of the Gaussian bar are chosen to minimize the mean-square difference between the bar and the dumbbell, the normalization suggested by Hanson and Myers [1991b]. An example of the two objects is shown in figure 3.4.

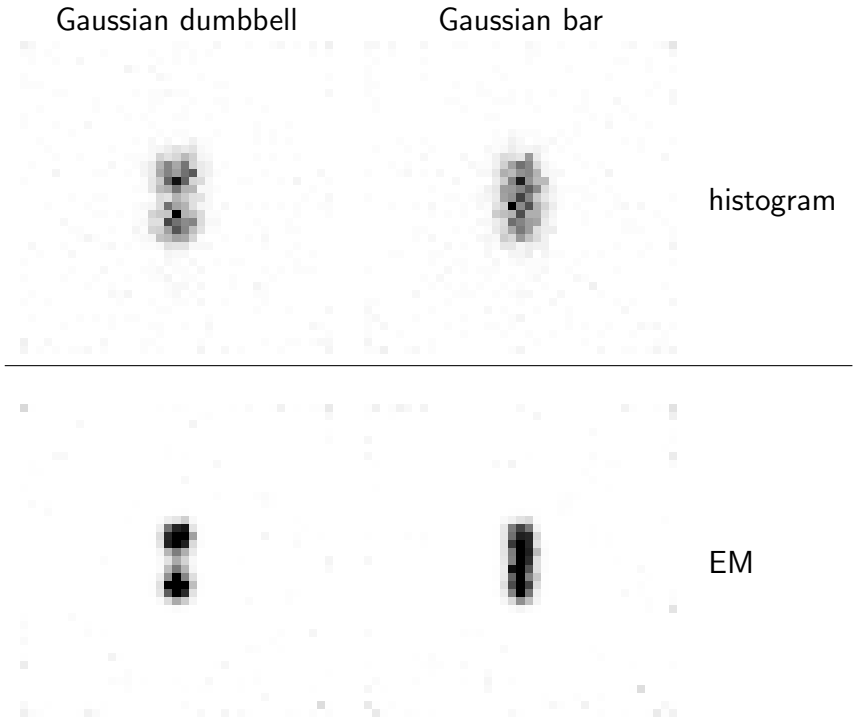

FIGURE 3.4. **Sample signals used in the Rayleigh task.** The left column are a pair of Gaussians (dumbbell) and the right column are Gaussian bars. The top row was reconstructed using the histogram algorithm, the bottom using 10 iterations of the EM algorithm. The observer must distinguish between the dumbbell and bar. These images use a dumbbell length of 17, considerably longer than used in the CHO study. The length was increased for this illustration because the intensity dip at the center of the dumbbell is harder to see on printouts than on the screen.

To determine the mean number of photons emitted at each grid point of the simulated fluence pattern, the analytic expressions for the dumbbell and bar were sampled on a high-resolution (0.5 mm × 0.5 mm) grid and added to a background image. Poisson noise was added about this mean. The required number of list entries were randomly selected, with replacement, from a special high-resolution, high-count (c. 19,000 events at each grid point) MDRF dataset recorded for this study.[4]

The simulated data were then reconstructed using both algorithms to produce projection images. The two algorithms use the same likelihood model, calibrated by an ordinary low-resolution (1.5 mm × 1.5 mm) low-count (c. 5000 events at each grid point) MDRF. Therefore pixelized objects are sampled on a grid that is three-times finer in each dimension (nine times finer in area) than the reconstruction grid of the images displayed to the observer.

Performance was measured by an anthropomorphic channelized Hotelling observer, using the 80 Gabor channels introduced in section 1.11.6.2. No internal noise was used in the CHO. (Internal noise is needed to quantitatively predict human performance, but has not been observed to change the rank order of the algorithms.) Area under the ROC curve was estimated using the percentage of correct identifications in a 2-AFC experiment. Error bars were computed using the techniques described in section 3.3.

Pseudo-random numbers needed for the randomized backgrounds, Poisson noise, and resampling of the high-resolution MDRF were provided by subroutines from version 1.5 of the GNU Scientific Library (GSL[5]) [Galassi et al., 2003]. To ensure processes running on different CPUs had different sequences of random numbers, the MT19937[6] random-number generator was seeded using the Linux `/dev/urandom`

[4] I am grateful to Jean Chen for help measuring these data.

[5] I highly recommend GSL's for those doing scientific programming in C.

[6] The algorithm is named MT19937 because it uses the "Mersenne Twister" technique and has a period of $2^{19937} - 1$ numbers. Ironically MT19937 was developed at Hiroshima University, while GSL was written at Los Alamos National Lab.

entropy pool.

Objects were sampled from ensembles at a variety of lengths for three background types:

Zero background: The simulated object consisted of one of the two signal types (Gaussian dumbbell or Gaussian bar) against a zero background. There are two sources of randomness in the data. The first is the Poisson noise in the number of photons emitted at each grid point of the object. The second is caused by sampling from the high-resolution MDRF to simulate noise in the scintillation crystal and PMT. 1000 pairs of images were created for dumbbell lengths of 6, 9, and 12 pixels, with a mean of 2000 detected photons from the dumbbell. These lengths were chosen to span the interval in which performing the task changes from being difficult to easy for the CHO. 1000 pairs of images at each length were generated using the zero background.

Flat background: The simulated object consisted of one of the two signal types added to a flat background. The background had a mean of five background counts per object pixel (corresponding to a mean of 45 background counts per reconstructed image pixel). These scenes have the same two sources of noise as the zero background case. The effect of the flat background is to raise the variance of the Poisson noise and to make the intensity dip at the center of the dumbbell less obvious. As in the zero background case, 1000 pairs of images were created for dumbbell lengths of 6, 9, and 12 pixels, with a mean of 2000 photons detected photons from the dumbbell. 1000 pairs of images at each length were generated using the flat background.

Lumpy backgrounds: The object consisted of one of the two signals added to a lumpy background [Rolland and Barrett, 1992]. The background consisted of a Poisson number (mean of 100) of symmetric Gaussian lumps (standard deviation

(width) of $w \in \{2, 3, 4\}$ pixels, and maximum amplitude of 40 counts—note that for $w = 2$ the background lump is the same width as the foreground signal) uniformly distributed in a region around the signal. As in the zero and flat background cases, the lumpy background scenes have randomness due to the number of events at each grid point in the simulated object and due to sampling. However the random number and random locations of the lumps introduce a third source of randomness: the background. 2000 pairs of images at each length were generated using each of the three lumpy background widths. Each image had its own randomly-generated lumpy background; backgrounds were independent within each pair and from pair to pair. To produce smaller error bars, more pairs were generated than for the other two backgrounds.

3.3 CHO error bars

To conclude that one algorithm is better than another it is not sufficient to compute their respective SNRs or AUCs and see which is bigger. We must also know their respective error bars in order to determine if the difference is statistically significant. For our version of the Rayleigh task we don't know analytic expressions for the CHO or its corresponding SNR/AUC. Therefore we must estimate the CHO template from training data and then estimate the AUC from testing data, which means the measured SNR or AUC is a random variable. This section explains how to compute error bars on the AUC.

There are two ways in which we can proceed: either assume a parametric distribution for the SNR, or we use a nonparametric approach. I used the nonparametric approach described below because I didn't learn that the distribution of the estimated Hotelling SNR was known until after completing the analysis. The two techniques will usually give similar answers, so there is no reason to prefer the more compute-intensive nonparametric approach.

The parametric approach is briefly mentioned for completeness. Recall from chapter 1 that the CHO has

$$\mathrm{SNR}^2 = \Delta\overline{\boldsymbol{x}}^t \boldsymbol{K}^{-1} \Delta\overline{\boldsymbol{x}}. \tag{1.83}$$

Because the channel outputs are the linear combination of many pixels, by the Central Limit Theorem it is reasonable to assume that $\Delta\overline{\boldsymbol{x}}$ is Gaussian distributed. It is also reasonable to assume that the covariance $\boldsymbol{K}$ is Wishart distributed. Therefore SNR^2 is a random variable with Hotelling's T^2 distribution, which makes it easy to compute a classical confidence interval [Johnson and Wichern, 1982].

Note that the T^2 distribution may not be appropriate for the naive list-mode Hotelling observer of section 1.11.5, as the raw PMT outputs are not Gaussian distributed, and the Central Limit Theorem does not apply.

3.3.1 Nonparametric error bars: bootstrapping and shuffling

For the purposes of an example, let us assume we have a thousand pairs of images available, such that exactly one image in each pair corresponds to each of the two possibilities in the Rayleigh task, i.e. that one image in each pair is a Gaussian dumbbell while the other is a Gaussian bar. Further, let us assume that all the pairs are statistically independent. If we don't care about error bars, computing the CHO is easy—simply train it on all thousand images. (By train we mean use the available images to estimate the covariance and mean of the channel outputs, K and $\Delta\overline{\boldsymbol{x}}$ in (1.79).)

We could also use all 1000 images to estimate the AUC, but this would result in an overestimate, as the testing data wouldn't be independent of the training data. Better would be to partition the set of images into two halves, using the first as trainers to estimate the CHO and SNR, and using the second half to as testers to estimate the AUC.

The preferred way to compute error bars is to create many (say several hundred)

other sets of 1000 images and use those to study the variance in the CHO's performance. Unfortunately use of this gold standard is not computationally feasible for our problem.

Another alternative is to partition the full set of a thousand images into five smaller sets of 200 images. From each of these subsets we could pick 100 images as trainers and use the other 100 as testers. This would give us five estimates of the AUC, from which we could compute a mean and variance. Because the pairs are independent, there is nothing special about the way we partitioned the original images. A different partitioning, say subsets of 100 images would have worked just as well.

This suggests using a class of statistical procedures known as resampling. By creating many subsets from our original set, we can estimate the variability which would have occurred using the preferred approach of creating many sets of 1000 images. The most famous resampling approach is called the bootstrap. (Another, not used in this dissertation, is the jacknife.) Readers desiring a gentle introduction to resampling in general, and bootstrapping in particular, are referred to Diaconis and Efron [1983] and Efron and Tibshirani [1991]. The basic idea is simple: create a resample by picking random elements from the original sample. Bootstrapping is resampling with replacement, which means that some elements will appear in the resample more than once, and others will not appear at all.

The best way to to compute CHO error bars by resampling from a limited amount of training and testing data has been studied by Gallas [2003]. He considered two possible ways of resampling the data and compared them with the gold standard. The first, which he calls the shuffle technique, places the pairs in a random order, much the way one shuffles a deck of cards. The first half of the shuffled set is used to train the CHO, and the second half to estimate the AUC. This is repeated many times. The shuffle mean gives an unbiased estimate of the AUC, but the shuffle's variance is biased low as an estimate of the error bar.

The second technique studied by Gallas is the usual bootstrap. Unlike the shuffle technique, where each image occurs exactly once, sometimes in the training set and other times in the testing set, with the bootstrap some images will not be present in the resample, and others will be represented more than once. Thus the bootstrap may include some images in both the training set and the testing set, which yields more variability in AUC than the shuffle. As with the shuffle, the first half of the resample is used to train the CHO, the second half to estimate AUC, and the whole procedure is repeated many times. The bootstrap gives an unbiased estimate of the variance of the AUC, but a biased-high estimate of the AUC itself.

Based on Gallas's results, the channelized Hotelling observer performance results presented in section 3.4.2 use the shuffle mean to estimate the AUC and the bootstrap variance to estimate the AUC's standard error.

3.4 Hotelling observer results for the Rayleigh task

3.4.1 Naive list-mode-Hoteling observer

After generating the simulated data sets, all lists were processed to produce maximum-likelihood positions estimates. The naive list-mode Hotelling observer described in section 1.11.5 was computed for both types of list. To do this the short lists corresponding to each scene were concatenated together into one long list which was then used to train the Hotelling observer. SNR was computed using (1.85), with error bars provided by bootstrapping.

The results are shown in figure 3.5 for the zero background, figure 3.6 for the flat background, and figure 3.7 for the lumpy background. The SNR numbers are so low because they correspond to making a decision using only one list entry. (By (1.88) the SNR scales as $\sqrt{N_{\text{list}}}$.) The Hotelling SNR (1.83) was used instead of AUC because the values are so low.

