

Adaptive SPECT

Harrison H. Barrett*, *Fellow, IEEE*, Lars R. Furenlid, *Member, IEEE*, Melanie Freed, Jacob Y. Hesterman, Matthew A. Kupinski, Eric Clarkson, and Meredith K. Whitaker

Abstract—Adaptive imaging systems alter their data-acquisition configuration or protocol in response to the image information received. An adaptive pinhole single photon emission computed tomography (SPECT) system might acquire an initial scout image to obtain preliminary information about the radiotracer distribution and then adjust the configuration or sizes of the pinholes, the magnifications, or the projection angles in order to improve performance. This paper briefly describes two small-animal SPECT systems that allow this flexibility and then presents a framework for evaluating adaptive systems in general, and adaptive SPECT systems in particular. The evaluation is in terms of the performance of linear observers on detection or estimation tasks. Expressions are derived for the ideal linear (Hotelling) observer and the ideal linear (Wiener) estimator with adaptive imaging. Detailed expressions for the performance figures of merit are given, and possible adaptation rules are discussed.

Index Terms—Adaptive imaging, Hotelling observer, image quality, single photon emission computed tomography (SPECT), Wiener estimator.

I. INTRODUCTION

THERE is considerable interest in many fields of imaging in adaptive systems that autonomously alter their data-acquisition configuration or protocol in real time in response to the image information being received. The goal of the adaptation may be to correct for random, dynamic imperfections in the imaging system or to optimize the system for the particular (unknown) object being imaged.

Adaptive imaging is most highly developed in ground-based astronomical imaging [1]–[5], where the spatial resolution is severely limited by rapidly changing phase distortions produced by the turbulent atmosphere. All modern observatories attempt to overcome this problem with some form of adaptive optics.

Manuscript received xxxx; revised xxxx. This work was supported by the National Institutes of Health under Grant P41 EB 002035 and Grant R37 EB 000803. Asterisk indicates corresponding author.

*H. H. Barrett is with the Center for Gamma-Ray Imaging and the Department of Radiology, University of Arizona, Tucson, AZ 85724 USA and also with the College of Optical Sciences, University of Arizona, Tucson, AZ 85721 USA (e-mail: barrett@radiology.arizona.edu).

L. R. Furenlid, M. A. Kupinski, E. Clarkson, and M. K. Whitaker are with the Center for Gamma-Ray Imaging and the Department of Radiology, University of Arizona, Tucson, AZ 85724 USA and also with the College of Optical Sciences, University of Arizona, Tucson, AZ 85721 USA.

M. Freed was with the Center for Gamma-Ray Imaging and the Department of Radiology, University of Arizona, Tucson, AZ 85724 USA and the College of Optical Sciences, University of Arizona, Tucson, AZ 85721 USA. She is now with the Food and Drug Administration, Rockville, MD 20857 USA.

J. Hesterman was with the Center for Gamma-Ray Imaging and the Department of Radiology, University of Arizona, Tucson, AZ 85724 USA and the College of Optical Sciences, University of Arizona, Tucson, AZ 85721 USA. He is now with Bioscan, Inc., Washington, DC 20007 USA.

Digital Object Identifier 10.1109/TMI.2007.913241

Most commonly, an auxiliary device called a wavefront sensor [6] is used to analyze the wavefront produced by a bright point source (a so-called guide star) in the field-of-view, and a control element such as a deformable mirror is used to correct the phase distortions. It is also possible, however, to analyze short-exposure images of an unknown extended object and deduce the wavefront distortions [7], [8] or derive various sharpness metrics [9], [10] that can be maximized by changing the signals applied to the control element.

Spurred by the success of astronomical adaptive optics, researchers in another phase-sensitive imaging modality, medical ultrasound, soon began using similar adaptive techniques [11]–[17]. In addition, adaptive control of the pulse sequence in magnetic resonance imaging, first proposed in 1993 by Cao and Levin [18], is currently an active area of study [19]–[25].

Much less has been done with adaptive data acquisition with ionizing radiation. A European consortium [26], [27] is currently developing a digital radiography system that allows active control of beam filtration. Some commercial computed tomography (CT) systems adapt the current in the X-ray tube in response to measured attenuation variations through the body, but with this exception adaptive data-acquisition methods have not been used in single photon emission computed tomography (SPECT), positron emission tomography (PET), or CT.

Adaptive data acquisition in tomography is difficult for several reasons. First, the computational requirements of a reconstruction algorithm may not be compatible with rapid control of acquisition parameters, and second, mechanical control of the imaging configuration is slower and more complex than the electronic control used in adaptive optics, ultrasound, or medical resonance imaging (MRI). More importantly, there has been no compelling argument that adaptation would be helpful and no rationale for performing the adaptation.

Dedicated parallel processors should solve the computational problems, and we have recently proposed and built [28]–[30] a prototype adaptive SPECT system, described briefly in Section II, to address the mechanical issues. The main thrust of the present paper is to present a framework for assessing the performance of adaptive SPECT systems and optimizing the adaptation strategy.

The optimization is based on the paradigm of objective or task-based assessment of image quality in which image quality is defined by the performance of a specific observer performing a specific task on a predetermined class or ensemble of patients [31], [32]. Important tasks include detection of a signal such as a tumor, estimation of parameters associated with the signal, or a combination of detection and estimation. The observer can be a human analyzing the images or a computer algorithm. Particular attention has been given in the literature to linear numerical observers which compute linear functions of the data and use them

to perform the task. Optimal linear observers for both detection and estimation have been studied thoroughly for conventional, nonadaptive imaging systems, and we shall extend that theory to adaptive imaging in this paper. Then we shall use the results to discuss practical adaptation strategies and show how to assess the improvement in task performance.

For context, we begin in Section II by discussing the tradeoffs in SPECT imaging, and then we describe two specific hardware platforms for implementing task-based system optimization in small-animal SPECT: the prototype adaptive system mentioned above and a more conventional but very flexible system called the multimodule, multiresolution (M^3R) system [33]–[35].

In Section III, we present the basic theory of image formation and task performance for adaptive systems, contrasting the results with more familiar ones for conventional, nonadaptive, systems. As we shall see there, implementation of the important linear observers and calculation of figures of merit require knowledge of the first-order and second-order statistics (mean and covariance) of the data. Building on recent work related to astronomical adaptive optics [36], Section IV develops a methodology for deriving these statistical properties of the data.

In Section V, we apply the theory from Section IV to calculating the performance of ideal linear observers with adaptive data. An important conclusion that will emerge is that the ideal linear observers require knowledge of object and data statistics when the objects are known to be consistent with a measured preliminary or scout image. We refer to this set of objects as the posterior ensemble.

In Section VI, we discuss the practical issues in estimating these posterior statistics and assessing the resulting task performance. Section VII summarizes the results presented and discusses some remaining challenges in adaptive imaging.

II. HARDWARE PLATFORMS

A. Controllable Variables and Tradeoffs in Pinhole SPECT

Pinhole SPECT systems are reviewed in [37] and from the viewpoint of small-animal imaging in [38]. Conventional single-camera SPECT systems use just one imaging detector, but several groups [39]–[42] are building small-animal SPECT systems with multiple detectors. At the University of Arizona, we use four imaging detectors in M^3R , eight in our SemiSPECT system [43], 16 in FastSPECT II [44], [45], and 24 in the original FastSPECT.

Conceptually, the simplest approach to pinhole SPECT is to have multiple image detectors, one pinhole per detector and no motion of the object or detector system. In this configuration, which we use routinely for FastSPECT and FastSPECT II, the controllable variables are the pinhole diameter, the distance s_1 from the center of the object to the pinhole plane, and the distance s_2 from the pinhole plane to the detector. The system magnification is defined as s_2/s_1 , and the system sensitivity is controlled by the pinhole diameter and s_1 . Spatial resolution, which depends on pinhole diameter, detector resolution, and magnification, can be traded off for sensitivity and field-of-view.

If we also allow rotation of the object or detector/aperture assembly, the projection angles are additional controllable variables, and the object can be translated parallel to the axis of rotation if desired, for example to perform helical cone-beam scans.

Next, we can consider multiple pinholes for each detector. These pinholes can form an arbitrary pattern, and shutter assemblies can be used to control which pinholes are open at any specific time. If all pinholes are open simultaneously, we have a classic coded-aperture system with a large photon-collection efficiency, but there is ambiguity about which pinhole captured a particular photon. If they are opened individually, each for a fraction of the same total acquisition time, we have lower collection efficiency but no ambiguity. It is also possible to open subsets of the pinholes simultaneously if that should prove advantageous for task performance.

We assume that the number and type of detectors will be fixed in an operational adaptive SPECT system, but that all of the other options listed above will be available, as they are in M^3R and our adaptive prototype.

B. Multimodule, Multiresolution System, M^3R

To investigate the tradeoffs listed above, we constructed the M^3R system. It uses four modular scintillation cameras, each of which has a 12 cm square, 5-mm-thick sodium iodide crystal, 15-mm-thick light guide, and nine photomultiplier tubes. Each camera is individually contained in an aluminum housing with lead shielding.

To allow rapid modification of the hardware configuration, we designed and built a Cerrobend structure that allows for interchangeable pinhole plates. There are five pinhole plate slots for each camera, providing for either magnification or minification depending on plate selection. Varying magnification and/or pinhole configuration is as simple as swapping pinhole plates. There is a central area able to accommodate a mouse or a phantom, and a rotary stage situated beneath the structure allows for object rotation.

Current research with M^3R includes comparison of overlapping with nonoverlapping pinhole images, study of the effect of pinhole diameter on tumor-detection performance, and choice of the optimum aperture for tumor-volume estimation [34], [35].

C. Adaptive SPECT Prototype

A prototype adaptive gamma-ray imaging system for small animals has been designed and constructed [30]. It places camera location, aperture location, aperture size, and integration times under flexible software control. The projections that are acquired and included for tomographic reconstruction have different magnifications, resolutions, sensitivities, fields-of-view, and dwell times.

The current prototype system has a single modular camera, a single on-axis aperture with an array of four selectable diameters, and a vertically-oriented object. The camera position, aperture position, aperture diameter, and object rotation angle are all controlled by motorized stages and thus can adopt an essentially continuous range of object-aperture and aperture-detector configurations allowing magnifications from less than one to about 21. The aperture diameter is selectable between four different pinhole diameters that are currently 0.25, 0.5, 1.0, and 1.5 mm.

In operation, the system acquires an initial scout image in order to obtain preliminary information on the distribution of radiotracer in the particular subject being imaged, and the system parameters are then altered in real time to optimize the system, within practical constraints, for best performance with that subject and for a specific clinical or biomedical task. Fixed geometry data can also be taken with the system to allow a more accurate comparison of fixed versus adaptive geometry optimizations.

Preliminary simulations indicate that the system can achieve sensitivities and resolutions greatly improved over comparable fixed-geometry systems [30].

III. IMAGE FORMATION AND TASK PERFORMANCE

A. Image Formation

The raw projection data in a conventional (nonadaptive) SPECT system consist of a set of M measurements, $\{g_m, m = 1, \dots, M\}$, where m is a composite index specifying a particular detector pixel, a particular projection angle, and possibly the temporal frame in dynamic studies. These measurements constitute the elements of an $M \times 1$ data vector \mathbf{g} .

Image formation in conventional SPECT can be described abstractly as [32]

$$\mathbf{g} = \mathcal{H}\mathbf{f} + \mathbf{n} \quad (1)$$

where \mathbf{f} is the radiotracer distribution being imaged, \mathbf{n} is a zero-mean noise vector, and \mathcal{H} is the system operator mapping the object to the mean image. The object is a function of continuous spatial variables and possibly time, and the image is a discrete vector; we refer to \mathcal{H} , therefore, as a continuous-to-discrete (CD) operator. It is a linear operator if we ignore the effects of detector saturation at high count rate.

It is sometimes useful to separate the effects of the imaging aperture and detector from the patient-dependent effects of attenuation and scatter. We can describe the latter by a linear operator \mathcal{A} , so that the overall system operator is $\mathcal{H} = \mathcal{H}_0\mathcal{A}$ and

$$\mathbf{g} = \mathcal{H}_0\mathcal{A}\mathbf{f} + \mathbf{n}. \quad (2)$$

The operator \mathcal{A} maps the primary-energy radiation emitted by the object to a phase-space distribution function which specifies the photon density as a function of spatial position, photon propagation direction, and energy. The spectrum of energies and propagation directions results from Compton scattering in the object. The effective object $\mathcal{A}\mathbf{f}$ is thus a function of energy as well as spatial and angular variables. The operator \mathcal{H}_0 has an energy sensitivity based on the energy resolution of the detector, and it has some finite exposure time, so any particular element of \mathbf{g} is obtained as an integral of $\mathcal{A}\mathbf{f}$ over spatial position, direction, time, and energy. For details, see [32, Sec. 10.4 and Sec. 16.2].

Image reconstruction maps the $M \times 1$ data vector \mathbf{g} to a discrete image with N voxels, which we can regard as an $N \times 1$ vector $\hat{\mathbf{f}}$. Denoting this mapping by the operator \mathbf{O} , we have

$$\hat{\mathbf{f}} = \mathbf{O}\mathbf{g} = \mathbf{O}[\mathcal{H}_0\mathcal{A}\mathbf{f} + \mathbf{n}]. \quad (3)$$

For linear algorithms such as filtered backprojection, \mathbf{O} is an $N \times M$ matrix, and the overall mapping from object to mean reconstruction, $\mathbf{O}\mathcal{H}_0\mathcal{A}$, is a linear CD operator.

Adaptive systems are inherently nonlinear because the system depends on the object being imaged. In principle, the system could be altered continuously during acquisition as in adaptive optics, but we assume here that there is only one adaptation step; a scout image \mathbf{g}_s is acquired, and some information about the object is derived from it and used to control the system for acquisition of the final data \mathbf{g} .

If the scout image is obtained with a linear system \mathcal{H}_s , we can write

$$\mathbf{g}_s = \mathcal{H}_s\mathbf{f} + \mathbf{n}_s. \quad (4)$$

The scout image is used to modify the characteristics of the final system, so we have

$$\mathbf{g} = \mathcal{H}(\mathbf{g}_s)\mathbf{f} + \mathbf{n}. \quad (5)$$

As in (2), we can decompose the scout-system operator as $\mathcal{H}_s = \mathcal{H}_{0s}\mathcal{A}$, where the patient-specific part, \mathcal{A} is common to both \mathcal{H} and \mathcal{H}_s . The resulting raw data vector from the final system is thus given by

$$\mathbf{g} = \mathcal{H}_0(\mathbf{g}_s)\mathcal{A}\mathbf{f} + \mathbf{n} = \mathcal{H}_0(\mathcal{H}_{0s}\mathcal{A}\mathbf{f} + \mathbf{n}_s)\mathcal{A}\mathbf{f} + \mathbf{n}. \quad (6)$$

The operator-valued function $\mathcal{H}_0(\mathbf{g}_s)$ is called the adaptation rule.

All reconstruction algorithms require information about the system operator (usually in the form of a matrix approximation to \mathcal{H}_0), so the final reconstructed image has the form

$$\hat{\mathbf{f}} = \mathbf{O}(\mathbf{g}_s)\mathbf{g} = \mathbf{O}(\mathbf{g}_s)[\mathcal{H}_0(\mathcal{H}_{0s}\mathcal{A}\mathbf{f} + \mathbf{n}_s)\mathcal{A}\mathbf{f} + \mathbf{n}]. \quad (7)$$

For linear reconstruction algorithms, $\mathbf{O}(\mathbf{g}_s)$ is an $N \times M$ matrix with elements dependent on the scout image \mathbf{g}_s in some way, at least because it uses $\mathcal{H}_0(\mathbf{g}_s)$.

B. Linear Observers for Detection Tasks

A linear observer for a detection task is one that computes a scalar product of the data vector with another vector \mathbf{w} called the template or discriminant function.

If the task is to be performed on the raw data \mathbf{g} from a conventional system, the test statistic has the form

$$t(\mathbf{g}) = c + \sum_{m=1}^M w_m g_m = c + \mathbf{w}^t \mathbf{g} \quad (8)$$

where c is a constant independent of the data, \mathbf{w} is an $M \times 1$ vector, and superscript t denotes transpose. A similar form is used when the observer has access only to a reconstructed image, but there the template is an $N \times 1$ vector to be applied to $\hat{\mathbf{f}}$.

For adaptive systems, the template must depend on the scout image, at least through knowledge of $\mathcal{H}_0(\mathbf{g}_s)$, so (8) is modified to

$$t(\mathbf{g}, \mathbf{g}_s) = c(\mathbf{g}_s) + \mathbf{w}^t(\mathbf{g}_s)\mathbf{g}. \quad (9)$$

The constant c in (8) must now depend on the scout image because different adapted systems have different sensitivities (see Section V-B).

Common choices for the template in the nonadaptive case include the matched filter, the prewhitening matched filter, or Hotelling observer discussed below, and various models based on preprocessing the data through spatial-frequency-selective filters or channels. Noteworthy in this last category is the channelized Hotelling observer (CHO), which is often an excellent predictor of human performance. When the signal location is random, the linear observer can be scanned to different positions, giving rise to scanning Hotelling observers, scanning CHOs, etc.

Each of these choices has a counterpart in the adaptive case. For example, the matched filter for a nonadaptive system is given by $\mathbf{w} = \mathcal{H}\Delta\bar{\mathbf{f}}$, where $\Delta\bar{\mathbf{f}}$ is the difference between the mean objects for the two hypotheses. For an adaptive system, the matched filter generalizes to $\mathcal{H}(\mathbf{g}_s)\Delta\bar{\mathbf{f}}$.

A generally applicable figure of merit (FOM) for detection tasks is the area under the receiver operating characteristic (ROC) curve. A common surrogate FOM, however, is a scalar separability measure or signal-to-noise ratio (SNR) defined by

$$\text{SNR}_{\mathbf{w}}^2 \equiv \frac{[\langle t \rangle_1 - \langle t \rangle_0]^2}{\frac{1}{2}\text{Var}_1(t) + \frac{1}{2}\text{Var}_0(t)} \quad (10)$$

where $\langle t \rangle_j$ is the expectation of the test statistic when hypothesis H_j is true (where, for example, H_0 is tumor-absent and H_1 is tumor-present), and $\text{Var}_j(t)$ is the corresponding variance. If the test statistic is normally distributed under both hypotheses, as it often is for linear discriminants, the area under the ROC curve is a simple monotonic function of the SNR [32].

The SNR as defined in (10) is applicable to both adaptive and nonadaptive systems, so long as the means and variances include all relevant random effects (including the statistics of \mathbf{g}_s in the adaptive case).

The linear observer that maximizes the SNR is called the Hotelling observer [46]–[48]. If this observer operates on the raw data from a nonadaptive system, its template is given by

$$\mathbf{w}_{\text{Hot}} = \mathbf{K}_{\mathbf{g}}^{-1}\mathcal{H}\Delta\bar{\mathbf{f}} \quad (11)$$

where $\mathbf{K}_{\mathbf{g}}$ is the covariance matrix of the data (usually approximately the same under both hypotheses for weak signals) and $\Delta\bar{\mathbf{f}}$ is the difference in the mean object under the two hypotheses. The corresponding FOM for the Hotelling observer is obtained by evaluating (10) for template (11); the result is

$$\text{SNR}_{\text{Hot}}^2 = [\mathcal{H}\Delta\bar{\mathbf{f}}]^t \mathbf{K}_{\mathbf{g}}^{-1} [\mathcal{H}\Delta\bar{\mathbf{f}}]. \quad (12)$$

The form of the Hotelling observer in the case of adaptive imaging is discussed in Section V-B.

C. Linear Observers for Estimation Tasks

If P parameters are to be estimated, we can assemble them into a $P \times 1$ vector $\boldsymbol{\theta}$. For a nonadaptive system, an estimate of this parameter vector, derived from an image vector \mathbf{g} , is denoted $\hat{\boldsymbol{\theta}}(\mathbf{g})$, and a common figure of merit is the ensemble mean-square error defined by

$$\text{EMSE} \equiv \left\langle \left\| \hat{\boldsymbol{\theta}}(\mathbf{g}) - \boldsymbol{\theta} \right\|^2 \right\rangle \quad (13)$$

where the angle brackets denote an average over all sources of randomness in \mathbf{g} and also over some prior probability density on $\boldsymbol{\theta}$ itself.

For an adaptive system, the estimator depends also on the scout image, so we denote it as $\hat{\boldsymbol{\theta}}(\mathbf{g}, \mathbf{g}_s)$ and write the EMSE as

$$\text{EMSE} \equiv \left\langle \left\| \hat{\boldsymbol{\theta}}(\mathbf{g}, \mathbf{g}_s) - \boldsymbol{\theta} \right\|^2 \right\rangle \quad (14)$$

where now the angle brackets must include the average over noise in the scout image.

In the nonadaptive case, an estimator is said to be linear (or more properly, affine), if it has the form

$$\hat{\boldsymbol{\theta}}(\mathbf{g}) = \mathbf{a} + \mathbf{W}^t \mathbf{g} \quad (15)$$

where \mathbf{a} is a $P \times 1$ vector and \mathbf{W}^t is a $P \times M$ matrix. The vector \mathbf{a} can be chosen so that the estimator is globally unbiased, i.e., $\langle \hat{\boldsymbol{\theta}}(\mathbf{g}) \rangle = \bar{\boldsymbol{\theta}}$, where $\bar{\boldsymbol{\theta}}$ is the average of the parameter vector over the ensemble of objects.

The linear estimator that minimizes the EMSE in the nonadaptive case is the generalized Wiener estimator (not to be confused with a Wiener filter), given by [32] and [36]

$$\hat{\boldsymbol{\theta}}_W(\mathbf{g}) = \bar{\boldsymbol{\theta}} + \mathbf{K}_{\boldsymbol{\theta}\mathbf{g}} \mathbf{K}_{\mathbf{g}}^{-1} [\mathbf{g} - \langle \mathbf{g} \rangle] \quad (16)$$

where $\langle \mathbf{g} \rangle$ is the data vector averaged over all sources of randomness, and $\mathbf{K}_{\boldsymbol{\theta}\mathbf{g}}$ is the $P \times M$ cross covariance between the parameter being estimated ($\boldsymbol{\theta}$) and the data.

The resulting EMSE is given by [32]

$$\text{EMSE}_W = \text{tr} \mathbf{K}_{\boldsymbol{\theta}} - \text{tr} \mathbf{K}_{\boldsymbol{\theta}\mathbf{g}} \mathbf{K}_{\mathbf{g}}^{-1} \mathbf{K}_{\boldsymbol{\theta}\mathbf{g}}^t \quad (17)$$

where tr denotes the trace (sum of the diagonal elements) of the matrix, and $\mathbf{K}_{\boldsymbol{\theta}}$ is the prior covariance of $\boldsymbol{\theta}$, based on the prior PDF before any measurements are made.

As with the Hotelling observer for a detection task, the Wiener estimator requires that we know the overall data covariance matrix, and it also requires the cross covariance $\mathbf{K}_{\boldsymbol{\theta}\mathbf{g}}$.

The adaptive counterpart of (15) is [cf. (9)]

$$\hat{\boldsymbol{\theta}}(\mathbf{g}, \mathbf{g}_s) = \mathbf{a}(\mathbf{g}_s) + \mathbf{W}^t(\mathbf{g}_s)\mathbf{g} \quad (18)$$

which is a linear (affine) function of \mathbf{g} but a nonlinear function of \mathbf{g}_s . Expressions for $\mathbf{a}(\mathbf{g}_s)$ and $\mathbf{W}(\mathbf{g}_s)$ are given in Section V-C.

IV. IMAGE STATISTICS

The image \mathbf{g} from an adaptive or nonadaptive system must be regarded as a random vector. In a nonadaptive system described by (1), \mathbf{g} is random because of the noise vector \mathbf{n} , because a

random ensemble of objects \mathbf{f} will be imaged, and also because the system itself can be considered random. Sources of randomness in \mathcal{H} include practical issues such as mechanical errors, electronic drift, and patient motion as well as scatter and attenuation; if we consider random patients to be imaged, and we treat the attenuation and scatter as part of the system, then each randomly chosen patient will correspond to a different random system acting on a different radiotracer distribution \mathbf{f} .

For an adaptive system, there is the additional randomness from the noise \mathbf{n}_s in the scout image. That is, if we repeatedly imaged the same patient through the same scout system, we would get a different final system $\mathcal{H}(\mathbf{g}_s)$ each time. Even if the adaptation rule $\mathcal{H}_0(\mathbf{g}_s)$ were completely deterministic (no noise or mechanical errors in setting the parameters of the acquisition system), the overall system would still be described by a random operator $\mathcal{H}_0(\mathbf{g}_s)\mathcal{A}$ since the patient-specific operator \mathcal{A} is unknown and must be treated as random.

A. Nested Averages

A general method for analyzing all of these sources of randomness is the use of nested conditional averages [31], [32], [36]. To illustrate the procedure and notation, consider a random vector \mathbf{x} that depends on two random parameter vectors \mathbf{a} and \mathbf{b} . By elementary probability theory, the probability density function of \mathbf{x} is given by

$$\text{pr}(\mathbf{x}) = \int d\mathbf{a} \int d\mathbf{b} \text{pr}(\mathbf{x}|\mathbf{a}, \mathbf{b}) \text{pr}(\mathbf{a}|\mathbf{b}) \text{pr}(\mathbf{b}) \quad (19)$$

where the integrals run over all values of all components of the vectors involved.

Random vectors with PDFs of this form are said to be multiply stochastic; they are random (because of measurement noise, for example), even if the parameters \mathbf{a} and \mathbf{b} are fixed and the randomness of each of the parameters also influences the overall PDF.

The expectation of an arbitrary function $\phi(\mathbf{x})$ is defined by

$$\begin{aligned} \langle \phi(\mathbf{x}) \rangle_{\mathbf{x}} &\equiv \int d\mathbf{x} \text{pr}(\mathbf{x}) \phi(\mathbf{x}) \\ &= \int d\mathbf{x} \int d\mathbf{a} \int d\mathbf{b} \text{pr}(\mathbf{x}|\mathbf{a}, \mathbf{b}) \text{pr}(\mathbf{a}|\mathbf{b}) \text{pr}(\mathbf{b}) \phi(\mathbf{x}) \end{aligned} \quad (20)$$

which we can also write as

$$\langle \phi(\mathbf{x}) \rangle_{\mathbf{x}} = \left\langle \left\langle \langle \phi(\mathbf{x}) \rangle_{\mathbf{x}|\mathbf{a}, \mathbf{b}} \right\rangle_{\mathbf{a}|\mathbf{b}} \right\rangle_{\mathbf{b}}. \quad (21)$$

The advantage of this nesting is that the inner expectation might be very simple since both random parameter vectors are fixed.

We can define partial averages of \mathbf{x} itself by

$$\bar{\mathbf{x}}(\mathbf{a}, \mathbf{b}) \equiv \langle \mathbf{x} \rangle_{\mathbf{x}|\mathbf{a}, \mathbf{b}} \quad (22)$$

$$\bar{\bar{\mathbf{x}}}(\mathbf{b}) \equiv \langle \bar{\mathbf{x}}(\mathbf{a}, \mathbf{b}) \rangle_{\mathbf{a}|\mathbf{b}} = \left\langle \langle \mathbf{x} \rangle_{\mathbf{x}|\mathbf{a}, \mathbf{b}} \right\rangle_{\mathbf{a}|\mathbf{b}} \quad (23)$$

$$\bar{\bar{\bar{\mathbf{x}}}} \equiv \langle \bar{\bar{\mathbf{x}}}(\mathbf{b}) \rangle_{\mathbf{b}} = \left\langle \left\langle \langle \mathbf{x} \rangle_{\mathbf{x}|\mathbf{a}, \mathbf{b}} \right\rangle_{\mathbf{a}|\mathbf{b}} \right\rangle_{\mathbf{b}}. \quad (24)$$

With (21), the covariance matrix of \mathbf{x} can be written as

$$\begin{aligned} \mathbf{K}_{\mathbf{x}} &\equiv \left\langle [\mathbf{x} - \bar{\bar{\bar{\mathbf{x}}}}][\mathbf{x} - \bar{\bar{\bar{\mathbf{x}}}}]^t \right\rangle_{\mathbf{x}} \\ &= \left\langle \left\langle \left\langle [\mathbf{x} - \bar{\bar{\bar{\mathbf{x}}}}][\mathbf{x} - \bar{\bar{\bar{\mathbf{x}}}}]^t \right\rangle_{\mathbf{x}|\mathbf{a}, \mathbf{b}} \right\rangle_{\mathbf{a}|\mathbf{b}} \right\rangle_{\mathbf{b}} \end{aligned} \quad (25)$$

where an expression of the form \mathbf{xy}^t denotes an outer product, i.e., if \mathbf{x} and \mathbf{y} are $M \times 1$ vectors, then \mathbf{xy}^t is an $M \times M$ matrix with elements $[\mathbf{xy}^t]_{mn} = x_m y_n$.