For all three background types the position estimates yielded a lower SNR than

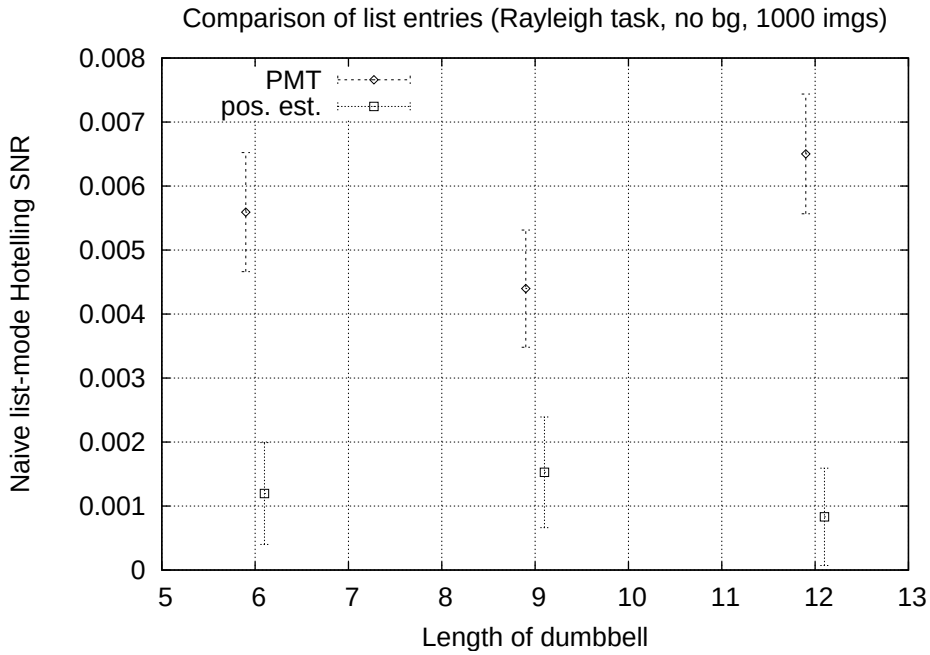

FIGURE 3.5. **SNR of the naive list-mode Hotelling observer performing the Rayleigh task against a zero background.** Plotted are the observer acting on the raw 9-dimensional PMT data and on the 2-dimensional ML position estimates. These are the SNRs for a single event, multiply by $\sqrt{N}$ to get the SNR for N events. The error bars have been horizontally displaced; the dumbbell lengths are exactly 6, 9, and 12 for both cases.

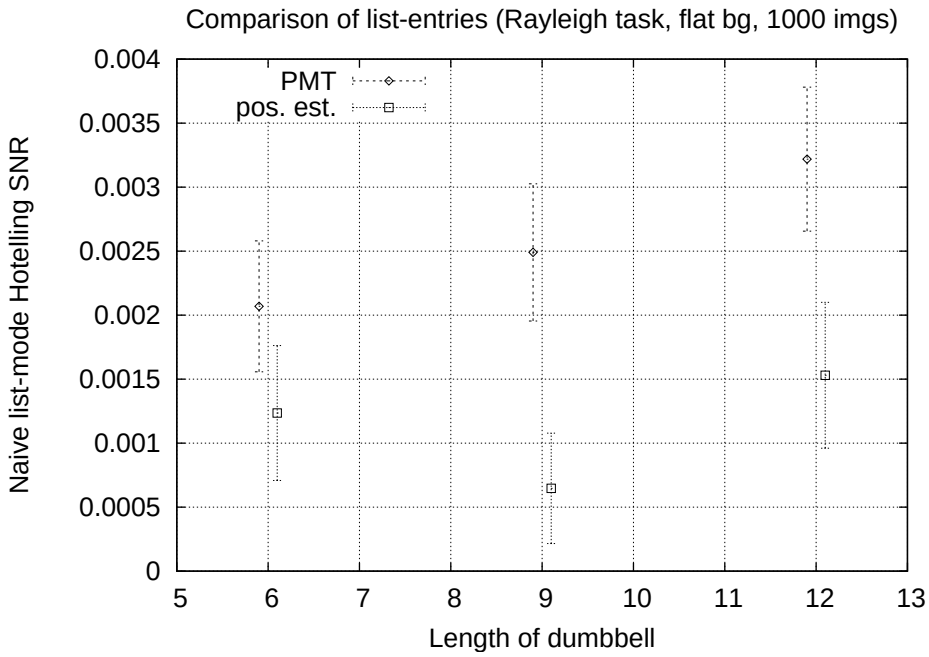

FIGURE 3.6. **SNR of the naive list-mode Hotelling observer performing the Rayleigh against a flat background.** Plotted are the observer acting on the raw 9-dimensional PMT data and on the 2-dimensional ML position estimates. These are the SNRs for a single event, multiply by $\sqrt{N}$ to get the SNR for N events. The error bars have been horizontally displaced; the dumbbell lengths are exactly 6, 9, and 12 for both cases.

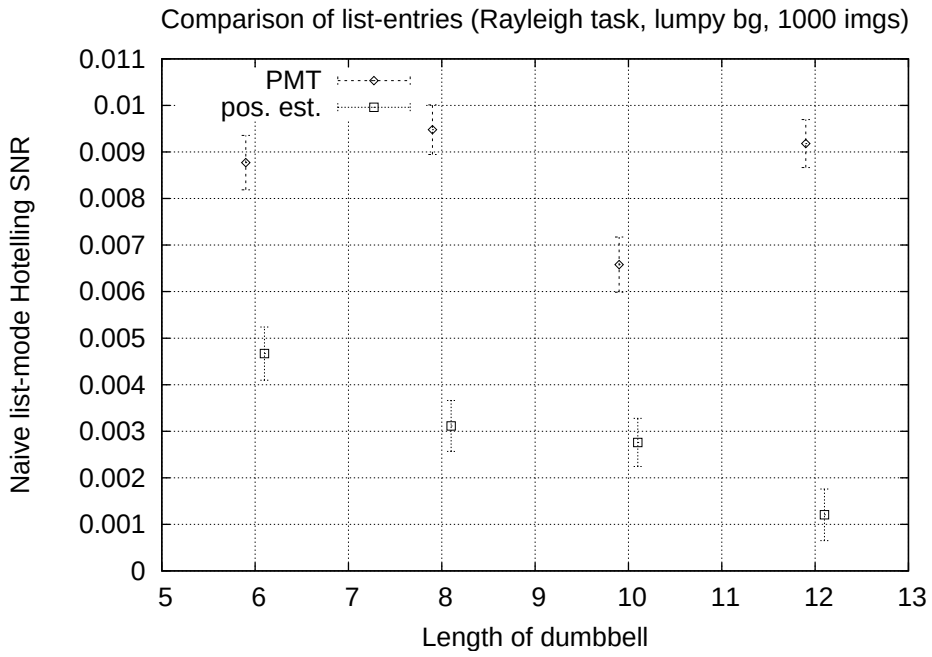

FIGURE 3.7. **SNR of the naive list-mode Hotelling observer performing the Rayleigh against a lumpy background,** $w = 2$**.** Plotted are the observer acting on the raw 9-dimensional PMT data and on the 2-dimensional ML position estimates. These are the SNRs for a single event, multiply by $\sqrt{N}$ to get the SNR for N events. The error bars have been horizontally displaced; the dumbbell lengths are exactly 6, 8, 10, and 12 for both cases. Note that the SNR is *decreasing* with length for unknown reasons.

the raw PMT values. Note that the naive list-mode Hotelling observer uses a linear decision strategy which weights each entry independently of the others. The ideal observer and the image reconstruction algorithms used to display the data to humans are nonlinear. These numbers give a lower bound on performance of the ideal list-mode observer.

3.4.2 Channelized-Hotelling observer

The data were also reconstructed into images using both algorithms. Thus each dumbbell or bar image was reconstructed twice. The images were then used to train the CHO and estimate its AUC.

For all background types the SNR of the EM algorithm reached a plateau after only a few iterations and then remained relatively constant for many iterations before beginning to decrease. (Abbey [1998] observed the same phenomenon while using the binned-mode EM algorithm to produce tomographic reconstructions.) In all cases the plateau had been reached by ten iterations.

Figure 3.8 shows the area under the channelized Hotelling observer's ROC curve for each algorithm. Here the CHO is performing the Rayleigh task against a zero background. At lengths 6 and 9 the fluence-estimation approach clearly produces higher-resolution reconstructions. At length 12 no difference in performance is seen because the task has become so easy for the CHO to perform that both algorithms get a nearly perfect score.

Figure 3.9 shows results for the flat background. As in the zero background case, the fluence estimation approach produces higher-resolution images, though the effects are much less pronounced. Note the considerable overlap in error bars at length 6. At length 9 there is also a slight overlap of the error bar.

Finally, figures 3.10, 3.11, and 3.12 compares both algorithm using lumpy backgrounds. For $w = 2$, where the lumps are the same width as the signal, statistically

significant difference between the two algorithms exists, as shown by the considerable overlap of error bars at all lengths. Note that at length 8 the EM algorithm has slightly lower AUC than the histogram algorithm. For $w = 3$ and $w = 4$ the background lump width is larger than the signal width, and a EM has better resolution at length 10.

3.5 Rayleigh task conclusions

The simulations in this chapter indicate that some information is lost during the dimensionality reduction from the 9-dimensional PMT vector to the 2-dimensional position estimate. Obvious candidates for the lost information are the amount of energy deposited in the crystal during the scintillation event and the depth of interaction. However the MDRF measurements used to collect data for the simulation are designed to minimize the impact of those two parameters.

The amount of energy deposited in the crystal depends to a large extent on the energy of the incoming gamma ray. In human imaging the photon may undergo Compton scatter within the subject, reducing the energy of the incoming photon. The MDRF source is not embedded in tissue, so Compton scatter is greatly reduced.

Estimating depth of interaction is most important in situations where one has an oblique ray interacting with a thick crystal, a common situation in PET imaging. The modular cameras have relatively thin crystals, so depth of interaction is not an important effect. Furthermore, the MDRF calibration beam is always perpendicular to the camera face.

Other candidates for the lost information are sub-pixel resolution and uncertainty about the true event location. The ML position-estimation technique used here assigns each event to a single grid point, throwing away any sub-grid information present in the raw data. EM, which is able to assign each event to more than one grid location, can better handle sub-pixel information and uncertainty about the true location.

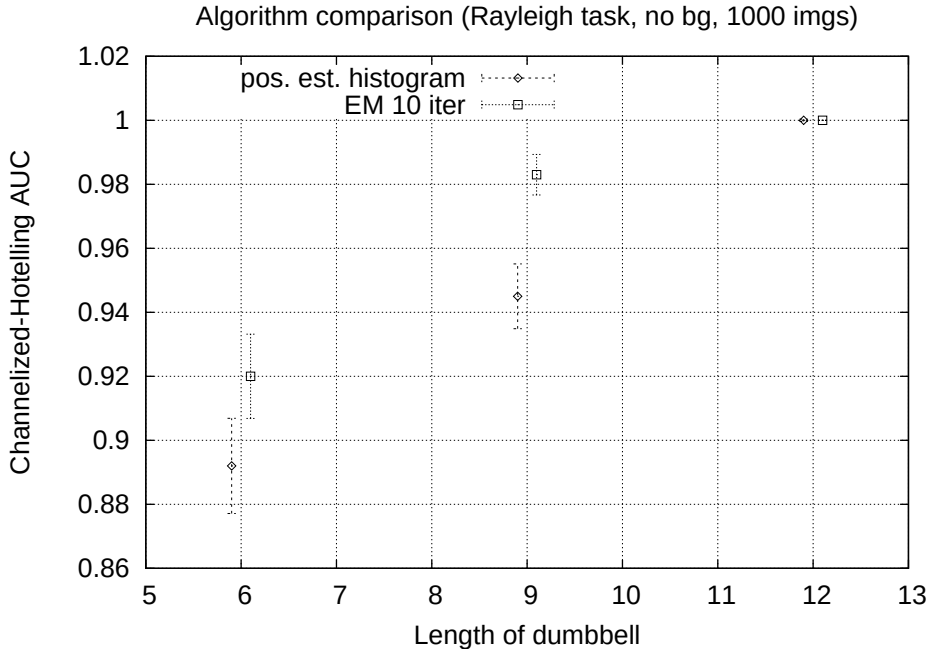

FIGURE 3.8. **Area under the ROC curve for the Rayleigh-task CHO with zero background.** Shown are results for the CHO acting on images reconstructed using both algorithms for a variety of lengths. The error bars have been horizontally displaced; the dumbbell lengths are exactly 6, 9, and 12 for both algorithms. The EM algorithm clearly produces higher-resolution reconstructions at lengths 6 and 9.