Adding and subtracting the two partial means in each factor yields

$$\begin{aligned} \mathbf{K}_{\mathbf{x}} &= \left\langle \left\langle \left\langle [\mathbf{x} - \bar{\mathbf{x}}(\mathbf{a}, \mathbf{b}) + \bar{\mathbf{x}}(\mathbf{a}, \mathbf{b}) - \bar{\bar{\mathbf{x}}}(\mathbf{b}) + \bar{\bar{\mathbf{x}}}(\mathbf{b}) - \bar{\bar{\bar{\mathbf{x}}}}] \right. \right. \\ &\quad \left. \left. \times [\mathbf{x} - \bar{\mathbf{x}}(\mathbf{a}, \mathbf{b}) + \bar{\mathbf{x}}(\mathbf{a}, \mathbf{b}) - \bar{\bar{\mathbf{x}}}(\mathbf{b}) + \bar{\bar{\mathbf{x}}}(\mathbf{b}) - \bar{\bar{\bar{\mathbf{x}}}}]^t \right\rangle_{\mathbf{x}|\mathbf{a}, \mathbf{b}} \right\rangle_{\mathbf{a}|\mathbf{b}} \right\rangle_{\mathbf{b}}. \end{aligned} \quad (26)$$

When we multiply this expression out, all cross-terms vanish identically. For example

$$\begin{aligned} &\left\langle \left\langle \left\langle [\bar{\mathbf{x}}(\mathbf{a}, \mathbf{b}) - \bar{\bar{\mathbf{x}}}(\mathbf{b})] [\bar{\bar{\mathbf{x}}}(\mathbf{b}) - \bar{\bar{\bar{\mathbf{x}}}}]^t \right\rangle_{\mathbf{x}|\mathbf{a}, \mathbf{b}} \right\rangle_{\mathbf{a}|\mathbf{b}} \right\rangle_{\mathbf{b}} \\ &= \left\langle \left\langle [\bar{\mathbf{x}}(\mathbf{a}, \mathbf{b}) - \bar{\bar{\mathbf{x}}}(\mathbf{b})] \right\rangle_{\mathbf{a}|\mathbf{b}} [\bar{\bar{\mathbf{x}}}(\mathbf{b}) - \bar{\bar{\bar{\mathbf{x}}}}]^t \right\rangle_{\mathbf{b}} = 0 \end{aligned} \quad (27)$$

where the last step follows from definition (23). By construction, $\bar{\mathbf{x}}(\mathbf{a}, \mathbf{b}) - \bar{\bar{\mathbf{x}}}(\mathbf{b})$ and $\bar{\bar{\mathbf{x}}}(\mathbf{b}) - \bar{\bar{\bar{\mathbf{x}}}}$ are uncorrelated even if \mathbf{a} and \mathbf{b} are correlated.

Thus, we have the rigorous decomposition

$$\mathbf{K}_{\mathbf{x}} = \bar{\bar{\bar{\mathbf{K}}}}_{\mathbf{x}|\mathbf{a}, \mathbf{b}} + \bar{\bar{\mathbf{K}}}_{\bar{\mathbf{x}}|\mathbf{b}} + \mathbf{K}_{\bar{\bar{\bar{\mathbf{x}}}}} \quad (28)$$

where

$$\begin{aligned} \bar{\bar{\bar{\mathbf{K}}}}_{\mathbf{x}|\mathbf{a}, \mathbf{b}} &= \langle \langle \mathbf{K}_{\mathbf{x}} \rangle_{\mathbf{a}|\mathbf{b}} \rangle_{\mathbf{b}} \\ &= \left\langle \left\langle \left\langle [\mathbf{x} - \bar{\mathbf{x}}(\mathbf{a}, \mathbf{b})][\mathbf{x} - \bar{\mathbf{x}}(\mathbf{a}, \mathbf{b})]^t \right\rangle_{\mathbf{x}|\mathbf{a}, \mathbf{b}} \right\rangle_{\mathbf{a}|\mathbf{b}} \right\rangle_{\mathbf{b}} \end{aligned} \quad (29)$$

$$\begin{aligned} \bar{\bar{\mathbf{K}}}_{\bar{\mathbf{x}}|\mathbf{b}} &= \langle \langle \mathbf{K}_{\bar{\mathbf{x}}|\mathbf{b}} \rangle_{\mathbf{b}} \\ &= \left\langle \left\langle [\bar{\mathbf{x}}(\mathbf{a}, \mathbf{b}) - \bar{\bar{\mathbf{x}}}(\mathbf{b})] [\bar{\mathbf{x}}(\mathbf{a}, \mathbf{b}) - \bar{\bar{\mathbf{x}}}(\mathbf{b})]^t \right\rangle_{\mathbf{a}|\mathbf{b}} \right\rangle_{\mathbf{b}} \end{aligned} \quad (30)$$

$$\mathbf{K}_{\bar{\bar{\bar{\mathbf{x}}}}} = \left\langle [\bar{\bar{\bar{\mathbf{x}}}}(\mathbf{b}) - \bar{\bar{\bar{\bar{\mathbf{x}}}}}] [\bar{\bar{\bar{\mathbf{x}}}}(\mathbf{b}) - \bar{\bar{\bar{\bar{\mathbf{x}}}}}]^t \right\rangle_{\mathbf{b}}. \quad (31)$$

Note that all three terms in (28) are averaged over all three sources of randomness, though some of the averages are before forming the relevant covariance and some after. In (29), for example, $\bar{\bar{\bar{\mathbf{K}}}}_{\mathbf{x}|\mathbf{a}, \mathbf{b}}$ (with no overbars) is averaged over \mathbf{x} conditional on \mathbf{a} and \mathbf{b} , so it is a function of those parameters. The contribution of this conditional average to $\mathbf{K}_{\mathbf{x}}$, i.e., the first term in (19), requires two additional averages, denoted by the two overbars on $\bar{\bar{\bar{\mathbf{K}}}}_{\mathbf{x}|\mathbf{a}, \mathbf{b}}$, and the result does not depend on \mathbf{a} and \mathbf{b} . All three terms in (28) have two overbars somewhere, either on the subscript or on \mathbf{K} , and all three terms involve a total of three

averages, one for each of the overbars and one by the definition of a covariance.

B. Multiply Stochastic Analysis of Nonadaptive Images

To illustrate this formalism and derive some results needed below, we compute the mean image vectors and covariance matrices for nonadaptive systems. We begin with the general expression for a nonadaptive image in (2), and we assume for simplicity that the aperture and detector are well characterized so that \mathcal{H}_0 is known and nonrandom, but the overall system operator \mathcal{H} is still random because of the object-specific part \mathcal{A} . In addition, the noise \mathbf{n} is random, with statistics that depend on the object \mathbf{f} , and \mathbf{f} itself is random when we consider an ensemble of objects.

For a detection task, we are interested in the expectation of \mathbf{g} conditional on hypothesis H_j , which we can write as

$$\bar{\bar{\bar{\mathbf{g}}}}_j \equiv \langle \mathbf{g} \rangle_{\mathbf{g}|H_j} = \left\langle \left\langle \left\langle \mathcal{H}_0 \mathbf{A} \mathbf{f} + \mathbf{n} \right\rangle_{\mathbf{g}|\mathcal{A},\mathbf{f}} \right\rangle_{\mathcal{A}|\mathbf{f},H_j} \right\rangle_{\mathbf{f}|H_j} \quad (32)$$

with the three overbars denoting the three levels of averaging. Note that we do not need to write the inner expectation as conditioned on hypothesis H_j ; when we have fixed \mathcal{A} and \mathbf{f} , it does not matter what ensemble they were drawn from.

The innermost expectation is easy since it is just an average over the zero-mean noise vector \mathbf{n} , which is the only thing that is random when we condition on \mathcal{A} and \mathbf{f} with \mathcal{H}_0 assumed known. Thus, the result of the inner expectation is

$$\bar{\mathbf{g}}(\mathcal{A}, \mathbf{f}) \equiv \langle \mathcal{H}_0 \mathbf{A} \mathbf{f} + \mathbf{n} \rangle_{\mathbf{g}|\mathcal{A},\mathbf{f}} = \mathcal{H}_0 \mathbf{A} \mathbf{f}. \quad (33)$$

The next level of averaging yields

$$\bar{\bar{\mathbf{g}}}_j(\mathbf{f}) \equiv \langle \bar{\mathbf{g}}(\mathcal{A}, \mathbf{f}) \rangle_{\mathcal{A}|\mathbf{f},H_j} = \langle \mathcal{H}_0 \mathbf{A} \mathbf{f} \rangle_{\mathcal{A}|\mathbf{f},H_j} = \mathcal{H}_0 \langle \mathbf{A} \mathbf{f} \rangle_{\mathcal{A}|\mathbf{f},H_j} \quad (34)$$

where the last step follows since \mathcal{H}_0 is a nonrandom linear operator. The subscript notation $\mathcal{A}|\mathbf{f}, H_j$ in (34) indicates the average over \mathcal{A} conditioned on both \mathbf{f} and H_j , so the result depends on both \mathbf{f} and the index j .

Finally, the overall (triple-bar) average is

$$\bar{\bar{\bar{\mathbf{g}}}}_j = \langle \bar{\bar{\mathbf{g}}}_j(\mathbf{f}) \rangle_{\mathbf{f}|H_j} = \mathcal{H}_0 \langle \langle \mathbf{A} \mathbf{f} \rangle_{\mathcal{A}|\mathbf{f},H_j} \rangle_{\mathbf{f}|H_j}. \quad (35)$$

Here, an additional average over \mathbf{f} conditioned on H_j has been performed, so the result depends on j but not \mathbf{f} . Thus, $\bar{\bar{\mathbf{g}}}_j(\mathbf{f})$ loses its argument \mathbf{f} and acquires another overbar in going from (34) to (35).

We might be able to assume that the attenuation and scattering are statistically independent of the activity distribution, though this is not guaranteed since they arise from the same patient. Sometimes, for example, a tumor seen in PET or SPECT may be denser than the surrounding normal tissue. We might also be able to assume that the statistics of \mathcal{A} are the same for signal-present and signal-absent cases. With those assumptions

$$\bar{\bar{\bar{\mathbf{g}}}}_j = \mathcal{H}_0 \langle \mathcal{A} \rangle_{\mathbf{f}|H_j} \quad (36)$$

which shows that the average image for hypothesis H_j in a nonadaptive system is an overall average system operator $\bar{\bar{\mathcal{H}}} \equiv \mathcal{H}_0 \langle \mathcal{A} \rangle$, acting on the average object $\bar{\bar{\mathbf{f}}}_j \equiv \langle \mathbf{f} \rangle_{\mathbf{f}|H_j}$.

For classification tasks we need the covariance matrix of the data under hypothesis H_j . A decomposition like (28) shows that

$$\mathbf{K}_{\mathbf{g}|H_j} = \bar{\bar{\bar{\mathbf{K}}}}_{\mathbf{g}|H_j}^{(\text{noise})} + \bar{\bar{\bar{\mathbf{K}}}}_{\mathbf{g}|H_j}^{(\text{sys})} + \mathbf{K}_{\bar{\bar{\mathbf{g}}}|H_j}^{(\text{obj})}. \quad (37)$$

The first term accounts for the Poisson noise in \mathbf{g} . For a fixed object and system, this noise is Poisson and its covariance matrix is diagonal, and it remains diagonal after averaging \mathcal{A} and \mathbf{f} . When these quantities are statistically independent as in (36), we find [32]

$$\begin{aligned} \bar{\bar{\bar{\mathbf{K}}}}_{\mathbf{g}|H_j}^{(\text{noise})} &= \left\langle \left\langle \left\langle \left\langle [\mathbf{g} - \bar{\mathbf{g}}(\mathcal{A}, \mathbf{f})] [\mathbf{g} - \bar{\mathbf{g}}(\mathcal{A}, \mathbf{f})]^t \right\rangle_{\mathbf{g}|\mathcal{A},\mathbf{f}} \right\rangle_{\mathcal{A}|\mathbf{f},H_j} \right\rangle_{\mathbf{f},H_j} \\ &= \text{diag} [\mathcal{H}_0 \langle \mathcal{A} \rangle_{\mathcal{A}|\mathbf{f},H_j} \langle \mathbf{f} \rangle_{\mathbf{f}|H_j}] \end{aligned} \quad (38)$$

where $\text{diag}[\mathbf{v}]$ denotes a diagonal matrix with vector \mathbf{v} as its diagonal elements.