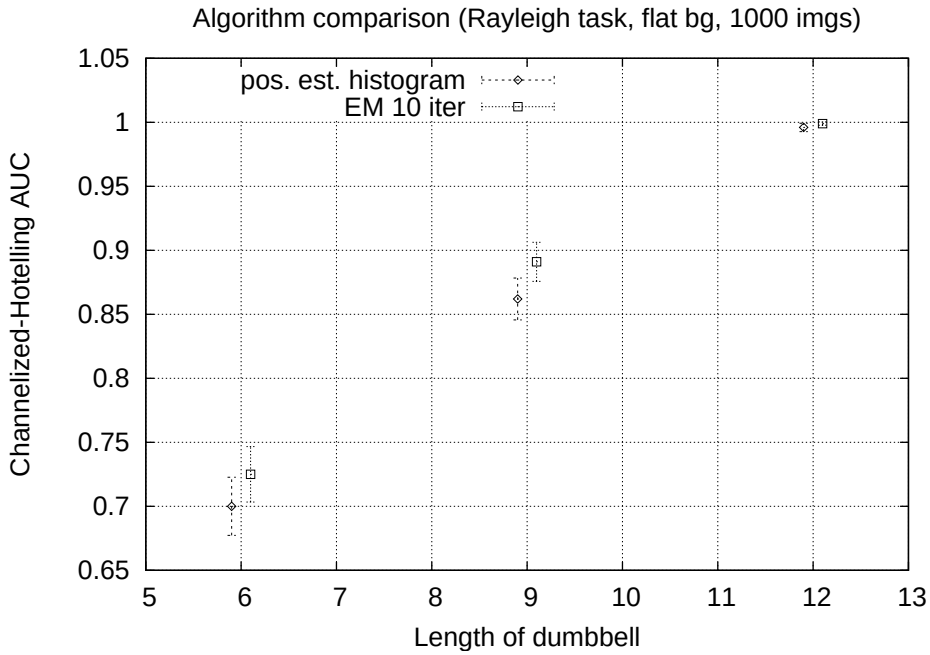

FIGURE 3.9. **Area under the ROC curve for the Rayleigh-task CHO with a flat background.** Shown are results for the CHO acting on images reconstructed using both algorithms for a variety of lengths. The error bars have been horizontally displaced; the dumbbell lengths are exactly 6, 9, and 12 for both algorithms. At length 9 the EM produces higher-resolution reconstructions.

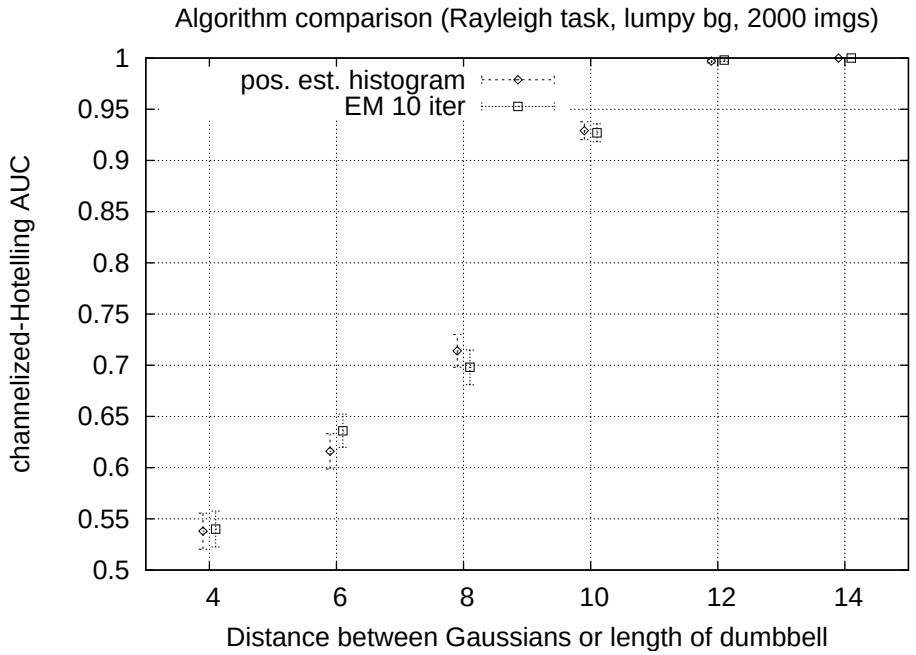

FIGURE 3.10. **Area under the ROC curve for the Rayleigh-task CHO with a lumpy background,** $w = 2$**.** Background lump width is the same as the signal. Shown are results for the CHO acting on images reconstructed using both algorithms for a variety of lengths. The error bars have been horizontally displaced; the dumbbell lengths are exactly 4, 6, 8, 10, 12, and 14 for both algorithms. Note that the rank-order of the algorithms changes for length 8.

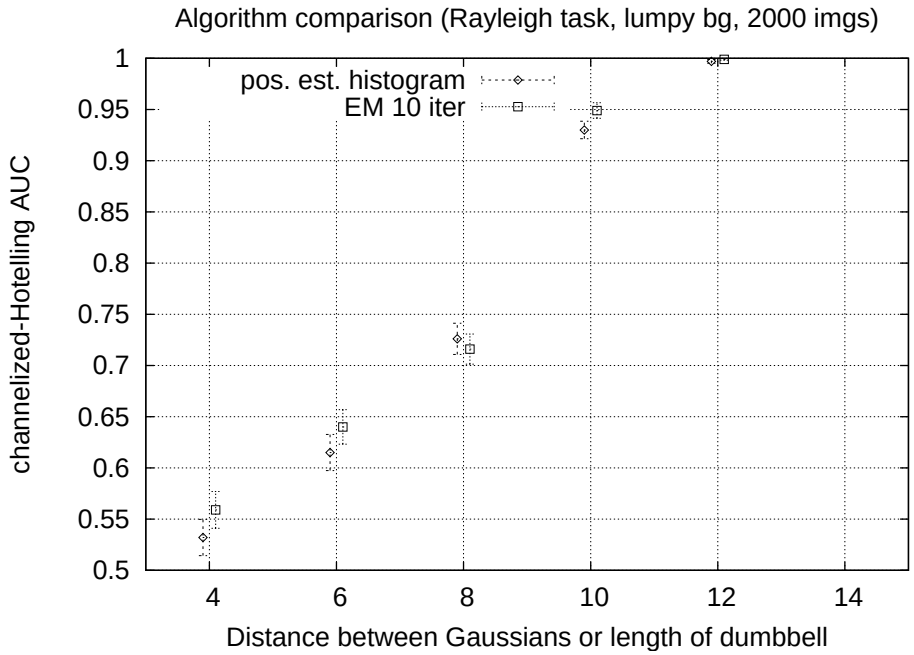

FIGURE 3.11. **Area under the ROC curve for the Rayleigh-task CHO with a lumpy background,** $w = 3$**.** Background lump width is the wider than the signal. Shown are Shown are results for the CHO acting on images reconstructed using both algorithms for a variety of lengths. The error bars have been horizontally displaced; the dumbbell lengths are exactly 4, 6, 8, 10, 12, and 14 for both algorithms.

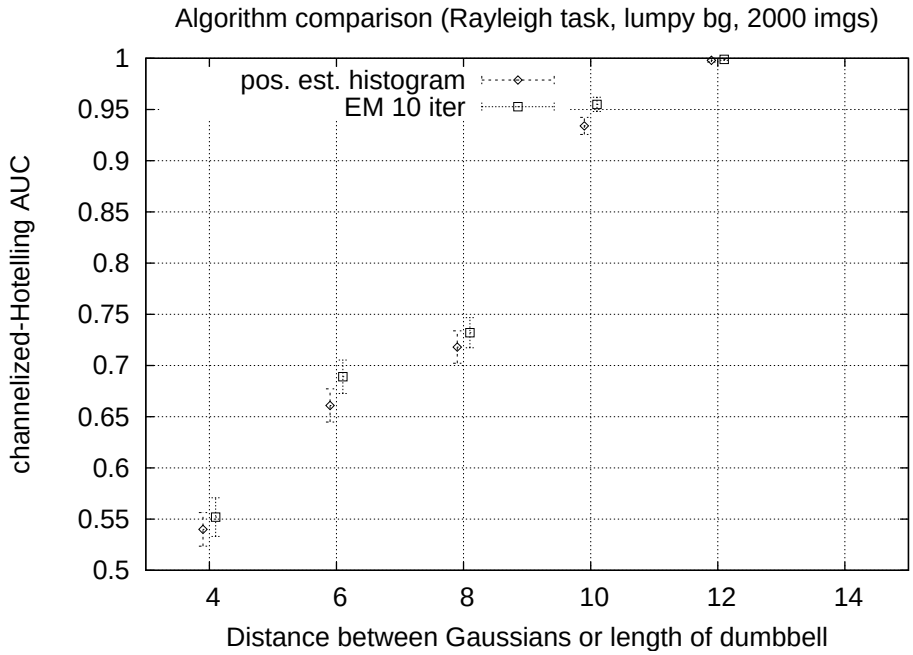

FIGURE 3.12. **Area under the ROC curve for the Rayleigh-task CHO with a lumpy background,** $w = 4$**.** Background lump width is the wider than the signal. Shown are Shown are results for the CHO acting on images reconstructed using both algorithms for a variety of lengths. The error bars have been horizontally displaced; the dumbbell lengths are exactly 4, 6, 8, 10, 12, and 14 for both algorithms.

The channelized Hotelling observer results show the EM algorithm yields higher resolution reconstructions than the histogram approach. The effect is especially apparent in the optimal (for EM) case of a zero background, and still exists for a flat background. Unfortunately the effect is greatly reduced against the lumpy background, which was designed to mimic the sorts of textures found in clinical images.

Because of the increased processing costs of the EM algorithm must be weighed against the slight improvement in resolution, it is unlikely to be of much value in generating projection images for regular use. In particular, the 3-dimensional EM algorithm can already compensate for the blur introduced during the histogram process when generating tomographic images. Therefore I expect the fluence estimation *not* to be useful when generating projection images prior to binned-mode tomographic reconstruction. (Verifying that is a possible avenue of future work.) The EM fluence estimates are most useful for the case when one wants to publish or present projection images, in which case a "prettier" image may have expository value.

Chapter 4

Estimability analysis for list-mode imaging systems

Note to the reader: This chapter was originally written as a stand-alone article intended for separate publication, and later added to the dissertation. As a result there are some duplications and differences in notation from the other chapters, and the text isn't integrated into the rest of the dissertation. A related discussion of null-space/measurement-space decomposition for a list-mode system and discretized objects is found in section 1.8.

4.1 Introduction

Most digital imaging detectors, ranging from cheap mass-produced cell phone cameras to expensive scientific instrumentation, operate in binned mode. A binned detector integrates for a preset amount of time, and then outputs a histogram. Bins typically correspond to regions on the camera face, for example detector pixels, but often also partition other dimensions of data space, such as spectrum or color. The values stored in each bin are proportional to the number of photons reaching that bin of the detector; in some cases bin values are literally a photon count.

Binned-mode detectors are popular because they work well in most situations, but they have serious drawbacks in regimes with very low count rates. The use of bins places an upper bound on the spatial, spectral, and temporal resolution of the imaging system. In many detector designs increasing the number of bins simultaneously increases readout time, and thus decreases temporal resolution. For a fixed number of incoming photons, in the high-resolution limit of many small bins most bins detect zero events, only a few bins detect one event, and probably no bins detect more than

one event. For this limit, which does occur in practice, having the detector output a sparse representation makes more sense than binned mode. For situations where the average count rate in each bin would be less than one, detectors are designed to output a list containing information about each detected event.

List-mode detectors are used in a variety of scientific disciplines. High-energy physics has been collecting data this way for decades [Solmitz, 1964], as has nuclear-medicine imaging [Snyder and Politte, 1983]. Although most closely associated with gamma-ray imaging, some optical imaging systems can be operated in list mode, for example the imaging spectrometer on the Hubble Space Telescope [Kim Quijano et al., 2003]. (Astronomers often refer to list mode as time-tagged mode.)

Each entry in the data list is an attribute vector containing information about a single detected photon. In the simplest case the vector components are the spatial coordinates of the event and perhaps a time stamp. For high-energy events, such as those occurring in nuclear medicine, instead of coordinates the attribute vector may consist of output from several detector elements activated by the event.

In recent years many issues questions to image reconstruction from list-mode data have been studied by the medical imaging community. Because of the low signal-to-noise ratio associated with low count rates, statistical techniques such as maximum-likelihood (ML) and maximum-a-posteriori (MAP) reconstruction have received the most attention [Barrett et al., 1997; Parra and Barrett, 1998; Byrne, 2001; Huesman et al., 2000]. Because such algorithms are iterative, improving the rate of convergence is of considerable interest [Reader et al., 2002; Khurd et al., 2004], as is the prospect of making only one iteration over the list [Reader et al., 2001]. Accuracy of reconstruction methods has also been studied [Qi and Huesman, 2004].

In this paper we consider which parameters of the object are estimable from list-mode data. Intuitively speaking, these are the parameters which can be accurately estimated in the limit of no noise, which in a list-mode system manifests itself as an infinite number of list entries. A rigorous definition of estimability is given in sec-

tion 4.2 together with a review of estimability theory for binned detectors. Section 4.3 extends the theory to list-mode detectors and suggests an image reconstruction algorithm based on the theory. Section 4.4 considers related questions in signal detection using a list-mode detector. Finally section 4.5 illustrates the theory using as an example time-of-flight positron-emission tomography (PET), a type of imaging system commonly used in nuclear medicine.

4.2 Brief review of linear theory and estimability for binned detectors

This section reviews the theory of estimable parameters for binned-mode linear imaging systems. Due to space constraints the discussion is brief; for an in-depth discussion, including review of the mathematical background and physics, see Barrett and Myers [2004]. We start with the familiar matrix model of imaging systems, in which the object and data are both discretized vectors. Then we look at the case where the system produces a discretized data vector from a continuum object.

4.2.1 Estimability for discretized objects and binned detectors

In the matrix model of imaging, the object being imaged is represented as a collection of V voxels, and thus is a V-dimensional vector $\boldsymbol{f}$. In nuclear-medicine applications f_d represents the amount of radiotracer activity within voxel d. The data consist of measurements from B detector bins, and are represented as a B-dimensional vector $\boldsymbol{g}$. (In many cases the components of $\boldsymbol{g}$ correspond to pixels on the detector.) In matrix form the fundamental linear imaging equation is

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{n}, \tag{4.1}$$

where $\boldsymbol{H}$ is the $B \times V$ system matrix and $\boldsymbol{n}$ is a random variable representing the noise. By definition, the mean data vector

$$\overline{\boldsymbol{g}} = \boldsymbol{H}\boldsymbol{f} \tag{4.2}$$

$$= \langle \boldsymbol{g} \rangle_{\boldsymbol{g}|\boldsymbol{f}}\,. \tag{4.3}$$

The angle brackets indicate conditional expected value. Therefore, by definition, $\boldsymbol{n} = \boldsymbol{g} - \overline{\boldsymbol{g}}$. In the common case of Poisson noise the statistics of the noise $\boldsymbol{n}$ depend on the object $\boldsymbol{f}$.

A standard photointerpretation task is to estimate the amount of the object contained within a region of interest (ROI):

$$\theta = \boldsymbol{\chi}^\dagger \boldsymbol{f}, \tag{4.4}$$

where $\boldsymbol{\chi}$ defines the ROI and † indicates the adjoint operator. The parameter θ is *estimable* if there exists an unbiased linear estimator $\hat{\theta}(\boldsymbol{g})$, in other words if $\hat{\theta}(\boldsymbol{g})$ is a linear function of the data $\boldsymbol{g}$ and if

$$\left\langle \hat{\theta}(\boldsymbol{g}) \right\rangle = \theta. \tag{4.5}$$

The estimable parameters are those which we will correctly estimate in the noise-free limit; however a parameter being estimable is no guarantee we will do a good job estimating it in the presence of noise.

It is well known that θ is estimable if, and only if, $\boldsymbol{\chi}$ can be written as a linear combination of rows of $\boldsymbol{H}$ [Albert, 1972, thm. 6.1.8]. It can be shown that if

$$\boldsymbol{\chi} = \boldsymbol{H}^\dagger \boldsymbol{w}, \tag{4.6}$$

where $\boldsymbol{w}$ is a weighting vector in data space, then

$$\hat{\theta} = \boldsymbol{w}^\dagger \boldsymbol{g} \tag{4.7}$$

is the unbiased linear estimator we seek, with variance

$$\mathrm{var}(\hat{\theta}) = \sum_m w_m^2 \overline{g}_m. \tag{4.8}$$

The solution to (4.6) provides the estimation template $\boldsymbol{w}$. When no solution to (4.6) exists the parameter θ is not estimable. *Even in the noise-free limit, a non-estimable parameter of the object cannot be accurately estimated from the data using a linear estimator, because any linear estimator must be biased!* More precisely, the estimator cannot be unbiased for all values of the parameter.

Rewritten in inner-product notation where $\left(\boldsymbol{x}\middle|\boldsymbol{y}\right)$ indicates inner product, we have

$$\hat{\theta} = \left(\boldsymbol{w}\middle|\boldsymbol{g}\right) \tag{4.9}$$

$$\left\langle \hat{\theta} \right\rangle = \left(\boldsymbol{w}\middle|\overline{\boldsymbol{g}}\right) \tag{4.10}$$

$$= \left(\boldsymbol{w}\middle|\boldsymbol{H}\boldsymbol{f}\right) \tag{4.11}$$

$$= \left(\boldsymbol{H}^\dagger \boldsymbol{w}\middle|\boldsymbol{f}\right) \tag{4.12}$$

$$= \left(\boldsymbol{\chi}\middle|\boldsymbol{f}\right) \tag{4.13}$$

$$= \theta. \tag{4.14}$$

Not only does the estimability analysis tell us which parameters of the object are estimable, (4.10) also gives us a practical technique for estimating them from data.

4.2.2 Estimability for continuum objects and binned detectors

Most objects in the real world are not made up of a finite number of voxels. A more realistic model is to represent the object as a function $f(\boldsymbol{r})$ defined on a continuum of points. Typically $\boldsymbol{r} \in \mathbb{R}^2$ or $\boldsymbol{r} \in \mathbb{R}^3$, with specified compact support. In nuclear-medicine imaging $f(\boldsymbol{r})$ usually represents the quantity of radiotracer at $\boldsymbol{r}$ and has units of disintegrations/cm^3. For a continuum object the fundamental linear imaging

equation is

$$\boldsymbol{g} = \mathcal{H} f(\boldsymbol{r}) + \boldsymbol{n}. \tag{4.15}$$

By definition, the mean data vector

$$\overline{\boldsymbol{g}} = \langle \boldsymbol{g} \rangle_{\boldsymbol{g}|f(\boldsymbol{r})} \tag{4.16}$$

$$= \mathcal{H} f(\boldsymbol{r}). \tag{4.17}$$

The Riesz representation theorem guarantees that, for a linear imaging system, the noise-free system operator $\mathcal{H}$ can always be written as an integral operator. Each component $\overline{g}_c$ of $\overline{\boldsymbol{g}}$ can be written as

$$\overline{g}_c = \int_\infty h_c(\boldsymbol{r}) f(\boldsymbol{r})\, d\boldsymbol{r}, \tag{4.18}$$

where the kernel $h_c(\boldsymbol{r})$ is the system sensitivity function for detector element c. Thus the system sensitivity functions play the same role as the rows of H in (4.2).

As in the matrix model, the amount of the object contained in the ROI defined by $\chi(\boldsymbol{r})$ is determined by an inner product. For continuum objects the inner product takes the form of an integral,

$$\theta = \int_\infty f(\boldsymbol{r}) \chi(\boldsymbol{r})\, d\boldsymbol{r}. \tag{4.19}$$

It can be shown that θ is estimable if, and only if, $\chi(\boldsymbol{r})$ can be written as the linear combination of sensitivity functions $h_c(\boldsymbol{r})$, in other words if

$$\chi(\boldsymbol{r}) = \mathcal{H}^\dagger \boldsymbol{w}, \tag{4.20}$$

for details see Barrett and Myers [2004]. The adjoint operator $\mathcal{H}^\dagger$ is defined by

$$\mathcal{H}^\dagger \boldsymbol{w} = \sum_{c=1}^{B} h_c^*(\boldsymbol{r}) w_c, \tag{4.21}$$

where * indicates complex conjugate. Because of (4.20) the sensitivity functions have been called the "natural pixels" with which to represent the object [Buonocore et al.,

1981]. As in the matrix model, θ is an estimable parameter if, and only if, the adjoint equation (4.20) has a solution $\boldsymbol{w}$. In inner-product notation we have:

$$\hat{\theta} = \Big(\boldsymbol{w}\Big|\boldsymbol{g}\Big) \tag{4.22}$$

$$\Big\langle\hat{\theta}\Big\rangle = \Big(\boldsymbol{w}\Big|\overline{\boldsymbol{g}}\Big) \tag{4.23}$$

$$= \Big(\boldsymbol{w}\Big|\mathcal{H}\boldsymbol{f}(\boldsymbol{r})\Big) \tag{4.24}$$

$$= \Big(\mathcal{H}^{\dagger}\boldsymbol{w}\Big|\boldsymbol{f}(\boldsymbol{r})\Big) \tag{4.25}$$

$$= \Big(\boldsymbol{\chi}(\boldsymbol{r})\Big|\boldsymbol{f}(\boldsymbol{r})\Big) \tag{4.26}$$

$$= \theta. \tag{4.27}$$

As in the matrix case, through (4.23) the estimability analysis suggests a method of estimating the parameter from data.

We can derive the matrix model from the continuum model by considering the special case where the ROI is the characteristic function $\phi_j(\boldsymbol{r})$ for the jth voxel. Thus reconstructing the voxelized object $\boldsymbol{f}$ is equivalent to estimating the amount of the object contained in each voxel. If the voxels are estimable parameters, the estimability analysis suggests a linear reconstruction operator.

4.3 List-mode estimability

The previous section reviewed the concept of estimability for traditional imaging systems, in which the detector consists of a finite number of elements. This section extends the concept to list-mode imaging systems. As discussed in the introduction, a list-mode system records a T-dimensional attribute vector $\boldsymbol{v}$ for each photon it detects. The measured data consist of a list of N_{list} attribute vectors, which we write $\{\boldsymbol{v}_i\}$, where $i \in \{1, \cdots, N_{\text{list}}\}$. The list entries $\boldsymbol{v}_i$ are independent identically-distributed random variables. We restrict our analysis to systems which collect data for a preset amount of time, therefore N_{list} is also a random variable.

One can think of a list-mode system as the limit of a binned-mode system when the size of detector elements decreases towards zero, simultaneously sending the number of detector elements to infinity. In this limit each detector element will collect at most one photon. We continue to model the object as a function $f(\boldsymbol{r})$.

In order to extend the linear-systems theory reviewed in the previous section to list-mode data, we must first convert the list into a function. We do so using a simple trick:

$$u(\boldsymbol{v}) = \sum_{i=1}^{N_{\text{list}}} \delta(\boldsymbol{v} - \boldsymbol{v}_i), \tag{4.28}$$

where $\delta()$ is the Dirac delta function. Notice that $u(\boldsymbol{v})$ is a Poisson random process in attribute space.

In the binned-mode cases $\overline{\boldsymbol{g}} = \boldsymbol{H}\boldsymbol{f}$ and $\overline{\boldsymbol{g}} = \mathcal{H}f(\boldsymbol{r})$ could both be thought of as the average data set over many noise realizations. For a list-mode system noise manifests itself as a finite number of list entries, and the corresponding average is over possible lists:

$$\overline{u}(\boldsymbol{v}) = \left\langle \left\langle u(\boldsymbol{v}) \right\rangle_{\{\boldsymbol{v}_i\}|N_{\text{list}}, f(\boldsymbol{r})} \right\rangle_{N_{\text{list}}|f(\boldsymbol{r})}. \tag{4.29}$$

Thus $\overline{u}(\boldsymbol{v})$ is the mean of the random point process $u(\boldsymbol{v})$.