The second term accounts for the randomness in the imaging system, here the result of attenuation and scatter; it is given by

$$\begin{aligned} \bar{\bar{\bar{\mathbf{K}}}}_{\mathbf{g}|H_j}^{(\text{sys})} &= \left\langle \left\langle \left\langle [\bar{\mathbf{g}}(\mathcal{A}, \mathbf{f}) - \bar{\bar{\mathbf{g}}}(\mathbf{f})] [\bar{\mathbf{g}}(\mathcal{A}, \mathbf{f}) - \bar{\bar{\mathbf{g}}}(\mathbf{f})]^t \right\rangle_{\mathcal{A}|\mathbf{f},H_j} \right\rangle_{\mathbf{f}|H_j} \\ &= \mathcal{H}_0 \left\langle \left\langle [\mathcal{A} - \bar{\mathcal{A}}_j] \mathbf{f} \mathbf{f}^t [\mathcal{A} - \bar{\mathcal{A}}_j]^t \right\rangle_{\mathcal{A}|\mathbf{f},H_j} \right\rangle_{\mathbf{f}|H_j} \\ &\quad \times \mathcal{H}_0^\dagger \end{aligned} \quad (39)$$

where $\bar{\mathcal{A}}_j \equiv \langle \mathcal{A} \rangle_{\mathcal{A}|\mathbf{f},H_j}$ and \mathcal{H}_0^\dagger is the adjoint or backprojection operator (corresponding to the transpose for a real matrix). Note that $\bar{\bar{\bar{\mathbf{K}}}}_{\mathbf{g}|H_j}^{(\text{sys})}$ vanishes if \mathcal{A} is not random, either because attenuation and scatter are negligible (as they might be in mouse studies) or because they have been accurately measured (for example, in dual-modality SPECT/CT). Any inaccuracies in the measurement of \mathcal{A} must, however, be accounted for with the system covariance term.

The final term in the covariance decomposition accounts for object variability. If \mathcal{A} is independent of \mathbf{f} , it is given by

$$\begin{aligned} \mathbf{K}_{\bar{\bar{\mathbf{g}}}|H_j}^{(\text{obj})} &= \left\langle \left[\bar{\bar{\mathbf{g}}}(\mathbf{f}) - \bar{\bar{\bar{\mathbf{g}}}} \right] \left[\bar{\bar{\mathbf{g}}}(\mathbf{f}) - \bar{\bar{\bar{\mathbf{g}}}} \right]^t \right\rangle_{\mathbf{f},H_j} \\ &= \mathcal{H}_0 \bar{\bar{\mathcal{A}}}_j \mathcal{K}_{\mathbf{f}|H_j} \bar{\bar{\mathcal{A}}}_j^\dagger \mathcal{H}_0^\dagger \end{aligned} \quad (40)$$

where $\mathcal{K}_{\mathbf{f}|H_j}$ is the integral operator whose kernel is the spatial autocovariance function of the object [32].

V. STATISTICAL ANALYSIS OF LINEAR OBSERVERS

In this section, we apply the results of Section IV to the analysis of linear discriminants and estimators. For simplicity, we assume throughout that the activity distribution \mathbf{f} and the attenuation and scattering properties contained in \mathcal{A} are statistically

independent, and that \mathbf{f} can be decomposed into statistically independent signal and background components.

For detection problems, we assume further that the signal is weak in the sense that it does not influence variances or covariances appreciably and that an adaptive system responds just to the background. For estimation of parameters associated with the signal, however, we consider the contribution of the signal to the covariances, and we allow the system to adapt to the signal.

A. Linear Discriminants for Nonadaptive Systems

For a nonadaptive system and a general template \mathbf{w} , the numerator in the SNR expression (10) is obtained from (36); with (8) and the assumptions listed above, we have

$$\Delta\langle t \rangle \equiv \langle t \rangle_1 - \langle t \rangle_0 = \mathbf{w}^t \mathcal{H}_0 \overline{\mathbf{A}} \overline{\mathbf{f}}_{\text{sig}} \quad (41)$$

where the constant c in (8) has cancelled.

The denominator of SNR^2 is the variance of t , now assumed to be the same under the two hypotheses; it contains three terms, one for each term in the covariance expression (37):

$$\text{Var}(t) = \mathbf{w}^t \mathbf{K}_g \mathbf{w} = \mathbf{w}^t \left[\overline{\mathbf{K}}_g^{(\text{noise})} + \overline{\mathbf{K}}_g^{(\text{sys})} + \mathbf{K}_g^{(\text{obj})} \right] \mathbf{w} \quad (42)$$

where again the constant c is irrelevant as it does not influence the variance.

Thus

$$\text{SNR}_w^2 = \frac{[\mathbf{w}^t \mathcal{H}_0 \overline{\mathbf{A}} \overline{\mathbf{f}}_{\text{sig}}]^2}{\mathbf{w}^t \mathbf{K}_g \mathbf{w}}. \quad (43)$$

The optimal or Hotelling template \mathbf{w}_{Hot} can be derived by several well-known methods; here we sketch a variational derivation that will be extended to adaptive imaging below.

If we differentiate (43) with respect to \mathbf{w} , use the identities $\partial(\mathbf{w}^t \mathbf{a})/\partial \mathbf{w} = \mathbf{a}$ and $\partial(\mathbf{w}^t \mathbf{A} \mathbf{w})/\partial \mathbf{w} = 2\mathbf{A} \mathbf{w}$, and set the result to zero, we obtain

$$\mathbf{K}_g \mathbf{w} = \left[\frac{\mathbf{w}^t \mathbf{K}_g \mathbf{w}}{\mathbf{w}^t \mathcal{H}_0 \overline{\mathbf{A}} \overline{\mathbf{f}}_{\text{sig}}} \right] \mathcal{H}_0 \overline{\mathbf{A}} \overline{\mathbf{f}}_{\text{sig}}. \quad (44)$$

The factor in large brackets is a scalar, so

$$\mathbf{w} = \left[\frac{\mathbf{w}^t \mathbf{K}_g \mathbf{w}}{\mathbf{w}^t \mathcal{H}_0 \overline{\mathbf{A}} \overline{\mathbf{f}}_{\text{sig}}} \right] \mathbf{K}_g^{-1} \mathcal{H}_0 \overline{\mathbf{A}} \overline{\mathbf{f}}_{\text{sig}}. \quad (45)$$

Moreover, the scalar factor affects the magnitude but not the direction of the vector \mathbf{w} , and SNR_w^2 is invariant to changes in magnitude. Therefore, we can drop the scalar factor and write the optimum template as

$$\mathbf{w}_{\text{Hot}} = \mathbf{K}_g^{-1} \mathcal{H}_0 \overline{\mathbf{A}} \overline{\mathbf{f}}_{\text{sig}}. \quad (46)$$

With this template, the numerator in (43) is the square of the denominator, and we find

$$\text{SNR}_{\text{Hot}}^2 = \overline{\mathbf{f}}_{\text{sig}}^\dagger \overline{\mathbf{A}}^\dagger \mathcal{H}_0^\dagger \mathbf{K}_g^{-1} \mathcal{H}_0 \overline{\mathbf{A}} \overline{\mathbf{f}}_{\text{sig}}. \quad (47)$$

B. Linear Discriminants for Adaptive Systems

With the general adaptive template $\mathbf{w}(\mathbf{g}_s)$ as in (9) and the assumptions listed at the beginning of this section, the difference of means, which appears in the numerator of SNR^2 , becomes [cf. (41)]

$$\Delta\langle t \rangle = \left\langle \left\langle \mathbf{w}^t(\mathbf{g}_s) \mathcal{H}_0(\mathbf{g}_s) \right\rangle_{\mathbf{g}_s | \mathbf{f}_{bg}} \right\rangle_{\mathbf{f}_{bg}} \overline{\mathbf{A}} \overline{\mathbf{f}}_{\text{sig}} \quad (48)$$

where the term $c(\mathbf{g}_s)$ has cancelled on the assumption that the scout system responds just to the background.

For deriving the optimum template, it is convenient to recast (48) by use of Bayes' rule: $\text{pr}(\mathbf{g}_s | \mathbf{f}_{bg}) \text{pr}(\mathbf{f}_{bg}) = \text{pr}(\mathbf{f}_{bg} | \mathbf{g}_s) \text{pr}(\mathbf{g}_s)$. Thus

$$\begin{aligned} \Delta\langle t \rangle &= \left\langle \left\langle \mathbf{w}^t(\mathbf{g}_s) \mathcal{H}_0(\mathbf{g}_s) \right\rangle_{\mathbf{f}_{bg} | \mathbf{g}_s} \right\rangle_{\mathbf{g}_s} \overline{\mathbf{A}} \overline{\mathbf{f}}_{\text{sig}} \\ &= \left\langle \mathbf{w}^t(\mathbf{g}_s) \mathcal{H}_0(\mathbf{g}_s) \right\rangle_{\mathbf{g}_s} \overline{\mathbf{A}} \overline{\mathbf{f}}_{\text{sig}} \end{aligned} \quad (49)$$

where the last step follows because the quantity being averaged does not depend explicitly on \mathbf{f}_{bg} . The remaining average in (49) could be computed, in principle, by choosing many backgrounds at random, constructing the corresponding noisy \mathbf{g}_s by (4), forming $\mathbf{w}^t(\mathbf{g}_s)$ and $\mathcal{H}_0(\mathbf{g}_s)$, and then averaging. The randomness of the background and the noise in the scout image thus appear, but the average is with respect to the resulting distribution of \mathbf{g}_s .

If we put the average over \mathbf{g}_s last, the variance of the test statistic can be written as

$$\text{Var}(t) = \left\langle \left\langle \left[t(\mathbf{g}, \mathbf{g}_s) - \bar{t} \right]^2 \right\rangle_{\mathbf{g} | \mathbf{g}_s} \right\rangle_{\mathbf{g}_s}. \quad (50)$$

We can define a partial average of the test statistic (under hypothesis H_0) by

$$\bar{t}(\mathbf{g}_s) = c(\mathbf{g}_s) + \left\langle \mathbf{w}^t(\mathbf{g}_s) \mathbf{g} \right\rangle_{\mathbf{g} | \mathbf{g}_s} = c(\mathbf{g}_s) + \mathbf{w}^t(\mathbf{g}_s) \overline{\mathbf{g}}(\mathbf{g}_s) \quad (51)$$

where we have used (9). The quantity $\overline{\mathbf{g}}(\mathbf{g}_s)$ is the mean of \mathbf{g} with respect to the conditional density $\text{pr}(\mathbf{g} | \mathbf{g}_s)$, which implicitly includes the conditional randomness from \mathbf{n} , \mathbf{A} , and \mathbf{f} ; it gets a single overbar here because we do not yet need to decompose the conditional average into three separate averages. Explicitly, with our current assumptions

$$\overline{\mathbf{g}}(\mathbf{g}_s) = \mathcal{H}_0(\mathbf{g}_s) \overline{\mathbf{A}} \overline{\mathbf{f}}_{bg}(\mathbf{g}_s) \quad (52)$$

where $\overline{\mathbf{f}}_{bg}(\mathbf{g}_s)$ is the posterior mean background object, obtained from the density $\text{pr}(\mathbf{f}_{bg} | \mathbf{g}_s)$.

Adding and subtracting the partial average $\bar{t}(\mathbf{g}_s)$, as in Section IV-A, we obtain

$$\text{Var}(t) = \left\langle \left\langle \left[t(\mathbf{g}, \mathbf{g}_s) - \bar{t}(\mathbf{g}_s) \right]^2 \right\rangle_{\mathbf{g} | \mathbf{g}_s} \right\rangle_{\mathbf{g}_s} + \left\langle \left[\bar{t}(\mathbf{g}_s) - \bar{t} \right]^2 \right\rangle_{\mathbf{g}_s} \quad (53)$$

where again the cross terms vanish by construction.

The second term in (53) is zero if we take

$$c(\mathbf{g}_s) = -\mathbf{w}^t(\mathbf{g}_s) \overline{\mathbf{g}}(\mathbf{g}_s) \quad (54)$$

in which case $\bar{t}(\mathbf{g}_s) = \bar{t} = 0$. Neither the difference in means nor the first term in (53) depends on $c(\mathbf{g}_s)$, so (54) is the optimum choice; choosing $\bar{t}(\mathbf{g}_s) = 0$ avoids spreading out the density $\text{pr}(t|H_0)$ when considering systems with different $\bar{\mathbf{g}}(\mathbf{g}_s)$, for example systems that collect differing numbers of photons.

With (54), (53) becomes

$$\begin{aligned} \text{Var}(t) &= \left\langle \mathbf{w}^t(\mathbf{g}_s) \left\langle [\mathbf{g} - \bar{\mathbf{g}}(\mathbf{g}_s)] [\mathbf{g} - \bar{\mathbf{g}}(\mathbf{g}_s)]^t \right\rangle_{\mathbf{g}|\mathbf{g}_s} \mathbf{w}(\mathbf{g}_s) \right\rangle_{\mathbf{g}_s} \\ &= \left\langle \mathbf{w}^t(\mathbf{g}_s) \mathbf{K}_{\mathbf{g}|\mathbf{g}_s} \mathbf{w}(\mathbf{g}_s) \right\rangle_{\mathbf{g}_s} \end{aligned} \quad (55)$$

where we have used $[\mathbf{a}^t \mathbf{b}]^2 = \mathbf{a}^t \mathbf{b} \mathbf{b}^t \mathbf{a}$.