In nuclear-medicine, as in many other applications, the physics dictates that the number of list entries N_{list} must obey Poisson statistics conditioned on the object being imaged:

$$p\big(N_{\text{list}} = n | f(\boldsymbol{r})\big) = \frac{\overline{N}^n e^{-\overline{N}}}{n!}, \tag{4.30}$$

where $\overline{N}$ is the mean of N_{list}. Let $s(\boldsymbol{r})$ represent the system sensitivity function, in other words the probability that a gamma ray emitted at point $\boldsymbol{r}$ inside the object is detected by the system. Then $\overline{N}$ is related to the object by

$$\overline{N} = \int_{\infty} f(\boldsymbol{r}) s(\boldsymbol{r})\, d\boldsymbol{r}. \tag{4.31}$$

To fully evaluate the expected values in (4.29) we need to know details about $p\big(\boldsymbol{v}|f(\boldsymbol{r})\big)$, the probability of recording attribute vector $\boldsymbol{v}$ when imaging the object

$f(\boldsymbol{r})$. The form of $p\big(\boldsymbol{v}|f(\boldsymbol{r})\big)$ is specific to the imaging system of interest; in section 4.5 we will discuss the probability law for a PET system. For now, note that

$$p\big(\boldsymbol{v}|f(\boldsymbol{r})\big) = \int_{\infty} p\big(\boldsymbol{v}|\boldsymbol{r}\big)p\big(\boldsymbol{r}|f(\boldsymbol{r})\big)\, d\boldsymbol{r}, \tag{4.32}$$

where $p(\boldsymbol{v}|\boldsymbol{r})$ is the probability a photon originating at point $\boldsymbol{r}$ will produce a list entry $\boldsymbol{v}$, and $p\big(\boldsymbol{r}|f(\boldsymbol{r})\big)$ is the probability of a detected photon being emitted from $\boldsymbol{r}$ by the object $f(\boldsymbol{r})$. Because

$$p\big(\boldsymbol{r}|f(\boldsymbol{r})\big) = \frac{f(\boldsymbol{r})s(\boldsymbol{r})}{\int_{\infty} f(\boldsymbol{r}')s(\boldsymbol{r}')\, d\boldsymbol{r}'} \tag{4.33}$$

$$= \frac{f(\boldsymbol{r})s(\boldsymbol{r})}{\overline{N}}, \tag{4.34}$$

we have

$$p\big(\boldsymbol{v}|f(\boldsymbol{r})\big) = \int_{\infty} p\big(\boldsymbol{v}|\boldsymbol{r}\big)\frac{f(\boldsymbol{r})s(\boldsymbol{r})}{\overline{N}}\, d\boldsymbol{r}. \tag{4.35}$$

Using these identities we can partially evaluate the expected values in (4.29). First we evaluate the inner expected value:

$$\big\langle u(\boldsymbol{v})\big\rangle_{\{\boldsymbol{v}_i\}|N_{\text{list}},f(\boldsymbol{r})} = \left\langle \sum_{i=1}^{N_{\text{list}}} \delta\left(\boldsymbol{v}-\boldsymbol{v}_i\right) \right\rangle_{\{\boldsymbol{v}_i\}|N_{\text{list}},f(\boldsymbol{r})} \tag{4.36}$$

$$= \int d\boldsymbol{v}_1 p\big(\boldsymbol{v}_1|f(\boldsymbol{r})\big) \cdots \int d\boldsymbol{v}_{N_{\text{list}}} p(\boldsymbol{v}_{N_{\text{list}}}|f(\boldsymbol{r})\big) \sum_{i=1}^{N_{\text{list}}} \delta\left(\boldsymbol{v}-\boldsymbol{v}_i\right) \tag{4.37}$$

$$= N_{\text{list}}\, p\big(\boldsymbol{v}|f(\boldsymbol{r})\big). \tag{4.38}$$

Next we evaluate the outer expected value:

$$\Big\langle N_{\text{list}}\, p\big(\boldsymbol{v}|f(\boldsymbol{r})\big)\Big\rangle_{f(\boldsymbol{r})} = \sum_{N_{\text{list}}=1}^{\infty} p\big(\boldsymbol{v}|f(\boldsymbol{r})\big)\, N_{\text{list}}\, p\big(N_{\text{list}}|f(\boldsymbol{r})\big) \tag{4.39}$$

$$= \overline{N} p\big(\boldsymbol{v}|f(\boldsymbol{r})\big). \tag{4.40}$$

Substituting (4.35) into (4.40) yields

$$\overline{u}(\boldsymbol{v}) = \overline{N_{\text{list}}} p\big(\boldsymbol{v}|f(\boldsymbol{r})\big) \tag{4.41}$$

$$= \int_{\infty} p(\boldsymbol{v}|\boldsymbol{r}) s(\boldsymbol{r}) f(\boldsymbol{r})\, d\boldsymbol{r} \tag{4.42}$$

$$\equiv \int_{\infty} l(\boldsymbol{v}, \boldsymbol{r}) f(\boldsymbol{r})\, d\boldsymbol{r} \tag{4.43}$$

$$\equiv [\mathcal{L}f](\boldsymbol{v}). \tag{4.44}$$

Equations (4.43) and (4.44) define the list-mode system operator $\mathcal{L}$, relating the object $f(\boldsymbol{r})$ to the noise-free data $\overline{u}(\boldsymbol{v})$. The list-mode operator $\mathcal{L}$ plays the same role for a list-mode system as $\boldsymbol{H}$ in (4.2) and $\mathcal{H}$ in (4.17) do for a system with binned detectors. The list-mode sensitivity function $l(\boldsymbol{v}, \boldsymbol{r})$ is the analogue of the binned-mode sensitivity function $h_c(\boldsymbol{r})$ in (4.18).

The amount of $f(\boldsymbol{r})$ contained in an ROI is still given by (4.19). Now our weight function is the solution to the integral equation

$$\chi(\boldsymbol{r}) = \mathcal{L}^{\dagger} w(\boldsymbol{v}) \tag{4.45}$$

$$\equiv \int_{\infty} p(\boldsymbol{v}|\boldsymbol{r}) s(\boldsymbol{r}) w(\boldsymbol{v})\, d\boldsymbol{v}. \tag{4.46}$$

The object's parameter θ is estimable if, and only if, (4.45) has a solution. Because $w(\boldsymbol{v})$ is a function in T dimensions, a list-mode system has considerably more estimable parameters than a binned-mode system. (It bears repeating: the estimability analysis considers the noise-free limit, and thus gives a best-case scenario. In practice, measurement noise will limit how many parameters of the object we can jointly estimate, because the accuracy of the estimates decreases as the number of parameters increases.)

In inner-product notation we have:

$$\left(w(\boldsymbol{v})\middle|\overline{u}(\boldsymbol{v})\right) = \left(w(\boldsymbol{v})\middle|\mathcal{L}f(\boldsymbol{r})\right) \tag{4.47}$$

$$= \left([\mathcal{L}^\dagger w](\boldsymbol{r})\middle|f(\boldsymbol{r})\right) \tag{4.48}$$

$$= \left(\chi(\boldsymbol{r})\middle|f(\boldsymbol{r})\right) \tag{4.49}$$

$$= \theta. \tag{4.50}$$

If we can solve (4.45), then the data list $\{\boldsymbol{v}_i\}$ can be used to estimate the inner product $\int_\infty w(\boldsymbol{v})\overline{u}(\boldsymbol{v})\,d\boldsymbol{v}$ using Monte-Carlo integration:

$$\theta = \left(w(\boldsymbol{v})\middle|\overline{u}(\boldsymbol{v})\right) \tag{4.51}$$

$$\approx \left(w(\boldsymbol{v})\middle|u(\boldsymbol{v})\right) \tag{4.52}$$

$$= \sum_{i=1}^{N_{\text{list}}} w(\boldsymbol{v}_i) \tag{4.53}$$

$$\equiv \hat{\theta}\left(\{\boldsymbol{v}_i\}\right). \tag{4.54}$$

Furthermore, $\hat{\theta}\left(\{\boldsymbol{v}_i\}\right)$ is an unbiased linear estimator of θ, with variance

$$\int_\infty w^2(\boldsymbol{v})\overline{u}(\boldsymbol{v})\,d\boldsymbol{v}. \tag{4.55}$$

(Note that $\hat{\theta}$ is not a linear combination of the attribute vectors $\boldsymbol{v}_i$, because for most imaging systems the weight function $w(\boldsymbol{v})$ is nonlinear.) Therefore, just as for binned-mode systems, the estimability analysis not only tells us which parameters are estimable, but also suggests a practical way to estimate them. Note in particular that (4.53) requires only one pass through the list to estimate the parameter. The potentially difficult task of solving (4.45) need only be done once per imaging system; section 4.5 does so for a PET system, for which the solution turns out to be simple.

4.4 List-mode Hotelling observer

A common photointerpretation task is to decide whether or not an image contains a signal. Lesions and tumors are typical examples of signals found in medical images.

Because of the ubiquity of signal detection tasks, the accuracy with which an observer can perform the task on images produced by a given imaging system is often used as a figure of merit to measure the quality of the imaging system [Barrett, 1990]. The Hotelling observer, a linear algorithm that performs a signal-detection task, is widely used to model observer performance on a signal-detection task. Its associated figure of merit, the Hotelling trace, also called the Hotelling signal-to-noise ratio (SNR), is used to assess the quality of an imaging system. This section derives the Hotelling observer and Hotelling trace for a list-mode detector.

The Hotelling observer makes its decision by computing a linear test statistic, or discriminant, which is then compared with a threshold [Barrett and Myers, 2004]. (The Hotelling discriminant is the population equivalent of the Fisher linear discriminant.) In the familiar matrix model of a binned-mode imaging system, the Hotelling test statistic is computed by

$$t_{\text{bin}} = \Delta\overline{\boldsymbol{g}}\boldsymbol{K}_{\boldsymbol{g}}^{-1}\boldsymbol{g}, \tag{4.56}$$

where

$$\Delta\overline{\boldsymbol{g}} = \langle \boldsymbol{g} \rangle_{\text{p}} - \langle \boldsymbol{g} \rangle_{\text{a}} \tag{4.57}$$

is the difference between the mean data when the signal is present and the mean data when the signal is absent, and the covariance

$$\boldsymbol{K}_{\boldsymbol{g}} = \frac{1}{2}\left(\boldsymbol{K}_{\boldsymbol{g}|\text{p}} + \boldsymbol{K}_{\boldsymbol{g}|\text{a}}\right). \tag{4.58}$$

The figure of merit associated with the Hotelling observer is the Hotelling trace:

$$SNR_{\text{bin}}^2 = \Delta\overline{\boldsymbol{g}}^{\dagger}\boldsymbol{K}_{\boldsymbol{g}}^{-1}\Delta\overline{\boldsymbol{g}} = \text{trace}\left(\boldsymbol{K}_{\boldsymbol{g}}^{-1}\Delta\overline{\boldsymbol{g}}\Delta\overline{\boldsymbol{g}}^{\dagger}\right). \tag{4.59}$$

As in the binned-mode case, the list-mode Hotelling observer first applies the inverse-covariance operator, which now takes the form of an integral transform, to the data $u(\boldsymbol{v}) = \sum_{i=1}^{N_{\text{list}}} \delta(\boldsymbol{v} - \boldsymbol{v}_i)$, and then computes its inner product with the signal difference. The same tricks used to derive $\overline{u}(\boldsymbol{v})$ can be used to show that for a fixed

object $f(\boldsymbol{r})$ the autocovariance of $u(\boldsymbol{v})$ is given by

$$\left[\mathcal{K}u(\boldsymbol{v})\right](\boldsymbol{v}') = \overline{u}(\boldsymbol{v})\delta(\boldsymbol{v}-\boldsymbol{v}'). \tag{4.60}$$

The signal difference is given by

$$\Delta\overline{u}(\boldsymbol{v}) = \langle u(\boldsymbol{v})\rangle_{\text{present}} - \langle u(\boldsymbol{v})\rangle_{\text{absent}}\,, \tag{4.61}$$

where the expected value encompasses both noise and variability within the signal class. To avoid having to average over object ensembles, since that would break the delta correlation of the autocovariance, let us assume fixed objects $f_{\text{p}}(\boldsymbol{r})$ and $f_{\text{a}}(\boldsymbol{r})$ for signal present and absent respectively. Then

$$\left[\mathcal{K}_{\text{dat}}u(\boldsymbol{v})\right](\boldsymbol{v}') = \frac{1}{2}\big(\overline{u}_{\text{p}}(\boldsymbol{v}) + \overline{u}_{\text{a}}(\boldsymbol{v})\big)\delta(\boldsymbol{v}-\boldsymbol{v}') \tag{4.62}$$

is the average autocovariance of the data.

A possible avenue of future work is determining the average autocovariance in the presence of object variability. For the binned-mode case it is possible to partition the covariance matrix into a diagonal component due to noise, which can be determined analytically, and a nondiagonal component due to object variability, which can be estimated using noise-free data [Barrett and Myers, 2004, sec. 14.3.2] [Fiete et al., 1987]. It may be possible to similarly partition the autocovariance operator in the list-mode case, where the noise-free data from object $f_k(\boldsymbol{r})$ is now given by the probability law $p\big(\boldsymbol{v}|f_k(\boldsymbol{r})\big)$. Of course this requires us to know this expression for an ordinary object. Alternatively, it may be possible to sample list-entries produced by $f_k(\boldsymbol{r})$, and then reconstruct the probability using density-estimation techniques [Silverman, 1986].

The list-mode Hotelling test statistic is given by

$$t_{\text{LM-Hotelling}} = \Delta\overline{u}(\boldsymbol{v})^{\dagger}\left[\mathcal{K}_{\text{dat}}^{-1}u(\boldsymbol{v}')\right](\boldsymbol{v}). \tag{4.63}$$

In this notation the adjoint indicates inner product, and the covariance is an integral operator:

$$\Delta\overline{u}(\boldsymbol{v})^\dagger[\mathcal{K}_{\text{dat}}^{-1}u](\boldsymbol{v}) = \int_\infty \Delta\overline{u}(\boldsymbol{v})[\mathcal{K}_{\text{dat}}^{-1}u](\boldsymbol{v})\,d\boldsymbol{v} \tag{4.64}$$

$$= \int_\infty \Delta\overline{u}(\boldsymbol{v}) \int_\infty \frac{2\delta(\boldsymbol{v}-\boldsymbol{v'})}{\overline{u}_{\text{p}}(\boldsymbol{v'})+\overline{u}_{\text{a}}(\boldsymbol{v'})}u(\boldsymbol{v'})\,d\boldsymbol{v'}\,d\boldsymbol{v} \tag{4.65}$$

$$= \int_\infty \Delta\overline{u}(\boldsymbol{v})\frac{2u(\boldsymbol{v})}{\overline{u}_{\text{p}}(\boldsymbol{v})+\overline{u}_{\text{a}}(\boldsymbol{v})}\,d\boldsymbol{v} \tag{4.66}$$

$$= \int_\infty \frac{2\Delta\overline{u}(\boldsymbol{v})}{\overline{u}_{\text{p}}(\boldsymbol{v})+\overline{u}_{\text{a}}(\boldsymbol{v})}\sum_{i=1}^{N_{\text{list}}}\delta(\boldsymbol{v}-\boldsymbol{v}_i)\,d\boldsymbol{v} \tag{4.67}$$

$$= \sum_{i=1}^{N_{\text{list}}}\frac{2\Delta\overline{u}(\boldsymbol{v}_n)}{\overline{u}_{\text{p}}(\boldsymbol{v}_n)+\overline{u}_{\text{a}}(\boldsymbol{v}_n)}. \tag{4.68}$$

The list-mode Hotelling SNR for detecting between two known objects $f_{\text{p}}(\boldsymbol{r})$ and $f_{\text{a}}(\boldsymbol{r})$ is given is given by

$$\text{SNR}_{\text{LM}}^2 = \Delta\overline{u}(\boldsymbol{v})^\dagger\left[\mathcal{K}_{\text{dat}}^{-1}\Delta\overline{u}(\boldsymbol{v'})\right](\boldsymbol{v}) \tag{4.69}$$

$$= \int \frac{2\Delta\overline{u}(\boldsymbol{v})^2}{\overline{u}_{\text{p}}(\boldsymbol{v})+\overline{u}_{\text{a}}(\boldsymbol{v})}\,d\boldsymbol{v}. \tag{4.70}$$

4.5 Estimability theory illustrated for PET

This section illustrate the ideas from earlier sections using positron-emission tomography (PET) as an example. In PET a molecule of physiological interest is tagged with a positron-emitting radioisotope, and is then administered to the subject. Radiation emitted during the decay process is detected by the imager and used to reconstruct an image of the radiotracer density. This radiotracer map is then used to determine where the tagged molecule is concentrating.

During radioactive decay the tracer isotope emits a positron, which travels a short distance before annihilating with an electron, in the process producing two 511 keV gamma rays traveling in nearly-opposite directions. A ring of detectors surrounding the patient records a large percentage of the emitted pairs. We assume either both

photons are detected or neither is detected (no random coincidences). The locations at which the two photons are detected define a line along which the annihilation event occurred, and the time difference between the detections determines the point on the line at which the annihilation occurred. This position estimation is done in hardware, so the attribute vectors in the data list consist of the coordinates at which each detected event is estimated to have been emitted. For this illustration we restrict ourselves to the two-dimensional case. This model for list-mode PET was introduced by Parra and Barrett [1998].

The probability of the annihilation location can be modeled by an exponential centered at the positron-emission location. The measurement errors in the position estimation of the gamma-ray detection and the time difference between detections are well modeled by a Gaussian. Because the spread of the Gaussian is much wider than the spread of the exponential, the Gaussian is a good model relating the data to the true locations.

Let the attribute vector $\boldsymbol{v} \in \mathbb{R}^2$ be the estimated coordinates of each decay event. Let $\boldsymbol{r} \in \mathbb{R}^2$ be the true coordinates. The position estimate is related to the true location by a Gaussian model:

$$p(\boldsymbol{v}|\boldsymbol{r}) = \frac{1}{N} \exp\left[-\frac{1}{2}(\boldsymbol{v}-\boldsymbol{r})^\dagger \boldsymbol{S}^{-1}(\boldsymbol{v}-\boldsymbol{r})\right], \tag{4.71}$$

where

$$N = \sqrt{4\pi^2 \det(\boldsymbol{S})}. \tag{4.72}$$

If we assume uniform sensitivity, $s(\boldsymbol{r}) = 1$, then by (4.44) we have

$$\overline{u}(\boldsymbol{v}) = [\mathcal{L}f](\boldsymbol{r}) \tag{4.73}$$

$$= \int_\infty \frac{1}{N} \exp\left[-\frac{1}{2}(\boldsymbol{v}-\boldsymbol{r})^\dagger \boldsymbol{S}^{-1}(\boldsymbol{v}-\boldsymbol{r})\right] f(\boldsymbol{r})\, d\boldsymbol{r} \tag{4.74}$$

Notice that because $\boldsymbol{v}$ and $\boldsymbol{r}$ are both in $\mathbb{R}^2$, the system operator $\mathcal{L}$ defined by (4.74) takes the form of a convolution operator

$$\overline{u}(\boldsymbol{r}) = l(\boldsymbol{r}) \circledast f(\boldsymbol{r}) \tag{4.75}$$

with kernel

$$l(\boldsymbol{r}) = \frac{1}{N} \exp\left[-\frac{1}{2}\boldsymbol{r}^\dagger \boldsymbol{S}^{-1}\boldsymbol{r}\right]. \tag{4.76}$$

This simplifies things greatly over the more general case where $\boldsymbol{v}$ and $\boldsymbol{r}$ have different dimensionality. Therefore (4.45) takes the form

$$\chi(\boldsymbol{r}) = [\mathcal{L}^\dagger w](\boldsymbol{r}) \tag{4.77}$$

$$= \frac{1}{N} \exp\left[-\frac{1}{2}\boldsymbol{r}^\dagger \boldsymbol{S}^{-1}\boldsymbol{r}\right] \circledast w(\boldsymbol{r}). \tag{4.78}$$

By the convolution theorem, the solution to (4.78) is

$$w(\boldsymbol{r}) = \mathcal{F}^{-1}\left\{\frac{X(\boldsymbol{\xi})}{L(\boldsymbol{\xi})}\right\}, \tag{4.79}$$

where $X(\boldsymbol{\xi})$ is the Fourier transform of $\chi(\boldsymbol{r})$,

$$L\left(\boldsymbol{\xi}\right) = \det(\boldsymbol{S}) \exp\left[-\frac{1}{2}\boldsymbol{\xi}^\dagger \boldsymbol{S}^2 \boldsymbol{\xi}\right], \tag{4.80}$$

and $\mathcal{F}^{-1}$ indicates the inverse Fourier transform operator.

We are now in a position to answer an obvious question about PET systems: is the average activity within a square pixel an estimable parameter from list-mode PET data? We are interested in the activity within the region

$$\chi(\boldsymbol{r}) = \mathrm{rect}\left(\boldsymbol{r}\right). \tag{4.81}$$

To find the weight function we must solve (4.78):

$$\mathrm{rect}\left(\boldsymbol{r}\right) = [\mathcal{L}^\dagger w](\boldsymbol{r}) \tag{4.82}$$

$$= \frac{1}{N} \exp\left[-\frac{1}{2}\boldsymbol{r}^\dagger \boldsymbol{S}^{-1}\boldsymbol{r}\right] \circledast w(\boldsymbol{r}). \tag{4.83}$$

Taking the Fourier transform of both sides gives

$$\mathrm{sinc}(\boldsymbol{\xi}) = L(\boldsymbol{\xi}) W(\boldsymbol{\xi}). \tag{4.84}$$

Therefore

$$W(\boldsymbol{\xi}) = \frac{\mathrm{sinc}(\boldsymbol{\xi})}{L(\boldsymbol{\xi})}. \tag{4.85}$$

Because $L(\boldsymbol{\xi})$ decays to zero faster than $\text{sinc}(\boldsymbol{\xi})$ does, $W(\boldsymbol{\xi})$ blows up for high frequencies. As a result the inverse-Fourier transform will not be a function, and there is no solution to (4.82). Thus square pixels are not estimable parameters; this shouldn't come as a big surprise; pixels and voxels are almost never estimable parameters in real imaging systems. (One could interpret the inverse transform of $W(\boldsymbol{\xi})$ as a generalized function, but it would involve delta functions, in practice making it impossible to estimate (4.53) by Monte Carlo.)

4.6 Discussion and Conclusions

We have extended the theory of estimable parameters from binned-mode linear imaging systems to list-mode systems.

To do so we introduced the notion of a list-mode system operator. Like its counterpart the familiar binned-mode system operator, the list-mode operator maps the object to the mean data. As in the binned-mode case, the estimable parameters are determined by the range of the adjoint of the system operator. The operator can also be used to calculate the list-mode Hotelling observer.

To illustrate the theory we provided an example from PET imaging. In the future we plan to use (4.53) to do PET reconstruction, and we also hope to solve (4.45) for the more difficult case of SPECT.

Chapter 5
Conclusion

This chapter summarizes the main new results of the dissertation and suggests several avenues for future work.

5.1 Summary

Most of chapter 1 was review of previous work by others, describing accepted models for various steps of the nuclear-medicine imaging chain. Only a few parts of the chapter are original: section 1.7.4 showed that the histogram approach produces maximum-likelihood projection images in the limiting case of no detector blur, sections 1.7.6.3 and 1.8 provided an analysis of when discretized objects are identifiable from list-mode data, and section 1.11.5 introduced the naive list-mode Hotelling observer.

Chapter 2 reviewed several ways of constructing parametric likelihood models for list-mode imaging systems, giving extra attention to the case when extensive calibration data is available. The main new contribution of the chapter is introducing the concept of double-list-mode reconstruction. In double list-mode the calibration list is first used to reconstruct a non-parametric likelihood model, which is then used to process the data list into a reconstruction. Because double list-mode conceptually involves comparing each calibration event with each data event it has horrendous time complexity. Even with the various available speedups, double list-mode is too slow to be of practical use in SPECT given current computing technology.

The Rayleigh task study in chapter 3 compared two algorithms for reconstructing projection images from list-mode PMT data. One was the usual histogram approach, the other the list-mode EM algorithm. To my knowledge this is the first time that list-

mode EM has been used to produce projection (as opposed to tomographic) images. Rayleigh task performance shows that the EM algorithm produces higher-resolution images than histograming ML position estimates. A task-based measure of resolution is required because position estimates are a nonlinear function of modular-camera data. The EM algorithm is also nonlinear. Because of the nonlinearities the usual linear-system PSF analysis fails. The amount of resolution improvement afforded by EM is dependent on background activity in the scene; it is not measurable against a lumpy background.