We have actually already computed $\mathbf{K}_{\mathbf{g}|\mathbf{g}_s}$ in detail because fixing \mathbf{g}_s removes the randomness in \mathbf{g} arising from the scout system, though it also replaces averages with conditional averages. By appropriately modifying (37)–(40), we obtain

$$\mathbf{K}_{\mathbf{g}|\mathbf{g}_s} = \bar{\mathbf{K}}_{\mathbf{g}|\mathbf{g}_s}^{(\text{noise})} + \bar{\mathbf{K}}_{\mathbf{g}|\mathbf{g}_s}^{(\text{sys})} + \mathbf{K}_{\mathbf{g}|\mathbf{g}_s}^{(\text{obj})} \quad (56)$$

where with the assumptions made at the beginning of this section

$$\bar{\mathbf{K}}_{\mathbf{g}|\mathbf{g}_s}^{(\text{noise})} = \text{diag} \left[\mathcal{H}_0(\mathbf{g}_s) \bar{\mathbf{A}} \mathbf{f}(\mathbf{g}_s) \right] \quad (57)$$

$$\begin{aligned} \bar{\mathbf{K}}_{\mathbf{g}|\mathbf{g}_s}^{(\text{sys})} &= \mathcal{H}_0(\mathbf{g}_s) \left\langle [\mathbf{A} - \bar{\mathbf{A}}] (\mathbf{f} \mathbf{f}^t)_{\mathbf{f}|\mathbf{g}_s} [\mathbf{A} - \bar{\mathbf{A}}]^t \right\rangle_{\mathbf{A}} \\ &\quad \times \mathcal{H}_0^\dagger(\mathbf{g}_s) \end{aligned} \quad (58)$$

and

$$\mathbf{K}_{\mathbf{g}|\mathbf{g}_s}^{(\text{obj})} = \mathcal{H}_0(\mathbf{g}_s) \bar{\mathbf{A}} \mathcal{K}_{\mathbf{f}|\mathbf{g}_s} \bar{\mathbf{A}}^\dagger \mathcal{H}_0^\dagger(\mathbf{g}_s). \quad (59)$$

If the adaptation rule $\mathcal{H}_0(\mathbf{g}_s)$ and the template function $\mathbf{w}(\mathbf{g}_s)$ are specified, then in principle (49) and (56)–(59) can be used to compute the SNR. In practice, actually performing the calculation in this manner would require knowledge of the posterior mean object $\bar{\mathbf{f}}(\mathbf{g}_s)$ and the posterior autocovariance function $\mathcal{K}_{\mathbf{f}|\mathbf{g}_s}$.

We shall return to the question of obtaining or estimating the posterior mean object and covariance function in the next section. For now, we assume that all requisite means and covariances are known, and we use this information to find the Hotelling observer for adaptive imaging. The procedure is to retrace the steps used in Section V-A to derive the nonadaptive Hotelling observer, but with the appropriate adaptive expressions (59) and (55) for the difference in means and variance, respectively, of the test statistic.

In the nonadaptive case, the difference in means had the form $\mathbf{w}^t \mathbf{a}$ and the variance had the form $\mathbf{w}^t \mathbf{A} \mathbf{w}$, so we used the identities $\partial(\mathbf{w}^t \mathbf{a})/\partial \mathbf{w} = \mathbf{a}$ and $\partial(\mathbf{w}^t \mathbf{A} \mathbf{w})/\partial \mathbf{w} = 2\mathbf{A} \mathbf{w}$. In the adaptive case, we must perform a functional or Fréchet derivative (see [32, Sec. 15-3-5]). Differentiating a scalar-valued function with respect to the $M \times 1$ vector \mathbf{w} gave another $M \times 1$ vector, and differentiating with respect to the vector-valued function $\mathbf{w}(\mathbf{g}_s)$ gives another vector-valued function. Specifically [32]

$$\frac{\partial}{\partial \mathbf{w}(\mathbf{g}_s)} \left\langle \mathbf{w}^t(\mathbf{g}_s) \mathbf{a}(\mathbf{g}_s) \right\rangle_{\mathbf{g}_s}$$

$$\begin{aligned} &= \frac{\partial}{\partial \mathbf{w}(\mathbf{g}_s)} \int d\mathbf{g}_s \text{pr}(\mathbf{g}_s) \mathbf{w}^t(\mathbf{g}_s) \mathbf{a}(\mathbf{g}_s) \\ &= \mathbf{a}(\mathbf{g}_s) \text{pr}(\mathbf{g}_s) \end{aligned} \quad (60)$$

and

$$\frac{\partial}{\partial \mathbf{w}(\mathbf{g}_s)} \left\langle \mathbf{w}^t(\mathbf{g}_s) \mathbf{A}(\mathbf{g}_s) \mathbf{w}(\mathbf{g}_s) \right\rangle_{\mathbf{g}_s} = 2\mathbf{A}(\mathbf{g}_s) \mathbf{w}(\mathbf{g}_s) \text{pr}(\mathbf{g}_s). \quad (61)$$

When we now retrace the steps that led to (46), all factors of $\text{pr}(\mathbf{g}_s)$ cancel, and the optimal template is given by

$$\mathbf{w}_{\text{Hot}}(\mathbf{g}_s) = \mathbf{K}_{\mathbf{g}|\mathbf{g}_s}^{-1} \mathcal{H}_0(\mathbf{g}_s) \bar{\mathbf{A}} \mathbf{f}_{\text{sig}}. \quad (62)$$

The randomness of the measurement noise, object, and system are included in $\mathbf{K}_{\mathbf{g}|\mathbf{g}_s}$, but the noise in the scout image does not enter explicitly because $\mathcal{H}_0(\mathbf{g}_s)$ is presumed to be known precisely by the ideal observer.

With this template, the numerator in SNR^2 is again the square of the denominator, and we find [cf. (47)]

$$\text{SNR}_{\text{Hot}}^2 = \bar{\mathbf{f}}_{\text{sig}}^\dagger \bar{\mathbf{A}}^\dagger \left\langle \mathcal{H}_0^\dagger(\mathbf{g}_s) \mathbf{K}_{\mathbf{g}|\mathbf{g}_s}^{-1} \mathcal{H}_0(\mathbf{g}_s) \right\rangle_{\mathbf{g}_s} \bar{\mathbf{A}} \mathbf{f}_{\text{sig}}. \quad (63)$$

If the scout image provides information about either the signal or the attenuation and scatter, this expression remains valid if we simply replace $\bar{\mathbf{A}}$ and $\bar{\mathbf{f}}_{\text{sig}}$ with posterior means and move them inside the expectation brackets.

C. Linear Estimators

In this subsection, we shall derive the optimum linear estimator for the nonadaptive case and then extend the procedure to adaptive imaging. In both cases, a two-step procedure is used. First, the affine term in the general linear estimator [\mathbf{a} in (15) or $\mathbf{a}(\mathbf{g}_s)$ in (18)] is chosen to make the estimator globally unbiased, and then the estimator matrix (\mathbf{W} or $\mathbf{W}(\mathbf{g}_s)$) is chosen to minimize the EMSE.

In the nonadaptive case, the estimator is globally unbiased if

$$\left\langle \left\langle \hat{\boldsymbol{\theta}}(\mathbf{g}) \right\rangle_{\mathbf{g}|\boldsymbol{\theta}} \right\rangle_{\boldsymbol{\theta}} = \bar{\boldsymbol{\theta}} \quad (64)$$

which requires that

$$\hat{\boldsymbol{\theta}}(\mathbf{g}) = \bar{\boldsymbol{\theta}} + \mathbf{W}^t(\mathbf{g} - \bar{\mathbf{g}}) \quad (65)$$

where $\bar{\mathbf{g}} = \langle \langle \mathbf{g} \rangle_{\mathbf{g}|\boldsymbol{\theta}} \rangle_{\boldsymbol{\theta}}$.

The EMSE can be stated as

$$\begin{aligned} \text{EMSE} &= \left\langle \left\langle \left\| \hat{\boldsymbol{\theta}}(\mathbf{g}) - \boldsymbol{\theta} \right\|^2 \right\rangle_{\mathbf{g}|\boldsymbol{\theta}} \right\rangle_{\boldsymbol{\theta}} \\ &= \text{tr} \left\langle \left\langle \left[\hat{\boldsymbol{\theta}}(\mathbf{g}) - \boldsymbol{\theta} \right] \left[\hat{\boldsymbol{\theta}}(\mathbf{g}) - \boldsymbol{\theta} \right]^t \right\rangle_{\mathbf{g}|\boldsymbol{\theta}} \right\rangle_{\boldsymbol{\theta}} \end{aligned} \quad (66)$$

where we have used $\|\mathbf{x}\|^2 = \mathbf{x}^t \mathbf{x} = \text{tr} \mathbf{x} \mathbf{x}^t$.

The expectation in (66) is not a covariance because $\boldsymbol{\theta}$ is not the mean of $\hat{\boldsymbol{\theta}}(\mathbf{g})$. We can, however, add and subtract the mean $\bar{\boldsymbol{\theta}}$ in each factor and do some algebra to obtain

$$\text{EMSE} = \text{tr} \mathbf{W}^t \left\langle \left\langle (\mathbf{g} - \bar{\mathbf{g}})(\mathbf{g} - \bar{\mathbf{g}})^t \right\rangle_{\mathbf{g}|\boldsymbol{\theta}} \right\rangle_{\boldsymbol{\theta}} \mathbf{W}$$

$$\begin{aligned}
 & -2\text{tr} \left\langle \left\langle (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})(\mathbf{g} - \bar{\mathbf{g}})^t \right\rangle_{\mathbf{g}|\boldsymbol{\theta}} \right\rangle_{\boldsymbol{\theta}} \mathbf{W} \\
 & + \text{tr} \left\langle (\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})(\boldsymbol{\theta} - \bar{\boldsymbol{\theta}})^t \right\rangle_{\boldsymbol{\theta}}. \quad (67)
 \end{aligned}$$

Unlike all of the covariance expansions above, the cross term does not cancel in this EMSE expansion.

Each of the terms in (67) is a covariance or cross covariance, and we can write

$$\text{EMSE} = \text{tr} \mathbf{W}^t \mathbf{K}_g \mathbf{W} - 2\text{tr} \mathbf{K}_{\boldsymbol{\theta}g} \mathbf{W} + \text{tr} \mathbf{K}_{\boldsymbol{\theta}}. \quad (68)$$

Differentiating the EMSE with respect to the matrix \mathbf{W} yields

$$\frac{d}{d\mathbf{W}} \text{EMSE} = 2\mathbf{K}_g \mathbf{W} - 2\mathbf{K}_{\boldsymbol{\theta}g}^t \quad (69)$$

where we have used the identities $\partial[\text{tr} \mathbf{W}^t \mathbf{K} \mathbf{W}]/\partial \mathbf{W} = 2\mathbf{K} \mathbf{W}$ (valid if \mathbf{K} is real and symmetric) and $\partial \text{tr}[\mathbf{B} \mathbf{W}]/\partial \mathbf{W} = \mathbf{B}^t$. The derivative in (69) must be zero for the optimum \mathbf{W} , and it follows that the Wiener estimator is indeed given by (16) and the optimized EMSE is given by (17).

In the adaptive case, a general estimator of the form (18) is globally unbiased if

$$\left\langle \hat{\boldsymbol{\theta}}(\mathbf{g}, \mathbf{g}_s) \right\rangle_{\mathbf{g}, \mathbf{g}_s} = \langle \mathbf{a}(\mathbf{g}_s) \rangle_{\mathbf{g}_s} + \langle \mathbf{W}^t(\mathbf{g}_s) \mathbf{g} \rangle_{\mathbf{g}, \mathbf{g}_s} = \bar{\boldsymbol{\theta}}. \quad (70)$$

This condition will hold if

$$\hat{\boldsymbol{\theta}}(\mathbf{g}, \mathbf{g}_s) = \bar{\boldsymbol{\theta}}(\mathbf{g}_s) + \mathbf{W}^t(\mathbf{g}_s) [\mathbf{g} - \bar{\mathbf{g}}(\mathbf{g}_s)] \quad (71)$$

where

$$\bar{\mathbf{g}}(\mathbf{g}_s) = \langle \langle \mathbf{g} \rangle_{\mathbf{g}|\mathbf{g}_s} \rangle_{\boldsymbol{\theta}|\mathbf{g}_s} \quad (72)$$

and similarly, $\bar{\boldsymbol{\theta}}(\mathbf{g}_s) = \langle \boldsymbol{\theta} \rangle_{\boldsymbol{\theta}|\mathbf{g}_s}$. Notice that the outer average on the right-hand side of (72) must be calculated with the posterior PDF of $\boldsymbol{\theta}$ conditioned on the scout data \mathbf{g}_s .

The EMSE of a globally unbiased estimator for adaptive data is given by [cf. (67)]

$$\text{EMSE} = \text{tr} \langle \mathbf{W}^t(\mathbf{g}_s) \mathbf{K}_{g|\mathbf{g}_s} \mathbf{W}(\mathbf{g}_s) \rangle_{\mathbf{g}_s} - 2\text{tr} \langle \mathbf{K}_{\boldsymbol{\theta}g|\mathbf{g}_s} \mathbf{W}(\mathbf{g}_s) \rangle_{\mathbf{g}_s} + \text{tr} \mathbf{K}_{\boldsymbol{\theta}}. \quad (73)$$

The Fréchet derivative now yields a matrix-valued function [cf. (60) and (61)]

$$\frac{\partial}{\partial \mathbf{W}(\mathbf{g}_s)} \text{EMSE} = \left[2\mathbf{K}_{g|\mathbf{g}_s} \mathbf{W}(\mathbf{g}_s) - 2\mathbf{K}_{\boldsymbol{\theta}g|\mathbf{g}_s}^t \right] \text{pr}(\mathbf{g}_s). \quad (74)$$

This derivative will be zero if $\mathbf{W}(\mathbf{g}_s) = \mathbf{K}_{g|\mathbf{g}_s}^{-1} \mathbf{K}_{\boldsymbol{\theta}g|\mathbf{g}_s}^t$, which means that the Wiener estimator is

$$\hat{\boldsymbol{\theta}}(\mathbf{g}, \mathbf{g}_s) = \bar{\boldsymbol{\theta}}(\mathbf{g}_s) + \mathbf{K}_{\boldsymbol{\theta}g|\mathbf{g}_s} \mathbf{K}_{g|\mathbf{g}_s}^{-1} [\mathbf{g} - \bar{\mathbf{g}}(\mathbf{g}_s)]. \quad (75)$$

This estimator is identical to the nonadaptive Wiener estimator of (16) except that all averages are computed with PDFs conditional on the scout image.

The resulting EMSE is given by [cf. (17)]

$$\text{EMSE} = \text{tr} \mathbf{K}_{\boldsymbol{\theta}} - \text{tr} \left\langle \mathbf{K}_{\boldsymbol{\theta}g|\mathbf{g}_s} \mathbf{K}_{g|\mathbf{g}_s}^{-1} \mathbf{K}_{\boldsymbol{\theta}g|\mathbf{g}_s}^t \right\rangle_{\mathbf{g}_s}. \quad (76)$$

To get more explicit expressions for the means and covariances in (75) and (76), let us assume that the object can be separated into background and signal parts and that the parameter $\boldsymbol{\theta}$ is associated only with the signal (perhaps a tumor). Thus, [cf. (21) **[AUTHOR: CORRECT?]**],

$$\mathbf{f} = \mathbf{f}_{bg} + \mathbf{f}_{\text{sig}}(\boldsymbol{\theta}). \quad (77)$$

Continuing to assume that \mathcal{A} is independent of the object and not influenced by the scout image, we get

$$\bar{\mathbf{g}}(\mathbf{g}_s) = \mathcal{H}_0(\mathbf{g}_s) \bar{\mathcal{A}} [\bar{\mathbf{f}}_{bg}(\mathbf{g}_s) + \bar{\mathbf{f}}_{\text{sig}}(\mathbf{g}_s)] \quad (78)$$

where $\bar{\mathbf{f}}_{\text{sig}}(\mathbf{g}_s) = \langle \mathbf{f}_{\text{sig}}(\boldsymbol{\theta}) \rangle_{\boldsymbol{\theta}|\mathbf{g}_s}$.

The posterior cross covariance, which expresses the sensitivity of the data to changes in the parameters, is given by

$$\begin{aligned} \mathbf{K}_{\boldsymbol{\theta}g|\mathbf{g}_s} &= \left\langle \left[\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}(\mathbf{g}_s) \right] \left\{ \mathcal{H}_0(\mathbf{g}_s) \bar{\mathcal{A}} [\mathbf{f}_{\text{sig}}(\boldsymbol{\theta}) - \bar{\mathbf{f}}_{\text{sig}}(\mathbf{g}_s)] \right\}^t \right\rangle_{\boldsymbol{\theta}|\mathbf{g}_s} \\ &= \mathbf{K}_{\boldsymbol{\theta}f|\mathbf{g}_s} \bar{\mathcal{A}}^\dagger \mathcal{H}_0^\dagger(\mathbf{g}_s) \end{aligned} \quad (79)$$

where $\mathbf{K}_{\boldsymbol{\theta}f|\mathbf{g}_s}$ is an operator defined by

$$\mathbf{K}_{\boldsymbol{\theta}f|\mathbf{g}_s} \equiv \left\langle \left[\boldsymbol{\theta} - \bar{\boldsymbol{\theta}}(\mathbf{g}_s) \right] [\mathbf{f}_{\text{sig}}(\boldsymbol{\theta}) - \bar{\mathbf{f}}_{\text{sig}}(\mathbf{g}_s)]^\dagger \right\rangle_{\boldsymbol{\theta}|\mathbf{g}_s}. \quad (80)$$

This operator maps a function of continuous spatial variables in object space to a $P \times 1$ vector in parameter space.

Finally, the posterior covariance $\mathbf{K}_{g|\mathbf{g}_s}$, which accounts for measurement noise, system randomness and object variability, can still be expressed by (56) if we choose to include randomness in both the background and the parameterized signal in the object term. If, however, we use separate posterior averaging steps for background and signal, (56) takes the form

$$\mathbf{K}_{g|\mathbf{g}_s} = \bar{\mathbf{K}}_{g|\mathbf{g}_s}^{(\text{noise})} + \bar{\mathbf{K}}_{g|\mathbf{g}_s}^{(\text{sys})} + \mathbf{K}_{\bar{\mathbf{g}}|\mathbf{g}_s}^{(bg)} + \mathbf{K}_{\bar{\mathbf{g}}|\mathbf{g}_s}^{(\text{sig})}. \quad (81)$$

VI. IMPLEMENTATION AND ASSESSMENT

In this section, we examine some of the issues that arise in implementing the mathematics above in practice and assessing the improvements attained.

A. Goals and Figures of Merit

The goal of adaptive imaging is to choose an adaptation rule $\mathcal{H}_0(\mathbf{g}_s)$ in order to maximize the performance of a detection or estimation task. For a detection task, we take the performance figure of merit as the Hotelling SNR², given by (63) along with the posterior covariance decomposition of (56)–(59). To simplify the expressions, we ignore the system covariance term (58) by assuming the attenuation and scatter are known to the observer, and we include the attenuation and scatter in the adapted system operator, $\mathcal{H}(\mathbf{g}_s) = \mathcal{H}_0(\mathbf{g}_s) \mathcal{A}$. Then we can rewrite the FOM as

$$\text{FOM}_{\text{det}} = \text{SNR}_{\text{Hot}}^2 = \langle \text{FOM}_{\text{det}}(\mathbf{g}_s) \rangle_{\mathbf{g}_s} \quad (82)$$

where

$$\text{FOM}_{\text{det}}(\mathbf{g}_s) = [\mathcal{H}(\mathbf{g}_s)\bar{\mathbf{f}}_{\text{sig}}]^t \left\{ \text{diag} [\mathcal{H}(\mathbf{g}_s)\bar{\mathbf{f}}_{bg}(\mathbf{g}_s)] + \mathcal{H}(\mathbf{g}_s)\mathcal{K}_{\mathbf{f}|\mathbf{g}_s}\mathcal{H}^\dagger(\mathbf{g}_s) \right\}^{-1} [\mathcal{H}(\mathbf{g}_s)\bar{\mathbf{f}}_{\text{sig}}]. \quad (83)$$

The goal of adaptation for an estimation task is to minimize the EMSE given in (76). The first term in that equation, $\text{tr}\mathbf{K}_\theta$, expresses the prior knowledge, before the scout image, so we want to make $\text{tr}\langle \mathbf{K}_{\theta\mathbf{g}|\mathbf{g}_s}\mathbf{K}_{\mathbf{g}|\mathbf{g}_s}^{-1}\mathbf{K}_{\theta\mathbf{g}|\mathbf{g}_s}^t \rangle$ as large as possible, and this term can be taken as the FOM for estimation. Thus, by analogy to (82) and (83), we write

$$\text{FOM}_{\text{est}} = \langle \text{FOM}_{\text{est}}(\mathbf{g}_s) \rangle_{\mathbf{g}_s} \quad (84)$$

where with (79) and (81)

$$\begin{aligned} \text{FOM}_{\text{est}}(\mathbf{g}_s) &= \text{tr}\mathbf{K}_{\theta\mathbf{g}|\mathbf{g}_s}\mathbf{K}_{\mathbf{g}|\mathbf{g}_s}^{-1}\mathbf{K}_{\theta\mathbf{g}|\mathbf{g}_s}^t \\ &= \text{tr} \left[\mathcal{K}_{\theta\mathbf{f}|\mathbf{g}_s}\mathcal{H}^\dagger(\mathbf{g}_s) \right] \\ &\quad \times \left\{ \text{diag} [\mathcal{H}(\mathbf{g}_s) (\bar{\mathbf{f}}_{bg}(\mathbf{g}_s) + \bar{\mathbf{f}}_{\text{sig}}(\mathbf{g}_s))] \right. \\ &\quad \left. + \mathcal{H}(\mathbf{g}_s) \left[\mathcal{K}_{\mathbf{f}|\mathbf{g}_s}^{(bg)} + \mathcal{K}_{\mathbf{f}|\mathbf{g}_s}^{(\text{sig})} \right] \mathcal{H}^\dagger(\mathbf{g}_s) \right\}^{-1} \\ &\quad \times \left[\mathcal{H}(\mathbf{g}_s)\mathcal{K}_{\theta\mathbf{f}|\mathbf{g}_s}^\dagger \right]. \end{aligned} \quad (85)$$

The trace operator (tr) applies to the $P \times P$ matrix that follows, with P being the number of parameters to be estimated. (Recall that $\mathcal{K}_{\theta\mathbf{f}|\mathbf{g}_s}\mathcal{H}^\dagger(\mathbf{g}_s)$ is a $P \times M$ matrix, and of course the inverse covariance matrix is $M \times M$.)

A scout imaging system and an adaptation rule will thus be effective in increasing task performance, compared to a non-adaptive system, if they either 1) increase the signal $\mathcal{H}(\mathbf{g}_s)\bar{\mathbf{f}}_{\text{sig}}$ in (83), 2) increase the sensitivity of the data to the parameters as expressed by $\mathcal{K}_{\theta\mathbf{f}|\mathbf{g}_s}\mathcal{H}^\dagger(\mathbf{g}_s)$ in (85), or 3), reduce one of the covariance terms in either equation. With these observers, there are no other ways to improve task performance.

B. Prior Ensemble Information

Before making any measurements on the particular subject of interest, we will almost always have available some information about the ensemble from which that subject is drawn. One important piece of information is the prior mean $\bar{\mathbf{f}}$, which enters into the noise covariance term [see (38)]. This 3-D function represents the average distribution of a given tracer for all patients or animals that might receive it. If we have previously imaged many subjects with this tracer and a well sampled but nonadaptive SPECT system, we can average the images to get an estimate of $\bar{\mathbf{f}}$ on a voxel grid. Because of the large intersubject variability, we anticipate that this estimate will be diffuse and relatively nondescriptive of the particular subject to which we wish to adapt.

We might also have some prior information about the fine structure or texture of the tracer distribution. Considerable work has been done on describing the fine structure seen in medical images, using models such as the lumpy or clustered lumpy backgrounds, Markov random fields, and wavelet-based descriptions (see [32, Ch. 8] and references cited therein). If one of these models is known to be applicable to the tracer under consideration, it will give a good approximation to the object

covariance term in (40). If the model is applicable in general but contains some unknown parameters, a method described by Kupinski *et al.* [49] can be used to estimate the parameter values from a set of sample images.

Of course object covariance also has contributions from large-scale variations in addition to the fine structure. Most texture models in the literature are applicable to spatially stationary random processes, but a quasistationary model will usually be an improvement. The approach here is to express the object autocovariance function, which is the kernel of the integral operator in (40), in the approximate form (see [32, Sec. 8.2])

$$k(\mathbf{r}_1, \mathbf{r}_2) = a(\Delta\mathbf{r})b(\mathbf{r}_0) \quad (86)$$

where $\mathbf{r}_0 = (1/2)(\mathbf{r}_1 + \mathbf{r}_2)$ and $\Delta\mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$. The function $a(\Delta\mathbf{r})$ describes the rapidly varying fine structure and can be represented by mathematical texture models as mentioned above, while $b(\mathbf{r}_0)$, representing large-scale variations, can be estimated from sample images.

In tumor-detection problems, we will probably have a good idea what a tumor looks like, so we will have some representation for the function $\bar{\mathbf{f}}_{\text{sig}}$. Similarly, for estimation of tumor parameters, we will necessarily have to assume some prior mean and covariance of the parameters in order to employ Wiener estimation in the first place.

C. Scout Strategies

A scout image should use up only a small fraction of the available imaging time, so it will necessarily be limited in the number of projection angles and/or the exposure time per projection compared to the final, adapted system. To obtain reasonable image statistics with this limitation, it will probably be necessary to use relatively large pinholes. With these restrictions, what can we learn about the object?

One thing we can do with the scout system is to refine our estimate of the prior mean $\bar{\mathbf{f}}$, which appears in the noise covariance term. If we take short-exposure projection images with large pinholes and an adequate number of projection angles, a heavily smoothed reconstruction should provide a much better approximation to \mathbf{f} than the prior mean $\bar{\mathbf{f}}$. Of course we do not know the actual object, but we can regard the smoothed scout reconstruction as an estimate of the posterior mean $\bar{\mathbf{f}}(\mathbf{g}_s)$ (where the word ‘‘posterior’’ in this context means after the scout image but still prior to the final measurement).

We can also try to refine our estimate of the object covariance terms. For example, with the quasistationary model (82), we can estimate the long-range factor $b(\mathbf{r}_0)$ from a smoothed scout reconstruction, thereby reducing the object covariance term even if we rely on the prior model for $a(\Delta\mathbf{r})$.

Another thing we can get from a scout image is an estimate of the sparsity of the tracer distribution. A sparse object in nuclear medicine is one where the tracer is concentrated within a volume that is small compared to the field of view of the system.

If we know the object is sparse, it is probably advantageous to use a large number of pinholes, thereby reducing the relative size of the noise term in the covariance and increasing the norm of the signal, but the performance advantage is much less for large, diffuse objects [50]–[53]. In addition, sparse objects can

be more readily reconstructed from a few projections than can nonsparse ones [54], [55], so fewer projection angles can be used in that case.

A quantitative measure of degree of object sparseness is its entropy, defined by

$$S = - \int d^3r p(\mathbf{r}) \ln [p(\mathbf{r})] \quad (87)$$

where $p(\mathbf{r}) \equiv f(\mathbf{r}) / \int d^3r' f(\mathbf{r}')$ is the tracer distribution normalized as a probability density function. Estimates of S can be obtained from a low-resolution reconstruction or directly from the projection images in the scout data set.

From a reconstructed scout image, we can also estimate the overall support of the tracer distribution, defined as the set of points over which $f(\mathbf{r})$ is nonzero, or where it exceeds some small threshold. If this support set is small, we can use large magnifications in the final image without truncating the projections, and if we know the 3-D coordinates of the centroid of the support, we can choose the pinhole location in each projection view to center the support on the detector.

For a pure detection problem, we cannot use the scout image to localize the signal as this would imply that the task could be performed from the scout data alone, obviating the adaptation. If, however, we wish to estimate the volume or activity of a previously detected tumor, say to assess response to therapy, we can use the scout image to get an estimate of the tumor location. Then we can adjust the magnification and pinhole locations so that the tumor projections optimally fill the detector field, thereby increasing cross-covariance $\mathbf{K}_{\theta\mathbf{g}}$ and reducing the parameter covariance term in (80). For this task, we might allow truncated projections so long as the final adapted system operator $\mathcal{H}_0(\mathbf{g}_s)$ is correctly modeled and used in the estimation.

A different scout strategy can be used for rapid dynamic studies, where the task might be estimation of the activity in a region of interest as a function of time. We can administer a low-activity bolus of tracer and image it with large pinholes and a small number of projection angles (maybe just two) in order to determine the general trajectory of the bolus and how rapidly the activity varies in the region of interest. With this prior information, we can configure the system for optimal estimation performance and then administer a larger bolus for the final data acquisition.