The notion that only certain parameters of an object may be estimated without bias is familiar from the traditional linear analysis of binned-mode imaging systems. Chapter 4 extended this idea to list-mode systems. The concept of a list-mode system operator was developed, and used to determine the estimable parameters and to define the list-mode Hotelling observer.

5.2 Future Work

5.2.1 Double list-mode

Two obvious areas of further research exist for double list-mode. The first is to find an imaging system for which producing double-list-mode reconstructions is practical. The time complexity of comparing each data list entry to every calibration list entry (or even just a subset) is only a problem in SPECT because we have so many list entries. A photon-scarce system, perhaps one used in astronomy to observe gamma-ray bursts, would be a better candidate for double list-mode.

The second is to do a task-based study comparing parametric likelihood models with double list-mode. I had hoped to include such a study here, but it is not computationally feasible today; it may be possible in a few years. The ideal system for such a study would be one in which we don't trust parametric models to do a good job.

5.2.2 Projection image reconstruction

An interesting question not addressed by the Rayleigh task study in chapter 3 is whether or not using list-mode EM is necessary to obtain the resolution improvement. Marcotte [1993] obtained impressive results using binned-mode EM to post-process histogram reconstructions into higher-resolution projection images. Her method is significantly faster than list-mode EM; it would be nice to know if it produces a comparable increase in resolution.

Another unresolved matter is whether or not improving the projection images also improves binned-mode tomographic reconstructions starting from those images. I predict that if one is using a well-calibrated system model for tomographic reconstruction, for example the measured $\boldsymbol{H}$ matrix used here in Arizona, there will *not* be an improvement when reconstructing using the 3-dimensional binned-mode EM algorithm. The 3-dimensional EM algorithm has similar information available to it as that used by Marcotte's technique, so it should be able to back out the blur on its own. However if one is using a poor system model to reconstruct there may be a noticeable improvement.

5.2.3 Estimability

A number of avenues are suggested by the estimability results of chapter 4. One is figuring out how to apply them to the modular gamma camera. To do this we must solve

$$\chi(\boldsymbol{r}) = \mathcal{L}^{\dagger} w(\boldsymbol{v}) \tag{4.45}$$

for the weight function $w(\boldsymbol{v})$. Note that $\boldsymbol{v} \in \mathbb{R}^9$ and $\boldsymbol{r} \in \mathbb{R}^2$. The PET example was tractable only because both vectors were the same dimension, so the system operator $\mathcal{L}$ worked out to be a convolution. This trick doesn't work for the modular camera. If we can figure out how to solve (4.45) several nifty things become possible. Most interesting to me is the prospect of one-pass linear reconstructions using (4.53). If

the weight function is cheap to evaluate this could be done quickly, perhaps even in real time.

Also useful would be computing the list-mode Hotelling observer, and comparing it to the list-mode ideal observer for problems in which we know how to evaluate the later. This then should allow us to estimate the efficiency of reconstruction algorithms, by comparing human (or anthropomorphic CHO) performance on the reconstructed image with ideal performance on the raw data list.

Many of these things are already possible using the PET model. One issue that still needs to be resolved is how best to use (4.53) to reconstruct objects. Choosing $\chi(\boldsymbol{r})$ to be elements of the familiar pixel and voxels won't work, because their sharp edges make them non-estimable parameters. But perhaps through clever use of an apodizing filter something similar can be made to work. High speed reconstruction algorithms remain much sought after, and may have clinical applications even if they produce lower quality reconstructions when time isn't factored into the task.

As currently formulated the PET model does not account for random coincidences. Seeing how that changes the estimability analysis is another possible avenue of work.

Appendix A

Glossary of abbreviations and acronyms

2-AFC Two Alternative Forced Choice

A/D Analog to Digital

AI Artificial Intelligence

a.k.a. Also Known As.

AUC Area Under the Curve

BKE Background Known Exactly

c. circa

CAD Computer aided diagnosis

CHO Channelized Hotelling Observer

EKG Electrocardiogram[1]

EMSE Ensemble Mean Square Error

FNF False Negative Fraction

FOM Figure Of Merit

FPF False Positive Fraction

FWHM Full-Width Half-Maximum

GSL GNU Scientific Library

[1]The K comes from German "Elektrokardiogramm".

HMPAO Hexamethylpropyleneamine Oxime

i.i.d. independent and identically distributed

MDP Methyenediphosonate

MDRF Mean Detector Response Function

MRI Magnetic Resonance Imaging

MSE Mean Square Error

MiB Mebibytes (2^{20} bytes), abbreviated from megabinary bytes

nucmed nuclear medicine

PET Positron Emission Tomography

PMT PhotoMultiplier Tube

PSF Point Spread Function

PSPMT Position-sensitive PhotoMultiplier Tube.

ROC Receiver Operating Characteristic

RV Random Variable

SKE/BKE Signal Known Exactly, Background Known Exactly

SKE Signal Known Exactly

SPECT Single Photon Emission Computed Tompgraphy

TNF True Negative Fraction

TPF True Positive fraction

UMC University Medical Center

References

J. N. Aarsvold, H. H. Barrett, J. Chen, A. L. Landesman, T. D. Milster, D. D. Patton, T. J. Roney, R. K. Rowe, R. H. Seacat III, and L. M. Strimbu. Modular scintillation cameras: a progress report. In Roger H. Schneider and Samuel J. Dwyer III, editors, *Medical imaging II: Image formation, detection, processing and interpretation*, volume 914 (part A) of *Proceedings of SPIE*, pages 319–25, 1988.

Craig K. Abbey and François Bochud. Modeling visual detection tasks in correlated image noise with linear model observers. In Jacob Beutel, Harold L. Kundel, and Richard L. Van Metter, editors, *Progress in medical physics and psychophysics*, volume 1 of *Handbook of medical imaging*, pages 629–54. SPIE, Bellingham, WA, 2000.

Craig Kendall Abbey. *Assessment of Reconstructed Images*. PhD thesis, University of Arizona, Graduate Interdisciplinary Program in Applied Mathematics, Tucson, AZ, 1998.

Arthur E. Albert. *Regression and the Moore-Penrose pseudoinverse*. Academic Press, New York, 1972.

Hal O. Anger. Scintillation camera. *The Review of Scientific Instruments*, 29(1): 27–33, 1958.

F. J. Anscombe. The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, 35(3/4):246–54, December 1948. URL `http://links.jstor.org/sici?sici=0006-3444%28194812%2935%3A3%2F4%3C246%3ATTOPBA%3E2.0.CO%3B2-W`.

Harrison H. Barrett. Objective assessment of image quality: effects of quantum noise and object variability. *Journal of the Optical Society of America A – Optics, Image-Science and Vision*, 7(7):1266–1278, July 1990.

Harrison H. Barrett and Kyle J. Myers. *Foundations of Image Science*. Wiley Interscience, Hoboken, NJ, 2004. ISBN 0471153001.

Harrison H. Barrett, Kyle J. Myers, Brandon D. Gallas, Eric Clarkson, and Hongbin Zhang. Megalopinakophobia: Its symptoms and cures. In Larry E. Antonuk and Martin J. Yaffe, editors, *Medical Imaging 2001: Physics of Medical Imaging*, volume 4320 of *Proceedings of SPIE*, pages 299–307, June 2001.

Harrison H. Barrett and William Swindell. *Radiological imaging: The theory of image formation, detection, and processing*. Academic Press, New York, 1981. ISBN 0120796031.

Harrison H. Barrett, Timothy White, and Lucas C. Parra. List-mode likelihood. *Journal of the Optical Society of America A – Optics, Image-Science and Vision*, 14(11):2914–23, November 1997.

M. S. Bartlett. The square root transformation in analysis of variance. *Supplement to the Journal of the Royal Statistical Society*, 3(1):68–78, 1936. URL `http://links.jstor.org/sici?sici=1466-6162%281936%293%3A1%3C68%3ATSRTIA%3E2.0.CO%3B2-V`.

Richard Ernest Bellman. *Adaptive control processes: a guided tour*. Princeton University Press, 1962.

Michael H. Buonocore, William R. Brody, and Albert Macovski. A natural pixel decomposition for two-dimensional image reconstruction. *IEEE Transactions on Biomedical Engineering*, 28(2):69–78, February 1981.

Jerrold T. Bushberg, J. Anthony Seibert, Edwin M. Leidholdt, and John M. Boone. *The Essential Physics of Medical Imaging*. Williams & Wilkins, Baltimore, 1994. ISBN 0683011405.

Charles Byrne. Likelihood maximization for list-mode emission tomographic image reconstruction. *IEEE Transactions on Medical Imaging*, 20(10):1084–92, October 2001.

Charles L. Byrne. Iterative image reconstruction algorithms based on cross-entropy minimization. *IEEE Transactions on Image Processing*, 2(1):96–103, January 1993.

George Casella and Roger L. Berger. *Statistical Inference*. Duxbury Press, Belmont, CA, 1990. ISBN 0-534-11958-1.

Benedict Cassen, Lawrence Curtis, Clifton Reed, and Raymond Libby. Instrumentation for I^{131} use in medical studies. *Nucleonics*, 9(2):46–50, August 1951.

Ramesh Chandra. *Introductory Physics of Nuclear Medicine*. Lea & Febiger, Philadelphia, 1992. ISBN 0812114426.

Mark R. Chartrand and Wil Tirion. *The Audubon Society field guide to the night sky*. A. A. Knopf, New York, 1991. ISBN 0679408525.

Eric Clarkson and Harrison Barrett. A bound on null functions for digital imaging systems with positivity constraints. *Optics Letters*, 22(11):814–815, June 1997.

Eric Clarkson and Harrison Barrett. Bounds on null functions of linear digital imaging systems. *Journal of the Optical Society of America A – Optics, Image-Science and Vision*, 15(5):1355–1360, May 1998.

Pierre Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, April 1994.

Thomas H. Cormen, Charles E. Leiserson, and Ronald L. Rivest. *Introduction to Algorithms*. MIT Press, Cambridge, MA, 1990. ISBN 0262031418.

William Rutter Dawes. Catalogue of micrometrical measurements of double stars. *Memoirs of the Royal Astronomical Society*, 35:137–502, 12 April 1867. URL `http://adsabs.harvard.edu/cgi-bin/nph-bib_query?bibcode=1867MNRAS..27..217D&db_key=AST&high=416229f11f25908`.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977. URL `http://links.jstor.org/sici?sici=0035-9246%281977%2939%3A1%3C1%3AMLFIDV%3E2.0.CO%3B2-Z`.

Luc Devroye. *A course in density estimation*. Birkhäuser, Boston, 1987. ISBN 0-8176-3365-0.

Persi Diaconis and Bradley Efron. Computer-intensive methods in statistics. *Scientific American*, 248(5):116–30, 1983.

David L. Donoho. High-dimensional data analysis: Curses and blessings of dimensionality. Unpublished manuscript, available from `http://www-stat.stanford.edu/~donoho/Lectures/AMS2000/AMS2000.html`, 8 August 2000.

Miguel P. Eckstein, Craig K. Abbey, Francois O. Bochud, Jay L. Bartroff, and James S. Whiting. Effect of image compression in model and human performance. In Elizabeth A. Krupinski, editor, *Medical Imaging 1999: Image Perception and Performance*, volume 3663 of *Proceedings of SPIE*, pages 243–252, February 1999.

Miguel P. Eckstein, Craig K. Abbey, and James S. Whiting. Human vs. model observers in anatomic backgrounds. In Harold L. Kundel, editor, *Medical Imaging 1998: Image Perception*, volume 3340 of *Proceedings of SPIE*, pages 16–26, April 1998.

Bradley Efron and Carl Morris. Stein's paradox in statistics. *Scientific American*, 236(5):119–127, 1977.

Bradley Efron and Robert Tibshirani. Statistical data analysis in the computer age. *Science*, 253(5018):390–395, July 1991.

Encyclopædia Britannica Online. Mizar, 21 October 2004. URL `http://search.eb.com/eb/article?tocId=9053081`.

R. D. Fiete, H. H. Barrett, W. E. Smith, and K. J. Myers. Hotelling trace criterion and its correlation with human-observer performance. *Journal of the Optical Society of America A – Optics, Image-Science and Vision*, 4(5):945–953, May 1987.

B. Roy Frieden. The importance of being positive. In *Processing of Images and Data from Optical Sensors*, volume 292 of *Proceedings of SPIE*, pages 151–159, 1981.