D. Modes of Adaptation

There are several distinct ways in which the information from the scout images can be used to optimize the final imaging system. All of them are based on approximating the system operator \mathcal{H} with a matrix \mathbf{H} and making further approximations in order to compute figures of merit rapidly. These computational issues are discussed below, but here we discuss the kinds of system variations to be considered in choosing the final system.

1) *Choice Among Known Systems*: First, we might have some relatively small number of candidate systems and merely want to choose the one that will give best task performance for a particular subject. For example, if the only variable is the magnification, we can precompute the \mathbf{H} matrices for each and then compute the corresponding figures of merit.

2) *Linear Combinations of Known Systems*: If we have multiple precalibrated systems with known \mathbf{H} , we can consider using combinations of them. There are two distinct forms of linear combination, which we can refer to as operator addition and concatenation. As an example, suppose we have pinhole plates with shutters, as mentioned in Section II for the adaptive prototype. Each pinhole in a plate can have its own, precomputed \mathbf{H} matrix. If some subset of the pinholes is open for the entire exposure, we simply add the matrices. If, however, the shutters operate L times during the exposure and J projection angles are used, and the detector has K pixels and the object representation uses N voxels, the final concatenated matrix is $JKL \times N$.

Another example of concatenation is when we choose some combination of projection angles, each of which has its own pre-computed \mathbf{H} . If the object representation consists of N voxels, and J angles are selected for a detector with K pixels, then each individual matrix \mathbf{H}_j is $K \times N$ but the overall matrix with concatenated rows is $JK \times N$.

3) *One-Parameter Continuous Optimization*: As noted in Section II-A, key parameters in pinhole SPECT include the pinhole diameter and the magnification. If multiple pinholes are used but they are all constrained to have the same diameter D , then D is an important single parameter to vary for best task performance. Similarly, if all projection angles are constrained to the same magnification, then that is an important parameter to vary.

4) *Multiparameter Optimization*: In the example just given, the pinhole diameter and the magnification can be varied together, giving us a two-parameter optimization space. Similarly, if we use J projection angles and vary the exposure time and magnification at each, we have a total of $2J$ parameters to vary, greatly increasing the computational complexity of the optimization.

5) *Empirical Rules*: Rather than trying to find a true optimum system configuration, we can also develop empirical rules that allow us to choose broad system characteristics. One example already mentioned is to choose the degree of multiplexing based on some measure of sparsity of the object. Another example would be a rule for choosing the allowable degree of image truncation for an estimation task where we try to fill the detector field of view by placing a pinhole close to the signal of interest [30].

E. Computational Issues

Methods for computing task-based figures of merit in non-adaptive imaging systems are discussed in [32, Ch. 14]. The greatest difficulty is that the covariance matrices are enormous ($M \times M$, where M is the total number of measurements, $\sim 10^5 - 10^6$ in SPECT), but there are two important techniques that allow a huge dimensionality reduction, often with little loss of accuracy in the final estimates of figures of merit.

First, we can use spatial-frequency-selective channels, similar to those known to exist in the human visual system, and use the channel outputs as new data values. The dimensionality reduction can be extreme; with so-called efficient channels, as few as 5–10 channels sometimes give excellent approximations for figure of merit [56], [57], and the size of the covariance matrix is then just 5×5 or 10×10 .

The second trick, especially useful for the object term in the covariance expansion, is to use a simulation program to create noise-free images of randomly generated objects through a specified \mathbf{H} matrix. The object covariance is then estimated by a low-rank sample covariance matrix, but the overall covariance matrix remains invertible because of the diagonal noise term. The Woodbury matrix inversion lemma [58] can then be used to reduce the size of the matrix to be inverted to $S \times S$, where S is the number of sample images (~ 100 – 1000), instead of $M \times M$.

The computational tasks needed to optimize system performance for an adaptive system are not fundamentally different from those required for any task-based system optimization, but they have to be done very rapidly in order to be able to make system changes before the patient leaves the room, and they have to be done many times in any iterative optimization scheme. The tools that give us any hope of meeting these demands are precomputation, interpolation, approximation, and parallelization.

1) *Precomputation*: As noted in Section VI-D, \mathbf{H} matrices for candidate systems can be precomputed or measured, but other key components of the FOMs can be precomputed as well. For example, a voxel representation of the object covariance function of (86) can be computed ahead of time, and the slowly varying factor can then be modified after the scout data are obtained.

Similarly, the cross-covariance operator $\mathcal{K}_{\theta_f|g_s}$ defined in (80) can be computed in a voxel representation and stored as a set of P images, one for each of the parameters to be estimated. These images are readily modified after preliminary parameter estimates are obtained from the scout data; for example, in tumor-volume estimation, a refined estimate of the location of a tumor requires only a recentering of the images that make up the rows of $\mathcal{K}_{\theta_f|g_s}$.

2) *Interpolation*: Suppose the \mathbf{H} matrix is known from precomputation or measurements for a sparse set of projection angles, but the optimization algorithm needs matrices for intermediate angles. It is easy to see that conventional linear or spline interpolation does not work by considering a single pinhole; a point in object space produces a blob at location \mathbf{r}_1 on the detector for projection angle ϕ_1 and another blob of different size or shape at location \mathbf{r}_2 on the detector for projection angle ϕ_2 . Linear interpolation would produce a linear combination of the two blobs, while the correct \mathbf{H} matrix for an intermediate angle would exhibit a single blob at an intermediate location.

Chen [45], [49] has developed two algorithms to overcome this problem and provide highly accurate interpolated \mathbf{H} matrices. One of them describes the blobs parametrically and interpolates the parameters, and the other uses 2-D Fourier transforms of the blob images and interpolates the Fourier magnitude and phase.

Similarly, if \mathbf{H} is known for a restricted set of pinhole diameters, interpolation schemes can be devised to obtain it for intermediate diameters.

These interpolation schemes are particularly valuable when we wish to optimize the system over a continuous range of parameters. The only alternative to interpolation of the \mathbf{H} matrix in this case is to represent it a parametric form and interpolate

the parameters. For example, we could use a geometrical model parameterized by the distances from the center of rotation to the pinhole plate and from the pinhole plate to the detector. When combined with a measured model of the scintillation detector, this geometrical model would then provide an overall \mathbf{H} that is readily evaluated for any desired value of the two distances. This approach requires a high degree of mechanical precision (which we indeed tried to build in to the adaptive prototype) and distortion-free performance of the scintillation camera (which we achieve by using nearly unbiased maximum-likelihood position estimation).

3) *Approximation*: After the scout data are obtained, the goal is to modify the final imaging system so as to optimize the figure of merit (83) or (85). Doing so requires not only operational approximations to the posterior means and covariances, but also very rapid ways of performing the necessary inverses. As with nonadaptive systems, channels and use of the matrix inversion lemma can be enormously helpful, but further computational benefit can be obtained by replacing the actual FOM, (83) or (85), with some surrogate function which is easier to calculate. If it can be established (by offline simulation studies, for example) that system modifications that increase the surrogate function also increase the actual task performance, then the real-time computational requirements are eased.

A prime example of this strategy is when the scout image is obtained with large pinholes and short exposures at a large number of projection angles. A rapid, highly regularized reconstruction can be performed to get a blurred version of object, and the surrogate optimization function can be obtained by taking this reconstruction as the mean object and neglecting the object covariance term completely. In other words, a background known exactly (BKE) task is used as a surrogate for a BKS (background known statistically) task. The requisite matrix inverse in this case is very easy because the noise covariance matrix is diagonal and readily computed for any proposed system configuration.

Other potential surrogate functions include expressions based on a Neumann series expansion of the overall covariance (see [32, Ch. 14] and ones based on Fisher information [60]; neither of these methods requires inversion of nondiagonal matrices.

4) *Parallelization*: The search for an optimum—or at least improved—system configuration requires evaluating the figure of merit for many candidate systems. If multiple processors with adequate memory are available online, the computations can be performed in parallel.

F. Performance Validation

The figures of merit in (82) and (84) are defined as posterior averages, where here the term posterior refers to ensembles of objects consistent with the scout image. Averages with respect to this posterior ensemble are decidedly Bayesian in the sense that we cannot know the PDF $\text{pr}(\mathbf{f}|g_s)$ analytically; we must regard it as a measure of our degree of belief in the properties of the object after we have obtained the scout image.

In a pure Bayesian paradigm, only this subjective density would be used in assessing different adaptation strategies. By contrast, a pure frequentist view would require that all probabilities be regarded as frequencies of occurrence, which is surely

not possible for either the prior $\text{pr}(\mathbf{f})$ or the posterior $\text{pr}(\mathbf{f}|\mathbf{g}_s)$, both of which are infinite-dimensional densities.

To resolve this dilemma, we adopt a hybrid view espoused in [32]; we allow subjective or approximate densities in performing the task, justifying them where we can by appeal to experiment, but in the end we estimate performance metrics by repeated experiments.

The experiments are best carried out in simulation. If we have an accurate model for the prior object class, including variations in object size and shape, object texture or signal parameters, we can create multiple sample objects from this prior on a voxel grid. We can then evaluate a particular combination of scout system \mathcal{H}_s and adaptation rule $\mathcal{H}_0(\mathbf{g}_s)$ by generating scout and final images for each sample object and performing the task. For a detection task that means generating many sample objects with and without signals, computing the test statistic $t(\mathbf{g}, \mathbf{g}_s)$ for each, and then either estimating SNR^2 from the definition (10) or varying the decision threshold and generating a sample ROC curve. For an estimation task, many random backgrounds would be generated, each with a signal having a parameter vector $\boldsymbol{\theta}$ drawn from the prior $\text{pr}(\boldsymbol{\theta})$, and forming an estimate $\hat{\boldsymbol{\theta}}(\mathbf{g}, \mathbf{g}_s)$ for each; the EMSE can then be estimated directly from its definition, (14).

Obviously the approach just outlined requires an enormous amount of computation, but it is offline and does not have to be done as part of the adaptation process.

VII. SUMMARY AND CONCLUSION

Task-based methods provide rigorous definitions of image quality, which can be used to optimize a system for a predefined ensemble of patients. In this paper, we have shown that preliminary scout measurements on a given patient can be used to narrow down the ensemble and hence reduce task-related uncertainty.

We considered in detail two specific linear observers, the ideal linear detector known as the Hotelling observer and the ideal linear estimator known as the generalized Wiener estimator. For nonadaptive systems, we showed that the performance of these observers is limited by three sources of randomness in the data, which in SPECT are mainly Poisson measurement noise, randomness in the radionuclide distribution being imaged, and unknown attenuation and scatter in the patient's body. We showed that the data covariance matrix could be decomposed rigorously into a sum of three component matrices, one associated with each source of randomness.

With this decomposition, we derived expressions for the figures of merit for the Hotelling and Wiener observers for nonadaptive systems, and then we showed how to extend the derivation to account for the adaptation. The expressions found for the observers and their figures of merit for the adaptive case were surprisingly similar to those for nonadaptive SPECT, but the mean vectors and covariance matrices had to be interpreted as posterior averages, conditional on the scout image, in the adaptive case.

Some of the practical issues in implementing adaptive SPECT and assessing its value were discussed qualitatively; more detailed studies based on this theory and the two imaging systems described in Section II will be published separately.

Two important issues that were not treated here are the use of nonlinear observers for detection or estimation and the problem of assessing and optimizing systems for joint detection-estimation tasks.

Nonlinear observers, such as the Bayesian ideal observer for detection or the maximum-likelihood estimator for estimation tasks are harder to analyze than the linear observers treated here. There is no closed-form rule for performing the task and no counterpart to the linear template or estimation matrix, so the Fréchet derivative on which we relied is not applicable. Nevertheless, there are well-established methods for implementing nonlinear observers and assessing the outcome, and these methods are applicable even when different systems are used with different objects as in the adaptive paradigm proposed here. A reasonable way of evaluating adaptation rules and scout strategies for nonlinear observers, therefore, would be to select several candidate adaptation methods that exhibit good performance with linear observers and then evaluate them with nonlinear observers; the approach to performance validation outlined in Section VI-F is still applicable with nonlinear observers.

Recent work by Clarkson [61] provides a basis for extending the methodology of this paper to joint detection-estimation tasks. He extended the familiar concept of localization receiver operating characteristic (LROC) curves to estimation ROC (EROC) curves and determined the form of the optimum observer for maximizing the area under the EROC curve. For Gaussian noise assumptions, this optimum observer is the scanning Hotelling observer, now scanned over all parameters of interest and not just over spatial coordinates as in the LROC problem. The adaptive Hotelling observer derived in the present paper can thus be extended to any detection-estimation problem, and the area under the EROC curve is a relevant figure of merit for assessing scout strategies and adaptation rules.

Finally, we note that the methodology in this paper is applicable, in principle, to any imaging system, and it is also applicable to multimodality imaging.

ACKNOWLEDGMENT

The authors would like to thank K. Myers for numerous interactions on the subjects of image quality and adaptive imaging and C. Dainty and N. Devaney for discussions on adaptive optics.

REFERENCES

- [1] J. M. Beckers, "Adaptive optics for astronomy: Principles, performance, and applications," *Annu. Rev. Astron. Astrophys.*, vol. 31, pp. 13–62, 1993.
- [2] D. Enard, A. Marechal, and J. Espiard, "Progress in ground-based optical telescopes," *Rep. Prog. Phys.*, vol. 59, pp. 601–656, 1996.
- [3] R. K. Tyson, *Principles of Adaptive Optics*. Boston, MA: Academic, 1998.
- [4] F. Roddier, Ed., *Adaptive Optics in Astronomy*. Cambridge, U.K.: Cambridge Univ. Press, 1999.
- [5] O. Guyon, "Limits of adaptive optics for high-contrast imaging," *Astrophys. J.*, vol. 629, pp. 592–614, 2005.
- [6] G. Rousset, "Wavefront sensing," in *Adaptive Optics in Astronomy*, F. Roddier, Ed. Cambridge, U.K.: Cambridge Univ. Press, 1999.
- [7] R. A. Gonsalves, "Phase retrieval and diversity in adaptive optics," *Opt. Eng.*, vol. 21, no. 5, pp. 829–832, 1982.

- [8] R. G. Paxman, T. J. Schulz, and J. R. Fienup, "Joint estimation of object and aberrations using phase diversity," *J. Opt. Soc. Amer.*, vol. 7, pp. 1072–1085, 1992.
- [9] R. A. Muller and A. Buffington, "Real-time correction of atmospherically degraded telescope images through image sharpening," *J. Opt. Soc. Amer.*, vol. 64, no. 9, pp. 1200–1210, 1974.
- [10] D. H. Kim, K. Kolesnikov, A. A. Kostrzewski, G. D. Savant, A. A. Vasiliev, and M. A. Vorontsov, "Adaptive imaging system using image quality metric based on statistical analysis of speckle fields," in *Proc. SPIE*, D. P. Casasent and A. G. Tescher, Eds., Jul. 2000, vol. 4044, pp. 177–186.
- [11] G. C. Ng, P. D. Freiburger, W. F. Walker, and G. E. Trahey, "A speckle target adaptive imaging technique in the presence of distributed aberrations," *IEEE Trans. Ultrason. Ferroelect. Freq. Control*, vol. 44, no. 1, pp. 140–151, Jan. 1997.
- [12] R. C. Gauss, G. E. Trahey, and M. S. Soo, "Adaptive imaging in the breast," in *Proc. Int. Symp. IEEE Ultrason. Symp.*, Caesars Tahoe, NV, 1999, pp. 1563–1569.
- [13] A. T. Fernandez, J. J. Dahl, K. Gammelmark, D. M. Dumont, and G. E. Trahey, "High resolution ultrasound beamforming using synthetic and adaptive imaging techniques," in *2002 IEEE Int. Symp. Biomed. Imag.*, Washington, DC, 2002, pp. 433–436.
- [14] P. C. Li and M. L. Li, "Adaptive imaging using the generalized coherence factor," *IEEE Trans. Ultrason. Ferroelect. Freq. Control*, vol. 50, no. 2, pp. 128–141, Feb. 2003.
- [15] S. W. Huang and P. C. Li, "Computed tomography sound velocity reconstruction using incomplete data," *IEEE Trans. Ultrason. Ferroelect. Freq. Control*, vol. 51, no. 3, pp. 329–342, Mar. 2004.
- [16] M. Claudon, F. Tranquart, D. H. Evans, F. Lefèvre, and J. M. Correas, "Advances in ultrasound," *Eur. Radiol.*, vol. 12, pp. 7–18, 2002.
- [17] M. L. Li, S. W. Huang, K. Ustiner, and P. C. Li, "Adaptive imaging using an optimal receive aperture size," *Ultrason. Imag.*, vol. 27, no. 2, pp. 111–127, Apr. 2005.
- [18] Y. Cao and D. N. Levin, "Feature-guided acquisition and reconstruction of MR images," in *Proceedings of the International Conference on Information Processing in Medical Imaging*, H. H. Barrett and A. F. Gmitro, Eds. New York: Springer, 1993, vol. 687, Lecture Notes in Computer Science, pp. 278–292, Springer.
- [19] G. P. Zientara, L. P. Panych, and F. A. Jolesz, "Dynamically adaptive MRI with encoding by singular value decomposition," *Magn. Reson. Med.*, vol. 32, no. 2, pp. 268–274, 1994.
- [20] G. P. Zientara, "Fast Imaging Techniques for Interventional MRI," in *Interventional MR*, I. Young and F. A. Jolesz, Eds. London, U.K.: Martin Dunitz, 1995, ch. 2, pp. 25–52.
- [21] Y. Cao and D. N. Levin, "Using prior knowledge of human anatomy to constrain MR image acquisition and reconstruction: Half k-space and full k-space techniques," *Magn. Reson. Imag.*, vol. 15, no. 6, pp. 669–677, 1997.
- [22] G. P. Zientara, L. P. Panych, and F. A. Jolesz, "Applicability and efficiency of near-optimal spatial encoding for dynamically adaptive MRI," *Magn. Reson. Med.*, vol. 39, no. 2, pp. 204–213, Feb. 1998.
- [23] M. Wendt, "Dynamic tracking in interventional MRI using wavelet-encoded gradient-echo sequences," *IEEE Trans. Med. Imag.*, vol. 17, no. 10, pp. 803–809, Oct. 1998.
- [24] S. Yoo, C. R. G. Guttman, L. Zhao, and L. P. Panych, "Real-time adaptive functional MRI," *NeuroImage*, vol. 10, pp. 596–606, 1999.
- [25] N. A. Ablitt, J. X. Gao, J. Keegan, D. N. Firmin, L. Stegger, and G. Z. Yang, "Predictive motion modelling for adaptive imaging," *Proc. Int. Soc. Mag. Reson. Med.*, vol. 11, p. 1570, 2003.
- [26] J. A. Griffiths, M. G. Metaxas, G. J. Royle, C. Venanzi, C. Esbrand, P. F. van der Stelt, H. Verheij, G. Li, R. Turchetta, A. Fant, P. Gasiorek, S. Theodoridis, H. Georgiou, D. Cavouras, G. Hall, M. Noy, J. Jones, J. Leaver, D. Machin, S. Greenwood, M. Khaleeq, H. Schulerud, J. Ostby, F. Triantis, A. Asimidis, D. Bolanakis, N. Manthos, R. Longo, A. Bergamaschi, and R. D. Speller, "A multi-element detector system for intelligent imaging: I-IMAS," in *Conf. Rec. IEEE Med. Imag. Conf.*, Nov. 2006, pp. 2554–2558.
- [27] A. Fant, P. Gasiorek, R. Turchetta, B. Avset, A. Bergamaschi, D. Cavouras, I. Evangelou, M. J. French, A. Galbiati, H. Georgiou, G. Hall, G. Iles, J. Jones, R. Longo, N. Manthos, M. G. Metaxas, M. Noy, J. M. Ostby, F. Psomadellis, G. J. Royle, H. Schulerud, R. D. Speller, P. F. van der Stelt, S. Theodoridis, F. Triantis, and C. Venanzi, "I-IMAS: A 1.5D sensor for high-resolution scanning," *Nucl. Instrum. Meth. Res.*, vol. 1573, pp. 27–29, 2007.
- [28] **[Please provide location/date of conference and/or page numbers]** L. R. Furenlid, M. Freed, J. Y. Hesterman, M. A. Kupinski, E. Clarkson, and H. H. Barrett, "Adaptive imaging techniques for nuclear medicine," presented at the Soc. Nucl. Med. Annu. Meeting, 2006.
- [29] M. Freed, M. A. Kupinski, L. R. Furenlid, and H. H. Barrett, "Design of an adaptive SPECT imager," presented at the Acad. Molecular Imag. Annu. Conf., Orlando, FL, Mar. 2006.
- [30] M. Freed, M. A. Kupinski, L. R. Furenlid, M. K. Whitaker, and H. H. Barrett, "A prototype instrument for adaptive SPECT imaging," *Proc. SPIE*, vol. 6510, pp. 6510–6530, 2007.
- [31] H. H. Barrett, "Objective assessment of image quality: Effects of quantum noise and object variability," *J. Opt. Soc. Amer.*, vol. 7, no. 7, pp. 1266–1278, Jul. 1990.
- [32] H. H. Barrett and K. J. Myers, *Foundations of Image Science*. New York: Wiley, 2004.
- [33] J. Y. Hesterman, M. A. Kupinski, L. R. Furenlid, D. W. Wilson, and H. H. Barrett, "The multi-module, multi-resolution system—a novel small-animal SPECT system," *Med. Phys.*, vol. 34, pp. 983–987, 2007.
- [34] **[Please provide volume, issue and page numbers]** J. Y. Hesterman, M. A. Kupinski, E. Clarkson, and H. H. Barrett, "Hardware assessment using the multi-module, multi-resolution system (M3R). A signal-detection study," *Med. Phys.*, 2007.
- [35] J. Hesterman, "Ph.D. dissertation, Univ. Arizona, Tucson, 2007 [**AUTHOR: Title?**].
- [36] H. H. Barrett, K. J. Myers, N. Devaney, and C. Dainty, "Objective assessment of image quality. IV. Application to adaptive optics," *J. Opt. Soc. Amer.*, vol. 23, pp. 3080–3105, 2006.
- [37] S. R. Meikle, P. Kench, M. Kassiou, and R. B. Banati, "Small animal SPECT and its place in the matrix of molecular imaging technologies," *Phys. Med. Biol.*, vol. 50, pp. R45–R61, 2005.
- [38] M. A. Kupinski and H. H. Barrett, Eds., *Small-Animal SPECT Imaging*. New York: Springer, 2005.
- [39] N. Schramm, G. Ebel, U. Engeland, T. Schurrat, M. Behe, and T. Behr, "High-resolution SPECT using multipinhole collimation," *IEEE Trans. Nucl. Sci.*, vol. 50, no. 3, pp. 315–320, Jun. 2003.
- [40] F. J. Beekman, F. van der Have, B. Vastenhouw, A. J. A. van der Linden, P. P. van Rijk, J. P. H. Burbach, and M. P. Smidt, "U-SPECT-I: A novel system for submillimeter-resolution tomography with radiolabeled molecules in mice," *J. Nucl. Med.*, vol. 46, pp. 1194–1200, 2005.
- [41] T. Funk, P. Depres, W. C. Barber, K. S. Shah, and B. H. Hasegawa, "A multipinhole small animal spect system with submillimeter spatial resolution," *Med. Phys.*, vol. 33, no. 5, pp. 1259–1268, 2006.
- [42] R. Zimmerman, S. Moore, and A. Mahmood, "Performance of a triple-detector, multiple-pinhole SPECT system with iodine and indium isotopes," in *IEEE Nucl. Sci. Symp. Conf. Rec. 2004 IEEE*, Rome, Italy, 2004, vol. 4, pp. 2427–2429.
- [43] H. Kim, L. R. Furenlid, M. J. Crawford, D. W. Wilson, H. B. Barber, T. E. Peterson, W. C. J. Hunter, Z. Liu, J. M. Woolfenden, and H. H. Barrett, "SemiSPECT: A small-animal SPECT imager based on eight CZT detector array," *Med. Phys.*, vol. 33, pp. 465–474, Feb. 2006.
- [44] L. R. Furenlid, D. W. Wilson, C. Yi-chun, K. Hyunki, P. J. Pietraski, M. J. Crawford, and H. H. Barrett, "FastSPECT II: A second-generation high-resolution dynamic SPECT imager," *IEEE Trans. Nucl. Sci.*, vol. 51, no. 3, pp. 631–635, Jun. 2004.
- [45] Y. Chen, "System calibration and image reconstruction for a new small-animal SPECT system," Ph.D. dissertation, Univ. Arizona, Tucson, 2006.
- [46] W. E. Smith and H. H. Barrett, "Hotelling trace criterion as a figure of merit for the optimization of imaging systems," *J. Opt. Soc. Amer.*, vol. 3, pp. 717–725, 1986.
- [47] R. D. Fiete, H. H. Barrett, W. E. Smith, and K. J. Myers, "Hotelling trace criterion and its correlation with human-observer performance," *J. Opt. Soc. Amer.*, vol. 4, pp. 945–953, 1987.
- [48] H. Hotelling, "The generalization of student's ratio," *Ann. Math. Stat.*, vol. 2, pp. 360–378, 1931.
- [49] M. A. Kupinski, E. Clarkson, J. Hoppin, L. Chen, and H. H. Barrett, "Experimental determination of object statistics from noisy images," *J. Opt. Soc. Amer.*, vol. 20, pp. 421–429, 2003.
- [50] H. H. Barrett and G. D. DeMeester, "Quantum noise in fresnel zone plate imaging," *Appl. Opt.*, vol. 13, no. 5, pp. 1100–1109, 1974.
- [51] H. H. Barrett and F. A. Horrigan, "Fresnel zone plate imaging of gamma rays: Theory," *Appl. Opt.*, vol. 12, no. 11, pp. 2686–2702, 1973.

- [52] H. H. Barrett and W. Swindell, *Radiological Imaging*. New York: Academic, 1981.
- [53] K. J. Myers, J. P. Rolland, H. H. Barrett, and R. F. Wagner, "Aperture optimization for emission imaging: Effect of a spatially varying background," *J. Opt. Soc. Am. A.*, vol. 7, no. 7, pp. 1279–1293, 1990.
- [54] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies," *IEEE Trans. Inform. Theor.*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
- [55] D. L. Donoho, I. M. Johnstone, J. C. Hoch, and A. S. Stern, "Maximum entropy and the nearly black object," *J. Roy. Stat. Soc.*, vol. B54, pp. 41–81, 1992.
- [56] H. H. Barrett, C. Abbey, B. D. Gallas, and M. Eckstein, "Stabilized estimates of hotelling-observer detection performance in patient-structured noise," in *Proc. SPIE, Med. Imag. 1998: Image Perception*, 1998, vol. 3340, pp. 27–43.
- [57] B. D. Gallas and H. H. Barrett, "Validating the use of channels to estimate the ideal linear observer," *J. Opt. Soc. Amer.*, vol. 20, pp. 1725–1739, 2003.
- [58] D. J. Tylavsky and G. R. Sohie, "Generalization of the matrix inversion lemma," *Proc. IEEE*, vol. 74, no. 7, pp. 1050–1052, Jul. 1986.
- [59] Y. C. Chen, D. W. Wilson, L. R. Furenlid, and H. H. Barrett, "Calibration of scintillation cameras and pinhole SPECT imaging systems," in *Small-Animal SPECT Imag.*, M. Kupinski and H. Barrett, Eds. New York: Springer, 2005, ch. 12, pp. 195–201.
- [60] F. Shen and E. Clarkson, "Using fisher information to approximate ideal-observer performance on detection tasks for lumpy-background images," *J. Opt. Soc. Amer.*, vol. 23, pp. 2406–2414, 2006.
- [61] E. Clarkson, "Estimation ROC curves and their corresponding ideal observers," in *Proc. SPIE*, 2007, vol. 6515, p. 651504.

IEEE
PROOF