Keinosuke Fukunaga and Patrenahalli M. Narendra. A branch and bound algorithm for computing k-nearest neighbours. *IEEE Transactions on Computers*, C-24(7): 750–3, July 1975.

Lars R. Furenlid, Donald W. Wilson, Yi-chun Chen, Hyunki Kim, Philip J. Pietraski, Michael J. Crawford, and Harrison H. Barrett. FastSPECT II: A second-generation high-resolution dynamic SPECT imager. *IEEE Transactions on Nuclear Science*, 51(3):631–5, June 2004.

D Gagnon, A Todd-Pokropek, A Arsenaualt, and G. Dupras. Introduction to holospectral imaging in nuclear medicine for scatter subtraction. *IEEE Transactions on Medical Imaging*, 8(3):245–50, September 1989.

Mark Galassi, Jim Davies, James Theiler, Brian Gough, Gerard Jungman, Michael Booth, and Fabrice Rossi. *GNU Scientific Library Reference Manual*. Network Theory Ltd., Bristol, UK, second edition, 2003. ISBN 0954161734. URL `http://www.gnu.org/software/gsl/`.

Brandon D. Gallas. Variance of the channelized-hotelling observer from a finite number of trainers and tester. In Dev P. Chakraborty and Elizabeth A. Krupinski, editors, *Medical Imaging 2003: Image Perception, Observer Performance, and Technology Assessment*, volume 5034 of *Proceedings of SPIE*, pages 100–111, 2003.

Brandon D. Gallas and Harrison H. Barrett. Validating the use of channels to estimate the ideal linear observer. *Journal of the Optical Society of America A – Optics, Image-Science and Vision*, 20(9):1725–38, September 2003.

David M. Green and John A. Swets. *Signal detection theory and psychophysics*. Wiley, New York, 1966.

K. M. Hanson and K. J. Myers. Performance of the Rayleigh task based on the posterior probability of tomographic reconstructions. In *Conference Record of the 1991 IEEE Nuclear Science Symposium and Medical Imaging Conference*, pages 2049–2053, November 1991a.

Kenneth M. Hanson and Kyle J. Myers. Rayleigh task performance as a method to evaluate image reconstruction algorithms. In W. T. Grandy, Jr. and L. H.

Schick, editors, *Maximum Entropy and Bayesian Methods*, volume 43 of *Fundamental Theories of Physics*, pages 303–312. Kluwer Academic, Dordrecht, 1991b. URL `http://public.lanl.gov/kmh/publications/maxent90.abs.html`.

J. L. Harris. Resolving power and decision theory. *Journal of the Optical Society of America*, 54(5):606–611, May 1964.

Ronald H. Huesman, Gregory J. Klein, William W. Moses, Jinyi Qi, Bryan W. Reutter, and Patrick R. G. Virador. List-mode maximum-likelihood reconstruction applied to positron emission mammography (PEM) with irregular sampling. *IEEE Transactions on Medical Imaging*, 19(5):532–7, May 2000.

Richard Arnold Johnson and Dean W. Wichern. *Applied multivariate statistical analysis*. Prentice-Hall, Englewood Cliffs, NJ, second edition, 1982.

Parmeshwar Khurd, Ing-Tsung Hsiao, Anand Rangarajan, and Gene Gindi. A globally convergent regularized ordered-subset EM algorithm for list-mode reconstruction. *IEEE Transactions on Nuclear Science*, 51(3):719–25, June 2004.

J. Kim Quijano et al. *STIS Instrument Handbook Version 7.0*. Space Telescope Science Institute, Baltimore, MD, 2003. URL `http://www.stsci.edu/hst/stis/documents/handbooks/`.

J.E. Koss, D.L. Kirch, E.P. Little, T.K. Johnson, and P.P. Steele. Advantages of list-mode acquisition of dynamic cardiac data. *IEEE Transactions on Nuclear Science*, 44(6):2431–2438, December 1997.

G. Krinke. *The laboratory rat*. Academic Press, San Diego, CA, 2000. ISBN 012426400X.

Kenneth Lange. *Numerical Analysis for Statisticians*. Springer-Verlag, New York, 1999. ISBN 0387949798.

Kenneth Lange and Richard Carson. EM reconstruction algorithms for emission and transmission tomography. *Journal of Computer Assisted Tomography*, 8(2):306–16, April 1984.

Carolyn E. Lehner, Zhong He, and Feng Zhang. 4π Compton imaging using a 3-D position-sensitive CdZnTe detector via weighted list-mode maximum likelihood. *IEEE Transactions on Nuclear Science*, 51(4):1618–22, August 2004.

Andre Lehovich, Harrison H. Barrett, Eric W. Clarkson, and Arthur F. Gmitro. Estimability of spatio-temporal activation in fMRI. In Michael F. Insana and Richard M. Leahy, editors, *Information Processing in Medical Imaging (IPMI 2001 Proceedings)*, volume 2082 of *Lecture Notes in Computer Science*, pages 259–271,

2001. URL http://link.springer.de/link/service/series/0558/bibs/2082/20820259.htm.

D. O. Loftsgaarden and C. P. Quesenberry. A nonparametric estimate of a multivariate density function. *Annals of Mathematical Statistics*, 36:1049–1051, June 1965. URL http://links.jstor.org/sici?sici=0003-4851%28196506%2936%3A3%3C1049%3AANEOAM%3E2.0.CO%3B2-M.

Hope Ann Marcotte. Expectation maximization methods for processing SPECT images. Master's thesis, University of Arizona, Committee on Optical Sciences, Tucson, AZ, 1993.

Geoffrey J. McLachlan and Thriyambakam Krishnan. *The EM Algorithm and Extensions*. John Wiley & Sons, New York, 1997. ISBN 0-471-12358-7.

T. D. Milster, J. N. Aarsvold, H. H. Barrett, A. L. Landesman, L. S. Mar, D. D. Patton, T. J. Roney, R. K. Rowe, and R. H. Seacat 3rd. A full-field modular gamma camera. *The Journal of Nuclear Medicine*, 31(5):632–9, May 1990.

John F. Monahan. *Numerical methods of statistics*. Cambridge University Press, Cambridge, 2001. ISBN 0521791685.

Gordon E. Moore. Cramming more components onto integrated circuits. *Electronics*, 38(8):114–7, 19 April 1965. URL ftp://download.intel.com/research/silicon/moorespaper.pdf.

K. J. Myers, J. P. Rolland, H. H. Barrett, and R. F. Wagner. Aperture optimization for emission imaging: effect of a spatially varying background. *Journal of the Optical Society of America A – Optics, Image-Science and Vision*, 7(7):1279–93, July 1990.

Kyle J. Myers and Harrison H. Barrett. Addition of a channel mechanism to the ideal-observer model. *Journal of the Optical Society of America A – Optics, Image-Science and Vision*, 4(12):2447–2457, December 1987.

Kyle J. Myers, Robert F. Wagner, and Kenneth M. Hanson. Rayleigh task performance in tomographic reconstructions: comparison of human and machine performance. In Murray H. Loew, editor, *Medical Imaging 1993: Image Processing*, volume 1898 of *Proceedings of SPIE*, pages 628–637, 1993.

OED 1989. *The Oxford English Dictionary*. Clarendon Press, Oxford, second edition, 1989. URL http://dictionary.oed.com.

Lucas Parra and Harrison H. Barrett. List-mode likelihood: EM algorithm and image quality estimation demonstrated on 2-D PET. *IEEE Transactions on Medical Imaging*, 17(2):228–35, April 1998.

Jinyi Qi and Ronald H. Huesman. Propagation of errors from the sensitivity image in list mode reconstruction. *IEEE Transactions on Medical Imaging*, 23(9):1094–9, September 2004.

Lord Rayleigh, John William Strutt. Investigations in optics, with special reference to the spectroscope (sections 1-6). *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 8:261–274, 403–411, 477–486, 1879.

Lord Rayleigh, John William Strutt. Investigations in optics, with special reference to the spectroscope (section 7). *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 9:40–55, 1880.

Andrew J. Reader, Stijn Ally, Filippos Bakatselos, Roido Manavaki, Richard J. Walledge, Alan P. Jeavons, Peter J. Julyan, Sha Zhao, David L. Hastings, and Jamal Zweit. Regularized one-pass list-mode EM algorithm for high resolution 3D PET image reconstruction into large arrays. In J. Anthony Seibert, editor, *2001 IEEE Nuclear Science Symposium Conference Record Volume 4*, pages 1853–1858, 2001. URL `http://dias.umist.ac.uk/AJR/rople.pdf`.

Andrew J. Reader, Roido Manavaki, Sha Zhao, Peter J. Julyan, David L. Hastings, and Jamal Zweit. Accelerated list-mode EM algorithm. *IEEE Transactions on Nuclear Science*, 49(1):42–49, February 2002.

J. P. Rolland and H. H. Barrett. Effect of random background inhomogeneity on observer detection performance. *Journal of the Optical Society of America A – Optics, Image-Science and Vision*, 9(5):649–58, May 1992.

H. L. Royden. *Real Analysis*. Macmillan, London, second edition, 1968.

John David Sain. *Optical Modeling, Design Optimization, and Performance Analysis of a Gamma Camera for Detection of Breast Cancer*. PhD thesis, Unversity of Arizona, Committee on Optical Sciences, Tucson, AZ, 2001.

David W. Scott. *Multivariate density estimation: theory, practice, and visualization*. Wiley-Interscience, New York, 1992.

L. A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, MI-1(2):113–22, October 1982.

Steven S. Shimozaki, Miguel P. Eckstein, and Craig K. Abbey. An ideal observer with channels versus feature-independent processing of spatial frequency and orientation in visual search performance. *Journal of the Optical Society of America A – Optics, Image-Science and Vision*, 20(12):2197–2215, December 2003.

B. W. Silverman. *Density estimation for statistics and data analysis*. Chapman and Hall, London & New York:, 1986. ISBN 0412246201.

Donald L. Snyder and David G. Politte. Image reconstruction from list-mode data in an emission tomography system having time-of-flight measurements. *IEEE Transactions on Nuclear Science*, NS-30(3):1843–9, June 1983.

Frank T. Solmitz. Analysis of experiments in particle physics. *Annual Review of Nuclear Science*, 14:375–402, 1964.

Mark A. Suckow, Peggy Danneman, and Cory Brayton. *The laboratory mouse*. CRC Press, Boca Raton, FL, 2001. ISBN 0849303222.

John A. Swets, Robyn M. Dawes, and John Monahan. Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest (Supplement to Psychological Science)*, 1(1):1–26, May 2000. URL `http://www.psychologicalscience.org/journals/pspi/1_1.html`.

R. F. Wagner, K. J. Myers, D. G. Brown, and M. P. Anderson. Toward optimal observer performance of detection and discrimination tasks on reconstructions from sparse data. In K. M. Hanson and R. N. Silver, editors, *Maximum Entropy and Bayesian Methods*, volume 79 of *Fundamental Theories of Physics*, pages 211–220. Kluwer Academic, Dordrecht, 1996.

Robert F. Wagner, David G. Brown, and Charles E. Metz. On the multiplex advantage of coded source/aperture photon imaging. In William R. Brody, editor, *Digital Radiography*, volume 314 of *Proceedings of SPIE*, pages 72–76, January 1981.

A. B. Watson. Detection and recognition of simple spatial forms. In O. J. Braddick and A. C. Sleigh, editors, *Physical and Biological Processing of Images*, volume 11 of *Springer Series in Information Sciences*, pages 100–114, 1983. URL `http://vision.arc.nasa.gov/publications/DetectionRecognitionSimple.pdf`.

Donald William Wilson. *Noise and Resolution Properties of FB and ML-EM Reconstructed SPECT Images*. PhD thesis, University of North Carolina at Chapel Hill, Department of Biomedical Engineering, 1994.

Yani Zhang, Binh T. Pham, and Miguel P. Eckstein. Automated optimization of JPEG 2000 encoder options based on model observer performance for detecting variable signals in X-ray coronary angiograms. *IEEE Transactions on Medical Imaging*, 23(4):459–474, April 2004a.

Yani Zhang, Binh T. Pham, and Miguel P. Eckstein. Evaluation of JPEG 2000 encoder options: human and model observer detection of variable signals in X-ray coronary angiograms. *IEEE Transactions on Medical Imaging*, 23(5):613–632, May 2004b.